

类植物乳杆菌 L-XM1 *plnBCD* 基因的生物信息学分析

张香美^{1,2} 李平兰^{2*}

(1. 河北经贸大学 生物科学与工程学院, 石家庄 050061; 2. 中国农业大学 食品科学与营养工程学院, 北京 100083)

摘 要 为了解类植物乳杆菌 L-XM1 *plnBCD* 编码蛋白的生物学特性, 并推测其功能, 综合运用 RPS-Blast、InterProscan 和 TMHMM Server 等多种生物信息学软件预测分析 *plnBCD* 编码蛋白的理化性质、疏水性、磷酸化位点、保守性功能域及二级结构, 并用 Swiss-model 同源建模。结果表明: *plnB* 编码蛋白为疏水性蛋白, 定位于细胞质膜上, 有 6 个跨膜区、21 个磷酸化位点和 1 个 HATPase_c 结构域, 具有大多数组氨酸蛋白激酶的基本性质。其二级结构以 α -螺旋和延伸链为主要构件, 其空间结构与组氨酸蛋白激酶的结构特征一致。 *plnC* 及 *plnD* 编码蛋白均为亲水性蛋白, 不具有跨膜区, 定位于细胞质中, 均具有 1 个 REC 和 1 个 LytTR 结构域。 *plnC* 及 *plnD* 编码蛋白与感应调节蛋白的空间结构具有较高的相似度。推测菌株 L-XM1 *plnBCD* 基因编码蛋白为群体感应调控系统组分。对 *plnBCD* 基因的生物信息学分析的研究可为发现新的群体感应系统和深入研究群体感应系统各组分的功能提供参考。

关键词 类植物乳杆菌; 生物信息学分析; 同源建模; 群体感应系统

中图分类号 Q 936

文章编号 1007-4333(2014)03-0175-05

文献标志码 A

Bioinformatics analysis of the *plnBCD* gene from *Lactobacillus paraplantarum* L-XM1

ZHANG Xiang-mei^{1,2}, LI Ping-lan^{2*}

(1. College of Biological Science and Engineering, Hebei University of Economics & Business, Shijiazhuang 050061, China;

2. College of Food Science and Nutritional Engineering, China Agricultural University, Beijing 100083, China)

Abstract Bioinformatics softwares were used to predict and analyse the biological properties of *plnBCD* encoded proteins in *Lactobacillus paraplantarum* L-XM1, the functions of *PlnBCD* were also predicted. The physical and chemical properties, hydrophobicity, phosphorylation sites, conserved domain, secondary structure information of the *plnBCD* encoded proteins were predicted using RPS-Blast, InterProscan, TMHMM Server, etc. . Three-dimensional structure models were constructed using Swiss-Model homology modeling. The results showed that *plnB* encoded protein was a hydrophobic protein, and located in the plasma membrane, with six transmembrane regions and 21 phosphorylation sites, one HATPase_c domain, which was consistent with the basic nature of most histidine protein kinase. α -helix and extended chains were the main secondary structure elements of *plnB* encoded protein. The spatial model of this protein was consistent with the histidine protein kinase. Both *plnC* and *plnD* encoded proteins were hydrophilic proteins, without transmembrane regions, located in the cytoplasm, and both were found with a REC domain and a LytTR domain. Both spatial models of *plnC* and *plnD* encoded proteins were similar to the response regulatory protein. Therefore, we predict that proteins encoded by *plnBCD* of strain L-XM1 are components of quorum sensing system. The bioinformatics analysis of *plnBCD* provide a reference for finding a new quorum sensing system and a deep understanding of the functions of quorum-sensing components.

