

右)或免疫电泳^[6]检查都显示是纯品;(2)柱容量大,每毫升凝胶可吸附样品 700 毫克,上样 200 毫克只需 5 厘米左右一根短柱,既省时又省料;(3)回收率几近定量,在其它几个组份中几乎不含可检查出来的胆铁质;(4)柱再生十分容易,只需以 CuSO_4 溶液重新通过一次即可再次上样。

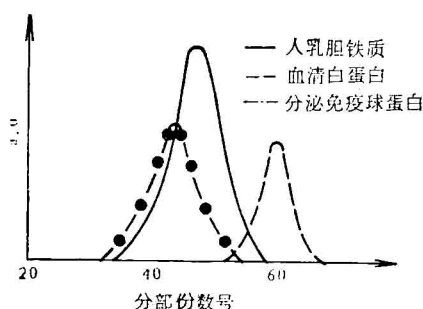


图 5 图 4 中标以斜线的峰位组份的免疫电泳

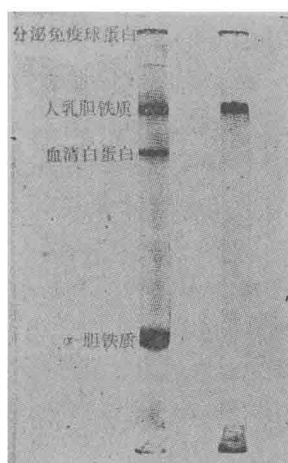


图 6 人乳胆铁质纯化前后的凝胶电泳图

小 结

通过以上介绍,可将金属螯合亲和层析操

作程序归纳为(1)泡胀及洗涤需要量的环氧活化的琼脂糖(Sephrose 4B, 6B)衍生物,过滤待用;(2)将亚氨基二乙酸钠溶解于约 2M 的 Na_2CO_3 中(0.2 克/毫升),再将其加到上述滤干的凝胶中;(3)保温 65°C ,使之偶合约 24 小时;(4)过滤,以蒸馏水充分洗涤;(5)装柱,然后将相应的金属离子溶液(1 毫克/毫升 ZnCl_2 或 $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$)以约 10 厘米/小时的流速通过(若用 Cu^{2+} ,只需通过全柱的 2/3 左右即可);(6)上样,再以至少 10 倍体积的起始缓冲剂平衡及洗脱不吸附物质;(7)用递增离子强度或递减 pH 值的缓冲液进行梯度洗脱,必要时可在其中加入 50 mM 的 EDTA 溶液用来洗脱吸附较强的组份。

本法的主要特点是:(1)设备简易,材料易得,环氧活化的琼脂糖已有商品(Pharmacia 公司)。此外,只需几种简单试剂;(2)操作方便,时间短(本法实际上只是一次柱层析;又因柱效高,容量大,所需层析时间自然较少);(3)回收率很高,一般达 80% 以上,有的几近定量回收;(4)样品纯度高,而且因层析条件温和,时间少,所以整个过程中没有蛋白质变性或水解现象发生。

参 考 文 献

- [1] Porath, J. et al.: *Nature*, **258**, 598—599, 1975.
- [2] Lebreton, J. P.: *FEBS Letter*, **80**, 351—354, 1977.
- [3] Edy, V. G.: *J. Biol. Chem.*, **252**, 5934—5935, 1977.
- [4] Bolonnerdal, J.: *FEBS Letter*, **75**, 89—92, 1977.
- [5] Scheidegger, J. J.: *Int. Arch. Allergy*, **7**, 103—110, 1955.
- [6] Scheidegger, J. J.: *Int. Arch. Allergy*, **7**, 103—110, 1955.

[本文于 1979 年 10 月 29 日收到]

遗传密码中的概率与信息问题

王 身 立

(复旦大学遗传研究所)

60 年代国外应用人工合成的、包含一定核苷酸序列的 RNA 作为模板,控制多肽的合成,

从模板 RNA 中核苷酸成份与多肽中氨基酸成份对应关系的分析,阐明了遗传密码,解读了大

自然的“天书”，弄清楚了为氨基酸编码的 mRNA 中三联体密码子 (Triplet codon) 的核苷酸排列顺序 (表 1)。

从表 1 可看出，遗传密码中是存在着简并的。例如亮氨酸、精氨酸和丝氨酸各有六个简并码，丙氨酸等五个氨基酸各有四个简并码，异亮氨酸有三码，赖氨酸等九个氨基酸各有二码。

