

如果线段刺激的方位角为 α ($\alpha \neq \theta$),

$$E_{\alpha} = 0 \quad (12)$$

对一定感受野范围的 CX 细胞, 发生反应仅在线段的视网膜影象方位角与此 CX 细胞的本征方位角相同时。

三、数学模型的应用意义讨论

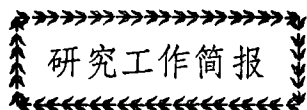
建立的数学模型概念除了能用于分析视觉脑的信息处理机制, 在图象识别的模拟研究中也有一定价值。人们早就认识到神经网络信息处理的“平行过程”不同于计算机的“顺序过程”, 并进行过电子学模拟^[15]。模拟装置中的神经单元能够再现真实神经元的某些生理特性, 如感受野大小和方位特异性等。从神经元的功能联系来看, 模拟装置却有很大区别。神经元对输入信息处理产生的输出不仅仅是导致新的生理特性出现, 更重要的, 是通过突触作用于下级神经元。对于每个神经元, 其下级神经元往往不止一个。如前所分析的, S 细胞与相同感受野内所有不同方位角的 CX 细胞都有抑制性输入。许多不同方位角的 CX 细胞对某线段不发生反应只是由于一个 S 细胞抑制的结果。这

样使用少数元件完成多种功能的高利用率反映了生物体结构功能的普遍现象。因此, 根据这一思想设计图象识别的模拟装置, 对于分析复杂影象无疑是有帮助的。

参 考 文 献

- [1] Hubel, D. H.: *J. Physiol.*, **160**, 106, 1962.
- [2] Hubel, D. H. et al.: *J. Neurophysiol.*, **28**, 229, 1965.
- [3] Schiller, P. H. et al.: *J. Neurophysiol.*, **39**, 1288, 1976.
- [4] Nagano, T. et al.: *Biol. Cybern.*, **38**, 103, 1980.
- [5] Henry, G. H.: *Brain Res.*, **113**, 1, 1977.
- [6] Bullier, J. et al.: *J. Neurophysiol.*, **47**, 417, 1982.
- [7] Toyama, K. et al.: *J. Neurophysiol.*, **46**, 191, 1981.
- [8] Toyama, K. et al.: *J. Neurophysiol.*, **46**, 202, 1981.
- [9] Singer, W. et al.: *Neurophysiol.*, **38**, 1080, 1975.
- [10] Wiesel, T. N. et al.: *J. Neurophysiol.*, **29**, 1115, 1966.
- [11] Ferster, D. et al.: *J. Comp. Neurol.*, **182**, 923, 1978.
- [12] Sillito, A. M.: *J. Physiol.*, **250**, 305, 1975.
- [13] Cleland, G. et al.: *J. Physiol.*, **255**, 299, 1976.
- [14] Springer, G.: *Calculus*, (2nd Edition) Vol. 1, by J. Wiley et al., 445, 1967.
- [15] Muller, F. J. et al.: *Pattern Recognition*, **8**, 47, 1976.

[本文于 1983 年 3 月 25 日收到]



研究工作简报

人、鼠、兔 β -珠蛋白基因的信息分析

王 玲 华

(第二军医大学生物教研室, 上海)

信息熵的计算是分析 DNA 分子结构和进化规律的有效方法之一。人们通过计算二联碱基水平的熵的偏离 D_2 揭示了 DNA 分子进化的方式, 同时表明, D_2 可作为进化指标。随着基因的核苷酸序列的不断确定, 有可能进一步分析 DNA 的信息特性。本文就从三联碱基水平计算人、鼠、兔 β -珠蛋白基因的各项熵值, 并就这些结果在 DNA 的结构和进化特性上作些

