

结核分枝杆菌的后基因组研究和新型疫苗

谢建平 王洪海* 陈永青

(复旦大学 生命科学院 上海 200433)

POST-GENOME STUDIES ON *MYCOBACTERIUM TUBERCULOSIS* AND NOVEL ANTI-TB VACCINES

Xie Jianping Wang Honghai Chen Yongqing

(School of Life Science Fudan University, Shanghai 2000433, China)

关键词：结核分枝杆菌，后基因组，疫苗，生物信息学

中图分类号：R394 文献标识码：A 文章编号：0001-6209 (2001) 02-0252-06

结核病仍然是全球人类健康的威胁。全球人口的 1/3(约 20 亿人)感染过结核分枝杆菌, 每年 300 万人死于结核病, 死于结核病的人数是其他传染病死亡人数的总和^[8], 因此, 世界卫生组织在 1993 年和 1997 年两度发出警告。去年 3 月, 我国卫生部的专家也宣布中国进入结核病的紧急状态, 中国是除印度外的世界第二大结核病重灾区, 如果不予以足够的重视, 采取切实可行的措施, 结核病将严重影响我国的现代化进程。造成结核病重新在世界抬头的原因很多, 除了社会经济的因素外, 结核分枝杆菌的耐药性和当前唯一可用的疫苗—卡介苗 BCG 的免疫效果下降是两个主要的因素。治疗结核病的药物很少, 而且, 结核分枝杆菌 (*Mycobacterium tuberculosis*, 以下简称 MTB) 对所有的一线药物均产生了不同程度的耐药性, 耐至少两种药的多药耐性菌株的出现使问题更加严峻; 尽管约 88% 的新生儿均接种过卡介苗^[4], 但是对过去发表的文献分析表明: 接种卡介苗只有 50% 的保护效果^[3], 对成年人是否具保护效果还不得而知。现在, 全球已经达成共识: 对付结核病需要研究开发新的抗结核药物和新型预防、治疗疫苗^[1, 2]。2000 年 2 月 10 日, 由洛克非勒基金, 美国国立卫生研究院, WELICOME 信托基金, WHO, 世界银行, 美国疾病控制中心 CDC 以及一些非政府机构等召集一流的科学家, 大的制药公司在南非的开普敦就投资于开发治疗 TB 的新型药物达成共识。由 WHO 协调, 计划 2007 年提出一种新药申请, 2012 年提出第二种新药申请^[1], 后基因组研究将有助于上述目标的实现。

1 基因组研究^[2]

结核分枝杆菌的基因组序列 1998 年由英国剑桥 Sanger 中心完成, 是该中心“致病菌基因组项目”第一个完成全序列测定的致病菌。序列发表在 NATURE 杂志。测序所用的菌株是 H37Rv, 选用该菌株的原因是: 自 1905 年分离该菌株后, 一直是全世界广泛使用的研究材料; 它与一些临床分离的菌株不同; 该菌株在动物模型中仍然保留全部的毒力; 对药物敏感; 相对容易进行遗传操作; 该菌株的物理图谱、粘

*联系作者

作者简介: 谢建平(1971-), 男, 重庆市人, 微生物学博士研究生, 方向: 分子免疫, 主要兴趣: 结核分枝杆菌耐药机理及抗结核新药、疫苗的研发。

收稿日期: 2000-07-10, 修回日期: 2000-09-14

粒库、细菌人工染色体已经构建成功。

结核分枝杆菌全基因组序列(约 4.41Mb,3924 个开放阅读框,4000 个基因)的测定是攻克结核顽症的一个重要里程碑。基因组明显的特征是:G+C 含量高达 65.6%,对基因组参数的影响至少体现在三个方面(1)GTG 为起始密码为~35%,而 *Bacillus subtilis* ~9%,*E. coli* ~14%;(2)富含 GC 的密码子编码的氨基酸如 Ala Gly Pro Arg 和 trp 较多,富含 AT 的密码子编码的氨基酸如 Asn Ile Lys Phe 和 Tyr 较少;(3)tRNA 反密码子第一位较少 A,在翻译时可能存在广泛的摆动。G+C 含量作图发现:整个基因组的 G+C 含量分布极不均一,说明可能存在水平转移。几个高 G+C 含量区域属于 PGRS(Polymorphic G+C rich sequence),AT 含量高于平均值的区域主要编码聚乙酰类合成酶和跨膜蛋白。

