

文章编号:1000-2367(2020)01-0082-08

DOI:10.16366/j.cnki.1000-2367.2020.01.014

鞘氨醇单胞菌中威兰胶合成关键基因 *welE* 和 *welC* 的生物信息学分析

钱锦¹, 李丽², 常爱平², 陆钰佳², 何侨妹², 沈亚领¹, 朱虎²

(1.华东理工大学 生物工程学院, 上海 200237; 2.福建师范大学 工业生物催化
福建省高校工程研究中心; 福建省先进材料化工基础重点实验室; 化学与材料学院, 福州 350007)

摘要:为探究威兰胶合成途径中关键基因 *welE* 和 *welC* 的结构特性及功能, 运用生物信息学手段分析了 *welE* 和 *welC* 基因的序列信息, 预测了其编码蛋白的理化性质、跨膜区、信号肽、磷酸化修饰、结构域及高级结构等。研究表明, *WelE* 蛋白是由 234 个氨基酸组成的稳定亲水性蛋白质, 无跨膜结构和信号肽, 存在 20 个磷酸化位点, 含有一个 BY-kinase 结构域, 说明 *WelE* 可能参与威兰胶的生产与转运, *WelE* 二级结构主要由 α -螺旋和无规则卷曲组成, 三级结构呈“圆环状”; *WelC* 蛋白是由 449 个氨基酸组成的不稳定亲水性跨膜蛋白质, 含有 2 个跨膜螺旋、无信号肽, 存在 39 个磷酸化位点, 含有一个 Wzz 结构域, 这表明 *WelC* 可能与威兰胶链长调节有关, *WelC* 二级结构以 α -螺旋和无规则卷曲为主, 三级结构呈“条状”, 这些结果为解析 *welE* 和 *welC* 基因的功能以及建立威兰胶代谢调控方法奠定了理论基础。

关键词:威兰胶; *welE* 基因; *welC* 基因; 生物信息学; 蛋白结构

中图分类号:Q503

文献标志码:A

威兰胶(Welan gum)又名韦兰胶或威伦胶, 是由产碱杆菌(*Alcaligenes* sp. ATCC 31555)^[1]和鞘氨醇单胞菌(*Sphingomonas* sp. WG)^[2-3]在适当的条件下分泌的一种可溶性胞外多聚糖, 是美国 KELCO 公司继黄原胶、结冷胶之后开发的最具市场前景的微生物多糖之一。威兰胶的骨架结构同结冷胶和鼠李胶等鞘氨醇胶类似, 由 D-葡萄糖、D-葡糖醛酸、D-葡萄糖和 L-鼠李糖组成, 侧链为单链 L-鼠李糖(占 2/3)或单链 L-甘露糖(占 1/3)^[4], 此外, 约有半数的四糖片段上还带有乙酰基和甘油基团。在威兰胶的化学组成中, 乙酰基(连接在 1→3 结合的吡喃型葡萄糖基团的 2 号碳原子上)占 2.8%~7.5%, 葡萄糖醛酸占 11.6%~14.9%, 中性糖甘露糖、葡萄糖和鼠李糖的摩尔比约为 1:2:2^[5], 其中大多数鼠李糖以 1→4 连接, 约是以末端连接的 2 倍, 而大部分的葡萄糖以 1→3 键连接^[6]。

威兰胶作为一种天然合成的高分子聚合物, 既有同植物多糖相似的性能, 又具有聚乙烯、聚丙烯等化学高分子物质的物理性能^[7]。即使温度升高至 150 °C, 溶液的黏度也不会有明显变化, 除了对热稳定以外, 威兰胶水溶液对酸碱的耐受性也很强, 其黏度在 pH=1~12 范围内仅有轻微的变化^[8-9]。同时它还具有易分解、无毒、无污染等优点, 因此, 威兰胶可作为驱油剂、增稠剂、悬浮剂等应用于石油、食品、石墨、医药和混凝土等诸多领域^[10-11]。

威兰胶的合成途径与结冷胶、S-88 等鞘氨醇胶类似, 主要包括核苷酸糖前体合成、四糖重复单元组装、长链聚合和分泌 3 个步骤^[12]。在威兰胶合成过程中, 多糖基本单元的聚合和分泌步骤中, 涉及 4 个关键的基因, 分别为 *welC*, *welE*, *welS* 和 *welG*, 前期研究发现这些基因可能与威兰胶聚合、链长调节和分泌有关^[13-14], 深入分析这些基因的结构与功能, 对于建立威兰胶合成调控新策略, 提高威兰胶的产量至关重要。

收稿日期:2019-07-24; 修回日期:2019-12-10.

基金项目:国家“863 计划”(2015AA020925); 国家自然科学基金(U1805234; 31800075); 福建省自然科学基金(2019J01264).

作者简介:钱锦(1995-), 女, 湖南慈利人, 华东理工大学硕士研究生, 主要从事微生物分子生物学研究, E-mail: y30171188@mail.ecust.edu.cn.

通信作者:沈亚领, 教授, 博士生导师, 长期从事生物工程科学研究, E-mail: ylshen@ecust.edu.cn; 朱虎, 福建师范大学教授, 博士生导师, 主要从事生物化工研究, E-mail: zhuhu@fjnu.edu.cn.