Key words *Lactobacillus paraplantarum*; bioinformatics analysis; homology modeling; quorum-sensing system

收稿日期: 2013-09-25

基金项目: 国家自然科学基金资助项目(31071591)

第一作者: 张香美, 副教授, 博士, 主要从事食品微生物、食品生物技术等研究, E-mail: zxm_bio@sina.com

通讯作者: 李平兰, 教授, 主要从事食品微生物、食品生物技术研究, E-mail: lipinglan420@126.com

Lactobacillus paraplantarum L-XM1 是分离自卡萨莫迪娜那不勒斯风味萨拉米香肠的一株Ⅱb类细菌素产生菌。乳杆菌Ⅱ类细菌素的合成通常由群体感应系统调控,目前已知的主要有 *plnABCD*^[1]、*plnNC8IF-HK-D*^[1]、*orf4-sapK-sapR*^[2] 和 *orfY-sppK-sppR*^[3] 等系统。经 PCR 扩增,从 *L. paraplantarum* L-XM1 中克隆到了群体感应系统编码基因的同源基因 *plnBCD*,了解 *plnBCD* 编码蛋白的分子特征及其结构,对群体感应调控机制的研究及运用具有重要的意义。

Man 等^[4] 用生物信息学方法研究了 *L. plantarum* KLD51.0391 的群体感应系统,Cui 等^[5] 对 *L. acidophilus* 群的群体感应系统进行了生物信息学分析,葛菁萍等^[6] 对副干酪乳杆菌的感应调节蛋白基因序列进行了分析,但尚未见对 *L. paraplantarum* 群体感应系统组分进行生物信息学分析的相关报道。本研究运用序列同源性比较及功能域分析等生物信息学手段对 *L. paraplantarum* L-XM1 的 *plnBCD* 基因进行理化性质分析预测、保守性功能域分析、二级结构预测以及空间结构同源建模,并对其功能进行预测,旨在丰富Ⅱ类细菌素乳杆菌群体感应调控系统的生物信息库,为该菌群体感应调控机制的研究奠定基础,为发现新的群体感应调控系统提供参考。

1 材料与方法

1.1 数据来源

本研究使用的类植物乳杆菌 L-XM1(中国农业大学应用微生物学实验室分离并保存) *plnBCD* 数据来源于 NCBI, GenBank accession number: JQ974915-JQ974917。

1.2 生物信息学分析

利用 NCBI 网站上的 BLASTN 软件对 *plnBCD* 进行序列同源性分析。运用 DNAMAN 软件对 *plnBCD* 核苷酸序列进行翻译,然后将编码的氨基酸序列进行同源性分析(NCBI 网站上的 BLASTP 软件)、蛋白质的分子质量和等电点预测(<http://web.expasy.org/protparam/>)、保守功能结构域检索(NCBI 网站上的 RPS-Blast 工具以及 EBI 网站上的 InterProscan 工具)、跨膜螺旋区域分析(TMHMM Server v. 2.0)、疏水性分析(<http://www.expasy.org/tools/protscale.html> (Hphob./Kyte&Doolittle))、磷酸化位点预测([\[dtu.dk/services/NetPhos\]\(http://dtu.dk/services/NetPhos\)\)、亚细胞定位\(<http://psort.hgc.jp/form.html>\)以及蛋白质二级结构预测\(\[http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html\]\(http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html\)\)。采用 Swiss-model 软件对 *L. paraplantarum* L-XM1 *plnBCD* 的保守性功能域进行空间结构建模。](http://cbs.</p></div><div data-bbox=)

2 结果与分析

2.1 *plnB* 生物信息学

组氨酸蛋白激酶是群体感应系统的重要组成部分,一方面它与群体感应信号分子结合启动群体感应调控,另一方面,它将感应调节蛋白磷酸化,引起下游应答。对 HPK 空间结构的研究,无疑将有利于更进一步解析群体感应系统各组分的间相互作用机制。

