

现代生物技术前沿

张今编著

进化生物技术

—酶定向分子进化

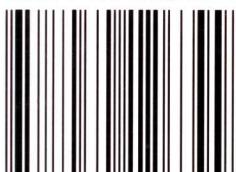


科学出版社
www.sciencep.com

现代生物技术前沿丛书

书名	书号 ISBN	定价
基因免疫的原理和方法	7-03-012588-6	38元
RNAi: 基因沉默指南 (影印版)	7-03-012685-8	65元
从基因到基因组 ——DNA技术概念和应用 (影印版)	7-03-012477-4	46元
结构生物学与药学研究	7-03-011688-7	48元
生物芯片分析(影印版)	7-03-012247-X	80元
DNA芯片和基因表达 (影印版)	7-03-012248-8	29元
计算分子生物学导论 (翻译版)	7-03-011493-0	36元
植物生物技术导论 (影印版)	7-03-012799-4	72元
蛋白质化学与蛋白质组学	7-03-012401-4	75元
进化生物技术 ——酶定向分子进化	7-03-012639-4	38元

ISBN 7-03-012639-4



9 787030 126399 >

生命科学编辑部

联系电话: 010-64012501
<http://www.lifescience.com.cn>
e-mail:info@lifescience.com.cn

ISBN 7-03-012639-4

定 价: 38.00 元

Q55
1280
2

现代生物技术前沿

张今编著

进化生物技术

——酶定向分子进化



SAH27/05



中纺院图书馆Z05995

科学出版社
北京

44064

内 容 简 介

本书较为全面地介绍了酶定向分子进化原理、策略、方法、应用以及发展趋势。主要内容包括定向分子进化思想、概念和原理；定向分子进化的基本策略和方法；定向分子进化的筛选/选择方法；核酶和脱氧核酶定向分子进化；蛋白质酶定向分子进化；外显子改组定向进化酶；抗体酶定向分子进化；蛋白质组、代谢途径、病毒和细胞的定向进化。

本书可供从事生物化学、分子生物学、生物工程学、化学和药学等相关专业的科研人员和教学工作者参考，也适用于综合性大学、医学和农业院校相关学科的高年级本科生和研究生阅读。

图书在版编目(CIP)数据

进化生物技术：酶定向分子进化 / 张今编著. —北京：科学出版社, 2004
(现代生物技术前沿)

ISBN 7-03-012639-4

I . 进… II . 张… III . 酶-定向分子-分子进化 IV . Q55

中国版本图书馆 CIP 数据核字(2003)第 120917 号

责任编辑：马学海 李久进 贾学文 / 责任校对：柏连海 / 封面设计：王浩
排版制作：科学出版社编务公司 / 责任印制：安春生

科 学 出 版 社 出 版

北京东黄城根北街16号

邮 政 编 码：100717

<http://www.sciencep.com>

新 蕉 印 刷 厂 印 刷

科学出版社发行 各地新华书店经销

*

2004 年 4 月第 一 版 开本：B5(720×1000)

2004 年 4 月第一次印刷 印张：16 1/2

印数：1—3 000 字数：309 000

定 价：38.00 元

(如有印装质量问题，我社负责调换〈新欣〉)

序

酶定向分子进化旨在模拟突变、重组和选择的自然进化机制，体外进化新酶。定向分子进化实质上是达尔文进化论在分子水平上的延伸和应用，是一个新兴的技术科学领域。过去 10 年，特别是近两年，定向分子进化在生物分子结构与功能研究方面，在工业、农业和医药等领域呈现强劲的发展势头。这表明进化论不再仅仅是一种理论，而是逐渐发展为一门应用科学——“进化生物技术”。进化生物技术关系人类命运（例如，HIV 是世界上最积极进化的生物体，如何用进化生物技术制服它；如何用进化生物技术战胜细菌耐药性；如何用进化生物技术解决生物品种灭亡等重大问题），定向分子进化则是进化生物技术中的关键技术。

张今教授编著的《进化生物技术——酶定向分子进化》一书对酶定向分子进化原则、策略和方法，现状和发展趋势进行了比较全面的介绍。他在广泛收集国内外资料的基础上，结合自己多年的实践经验，使本书内容具有新颖性和实用性，对读者颇有参考价值和启迪作用。我相信《进化生物技术——酶定向分子进化》一书的出版将会推动我国进化生物技术的发展。