探讨。

一、材料和方法

我们用的材料是人、兔 β -珠蛋白基因和鼠 β^{maj} ——珠蛋白基因, 它们的核苷酸序列已完全确定^[1,2,3]。

计算方法

(1) 碱基实际出现概率的计算 取 DNA

单链接 5'→3' 端的方向,对单个碱基 A、T、C、G, 二联碱基 AA、AT……, 三联碱基 AAA、AAT…… 分别计数得到其出现总数 N_i 、 N_{ij} 、 N_{ijk} (i, j, k 为 A、T、C、G 中任一) 用

$$P_m = \frac{N_m}{\sum_m N_m} \quad (m = i, ij, ijk)$$

求得其出现频率,并认为这就是概率。

(2) 碱基预计概率的计算 假定碱基的出现是独立的。根据独立的性质,即可求得各碱基的预计概率 $P'_{ij} = P_i \cdot P_j$, $P'_{ijk} = P_{ij} \cdot P_k$, $P^2_{i,j,k} = P_{i,j,k} \cdot P_i$, $P'_{i,j,k} = P_{jk} \cdot P_i$ 其中

$$P_{i,j,k} = \sum_j P_{i,j,k}$$

而 $P_{i,j,k}$ 右上方数字表示三联碱基的第几碱基独立出现。

(3) 概率差的计算: 为了衡量三联碱基中由不同碱基引起的熵的改变的程度, 还需计算概率差。所用公式:

$$X^n_{i,j,k} = \frac{(P_{i,j,k} - P^n_{i,j,k})^2}{P^n_{i,j,k}} \quad (n = 1, 2, 3)$$

再归一化

$$P^n_{i,j,k} = \frac{x^n_{i,j,k}}{\sum_{i,j,k} x^n_{i,j,k}}$$

使之成为一完全信息概型,以便求熵。

(4) 信息熵的计数 利用公式

$$H(s) = - \sum_i P_i \log P_i$$

把已求得的 P_i 、 P_{ij} 、…… 分别代入公式即得对应的熵值 $H(s^1)$ 、 $H(s^2)$ 、……。为了比较, 可把熵值标准化。先求最大熵 $H_{\max}(s^i) = \log N_i$, (在此 $i = 1, 2, 3$, $N_i = 4^i$) 再求标准熵

$$H_i(s^i) = \frac{H(s^i)}{H_{\max}(s^i)}$$

由于习惯,其中的对数底都取 2, 对应的熵的单位是比特。

(5) D 值的计算 D 值表示由于碱基间相互作用造成的熵的偏离,它是一种信息量。各项 D 值为 $D_1 = H_{\max}(s^1) - H(s^1)$, $D_2 = H'(s^2) - H(s^2)$, $D_3 = H'_x(s^3) - H_x(s^3)$ 。(由于

$$P'_{i,j,k} = \sum_j P'_{i,j,k} = \sum_j P_{i,j} \cdot P_k = P_i \cdot P_k$$

$$\begin{aligned} \text{故 } H'_x(s^3) &= - \sum_{i,j,k} P'_{i,j,k} \log P'_{i,j,k} = - \sum_{i,j,k} P'_{i,j,k} \log P_{i,j,k} \\ &= H'(s^2) \text{ 即有 } D_3 = H'(s^2) - H_x(s^3) \end{aligned}$$

二、结果和讨论

以上各项结果见表 1 至 4。

表 1 单个碱基的出现概率 P_i

	人				兔				鼠			
	A	T	C	G	A	T	C	G	A	T	C	G
编码序列	0.1943	0.2340	0.2759	0.2558	0.2103	0.2661	0.2318	0.2918	0.2087	0.2638	0.2821	0.2454
非编码序列	0.2798	0.3548	0.1815	0.1840	0.2466	0.3472	0.2021	0.2041	0.2529	0.3277	0.2075	0.2119
全 段	0.2636	0.3341	0.2005	0.2093	0.2348	0.3267	0.2051	0.2259	0.2337	0.3081	0.2294	0.2288

表 2 二联碱基的出现概率 P_{ij}

第一碱基 \ 第二碱基 概率	人				兔				鼠			
	A	T	C	G	A	T	C	G	A	T	C	G
A	0.0774	0.0841	0.0467	0.0554	0.0652	0.0630	0.0489	0.0578	0.0670	0.0577	0.0398	0.0674
T	0.0718	0.1210	0.0538	0.0889	0.0349	0.1237	0.0659	0.0985	0.0428	0.0973	0.0781	0.0899
C	0.0651	0.0899	0.0597	0.0056	0.0644	0.0844	0.0481	0.0081	0.0688	0.0936	0.0645	0.0096
G	0.0446	0.0518	0.0518	0.0610	0.0467	0.0615	0.0489	0.0689	0.0576	0.0620	0.0508	0.0583