相似性比对结果发现,*L. paraplantarum* L-XM1 *plnB* 与 *L. plantarum* WCFS1 以及 C11 中的组氨酸蛋白激酶编码基因 *plnB*^[7] 同源性最高,达到 97.51%,与 *L. plantarum* J51 组氨酸蛋白激酶编码基因 *plnB*^[8] 的同源性也高达 97.29%。菌株 L-XM1 的 *plnB* 与 *L. pentosus* MP-10 的组氨酸蛋白激酶编码基因 *plnB* 同源性为 70.59%,与 *L. sakei* 中调控细菌素合成的组氨酸蛋白激酶编码基因 *sppK*^[9] 的同源性为 23.82%。

plnB 基因共编码 442 个氨基酸,其氨基酸序列与 *L. plantarum* 细菌素合成相关的组氨酸蛋白激酶(GenBank accession No. FJ809773.1)的同源性为 97%。

序列分析表明,*plnB* 编码蛋白 PlnB 含有 38 个强碱性氨基酸(K、R);45 个强酸性氨基酸(D、E)。预测分子质量为 50 454.7,等电点 pI=5.59,为一疏水、稳定蛋白,GRAVY (Grand average of hydropathicity) 值为 0.301,不稳定系数为 31.04。该蛋白具有 21 个磷酸化位点,符合大多数组氨酸蛋白激酶的基本性质。保守功能结构域检索结果显示,在 PlnB 的 C 端存在一个结合 ATP 的激酶活性功能域 HATPase_c,跨膜区分析显示,PlnB 是一个膜蛋白,有 6 个跨膜区(图 1),定位于细胞质膜上,与前人对组氨酸蛋白激酶的描述一致^[10]。

以上结果显示,PlnB 具有典型的组氨酸蛋白激酶的结构特点,故可以认为 PlnB 具有组氨酸蛋白激酶的功能。该蛋白在群体感应系统中,作为信号分子的受体,可以通过自身的磷酸化和去磷酸化,起到信号转导的作用。

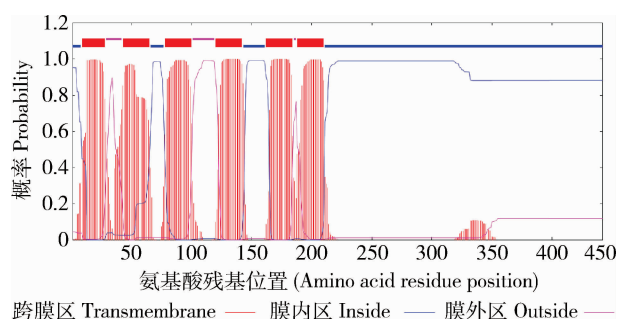


图1 PlnB 的跨膜螺旋区域分析

Fig. 1 Predictions of transmembrane helices in PlnB

二级结构预测结果表明 *L. paraplantarum* L-XM1 的 PlnB 蛋白主要由 α 螺旋、延伸链、 β 转角和无规卷曲组成。其中 α 螺旋占 64.71%、延伸链占 15.61%、 β 转角占 4.98%、无规卷曲占 14.71%。采用 SWISS-MODEL server^[11] 的自动选择模式预测 PlnB 的三维结构,程序自动选择 *Thermotoga maritima* Sensor histidine kinase (PDB code: 2c2aA) 作为模板。同源建模得到的从第 246 到第 442 位氨基酸残基的三维结构见图 2,与模板蛋白的序列相似性为 16.071,与二级结构预测结果基本一致,从而进一步验证了结构的正确性。该结构域由 5 个 α 螺旋和 5 个反向平行的 β 片层组成,符合群体感应系统组氨酸蛋白激酶的结构特征^[12-13]。该结构域有 3 个位于 C 末端的激酶催化区域的盒: H 盒、N 盒和 G 盒,围绕形成了一个结合 ATP 的口袋,位于 H 盒的 His-254 为磷酸化位点,是磷酸信号转导的中心区域,这与前人对 HPK10 亚家族保守区域的描述相一致^[14-15]。已报道的产 II 类细菌素乳酸菌群体感应系统中的组氨酸蛋白激酶组分均为 HPK10 亚家族成员^[16],故可认为 *plnB* 为群体感应系统的组氨酸蛋白激酶。