孙曼寧

前　　言

定向分子进化(directed molecular evolution)又称实验室进化(laboratory evolution)或进化生物技术(evolutionary biotechnology)。它旨在体外模拟自然进化机制(突变、重组和选择),使进化过程朝着人们需要的方向发展。定向分子进化的实质是达尔文进化论思想在分子水平上的延伸和应用,用进化和组合技术研究和解决复杂生物学和化学问题已成为化学和生物学迅猛发展的领域之一。

定向分子进化的思想是1967年由Spiegelman及其同事提出的,他们当时定向进化的目标是RNA(Spiegelman S et al. Proc Natl Acad Sci USA, 1967, 58: 217~224)。1984年Eigen和Gardiner提出了分子进化的理论(Eigen M, Gardiner W. Pure Appl Chem, 1984, 56: 967~978)。1993年Arnold研究组首先应用定向分子进化的原理改进酶,提出定向分子进化的方法——易错PCR(error-prone PCR, epPCR)(Chen K, Arnold F H. Proc Natl Acad Sci USA, 1993, 90: 5618~5622)。1994年Stemmer发展了一种方法,称之为DNA改组(DNA shuffling),有力地推动了定向分子进化的发展(Stemmer W P. Nature, 1994, 370: 389~391)。现在定向进化已由基因和蛋白质分子向基因组和蛋白质组、代谢途径和病毒,甚至向细胞定向进化方面努力延伸,细胞的定向分化和定向进化可能是生物工程的未来。

近10年,定向分子进化在生物分子结构与功能研究方面,在工业、农业和医药等领域已展示出广阔的应用前景,高通量筛选是发现新酶的重要工具。可以预见,在不久的将来,定向分子进化的发展将使生物催化剂及催化过程以崭新的面貌出现在生命科学和生物技术“世界”。

本书的意图是向读者介绍定向分子进化原理、策略、方法、应用以及发展趋势,以期促进定向分子进化在我国的发展。本书1~7章围绕核酸和蛋白质定向分子进化展开,第8章介绍蛋白质组、代谢途径、病毒和细胞定向进化。本书编写有两个指导原则,一个是讨论一般原理和采用特殊实例并举的原则;另一个是讨论学科交叉和技术集成在定向分子进化领域的现状。每章末所引参考文献尽量选择最近发表的文章和能提供更广泛的文献目录的文献,便于读者进一步阅读。在编写过程中,力求在有限篇幅中尽可能介绍基本原理,同时也力求最大限度反映出近年来本领域的进展。在内容取舍方面,既注意到避免与相关著作内容重复,又尽可能自成系统。本书从基因和蛋白质→基因组和蛋白质组→代谢途径、病毒和细胞,结合自己的实践经验(本课题组自1990年至今主要从事酶定向分子进化方面的研究工作),比较详细地介绍了酶定向分子进化的理论和实践。

由于本人学术水平有限,错误和欠妥之处在所难免,诚恳地希望广大读者批评指正,不胜企盼之至。

在本书编写过程中,美国西南医学中心张雷博士、加州大学孔祥铎博士、麻省理工学院(MIT)刘文胜博士提供了有效的文献资料;书稿的电脑输入、图表制作和校对由孙妍红工程师、李爽博士(参加第2章和第3章部分编写)和李宾硕士完成,特此致谢!

中国科学院院士孙曼霁先生为本书作序,在此表示深深的谢意!

吉林大学分子酶学工程教育部重点实验室

张今

2003年8月18日

目 录

序

前言

第1章 定向分子进化的思想、概念和原理	(1)
1.1 引言	(1)
1.2 体内进化	(2)
1.3 体外进化	(4)
1.4 <i>in silico</i> 进化	(8)
参考文献	(9)
第2章 定向分子进化的基本策略和方法	(11)
2.1 引言	(11)
2.2 基本策略	(11)
2.3 基本方法	(12)
2.3.1 易错 PCR	(12)
2.3.2 DNA 改组和簇改组	(13)
2.3.3 外显子改组	(16)
2.3.4 合成改组	(16)
2.3.5 体外随机引发重组	(17)
2.3.6 交错延伸	(18)
2.3.7 酶法体外随机-定位诱变	(19)
2.4 非同源基因改组	(19)
2.4.1 过渡模板随机嵌合	(19)
2.4.2 增长截短	(20)
2.4.3 非顺序同源蛋白质重组	(22)
2.4.4 合理设计与随机组合方法	(22)
2.4.5 模块组装	(24)
2.4.6 任意数量的碱基随机插入和缺失	(26)
2.4.7 片段突变	(28)
2.4.8 设计的寡核苷酸组装	(28)
2.4.9 体外区室化	(31)
参考文献	(31)