表 1 遗传密码

第一核苷酸	第二核苷酸				第三核苷酸
	U	C	A	G	
U	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	U
	苯丙氨酸	丝氨酸	酪氨酸	半胱氨酸	C
	亮氨酸	丝氨酸	终止信号	终止信号	A
	亮氨酸	丝氨酸	终止信号	色氨酸	G
C	亮氨酸	脯氨酸	组氨酸	精氨酸	U
	亮氨酸	脯氨酸	组氨酸	精氨酸	C
	亮氨酸	脯氨酸	谷酰胺	精氨酸	A
	亮氨酸	脯氨酸	谷酰胺	精氨酸	G
A	异亮氨酸	苏氨酸	门冬酰胺	丝氨酸	U
	异亮氨酸	苏氨酸	门冬酰胺	丝氨酸	C
	异亮氨酸	苏氨酸	赖氨酸	精氨酸	A
	甲硫氨酸	苏氨酸	赖氨酸	精氨酸	G
G	缬氨酸	丙氨酸	门冬氨酸	甘氨酸	U
	缬氨酸	丙氨酸	门冬氨酸	甘氨酸	C
	缬氨酸	丙氨酸	谷氨酸	甘氨酸	A
	缬氨酸	丙氨酸	谷氨酸	甘氨酸	G

另外有三个密码子 UAA、UAG 和 UGA 不对应任何氨基酸，仅仅作为 mRNA 指导的蛋白质或多肽合成的终止信号(标点)。

本文讨论遗传密码中有关概率和信息方面的数学问题，所应用的数学方法参考了徐京华的“生物体负熵输入的计算”一文^[3]。

设有甲乙两方进行通讯。甲方为发信者，以相等的概率发出 A、U、G、C 四种核苷酸信号。每种核苷酸的出现概率为 1/4，信息量为

$$-\sum_{i=1}^4 p_i \log p_i = -\log \frac{1}{4} = 2 \text{ bit} \quad (1)^*$$

但发信者不是发送单个的核苷酸信号，而是以三个核苷酸为一组发送三联体密码子的信号。四种核苷酸一共可组成 $4^3 = 64$ 种密码子。每个密码子的出现概率为

$$p(x) = \left(\frac{1}{4}\right)^3 = \frac{1}{64}^{**}$$

信息量

$$\begin{aligned} H(x) &= -\sum_{x=1}^{64} p(x) \log p(x) \\ &= -\log \frac{1}{64} = 6 \text{ bit} \end{aligned} \quad (2)$$

乙方为受信者，接受三联体密码子转译成的氨基酸消息。命消息的概率为 $p(y)$ 。例如收到甲硫氨酸或色氨酸消息的概率为 1/64，收到赖氨酸等九个氨基酸消息的概率各为 2/64 (它们每一个均对应于两个三联码)，收到异亮氨酸或终止密码子消息的概率各为 3/64，……等等。则接收到消息的信息量

$$\begin{aligned} H(y) &= -\sum_{y=1}^{21} p(y) \log p(y) \\ &= -\left(2 \cdot \frac{1}{64} \log \frac{1}{64} + 9 \cdot \frac{2}{64} \log \frac{2}{64} \right. \\ &\quad \left. + 2 \times \frac{3}{64} \log \frac{3}{64} + 5 \times \frac{4}{64} \log \frac{4}{64} \right. \\ &\quad \left. + 3 \times \frac{6}{64} \log \frac{6}{64}\right) = 4.22 \text{ bit} \end{aligned} \quad (3)$$

甲乙两方信号各为 x (三联体密码子)与 y (氨基酸)的联合概率 $p(x, y)$ 如下面的矩阵(联合矩阵)所示 (表 2)。

联合信息量

$$\begin{aligned} H(x, y) &= -\sum_{x=1}^{64} \sum_{y=1}^{21} p(x, y) \log p(x, y) \\ &= -\sum_{x=1}^{64} p(x) \log p(x) \\ &= H(x) = 6 \text{ bit} \end{aligned} \quad (4)$$

