

# 酵母基因上游与内含子可能存在的 转录协同作用\*

张昆林<sup>1)</sup> 张静<sup>2)\*\*</sup> 罗静初<sup>1)</sup>

<sup>1)</sup> 北京大学生物信息中心, 北京大学生命科学学院, 蛋白质工程和植物基因工程国家重点实验室, 北京 100871;

<sup>2)</sup> 云南大学统计系, 云南大学应用统计中心, 昆明 650091)

**摘要** 前期研究表明, 酵母高转录基因内含子与低转录基因内含子的序列结构有较大差异, 高转录基因内含子中存在一些潜在的正调控位点, 而且这些内含子非常靠近基因上游区, 有的内含子甚至位于 5'-UTR 区. 这些结果提示, 高转录基因内含子可能参与转录调控, 并可能与上游转录调控有协同作用. 为探索内含子与上游区的协同作用, 将基因上游序列 (翻译起始点上游 800 bp 或两个相邻基因之间部分) 取出, 仍使用寡核苷酸频率比较方法, 抽提出高转录基因上游区可能的转录正调控元件, 这些元件与实验所得结果吻合得较好. 然后定义“寡核苷酸对”, 其中一个位于上游区, 另一个位于内含子, 对“最近距离”在一定范围内 (84 bp) 的寡核苷酸对进行分析, 抽提出匹配基因数显著高于随机匹配数的寡核苷酸对 (主要是四核苷酸对和五核苷酸对), 分析这些寡核苷酸对的相互作用模式 (位置分布及可能的作用因子, 如 RAP1, ABF1 和 TAF 等), 获得了酵母基因上游与内含子之间可能存在的一些转录协同作用模式. 这些结果有助于对基因转录调控机制的认识.

**关键词** 基因上游, 内含子, 寡核苷酸对, 协同作用, 酵母基因

**学科分类号** Q61

基因转录调控是整个基因表达调控体系的一个重要方面, 过去对基因转录调控的研究主要集中在基因上游. 近年来, 越来越多的实验表明内含子 (intron) 区也有增强子 (enhancer) 或启动子 (promoter), 或者二者皆有<sup>[1-4]</sup>. 我们前期的研究工作表明, 高转录酵母基因的内含子通常比低转录酵母基因的内含子长. 通过详细分析, 在高转录基因内含子中找到了一些可能的正转录调控位点<sup>[5,6]</sup>. 进一步分析还显示, 高转录酵母基因的内含子较多靠近基因的 5' 端, 有的甚至位于 5' 非翻译区 (5'-UTR), 相对地, 低转录酵母基因的内含子不具有上述明显的特征<sup>[7]</sup>.

真核基因的表达调控是一个很复杂的过程. 这个过程涉及到很多转录因子以及 DNA 上的调控元件 (位点) 之间的相互作用. 因此, 转录因子和调控位点是研究真核基因转录调控的两个重点. 一般来说, 转录过程中转录因子不是单独起作用. 一些转录因子聚集在一起形成复合体, 如 Enhanceosome<sup>[8,9]</sup>. 结合位点邻近的两个转录因子, 彼此之间往往会有相互作用. 这样的相互作用可使转录因子更加稳定地绑定在结合位点上. 这种现象称为协同作用 (cooperativity 或 synergy), 是组合调控 (combinatorial control) 的一种<sup>[10,11]</sup>. 实验表明, 协

同作用能极大地提高基因转录的效率<sup>[12]</sup>. 因此, 高转录基因内含子的位置偏向基因上游的现象似乎提示, 这些内含子与基因上游很可能存在着某些能增强基因转录的协同作用.

本文将探索高效转录基因内含子与基因上游调控转录位点间可能存在的协同作用. 对于两个调控位点间的协同作用来说, 位点间的距离一般不能太长<sup>[13]</sup>, 通常不超过 8 个螺距 (约 84 bp), 并为整数螺旋 (即 10.5 bp 的整数倍)<sup>[14]</sup>. 但是, 也有协同作用在 20 个螺距 (约 210 bp) 时被观测到的报道, 且位点间的距离不一定是整数螺旋<sup>[11]</sup>. 本文主要考虑两位点间的距离在 84 bp 以内时的协同作用. 当然, 更长距离上的协同作用将进一步研究.

