

罗非鱼源无乳链球菌(*Streptococcus agalactiae*) Sip 基因的克隆、鉴定及分子特性分析*

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

(1. 四川农业大学鱼病研究中心 雅安 625014; 2. 动物疫病与人类健康四川省重点实验室
四川农业大学 雅安 625014; 3. 通威股份有限公司 成都 610041)

提要 利用设计的特异性引物, 扩增出了分离自患病罗非鱼无乳链球菌强毒株 Sip 基因, 并将其克隆到 pMD19-T 载体上, 之后对重组质粒进行了 PCR 和双酶切(*Bam*H + *Hind*) 鉴定, 并利用多种生物信息学软件对 Sip 基因进行分子特性分析。结果显示, 罗非鱼源无乳链球菌 Sip 编码氨基酸序列具有极高保守性, 与人源、哺乳动物源无乳链球菌亲缘性达 100%, 存在 1 个由 25 个氨基酸组成的信号肽, 具有 1 个与免疫调节功能相关的 LysM 结构域, 具有与蛋白翻译后修饰功能相关的磷酸化位点 33 个和 N-糖基化位点 2 个, 编码多肽链中疏水区大于亲水区, 是一种膜外蛋白。密码子偏爱性分析表明, 罗非鱼源无乳链球菌 Sip 基因密码子使用频率差异较大, 密码子偏爱性与真核生物较为接近。

关键词 无乳链球菌, Sip 基因, 克隆, 分子特性

中图分类号 Q7

无乳链球菌(*Streptococcus agalactiae*), 也称 B 组链球菌(group B streptococci, GBS), 是一种重要的人、畜、鱼共患的病原菌(邓江红等, 2005; Trigo *et al.*, 2008; Suanyuk *et al.*, 2008; 祝璟琳等, 2010)。该菌表面含有两类特定的糖抗原, 即聚合糖类和荚膜多糖, 而这两类抗原物质都具有型特异性(Vidová *et al.*, 2009)。Brodeur 等(2000)对人源无乳链球菌基因文库进行免疫学筛选时, 发现 9 种血清型的代表菌株都存在一种非荚膜表面蛋白, 并命名为表面免疫原性蛋白(Surface Immunogenic Protein, Sip), 该蛋白对受不同血清型的无乳链球菌的致死剂量攻击的老鼠具有极强的交叉免疫保护作用, 因此被列为预防无乳链球菌感染候选疫苗的重要研究对象。Vidová 等(2009)通过多重 PCR 对 147 株牛源无乳链球菌 Sip 基因进行鉴定, 发现其中 4 株的 Sip 基因的推导氨基酸序列

中出现了 LysM 序列的缺失, 其编码蛋白对老鼠的免疫保护力也显著降低。可见不同分离来源的无乳链球菌 Sip 基因序列存在功能基因缺失的可能, 并将对其免疫保护作用产生显著影响。因此, 对 Sip 编码基因进行系统的生物信息学分析, 可为不同来源的无乳链球菌 Sip 基因分子特性和相关功能的比较分析提供重要理论支撑。目前, 有关无乳链球菌 Sip 基因的分子特性分析的研究几乎处于理论空白状态。

本文中, 作者通过对从海南文昌链球菌病暴发疫区的患病罗非鱼分离到的无乳链球菌强毒株进行基因组 DNA 提取, 利用参考 GenBank 中无乳链球菌 Sip 核酸序列设计的特异性引物进行 PCR 扩增, 并将扩增产物与 pMD19-T 载体连接, 经 PCR 和双酶切鉴定以及序列的进化关系比对验证所扩增的基因即为 Sip 基因, 随后应用生物信息学分析工具对其进行了

* 教育部“长江学者和创新团队发展计划”创新团队项目, IRT0848 号; 通威股份有限公司重点资助项目, 2006—2009。黄锦炉, 博士, E-mail: hgl25@163.com

通讯作者: 汪开毓, 教授, 博导, E-mail: kywangsicau@126.com

收稿日期: 2010-07-25, 收修改稿日期: 2010-09-23

分子特性分析,旨在为进一步研究罗非鱼源无乳链球菌 Sip 的相关功能和作用机理以提供分子生物学依据。

1 材料与方法

1.1 菌株

无乳链球菌(*Streptococcus agalactiae*)分离自海南文昌水产养殖区链球菌病暴发疫区的患病罗非鱼,由本实验室分离、鉴定和保存,以下称鱼源无乳链球菌;大肠杆菌 *Escherichia coli* DH5 菌株购自天根生化科技(北京)有限公司。

