

- chemiluminescence flow-through sensor for the determination of analgin. *Fresenius J Anal Lett*, 1999, **365** (4): 381~ 383
- 5 Huang Y M, Zhang C, Zhang X R, *et al.* Chemiluminescence analysis of menadione sodium bisulfite and analgin in pharmaceutical preparations and biological fluids. *J Pharmaceut Biomed*, 1999, **21** (4): 817~ 825
- 6 Zhao Y, Baeyens W R G, Zhang X, *et al.* Chemiluminescence determination of tiopronin by flow injection analysis based on cerium (IV) oxidation sensitized by quinine. *Analyst*, 1997, **122** (2): 103~ 106
- 7 Zhang X R, Baeyens W R G, Van der Weken G, *et al.* Chemiluminescence determination of some local anaesthetics. *Anal Chim Acta*, 1995, **303** (1): 137~ 142

Ultra weak Chemiluminescence Analytical Technology Principle and Application. ZHANG Zhong-Lun (*Institute of Biophysics, The Chinese Academy of Sciences, Beijing 100101, China*); YANG Feng-

Zhen, ZHANG Xin-Rong (*Department of Chemistry, Tsinghua University, Beijing 100084, China*).

Abstract Ultra weak chemiluminescence analytical technology has been developed rapidly in recent years. The technique in pharmaceutical analysis was introduced by using two examples. One is based on the chemiluminescence reaction of Ce (IV) -analgin-rhodamine 6G and the other is based on the auto-oxidation of analgin in micellar medium. An ultra weak chemiluminescence analyzer has been coupled to flow-injection device for the determinations.

Key words chemiluminescence, flow-injecton, analgin, Ultra weak chemiluminescence analytical technology

充分利用 EST 数据库资源

王华春 陈清轩

(中国科学院发育生物学研究所, 北京 100080)

摘要 表达序列标签 (EST) 数据库作为一种重要的基因组数据库, 已成为新基因的发现、基因表达及重组蛋白表达等研究的有力分子生物学工具. 介绍了如何充分利用 EST 数据库.

关键词 表达序列标签, 电脑, 克隆

学科分类号 Q7

近年来, 随着人类基因组计划 (HGP) 在世界范围内的开展, 生物信息学作为一门热门交叉学科, 不断地完善和发展起来^[1]. 作为一种强有力的工具, 它帮助我们对巨量的生物信息进行归纳和理解, 从而揭示生命的奥秘. 然而面对复杂和庞大的数据库, 如何驾轻就熟地获取我们所需要的信息, 已成为每个研究者关注的焦点.

表达序列标签 (expressed sequence tag, ESTs) 来自随机选取的 cDNA 克隆的末端序列. 简单地说, 一个 EST 就是对应于某一种 mRNA 的一个 cDNA 克隆的一段序列. 一般长于 150 bp 的 EST 在同源查找和基因作图中的作用较大. EST 方法发展几年来, 使分子遗传学家识别和克隆新基因的策略发生了革命性的变化^[2]. 现在在 NCBI 的 EST 数据库中 (dbEST) (<http://www.ncbi.nlm.nih.gov/dbEST/index.html>) 已录入的来自不同物种的不同组织的 EST 有 2 690 828 条. 当然, 人和小鼠的 EST 占绝大部分. 对绝大部分克隆来说,

EST 取自其 cDNA 的两端, 而且很容易钓取该 EST 相应克隆的其他序列进行分析 (<http://www.genome.wustl.edu/gsc/>). 另外也可以在几天内从它们的国际分送者 (distributor) 那儿获得大部分 EST 的 cDNA 克隆 (<http://www-bio.llnl.gov/bbrp/image/istributors.html>). EST 克隆已成为基因定性 (characterization)、基因表达和重组蛋白表达研究的有力的分子生物学工具. 多数情况下, 某一个 EST 克隆 (平均插入片段大小约 1.5 kb) 并不跨越某一基因的全长区. 然而, 现在有许多生物信息学工具可以帮助我们获取某一基因的全长转录本信息及在染色体作图中的精确定位. 所以, 当我们着手去分离某一转录本的剩余部分时, 必需先借助这些生物信息学工具, 核查一下该基因转录本全序列是否已经被分离.