## 1 材料和方法

### 1.1 材料

酵母内含子数据库 (YIDB, <http://www.imb-jena.de/RNA.html>) 列出了一些酵母基因的内含

\*国家自然科学基金资助项目 (30360027), 国家重点基础研究发展规划项目 (973) (2003CB715900).

\*\* 通讯联系人.

Tel: 0871-6541419, E-mail: zhangjing@ynu.edu.cn

收稿日期: 2004-07-07, 接受日期: 2004-07-31

子和转录频率. 按照转录频率, 挑出了高转录频率基因 (转录频率 > 30 mRNA/h) 和低转录频率基因 (转录频率 ≤ 10 mRNA/h), 与前期工作一致<sup>[9]</sup>. 鉴于主要研究对象为上游区和内含子的协同调控作用, 从 NCBI 酵母基因组数据库中取出包括上游序列的酵母基因, 共 67 个高转录基因和 71 个低转录基因 (内含子在 5'-UTR 的基因均未考虑). 根据 van Helden 等<sup>[10]</sup>的分析, 酵母基因上游的转录调控位点一般位于翻译起始点上游 800 bp 区域内. 但是, 我们得到的这两组基因中, 很多基因的翻译起始点上游 800 bp 区域中含有其他基因的编码区. 于是, 分两种情况来确定上游区: 若翻译起始点上游 800 bp 区域内不含其他基因的编码区, 上游区就取翻译起始点上游 800 bp; 若翻译起始点上游 800 bp 区域内含有其他基因的编码区, 上游区就取两个基因之间的部分 (小于 800 bp).

我们的目的是研究高转录基因的序列结构和转录调控机制, 而得到的高转录基因编码的几乎都是核糖体蛋白. 实验证明, 三类转录因子, RAP1 (repressor-activator protein)、ABF1 (ARS-binding factor) 和 TAF (TATA-binding protein (TBP)-associated factor), 参与核糖体蛋白基因的转录调控. 于是, 为了验证结果和研究协同作用的机制, 从专业的转录因子数据库 TRANSFAC<sup>[16,17]</sup>中取出以上 3 种转录因子的所有结合位点 (仅限于酵母的), 称之为“检验位点”(“test sites”).

### 1.2 提取上游区的潜在正调控位点和内含子中的潜在正调控位点

为了分析基因上游区与内含子的协同调控作用, 需要抽提出这两个区域中可能的正调控位点.

仍用与前期工作相同的方法<sup>[9]</sup>, 比较两类基因上游区寡核苷酸的使用频率, 寻找高转录基因上游区中可能的正调控位点. 用  $n_1(b)$  和  $n_2(b)$  分别表示寡核苷酸 (b) 在高转录基因上游区和低转录基因上游区中出现的次数;  $n_1$  和  $n_2$  分别表示确定长度 (4 或 5 或 6) 的所有寡核苷酸在高转录基因上游区和低转录基因上游区出现的总数, 即

$$n_1 = 2 \times \sum_1^{k_1} (L_{1i} - l + 1), \quad n_2 = 2 \times \sum_1^{k_2} (L_{2i} - l + 1)$$

$L_{1i}$  和  $L_{2i}$  分别表示高转录基因上游区和低转录基因上游区中第  $i$  个上游区的长度;  $l$  表示寡核苷酸的长度;  $k_1$  和  $k_2$  分别代表高转录基因上游区和低转录基因上游区的个数, 本文中  $k_1=67$ ,  $k_2=71$ . 因子 2 表示对寡核苷酸的统计是在 DNA 双链上

(方向均为 5'→3')进行的. 寡核苷酸 (b) 在两组上游区中的出现频率分别为  $n_1(b)/n_1$  和  $n_2(b)/n_2$ , 记为  $f_1(b)$  和  $f_2(b)$ . 采用单边假设检验法进行分析, 即零假设  $H_0: f_1(b) \leq f_2(b)$ , 其备择假设为  $H_a: f_1(b) > f_2(b)$ . 寡核苷酸 (b) 在两组上游区中出现频率差异的标准差为:

$$s = \sqrt{\frac{n_1(b) + n_2(b)}{n_1 + n_2} \left(1 - \frac{n_1(b) + n_2(b)}{n_1 + n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

计算  $u$  值:

$$u = \frac{n_1(b)/n_1 - n_2(b)/n_2 - \delta_1 \times 0.5/n_1 - \delta_2 \times 0.5/n_2}{s}$$