1.2 主要试剂

pMD19-T(Simple)克隆载体、Premix、T₄ DNA 连接酶、限制性内切酶 *Bam*H 和 *Hind*、DNA Marker、DNA 凝胶回收试剂盒和质粒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)购自天根生化科技(北京)有限公司。

1.3 引物设计与合成

根据 GenBank 中无乳链球菌 Sip 基因序列(GenBank 登录号为 DQ650634),运用生物软件 Oligo6.0 和 primer5.0 对其上下游引物进行设计。引物由生工生物工程(上海)有限公司合成,预期扩增片段大小为 1302bp,包含完整的 Sip 基因序列。引物序列分别为:上游 5'-CGCGGATCCATGGAAATGAATAAAAGGTAC-3' 划线部分为添加的 *Bam*H 酶切位点,下游 5'-CCCAAGCTTTTAAAGGATACGTGAACGTGG-3' 划线部分为添加的 *Hind* 酶切位点,划线前端部分为保护性碱基。

1.4 Sip 基因的 PCR 扩增

挑取低温保存的无乳链球菌菌种接种于脑心浸液营养肉汤中,37℃ 恒温床 120r/min 培养 24h,5000g 离心 3min 收集菌体,并按天根细菌基因组 DNA 提取试剂盒(DP302-02)的操作说明提取菌体基因组 DNA,以 DNA 为模板,按以下参数对 Sip 基因进行 PCR 扩增:PCR 反应体系为 25μl,其中 Premix 12.5μl, DNA 模板 1.0μl,上、下游引物(20pmol)各 0.5μl,无菌超纯水 10.5μl;反应条件为 94℃ 预变性 5min,94℃ 30s,55℃ 35s,72℃ 1min 18s,共 30 个循环,最后 72℃ 延伸 10min。取 5μl 反应产物在 1% 琼脂糖凝胶进行电泳,凝胶成像系统检测结果。

1.5 Sip 基因的 T 克隆、测序及序列分析

PCR 扩增产物经 1% 琼脂糖凝胶电泳分离后,胶回收目的片段,再与 pMD19-T 载体进行连接反应,构建 pMD19-T-Sip 重组质粒,并将重组质粒转化到大

肠杆菌(*E. coli*)DH5 株,筛选阳性克隆并进行 PCR 和酶切鉴定。同时将 pMD19-T-Sip 重组质粒送生工生物工程(上海)有限公司测序。

1.6 核酸序列特征分析

利用 NCBI 的 ORF Finder 工具进行阅读框架分析,再利用 NCBI 的 BLASTN 工具进行序列的相似性搜索。

1.7 蛋白质序列特征分析

应用瑞士蛋白质专家网在线翻译工具将鱼源无乳链球菌 Sip 基因序列翻译成氨基酸序列,应用 NCBI Conserved Domains 查找工具分析蛋白质结构域;通过 NCBI 的 BLASTP 程序进行氨基酸序列同源性检索,利用 Clustal X 2.0 软件和 MEGA 4.1 软件构建进化树;应用 Bioedit7.0 软件进行疏水性分析,应用 TMHMM 进行跨膜区进行预测(Moller *et al.*, 2001), NetPhos2.0 及 NetNGlyc1.0 在线程序分析磷酸化位点和糖基化位点(Blom *et al.*, 1999; Chen *et al.*, 2008), SignalP 3.0 Server 在线搜索查找信号肽序列(Bendtsen *et al.*, 2004); PSIPred 程序进行蛋白质二级结构预测(Jones, 1999), SAM_T08 自动蛋白同源建模数据库进行 3D 结构预测(Karplus, 2009),最后在 <http://www.biotech.ou.edu/> 进行鱼源无乳链球菌 Sip 重组蛋白溶解度预测。

1.8 密码子偏爱性分析

应用 EMBOSS 生物信息学在线分析工具的 CUSP 程序计算鱼源无乳链球菌 Sip 基因中不同密码子在同一个氨基酸编码中所占的比例,并通过外延法计算密码子在编码基因中出现的频率。将结果与大肠杆菌、酵母和人的密码子偏爱性进行比较(Nakamura *et al.*, 2000)。

2 结果

2.1 无乳链球菌 SIP 基因的扩增和 T 克隆鉴定

以鱼源无乳链球菌基因组 DNA 为模板,根据所设计的 PCR 引物,经高保真 DNA 聚合酶特异性的扩增出一条位于 1000—1500bp 区间的目的片段(图 1a)。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-Sip,重组质粒经 *Bam*H /*Hind* 以双酶切和 PCR 鉴定正确(图 1b)。