第3章 定向分子进化的选择/筛选策略和方法	(34)
3.1 引言	(34)
3.2 核糖体和 mRNA 展示技术	(36)
3.3 噬菌体展示技术	(39)
3.3.1 噬菌体展示作为酶库筛选的工具	(39)
3.3.2 噬菌体-酶	(43)
3.3.3 噬菌体-酶的选择	(44)
3.4 酵母表面展示系统	(57)
3.5 酵母 n-杂交系统	(59)
3.5.1 酵母双杂交检验	(59)
3.5.2 酵母双杂交系统	(61)
3.6 细菌双杂交(B2H)系统	(66)
3.7 蛋白质互补检验	(67)
3.8 酵母 n-杂交系统的应用	(68)
3.8.1 蛋白质-蛋白质相互作用	(68)
3.8.2 蛋白质-DNA 相互作用	(68)
3.8.3 蛋白质-RNA 相互作用	(68)
3.8.4 蛋白质-小分子相互作用	(68)
3.9 环境库的构建和筛选	(69)
参考文献	(73)
第4章 核酶和脱氧核酶定向分子进化	(76)
4.1 引言	(76)
4.2 核酶	(81)
4.2.1 锤头型核酶	(81)
4.2.2 发夹型核酶	(84)
4.2.3 蛋白质-RNA 复合酶(RNaseP)	(84)
4.2.4 组 I 内含子和组 II 内含子	(86)
4.2.5 环状核酶	(87)
4.3 脱氧核酶	(89)
4.3.1 10-23 结构脱氧核酶	(90)
4.3.2 8-17 结构脱氧核酶	(90)
4.3.3 手枪型结构脱氧核酶	(91)
4.3.4 “二分”型结构脱氧核酶	(91)
4.3.5 环状结构脱氧核酶	(92)
4.4 核酶和脱氧核酶的体外进化	(93)

4.4.1 拓展核酸顺序空间的新方法	(93)
4.4.2 体外选择别构核酶	(97)
4.4.3 核酶的连续体外进化	(101)
4.4.4 核蛋白酶的体外选择	(102)
4.4.5 延伸蛋白质编码的核酶体外选择	(104)
4.4.6 体外选择环状核酶	(107)
4.4.7 核酶/脱氧核酶的催化潜能和进化策略	(110)
参考文献	(113)
第 5 章 蛋白质酶定向分子进化	(118)
5.1 引言	(118)
5.2 自然界蛋白质进化机制	(119)
5.2.1 基因复制	(119)
5.2.2 串联复制	(120)
5.2.3 环状变换	(121)
5.2.4 寡聚合	(122)
5.2.5 基因融合	(122)
5.2.6 结构域募集	(123)
5.2.7 外显子改组	(124)
5.3 定向分子进化	(125)
5.3.1 组合定向进化	(125)
5.3.2 模块定向进化	(130)
5.3.3 杂合酶	(140)
5.3.4 定向共进化	(145)
5.3.5 适应性进化	(146)
5.3.6 稳定性和活性的进化	(150)
5.3.7 底物特异性的进化	(151)
5.3.8 对映体选择性的进化	(155)
5.3.9 $(\beta/\alpha)_8$ -桶酶的功能进化	(170)
5.4 应用进化策略研究酶的结构和功能	(172)
5.4.1 应用进化策略研究分支酸变位酶的结构和功能	(172)
5.4.2 应用进化策略研究 DNA 聚合酶的结构和功能	(178)
参考文献	(184)
第 6 章 外显子改组定向进化酶	(186)
6.1 引言	(186)
6.2 外显子改组的分子机制	(187)

