

细菌外毒素序列中特有模体的识别及其基因本体注释分析

牛超^{1,2}, 王月兰², 岳俊杰², 宁保安¹, 高志贤^{1*}, 梁龙^{2*}

(¹ 军事医学科学院卫生学环境医学研究所卫生检验室, 天津 300050)

(² 军事医学科学院生物工程研究所微生物基因组学研究室, 病原微生物生物安全国家重点实验室, 北京 100071)

摘要:【目的】识别细菌外毒素序列中特有模体, 进一步理解外毒素的致病机制。【方法】构建非致病性细菌蛋白质数据库, 利用 InterProScan 对数据库中非致病菌蛋白质序列以及收集的经实验确认的 89 条细菌外毒素蛋白质序列进行模体搜索。【结果】在 89 条细菌外毒素序列中, 分析得到了 39 个细菌外毒素特有模体。【结论】得到的外毒素特有模体与外毒素功能密切相关, 为在致病性细菌基因组内搜索外毒素序列奠定了基础; 同时通过对外毒素特有模体的基因本体 (Gene ontology, GO) 注释分析, 进一步阐明了细菌外毒素的致病机制。

关键词: 外毒素; 模体; 基因本体

中图分类号: Q933 **文献标识码:** A **文章编号:** 0001-6209 (2010) 09-1246-05

细菌在人体内寄生、增殖并引起疾病的特性称为细菌的致病性或病原性 (Pathogenicity), 细菌外毒素 (Exotoxin) 与细菌致病性密切相关^[1]。产生外毒素的细菌主要是一些革兰氏阳性细菌和少数革兰氏阴性菌, 例如金黄色葡萄球菌、白喉杆菌、破伤风杆菌、霍乱弧菌和产毒性大肠杆菌等。外毒素对组织的毒性作用有高度的选择性, 可以引起特殊的临床症状。例如破伤风杆菌产生的是破伤风外毒素, 作用到脊髓和脑, 引起肌肉的痉挛和强直^[2]; 白喉杆菌产生的白喉外毒素, 能抑制人体细胞蛋白质的合成, 使细胞变性死亡, 导致心肌炎、肾上腺出血和神经麻痹^[3]; 霍乱弧菌产生的肠毒素作用到小肠粘膜, 使粘膜细胞分泌功能加强, 引起严重的呕吐和腹泻^[4]。

模体 (motif/domain) 是蛋白质具有的局部保守序列区域, 它是蛋白质的基本结构单位和功能单位, 决定着蛋白质的主要功能^[5]。模体作为一个重要的进化单元, 可以在基因组中重复出现, 这表明模体

在基因组中的重复和重排可能是影响该基因编码的蛋白质功能的一个重要因素; 模体的核苷酸序列比较保守, 可以经若干代的进化比较完整地保留下来, 这也证明了模体对于蛋白质的功能具有无可替代的重要性。在进行蛋白质功能注释时, 在序列整体同源性不明显的情况下, 通过进行模体搜索来查找序列上的局部特征, 可以提高功能预测的灵敏度^[6]。

细菌毒素与宿主细胞发生特异性的相互作用是细菌致病的关键, 蛋白质之间的相互作用实际上是组成蛋白质的模体之间的相互作用^[7]。我们借助生物信息学方法, 对已经经过实验确定的外毒素序列进行了序列和模体分析, 得到了一些与外毒素功能相关的特有模体, 并进一步对相应基因本体 (Gene Ontology, GO) 注释进行分析, 旨在更深入地认识外毒素, 为理论研究和生物学实验提供参考。

* 通信作者。高志贤, Tel: +86-22-84655191, E-mail: gaozhx@163.com; 梁龙, Tel: +86-10-66948874, E-mail: LL@bioinflab.org

作者简介: 牛超 (1981-), 男, 河南焦作人, 博士研究生, 主要从事食品安全关键技术研究。E-mail: niuchao601@126.com

收稿日期: 2010-03-16; 修回日期: 2010-04-21

1 材料和方法

1.1 材料

1.1.1 主要试剂和仪器:大型计算服务器一台:型号 Altix 3700bx 64CPU 64G, 由美国硅图公司 (Silicon Graphics, SGI) 生产。