2.2 核苷酸序列特征分析

对重组质粒 pMD19-T-Sip 测序结果表明,鱼源无乳链球菌 Sip 基因是一个由 1302 个碱基组成的完整开放性阅读框(ORF),编码一个长度为 433 个氨基酸

的蛋白。全序列中 GC 含量为 41%。序列已提交至 GenBank, 登录号为 HQ878436。

2.3 蛋白质序列特征分析

瑞士蛋白质分析专家系统 protparam 工具预测表明, 鱼源无乳链球菌 Sip ORF 编码一条由 433 个氨基酸残基组成的多肽, 相对分子量为 45.6kDa, 等电点理论值为 6.46, 分子式为: C₁₉₈₁H₃₁₉₂N₅₄₄O₆₅₈S₁₀。SIP 编码蛋白的氨基酸组成成分分析表明(表 1), 蛋白质中含量较高的氨基酸为 Ala(14.3%)、Thr(11.8%)和 Val(11.8%), 含量较少的氨基酸为 Trp(0.7%)、Phe

(0.9%)、Arg(1.4%)和 His(1.8%), 而 Sip 氨基酸组成中则不含有 Cys、Phe 和 Sec 这三种。其中, 疏水氨基酸 201 个, 占氨基酸总数的 46.4%, 亲水性氨基酸有 154 个, 占 35.6%, 酸性氨基酸(Asp + Glu) 36 个, 占 8.3%, 碱性氨基酸(Arg + Lys) 34 个, 占 7.8%。

NCBI 的 CDD(conserved domain)工具分析发现, 鱼源无乳链球菌 Sip 氨基酸序列 N-末端的第 53 和第 95 个氨基酸残基之间含有 1 个溶解素(LysM)超家族的保守结构域。经一次迭代分析, 该结构域序列与链球菌属的其它成员菌, 如酿脓链球菌(*S. pyogenes*)、唾液链球菌(*S. salivarius*)、前庭链球菌(*S. vestibularis*)和血链球菌(*S. sanguinis*)具有极高相似性, 其中与编码部分表面蛋白的酿脓链球菌 LysM 结构域氨基酸序列相似性最高。

利用 Clustal X 2.0 软件对 17 个相关菌株 Sip 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 2)。可见, 鱼源无乳链球菌 Sip 编码氨基酸序列与血清型 Ia、Ia/c、Ib、以及未知血清型的无乳链球菌的亲缘关系均为 100%; 与同属链球菌属的未知血清型的唾液链球菌、前庭链球菌和口腔链球菌(*S. oralis*)等的亲

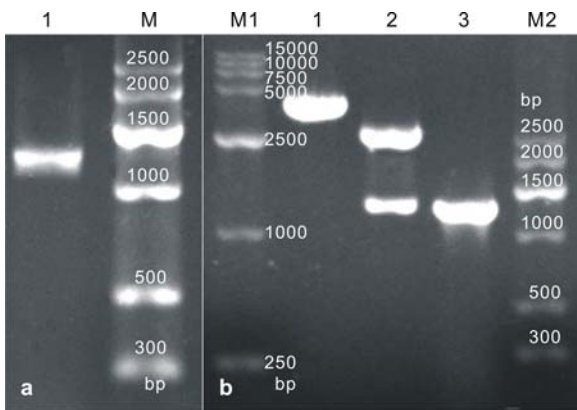


图 1 鱼源无乳链球菌 Sip 基因的 PCR 扩增结果(a)以及重组质粒 pMD19-T-Sip 的酶切和 PCR 鉴定结果(b)

Fig.1 The results of PCR amplification of the Sip gene in *S. agalactiae* (a) and restriction digestion and PCR amplification of the recombinant plasmid pMD19-T-Sip (b)

a: M. DL2500Marker; 1. Sip 基因扩增产物。b: M1. DL15000 Marker; M2. DL2500 Marker; 1. pMD19-T-Sip/*Bam*H ; 2. pMD19-T-Sip/*Bam*H +*Hind* ; 3. pMD19-T-Sip 重组质粒的 PCR 产物

表 1 鱼源无乳链球菌 Sip 基因序列推导肽链的氨基酸组分分析

Tab.1 Amino acid composition of the peptide deduced from the *S. agalactiae* Sip gene