与已知序列的其他生物的数据库比较,在 3924 个开放阅读框中,约 40% 有功能,44% 可能有功能,16% 称为孤儿序列(orphan sequence),与其他微生物的序列无相似性。结核分枝杆菌的基因有广泛的功能冗余,序列高度保守,对大量菌株进行 26 个位点的序列测定表明这些菌株在遗传组成上相当均一,这可能和结核分枝杆菌的 DNA 修复机制非常忠实有关。分析修复有关的基因表明有以下特点:(1)核苷前体均一,可能和 mutT(功能是除去被氧化的鸟苷,因为复制时被氧化的鸟苷的整合会导致碱基错配)有 3 拷贝有关;(2)无错配修复系统存在,因为该系统的几个关键基因 mutH,L,S,recJ 均不存在,这可能和 mutT 清除能力强,没有 mutH,L,S 系统的底物-错配对的碱基存在有关。

染色体的 20% 用于编码两类不同的蛋白。一类用于编码~250 种脂类代谢的酶(*E. coli* 仅 50 种);另一类用于编码两类大、不相关的富含 Gly 的酸性蛋白(PE, PPE 家族),功能未知,可能是产生抗原变异、逃避宿主免疫的主要来源。

结核分枝杆菌高度疏水的细胞壁特殊、复杂;在肽聚糖层外有厚的脂层,含有分支菌酸、结核菌醇,脂阿拉伯甘露聚糖、阿拉伯半乳聚糖等复杂的成分,其中许多成分的半衰期较长,是重要的致病成分。基因组中编码相当多的酶蛋白,除细菌常见的酶外,还包括了哺乳动物、植物的脂类代谢的酶类。部分蛋白可能是在特定的时间、空间表达,如在感染时的巨噬细胞中的表达。降解脂类的酶远比合成脂类的酶多,有 100 多个基因编码 β 氧化中催化每一步反应的酶。这可能和降解宿主的脂类及能量代谢有关。

结核分枝杆菌测序中最大的意外是:发现了两类大不相关的富含 Gly 的酸性蛋白(PE, PPE 家族),其 N-端高度保守,因为富含 Pro-Glu 和 Pro-Pro-Glu 基序,分别被称作 PE 和 PPE 家族。PE 家族有 99 个成员,其中 61 个属于 PGRS 亚类,含多个 Gly-Gly-Ala 串联重复,蛋白也含至少 1400 个氨基酸;其他的成员的 C-端变化很大。PPE 家族有 68 个成员,至少可以分成 3 个亚类,最意外的是 MPTR(major polymorphic tandem repeats),编码的蛋白含高达 3 000 个氨基酸,主要含有 Asn-x-Gly-x-Gly-Asn-x-Gly 基序(motif)(x 代表氨基酸)。PE 和 PPE 家族基因具有高度的多态性,这可能是上述基序在复制时链剪切的结果。通过改变表达蛋白的类型,提供宿主免疫系统一个活动的靶子,逃避免疫。

另外,基因组序列测定还发现了一些毒力因子。序列测定前确定的毒力因子仅 3 个:过氧化物酶,其功能是抵御宿主巨噬细胞产生的活性氧,mce 编码巨噬细胞集落因子,sigma A 因子,其突变将导致减毒。测序后发现 mce 有 4 个拷贝,均位于同一个操纵子内,(编码 8 个组织方式完全一致的基因,mce 之前两个编码膜整合蛋白,之后的 5 个编码带疏水信号肽的蛋白,说明该蛋白可能是分泌或者是定位在细胞的表面)。SmpB 的同系物与 *Salmonella typhimurium* bfrA 在宿主体内存活有关。其它可作为毒力因子的分泌蛋白如:磷酯酶 C(其中之一为细胞接触依赖的溶血素)、脂酶、酯化酶(攻击细胞膜)、蛋白酶、贮存蛋白、血红蛋白样蛋白(适应厌氧环境)、转铁素样蛋白(在铁饥饿时与宿主竞争铁),bfrA 和 B。