在甲方出现信号 x 的条件下，乙方收到某个消息 y 的条件概率为 $p(y/x)$ 。例如甲方出现 UUU 的条件下，乙方收到苯丙氨酸的概率为 1，而收到其余氨基酸的概率为 0 (如果通讯与翻译中不发生错误的话)。这些条件概率如下面的“消散矩阵”所示 (表 3)。这种矩阵就是信息

* 本文各式中的对数 (log) 均以 2 为底，算式中不另一一标明。
 ** 此文暂假定三联体密码子以等概出现。但实际的 mRNA 并不完全如此，参见本文末的讨论。

表2 三联体密码子-氨基酸的联合矩阵式 $\langle p(x, y) \rangle$

$p(x, y)$ 密码 \ 氨基酸	苯丙, 亮..... 半胱..... 蛋..... 终止信号	$p(x) = \sum_{y=1}^{21} p(x, y)$
UUU	1/64, 0 0 0 0	1/64
UUC	1/64, 0 0 0 0	1/64
UUA	0 1/64..... 0 0 0	1/64
UUG	0 1/64..... 0 0 0	1/64
CUU	0 1/64..... 0 0 0	1/64
CUC	0 1/64..... 0 0 0	1/64
CUA	0 1/64..... 0 0 0	1/64
CUG	0 1/64..... 0 0 0	1/64
.....		
.....		
.....		
UGU	0 0 1/64..... 0 0	1/64
UGC	0 0 1/64..... 0 0	1/64
.....		
.....		
.....		
AUG	0 0 0 1/64..... 0	1/64
.....		
.....		
.....		
UAA	0 0 0 0 1/64	1/64
UAG	0 0 0 0 1/64	1/64
UGA	0 0 0 0 1/64	1/64
$p(y) = \sum_{x=1}^{64} p(x, y)$	2/64, 6/64 2/64..... 1/64..... 3/64	$\sum_{x=1}^{64} p(x) = \sum_{y=1}^{21} p(y) = 1$

论中的“信道矩阵”，它的每一行相应于信道的一个输入，每一列对应于一个信道输出。它的任何一行中各项概率之和总是等于1，这是因为对于任何一个输入信号 x ，总可以得到信道矩阵相应的一列中若干个消息中的某一个^[1]。

于是我们有

条件熵

$$\begin{aligned}
 H(y/x) &= H(x, y) - H(x) \\
 &= - \sum_{x=1}^{64} \sum_{y=1}^{21} p(x, y) \log \frac{p(x, y)}{p(x)} \\
 &= - \sum_{x=1}^{64} p(x) \sum_{y=1}^{21} p(y/x) \log p(y/x) = 0
 \end{aligned}
 \tag{5}$$

在乙方收到消息 y 的条件下，甲方出现信号 x 的条件概率为 $p(x/y)$ 。例如乙方收到苯丙氨酸的条件下，甲方出现 UUU 或 UUC 的概率

各为 1/2，而出现任何其它三联体密码子的概率为0(如果通讯与翻译中不发生错误的话)，这些条件概率如下面的“相当矩阵”所示(表4)。

这样，从接受方面看，条件熵

$$\begin{aligned}
 H(x/y) &= H(x, y) - H(y) = 6 - 4.22 \\
 &= 1.78 \text{ bit}_0
 \end{aligned}$$

$$\begin{aligned}
 H(x/y) &= - \sum_{x=1}^{64} \sum_{y=1}^{21} p(x, y) \log \frac{p(x, y)}{p(y)} \\
 &= - \sum_{y=1}^{21} p(y) \sum_{x=1}^{64} p(x/y) \log p(x/y) \\
 &= - \left(2 \times \frac{2}{64} \log 1 + 9 \times \frac{2}{64} \log \frac{1}{2} \right. \\
 &\quad \left. + 2 \times \frac{3}{64} \log \frac{1}{3} + 5 \times \frac{4}{64} \log \frac{1}{4} \right. \\
 &\quad \left. + 3 \times \frac{6}{64} \log \frac{1}{6} \right) = 1.78 \text{ bit}
 \end{aligned}
 \tag{6}$$