氨基酸	数量	摩尔比(%)	氨基酸	数量	摩尔比(%)
Ala	62	14.3	Lys	28	6.5
Arg	6	1.4	Met	10	2.3
Asn	24	5.5	Phe	4	0.9
Asp	15	3.5	Pro	27	6.2
Cys	0	0	Ser	42	9.7
Gln	22	5.1	Thr	51	11.8
Glu	21	4.8	Trp	3	0.7
Gly	15	3.5	Tyr	13	3.0
His	8	1.8	Val	51	11.8
Ile	14	3.2	Phe	0	0
Ieu	17	3.9	Sec	0	0

表 2 相关菌株的 Sip 氨基酸序列

Tab.2 The amino acid sequence deduced from the Sip gene from related strains

蛋白	细菌名称	GenBank 登录号	血清型
Sip	<i>S. agalactiae</i>	AAG18474	Ia/c
Sip	<i>S. agalactiae</i>	AAG18476	
Sip	<i>S. agalactiae</i>	ABG37993	NT
Sip	<i>S. agalactiae</i>	ABK56861	
Sip	<i>S. agalactiae</i>	ABK56878	
Sip	<i>S. agalactiae</i>	ABK56880	
Sip	<i>S. agalactiae</i>	ABK56886	Ia
Sip	<i>S. agalactiae</i>	ABK56888	Ib
Sip	<i>S. agalactiae</i>	ABK56889	
Sip	<i>S. agalactiae</i>	YP328758	NT
Sip	<i>S. oralis</i>	ZP06611288	NT
Sip	<i>S. mitis</i>	ZP07463433	NT
Sip	<i>S. salivarius</i>	CAP16604	NT
Sip	<i>S. vestibularis</i>	CAP16606	NT
Sip	<i>S. thermophilus</i>	YP_140861	NT
Sip	<i>A. marinum</i>	ZP_06807629	NT
Sip	<i>A. viridans</i>	ZP_06807629	NT
Sip	<i>B. mcbrellneri</i>	ZP_06805573	NT

注: 血清型一栏中, “NT” 表示对应菌株的血清型为未知

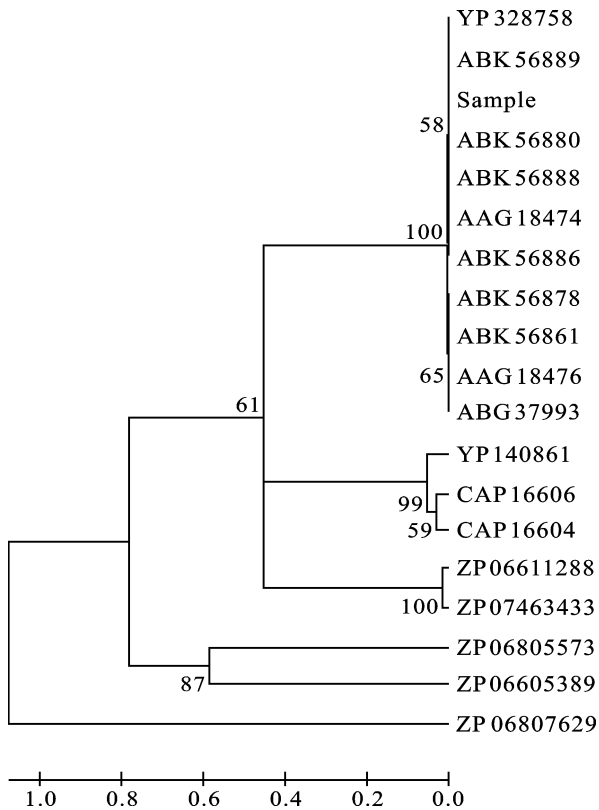


图2 鱼源无乳链球菌 Sip 与参考菌株同源基因氨基酸序列进化树分析

Fig.2 The phylogenetic tree analysis between the *S. agalactiae* Sip and its homologous reference strains

注: “Sample”代表鱼源无乳链球菌 Sip 编码氨基酸序列

缘关系较远, 而与绿浅气球菌(*Aerococcus viridans*)和苛养颗粒链菌(*Granulicatella elegans*)等的亲缘关系最远。

亲/疏水性预测分析表明, 鱼源无乳链球菌 Sip 多肽链有 8 个主要的疏水区, 分别位于氨基酸序列的 9—22 位、59—91 位、151—159 位、193—201 位、222—255 位、268—277 位、291—297 位和 359—368 位。整个蛋白质疏水性最大值为 1.944, 最小值为 -2.433, 然而亲水区在蛋白质多肽链占据的区域小于疏水区域。NetPhos2.0 程序分析表明, 当阈值为 0.5 时, 鱼源无乳链球菌 Sip 氨基酸序列中共有 33 个潜在的磷酸化位点, 其中包含 17 个丝氨酸(Ser)磷酸化位点、12 个苏氨酸(Thr)磷酸化位点和 4 个酪氨酸(Tyr)磷酸化位点, 而且 N-末端氨基酸磷酸化的数量明显多于 C-末端。NetNGlyc1.0 程序预测表明, 鱼源无乳链球菌 Sip 氨基酸序列中共有 2 个潜在的 N-糖基化位点, 当取阈值为 0.5 时, 第 49 位、387 位的天冬酰胺(Asn)可能被糖基化。SignalP V3.0 程序的 SignalP-NN 和