因此,本研究运用生物信息学手段对 *welC* 和 *welE* 基因及其编码蛋白的性质、结构及功能进行了研究。

1 材料与方法

1.1 实验材料

welE 和 *welC* 基因序列均从课题组前期测定的鞘氨醇单胞菌(*Sphingomonas* sp. WG)全基因组序列(NCBI ACCESSION NO: NZ_LNOS00000000)中获得^[13], 位置在 contig_62 和 contig_36 中。

1.2 序列分析

使用 NovoPro(<http://www.novopro.cn/tools/translate.html>)在线软件将 *welE* 和 *welC* 基因序列翻译为对应的氨基酸序列, 利用 ProtParam 程序(<http://web.expasy.org/protparam/>)^[15] 获取各蛋白质的理论等电点、相对分子质量、氨基酸组成、消光系数和稳定系数, 蛋白亲水/疏水性分析利用在线软件 ProtScale(<http://web.expasy.org/protscale/>)^[15] 完成, 使用 TMHMM v2.0(<http://www.cbs.dtu.dk/services/TMHMM/>)^[16] 在线软件预测蛋白质的跨膜螺旋结构域, 使用 SignalP 5.0 Server(<http://www.cbs.dtu.dk/services/SignalP>)在线软件预测蛋白是否含有信号肽, 蛋白质结构域组织信息通过 NCBI Conserved Domain Search(CDD, <https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>)^[17] 平台获取, 利用 NetPhos 3.1 Server(<http://www.cbs.dtu.dk/services/NetPhos/>)^[18] 在线预测蛋白潜在的磷酸化位点, 利用 ExPASy 在线服务器上的 SOPMA(https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html)^[19] 分析蛋白质的二级结构, 通过 SWISS-MODEL(<https://www.swissmodel.expasy.org/interactive>)^[20] 软件对蛋白的三级结构进行预测。

2 实验结果

2.1 蛋白质理化性质和一级结构分析

2.1.1 蛋白质理化性质分析

蛋白质一级结构是指多肽链内的氨基酸残基由 N 末端到 C 末端的顺序排列, 利用在线工具 ExPASy ProtParam 分析 *WelE* 和 *WelC* 蛋白的氨基酸组成及其理化性质, 如表 1 所示, *WelE* 含有 234 个氨基酸, 其中含量最高的前两位氨基酸为丙氨酸(Ala 32, 13.7%)和亮氨酸(Leu 25, 10.7%), 含量最少的是组氨酸(His 3, 1.3%)和赖氨酸(Lys 3, 1.3%), 色氨酸(Trp)的含量为 0。*WelE* 蛋白的分子式为 $C_{1085}H_{1776}N_{316}O_{347}S_8$, 原子总数 3 532, 相对分子质量为 25.056 kDa, 带负电荷的残基(Asp+Glu)28 个, 带正电荷的残基(Arg+Lys)22 个, 理论等电点(pI)值为 5.12, 属于酸性蛋白, 消光系数为 0.297, 不稳定系数为 28.64(<40), 为稳定蛋白, 脂肪系数为 98.08, 总平均亲水系数(GRAVY)为 $-0.004 < 0$, 预测该蛋白为亲水性蛋白。

WelC 含有 449 个氨基酸, 其中含量最高的前两位氨基酸为丙氨酸(Ala 83, 18.5%)和亮氨酸(Leu 39, 8.7%), 含量最少的是组氨酸(His 3, 0.7%), 半胱氨酸(Cys)的含量为 0。*WelC* 蛋白的分子式为 $C_{2122}H_{3408}N_{626}O_{652}S_8$, 原子总数为 6 816, 相对分子质量为 48.379 kDa; 带负电荷的残基(Asp+Glu)45 个, 带正电荷的残基(Arg+Lys)44 个, 理论等电点(pI)值为 6.45, 属于酸性蛋白, 消光系数为 0.876, 不稳定系数为 45.82(>40), 属于不稳定蛋白, 脂肪系数为 92.78, 平均亲水系数(GRAVY)为 $-0.104 < 0$, 预测该蛋白为亲水性蛋白。

表 1 *WelE* 和 *WelC* 的理化性质分析

Tab.1 Physic-chemical property analysis of *WelE* and *WelC*

理化性质	<i>WelE</i>	<i>WelC</i>	理化性质	<i>WelE</i>	<i>WelC</i>
氨基酸总数/个	234	449	理论等电点	5.12	6.45
原子总数/个	3 532	6 816	消光系数	0.297	0.876
相对分子质量/kDa	25.056	48.379	脂肪系数	98.08	92.78
带负电荷的残基/个	28	45	不稳定系数	28.64	45.82
带正电荷的残基/个	22	44	平均亲水系数	-0.004	-0.104

2.1.2 蛋白质的亲水性和疏水性分析

蛋白质亲疏水性氨基酸的组成是蛋白质折叠的主要驱动力,通常情况下,可以根据蛋白质的亲疏水性分布图分析出蛋白质的折叠情况.利用 ExPAXy ProtScale 在线软件进一步分析了蛋白亲疏水性,分析结果如图 1 所示,WeIE 蛋白分布于阈值 0 以下的氨基酸数目明显高于分布于阈值 0 以上的氨基酸数目(0 以上表示疏水区,0 以下表示亲水区),其最大疏水值 2.267 出现在第 198 位氨基酸处,最大亲水值 -2.678 出现在第 11 位氨基酸处,平均亲水系数为 $-0.004 < 0$,推测 WeIE 蛋白为亲水蛋白.