表 3 三联碱基的出现概率 P_{ijk}

第一碱基 第二碱基 概率	人				兔				鼠				第三碱基
	A	T	C	G	A	T	C	G	A	T	C	G	
A	0.0221	0.0210	0.0179	0.0118	0.0192	0.0130	0.0199	0.0199	0.0174	0.0174	0.0180	0.0192	A
	.0231	.0292	.0169	.0123	.0192	.0207	.0184	.0130	.0196	.0149	.0099	.0155	T
	.0133	.0118	.0097	.0103	.0130	.0138	.0092	.0084	.0081	.0105	.0093	.0167	C
	.0190	.0200	.0021	.0210	.0156	.0176	.0023	.0184	.0229	.0149	.0024	.0180	G
T	.0252	.0262	.0169	.0169	.0130	.0138	.0230	.0207	.0167	.0118	.0198	.0229	A
	.0262	.0497	.0251	.0205	.0146	.0544	.0291	.0291	.0118	.0322	.0316	.0254	T
	.0122	.0215	.0103	.0226	.0138	.0153	.0130	.0253	.0074	.0267	.0236	.0186	C
	.0072	.0236	.0015	.0200	.0061	.0199	.0031	.0268	.0068	.0267	.0031	.0229	G
C	.0164	.0159	.0128	.0005	.0146	.0069	.0084	.0015	.0136	.0099	.0161	.0031	A
	.0195	.0256	.0210	.0031	.0146	.0261	.0199	.0031	.0124	.0267	.0298	.0037	T
	.0113	.0138	.0149	.0015	.0184	.0169	.0176	.0015	.0167	.0298	.0161	.0018	C
	.0179	.0246	.0010	.0005	.0192	.0375	.0007	.0023	.0254	.0298	.0025	.0012	G
G	.0133	.0092	.0179	.0128	.0141	.0077	.0207	.0077	.0174	.0081	.0161	.0130	A
	.0118	.0164	.0200	.0149	.0069	.0184	.0230	.0184	.0243	.0198	.0192	.0167	T
	.0072	.0087	.0128	.0159	.0107	.0153	.0092	.0156	.0099	.0136	.0136	.0149	C
	.0123	.0174	.0010	.0174	.0156	.0222	.0023	.0176	.0164	.0205	.0019	.0136	G

表 4 各项熵值(单位: 比特)

	$H(s^1)$	$H(s^2)$	$H'(s^2)$	$H'(s^3)$	$H'(s^3)$	$H^3(s^3)$	$H_x(s^3)$	$H(s^3)$	D_1	D_2	D'_3
人	1.9719	3.8769	3.9427	4.7565	4.8736	4.8985	3.9308	5.7434	0.0281	0.0658	0.0119
兔	1.9725	3.8821	3.9425	4.7848	5.1406	4.9757	3.9359	5.7669	0.0275	0.0604	0.0066
鼠	1.9869	3.9222	3.9767	4.8185	4.9260	4.9371	3.9669	5.8001	0.0131	0.0545	0.0098

(1) 从表 1 可知, 各碱基的出现是非随机的。把表 2、表 3 的概率与相应预计概率比较, 结果表明, 各碱基的出现亦非独立。两者之差按单个碱基、二联碱基、三联碱基的顺序逐渐增大。从某一碱基水平看, 同一种碱基组合的不同排列也可有明显差别。如人的 CG 出现概率为 0.0056, 而 GC 为 0.0518。总的看, A、T 出现较多, 而 C、G 出现较少。如三联碱基水平最多的 TTT 为 0.0544 (人), 最少的 CGG 仅为 0.0005。但这种现象主要在于非编码序列中, 而编码序列不甚明显(见表 1)。由于 A-T 对缔合能小, 而 C-G 对缔合能大, 这就使编码区较稳定, 具有较高的抗干扰性, 非编码区较松散, 对内外环境的变化较敏感; 在复制转录时也可节省能量。

(2) 从表 4 的 D 值可知, 碱基的相互影响,

提高了 DNA 的有序性。这种现象按单个、二联、三联碱基的顺序递增。根据 D'_3 值和 D'_3 的定义, 可知 DNA 分子中不但相邻碱基有相互作用, 相间的也有作用。这些作用都可提高 DNA 分子的有序性。