SignalP-NNSIP-HMM 两个模型预测发现, 鱼源无乳链球菌 Sip 氨基酸序列 N 末端有一个由 25 个氨基酸组成的信号肽, 信号肽切割位点位于序列的第 25 和 26 个氨基酸残基之间。跨膜预测结果显示该多肽链没有跨膜区域, 是一种膜外蛋白。重组蛋白可溶度预测结果显示, 当选择大肠杆菌(*E. coli*)为表达宿主时, 鱼源无乳链球菌 Sip 重组蛋白的不溶解度达 75.2%。

PSIpred 程序预测结果表明, 鱼源无乳链球菌 Sip 二级结构组成中, 无规则卷曲(Coil)的含量最大, 高达 48.96%, 而 α -螺旋(Alpha-Helix)和 β -折叠(Beta-Strand)含量分别为 26.08%和 24.94%。SAM_T08 自动蛋白同源建模数据库进行 3D 结构预测, 结果找到与鱼源无乳链球菌 Sip 基因较高同源性的已知 3D 结构模型, 其三维构象主要包含无规则卷曲、 α -螺旋和 β -转角(图 3)。

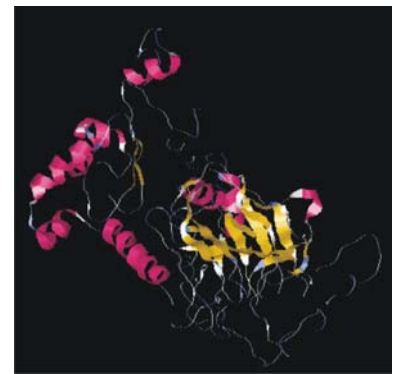


图3 鱼源无乳链球菌 Sip 三级结构预测图

Fig.3 The tertiary structure of *S. agalactiae* Sip

注: α -螺旋标记为红色, β -折叠标记为黄色, 无规则卷曲为蓝色, 其它为灰色

2.4 密码子偏爱性分析

经过 CUSP 程序分析, 鱼源无乳链球菌 Sip 基因序列中, 编码 I(Ile)、T(Thr)、A(Ala)、S(Ser)、V(Val)等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 I(Ile)的偏爱密码子为 ATT, 比例为 0.786, 而 ATA 的比例只有 0.071; T(Thr)偏爱密码子为 ACA 和 ACT, 其比例分别为 0.549 和 0.333, 而 ACC 的比例只有 0.020; A(Ala)的密码子中偏爱密码子为 GCA, 比例为 0.532, 而 GCC 和 GCG 的比例均只有 0.065; S(Ser)偏爱密码子为 TCA, 比例为 0.476, 而 TCC 的比例只有 0.024; V(Val)偏爱密码子为 GTT 和 GTA, 其比例分别为 0.431 和 0.392, 而 GTG 的比例仅有 0.078。而序列中包括 TGC、TGT、CCC、CGA、CGC、CGG 等在内的 8 个密码子出现的频率则为零。而通过将鱼源无乳链球菌 Sip 的密码子与大肠杆菌、酵母及人的密码子偏性进行比较发现, 鱼源无乳链球菌 Sip 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 37 个, 酵母有 31 个, 人有 42 个。

3 讨论

3.1 Sip 基因分子特性分析的研究意义

目前,无乳链球菌已经发现有 10 种荚膜多糖血清型(Slotved *et al.*, 2007),其中 a、 和 是已知的对人类具有致病力的最常见血清型(Persson *et al.*, 2004),而 a、 b 和 这三种血清型的也已经证明能引起鱼类感染发病(Vandamme *et al.*, 1997; Suanyuk *et al.*, 2008)。无乳链球菌的毒力因子主要分布于菌体的表面,菌株表达分泌的特异性蛋白具有严格的免疫保护力,如荚膜多糖虽具有良好的抗原性,但产生的免疫保护力却具有型特异性(Kasper *et al.*, 1996);某些表面蛋白如 α 蛋白和 Rib 等物质虽也可产生具有良好的免疫原性,然而只有约 50% 的无乳链球菌能表达 α 蛋白,大部分血清型 的无乳链球菌能表达 Rib 蛋白(Ferrieri *et al.*, 1997; Stalhammar-Carlemalm *et al.*, 1993)。无乳链球菌 Sip 是一种表面蛋白,除了血清型

