

的碱基成分和假尿嘧啶核苷酸。

此外,合成的环核苷酸混合物以及鼠脑提取液中的 cGMP 和 cAMP 也可以用 HPLC 分离^[16]。

五、氨基酸及其它生物 活性物质的分析

近 20 年来,生物样品中氨基酸的测定大多使用氨基酸分析仪,这种仪器操作麻烦。最近报道,用反相 HPLC 分离和定量测定游离氨基酸,具有简单迅速、灵敏度高等优点。Fernstrom 等人^[17]用 HPLC 测定血清和脑脊液中游离氨基酸,首先把提取液中的氨基酸与邻苯二甲醛反应衍生,然后在 C18 反相柱同时进行分离,并用荧光检测器检测出谷氨酸、天门冬氨酸、甘氨酸、丙氨酸、蛋氨酸、亮氨酸、异亮氨酸、苯丙氨酸、丝氨酸、组氨酸、精氨酸、赖氨酸、色氨酸、缬氨酸和苏氨酸等 16 种氨基酸,灵敏度可达 10 微微克分子。

HPLC 测定脂肪酸酯^[18]、前列腺素^[19]、卟啉^[20]、尿酸^[21]、多胺^[22]、脂质^[23]、蛋白质、类固醇、碳水化合物、抗生素等也屡见报道。

此外,在生物测定或放射免疫测定之前也常用 HPLC 迅速部分提纯样品,以除去干扰物质、提高测定结果的准确性。

总之,在生物化学各个领域 HPLC 的应用越来越广泛,而且还将继续得到迅速的发展。

参 考 文 献

- [1] Flatmark, T et al.: *Anal. Biochem.*, **107**, 71, 1980.
- [2] Cross, A. J. et al.: *Life Sci.*, **28**, 499, 1981.
- [3] Hamaji, M. et al.: *J. Chromatogr.*, **163**, 329, 1979.
- [4] Beek, O, et al.: *Clin. Chim. Acta*, **79**, 149, 1977.
- [5] Jackman, G. P.: *Clin. Chem.*, **26**, 1623, 1980.
- [6] Martin, C. et al.: *J. Endocrinol.*, **77**, 67, 1978.
- [7] Molnar, I. et al.: *J. Chromatogr.*, **142**, 623, 1977.
- [8] Hancock, W. S, et al.: *Science*, **200**, 1168, 1978.
- [9] Ling, N.: *Biochem. Biophys. Res. Commun.*, **74**, 248, 1977.
- [10] Akagi, H, et al.: *Neuroscience Letters*, **20**, 259, 1980.
- [11] Loeber, J. G, et al.: *Biochem. Biophys. Res. Commun.*, **86**, 1288, 1979.
- [12] Graves, F. R, et al.: *J. Pharmacol. Exp. Ther.*, **206**, 492, 1978.
- [13] Rahman, M. D. K. et al.: *Life Sci.*, **28**, 485, 1981.
- [14] Brown, P. R. et al.: *Anal. Biochem.*, **99**, 1, 1979.
- [15] Vandenberghe, A. et al.: *Anal. Biochem.*, **107**, 369, 1980.
- [16] Krstulovic, A. M. et al.: *Clin. Chem.*, **25**, 235, 1979.
- [17] Fernstrom, M. H. et al.: *Life Sci.*, **29**, 2119, 1981.
- [18] Scholfield, C. R. et al.: *J. Amer. Oil Chem. Soc.*, **52**, 36, 1975.
- [19] Fitzpatrick, F. A, et al.: *Anal. Chem.*, **49**, 1032, 1977.
- [20] Evans, N. et al.: *J. Chromatogr.*, **125**, 345, 1976.
- [21] Molnar, I. et al.: *J. Chromatogr. Biomed. Appl.*, **143**, 391, 1977.
- [22] Adler, H. et al.: *J. Chromatogr. Biomed. Appl.*, **143**, 125, 1977.
- [23] Jungawala, F. B, et al.: *Biochem. J.* **155**, 55, 1976.