Tel: (010) 82614427

收稿日期: 1999-07-22, 修回日期: 1999-12-24

1 利用 EST 数据库进行电脑克隆

作为电脑克隆的第一步, 即找到与待克隆基因相关的 EST. 在 dbEST 中找到某一 EST 的最有效方法即通过序列同源寻找. 一般借助 NCBI 的 BLAST 软件即可以达到这一目的 (<http://www.ncbi.nlm.nih.gov/BLAST/>). 从核酸序列或氨基酸序列出发进行查询, 可以证实 dbEST 中是否存在与待查询项完全相同或相似的 ESTs. 待查询项既可以是基因组序列, 以鉴定相应的转录本 (putative transcript) 是否存在 (比如用于定位克隆计划), 也可以是一个已知的基因, 以鉴定与该基因相关的序列的存在 (例如一个基因家族中的新成员). 应用 BLAST 软件进行查询还可以将查询范围限定成 dbEST 的一部分, 例如鼠源或人源 EST 部分. 也可以通过关键词对 dbEST 进行搜索. 例如当查询到某一 EST 后, 可以以该 EST 所在克隆的克隆编号作为查询项, 以期获取对应于该 EST cDNA 克隆的另一端的 EST 信息.

为了进行电子 cDNA 文库筛选, 可以把通过以上方法获得的 EST 序列作为查询项通过 BLASTN 软件对 dbEST 进行搜寻, 寻找 EST 重叠群, 进而通过计算机程序整合成几条更长的 EST 序列. 现在一些生物信息中心, 例如 TIGR (<http://www.tigr.org/tdb/hgi/>) 和 NCBI (<http://www.ncbi.nlm.nih.gov/UniGene/index.html/>) 已经对所有人类 EST 和已知基因序列进行了分析, 识别并聚类属于某一相同转录本的序列. 特别是 NCBI 定期进行该项分析, 并将分析结果存放入 UniGene 数据库^[3]. 至 1999 年 7 月 1 日止该数据库包括 74 939 个人基因转录本重叠群. 这个数目相当于收入 dbESTs 中的人基因转录本的大致数目. 通过输入某一个人 EST 的登录号 (accession number) 有可能搜索到属于该 EST 所在转录本的其他序列. 此外, 对某一 UniGene 重叠群来说, 所收录的关于它的表达数据, 在理论上均可用于虚拟的 RNA 印迹 (virtual northern blots) 或者说进行电脑基因差别表达研究.

但 UniGene 存在一个局限, 即不能展现重叠群的彼此毗连的 DNA 序列. 这方面的信息可通过 Tigem 的 EST 归纳软件而获得 (<http://gcg.tigem.it/cgi-bin/uniestass.pl>): 通过输入某一个人的 EST 登录号, 有可能获取该重叠群的所有序列, 并通过毗邻序列组装程序将其整合成一个更长的

EST 序列^[4]. 一个相似的, 但功能更高级的工具——ESTblast 软件已发展起来, 从人类基因组作图项目资源中心 (HGMP-RC) 的服务器上可以获得 (<http://www.hgmp.mrc.ac.uk/ESTBlast>). 以某一待查询项通过 BLAST 查询软件对 dbEST 进行查询, 查询的结果可以自动归纳为一个毗邻序列群 (contig), 而对这一 contig 的共有序列又可以进行自动的重复的序列比较, 进而有可能找到其相应的全长转录本. 此外, 查询结果还可以以图解形式展现, 包括: a. 其他 WWW 超链接网址, 例如 EST 序列来源、克隆大小、与 UniGene 超链接等; b. 检测潜在 ORF 的工具; c. 提供以最终的一致序列作为查询项对其他多个数据库进行 BLAST 搜索的可能性.

2 从 EST 数据库中筛选 SNPs

除了用于电脑克隆外, EST 组装程序及 EST-blast 也是检查和分析 DNA 序列变异频率或转录序列多样性 (polymorphism) 的有力工具. 当某一重叠群中包含大量的来自不同 cDNA 文库的 EST 时, 该方法尤为适用. 单核苷酸多型性 (single nucleotide polymorphism, SNPs) 是基因组变异最丰富的一种形式也是对复杂遗传性状进行定位的重要源泉^[5,6]. 各领域的遗传学专家均从 SNPs 的研究与应用中获益匪浅. 近两年对 SNPs 研究热情大涨主要缘于以下几个领域的同步成熟与会合, 即: a. 大规模基因组序列分析及其相关技术的发展; b. 生物信息学及计算机技术的飞跃; c. 对简单或复杂的疾病性状的分析; d. 全球人类种群遗传学的发展. 从 EST 数据库中筛选 SNPs 的主要优点在于, 这样筛选出来的 SNPs Marker 直接与基因的编码区相对应, 即得到的往往是基因相关标记 (gene-associated markers)^[7]. 筛选的主要步骤为 a. cDNA 序列的组装. 有很多功能强大的程序可以完成这个任务 (<http://genome.washington.edu>); b. 识别出候选 SNPs, 简单地说就是通过对大量重复的 ESTs 进行序列比较找到 SNPs; c. 对候选 SNPs 进行确认, 具体操作步骤可参见文献 [6]. 总之, 通过对大量 EST 数据的归纳整理是寻找人类 SNPs 以构建高密度遗传图谱的最经济的方法.