其中

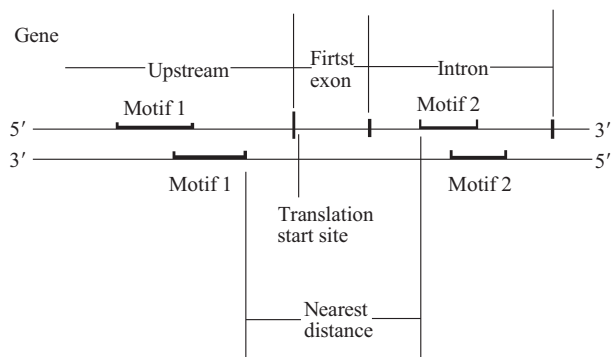
$$\delta_i = \begin{cases} 0, & \text{if } n_i(b) \geq 30 \\ 1, & \text{if } n_i(b) < 30 \end{cases}, \quad i=1,2$$

当  $n_1(b)$ 、 $n_2(b)$  小于 30 时, 进行了连续性矫正. 如果  $n_1(b)$  和  $n_2(b)$  都小于 5, 则直接用二项分布计算概率. 取显著水平  $\alpha=0.05$ , 当  $u > 1.64$ , 即  $P < 0.05$  时, 拒绝  $H_0$  而接受  $H_a$ . 即认为寡核苷酸 (b) 在高转录基因上游区中的出现频率显著高于在低转录基因上游区中的出现频率.  $u$  值越大, 差异越显著. 我们取所有  $u > 1.64$  的寡核苷酸作为结果.

用相同的方法, 通过比较两类基因内含子寡核苷酸的使用频率, 找出高转录基因内含子中潜在的正调控位点.

### 1.3 寻找潜在的协同作用模式

**1.3.1** 寡核苷酸对的“最近距离”的定义. 得到高转录基因两个区域可能的正调控位点之后, 就可以研究特定寡核苷酸对之间的协同作用. 定义一对寡核苷酸 (M1, M2) 在一条特定基因上的“最近距离”如下: a. 正反链都要考虑; b. 必须是 M1 位于上游, M2 位于内含子; c. 取最靠近第一外显子的 M1 和最靠近第一外显子的 M2; d. 最近距离为它



**Fig.1** The nearest distance (end-to-head) between two motifs, one in the upstream region and the other in the intron

们各自最靠近第一外显子的一端之间的碱基数 (图 1). 这种距离也可称为“尾-头”距离, 与实验上分析协同作用的研究一致<sup>[1]</sup>.

对于给定的长度 (本文为 84 bp), 当一对特定的寡核苷酸在一条基因上的最近距离小于或等于这一长度限制时, 我们说寡核苷酸对“匹配”这一基因.

**1.3.2** 抽取潜在的具有转录协同调控作用的寡核苷酸对.

具有协同作用的寡核苷酸对应该比较普遍地存在于高转录的基因中, 而在随机序列中较少出现. 即匹配的高转录基因数显著高于随机匹配数. 所以, 我们用与随机序列比较的方法来确定具有协同作用的寡核苷酸对. 具体方法如下:

记寡核苷酸对为(M1, M2), 其中 M1 位于上游区, M2 位于内含子. 考虑到统计显著性, 要求 (M1, M2) 至少匹配上 5 条高转录基因. 构造随机序列组如下: 用  $S$  表示高转录基因序列组,  $s_i$  表示其中的第  $i$  条序列, 则有  $S = \{s_i | i = 1, 2, \dots, 67\}$ . 对每一  $s_i$  做随机化处理 ( $i = 1, 2, \dots, 67$ ): 对上游区, 保持四种碱基含量不变, 将序列进行随机打乱 (shuffle); 对内含子做同样处理; 外显子不做处理. 这样, 就得到一个随机序列组  $S^{(1)} = \{s_i^{(1)} | i = 1, 2, \dots, 67\}$ . 上述处理进行  $n$  次, 就得到  $n$  组随机序列  $S^{(k)} = \{s_i^{(k)} | i = 1, 2, \dots, 67\}$ ,  $k = 1, 2, \dots, n$ . 设在 84 bp 范围内, (M1, M2) 匹配的高转录基因数为  $m_0$ , 在第  $k$  个随机序列组中匹配的基因数