结核分枝杆菌基因组序列的测定标志着人类与结核分枝杆菌的斗争进入到了一个新的阶段。结核分枝杆菌的经典微生物学研究经历了几十年的缓慢前进后,重新进入了快车道。每个可能的药物靶子和可用于疫苗设计的保护性抗原均可能从基因组序列中检索,结核分枝杆菌致病机理的研究手段将发生质的飞跃。综合利用后基因组的研究手段,人类彻底根除结核病这个顽症已指日可待。

2 后基因组研究

序列测定完成后,结核分枝杆菌研究工作的主要目标是:综合利用各学科的优势和技术,彻底阐明每个基因的功能,找出毒力因子和具有保护宿主功能有的因子,为预防、治疗结核病找到更加理想的途径。现将目前使用的主要技术及取得的结果综述如下。

2.1 DNA 芯片(DNA chip)

美国 1998 年 6 月 29 日宣布正式启动“BIOCHIP”计划。具体做法是:用机器人在固相支持物(硅片、玻片、聚丙烯、尼龙膜等)上密集成千上万的特异寡核苷酸探针,抽提不同条件下样本 mRNA,并以荧光或同位素标记,与探针杂交,共聚焦荧光显微镜扫描,记录杂交结果,分析杂交位点及其信号强弱,找出差异表达的基因(是否表达,表达多少)。这可用于基因诊断、设计基因药物等。

用含代表 3924 个开放阅读框中 97% 开放阅读框的 DNA 芯片监测 MTB 对 INH(异烟肼)反应的基因^[4],发现 INH 诱导表达的基因包括一个操纵子基因簇(共 5 个基因,fbpC,efpA,fadE23,fadE 24,ahpC)。生物合成途径未受到明显的影响。

利用 DNA 芯片,比较 *M. tuberculosis*H37Rv, *M. Bovis*, *M. Bovis* BCG 三种菌株之间的差异,研究 BCG 保护效果不一致的原因,得到了一些有意义的结果。试验所用的 DNA 芯片有 4896 个点,代表 3924 开放阅读框中 3902 个开放阅读框(99.4%)。*M. bovis* 比 H37Rv 少 11 个区(包括 91 个开放阅读框),具有 *M. tuberculosis*H37Rv 所缺少的 5 个区(包括 38 个开放阅读框),而这恰恰为所有 BCG 所缺少。这表明,BCG 菌株一直在不断演变。*M. bovis* 比 H37Rv 缺失了 3 个磷脂酶 C 基因(plcA,-B,-C),磷脂酶 C 是 *Pseudomonas aeruginosa* 引起的条件性肺炎感染中的毒力因子,*M. tuberculosis* plcA-plcC 基因簇可能在临幊上具有相似的重要作用。以 H37Rv 序列为參照进行的分析仅仅是一个方面。

2.2 体内表达技术(*in vivo* expression technology)^[5]

该技术首先要建立 DNA 库,然后将 DNA 片段与某些缺失了启动子的基因连接,只有插入的 DNA 片段具备启动子活性,菌才能在宿主体内生存。识别这些基因,可以初筛不同入侵阶段表达的基因及其对致病的影响程度、宿主如何控制致病菌基因的表达。现在正在进行的有:MTB 转录细胞器(transcriptional apparatus)对体内胁迫的反应,如铁调节子 IdeR 与其靶子相互作用调控铁的代谢,Sigma 因子的差异表达等。

2.3 基因中断(gene disruption)

这是后基因组研究中强有力分子遗传学工具,主要有两种方法。即签名标记转座子突变(Signature-tagged transposon)和等位交换(AAllelic exchange)

原理:能够作标记的可以是带不同特征的转座子(带标记的 DNA 片段称为 tag,一般约 80bp,分为两部分:臂和可变区,臂位于两端,固定,供设计 PCR 引物;可变区位于中间,ATGC 四种碱基随机组成,理论上具 10^{12} 个互不相同的序列)。用带-tag 的转座子或质粒构建不同基因被中断的突变株,所获突变株应各有一个特异的 tag。用一组带不同 tag 的突变株接种动物模型(通常是鼠),筛选那些突变后毒力减弱或消失的突变株。