(3) 把表 4 的熵值标准化后, 有 $H_i(s^1) > H_i(s^2) > H_i(s^3)$, 且 $H_x(s^3) > H_i(s^3)$ 。由于碱基间相互影响的结果是有序性增加即信息熵的减少, 因而可知二个碱基对另一个碱基的影响比任何一个碱基对另一碱基的影响大。这就是说, 碱基的相互作用是可以叠加的, 是互相协同的。这种作用可能是氢键堆积能, 构象能。

(4) 从表 4 的差熵 $H''(s^3)$ 可知, 二个碱基对另一碱基的影响按第一、第二、第三碱基的顺序递增。这和复制时子链的合成方向 5'→3' 端

(下转第 74 页)

与抗血清的工作配比,做 BK 抗血清的稀释度曲线,抗血清最终稀释度 1:70,000 时结合百分率为 55% (抗血清 -20℃ 稀释储存后,效价逐渐下降,结合率为 30—40%) 为最适宜。选择这一工作配比做 BK 放免测定得到典型的竞争性抑制曲线(图 2),25 pg 与 0 pg BK 管的结合百分率有明显差别,BK 含量的增加与结合百分率成负相关。

^{125}I -BK 的非特异性吸附是 $6.547 \pm 0.271\%$, 偏离系数 C. V 为 4.1%。批内测定取三个血样,每个血样取 5 管同时测定,偏离系数 C. V 8.0—9.98%,批间是取三个血样分批测定(见表 2)。

通过上述方法测得正常个体血浆 BK 浓度范围为 2.4—7.2 $\mu\text{g/L}$ 血浆 ($n = 23$) ($\bar{X} \pm SD = 4.20 \pm 1.35 \mu\text{g/L}$)。

表 2 血 BK 批内批间测定

血浆样品		BK 均值 $\mu\text{g/L}$	S. D	C. V%
批内	1	3.80	0.30	8.04
	2	3.00	0.24	8.00
	3	2.41	0.24	9.98
批间	1	3.48	0.28	8.28
	2	6.06	0.52	8.61
	3	3.70	0.27	7.21

三、讨 论

血浆激肽系受血浆中几种酶的调节,前激肽释放酶是一种蛋白酶元,它可受第 XII 凝血

因子(哈格曼因子)活化,作用于血浆中的激肽元,释放出 BK。这一系统还受激肽释放酶抑制剂及分解激肽的激肽酶的调节。因此,测定血中 BK 首先要抑制这些酶类,防止由于 BK 的生成或破坏而人为造成 BK 水平过高或过低。在血浆 BK 的制备过程中我们用聚凝胺抑制哈格曼因子,又在一切操作过程中避免接触玻璃(因为玻璃表面带负电荷,可激活哈格曼因子)。并利用乙二胺四乙酸二钠抑制激肽酶,兼抗凝。

1969 年 Talama 等首先报道用 RIA 测定血中 BK,其灵敏度达 1—25 $\mu\text{g/L}$ 血浆。但步骤多,工序麻烦,并且混有激肽元,它与 BK 抗体交叉反应,造成 BK 结果偏高。1972 年 Masbford 等报道的方法虽有改进,但因为脂类干扰使游离 BK 与结合 BK 不易分离,增加了 BK-RIA 测定的困难。

我们所用的测定血 BK 水平的方法,简便、快速,经 G.25 柱一步即分离出血 BK,一次可测定 20—30 个样品。其灵敏性及精确性均好,可用于各种研究工作及临床常规测定。

参 考 文 献

- [1] Pritam, S. V. et al.: *Clin. Chem.*, **26** (3), 429, 1980.
- [2] 陆以信等:《第四次中国生化学术会议论文摘要汇编》, 227, 1981.
- [3] Lennart, H.: *Acta Endocrinologica, Suppl.*, **235**, 94, 1980.

[本文于 1983 年 3 月 7 日收到]

(上接第 53 页)

的顺序是一致的。这种影响可以减少子链合成时碱基出现的不确定性。减少错误的发生。构象适应能和 DNA 聚合酶中的 3'→5' 核酸酶的存在可能是造成这种现象的基础。

参 考 文 献

- [1] Richard, M. et al: *Cell*, **21**:647, 1980.
- [2] David, A. et al: *Cell*, **18**:865, 1979.
- [3] Ross, C. et al: *Cell*, **18**:1285, 1979.

[本文于 1983 年 3 月 7 日收到]