为  $m_k (k = 1, 2, \dots, n)$ . 记  $\#(m) = \{m_k \geq m_0 (k = 1, 2, \dots, n)\}$  的个数, 即  $\#(m)$  表示 (M1, M2) 匹配的基因数大于或等于  $m_0$  的随机组的个数. 要检验 (M1, M2) 在高转录基因中的匹配数是否显著高于在随机序列组中的匹配数, 自然地有  $p = \#(m) / n$ . 取  $P < 0.05$  (要求在 95% 以上的随机序列组中, (M1, M2) 的匹配数小于在高转录基因中的匹配数), 抽提出潜在的具有协同作用的寡核苷酸对. 由于我们采用的是随机化的方法, 为了减少随机涨落的影响, 分别取  $n = 30, 100, 1000$ , 要求 3 种情况下的  $P$  值均小于 0.05. 取符合这样条件的 (M1, M2) 作为结果, 同时把  $n = 1000$  时的  $P$  值作为检验的  $P$  值. 我们对四、五、六核苷酸对都做了上述分析.

## 2 结 果

### 2.1 高转录基因上游区和内含子中的潜在正调控位点

我们首先抽提出高转录基因上游区和内含子中可能的正调控位点 (motif). 表 1 列出了四核苷酸和五核苷酸的结果 ( $u > 1.64$ ), 以黑体标注的是与实验得到的调控元件完全吻合的寡核苷酸 (通过对比“检验位点”). 限于篇幅, 我们没有列出六核苷酸的结果. 从表 1 可以看到, 大部分寡核苷酸都被实验所支持. 例如, 在上游区找到了 ACCCA, 这是一个很常见的调控元件, 被 Vignais 等<sup>[18]</sup>的实验所支持. 可见我们的结果可信度比较高.

**Table 1 The motifs (tetra-, penta-oligonucleotides) extracted from different data sets ( $u > 1.64$ )**

	Motifs in upstream regions	Motifs in introns
tetra-	ATGT TACC CCAA CCAT CCCA CCAG GTAC CACC ACCC TCCG CGGC CCGT TGGA GGAA AGGT GGGC TAGG CTAG GCCT CTCC GCGC CCGG GGCC	TTAA TAAA AAAT AATT TTAT TATT TGGT TTCA GAAT ATTA AGAC ATAT TATC CATG CACG GGTC
penta-	TATTG ATGTA ATACC CCATG ACATT TGTAC ACACC CACCC TGAAA AACCC ACCCA CCCAT CCAAG ATCCG GCCGT CCGTA CGTAC ACCGT CCGTC AAATG AGGTA CAAAC CGCTC CGGGC AAATT TATGG TCTAG CAGTG AGTGT CCTAG GTGGA GGATG ACGGA CTCCG TCCCA CCCAC CCTGG GGTGG CATA CCGCA GCACG CGCGC CCCC CCGGC GGCAG CCTAT GGGAG GCCTA CCTAC AGGCC GGCC GGCCT GCTAG TCCGC CCGCC CTCCA CGGCC GCTGG GGGCC	TAAAA AAATT TATTT ATTTG TAAAT ATTGA ATTTC TTTCA AGAAT AATTA TATCG ATTCA AATAT TATTTC CAAGC ATATC GATAA TCACG CACGC CTATT ACCAC TGCTA ATTAT ATAGT TATTA ACGTG GATTG TGGTG TAAAA GAATT

All motifs indicated by bold face agree with “test sites”. The reverse complements are not listed.

两类基因的上游区和内含子 GC 含量均较低, 其中高转录基因上游区 GC 含量约为 38%, 高转录基因内含子 GC 含量为 33%; 低转录基因上游区 GC 含量为 36%, 低转录基因内含子 GC 含量为 35%. GC 含量在上游和内含子中差异不大. 而不论是四核苷酸还是五核苷酸, 通过对比两类上游区和对比两类内含子得到的两组正调控位点的组成则有较大差异. 表 1 表明, 上游区的潜在正调控位点富含 G、C 两种碱基, 而内含子中的潜在正调控位点富含 A、T 两种碱基.