1.1.2 实验材料:从 NCBI GenBank (<ftp://ftp.ncbi.nih.gov/genbank>) 下载了截止 2007 年 12 月已经完成全基因组测序的 279 株对人及动物不致病的细菌菌株的全基因组蛋白质序列, 构建非致病菌蛋白质数据库, 数据库中包含了 991,618 条蛋白质序列。从细菌毒力因子数据库 (Virulence factor database, VFDB) (www.mgc.ac.cn/VFs/) 中下载经实验确认的 89 条外毒素蛋白质序列^[8]。

1.2 方法

采用 InterProScan 分析工具对 991618 条非致病菌蛋白质序列和 89 条外毒素蛋白质序列进行模体识别搜索。InterProScan 数据库整合了 PROSITE、Pfam、PRINTS、ProDom、SMART、TIGRFAMs、PIRSF、SUPERFAMILY、Gene3D、PANTHER 等多个关于蛋白质结构域、功能位点、家族保守位点等信息, 是目前内容最全、功能最强的蛋白质模体分析系统, 并且 InterProScan 把相似的模体归类在一个 InterProScan 条目下 (一个 IPR), 给出统一的功能描述、GO 注释以及相关文献^[9]。由于 InterProScan (www.ebi.ac.uk/Tools/InterProScan/) 在线分析检索时每次只限提交一条序列, 无法满足我们高通量的分析大批量数据, 所以通过网址 <ftp://ftp.ebi.ac.uk/pub/databases/interpro/iprscan> 下载相关文件安装到本地服务器上对数据进行分析, 安装说明见 [ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/Installing _ InterProScan.txt](ftp://ftp.ebi.ac.uk/pub/software/unix/iprscan/Installing_InterProScan.txt)。本研究采用的相关 InterProScan 版本是 19.0。

同时利用所获得特有模体的 GO 注释结果对外毒素的特性进行了总结分析。GO (Gene Ontology) 即基因本体论, 它主要是制定出一套结构化的、定义精确的通用受控词表, 可用来描述任何有机生物体中基因和基因产物的作用^[10]。目前 GO 主要包括 3 大本体: 生物过程 (biological process)、分子功能 (molecular function) 及细胞组分 (cellular component), 在注解基因/蛋白质时, 首先考虑的是其所属

细胞组分, 其次是在分子水平上所行使的功能, 最后是分子功能所直接参与的生物过程, 三者构成一个完整的描述。

2 结果

2.1 外毒素序列的聚类

利用 BLASTClust 程序, 对 89 条细菌外毒素序列按照序列相似性大小进行了聚类, 将相似性超过 30% 的序列归为一个家族, 89 条外毒素序列可以分为 82 个家族, 其中 75 个家族只有一个序列, 7 个家族只有 2 个序列, 说明这些细菌外毒素分子之间序列整体同源性不高, 同时也说明了细菌外毒素的复杂多样性。

2.2 外毒素序列中特有模体的识别与分析

利用 InterProScan 数据库对 89 条外毒素序列进行模体搜索: 结果在 83 条外毒素序列找到了 InterProScan 数据库中 93 条对应的模体; 同时利用 InterProScan 数据库对 991618 条非致病菌蛋白质序列进行模体搜索, 结果中得到了无冗余的 9592 条对应的模体。比较分析搜索到的外毒素序列中的 93 条模体与非致病菌序列中的 9592 条模体, 在毒素序列对应模体中去除非致病菌中存在的模体, 最终获得了 39 条细菌外毒素序列所特有的模体。另外本文中涉及到的 InterProScan 数据库中 IPR 登陆号, 均可登陆 www.ebi.ac.uk/interpro/ 进一步查询其详细相关信息。