用带 *secB-ts* 反选择标记的自杀性质粒构建突变株,该方法将 *sacB* 的反选择特性与 MTB 的温度敏感型特性结合,可以在 39℃,蔗糖培养基上有效反选择丢掉载体的插入突变株^[9]。该方法构建的转座突变株库的容量为 10^6 ,远大于理论上要求的每个基因的插入数,筛选到了结核菌的 *lysA* 突变株,在合成培养基上的培养,该营养缺陷型突变株比正常的菌株对赖氨酸的需求要高 25 倍且依赖于 Tween-80,可以作为疫苗候选物,研究细胞壁生物合成及氨基酸代谢^[6]。另外一个实验在 1927 个突变株中筛选到 16 个的毒力衰减,大部分的插入子能在染色体上定位,且与脂代谢或者跨膜相关,4 个突变株与毒力基因簇相关(参与结核菌醇及其衍生物的合成)^[7]。

第一次获得了脂类生物合成、运输与 MTB 的毒力、菌落形态有关的遗传证据^[10]。细胞壁一直被怀

疑是 MTB 的毒力因子,PDIM(phthiocerol dimycocerosate—结核菌醇二分支菌酸,该成分与细胞壁有关,结构复杂,仅致病分枝杆菌中存在)是细胞壁的主要脂类成分之一。最近,用签名的标记转座子(Signature-tagged transposon)诱变,获得了 3 个 MTB 突变株:2 个为转座子插入到参与 PDIM 合成的基因(ppsA-E 编码多亚基的聚乙酰合酶,fadD28 编码转移分支菌酸酰基)。1 个为中断编码 PDIM 进行准确亚细胞定位所需跨膜蛋白的基因(mmpL)。这 3 个突变株在体外、肝和脾里的生长不受影响。只有在肺中生长受影响,说明合成和运输 PDIM 是肺中 MTB 才需要。

中断抗原 85 复合物表明:85 复合物是纤粘素(fibronectin)结合蛋白,具枝菌酸酰转移酶活性,是一毒力因子,可能参与结核分枝杆菌复杂细胞壁的生物合成^[18]。

2.4 基因互补法(gene complement)

将毒力株的基因克隆到同种或近种无毒、弱毒株,分析后者毒力是否改变,改变程度,筛选使后者毒力增强的基因。该方法找到了基因 EIS(enhance intracellular survival),仅存在于 H37Rv, *M. bovis*, 非致病菌 *M. smegmatis* 中不存在^[13]),厌氧氮还原酶(有助于 MTB 在厌氧的溃疡和肉芽生存^[12])、谷氨酸合成酶 A(可以使 *M. smegmatis* 在人 THP-1 细胞中的存活率明显上升^[11]);结核分枝杆菌之所以能够持续感染的关键酶-异柠檬酸裂解酶^[20]。

2.5 差示荧光诱导法(Differential fluorescent induction)

识别被宿主特定细胞如 macrophage 和特定条件诱导的细菌基因及其启动子。一般以 GFP(Green Fluorescent Protein 是一种水母蛋白,体内表达对宿主无毒)为标记。可以自动化,高通量筛选,不涉及代谢营养要求^[17]。基于全基因组基序(motif)的生物信息学方法与分子遗传学方法的结合^[14,15],用生物信息学方法比较 MTB 与其他细菌的海藻糖生物合成途径的相关基因发现,MTB 海藻糖生物合成途径可能有三:(1)从麦芽糖起始,(2)从糖原样 $\alpha(1 \rightarrow 4)$ 连接葡萄糖聚合物起始,(3)从葡萄糖-6-磷酸,UDP-葡萄糖起始。经抽提 MTB 的 DNA 分析证实了上述生物信息学方法的分析。同时也证明:海藻糖(胞浆中的游离二糖,胞壁糖组分),参与 MTB 感染所致组织损伤,在 MTB 的生理及病理中具有重要的作用。

2.6 2D-电泳结合 MS 进行蛋白质组(proteome)研究

蛋白质是生命功能的主要体现者,多数疾体现在蛋白质水平,蛋白质也是致病菌主要的致病因子之一。蛋白质组是 1994 年澳大利亚 Macquarie 大学的 Marc Wilkins 和 Keith Williams 提出,指胞内全部蛋白质的存在及活动方式。蛋白质组学以蛋白质组为研究对象,从蛋白质整体水平上认识生命活动的规律。基因表达和蛋白表达之间并不总是呈现良好相关性,蛋白质组可以分析非转录水平控制的细胞过程,补充 DNA 芯片等技术的不足。同基因组比,蛋白质组在时间、空间上具多样性,变化性大,研究的切入点多。蛋白质组学的研究有助于设计新的防治 MTB 的方案^[16,17]。