**2.2 潜在的转录协同调控作用模式**

表 2 和表 3 分别列出了与随机序列做对比得到的潜在的转录协同调控作用寡核苷酸对. 表 2 是四核苷酸对的结果, 表 3 是五核苷酸对的结果. 表中每一行都列出了潜在的协同作用寡核苷酸对、统计

检验的 *P* 值、匹配的高转录基因数、位于上游区的潜在正调控位点到翻译起始点的距离的平均、寡核苷酸对的最近距离的平均, 以及可能与潜在正调控位点结合的转录因子 (RAP1、ABF1 和 TAF, 分别简记为 R、A、T). 所有这些信息, 构成了若干协同作用模式. 例如: 表 2 第 2 行的协同作用模式说明, 位于上游的 CCAA 与位于内含子的 TAAA 可能具有协同作用; 检验的 *P* 值是 0, 即所有 1 000 个随机组中, (CCAA, TAAA) 匹配的基因数均低于匹配的高转录基因数; CCAA 到翻译起始点的平均距离是 20 bp, 到 TAAA 的最近距离平均是 52 bp; RAP1/TAF 可能结合在 CCAA 上, RAP1/ABF1 可能结合在 TAAA 上; 23 条高转录基因均具有这种协同作用模式 (图 2).

**Table 2 Tetra-nucleotide pairs and the related information with nearest distance less than 84 bp**

Motif 1	Motif 2	<i>P</i> value	Match	Up_dis mean	Dis mean	Factor 1	Factor 2
CCAA	TTAA	3.00E-2	13	21	53		A/T
	TAAA	0	23	20	52		R/A
	AAAT	0	23	21	58		R/A
	AATT	1.00E-3	16	22	62		A
	TTAT	1.70E-2	18	23	61	R/T	R/A
	TTCA	4.00E-3	15	17	57		R/A/T
	GAAT	0	19	18	63		R/A/T
	ATTA	1.60E-2	18	26	59		R/A/T
	CATG	9.00E-3	7	16	60		R/A
	CACG	0	8	19	68		R/A/T

“Motif 1” and “Motif 2” denote the motifs in upstream regions and introns respectively. “Match” denotes the number of genes matched by the motif pair. “Up\_dis mean” represents the average of the distances between Motif 1 and the translation start site. “Dis mean” represents the average of the nearest distances between Motif 1 and Motif 2. “Factor 1” and “Factor 2” denote the transcription factors which may bind to Motif 1 and Motif 2, respectively. “R” denotes RAP1, “A” denotes ABF1, and “T” denotes TAF, respectively. Similarly in Table 3.

**Table 3 Penta-nucleotide pairs and the related information with nearest distance less than 84 bp**

Motif 1	Motif 2	<i>P</i> value	Match	Up_dis mean	Dis mean	Factor 1	Factor 2
AAATT	TTAAA	1.90E-2	8	30	49	A	-
CAAAC	TATTT	1.00E-2	6	30	72	R/A	R/A
CAAAC	AATTA	2.30E-2	5	15	69	R/A	A

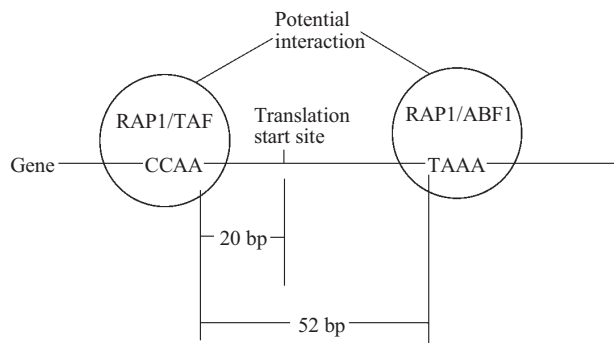


Fig.2 A synergistic pattern between two motifs “CCAA” and “TAAA” (Table 2)

在我们的结果中有两点值得注意. 其一是, 表 2 列出的潜在的具有协同作用的四核苷酸对, 大部分是由内含子中富含 A、T 的潜在正调控位点 (除了 CATG 和 CACG) 与上游区的 CCAA 组成, 而表 3 列出的潜在的具有协同作用的五核苷酸对, 是由内含子中富含 A、T 的潜在正调控位点和上游区的 AAATT 或 CAAAC 组成. 所以, 总体上看, 84 bp 范围内潜在的具有协同作用的寡核苷酸对, 可以分为 3 类: 第一类由内含子中富含 A、T 的潜在正调控位点和上游区由 A、C 两种碱基组成的潜在正调控位点构成, 共 10 对; 第二类由内含子中富含 G、C 的潜在正调控位点和上游区由 A、C 两种碱基组成的潜在正调控位点构成, 共 2 对——(CCAA, CATG) 和 (CCAA, CACG); 第三类是全由 A、T 组成的寡核苷酸对 (AAATT, TAAA). 另外值得注意的一点是, 有的协同作用模式在某些特定序列上可以合并, 如 (CCAA, TTAA) 和 (CCAA, TAAA), 从下面的例子中可以看到这一点.