表 3 三联体密码子-氨基酸的消散矩阵式(信道矩阵) $\langle p(y/x) \rangle$

$p(y/x)$ 氨基酸 密码	苯丙,亮.....	半胱.....	蛋.....	终止信号	$\sum_{y=1}^{21} p(y/x)$
UUU	1	0	0	0	1
UUC	1	0	0	0	1
UUA	0	1	0	0	1
UUG	0	1	0	0	1
CUU	0	1	0	0	1
CUC	0	1	0	0	1
CUA	0	1	0	0	1
CUG	0	1	0	0	1
.....					...
.....					...
.....					...
UGU	0	0	1	0	1
UGC	0	0	1	0	1
.....					...
.....					...
.....					...
AUG	0	0	0	1	1
.....					...
.....					...
UAA	0	0	0	1	1
UAG	0	0	0	1	1
UGA	0	0	0	1	1

$H(x) = H(x, y)$, 这说明甲方每发出一个信号 x (三联体密码子), 都能决定乙方所收到的一个特定的消息 y (氨基酸)。但

$H(y) = 4.22\text{bit} \approx H(x, y)$,

说明乙方每收到一个消息, 并不能完全确定甲方所发出的信号 x , 这是因为遗传密码存在简并的缘故。例如乙方收到苯丙氨酸的消息时, 不能确定甲方发出的信号究竟是 UUU 还是 UUC。也就是说, 在译码(翻译)过程中产生了疑义度。该疑义度即乙方收到氨基酸消息的条件下对甲方所发出的三联体密码子信号的不定度 $H(x/y) = 1.78 \text{ bit}$ 。

交互信息(乙方收到消息 y 后得到的每个符号 x 的平均信息量; 参见图 1)为

$I(x, y) = H(x) - H(x/y)$
 $= 4.22 \text{ bit} = H(y)$ (7)

乙方收到氨基酸消息后平均可得到

$H(y) \approx 4.22 \text{ bit}$

的信息, 而甲方发出的三联体密码子平均有

$H(x) = 6\text{bit}$

信息。因为, 冗余度为

$1 - \frac{4.22}{6} = 30\%$ (8)

我们发现, 某种氨基酸简并码数的多少, 与该氨基酸在生物蛋白质总体中出现的概率有很大程度的相关。根据 H. Neurath 等所著《蛋白

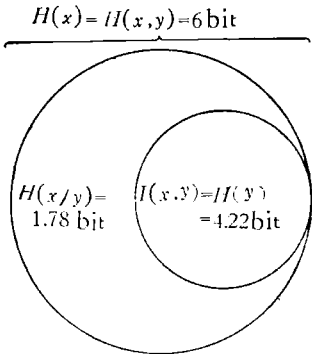


图 1

表 4 三联体密码子-氨基酸的相当矩阵式 $\langle p(x/y) \rangle$

$p(x/y)$ 氨基酸 密码		苯丙	亮		半胱		蛋		终止信号
UUU		1/2	0		0		0		0
UUC		1/2	0		0		0		0
UUA		0	1/6		0		0		0
UUG		0	1/6		0		0		0
CUU		0	1/6		0		0		0
CUC		0	1/6		0		0		0
CUA		0	1/6		0		0		0
CUG		0	1/6		0		0		0
.....									
.....									
UGU		0	0		1/2		0		0
UGC		0	0		1/2		0		0
.....									
.....									
AUG		0	0		0		1		0
.....									
.....									
UAA		0	0		0		0		1/3
UAG		0	0		0		0		1/3
UGA		0	0		0		0		1/3
$\sum_{x=1}^{64} p(x/y)$		1	1		1		1		1

质》(卷 1)一书中所提供的氨基酸出现频率的资料,我们把它与氨基酸的简并码数进行比较,列出表 5。H. Neurath 书中总结了十几种蛋白质,共计 5,883 个氨基酸。因为每一个蛋白质分子必须有而且仅有一个终止信号,故 19 种蛋白质共有 19 个终止信号。从表 5 计算出氨基酸(以及终止信号)的出现频率与简并码数的相关系数。

$$r = \frac{\sum p(x) \cdot p(y)}{\sqrt{\sum [p(x)]^2 \cdot \sum [p(y)]^2}}$$
$$= \frac{0.584}{\sqrt{0.598 \times 0.672}} \doteq 0.93 \quad (9)$$

图 2 是氨基酸出现概率对简并码概率的回归直线。回归系数

$$b = \frac{\sum [p(x) \cdot p(y)]}{\sum [p(x)]^2} = \frac{0.584}{0.598} \doteq 0.98 \quad (10)$$