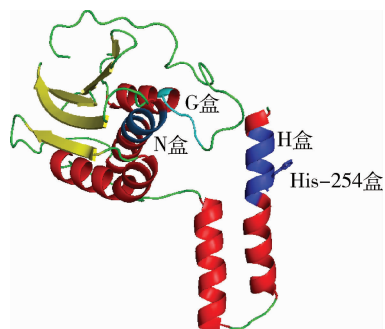


图2 PlnB 三级结构预测

Fig. 2 Tertiary structure of PlnB prediction result

2.2 *plnC* 生物信息学

感应调节蛋白可接受来自组氨酸蛋白激酶 His 残基上的磷酸基团,被磷酸化的感应调节蛋白激活子 PlnC 结合到启动子上,激活目的基因的表达。

通过比对发现,*plnC* 基因序列与 *L. plantarum* WCFS1 中编码感应调节蛋白激活子的 lp_0417 (GenBank accession No. AL935263.2) 相似度为 98.66%。*plnC* 基因共编码 247 个氨基酸,其氨基酸序列与 *L. plantarum* 中细菌素合成相关的感应调节蛋白激活子 (GenBank accession No. FJ809773.1) 的同源性为 97%。

序列分析表明,*plnC* 编码蛋白 PlnC 含有 34 个强碱性氨基酸(K、R);35 个强酸性氨基酸(D、E)。预测分子质量为 28 567.9,等电点 pI=6.21,为一亲水、稳定蛋白,GRAVY 值为 -0.319。不稳定系数为 34.69。该蛋白具有 16 个磷酸化位点。

保守功能结构域检索结果显示,PlnC 具有 2 个结构域,分别为 REC、LytTR。在 PlnC 的 N 端具有磷酸受体结构域 REC,包括 1 个激活位点,1 个磷酸化位点,1 个分子间识别位点和二聚体接触面,预测为群体感应系统中的应答调控蛋白或信号传递蛋白。在其 C 端有一个 LytTR 结构域,具有结合 DNA 靶序列的功能,起到转录调节的作用^[17-18]。PlnC 不具有跨膜区,定位于细胞质中。以上结果显示,PlnC 具有典型的感应调控因子的结构特点^[19],该蛋白在群体感应系统中,可以通过自身的磷酸化和去磷酸化,起到信号转导的作用^[20-21]。

二级结构预测结果表明 *L. paraplantarum* L-XM1 的 PlnC 蛋白主要由 α 螺旋、延伸链、 β 转角和无规卷曲组成。其中 α 螺旋占 47.77%、延伸链占 25.51%、 β 转角占 5.26%、无规卷曲占 21.46%。用 SWISS-MODEL server 的自动选择模式对 PlnC 的三维结构进行预测,程序自动选择 *Klebsiella pneumoniae* mrkE 蛋白的调节结构域 (PDB code: 2qv0B) 和 *Staphylococcus aureus* AgrA LytTR 结构域 (PDB code: 3bs1A) 作为模板。同源建模得到的从第 2~131 位氨基酸残基的三维结构见图 3(a),与模板蛋白的序列相似性为 24.615%,该结构域由 5 个 α 螺旋和 5 个平行的 β 片层组成,其 Asp-59 残基是组氨酸蛋白激酶磷酸基团传递的受体位点,当磷酸化后,其构象发生变化,引起下游应答^[22-23]。同源建模得到的从第 141~240 位氨基酸残基的三维结构见图 3(b),与模板蛋白的序列相似性为 21%,

该结构由 10 个反向平行的 β 片层和 1 个 α -螺旋结构组成,便于与 DNA 分子结合。图 3 显示,根据同源建模得到的 PlnC 的三维结构具有典型的群体感应系统中感应调节蛋白组分的结构特征。

