

·成果简介·

信使 RNA 三维遗传信息的研究

刘次全

(中国科学院昆明动物研究所细胞与分子进化开放实验室, 昆明 650223)

[关键词] mRNA, 蛋白质, 三维遗传信息

遗传信息在从 DNA 到 RNA 再到蛋白质的传递过程中, 翻译过程是一个关键的步骤, 通过翻译, 从成熟的 mRNA 序列可以得到蛋白质的氨基酸序列, 一般认为遗传信息就是这样得以表达的。应该注意的是, 这样的信息仅只是一维序列上的信息。蛋白质作为生命活动的主要物质, 其生物功能取决于它的三维构象。目前比较普遍的观点认为, 蛋白质的三维结构由它的一维结构(氨基酸顺序)所决定, 即通常说的“序列决定构象”。迄今, 蛋白质结构的理论预测主要就是基于这个观点。然而, 近年来的研究发现, 存在着氨基酸序列完全相同但构象不同的蛋白质, 这使得“序列决定构象”的观点受到挑战。那么决定蛋白质构象的关键因素究竟是什么? 对此, 生物学家们从很多方面进行了研究, 但似乎都没能从根本上解决问题。是否存在某种决定蛋白质构象的遗传信息呢? 回过头来重新考察遗传信息的传递过程, 我们把注意力集中到了合成蛋白质的直接模板——信使 RNA (mRNA)。

在活细胞中, 生物大分子一般都是以折叠的形式存在的, RNA 亦不例外。实验已证实, RNA 分子的功能与它们的结构密切相关^[1,2]。尤其是 RNA 的三维结构对于其生物功能的体现以及基因的表达调控起着很重要的作用。那么, 作为合成蛋白质的直接模板, 成熟 mRNA 特定的三维结构是否对蛋白质的三维结构起着某种重要作用呢? 从我们最近的研究结果看, 答案是肯定的。

在探索成熟 mRNA 三维遗传信息的过程中, 我们归纳出以下基本逻辑推理: (1) 遗传信息可分为一维信息(序列信息)和三维信息(空间信息), 相应地遗传信息编码为线性密码和空间密码。(2) 有效信息通常是反馈的, 对“中心法则”所描述的单向遗传信息流我们持保留态度。(3) 在翻译过程中, mRNA 与由其指导合成的蛋白质之间, 在某一阶段可能有“共折叠”效应。

基于以上认识和推理, 我们对 72 个成熟 mRNA 的折叠结构与它们所编码的蛋白质三维结构进行了比较。蛋白质的结构数据取自 Brookhaven PDB 数据库。分析的样本涉及到 11 个 α 蛋白, 20 个 β 蛋白, 20 个 α/β 蛋白, 17 个 $\alpha + \beta$ 蛋白以及根据 SCOP 库定义的 4 个小蛋白, 这些蛋白质结构的分辨率均在 2.8\AA 以下。相应的 mRNA 序列取自 GenBank 核酸数据库。由于实验测定的 mRNA 三维结构数据极少, 所以 mRNA 的折叠结构是根据自由能最小的原则进行预测而得到的, 结构预测所用的能量数据由热力学实验测定。我们用的预测程序是

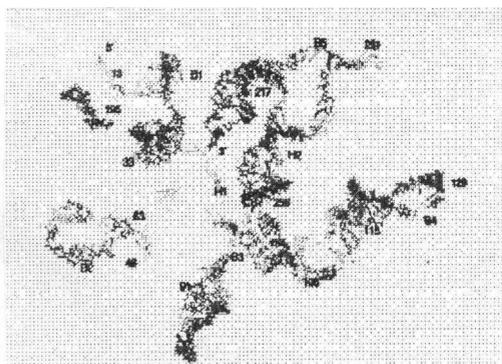
本文于 1997 年 5 月 15 日收到。

Zurker 的 MFOLD2.3。MFOLD 程序对短序列分子的预测准确率比较高, 例如对 tRNA 的预测准确率为 96%, 对 5s rRNA 是 88%。对于序列较长的 RNA 分子预测准确率大约是 70%^[3-5]。对 72 个样本进行分析研究后, 我们得到以下结果:

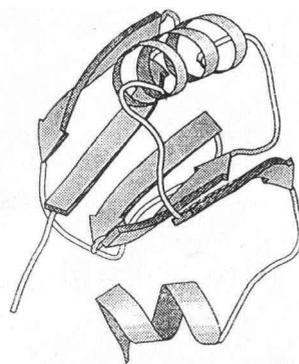
(1) mRNA 和蛋白质二级结构单元个数的相关性 RNA 二级结构主要包含发夹环、膨胀环、多分支环和碱基配对的茎, 由这些结构元素可以形成发夹或复合发夹结构。我们把这些发夹或复合发夹结构定义为 RNA 的二级结构单元。蛋白质的规则二级结构单元是螺旋(包括 α -螺旋和 3_{10} 螺旋)和折叠片。我们的研究表明, 大多数样本的 mRNA 二级结构单元个数与相应蛋白质的规则二级结构单元个数基本相同, 只有少数样本的差别略大一点。两组数据总的符合率为 81.94%, 相关系数为 0.95。对于序列较短的样本, 符合率更大。考虑到 RNA 二级结构预测的准确率, 这种相关性是很好的。