WeIC 蛋白分布于阈值 0 以下的氨基酸数目明显高于分布于阈值 0 以上的氨基酸数目,其最大疏水值 3.122 出现在第 391 位氨基酸处,最大亲水值 -2.211 出现在第 145 和第 435 位氨基酸处,平均亲水系数为 $-0.104 < 0$,推测 WeIC 为亲水蛋白.

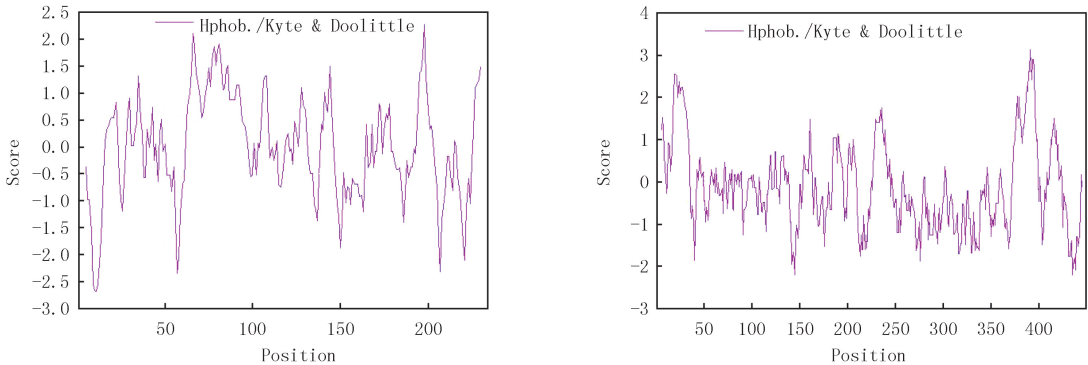


图 1 WeIE(左)和WeIC(右)的亲/疏水性曲线图

Fig.1 Curves of hydrophilicity or hydrophobicity of WeIE(Left) and WeIC(Right)

2.2 蛋白序列特征和功能分析

2.2.1 蛋白质的跨膜区分析

膜蛋白具有细胞间的信号传导、物质运输等重要功能,因此有必要对蛋白质是否存在跨膜结构进行预测.利用 TM-HMM Server v.2.0 在线软件分析预测 WeIE 和 WeIC 蛋白的跨膜结构,结果表明(图 2),WeIE 蛋白不具有跨膜结构,蛋白全部位于膜内. WeIC 含有 2 个跨膜螺旋(TMHS),分别位于第 13~35 个氨基酸和第 373~395 个氨基酸处,非跨膜螺旋区(ExpAA)概率为 46.438 54,N 端位于膜内侧的概率为 0.624 44,总体位于膜外,预测 WeIC 蛋白属于跨膜蛋白.

2.2.2 蛋白质的信号肽预测

信号肽是一段由 15~30 个氨基酸残基组成的短肽,通常位于分泌蛋白 N 端,在分泌蛋白合成结束后将其切除,因此可以通过分析蛋白的氨基酸序列是否具有信号肽来判断蛋白是否为分泌蛋白.利

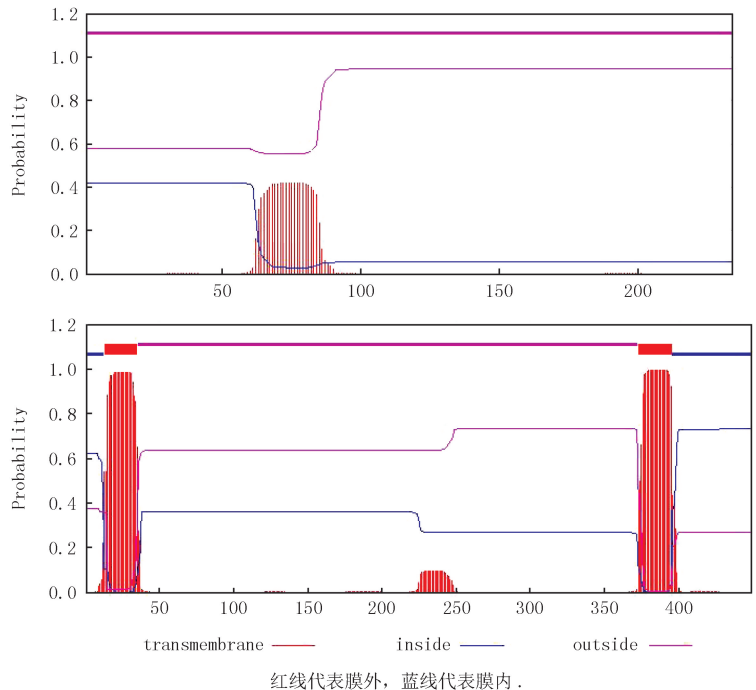


图 2 WeIE(上)和WeIC(下)的跨膜区分析

Fig.2 Transmembrane Region Analysis of WeIE(top) and WeIC(bottom)