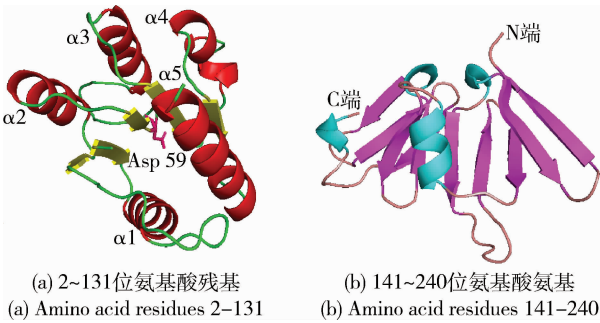


图 3 PlnC 三级结构预测

Fig. 3 Tertiary structure of PlnC prediction result

2.3 plnD 生物信息学

被磷酸化后的感应调节蛋白抑制子 PlnD 可特异性地结合到目标启动子上,抑制目的基因的表达。

登陆 NCBI 进行 BLAST 比对发现,plnD 基因序列与 *L. plantarum* WCFS1 中编码感应调节蛋白抑制子的 lp_0418 (GenBank accession No. AL935263.2) 相似度为 95.70%,与 *L. plantarum* NC8 中调控细菌素合成的 PlnD (GenBank accession No. EHS83009) 同源性为 96%。plnD 基因共编码 247 个氨基酸,其氨基酸序列与 *L. plantarum* 中细菌素合成相关的感应调节蛋白抑制子 (GenBank accession No. FJ809773.1) 的同源性为 96%,故预测其为群体感应系统重要组分。

PlnD 序列分析表明,该蛋白含有 35 个强碱性氨基酸(K,R);36 个强酸性氨基酸(D,E)。预测分子质量为 28 607.8,等电点 pI=6.35,为一亲水、稳定蛋白,GRAVY 值为-0.333,不稳定系数为 28.74,该蛋白具有 11 个磷酸化位点。

保守功能结构域检索结果显示,该蛋白具有 3 个结构域,分别为 REC、LytTR、LytT。二级结构预测结果表明 *L. paraplantarum* L-XM1 的 PlnD 蛋白主要由 α 螺旋、延伸链、 β 转角和无规卷曲组成。其中 α 螺旋占 47.37%、延伸链占 25.91%、 β 转角占 4.86%、无规卷曲占 21.86%。采用 SWISS-MODEL server 的自动选择模式对 PlnD 的三维结构进行预测,以 *Klebsiella pneumoniae* mrkE 感应调节子结构域(PBD code 为 2qv0B)和 *Staphylococcus aureus* AgrA LytTR 结构域(PBD code 为 3bs1A)

作为模板同源建模,得到从第 2~130 位氨基酸残基的三维结构见图 4(a),与模板蛋白的序列相似性为 24.031%,该结构域由 5 个 α -螺旋和 5 个平行的 β 片层组成。该结构域第 59 位的 Asp 为磷酸化位点,这与 Diep 等^[24]的研究结果一致。同源建模得到的从第 140~240 位氨基酸残基的三维结构见图 4(b),与模板蛋白的序列相似性为 26.733%,该结构由 10 个反向平行的 β 片层和 1 个 α -螺旋结构组成。以上结构特征与大多数群体感应调控系统的感应调节蛋白相同。

应用序列同源性比较及功能域分析等生物信息学手段对 plnB、plnC 和 plnD 编码蛋白进行预测分析发现,plnB 编码蛋白具有典型的组氨酸蛋白激酶的结构特征,plnC 和 plnD 编码蛋白具有典型的感应调节蛋白的结构特征,与 Straume 等^[25]的报道一致。故可以预测 plnB、plnC 和 plnD 编码蛋白是 *L. paraplantarum* L-XM1 群体感应系统的重要组分。