菌株尚未知是否存在该蛋白以外,已知的其余血清型的无乳链球菌株都存在该蛋白,而且在培养条件下的胞外产物中也可被检测到。由于已发现的不同血清型的无乳链球菌 Sip 基因具有极高的保守性,其编码蛋白也具有有良好的免疫原性,特别是它在受免小鼠抵御不同血清型的无乳链球菌以致死剂量攻击时表现出了良好的交叉保护效果,因此它的发现是人们在预防无乳链球菌感染研究方面获得的重大突破。虽然人们已经从感染罗非鱼等鱼类分离到无乳链球菌,但有关鱼源无乳链球菌 Sip 基因分子特性及编码蛋白的相关功能研究仍是空白,因此本研究的发现将为鱼源无乳链球菌 Sip 基因后续研究提供重要的理论支撑。

3.2 Sip 氨基酸序列的进化关系分析

病原菌在不同遗传背景下,其特异性毒力基因可能发生变异,而这些毒力基因可能与疾病的发病机制之间存在密切的联系(Brochet *et al.*, 2006)。而无乳链球菌在感染宿主的过程中,Sip 基因也可能发生某些变异以逃避宿主的免疫防御(Vidová *et al.*, 2009)。本研究中,测序结果显示,鱼源无乳链球菌 Sip 基因序列全长 1302bp,推导的氨基酸序列含有 433 个氨基酸,其中不同氨基酸的含量差异较大。与已报道的牛源乳链球菌 Sip 氨基酸序列长度相同,而在 C-末端却比首次发现的人源无乳链球菌 Sip 氨基酸序列少 1 个 Lys(K)。进化关系分析表明,鱼源无乳链球菌 Sip 氨基酸序列与人源、牛源不同血清型无乳链球菌 Sip 的

亲缘关系达到 100%,这说明鱼源无乳链球菌 Sip 氨基酸序列仍具有极高的保守性。

3.3 Sip 氨基酸序列的分子特性分析

LysM 结构域是一种常见的蛋白质功能模块,能与肽聚糖特异性结合,它的结合位点已被发现存在于多种不同细菌且类型不同的肽聚糖中(Desvaux *et al.*, 2006)。许多含有 LysM 结构域的蛋白,如葡萄球菌 IgG 结合蛋白和大肠杆菌黏附素,常常与细菌的致病机制有关(Bateman *et al.*, 2000),此外,含有该结构域的某些蛋白也具有重要功能,如刺激宿主的免疫应答等。这表明,LysM 结构域对某些蛋白生物学功能的发挥具有重要的影响。Vidová 等(2009)研究牛源无乳链球菌 Sip 重组蛋白对小鼠的免疫保护力时发现,与含有 LysM 结构域无乳链球菌 Sip 相比,缺失了 LysM 结构域的 Sip 对受致死剂量的无乳链球菌攻击的小鼠的免疫保护力也显著降低。由此可见,LysM 结构域在无乳链球菌 Sip 发挥生物学功能的过程中也具有极其重要的作用。而保守结构域预测结果显示,鱼源无乳链球菌 Sip 氨基酸序列含有 LysM 保守结构域,这个发现对作者接下来研究鱼源无乳链球菌 Sip 对罗非鱼的免疫保护和毒性评估方面具有重要的理论参考价值。

蛋白翻译后修饰(Post-translational modifications, PTMs)(如磷酸化和糖基化等)是一种在细胞内普遍存在的共价结合过程,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从而改变蛋白的特性(Mann *et al.*, 2003)。蛋白的丝氨酸/苏氨酸/酪氨酸(Ser/Thr/Tyr)磷酸化通常是调节蛋白翻译后修饰的重要环节;而糖基化是一个复杂的过程,共有 13 种不同种类的单糖与 8 种类型的氨基酸残基参与(Spiro, 2002),某些糖基化位点的缺失将可能会影响蛋白的正确折叠(Shental-Bechor *et al.*, 2008)。鱼源无乳链球菌 Sip 氨基酸序列中存在 33 个的磷酸化位点和 2 个糖基化位点,这些潜在的位点可能在蛋白翻译后得到一定程度的修饰,并且可能在 Sip 发挥其生物学功能中起着重要调控作用。疏水性是每种氨基酸所固有的特性,它是决定一个蛋白质最终折叠成的三维构象的一个重要因素(曹海燕等, 2010),从鱼源无乳链球菌 Sip 肽链的亲/疏水区分布情况分析,亲水区明显少于疏水区,表现出一定的疏水性。当选择大肠杆菌作为表达宿主时,鱼源无乳链球菌 Sip 基因重组蛋白不溶解度理论预测值为 75.2%,这说明在建构表达质粒时,在 Sip 基因两端选择加入合适的融合蛋白