6.3 蛋白质结构域是运动的功能模块	(189)
6.4 外显子改组产生模块酶	(192)
6.4.1 血液凝集和血纤维溶解的蛋白酶	(192)
6.4.2 模块交换的原理	(196)
6.4.3 模块交换的机制	(198)
6.5 体外外显子改组	(199)
6.5.1 策略和方法	(199)
6.5.2 组合结构域重组	(201)
6.5.3 体外外显子改组实例	(203)
参考文献	(204)
第7章 抗体酶定向分子进化	(206)
7.1 引言	(206)
7.2 抗体酶定向分子进化策略	(207)
7.2.1 突变策略	(207)
7.2.2 组合抗体库策略	(208)
7.2.3 噬菌体抗体库	(209)
7.2.4 合成抗体库策略	(209)
7.2.5 结合荧光素 scFv 抗体库	(210)
7.2.6 T-细胞受体策略	(210)
7.2.7 提高催化活性策略	(211)
7.3 抗体酶定向分子进化筛选/选择策略	(211)
7.3.1 结合、催化和免疫组分策略	(211)
7.3.2 理论检验策略	(212)
7.4 抗体酶定向分子进化筛选/选择方法	(216)
7.4.1 直接筛选	(216)
7.4.2 化学筛选	(216)
7.4.3 遗传选择	(217)
7.4.4 动力学选择	(219)
7.4.5 应用反应化合物体外选择抗体酶	(221)
7.4.6 催化抗-DNA 抗体的筛选方法	(222)
7.4.7 选择性感染噬菌体	(223)
参考文献	(224)
第8章 蛋白质组、代谢途径、病毒和细胞的定向进化	(225)
8.1 引言	(225)
8.2 结构基因组定向进化的策略和方法	(226)

8.2.1 融合报告	(227)
8.2.2 宿主细胞应激反应报告	(229)
8.2.3 直接检验	(229)
8.3 基因组改组	(230)
8.3.1 基因组改组改进泰乐菌素生产菌表型	(231)
8.3.2 基因组改组改进乳酸生产菌表型	(235)
8.4 生物合成途径定向进化	(237)
8.4.1 β -葡萄糖苷酸酶定向进化	(237)
8.4.2 三嗪水解酶类定向进化	(239)
8.4.3 类胡萝卜素生物合成途径定向进化	(240)
8.4.4 L-甲硫氨酸生物合成途径定向进化	(243)
8.5 病毒定向进化	(245)
8.6 微生物细胞定向进化	(248)
参考文献	(250)

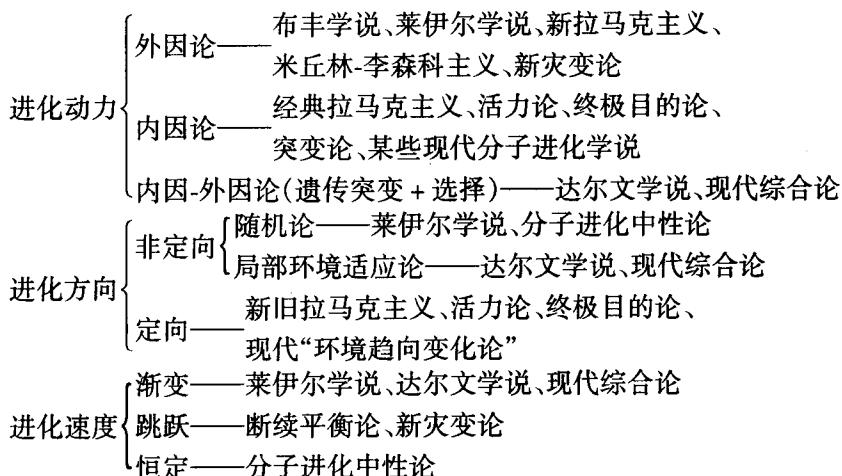
第1章 定向分子进化的思想、概念和原理

1.1 引言

过去10年,定向进化(directed evolution)在生物分子结构与功能研究方面及工业、农业、医药等领域的应用呈现出强劲的发展势头^[1]。特别是近两年,定向进化研究的论文以令人惊奇的数量和速度增加。

关于生物进化的新颖思路可以追溯到《物种起源》一书^[2]。在《物种起源》问世140多年中,进化理论已经有了很大发展。现代生物学不断地证实达尔文的惊人想像,尽管这位伟人根本不懂遗传学和分子生物学。

关于生物进化的学说可以围绕进化动力、进化方向、进化速度等3个方面归纳如下^[3]:



早期的进化学说基本倾向于决定论;近几十年的新学说基本倾向于随机论;近几年进化学说基本倾向于基因驱动(gene driver)说和新灾变论(新决定论)。100多年来,进化论的发展似乎在决定论和随机论之间摆动,达尔文学说似乎介于两者之间。

在分子水平上体外定向进化生物分子的思想是1967年Spiegelman及其同事^[4]提出的,当时他们定向进化的对象是RNA。1984年Eigen^[5]及1993年Kauffman^[6]提出分子进化的理论。1991年我们建立了“随机-定位诱变”定向进化的方

法^[7,8],又被称为 REM(random and extensive mutagenesis)法^[9,10]。1993年Arnold研究组^[11]应用分子进化的原理创造性地改进酶,提出易错PCR(epPCR)。1994年Stemmer^[12]发展了一种方法,称之为DNA改组(DNA shuffling)。1999年Stemmer等^[13]把DNA改组延伸到组改组(family shuffling),扩大了顺序空间(sequence space),进而提出分子育种(molecular breeding)。1998年在美国加利福尼亚州召开的“IBC第二届工业酶定向进化讨论会”,以及第14~16届国际酶工程会议都将定向进化列为会议专题。现在定向进化从基因和蛋白质→基因组和蛋白质组→代谢途径和病毒甚至细胞方向发展。

定向进化的实质是达尔文进化论在分子水平上的延伸和应用^[1]。定向进化(也称实验室进化或进化生物技术或定向分子进化)是在体外模拟突变、重组和选择的自然进化机制,使进化朝着人们需要的方向发展。定向进化虽然人工模拟自然进化,但两者却不相同:①进化动力不同。同源蛋白质的比较指出,保守突变是自然进化的驱动力^[14],而定向进化中非保守取代也可明显改进蛋白质的功能^[15];②进化方向不同。定向进化是人为制造特殊条件,模拟自然进化机制,使进化过程定向进行,主要是蛋白质突变的适应积累。定向进化的“适应度”(fitness,也称遗传漂移,drift)主要指生物分子的稳定性和活性。中性突变的组合可以产生新功能的嵌合蛋白质,称为“结构中性进化”^[16](construction neutral evolution)或“选择中性”^[17](selective neutrality)。自然进化是机体的适应突变(adaptive mutation)和中性突变(neutral mutation)的积累。自然进化的“适应度”主要指机体生存和繁殖的能力。③进化速度不同。定向进化只需几年、几月、几周、甚至几天的时间,而自然进化是自发出现的一非常缓慢的过程,需几百万年。自然选择使进化向有利于生物体适应生存和繁殖环境的方向发展。环境和适应方式的多样性决定进化方向的多样性。④定向进化的目标往往超越生物学意义的要求,以寻求自然环境中并不需要的功能,所有可能的氨基酸顺序,即顺序的多样性(X 个氨基酸链长的蛋白质顺序有 X^{20} 种)组成蛋白质的顺序空间。所有可能的顺序空间组成“适应度景观”(fitness landscape)。对于定向进化而言,将研究每一个性质或性质组合的变化,探索所希望的蛋白质在顺序空间的可及性^[17~19]。

1.2 体内进化^[20]

群体遗传学微分方程一般是为有性复制物种导出的,这种思想基本上以重组作为变异的主要来源,突变则被认为是特殊的事件。在生物多聚物进化设计中,与上述观点正相反,突变是变异的共同来源,重组则是特殊事件,例如基因改组。在排除无性情况下,通过微分适应度讨论选择的数学表达式如下:

$$\frac{dX_k}{dt} = X_k(f_k - \sum_{j=1}^n f_j X_j) = X_k(f_k - \Phi); k = 1, 2, \dots, n \quad (1)$$