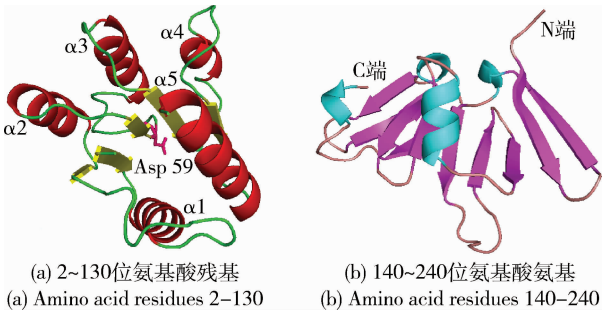


图 4 PlnD 三级结构预测

Fig. 4 Tertiary structure of PlnD prediction result

3 讨论

群体感应是调控细菌素合成的重要系统,了解群体感应系统重要组分的结构及与之相关的辅助信息,可为继续研究 PlnBCD 蛋白的结构与功能奠定基础,这对于解析群体感应调控机制,以及更好地调控细菌素合成具有重要的作用。

在对生物大分子结构和功能特性认识的基础上,用生物信息学方法通过计算机模拟和计算来“预测”出这些信息或提供与之相关的辅助信息,已成为目前研究的热点。本研究借助生物信息学工具对 plnB、plnC 和 plnD 编码蛋白进行了理化性质、结构和功能的分析预测,发现 plnB 编码蛋白具有典型的组氨酸蛋白激酶的结构特征,plnC、plnD 编码蛋白具有典型的感应调节蛋白的结构特征,推测

plnBCD 是 *L. paraplantarum* L-XM1 群体感应系统重要组分。在产细菌素乳酸菌中, 研究得较为透彻的双组份系统, 多数与自诱导肽组成群体感应系统, 参与细菌群体感应, 其 HPK 为 HPK10 亚族, RR 为 LytR 亚族, *L. paraplantarum* L-XM1 中 *plnBCD* 即属于此类。

调控乳杆菌Ⅱ类细菌素合成的群体感应系统的生物信息学研究虽然已有报道, 但仅限于植物乳杆菌、副干酪乳杆菌等少数菌种, 对 *L. paraplantarum* 群体感应系统生物信息学研究尚未见报道, L-XM1 菌株群体感应系统组分的生物信息学分析对于丰富乳杆菌群体感应系统的序列特征、理化性质、空间结构等具有重要的理论意义和实际价值。目前, 对乳杆菌群体感应系统组分的空间结构的研究相对较少, 多数研究是通过借鉴其他菌, 如金黄色葡萄球菌^[18]、大肠埃希菌^[21]等的双组份系统。类植物乳杆菌 L-XM1 PlnBCD 的生物信息学研究, 将为 PlnBCD 的功能研究提供理论依据, 为新的群体感应调控系统的发现提供参考。