标签显得尤为重要,同时也有利于提高重组蛋白的可溶解度和重组蛋白的纯化水平。

信号肽是引导翻译出的前体蛋白通过细胞膜分泌到细胞外的一段序列,经信号肽酶切除信号肽序列后的前体蛋白才可能成为具有正常功能的成熟蛋白(Nielsen *et al*, 1998)。不同生物的信号肽长度差异大,且序列没有保守性,但几乎所有的信号肽都有一个共同的三区结构:即带正电荷的 n 区,疏水性的 h 区以及切割位点所在的 c 区(Nakai, 2000)。信号肽预测结果表明,鱼源无乳链球菌 Sip 信号肽由 25 个氨基酸组成,其 n 区从起始 Met 开始到第 5 个氨基酸,其氨基酸片段为 MEMNK,带正电荷;c 区位于第 22 位氨基酸和 25 位氨基酸之间,其片段为 SVQA,不带电荷,第 25 位和 26 位氨基酸之间为切割位点;h 区则位于 n 区和 c 区之间,为疏水性氨基酸组成的疏水性区域。这说明鱼源无乳链球菌 Sip 前体蛋白经信号肽引导到达胞外并经信号肽酶切割后才能折叠形成成熟蛋白。以上发现与鱼源无乳链球菌 Sip 为一种膜外蛋白的预测结果也是相吻合。

3.4 Sip 基因的密码子偏爱性分析

关于鱼源无乳链球菌 Sip 在密码子使用上的偏爱性问题,目前国内外未有报道。鱼源无乳链球菌 Sip 编码基因的密码子偏爱性分析表明,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, Sip 的密码子与大肠杆菌和人的相差较大,其表达系统选择在真核系统(如酵母)较为合适。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、是否是毒素、是否需糖基化或其它修饰、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之一,通过密码子偏爱性分析不仅可为其基因表达选择合适的表达系统,同时也可以为改造密码子从而提高基因的表达量提供依据。

参 考 文 献

- 邓江红, 杨永弘, 2005. B 组链球菌的分子生物学诊断和基因分型研究进展. 中华儿科杂志, 43(11): 832—835
- 祝璟琳, 杨 弘, 邹芝英等, 2010. 海南养殖罗非鱼(*Oreochromis niloticus*)致病链球菌的分离、鉴定及其药敏试验. 海洋与湖沼, 41(4): 590—596
- 曹海燕, 王 丹, 张耀洲, 2010. 家蚕 BmLITAF 基因的克隆表达及生物信息学分析. 浙江理工大学学报, 27(1): 149—154
- Bateman A, Bycroft M, 2000. The structure of a LysM domain from *E. coli* membrane-bound lytic murein transglycosylase D (MltD). J Mol Biol, 299(4): 1113—1119
- Bendtsen J D, Nielsen H, von Heijne G *et al*, 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol, 340(4): 783—795
- Blom N, Gammeltoft S, Brunak S, 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J Mol Biol, 294(5): 1351—1362
- Brochet M, Couve E, Zouine M *et al*, 2006. Genomic diversity and evolution within the species *Streptococcus agalactiae*. Microbes Infect, 8(5): 1227—1243
- Brodeur B R, Boyer M, Charlebois I *et al*, 2000. Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. Infect Immun, 68(10): 5610—5618
- Chen Y Z, Tang Y R, Sheng Z Y *et al*, 2008. Prediction of mucin-type O-glycosylation sites in mammalian proteins using the composition of k-spaced amino acid pairs. BMC Bioinformatics, 9(101): 1—12
- Desvaux M, Dumas E, Chafsey I *et al*, 2006. Protein cell surface display in Gram-positive bacteria: from single protein to macromolecular protein structure. FEMS Microbiol Lett, 256(1): 1—15
- Ferrieri P, Flores A E, 1997. Surface protein expression in group B streptococcal invasive isolates. Adv Exp Med Biol, 418: 635—637
- Jones D T, 1999. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol, 292(2): 195—202
- Karplus K, 2009. SAM-T08, HMM-based protein structure prediction. Nucleic Acids Res, 37: 492—497
- Kasper D L, Paoletti L C, Wessels M R *et al*, 1996. Immune response to type III group B streptococcal polysaccharide-tetanus toxoid conjugate vaccine. J Clin Invest, 98(10): 2308—2314
- Mann M, Jensen O N, 2003. Proteomic analysis of post-translational modifications. Nat Biotechnol, 21(3): 255—261
- Moller S, Croning M D, Apweiler R, 2001. Evaluation of methods for the prediction of membrane spanning region. Bioinformatics, 17(7): 646—653
- Nakai K, 2000. Protein sorting signals and prediction of subcellular localization. Adv Protein Chem, 54: 277—344
- Nakamura Y, Gojobori T, Ikenura T, 2000. Condon usage tabulate form the international DNA sequence databases: status for the year 2000. Nucl Acids Res, 28(1): 292
- Nielsen H, Krogh A, 1998. Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol, 6: 122—130
- Persson E, Berg S, Trollfors B *et al*, 2004. Serotypes and clinical manifestations of invasive group B streptococcal infections in western Sweden 1998—2001. Clin Microbiol Infect, 10(9): 791—796
- Shental-Bechor D, Levy Y, 2008. Effect of glycosylation on pro-