将上游区和内含子所有潜在正调控位点在序列上定位会给出更多信息. 图 3 显示基因 YER074W 在翻译起始点附近的潜在正调控位点 (四核苷酸) 定位情况. 图 3 中第一外显子用方括号标明, 潜在正调控位点用大写表示. 可以看到, 上游区和内含子中都有离第一外显子较近的潜在正调控位点. 下划线标注的是本文找到的两个四核苷酸对 (CCAA, TTAA) 和 (CCAA, TAAA), 详细信息见表 2 第一、二行. 并且, 两个四核苷酸对可以合在一起构成 (CCAA, TTAAA). 在 CCAA 和 TTAAA 下面分别标注了可能与它们结合的转录因子, TTAAA 可能的结合因子取的是 A/T 和 R/A 的交集 A. 图 3 (CCAA, TTAA) 的最近距离是 24 bp, 比平均值 (53 bp) 要短, 原因是这一基因的第一外显子较短,

只有 3 bp. 包含 CCAA 上游区的潜在正调控位点, 以及包含 TTAA、TAAA 的内含子中潜在正调控位点都距离第一外显子最近.

5' -caagaaaCCAACATagattaagcagaa[-atg-]gtatgTTAAAaagtgcaga -3'  
R/T A

Fig.3 Illustration of tetra-nucleotide motif pairs with the nearest distance  $\leq 84$  bp in gene YER074W

The first exon (only 3 bp), which is just the start codon “atg”, is indicated as “[ ]”. The upper cases bases are potential motifs extracted by our method, and those underlined are two potential synergistic motif pairs in our results. The transcription factors which could bind to the two motif pairs are shown under the two motif pairs, respectively.

### 3 讨 论

很多实验证实, 真核基因内含子可以参与转录调控. 酵母中高转录基因内含子的序列特征和位置特征表明这些基因的内含子很可能参与转录调控. 一个很自然的问题是: 内含子如何参与调控? 一个可能的形式是上游区的调控位点与内含子的调控位点的协同作用. 这是很有可能的, 因为一个转录因子为了能稳定地绑定在 DNA 上, 往往会和附近的其他转录因子相互作用. 协同作用的前提之一就是结合位点在序列上要靠近. 实验也证明协同作用可以增强转录频率. 一个更深层次的问题是: 在酵母的高转录基因中是否存在一些特定的协同作用模式. 我们的研究给出了一个可供参考的答案.

详细的实验分析表明, RAP1 可以调控核糖体蛋白基因, 它的结合位点可以在基因间区, 也可以在编码区<sup>[9]</sup>. 除了 ACT1, 我们找到的酵母高转录基因都是编码核糖体蛋白的. 所以 RAP1 的结合位点应当是这些基因的一个普遍特征. 实际上, 我们发现每一条高转录基因的上游区都有 RAP1 潜在结合位点. 此外, 一些高转录基因的内含子中也有 RAP1 结合位点. RAP1 可能会和别的转录因子有协同作用. 我们的结果给出了一些可能的协同作用模式.

### 参 考 文 献

- Bhattacharyya N, Banerjee D. Transcriptional regulatory sequences within the first intron of the chicken apolipoprotein A I (apoA I) gene. *Gene*, 1999, **234** (2): 371~380
- Brinster R L, Allen J M, Behringer R R, *et al.* Introns increase transcriptional efficiency in transgenic mice. *Proc Natl Acad Sci USA*, 1988, **85** (3): 836~840