(2) 蛋白质二级结构的转角区与 mRNA 的二级结构有关 我们对序列长度较短(140 个氨基酸以下)的蛋白质样本仔细分析了转角区的氨基酸序列所对应的 mRNA 中的碱基序列, 发现这些碱基序列大多位于 mRNA 二级结构的发夹环、膨胀环和多分支环的区域。在这些部位, mRNA 分子的主链除了有螺旋桨状的旋转弯曲外, 还有不同程度的弯折。

(3) mRNA 和蛋白质三维结构中相应结构单元在空间排布上的对应 RNA 二级结构的二维图形只能反映碱基的配对情况, 为了了解 mRNA 分子的主链在空间的走向, 我们将一部分 mRNA 的二级结构在 SGI 图形工作站上做了三维模拟。将这些 mRNA 的三维结构与相应蛋白质的三维结构对比后, 我们发现很多相应的结构单元在空间上的位置排布是一致的。例如, 蛋白质中的 β -折叠结构是一些在一级结构上不连续的氨基酸序列片段由于在空间位置上很靠近而形成的, 有些序列片段在一维序列上甚至相距很远。而在我们分析的 mRNA 三维结构中, 很多编码 β -折叠片的碱基序列片段亦有相应的特征。这是由于 mRNA 的折叠使得一些在一维序列上相距较远的片段在空间上被拉近了。见图 1。



(a) 1SVR 的 mRNA 二级结构三维显示(用 INSIGHT II 软件模拟)。图中 H 表示编码 α -螺旋的核苷序列; B 表示编码 β -折叠片的核苷序列; H 和 B 后的数字分别表示螺旋和折叠片的序号, 其它数字表示核苷顺序



(b) 1SVR 蛋白质的三维结构(用 MOLSCRIPT 程序画出)

图 1 1SVR 蛋白质的三维结构及其编码序列 mRNA 二级结构的三维显示

(4) 密码子在 mRNA 的二级结构上的分布与氨基酸的性质 蛋白质的氨基酸序列是从 mRNA 序列按三联体密码翻译过来的。对蛋白质结构的研究表明, 各种氨基酸在蛋白质结构

中的位置有一定的倾向性。我们分析了各种氨基酸的密码子在 mRNA 二级结构中的分布,发现密码子的位置分布有与氨基酸相应的特点。例如,疏水氨基酸的密码子绝大多数位于二级结构的茎区,而亲水的则主要位于环区。赖氨酸(Lys)的密码子 AAA 有 90% 以上位于环区,且其中的 40% 位于发夹环(即末梢环)区。在这种部位,mRNA 分子与溶剂的接触面较大,而赖氨酸的特点正好倾向于在蛋白质表面,且柔性大。再如,空间上距离靠近的两个半胱氨酸(Cys)能够形成二硫键,我们对于有二硫键的蛋白质样本分子对应的 mRNA 二级结构做了分析,发现形成二硫键的每对半胱氨酸的密码子在 mRNA 二级结构中一般都非常靠近。

从以上结果可以看出,mRNA 与其所编码的蛋白质的二级(以及三维)结构之间,的确存在着有意义的相关性。我们的发现为前面所作逻辑推论提供了支持,这说明 mRNA 分子不仅携带有一维遗传信息,而且还蕴含着三维遗传信息。一维序列上的遗传问题已被解决,余下的问题是三维遗传信息是怎样被传递到蛋白质的,以及它是如何控制蛋白质折叠的。这些问题的解决对于蛋白质的翻译机制、折叠机制、蛋白质结构预测和功能研究,以及生命的起源和进化等问题的研究将起到极大的促进作用。此外,以上结果还提示我们,用 mRNA 的折叠结构来研究蛋白质的结构及相关的问题不失为一种新的思路。

参加本项研究工作的还有云南大学,北京大学及中国科学院化学研究所的张静、来鲁华、李维忠、白春礼、李任植、王任小、彭守礼和唐有琪等人。

参 考 文 献

- [1] Sorensen MA, Kurland CG, Pedersen S. Codon Usage Determines Translation Rate in *Escherichia coli*. J. Mol. Biol., 1989, 207: 365 - 377.
- [2] Wikström PM, Lind LK, Berg DE, Björk GR. Importance of mRNA Folding and Start Codon Accessibility in the Expression of Genes in a Ribosomal Protein Operon of *Escherichia coli*. J. Mol. Biol., 1992, 224: 946 - 966.
- [3] Zurker M. In Computer Analysis of Sequence Data, Part II, Vol. 25, Griffin AM, Griffin HG, eds., CRC Press, Inc, Totowa, NJ (1994), 267 - 294.
- [4] Zurker M, Jaeger JA, Turner DH. A Comparison of Optimal and Suboptimal RNA Secondary Structures Predicted by Free Energy Minimization with Structures Determined by Phylogenetic Comparison. Nucleic Acids Res., 1991, 19: 2707 - 2714.
- [5] Zurker M. On Finding All Suboptimal Foldings of an RNA Molecule. Science, 1989, 244: 48 - 52.

STUDY OF THE THREE-DIMENSIONAL GENETIC INFORMATION OF MESSENGER RNA

Liu Ciquan

(Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, CAS, Kunming 650223)

Key words mRNA, protein, three-dimensional genetic information