经统计分析 39 条细菌外毒素特有模体在致病菌外毒素中的分布, 发现有 22 个模体在致病菌菌株的分布上具有株特异性 (表 1), 例如炭疽芽孢杆菌所单独具有的炭疽毒素: 致死因子和水肿因子 (IPR015239 和 IPR014781), 这说明了致病菌菌株在分泌外毒素和宿主产生相互作用并导致宿主损伤时, 都分别采取了各自特有的相互作用。

同时, 还发现有 9 个外毒素特有模体分别在两个细菌致病菌菌株之间共享 (表 2), 经分析主要包括: 同一菌属下不同菌株的共享, 例如梭菌株之间、大肠埃希氏菌株之间等; 不同菌属之间的共享, 例如大肠埃希氏菌株和志贺氏菌株之间、霍乱弧菌菌株和大肠埃希氏菌株之间等。另外还发现 5 个模体在多个 (大于 2) 致病菌菌株上共享, 3 个模体在金黄色葡萄球菌和链脓球菌上超抗原毒素上分布具有特异性。这些都说明了不同致病菌在产生外毒素和宿主相互作用时同时也采用了某些共有的机制。

表 1 细菌菌株特异的 22 个外毒素特有模体在病原体上的分布

Table 1 The distribution of the 22 strain-specific exotoxin-specific motif/domain in bacterial pathogens

Bacterial Pathogen	InterPro ID	InterPro Description	Gi	Exotoxin Gene Name
<i>Bacillus anthracis str. 'Ames Ancestor'</i>	IPR015239	Anthrax toxin, lethal factor, central	47566484	lef
	IPR014781	Anthrax toxin, lethal/endema factor, N- and C-terminal	47566456	Lef
			47566484	cya
<i>Bordetella pertussis Tohama I</i>	IPR003898	Bordetella pertussis toxin A	33594638	ptxA
	IPR015355	Pertussis toxin, subunit S4	33594640	ptxD
	IPR015356	Pertussis toxin, subunit S5	33594641	ptxE
	IPR003899	Bordetella pertussis toxin	B33594639	ptxB
			33594642	ptxC
<i>Pseudomonas aeruginosa PAO1</i>	IPR003537	Yersinia virulence determinant YopE	15595242	exoT
	IPR015099	Exotoxin A catalytic	15596345	toxA
	IPR015185	Exotoxin A, binding	15596345	toxA
	IPR015186	Exotoxin A, targeting	15596345	toxA
	IPR014773	Yersinia virulence determinant YopE, C-terminal	15599036	exoS
			15595242	exoT
<i>Salmonella enterica (serovar typhimurium) LT2</i>	IPR003519	Salmonella virulence-associated 28 kDa protein	17233490	spvC
	IPR008834	Salmonella plasmid virulence SpvD	17233489	spvD
<i>Vibrio cholerae N16961</i>	IPR001835	Heat-labile enterotoxin, B chain	15641467	ctxB
	IPR011509	RtxA toxin	15641462	rtx
<i>Staphylococcus aureus MW2</i>	IPR008034	Delta lysin	21283688	hld
	IPR003963	Bi-component toxin, staphylococci	21282773	Hla
			21283107	lukF
			21283108	lukS
			21284071	lgA
			21284072	hlgC
			21284073	hlgB
<i>Clostridium tetani E88</i>	IPR011065	Kunitz inhibitor ST1-like	28373188	tetX
<i>Clostridium perfringens SM101</i>	IPR003897	Clostridium enterotoxin	110801889	cpe
<i>Streptococcus agalactiae 2603V/R</i>	IPR010860	CAMP factor	22538178	cfb
<i>Campylobacter jejuni NCTC 11168</i>	IPR015957	Cytolethal distending toxin A	15791469	cdtA
<i>Corynebacterium diphtheriae NCTC 13129</i>	IPR000512	Diphtheria toxin (NAD + -diphthamide ADP-ribosyltransferase)	38232848	tox

表 2 菌株间共享的 9 个外毒素特有模体在在病原体上的分布

Table 2 The distribution of the 9 exotoxin-specific motifs/domains shared by strains in bacterial pathogens