用 SignalP 5.0 Server 对 WelE 和 WelC 蛋白进行信号肽分析,结果如图 3 所示,WelE 和 WelC 的初始切位点分值(C-Score)、结合剪切位点分值(Y-Score)和信号肽分值(S-Score)均未达到阈值 0.5,所以推测这两个蛋白都不存在信号肽,均属于非分泌蛋白,不能引导蛋白质跨膜运输。

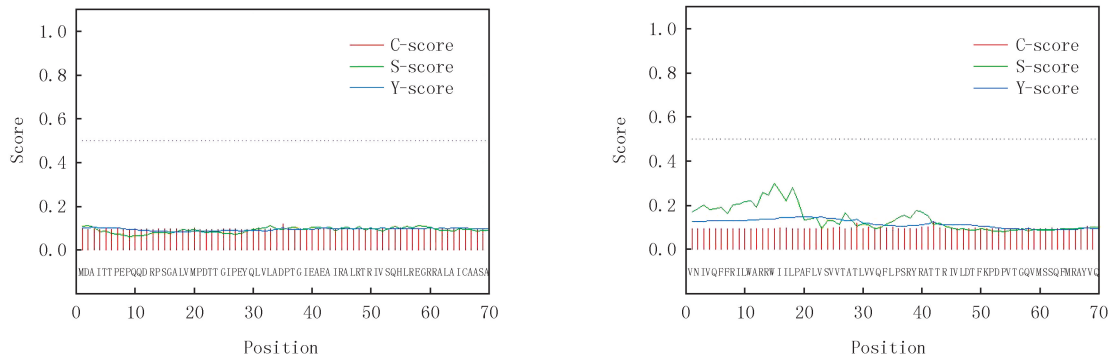


图 3 WelE (左) 和 WelC (右) 的信号肽预测

Fig.3 Signal Peptide Prediction of WelE (left) and WelC (right)

2.2.3 蛋白磷酸化修饰预测

蛋白质表达后,需要进行翻译后修饰才能发挥正常功能,磷酸化是蛋白最常见的修饰之一,在基因表达、信号传递和代谢等重要生命过程中起到关键调节作用。利用 NetPhos 3.1 Server 在线预测 WelE 和 WelC 蛋白潜在的磷酸化位点,分析结果(图 4, 阈值 0.5)表明,这两个蛋白在多个氨基酸位点存在不同程度的磷酸化,二者都有丝氨酸、苏氨酸、酪氨酸三种磷酸化位点,但数量存在差异(表 2),其中 WelE 存在 20 个潜在的磷酸化位点,包括 7 个丝氨酸(S),12 个苏氨酸(T)和 1 个酪氨酸(Y), WelC 存在 39 个潜在的磷酸化位点,包括 21 个丝氨酸(S),10 个苏氨酸(T)和 8 个酪氨酸(Y)。

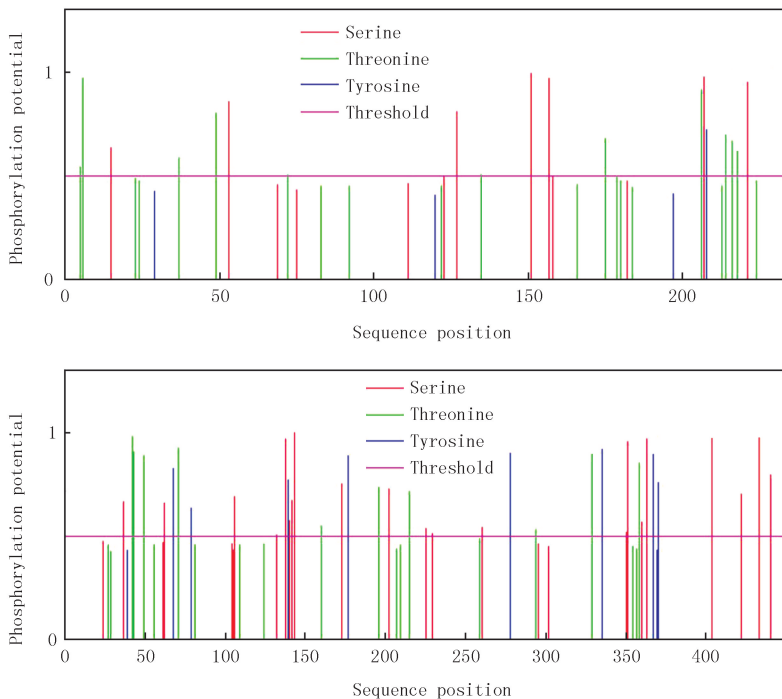


图 4 WelE(上) 和 WelC(下) 蛋白磷酸化位点

Fig.4 Phosphorylation sites of protein WelE (top) and WelC (bottom)

表 2 WelE 和 WelC 的磷酸化位点

Tab.2 Phosphorylation sites of protein WelE and WelC

磷酸化位点	WelE	WelC	磷酸化位点	WelE	WelC	磷酸化位点	WelE	WelC
丝氨酸(Serine)	7	21	酪氨酸(Tyrosine)	1	8	苏氨酸(Threonine)	12	10