[本文于1982年3月29日收到]

基 因 组 结 构 分 析

——DNA 复性动力学的应用

沈 建 华

(中国科学院上海生物化学研究所)

DNA 双螺旋结构互补链之间的氢链在一定条件下可被破坏,形成单链,是为变性过程。

这些拆开的单链在另一些条件下,互补的碱基间又可重新建立起氢链联系。双螺旋结构得恢

复。这个过程称为复性过程。复性过程的动力学是研究有关生物基因组结构 (genome organization) 的有效手段。

DNA 复性动力学的基本公式

现在已知 DNA 复性过程作为一个双分子反应是符合二级反应动力学公式的。任何一个不包括重复顺序的基因组 (如 *E. coli* 等原核生物的基因组) DNA 的复性动力学都可以用一个单一的二级反应动力学公式描述:

$$\frac{c}{C_0} = (1 + K_2 C_0 t)^{-1} \dots \dots (1) \text{式}^{[1]}$$

式中: C_0 为单链的初始浓度; c 为 t 时间仍为单链部分的浓度; C_0 和 c 的单位都以 $\text{mole} \cdot \text{L}^{-1}$ 表示。实际上为了方便都按 C_0 (或 c) = $A_{260} \times 1.47 \times 10^{-4}$ 计算。 t 为复性反应时间, 单位为秒。 K_2 为二级反应速度常数, 单位为 $\text{mole}^{-1} \times \text{sec}^{-1} \times \text{L}$ 。

DNA 复性反应的机制现在已比较清楚了。整个反应主要由两个分步反应组成: 首先是单链分子间由于热运动造成分子碰撞, 存在于不同单链上的一对或数对互补碱基间建立了氢键联系, 形成一个“核” (nucleation), 这是一个慢反应, 其速度常数记为 K_N ; 接着是一个快速的“拉链反应” (zippering reaction), 就是以已形成的“核”区域为起点向两个方向, 两条链上一互补的碱基很快地依次建立起氢键联系。这样一个复性的 DNA 分子因而产生。必须指出“拉链反应”是一个可逆反应。在通常的反应条件下 (温度、盐度、粘度), 第一步成“核”反应是速度限制的一步, 整个二级反应的速度常数 K_2 主要地是由成“核”反应的 K_N 决定的^[2,3]。

DNA 复性反应速度常数 K_2 首先反映了基因组 DNA 本身的特征—复杂性。所谓复杂性就是指不同 DNA 顺序总的长度。对病毒或 *E. coli* 等没有重复顺序的基因组来说, 其基因组大小 (Genome size) 就是它的复杂性^[4]。此外 K_2 还取决于一定的复性反应条件。

下面我们逐一讨论这些决定复性反应速度常数 K_2 的因素。

1) 复性反应速度常数 K_2 和 DNA 复杂性 N 大小成反比, 即 $K_2 \sim N^{-1}$ 。

这一关系是整个复性反应动力学实验的基础, 正因为它的成立, 我们才能从动力学数据中得到一系列有关基因组结构的信息。

对无重复顺序的基因组 DNA 来说, 则基因组大小 (G_i) 有如下关系式:

$$\frac{G_i(A)}{G_i(B)} = \frac{K_2(B)}{K_2(A)} \dots \dots (2) \text{式}$$

由此可用复性动力学方法, 由一个已知基因组大小的 DNA 测另一个未知基因组的大小, 在实际上一般都以大肠杆菌 *E. coli* 基因组大小为标准测定其他基因组大小, 即,

$$N_i = G_{E.coli} \times \frac{K_{E.coli}}{K_i} \dots \dots (3) \text{式}$$

当未知基因组是不含重复顺序的原核基因组时, 则

$$G_i = N_i \dots \dots (4) \text{式}$$

2) 反应速度和 DNA 片断长度 L 的平方根成正比, 也即 $K_2 \sim L^{0.5}$ 。

因此在每一个实验中必须严格控制 DNA 片断长度, 否则会造成对实验结果的错误解释, 尤其是涉及到所谓的短周期相嵌 (short period interspersion) 的情况时更要慎重。另一方面, 正因为有 $K_2 \sim L^{0.5}$ 关系存在, 比较同一 DNA 在不同单链长度情况下的 K_2 值可以提供单拷贝顺序和重复顺序相嵌情况的信息^[5,6]。

3) 温度 T 对反应速度常数 K_2 的影响:

复性温度对复性反应速度常数的影响呈现一个钟型曲线 (bell-shaped dependence), 当 T 在 $T_m - 10^\circ\text{C}$ 至 $T_m - 30^\circ\text{C}$ 之间时曲线出现一个平坦的 K_2 最大值区域, 因此一般复性反应都选择在 $T = T_m - 25^\circ\text{C}$ 左右即 60°C 进行^[2,3]。

4) 溶液 pH 对反应速度常数 K_2 的影响:

溶液 pH 太高或太低都会改变 DNA 链上碱基的电离状态, 从而对复性反应速度产生影响。在一般情况下, 当 pH 处于 5.0 到 9.0 之间, 在磷酸盐缓冲系统中复性反应速度变化不大^[2]。

5) 溶液的离子强度对反应速度常数 K_2 的

影响:

在低离子浓度下(一般复性反应的介质条件),盐浓度对反应速度有显著影响,这可能是由于溶液离子强度改变引起碱基之间静电相互作用的变化^[2]。

为了比较不同的实验数值,目前所采用的标准复性条件是: 溶液为 0.12M 的 Na_2HPO_4 — NaH_2PO_4 缓冲液(pH6.8),反应温度为 60℃。如果反应不是在这个条件下进行,则必须校正,表观的 C_0t 值必须换算成相应于标准条件下的 C_0t 值,即所谓的等值 C_0t (equivalent C_0t)^[4]。

6) 粘度 η 对 K_2 的影响:

粘度对反应速度的影响表现为 $K_2 \sim \eta^{-1}$ 。在太粘的介质条件下,“拉链反应”由已配对的“核”(nucleation)部位向两个末端扩散时,链的舒展受到影响,从而影响整个二级反应的速度常数 K_2 值^[2]。

在一般情况下,反应环境的粘度差别不大,不存在对 K_2 的校正问题,但在高 C_0t ($C_0t > 10^3$ — 10^4) 情况下,这时 DNA 的浓度很大(可达 2mg/ml 以上),所得到的表观 K_2 值(observed K_2 value)必须作粘度校正^[6]。

综上所述,可知简单 DNA 的复性反应是一个二级反应,二级反应速度常数 K_2 首先反映了该 DNA 的复杂性。另外复性反应速度还与一定的反应条件,如片段长度,反应温度,离子强度,溶液粘度等有关。

从二级反应公式出发,设 $c = \frac{1}{2} C_0$, 即复性一半时,有 $C_0t_{\frac{1}{2}} = K_2^{-1}$ 成立,即 $C_0t_{\frac{1}{2}} \sim N_0$ 。

真核生物 DNA 的复性动力学

和原核生物不同,真核生物 DNA 中有重复顺序存在,因此包括不止一个动力学组份,每一个组份的复杂性和重复性次数都不相同,每个组份都有其特征的复反应动力学参数,因此不能象处理原核基因组情况那样,用一个单一的二级反应动力学公式来描述真核基因 DNA 的复性过程。

真核基因组中各动力学组份虽然共存于一

个基因组内,但在把 DNA 切短成一定大小的片断(约 250—400bp)可以认为各组份的复性是相互独立的,总的复性动力学曲线可以认为是各组份曲线的迭加。

这里还有一个因素必须加以考虑,在真核基因中有一部分复性很快的 DNA,在反应时间还不足以形成起码的链间(interstrand)复性之前($C_0t < 10^{-3}$),就有相当部分已复性了^[7]。一般的解释是,这部分 DNA 是由于在一条链上存在着方向相反的重叠顺序(foldback 或 inverted repetitive sequences)^[5,7,8],由于这些反向重叠顺序之间建立了氢链联系,形成链内(intrastrand)复性。这个情况不同于双分子反应,对于它们复性的动力学研究很少。这部分 DNA 记为 f_0 ,作为一个常数项处理。

综上所述,我们用以下多项式描述真核基因组 DNA 复性动力学行为。

$$\frac{c}{C_0} = \sum_{i=1}^n f_i (1 + K_i C_0 t)^{-1} \dots \dots (5) \text{ 式}$$

而且

$$f_0 + \sum_{i=1}^n f_i = 1.0 \dots \dots (6) \text{ 式}$$

式中: f_i 代表各动力学组份在基因组中的比例, f_0 为回折 DNA 所占的比例。 K_i 为各动力学组份各自二级反应速度常数。 n 为基因组所包括二级反应动力学组份数。

这里几点说明:

A) 根据已报道的情况来看, $2 \leq n \leq 4$ (n 取整数),大多数情况下, $n = 3$, 即除了回折重复顺序外,还有三个动力学组份存在。

B) 由于长期保温中 DNA 的破坏,也由于链切断过程中有部分 DNA 被切得太短,以致于无法为羟基磷灰石吸附,因此最终高 C_0t 下复性的 DNA 始终达不到 100%, 故(6)式左项有时小于 1.0。

C) 第(1)式也可写为:

$$\frac{c}{C_0} = \sum_{i=1}^n f_i (1 + f_i K'_i C_0 t)^{-1} \dots \dots (7) \text{ 式}$$

在此时所得到的速度常数 K'_i 为各该组份独立存在下复性时的反应速度常数,即所谓“组

分 K_i (pure K_i value), 而式 (5) 中的 K_i 为总的 K_i (total K_i value), 很明显有以下式子成立:

$$K_{\text{pure}} = K_{\text{total}}/f_i \dots \dots (8) \text{ 式}$$

根据以上多项式来描述复性过程, 至少可以得到如下一些有关真核基因组结构的信息:

A) 各动力学组份的比例 f_i 和各自的复性反应速度常数 K_i 。用非线性最小二乘拟合, 可以得到 f_i 和 K_i 的最佳值^[4,9]。当然对应于一定的最小均方差 RMS 值, 可能得到的不是唯一解^[10]。必须分离不同 C_0t 值的 DNA, 分别作各组份的复性曲线 (minicot curve)^[4,10], 再用以上数学处理, 最后证实几个待定值 (f_i 和 K_i) 的确切值。

B) 各重复顺序的重复频率:

我们假定单拷贝顺序组份中每个顺序在单倍基因组 (haploid genome) 中仅出现一次, 则, 重复顺序的重复频率 $F_{\text{rep.}} = \frac{K_{\text{rep.}}}{K_{\text{s.c.}}} \times 1 \dots \dots (9) \text{ 式}$

式中: $K_{\text{rep.}}$ 及 $K_{\text{s.c.}}$ 分别为重复顺序及单拷贝顺序 DNA 复性反应速度常数。

当然, 如果已知某一重复顺序组份 A 的复性反应速度常数 K_A 和重复频率 F_A , 以及另一重复顺序 B 的速度常数 K_B , 则据此同样可以得到组份 B 的重复频率 F_B :

$$F_B = \frac{K_B}{K_A} \times F_A \dots \dots (10) \text{ 式}$$

显然, (9) 式是 (10) 式的特殊情况。

C) 整个单倍基因组的大小 (genome size) G_{io}

已知关系式 $N \sim K_i^{-1}$ 成立。对原核基因组来说, 其复杂性也就是其基因组大小, 即 $G_i = N$ 。因此可用一个已知基因组大小的原核生物 DNA, 把它切成所要测定的真核 DNA 在复性反应中所用长度一样的长度, 在相同条件下复性 (最好作为内标准在同一反应体系内复性), 可以得到一条参比 (reference) DNA 的复性曲线, 这一过程可用二级反应动力学公式描述, 即 (1) 式。

$$\frac{c}{C_0} = (1 + K_{\text{ref.}} C_0 t)^{-1}$$

由于参比 DNA 的复杂性必须是已知的, 在大多数情况下选用 *E. coli* DNA^[11]。

已知单拷贝顺序的 K_i 和 f_i , 由于

$$G_i = N_{\text{s.c.}}/f_{\text{s.c.}} \dots \dots (11)$$

以及

$$N_{\text{s.c.}} = G_{E.coli} \times \frac{K_{E.coli}}{K_{\text{s.c.}} (\text{组份})} \dots \dots (12)$$

因此,

$$\text{真核 } G_i = G_{E.coli} \times \frac{K_{E.coli}}{K_{\text{s.c.}} (\text{总})} \dots \dots (13)$$

由于真核基因组中 $G + C/A + T$ 未必一定为 1.0, 因此单拷贝组份的反应速度常数 $K_{\text{s.c.}}$ 在未作碱基组成校正^[2]前, 所得到的复杂性仅是动力学复杂性, 而不是真正的分析复杂性或称顺序复杂性 (sequence complexity), 只有在 $\frac{A+T}{G+C} = 1.0$ 的情况下, 动力学复杂性和分析复杂性才等值^[2]。但是迄今所发表的工作中, 如果不说全部的话, 至少大部分均未作这项校正。

D) 每个动力学组份的复杂性 N_i :

显然

$$N_i = G_i \cdot f_i \cdot F_i^{-1} \dots \dots (14) \text{ 式}$$

对于单拷贝组份来说, 由于 $F_{\text{s.c.}} = 1$, 因此式子也就相应地简化为:

$$N_{\text{s.c.}} = G_i \times f_{\text{s.c.}} \dots \dots (15) \text{ 式}$$

当然也可用式 (12) 类似的形式计算复杂性 N_{io} 。

综上所述, 通过某一真核 DNA 的复性动力学分析, 我们至少可以从中了解这个基因组的大小, 有几个动力学组份存在, 每个组份所占的比例, 各个组份复性反应速度常数 K_i , 和复性完成一半所要的时间, 以及该组份内相同或相似顺序在单倍基因组内的拷贝数和各组份的动力学复杂性。

参 考 文 献

- [1] Britten, R. J. and Kohne, D. E.: *Science*, **161**, 529, 1968.