Bacterial Pathogen	InterPro ID	InterPro Description	Gi	Exotoxin Gene Name
<i>Clostridium botulinum A str. all</i> 和	IPR012928	Clostridium neurotoxin, receptor binding N-terminal	153936924	tont/a
<i>Clostridium tetani E88</i>			28373188	tetX
	IPR013104	Clostridium neurotoxin, receptor-binding C-terminalI		
	PR012500	Clostridium neurotoxin, translocationI		
	PR000395	Peptidase M27, bontoxilysin		
<i>Escherichia coli CFT07</i> 和	IPR013550	RTX C-terminal	26249405	hlyA
<i>Escherichia coli O157:H7 EDL933</i>			75994494	hlyA
<i>Escherichia coli O157:H7 EDL933</i> 和	IPR016139	Ribosome-inactivating protein, subdomain 2	15800960	stx2A
			15802646	stx1A
<i>Shigella dysenteriae Sd197</i>	IPR016331	Shiga-like toxin, subunit A	82776676	stxA
			15800961	stx2B
	IPR003189	Shiga-like toxin, beta subunit	15802645	stx1B
			82776677	stxB
<i>Vibrio cholerae N16961</i> 和	IPR001144	Heat-labile enterotoxin, A chain	15641468	ctxA
<i>Escherichia coli E24377A</i>			157149353	eltA

2.3 外毒素相关 GO 注释分析

在对细菌外毒素所具有的 39 个特有模体的 GO 注释分析时,我们发现相关 GO 注释条目分别在细胞成分、分子功能、生物学过程中出现了富集。例如:外毒素在细胞成分中出现频次最高的是细胞外区域(GO:0005576, extracellular region),这充分说明了外毒素是病原菌在代谢过程中分泌到菌体外的物质;在分子功能方面主要是具有一些酶活性和连接功能等,这说明了外毒素往往具有特定的生物学功能,并且通过特异性的相互作用,特定地作用于宿主靶标;而在生物学过程中外毒素主要参与致病机理(GO:0009405, pathogenesis)和蛋白质水解(GO:0006508, proteolysis)等过程,这说明了外毒素主要参与宿主致病过程,通过蛋白质水解而阻止宿主蛋白质合成,最终对宿主造成损伤。

3 讨论

外毒素是致病菌毒力的重要组成部分,对其进行深入研究有助于从分子水平上了解毒素与细菌致病机理间的相关性,为细菌性传染病的诊断预防、疫苗研制等带来新的视野,从而为探索更多更有效的治疗方法奠定基础。

在本研究中,我们首先对数据库中搜集的比较确定的 89 条外毒素进行了序列比较,结果表明这些毒素彼此之间的同源性很低,说明了细菌毒素的复杂多样性。随后,利用 InterProScan 数据库对收集的外毒素序列以及完成全基因组测序的非致病菌蛋白质序列进行了模体查找,通过比较分析最终得到了 39 条细菌外毒素所特有的模体。进一步分析 39 条特有模体在细菌致病菌上的分布,我们发现细菌致病菌菌株不仅享有自己特有的模体,而且菌株之间也享有共同的模体,也就是说致病菌和宿主所产生的相互作用既有特异的相互作用,也有共同的相互作用,这些相互作用构成了致病的基础。总之,细菌外毒素特有模体的获得以及分布统计,为在细菌基因组内搜索新的外毒素序列奠定了基础;同时相应 GO 注释分析,进一步加深了对于外毒素致病机制