比较两无毒株(*M. bovis* BCG-Chicago/Copenhagen)与两有毒株(*M. tuberculosis*-H37Rv/Erdman)的蛋白质组^[16]发现细胞或上清液分别含 1800 点和 800 点,有 263 点为特征性(54 点来自上清液)。

以上方法为开发新的治疗结核药物和新型疫苗(包括治疗性和预防性疫苗)奠定了基础。尤其是为新型疫苗的发展拓展了新的空间。

3 新型疫苗

在后基因组研究的基础上,多种新型疫苗进入不同的实验阶段。

3.1 营养缺陷型活疫苗

利用中断技术构建 *Mycobacterium tuberculosis* *M. bovis* BCG 的嘌呤生物合成基因和赖氨酸生物合成途径被中断的营养缺陷型突变株^[19,6]。体外实验表明,两个突变株不能在鼠骨髓衍生的巨噬细胞中复制;在鼠及豚鼠体内模型中的毒性也减弱,在豚鼠中,可以诱导对 PPD(purified protein derivative-纯化蛋白衍生物)的强烈 DTH(Delayed type hypersensitivity-迟发型超敏反应);两个突变株对气溶胶态的 *M. tuberculosis* 有毒株攻击的保护效果和卡介苗一样。合理减毒株作为疫苗有广阔前景。

3.2 亚单位疫苗

所谓亚单位疫苗是指利用致病菌的亚单位，尤其是抗原表位作为疫苗，通常理想的亚单位疫苗应该由几个保护性抗原或抗原表位组成。保护效果优于卡介苗，又克服卡介苗的不足亚单位疫苗是卡介苗的理想替代品。寻找保护性抗原或表位的主要方法有：(1)比较蛋白组学^[16]，从上清和胞内复杂的蛋白混合物中寻找特异蛋白，尤其是上清中的许多分泌蛋白均可以诱发强烈的免疫反应；(2)构建结核分枝杆菌的基因组文库，用不同的T细胞(CD8+，CD4+)或不同遗传背景的动物模型从文库中筛选特定的表位^[22]。

以往困扰亚单位疫苗开发的主要问题之一是难以实现大规模生产。利用定点突变技术(site-directed mutagenesis)将其密码子修饰为宿主高频使用的密码，重组抗原85B在大肠杆菌中的产量提高了54倍，Ag85A和SOD提高了4~6倍，仍含有T、B细胞表位，蛋白产量的增加主要是因为翻译水平提高而不是转录水平提高^[21]。

3.3 DNA疫苗

DNA疫苗的实质是基因免疫。用含结核分枝杆菌38kd、Ag85T HSP65编码基因的质粒免疫，在动物模型中有很好的保护免疫结果^[23]。以人工合成的包括全部Ag85A的20肽刺激脾T细胞分泌IL-2、γ-IFN为指标，发现DNA免疫诱导产生的T细胞反应在强度和广度上均比皮下注射*M. tuberculosis*感染的效果好，CTL活性也上升。传统的DNA疫苗是逐个将DNA注射入肌肉。Johnston利用表达库免疫技术将传统的DNA疫苗概念进行了改变。他们认为：对于免疫复杂的机体，没有必要逐个去研究哪个蛋白是保护性的，免疫系统自己会选择。因此，他们抽提致病菌的DNA，随机打断，克隆片段至不同的质粒，每组约1000个质粒。将这些克隆质粒注射实验动物，随后，用致病菌攻击动物。不发病的动物即是受到保护的，找出有保护作用的一个或几个质粒。他们现正在对结核分枝杆菌的4000个基因进行筛选，一旦找到该质粒，即可将基因所编码的免疫原性蛋白放入任何释放系统^[24]。这将是对使用了近30年的鸡啄米式技术的重大改进，短时间内即可系统地对致病菌的相关抗原进行分类研究。当然，这只是加快了识别的速度、精度，进一步的工作仍离不开传统方法。