2.2.4 蛋白质结构域预测

使用 NCBI Conserved Domain Search 对 WelE 和 WelC 蛋白序列进行结构域分析,结果表明(图 5), WelE 蛋白含有一个 Bacterial tyrosine(BY)-kinase 结构域,位于第 42~233 位氨基酸,属于 SIMIBI 蛋白超

家族,大多数 BY-kinase 蛋白参与胞外多糖的生产的转运;WelC 蛋白含有一个 Wzz 结构域,位于第 1~76 位氨基酸,Wzz 蛋白是一种决定链长的蛋白质,属于 Wzz 蛋白超家族。

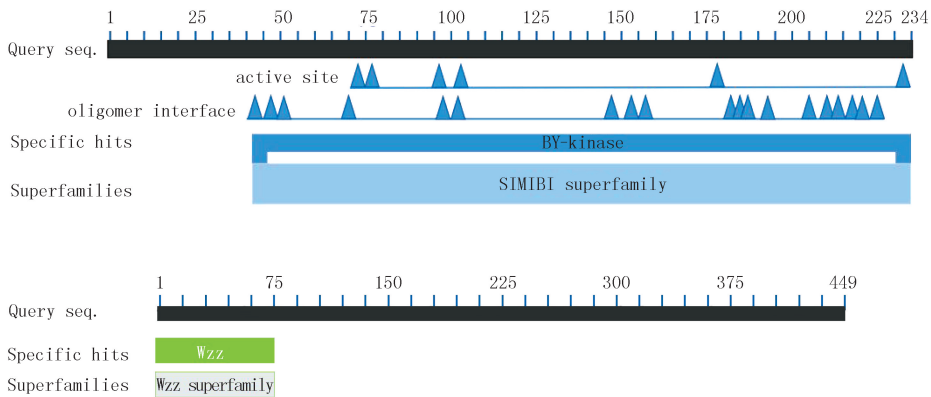


图5 WelE (上) 和 WelC (下) 蛋白保守结构域预测

Fig.5 Protein conserved domain prediction of WelE (top) and WelC (bottom)

2.3 蛋白质二级结构预测

蛋白质二级结构是由多肽主链有规则地盘绕、折叠而形成的一种重复性构象,其包括 α -螺旋、延伸链、 β -转角、无规则卷曲等,二级结构不仅是一级结构与三级结构之间的连接,还是预测三维空间结构的重要环节。

采用 SOPMA 在线软件预测 WelE 和 WelC 蛋白的二级结构(图 6),分析结果表明(表 3),WelE 和 WelC 蛋白二级结构均由 α -螺旋(Alpha helix)、延伸链(Extended strand)、 β -转角(Beta turn)和无规则卷曲(Random coil)构成.WelE 蛋白中有 91 个氨基酸参与形成 α -螺旋,占 38.89%;有 40 个氨基酸参与形成延伸链,占 17.09%;有 16 个氨基酸参与形成 β -转角,占 6.84%;有 87 个氨基酸参与形成无规则卷曲,占 37.18%。WelC 蛋白二级结构中含有 66.59%的 α -螺旋,8.24%的延伸链,1.78%的 β -转角和 23.39%的无规则卷曲,其中 α -螺旋和无规则卷曲是 WelC 和 WelE 蛋白的主要结构元件, β -转角则较少。

表 3 WelE 和 WelC 的二级结构

Tab.3 Secondary structure prediction of WelE and WelC

WelE			WelC		
类别	个数	比例/%	类别	个数	比例/%
Alpha heli(Hh)	91	38.89	Alpha heli(Hh)	299	66.59
Extended stran(Ee)	40	17.09	Extended stran(Ee)	37	8.24
Beta turn(Tt)	16	6.84	Beta turn(Tt)	8	1.78
Random coil(Cc)	87	37.18	Random coil(Cc)	105	23.39

2.4 蛋白质三级结构预测

蛋白质的三级结构是一种特定的空间结构,是多肽链侧链基团之间的相互作用发生卷曲折叠并靠次级键维系固定而形成的.预测蛋白质三级结构的方法有同源建模法、折叠识别法(串线法)和从头计算法,其中同源建模法是最常用的预测方法^[21],其原理是:具有相似序列的蛋白质会倾向于折叠成相似的空间结构,所以,如果两个蛋白质之间具有足够的序列相似性,那么未知蛋白质可以通过已知蛋白质的结构得到一个近似的三维模型。

利用在线软件 SWISS-MODEL 对 WelE 和 WelC 蛋白的三级结构进行模拟.通过同源建模的方法获得 WelE 蛋白的三级结构;将 WelE 蛋白的氨基酸序列提交至 SWISS-MODEL 软件,寻找其同源性较高的模板,在所给出的 50 个同源性较好的模板中,选取 3la6.1.A 为模板,在 31-232 位氨基酸同源建模,预测结果如图 7(左)所示,WelE 蛋白的氨基酸序列与模板序列的一致性(Seq Identity)为 30.24%,GMQE 为 0.63(可信