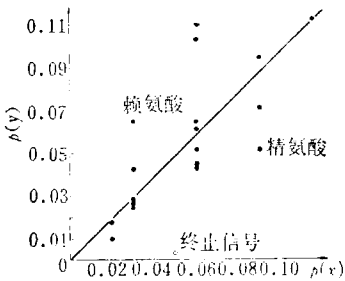


图 2 氨基酸出现概率 $p(y)$ 对三联体密码子概率 $p(x)$ 的回归直线

氨基酸出现概率与简并码数之间存在着如此程度的相关,一种可能的解释是,在生命发展的初期阶段(可能是生命起源时),RNA 分子中 64 种三联体密码子是以等概出现的,亦即三联体密码子是由核苷酸随机组合而成的。这样,简并码数愈多,出现的概率也就愈大(概率的相加法则,例如某一种三联体出现的概率为 $1/64$, 6

表 5 氨基酸出现频率与简并码数的对应关系

氨基酸	码数	$p(x)$	$[p(x)]^2$	出现频率	$p(y)$	$[p(y)]^2$	$p(x) \cdot p(y)$
亮氨酸	6	0.09	0.0081	56.7	0.095	0.0090	0.0086
丝氨酸	6	0.09	0.0081	426	0.071	0.0050	0.0064
精氨酸	6	0.09	0.0081	309	0.052	0.0027	0.0047
谷氨酸	}	4	0.06	661	0.110	0.0121	0.0066
谷酰胺							
门冬氨酸	}	4	0.06	625	0.104	0.0108	0.0062
门冬酰胺							
甘氨酸	4	0.06	0.0036	392	0.065	0.0042	0.0039
缬氨酸	4	0.06	0.0036	389	0.065	0.0042	0.0039
苏氨酸	4	0.06	0.0036	365	0.061	0.0037	0.0037
丙氨酸	4	0.06	0.0036	310	0.052	0.0027	0.0031
脯氨酸	4	0.06	0.0036	270	0.045	0.0020	0.0026
异亮氨酸	3	0.05	0.0025	261	0.044	0.0019	0.0022
赖氨酸	2	0.03	0.0009	391	0.065	0.0042	0.0020
苯丙氨酸	2	0.03	0.0009	259	0.043	0.0019	0.0013
半胱氨酸	2	0.03	0.0009	172	0.029	0.0008	0.0009
酪氨酸	2	0.03	0.0009	170	0.028	0.0008	0.0008
组氨酸	2	0.03	0.0009	155	0.026	0.0007	0.0008
甲硫氨酸	1	0.02	0.0004	99	0.017	0.0003	0.0003
色氨酸	1	0.02	0.0004	62	0.010	0.0001	0.0002
终止信号	3	0.05	0.0025	19	0.003	0.0001	0.0002
总和	64	1.00	0.0598	5902	1.000	0.0672	0.0584

个简并码出现的总概率则为 $6/64$)。随着生命的发展,在自然选择的作用下,这种随机性减小了。例如终止信号出现的总概率不再是 $3/64 \approx 0.05$,而减少到了 $19/5902 \approx 0.003$ (表 5)。我们可以由此推测,在生命发展初期,肽链合成过程中发生断裂的概率约为 $3/64 \approx 0.05$,亦即合成多肽分子时,平均每 100 个氨基酸有 5 个终止信号,所以多肽的平均长度为 $100/5 = 20$ 个氨基酸残基左右。而现代生物的蛋白质(或多肽)分子的长度平均终为 $5902 \div 19 \approx 300$ 个氨基酸左右。

近年来还发现某些微生物的 mRNA 中,有好几个终止密码子连续在一起,以确保蛋白质(或多肽)合成时在一定的位置中断肽链,从而形成固定长度的具有特定功能的蛋白质分子。因为连续的好几个终止密码子只能作为一个标点,亦即只能给出一个终止信号,而且中止密码子的连续出现也体现了随机性的减少,所以我们的讨论对于这种情况也仍然是适合的。

此外,在生命发展过程中, mRNA 分子中 AAA、AAG 等组合的出现概率增加了。在生命发展的初期,这两种三联体密码子出现的总概率可能仅为 $2/64 \approx 0.03$,因而那时蛋白质或多肽分子的赖氨酸含量仅为现代生物蛋白质中含量的一半左右,而现代生物蛋白质分子中赖氨酸出现的概率平均为 0.065。另一方面,精氨酸的含量则可能降低了十倍以上,由原来的 9% 降低到现代蛋白质的 0.81%。这表明 mRNA 中 CGX(x 表示 A、U、C、G 中的任何一个),AGA、AGG 等组合随着生命的发展,它们出现的概率显著减小了。