式中, X_k 表示突变体的分数; f_k 是适应值; Φ 作为群体的平均适应值。方程(1)直接解释为:当突变体的 $f_k - \Phi$ 是正或它的适应值是在平均值以上,即 $f_k > \Phi$,那么 dX_k/dt 则为正,并且这个突变体频率增加;相反,如果 $f_k < \Phi$,那么相应突变体的分数将减少,并且最终接近零,突变体消失。这样选择具有最大适应值 $f_{\max} = \max\{f_k, k = 1, 2, \dots, n\}$ 的突变体,并且在足够长时间之后,惟有该突变体将在群体中存在, $\lim_{t \rightarrow \infty} X_{\max} = 1$ 。换句话说,如果时间相当长,所有不太适应的突变体将消失,群体将变为均一。当新种由于偶然事件处于消失的危险之中,选择过程和结果只是由适应值的差所决定: $S = f_k - f_{\max}$, S 值是通过选择优势突变体所需要的世代数来反映(图 1.1)。在自然选择中,优势突变体一般很少出现,新突变体在可以被群体中接收之前需要进化上千代。

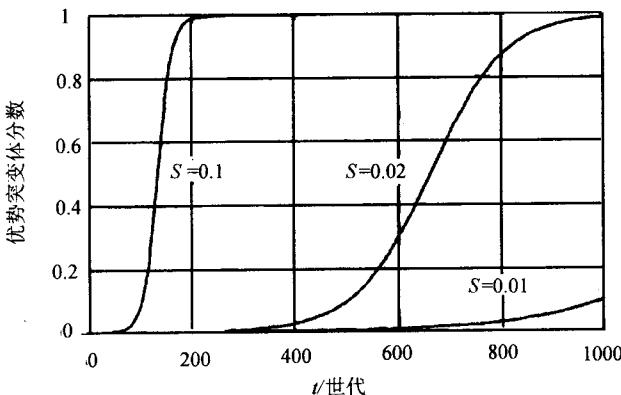


图 1.1 优势突变体的选择示意图

曲线表明根据方程 $X(t) = X_0 / [X_0 + (1 - X_0) \exp(-St)]$, 在 $N = 10000$ 个体的群体中突变体的选择和固定。时间 t 是以世代数或复制步骤计量, X_0 是群体中新突变体的最初频率, $S = f - f_{\max}$ 是它的选择优势。曲线表明在群体中采用单拷贝的初始条件, $X = 0.0001$

群体遗传学基本是 Motoo Kimura 思想的延伸。Kimura 指出适应突变是相当少的,大多数突变体在选择上是中性的,进化的主要作用是消除有害突变体。Kimura 观点是由蛋白质和核酸顺序比较分析资料支持的,这成为目前分子系统发育的基础。基因型是稳步地改变,并且当表型停滞时基因型也在稳步地改变。尽管分子资料以压倒的优势支持中性进化,但是第一个直接证据还是最近细菌进化

的实验:表型性质的改变,基因组 DNA 顺序也继续改变^[21];当表型停滞时,基因组 DNA 顺序以同样速度或者甚至在适应周期时改变更快^[22]。Kimura 途径是以选择过程随机描述为基础的:每个新形成的突变体达到一定的固定概率(随 S 值增加),它是相对于群体中当时优势类型的适应度。在中性情况下, $S = 0$, 群体通过随机漫步方式在顺序空间移动。随机漫步使基因型连续空间扩大(它与顺序空间概念相似)。20世纪 90 年代,中性进化的计算机模拟确定了 Kimura 观点。

1.3 体外进化^[4,18]

在试管中模拟自然进化的思想是 20 世纪 60 年代由 Spiegelman 及其同事在哥伦比亚大学提出的^[4]。他们的实验如图 1.2 所示。病毒 RNA 分子被转移到含有适于病毒 RNA 复制介质($\text{Q}\beta$ 复制酶及核苷三磷酸单体)的试管中,产生修饰的子代。70 次转移, RNA 合成的速率大约增加一个数量级(图 1.2)。这就是体外定向进化的经典实验。

在 Watson 和 Crick 揭示核酸分子的结构之后,进化的各种群体的空间性质是清楚的。顺序空间是指突变或重组的所有 DNA(或 RNA)顺序的分离空间(图 1.3)。顺序空间是以 Hamming 距离度量的超立方体^[23]。顺序空间有两个非常重要的性质:①它是具有各种长度基因组的高维物,λ 以核苷酸度量;②顺序空间的所有的点,即所有顺序是相等的。

关于分子进化, Eigen^[24]综合分子生物学和化学反应动力学的知识,提出通过动力学方程网络方式描述复制、突变和选择的模型,即

$$\frac{dX_k}{dt} = X_k(Q_{kk}f_k - \Phi) + \sum_{j=1, j \neq k}^n Q_{kj}f_j X_j; k = 1, 2, \dots, n \quad (2)$$