度较好), QMEAN 为 -1.07 (与模板的匹配度较好), 通过软件自带评分可知建模结果较为可靠。

用同样的方法获得 WelC 蛋白的三级结构: 在软件所提供的 50 个模板中, 以 5j1g.1.A 为模板进行建模, 可信度评分 GMQE 为 0.09 (可信度一般), 匹配度评分 QMEAN 为 -1.98 (与模板的匹配度较好), 通过软件自带评分可知建模结果合理, 模型预测结果如图 7 (右) 所示。

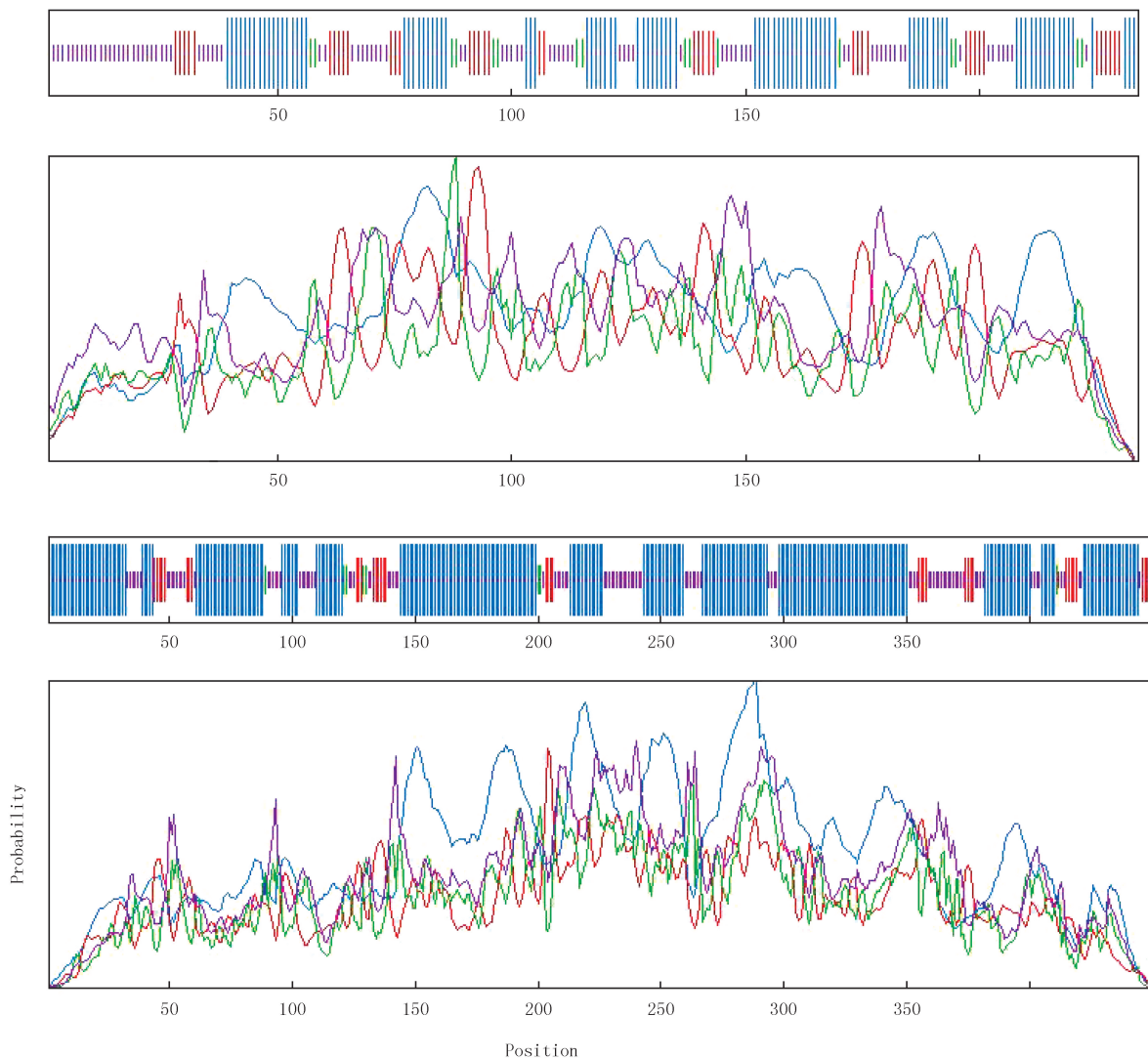


图 6 WelE (上) 和 WelC (下) 的二级结构

Fig. 6 Secondary structure prediction of WelE (top) and WelC (bottom)

通过建模所得到的三级结构可以看出, WelE 和 WelC 蛋白均以 α -螺旋和无规则卷曲为主, 与二级结构预测一致, 但二者的三级结构所呈现出的形状却截然不同, WelE 蛋白呈“圆环状”, WelC 蛋白呈“条状”。