参 考 文 献

- [1] Maldonado-Barragán A, Caballero-Guerrero B, Lucena-Padrós H, et al. Induction of bacteriocin production by coculture is widespread among plantaricin-producing *Lactobacillus plantarum* strains with different regulatory operons[J]. Food Microbiol, 2013, 33(1): 40-47
- [2] Axelsson L, Holck A. The genes involved in production of and immunity to sakacin A, a bacteriocin from *Lactobacillus sake* Lb706[J]. J Bacteriol, 1995, 177(8): 2125-2137
- [3] Eijssink V G H, Brurberg M B, Middelhoven P H, et al. Induction of bacteriocin production in *Lactobacillus sake* by a secreted peptide[J]. J Bacteriol, 1996, 178(8): 2232-2237
- [4] Man L L, Meng X C, Zhao R H, et al. The role of *plnC8HK-plnD* genes in bacteriocin production in *Lactobacillus plantarum* KLD51.0391[J]. Int Dairy J, 2014, 34(2): 267-274
- [5] Cui Y, Qu X. Comparative analysis of two component signal transduction systems of the *Lactobacillus acidophilus* group[J]. Braz J Microbiol, 2011, 42(1): 147-157
- [6] 葛菁萍, 赵晓龙, 张岚, 等. 副干酪乳杆菌响应调节蛋白基因克隆与序列分析[J]. 微生物学杂志, 2009, 29(1): 60-64
- [7] Diep D B, Håvarstein L S, Nes I F. A bacteriocin-like peptide induces bacteriocin synthesis in *Lactobacillus plantarum* C11[J]. Mol Microbiol, 1995, 18(4): 631-639
- [8] Navarro L, Rojo-Bezares B, Sáenz Y, et al. Comparative study of the *pln* locus of the quorum-sensing regulated bacteriocin-producing *L. plantarum* J51 strain[J]. Int J Food Microbiol, 2008, 128(2): 390-394
- [9] Mathiesen G, Axelsen G W, Axelsson L, et al. Isolation of constitutive variants of a subfamily 10 histidine protein kinase (SppK) from *Lactobacillus* using random mutagenesis[J]. Arch Microbiol, 2006, 184(5): 327-334
- [10] Johnsborg O, Diep D B, Nes I F. Structural analysis of the peptide pheromone receptor PlnB, a histidine protein kinase from *Lactobacillus plantarum*[J]. J Bacteriol, 2003, 185(23): 6913-6920
- [11] Schwede T, Sali A, Honig B, et al. Outcome of a workshop on applications of protein models in biomedical research[J]. Structure, 2009, 17(2): 151-159
- [12] West A H, Stock A M. Histidine kinases and response regulator proteins in two-component signaling systems[J]. Trends Biochem Sci, 2001, 26(6): 369-376
- [13] Gao R, Stock A M. Biological insights from structures of two-component proteins[J]. Annu Rev Microbiol, 2009, 63: 133-154
- [14] Grebe T W, Stock J B. The histidine protein kinase superfamily[J]. Adv Microb Physiol, 1999, 41: 139-227
- [15] Nakayama J, Akkermans A D L, De Vos W M. High-throughput PCR screening of genes for three-component regulatory system putatively involved in quorum sensing from low-G+C Gram-positive bacteria[J]. Biosci Biotech Bioch, 2003, 67(3): 480-489
- [16] 张香美, 李平兰. 产Ⅱ类细菌素乳酸菌群体感应及其应用[J]. 微生物学报, 2011, 51(9): 1152-1157
- [17] Nikolskaya A N, Galperin M Y. A novel type of conserved DNA-binding domain in the transcriptional regulators of the AgrR/AgrA/LytR family[J]. Nucleic Acids Res, 2002, 30(11): 2453-2459
- [18] Sidote D J, Barbieri C M, Wu T, et al. Structure of the *Staphylococcus aureus* AgrA LytTR domain bound to DNA reveals a beta fold with an unusual mode of binding[J]. Structure, 2008, 16(5): 727-735
- [19] Galperin M Y. Structural classification of bacterial response regulators: Diversity of output domains and domain combinations[J]. J Bacteriol, 2006, 188(12): 4169-4182
- [20] Gao R, Stock A M. Molecular strategies for phosphorylation-mediated regulation of response regulator activity[J]. Curr Opin Microbiol, 2010, 13(2): 160-167
- [21] Gao R, Mack T R, Stock A M. Bacterial response regulators: Versatile regulatory strategies for common domains[J]. Trends Biochem Sci, 2007, 32(5): 225-234
- [22] Patrucco E, Kraynik S, Beavo J A. Phosphodiesterase 8A, cAMP-specific[J]. UCSD Molecule Pages, 2012, 1(1): 27-30
- [23] Wang L, Tian X, Wang J, et al. Autoregulation of antibiotic biosynthesis by binding of the end product to an atypical response regulator[J]. P Natl Acad of Sci, 2009, 106(21): 8617-8622
- [24] Diep D B, Myhre R, Johnsborg O, et al. Inducible bacteriocin production in *Lactobacillus* is regulated by differential expression of the *pln* operons and by two antagonizing response regulators, the activity of which is enhanced upon phosphorylation[J]. Mol Microbiol, 2003, 47(2): 483-494
- [25] Straume D, Johansen R F, Bjørås M, et al. DNA binding kinetics of two response regulators, PlnC and PlnD, from the bacteriocin regulon of *Lactobacillus plantarum* C11[J]. BMC Biochem, 2009, 10(1): 17