- [2] Wetmur, J. G. and Davidson, N.: *J. Mol. Biol.*, **31**, 349, 1968.
- [3] Wetmur, J. G.: *Ann. Rev. Biophys. Bioeng.*, **5**, 337, 1976.
- [4] Britten, R. J., et al.: *Methods in Enzymology* (Grossman, L. and Moldave, K. eds), Academic Press, New York, **29E**, 363, 1974.
- [5] Goldberg, P. B.: *Biochem. Genet.*, **16**, 45, 1978.
- [6] Murray, M. G., et al.: *Biochem.*, **18**, 5259, 1979.
- [7] Flavell, R. B.: *Ann. Rev. Plant Physiol.*, **31**, 569, 1980.

- [8] Walbit, V. and Goldberg, R. B.: *Nucleic acids in Plants* (Hall, T. C. and Dannies, J. W. eds), CRC Press, 1979.
- [9] Pearson, W. R. and et al.: *Nucleic Acid Res.*, **4**, 1727, 1977.
- [10] Cullis, C. A.: *Biochem. Biophys. Acta*, **652**, 1, 1981.
- [11] Cairns, J.: *Gold Spr. Harb. Symp. Quan. Biol.*, **28**, 43, 1963.

[本文于1982年8月23日收到]

对生物进化理论新发展的一点认识

刘 端*

(内蒙药物检验所)

现在得到大多数人公认的生物进化的理论是达尔文的自然选择学说与孟德尔遗传理论相结合所发展起来的“现代达尔文主义”或“综合进化论”。这个理论认为经过基因的突变。重组及生殖隔离等因素的相互作用,在自然选择的压力下,生物群体渐次分化,造成了种群的多样性和适应性的变化,即生物的进化。其核心仍是自然选择。

六十年代后期分子生物学的发展,从分子水平上揭示出了与自然选择学说不同的进化图象。日本的木村和美国的金和朱克斯相继提出了“中性突变漂变学说”(Neutral mutation random drift hypothesis),简称中性学说,或称为“非达尔文进化学说”,它的要点是:生物的突变大部分对于进化来说不是有利的,也不是有害的,而是中性的;生物进化的速率由中性突变的速率所决定,突变后的基因通过随机的“遗传漂变”在群体内消失或固定,自然选择只起次要的作用。

“中性突变”学说是在大量实验结果和数学模拟的理论运算的基础上提出来的,是从分子水平提出的关于生物进化的一种新学说,不涉及表现型。至于表现型的演化,中性学说的提出人之一、日本的木村认为,还只能用达尔文自然选择说来解释。而且他还说:“今后的问题在于分子的进化同表现型的进化,将怎样联系在一起”。中性学说被认为是向以自然选择为基础的达尔文进化论的严峻挑战,它提出了达尔文进化观点所完全不能解释的一些问题。

显然,围绕达尔文主义的争论,预示着生物学孕育着一场革命,这场革命将使人类对生物发展规律的认识将提到一个新的高度。以下就这个问题谈一下笔者的一些认识。

生物进化的根本原因在于生物内部,而不是在外

部。生命在按其自身规律向前发展中受到环境很大的影响。自然选择在其中起相当的作用,但选择的基础不只是结构基因的突变,而在于生命本身自控系统的变化。

生命区别于非生命的主要特点在于它是一种具有高度秩序的物质系统,在于组成自身的化学物质都遵循精密的法则,保持相互之间以及与周围环境协调的和有条不紊的相互作用,并能不断地自我更新、自我调节,自我复制、有选择地进行反应。这表明生命存在一个自控系统。比如,现在已经能人工合成蛋白质晶体,这说明生命物质与非生命物质之间没有不可逾越的鸿沟,但是人工合成的蛋白质只是有机物质,决不是生命。因为它不具备生物体内蛋白质分子的功能,没有一套指挥它行动的“自控系统”。这好比一台电子计算机,不但要有机器系统(硬件),还要有程序系统(软件);只有机器部分未编有程序设计的“计算机”不过是一堆金属。生命也是如此,没有控制系统的“生物体”只是一堆有机物质,高分子化合物。

一般认为,生命的起源和进化可以分为三个阶段,一是化学进化阶段,这一阶段中产生了氨基酸、核苷酸等有机分子、并可能合成蛋白质、核酸等大分子;二是分子自组织阶段,这个过程导致普适密码的统一细胞结构;三是生物进化阶段,在这一阶段中原始的单细胞系统逐步进化到复杂的,高度组织起来的多细胞系统^[1]。与此同时,生命的自控系统也在逐步发展和完善。生物体结构的有序性必然带来生物大分子活动的更为奇妙的有序性。这里必须说明的是生命的有序性

* 作者是内蒙古大学生物系78级毕业生。此文原是她在校时写的一篇学习心得,现略加删节发表。