3 讨论

长期以来, 威兰胶的生产技术被美国垄断, 市场价格十分昂贵, 而国内威兰胶生产存在产率低、成本高、成品质量较差等问题, 我国对威兰胶的需求仍依赖于进口^[22], 因此, 深入分析威兰胶合成途径的关键基因及功能, 建立威兰胶合成代谢的调控手段, 构建威兰胶高产菌株, 开发工业化生产工艺具有重要的现实意义。本文利用课题组前期测定的鞘氨醇单胞菌 (*Sphingomonas* sp. WG) 全基因组序列中获得的 *welE* 和 *welC* 基因序列, 通过使用一系列在线软件, 对这两个基因进行了生物信息学分析, 结果表明, *welE* 基因编码蛋白是由 234 个氨基酸组成的稳定的亲水性蛋白质, 无跨膜结构、信号肽, 存在 20 个磷酸化位点, 含有一个 BY-kinase

结构域,说明 WelE 可能参与威兰胶的生产与转运,WelE 二级结构包括 38.89% α -螺旋、7.09% 延伸链、6.84% β -转角和 37.18% 无规则卷曲;*welC* 基因编码蛋白是由 449 个氨基酸组成的不稳定的亲水性跨膜蛋白质,含有 2 个跨膜螺旋、无信号肽,存在 39 个磷酸化位点,含有一个 Wzz 结构域,表明 WelC 与威兰胶的链长有关,其二级结构中 α -螺旋占 66.59%、延伸链占 8.24%、 β -转角占 1.78%、无规则卷曲占 23.39%。本文进行的一系列生物信息学分析为后期进一步挖掘和验证 *welE* 和 *welC* 基因的功能奠定了理论基础,同时也为建立威兰胶合成调控新策略提供了思路。

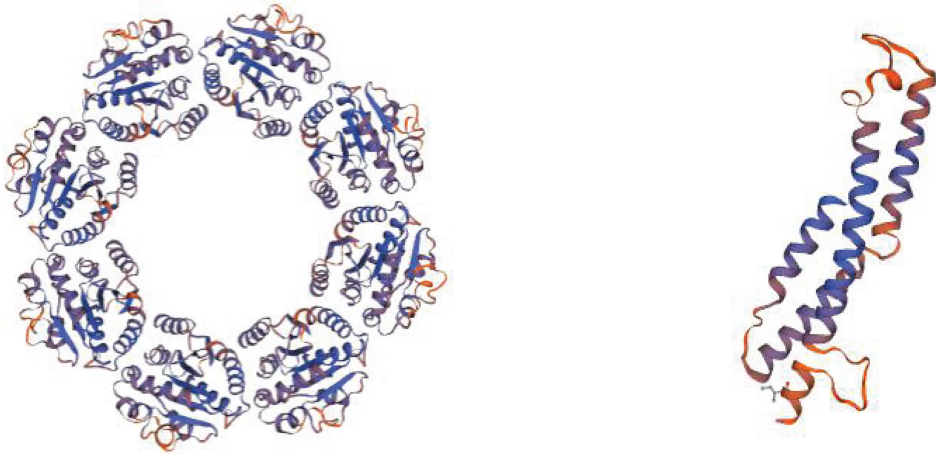


图7 WelE (左) 和 WelC (右) 的三级结构

Fig.7 Tertiary structure prediction of WelE (left) and WelC (right)

参 考 文 献

- [1] 刘敏.产碱杆菌 *Alcaligenes* sp.ATCC31555 发酵产韦兰胶的研究[D].广州:华南理工大学,2013.
LIU M.Study on the fermentation of Welan gum by *Alcaligenes* sp.ATCC31555[D].Guangzhou:South China University of Technology, 2013.
- [2] 周万龙,焦学,孙亚杰,等.1 株海洋来源鞘氨醇单胞菌胞外多糖的含量测定[J].中国海洋药物,2017(2):31-38.
ZHOU W L, JIAO X, SUN Y J, et al.Determination of extracellular polysaccharides produced by marine derived *Sphingomonas* sp. WG[J].Chinese Journal of Marine Drugs,2017,36(2):27-34.
- [3] 中国石油大学(华东).生产威兰胶的菌株及方法:中国,ZL201410485635.5[P/OL].2017-12-01[2019-07-24].<http://www.pss-system.gov.cn/>.
China University of Petroleum(East China).Welan gum producing strain and methods:China,ZL201410485635.5[P/OL].2017-12-01 [2019-07-24].<http://www.pss-system.gov.cn/>.
- [4] JANSSON P E, WIDMALM G.Welan gum(S-130)contains repeating units with randomly distributed l-mannosyl and l-rhamnosyl terminal groups,as determined by FABMS[J].Carbohydrate Research,1994,256(2):327-330.
- [5] ATTWOOL P T, ATKINS E D T, MILES M J, et al.X-Ray fibre diffraction results from *Alcaligenes*(ATCC 31555) microbial polysaccharide S-130 and a comparison with gellan gum[J].Carbohydrate Research,1986,148(1):c1-c4.
- [6] CLARKE A J, SARABIA V, KEENLEYSIDE W, et al.Compositional analysis of bacterial extracellular polysaccharides by high-performance anion-exchange chromatography[J].Analytical Biochemistry,1991,199(1):68-74.
- [7] LI H, LI J, ZHOU W L, et al.An efficient production of a novel carbohydrate polymer Sphingan WL[J].Journal of Chemical Technology & Biotechnology,2018,93:3472-3482.
- [8] 吉武科,赵双枝,董学前,等.新型微生物胞外多糖-韦兰胶的研究进展[J].中国食品添加剂,2011(1):210-215.
JI W K, ZHAO S Z, DONG X Q, et al.A novel microbial polysaccharide-the research progress of Welan Gum[J].China Food Additives, 2011,1:210-215.
- [9] 李莎.威兰胶发酵工艺优化及结构性能的研究[D].南京:南京工业大学,2005.
LI S.Study on optimization of fermentation process and structure property[D].Nanjing:Nanjing Tech University,2005.
- [10] 徐虹,欧阳平凯.生物高分子-微生物合成的原理与实践[M].北京:化学工业出版社,2010.
XU H, OUYANG P K.Biomacromolecule-principles and practice of microbial synthesis[M].Beijing:Chemical Industry Press,2010.
- [11] ZHU H, SUN S W, LI H, et al.Significantly improved production of Welan gum by *Sphingomonas* sp. WG through a novel quorum-sens-