的理解。

参考文献

- [1] Middlebrook JL, Dorland RB. Bacterial toxins: cellular mechanisms of action. *Microbiological Reviews*, 1984, 48 (3):199-221.
- [2] Grumelli C, Verderio C, Pozzi D, Rossetto O, Montecucco C, Matteoli M. Internalization and mechanism of action of clostridial toxins in neurons. *Neurotoxicology*, 2005, 26 (5):761-767.
- [3] Lee JW, Nakamura LT, Chang MP, Wisnieski BJ. Mechanistic aspects of the deoxyribonuclease activity of diphtheria toxin. *Biochimica et Biophysica Acta*, 2005, 1747(1):121-131.
- [4] Miller CE, Majewski J, Faller R, Satija S, Kuhl TL. Cholera toxin assault on lipid monolayers containing ganglioside GM1. *Biophysical Journal*, 2004, 86 (6):3700-3708.
- [5] Sonnhammer EL, Wolfsberg TG. Identification of motifs in protein sequences. *Current protocols in cell biology*, 2001, Appendix 1:Appendix 1C.
- [6] Sigrist CJ, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P. PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings in bioinformatics*, 2002, 3(3):265-274.
- [7] Dyer MD, Murali TM, Sobral BW. Computational prediction of host-pathogen protein-protein interactions. *Bioinformatics*, 2007, 23(13):159-166.
- [8] Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research*, 2005, 33(Database issue):D325-328.
- [9] Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. *Nucleic Acids Research*, 2005, 33(Web Server issue):W116-120.
- [10] Giglio MG, Collmer CW, Lomax J, Ireland A. Applying the Gene Ontology in microbial annotation. *Trends in microbiology*, 2009, 17(7):262-268.

Identification of exotoxin-specific motifs/domains in bacterial exotoxin sequences and corresponding Gene Ontology analysis

Chao Niu^{1,2}, Yuelan Wang², Junjie Yue², Bao'an Ning¹, Zhixian Gao^{1*}, Long Liang^{2*}

(¹Institute of Hygiene and Environmental Medicine, Academy of Military Medical Sciences, Tianjin 300050, China)

(²Biotechnology Institute of Beijing, Academy of Military Medical Sciences, State Key Laboratory of Medical Microbiology and Biosafety, Beijing 100071, China)

Abstract: [**Objective**] To identify the exotoxin-specific motifs/domains in bacterial exotoxin sequences, and to expand understanding of bacterial exotoxins pathogenic mechanisms. [**Methods**] We constructed a non-pathogenic bacterial proteins database and collected 89 bacterial exotoxin sequences from Virulence factor database (VFDB), then we analyzed these protein sequences by motif/domain search using InterProScan (www.ebi.ac.uk/Tools/InterProScan/). [**Results**] We identified 39 exotoxin-specific motifs/domains in 89 bacterial exotoxin sequences. [**Conclusion**] The identified exotoxin-specific motifs/domains were closely related to the functions of the exotoxins and could be used as template to search for new exotoxins by mining pathogenic bacterial genomes. The analysis of the acquired Gene Ontology (GO) items was to further expand our understanding of bacterial exotoxin pathogenic mechanisms.

Keywords: exotoxin; motif/domain; Gene ontology

(本文责编:王晋芳)

* Corresponding authors. Zhixian Gao, Tel: + 86-22-84655191, E-mail: gaozhx@163.com; Long Liang, Tel: + 86-10-66948874, E-mail: LL@bioinflab.org

Received: 16 March 2010/Revised: 21 April 2010

系统发育树的构建方法

构建系统树是为了鉴定菌株的分类学地位,应该使用正确的方法构建。具体要求如下:

1. 将鉴定菌的 16S rRNA 序列递交 GenBank,用 Blast 软件搜索相似的 16S rRNA,然后一起构树。
2. 采用能反应分支长度的软件(如 NJ 法),并用 Bootstrap 值分析分支聚类的稳定性。
3. 用国际较为通用的一些建树方法,如 Neighbour - Joining 等,这样结果就更为可靠,更直观。
4. 请严格按照下列具体要求写作[参见:微生物学报,2004,44(2):143.]

① 系统树中:菌名应列出全称,且属和种名应斜体,名称后再加括号,其内含序列号。

② 图注(本刊的图注要求用英文写作):应表明"树"上所有的内容,包括:括号中的序号、分支点上的数字涵义、0.01 代表的意义。

③ 作图要求:要求达到印刷清晰,字体为"Time New Roman",字号为"8p"。可以选用两种方式——(A)文件格式为"*.Tif",分辨率为 600 线;(B)文件格式为"word",画出树,输入文字。