3.4 结合树突细胞的疫苗

严格的讲，这是否是疫苗还有争议。树突细胞(dendritic cell DC)是强有力抗原递呈细胞和细胞免疫反应的诱导者，在启动早期的抗感染免疫中具有重要作用。体外培养的DC经BCG感染，被诱导成熟，气管施用的该DC对空气途径感染的结核分枝杆菌有明显的保护效果^[25]。同皮下注射BCG需100d，感染BCG的DC免疫只需7d，而两者的保护水平相近。而且DC不受衰老的影响，因此，对于经常发生在壮年及老年的结核病，DC可能是老年人免疫治疗和疫苗开发的有力工具^[26]。

彻底根除结核病，新型疫苗具有举足轻重的地位。疫苗的成败最终是一个免疫学问题，后基因组研究能够深化对结核分枝杆菌毒力因子表达及其调控、结核发病机制和免疫机制更的认识，为开发成功新型、有效、实用的疫苗奠定坚定基础。现有100多种疫苗处于不同的研究开发阶段。我们目前的工作主要是利用2D电泳和质谱、毛细管电泳、HPLC等技术，从结核分枝杆菌独特的脂类及其与蛋白质组、基因组水平的联系，与宿主信号传导的关系角度，开展结核分枝杆菌新耐药机理研究，并利用上述技术和成果，开发新型治疗结核病的自主知识产权药物和疫苗。

致谢 本文在写作过程中得到了陆德如教授的热忱指导，特此致谢。

参 考 文 献

- [1] Declan B. *Nature*, 2000, 403(6771):692.
- [2] Cole ST, Brosch R, Parkhill J, et al. *Nature*, 1998, 393:537~544.
- [3] www.sanger.ac.uk/Projects/M-tuberculosis/Gene-list/.

- [4] Hilson M, DeRisi J, Kristensen HH, et al. *P N A S*, 1999, **96**(22):12833~12838.
- [5] Smith I, Dussurget O, Rodriguez GM, et al. *Tuber Lung Dis*, 1998, **79**(2):91~97.
- [6] Pavelka MS Jr, Jacobs W R Jr. *J Bacteriol*, 1999, **181**(16):4780~4789.
- [7] Gamacho LR, Ensergueix D, Perez E, et al. *Mol Microbiol*, 1999, **34**(2):257~267.
- [8] Dye C., Scheele S., Doliin P. et al. *J Am Med Ass*, 1999, **282**:677~686.
- [9] Pelicic V, Jackson M, Reyrat JM, et al. *P N A S*, 1997, **94**(20):10955~10960.
- [10] Jeffery S. C, Bing C, Michae 1, et al. *Nature*, 1999, **402**:79~83.
- [11] Miller B H, Shinnick T M. *Infect Immun*, 2000, **68**(1):387~390.
- [12] Weber I, Fritz C, Ruttkowski S, et al. *Mol Microbiol*, 2000, **35**(5):1017~1025.
- [13] Wei J, Dahl J L, Moulder J W. *J Bacteriol*, 2000, **182**(2):377~384.
- [14] Gomez M, Johnson S, Gennaro M L. *Infect Immun*, 2000, **68**(4):2323~2327.
- [15] Desmet K A, Weston A, Broun IN, et al. *Microbiol*, 2000, **126**(Pt I):199~208.
- [16] Jungblut P R, Schaible U E. *Mol Microbiol*, 1999, **33**(6):1103~1117.
- [17] Via LE, cucic R J. *Bacteriol*, 1996, **178**(11):3314~3321.
- [18] Armitige L Y, Jagannath C, Wanger A R, et al. *Infect, Immunity*, 2000, **68**(2):767~768.
- [19] Jackson M, Phalen S W. *Infect Immun*, 1999, **67**(6):2867~2873.
- [20] McKinney J. *Nature*, 2000, **406**:735~738.
- [21] Lakey D L, Voladri R K, Edwards K M, et al. *Infect Immun*, 2000, **68**(1):233~238.
- [22] Geluk A, Taneja V. *P N A S*, 1998, **95**(18):10797~10802.
- [23] Denis O, Tnghe A, Palfleet K, et al. *Infect Immun*, 1998, **66**(4):1527~1533.
- [24] Taubes G. *Science*, 1997, **278**(5344):1713.
- [25] Demangel C, Bean A G, Martin E, et al. *Eur J Immunol*, 1999, **29**(8):1972~1979.