- ing-interfering dipeptide cyclo(L-Pro-L-Phe)[J].International Journal of Biological Macromolecules,2019,126:118-122.
- [12] LI H,XU H,XU H,et al.Biosynthetic pathway of sugar nucleotides essential for welan gum production in *Alcaligenes* sp.CGMCC2428 [J].Applied Microbiology and Biotechnology,2010,86(1):295-303.
- [13] LI H,FENG Z M,SUN Y J,et al.Draft genome sequence of *Sphingomonas* sp.WG,a welan gum-producing strain[J].Genome announcements,2016,4(1):e01709-15.
- [14] LI H,LI J,JIAO X,et al.Characterization of the biosynthetic pathway of nucleotide sugar precursor UDP-glucose during sphingan WL gum production in *Sphingomonas* sp.WG[J].Journal of Biotechnology,2019,302:1-9.
- [15] WILKINS M R,GASTEIGER E,BAIROCH A,et al.Protein Identification and Analysis Tools in the ExpASY Server[J].Methods in Molecular Biology,1999,112(112):531-552.
- [16] KROGH A,LARSSON B,HEIJNE G V,et al.Predicting transmembrane protein topology with a hidden markov model:application to complete genomes[J].Journal of Molecular Biology,2001,305(3):567-580.
- [17] BAUER A M,BO Y,HAN L Y,et al.CDD/SPARCLE:functional classification of proteins via subfamily domain architectures[J].Nucleic Acids Research,2017,45(D1):D200-D203.
- [18] BLOM N,GAMMELTOFT S,BRUNAK S.Sequence and structure-based prediction of eukaryotic protein phosphorylation sites[J].Journal of Molecular Biology,1999,294(5):1351-1362.
- [19] GEOURJON C,G Deléage.SOPMA:Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments[J].Computer applications in the biosciences,CABIOS,1996,11(6):681-684.
- [20] BIASINI M,BIENERT S,WATERHOUSE A,et al.SWISS-MODEL:modelling protein tertiary and quaternary structure using evolutionary information[J].Nucleic Acids Research,2014,42:252-259.
- [21] BAKER D,SALI A.Protein structure prediction and structural genomics[J].Science,2001,294(5540):93-96.
- [22] LIN M,LIU Y H,CHEN W W,et al.Use of bacteria-immobilized cotton fibers to absorb and degrade crudeoil[J].International Biodeterioration & Biodegradation,2014,88(1):8-12.

Bioinformatics analysis of key genes *welE* and *welC* involved in welan gum biosynthesis in *Sphingomonas* sp.WG

Qian Jin¹,Li Li²,Chang Aiping²,Lu Yujia²,He Qiaomei²,Shen Yaling¹,Zhu Hu²

(1.School of Bioengineering,East China University of Science and Technology,Shanghai 200237,China;

2.College of Chemistry and Materials Science,Fujian Normal University,Fuzhou 350007,China)

Abstract: In order to explore the structures and functions of key genes *welE* and *welC* involved in the synthesis pathway of welan gum,a series of bioinformatics methods were used in this research to analyze the sequence characteristics of *welE* and *welC*, and predict the physicochemical properties, transmembrane regions, signal peptides, phosphorylation modification, domains and spatial structures of their encoded *WelE* and *WelC* proteins. The results showed that the *WelE* is a stable hydrophilic protein. It consists of 234 amino acids and owns a structure without transmembrane structure and signal peptide. There are 20 phosphorylation sites and a BY-kinase domain in *WelE*. The structure indicates that the *WelE* may be involved in the production and transport of welan gum. Its secondary structure is mainly composed of α -helix and random coil, and the tertiary structure is annular. The *WelC* is an unstable hydrophilic transmembrane protein consisting of 449 amino acids. This protein contains two transmembrane helices, but no signal peptide. There are 39 phosphorylation sites in *WelC* and it contains a Wzz domain. The domain shows that the *WelC* may participate in the chain-length regulation of welan gum. The secondary structure is mainly composed of α -helix and random coil, and the tertiary structure is strip. These results will lay a theoretical foundation for revealing the detailed functions of *welE* and *welC* genes and further establishing new strategies for metabolic pathway regulation of welan gum biosynthesis.

Keywords: welan gum; *welE* gene; *welC* gene; bioinformatics; protein structure