式中, Q_{kk} 表示 $I_k \rightarrow 2I_k$ 的概率, Q_{kj} 表示突变体 ($I_k \rightarrow I_k + I_j$) 的概率; Φ 与适应度权重相同。复制过程是一种产生正确拷贝产物平行的各种反应网络。方程(2)可以提供一种研究分子进化的实验方案。非中性复制集合的最适基因型称为主顺序 (master sequence)(图 1.4)。如果突变是一种相当复杂的事件,主顺序会被密切相关或不相关的突变体云 (mutant cloud) 所围绕。在适当条件下,主顺序及它的突变体云接近稳态的基因型分布,称为准种 (quasi-species)。准种概念的提出,对于了解进化是重要的。

突变频率只取决于由 I_k 到 I_j 突变必须交换的单体数,它通过 Hamming 距离计算 $d(I_j, I_k)$,有

$$Q_{kj} = q^\lambda \epsilon^{d(I_j, I_k)} = q^\lambda \left(\frac{1-q}{q} \right)^{d(I_j, I_k)} \quad (3)$$

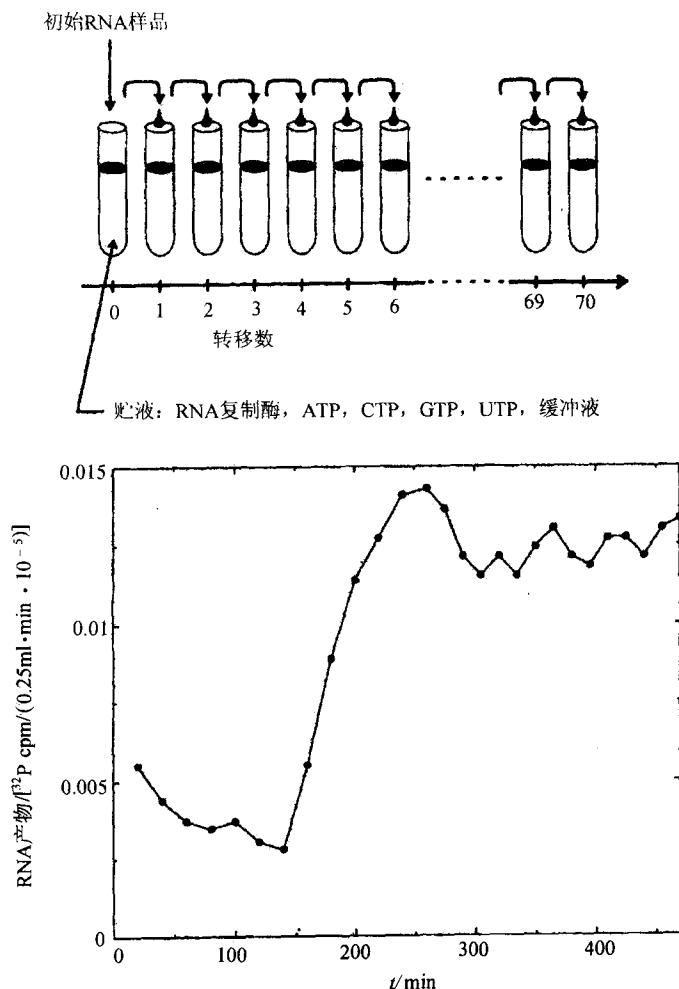


图 1.2 RNA 进化实验图

图上部分表示应用系列转移技术在试管中进化 RNA 分子。在一定的时间间隔之后，样品被转移到下一个含有新鲜贮液的试管中，补充 RNA 合成时所消耗的材料。贮液含有复制所需要的酶（例如 Q^B 复制酶）和活化的单体（ATP, UTP, GTP 和 CTP）。图下部分表示 RNA 合成速率。它是通过放射性 GTP 移入新产生的 RNA 分子来测量。复制速率表明逐步增加。早期降低是由于通过主顺序首先形成的准种适应度变低^[4]

式中，独立的参数 $\epsilon = (1 - q)/q$ 是突变速率和精确度之间的比率。在这个模型中，突变速率可以由 3 个量来表达，聚合物链长 λ ，复制的精确度 q （通常以 $p = 1 - q$ 表示）。