- 3 Chen J, Hayes P, Roy K, *et al.* Two promoters regulate transcription of the mouse folylpolyglutamate synthetase gene three tightly clustered Sp1 sites within the first intron markedly enhance activity of promoter B. *Gene*, 2000, **242** (1~2): 257~264
- 4 Surinya K H, Cox T C, May B K. Identification and characterization of a conserved erythroid-specific enhancer located in intron 8 of the human 5-Aminolevulinatase 2 gene. *J Biol Chem*, 1998, **273** (27): 16798~16809
- 5 张 静, 石秀凡. 酵母基因中转录正调控内含子序列特征的统计分析. *生物化学与生物物理进展*, 2003, **30** (2): 231~238  
Zhang J, Shi X F. *Prog Biochem Biophys*, 2003, **30** (2): 231~238
- 6 Zhang J, Hu J, Shi X F, *et al.* Detection of potential positive regulatory motifs of transcription in yeast introns by comparative analysis of oligonucleotide frequencies. *Comput Biol Chem*, 2003, **27** (4~5): 497~506
- 7 张 静, 石秀凡, 杨恒芬. 酵母内含子在基因序列中的分布对基因转录效率的影响. *生物化学与生物物理进展*, 2003, **30** (6): 945~949  
Zhang J, Shi X F, Yang H F. *Prog Biochem Biophys*, 2003, **30** (6): 945~949
- 8 Thanos D, Maniatis T. Virus induction of human IFN  $\beta$  gene expression requires the assembly of an enhanceosome. *Cell*, 1995, **83** (7): 1091~1100
- 9 Kim T K, Maniatis T. The mechanism of transcriptional synergy of an *in vitro* assembled interferon- $\beta$  enhanceosome. *Mol Cell*, 1997, **1** (1): 119~129
- 10 Carey M. The Enhanceosome and Transcriptional Synergy. *Cell*, 1998, **92** (1): 5~8
- 11 Carey M, Smale S T. *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques*. New York: Cold Spring Harbor Laboratory Press, 2001. 38~41, 466~467
- 12 薛 文, 王 进, 黄启来, 等. 真核基因转录激活的多位点协同调控. *生物化学与生物物理进展*, 2002, **29** (4): 510~513  
Xue W, Wang J, Huang Q L, *et al.* *Prog Biochem Biophys*, 2002, **29** (4): 510~513
- 13 Griffith J, Hochschild A, Ptashne M. DNA loops induced by cooperative binding of  $\lambda$  repressor. *Nature*, 1986, **322** (6081): 750~752
- 14 Hochschild A, Ptashne M. Cooperative binding of  $\lambda$  repressors to sites separated by integral turns of the DNA helix. *Cell*, 1986, **44** (5): 681~687
- 15 van Helden J, André B, Collado-Vides J. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. *J Mol Biol*, 1998, **281** (5): 827~842
- 16 Wingender E, Chen X, Fricke E, *et al.* The TRANSFAC system on gene expression regulation. *Nucl Acids Res*, 2001, **29** (1): 281~283
- 17 Matys V, Fricke E, Geffers R, *et al.* TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucl Acids Res*, 2003, **31** (1): 374~378
- 18 Vignais M L, Huet J, Buhler J M, *et al.* Contacts between the factor TUF and RPG sequences. *J Biol Chem*, 1990, **265** (24): 14669~14674
- 19 Lieb J D, Liu X L, Botstein D, *et al.* Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genetics*, 2001, **28** (4): 327~334

## Potential Transcriptional Synergy Between Upstream Regions and Introns of Yeast Genes\*

ZHANG Kun-Lin<sup>1)</sup>, ZHANG Jing<sup>2)\*\*</sup>, LUO Jing-Chu<sup>1)</sup>

1) *Center of Bioinformatics, College of Life Science and National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, Beijing 100871, China;*

2) *Department of Statistics, The Center of Applied Statistics, Yunnan University, Kunming 650091, China)*

**Abstract** Previous work indicated that there was a large difference of the sequence features of introns between highly- and lowly-transcribed yeast genes. Some potential positive transcriptional regulatory motifs in the highly-transcribed introns were extracted. It was found that these introns were very close to the upstream regions of genes, and several introns even located within 5'-UTR. These results show that introns of yeast genes may regulate the transcriptional efficiencies and cooperate with upstream regions. To understand the synergy between upstream regions and introns of highly-transcribed yeast genes, the upstream regions (800 bp upstream of translation start sites, or the regions of two adjacent genes) of these two sets of genes were retrieved, and some potential positive transcriptional motifs in the upstream of highly-transcribed genes were extracted using the statistical comparative analysis approach developed before. Most of the potential motifs extracted were supported by literature search of experimental analyses. Then, every pair of oligonucleotides, one in the upstream region and the other in the intron, was defined as an “oligonucleotide pair”, i.e. a motif pair. Considering the motif pairs with the “nearest distance” not larger than 84 bp, the motif pairs, mainly tetra- and penta-nucleotide pairs, occurring in the highly-transcribed genes non-randomly were extracted and the synergistic pattern of these motif pairs including

position distributions and binding transcription factors such as RAP1, ABF1 and TAF was analyzed. Some potential patterns of transcriptional synergy were observed. These results could help people to understand the mechanism of transcriptional regulation and provide evidence for biological verification.