3 电脑基因作图

除了可以进行电脑基因克隆外, 也可以通过电脑获得某个给定的人 EST 的作图信息. 这些基因

作图信息可以从 UniGene 和人类基因转录本作图数据库 (Human Transcript Map database) 中获得 (<http://www.ncbi.nlm.nih.gov/genemap98>)。另外, 还有其他一些基因作图联合会的网址可以利用 (<http://www.ncbi.nlm.nih.gov/SCIENCE96/Restools.html>)。特别是白头研究所基因组研究中心 (<http://www.genome.wi.mit.edu/>) 和斯坦福大学人类基因组中心 (<http://www.shgc.stanford.edu/>), 他们将 EST 作图信息整合成非常精确的物理和遗传图谱。

随着模式生物基因组测序工作的完成, 可以预见未来几年 EST 在基因识别及阐明基因功能的工作中将发挥越来越大的作用。而且也会有更多的功能更强大的生物信息学工具问世, 使该领域的研究工作更方便迅捷。

参 考 文 献

- 1 陈润生. 生物信息学. 生物物理学报, 1999, 15 (1): 5~12
Chen R S. Acta Biophysica Sinica, 1999, 15 (1): 5~12
- 2 Adams M D, Kelley J M, Gocayne J D, *et al.* Complementary DNA sequencing: expressed sequence tags and Human Genome Project. Science, 1991, 252 (5013): 1651~1656

- 3 Boguski M S, Schuler G D. EST ablishing a human transcript map. Nat Genet, 1995, 10 (4): 369~371
- 4 Huang X. An improved sequence assembly program. Genomics, 1996, 33 (1): 21~31
- 5 Brookes A J. The essence of SNPs. Gene, 1999, 234 (2): 177~186
- 6 Wang D G, Fan J B, Siao C J, *et al.* Large scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. Science, 1998, 280 (5366): 1077~1082
- 7 Picoult-Newberg L, Ldeker T E, Pohl M G, *et al.* Mining SNPs from EST databases. Genome Research, 1999, 9 (2): 167~174

Get the Best Use of dbEST. WANG Hua Chun, CHEN Qing-Xuan (*Institute of Developmental Biology, The Chinese Academy of Sciences, Beijing 100080, China*).

Abstract dbEST is an important division of Genbank. EST clones represent useful molecular tools for new gene identification, gene expression and recombinant protein expression studies. A description of dbEST and the way to get the best of it were given.

key words expressed sequence tag (EST), computer, cloning

质谱在寡核苷酸药物质量控制中的应用*

杨秉呼 王升启¹⁾

(军事医学科学院放射医学研究所, 北京 100850)

摘要 介绍了质谱在寡核苷酸质量控制方面的应用。合成寡核苷酸及类似物作为反义治疗剂在病毒感染和一些癌症治疗方面有良好的前景。寡核苷酸作为药物, 其结构特性必须进行确证。寡核苷酸浓度和纯度的分析可使用色谱或电泳技术, 对寡核苷酸的碱基组成、序列、同一性, 修饰基团, 色谱或电泳分析方法无能为力。质谱的高鉴别能力使其能有效、灵敏、快速和精确地确定寡核苷酸的这些特性。

关键词 寡核苷酸, 质谱, 质量控制

学科分类号 Q503

常用的寡核苷酸基本上全是由固相 DNA 自动合成仪合成的, 合成仪上合成效率的监测不能作为可靠的寡核苷酸切割和脱保护最终质量指标。确定合成寡核苷酸质量的唯一可靠方法是检测合成的每一个核苷酸。常用的寡核苷酸分析方法有聚丙烯酰胺凝胶电泳 (PAGE)^[1]、毛细管凝胶电泳 (CGE)^[1~3]、高效液相色谱 (HPLC)^[3,4]、高效薄层色谱 (HPTLC)^[5,6]。这些分析方法可以确定寡核苷酸的组分和纯度, 对于长度、同一性、序列、

修饰基团等结构特性的检测无能为力。质谱 (MS) 测定质量的高鉴别能力使它能够快速提供寡核苷酸的这些结构信息^[7~20]。

* 国家“863” (102-08-04-01) 及军队“九五”重点课题 (96Z007) 资助项目。

¹⁾ 通讯联系人。

Tel: (010) 66932211, E-mail: sqwang@nic.bmi.ac.cn

收稿日期: 1999-07-14, 修回日期: 1999-11-01