- tein folding: a close look at thermodynamic stabilization. *Proc Natl Acad Sci U S A*, 105(24): 8256—8261
- Slotved H C, Kong F, Lambertsen L *et al*, 2007. Serotype IX, a proposed new *Streptococcus agalactiae* serotype. *J Clin Microbiol*, 45(9): 2929—2936
- Spiro R G, 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*, 12(4): 43—56
- Stalhammar-Carlemalm M, Stenberg L, Lindahl G, 1993. Protein rib: a novel group B streptococcal cell surface protein that confers protective immunity and is expressed by most strains causing invasive infections. *J Exp Med*, 177(6): 1593—1603
- Suanyuk N, Kong F, Ko D *et al*, 2008. Occurrence of rare genotypes of *Streptococcus agalactiae* in cultured red tilapia *Oreochromis* sp. and Nile tilapia *O. niloticus* in Thailand-Relationship to human isolates. *Aquaculture*, 284: 35—40
- Trigo G, Ferreira P, Ribeiro N *et al*, 2008. Identification of immunoreactive extracellular proteins of *Streptococcus agalactiae* in bovine mastitis. *Can J Microbiol*, 54(11): 899—905
- Vandamme P, Devriese L A, Pot B *et al*, 1997. *Streptococcus difficile* is a nonhemolytic group B, type Ib *Streptococcus*. *Int J Syst Bacteriol*, 47(1): 81—85
- Vidová B, Chotár M, Godány A, 2009. N-terminal anchor in surface immunogenic protein of *Streptococcus agalactiae* and its influence on immunity elicitation. *Folia Microbiologica*, 52(2): 161—166

CHARACTERIZATION OF THE Sip GENE IN THE STRAIN *STREPTOCOCCUS AGALACTIAE* ISOLATED FROM TILAPIA

HUANG Jin-Lu^{1,2}, WANG Kai-Yu^{1,2}, XIAO Dan³, WANG Jun^{1,2},
FU Xi^{1,2}, WANG Hao-Cheng^{1,2}, LIAN Hai²

(1. Fish Disease Research Center of Fisheries Department of Sichuan Agricultural University, Ya'an, 625014;
2. Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Ya'an, 625014; 3. Tongwei Co. Ltd., Chengdu, 610041)

Abstract The Sip gene of a virulent strain of *Streptococcus agalactiae* isolated from Tilapia was amplified using the PCR method with specific primers and then cloned into pMD19-T vector. The presence of the recombinant plasmid was confirmed by the PCR method combined with the restriction enzyme digestion method (*Bam*H and *Hind*). The results showed that the amino acid sequence derived from the Sip of *S. agalactiae* in Tilapia was highly conserved and had a 100% of homology among strains isolated from humans and other mammals. The polypeptide contained a signal peptide consisting of 25 amino acids and a LysM functional domain which was related to immunoregulation. The polypeptide had a number of important sites related to post-translational modification, including 33 phosphorylation sites and 2 N-glycosylation sites. The hydrophobic regions of the Sip were larger than the hydrophilic regions and the Sip was predicted to be located outside of the cell membrane. An analysis of codon bias demonstrated that the codon usage frequency of Sip in the strain *S. agalactiae* was distinctly different and it resembles more closely to those of the eukaryotes.

Key words *Streptococcus agalactiae*, Sip gene, Clone, Molecular characteristics