**Key words** upstream region of gene, intron, oligonucleotide pair, synergy, yeast gene

\*This work was supported by grants from The National Natural Science Foundation of China (30360027), The Special Funds for Major State Basic Research of China (2003CB15900).

\*\*Corresponding author. Tel: 86-871-6541419, E-mail: zhangjing@ynu.edu.cn

Received: July 7, 2004 Accepted: July 31, 2004

## 科学出版社生命科学编辑部新书推介

### 精编分子生物学实验指南(第四版)(译)

Frederick Ausubel 等著

马学军、舒跃龙、颜子颖、王海林 等译

2005年1月出版, ISBN 7-03-014725-1/Q.1532, 定价: 130元

被誉为分子生物学“红宝书”的《精编分子生物学实验指南》(Short Protocols in Molecular Biology)是知名度很高、不断更新的《最新分子生物学实验方法汇编》(Current Protocols in Molecular Biology)系列的精简版本,其囊括了《汇编》中所有基本方法及其详细实验步骤,是一本实验室必备的工具书。在上一版推出5年后,新版即将面世。新版对原有内容进行了修订和更新,包括:大肠杆菌、质粒和噬菌体, DNA制备与分析, DNA和RNA的酶促操作, RNA的制备与纯化, 重组DNA文库, 重组DNA文库的筛选, DNA测序, 重组DNA诱变, DNA转染哺乳动物细胞方法的介绍, 蛋白质分析, 免疫学, DNA-蛋白质相互作用, 酿酒酵母, 原位杂交与免疫组织化学, 聚合酶链式反应, 蛋白质表达, 蛋白质磷酸化分析等; 又新增了生物信息学、蛋白质相互作用、统计分析等新内容。



### 重要农林外来入侵物种的生物学与控制

万方浩 郑小波 郭建英 主编

ISBN 7-03-014409-0/Q.1492

16开, 818面, 精装, 定价: 150元

本书系统介绍了我国重要农林外来入侵物种的生物学特性与控制方法。全书分为总论、各论和农林外来入侵物种名录三大部分。总论部分介绍生物入侵的有关概念和国内外生物入侵的发生、预防与控制研究现状以及入侵机制的研究发展趋势。各论部分详细介绍我国目前35种最重要的农林外来入侵生物的分布与起源、生物学特性、识别特征与早期诊断、传入途径与入侵成因、危害现状与生态经济影响评估、风险分析与监测、预防控制措施。农林外来入侵物种名录附有279种农林外来入侵物种名录及其信息分析。



本书适用于相关领域的科研人员、大专院校师生、从事动植物检疫和农林研究的科研人员、行政官员及管理人员参考。

### 转基因棉花(英文版)

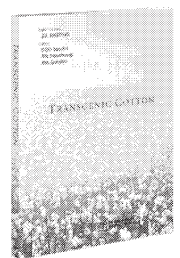
贾士荣 编著

2005年1月出版

ISBN 7-03-014470-8/Q.1500

定价: 120元

本书是“863”生物高技术丛书《转基因棉花》的英文版。全书共分14章,以转基因抗虫棉为重点,从理论和实践两方面详尽阐述了Bt杀虫晶体蛋白及其基因、蛋白酶抑制剂基因及其他抗虫基因的结构、功能及其抗虫机理、植物表达载体的构建、棉花遗传转化、抗虫棉的抗性及其分子生物学检测、抗虫棉品种的培育及遗传、抗虫棉抗虫性治理对策、安全性评估、国产抗虫棉的品种特性、栽培特点及综合防治技术、抗虫杂种棉的培育、双价转基因抗虫棉的研制等。同时,还增加了近几年的最新研究进展。



本书可供从事农学、分子生物学、遗传学、生物工程等专业研究和教学的科技人员、教师和学生参考,也可供从事农业技术推广和管理的工作人员参考。

欢迎各界人士邮购科学出版社各类图书(免邮费)

邮购地址: 100717 北京东黄城根北街16号科学出版社 科学分社, 联系人: 阮芯 联系电话: 010-64034622(带传真)

更多精彩图书请登陆网站 <http://www.lifescience.com.cn>, 欢迎致电索要书目, 010-64012501