如果在生命发展的早期阶段就已经出现从 DNA 转录 RNA 的机制的话,那末 RNA 分子中密码子出现的随机性也反映了 DNA 分子中核苷酸出现的随机性。终止信号随机地出现,起始信号也随机性出现。现代生物的 mRNA 从 DNA 链上开始转录的起点则是固定的,该起点由 RNA 多聚酶的辅因子(σ 因子)来辨认。 σ

因子不存在时,转录就随机地从 DNA 模板上任何一点开始,好象许许多多生理功能缺陷型的“重迭基因”,合成的蛋白质或多肽分子长短参差,丧失正常的生理功能。在生命的早期阶段,可能还没有发展出 σ 因子的辨认机制,因而一个 DNA 链就成为许多的“重迭基因”,这也造成了密码子出现的随机性。现代生物的重迭基因就不再是随机的了。即使是 $\phi x-174$ 、G 4

和 λ 噬菌体以及 SV-40 病毒的重迭基因,也都有固定的起点。

参 考 文 献

- [1] 徐京华:《生物化学与生物物理学报》,1962年,2卷,1期,第31—32页。
- [2] 王身立:《遗传与育种》1978年,第1期,31—32页。

[本文于1978年10月23日收到]

专题 讲 座

冰冻超薄切片法及其在超结构形态学上的应用

徐 伟 门绍杰

(中国科学院生物物理研究所)

为了克服常规超薄切片术中采用的强烈化学固定、有机溶剂脱水 and 塑料包埋等步骤所带来的结构损伤和分子活性丧失等缺点,人们发展了如水溶性树脂包埋法、冰冻干燥法、冰冻刻蚀法、冰冻置换法及冰冻超薄切片法等新制备技术。60年代中期 Bernhard 等首次建立冰冻超薄切片术的基本方法和仪器。后又经 Appleton^[1], Beruhard^[2], Christeusen^[3] 和 Tokuyasu, K. T.^[4] 等人的努力,在技术和应用方面都取得了很大发展。70年代初期,商品冰冻超薄切片装置,如 LKB “Cryokit”, Sorvall “FTS bowl” 及 Reichert “FC-150” 等问世,进一步促进了这一技术的发展。本文对冰冻超薄切片术加以介绍。

生物材料的冰冻

冰冻法固定的优点在于可制备含水样品,使生物结构尽可能保持接近活体的状态;防止了可溶性物质的抽取、流失和移位;保持生物大分子的活性;使样品便于切片。

最理想的是在瞬间使整个细胞中的各部分都迅速冻结。这样不仅使结构形态和大分子活性得到良好的保存,而且细胞中的水分不会冻结成大的结晶体,只出现结构损伤不明显的无定形微晶态,即所谓“玻璃态”。可是生物材料

的热传导率很低,达到这样理想的冰冻状态是很难的。在较慢的冰冻速度(1度/秒)下,细胞外层的溶液中的水分首先冰冻,从而使溶质向未冰冻区域迁移,形成相分离,造成了形态的畸变。同时,溶质的迁移造成溶液中 pH 值的重大变化,进而使蛋白质等分子的构型发生变化。冰晶界面上的表面张力也会使蛋白质变性。较高的冰冻速率(大于 10^4 度/秒)由于热传导快,因而形成玻璃态,也减小了溶质迁移的破坏作用。所以提高冰冻速率是保持形态完好的关键。然而,目前所采用的冰冻方法,即使是很小的样品,也只能使样品块表面以下几微米到几十微米的范围内达到所谓“玻璃态”。

通常的冰冻条件下,事先将生物样品用甘油、蔗糖或二甲亚砜一类溶液浸透,则可抑制冰晶产生。

冰冻超薄切片的制备

制备冰冻超薄切片的程序大体上可分为两类(图1)。

理论上讲,不经固定、包埋和染色等溶液处理制备冰冻切片是最合乎需要的。这正是无溶液制备法追求的目标。它利于生物材料中可溶性物质不致流失,有利于保存生物大分子的活