

密码子的分配

杨 雨 善

(第二军医大学, 上海)

提 要

运用密码子与反密码子相互作用的规律, 本文从理论上提出了密码子在遗传密码表中的分配原则, 它与迄今所发现的一些线粒体内和线粒体外的密码表都相符合。

一、简要的回顾

现在相距遗传密码的解译 (即首次排出遗传密码表) 和 Crick 提出变偶假说的 1966 年, 已经二十载了。

二十年来, 经过科学工作者们的努力, 以更多的事实证明了最早得到的“通用” (universal) 遗传密码表对于从原核生物到真核生物来说, 在一定范围内仍然是通用的。自 1965 年 Holley 等分析了第一个 tRNA 的单核苷酸的排列顺序以来, 截至 1984 年底为止, 共分析了 377 个天然 tRNA 的一级结构^[1], 目前数量仍在增加。在这期间, 我国王德宝等^[2]在体外人工合成了具有生物活性的酵母 tRNA^{Ala}。这些都能证明, 在一定范围内, 变偶假说也是正确的。

伴随历史进程, 科学不断发展。七十年代后半叶、特别是八十年代以来, 由于线粒体基因组和翻译系统研究的长足前进^[3-5], 刷新了一些传统的概念:

1. 某些线粒体密码表与“通用”密码表不完全通用;

2. 线粒体密码表之间一般也不完全通用;

3. 在线粒体内, 密码子与反密码子的相互识别, 已经突破了经典的变偶假说所包含的主要原则。

目前, 对线粒体基因组和翻译系统的研究正方兴未艾, 生命体内存在的一些规律, 不断地被人们所发现、所认识; 国内外一些研究者对密码子与反密码子的相互作用也更加重视。

在回顾这段历史之际, 借用现有的研究成果和理论, 本文对密码子在线粒体内和线粒体外 (主要指胞液) 遗传密码表中的分配问题, 进行一些探讨, 寻找一些有规律性的东西, 作为值此二十周年之纪念。

二、分配的原则

已经知道, 遗传密码表中的 64 个密码子被划分在 16 个方框内, 每个方框中容纳 4 个密码

[12] Iwasa, Y.: *Biochem. Biophys. Res. Commun.*, **96**, 180, 1980.

[13] Johnson, E. Y.: *Hand Exp. Pharm.*, **58**, 507, 1982.

[14] Mamoru, S.: *Biochem. Biophys. Res. Commun.*, **127**, 590, 1985.

[15] Ushiro, H.: *J. Biol. Chem.*, **255**, 8368, 1980.

[16] Miwako, K.: *Biochem. Biophys. Res. Commun.*, **129**, 375, 1985.

[17] Cochol, C.: *J. Biol. Chem.*, **259**, 2553, 1984.

[18] Davis, R. T.: *J. Biol. Chem.*, **259**, 8545, 1984.

[19] Iwashita, S.: *J. Biol. Chem.*, **259**, 2550, 1984.

[20] Miskimins, R.: *Exp. Cell. Res.*, **146**, 53, 1983.

[21] Mrpczkowski, B.: *Nature*, **309**, 270, 1984.

[22] Macora, I. G.: *Proc. Natl. Acad. Sci. USA*, **81**, 2728, 1984.

[23] Sugimoto, Y.: *Proc. Natl. Acad. Sci. USA*, **81**, 2117, 1984.

[本文于 1986 年 12 月 9 日收到]

子,这里统称为密码子组,可用通式 XYN (X、Y、N 分别为第 1、2、3 位碱基或核苷,它们可以为 U、C、A、G) 表示,见表 1。

表 1 密码子组

		Y			
X		X	Y	U	U
		X	Y	C	C
		X	Y	A	A
		X	Y	G	G

注: 4 个密码子可以缩写成 XYN。

经过综合分析,在同一细胞的相同亚微结构中,一组内的 4 个密码子所代表的涵义(氨基酸或终止信号)是受到严格限制的,必须符合一定的要求。本文把这种现象称之为“密码子的分配原则”(简称分配原则)。其要点如下:

1. 同组内的 4 个密码子最多代表 3 种涵义,而且仅有一种分配: XY_C^U 代表一种氨基酸,XYA 代表终止信号,XYG 代表另一种氨基酸。

2. 同组内的 4 个密码子代表两种不同的涵义时,可以是以下任何一种分配:

(1) XY_C^U 代表一种涵义,XYG 代表另一种涵义;

(2) XY_C^U 代表一种氨基酸,XYA 为终止信号;

(3) XY_C^U 代表一种涵义,XYA 代表另一种涵义。

3. 同组内的 4 个密码子可以代表一种涵义,此时常称 XYN 为密码子族 (codon families)。

这就是分配原则的主要内容。它适用于原核生物与真核生物,也适用于线粒体内和线粒体外。尽管在线粒体内涵义发生变化的密码子和新的密码表时有发现,但可以想象到,绝大多数情况下,密码子的分配应该符合上述原则。

三、分配原则的必要前提

对以上密码子的分配原则,下面 4 个前提是必要的:

1. 这些原则主要适用于绝大多数的天然情况;

2. 密码子与反密码子相互识别时,是按碱基(或核苷)的标准配对、变偶配对和三配二 (two out of three pairing) 原则^[6,7]进行的;

3. 密码子与反密码子相互识别时,除次黄苷 I 外,本文较少考虑其他的修饰成分;

4. 依赖于已经取得的研究成果。

现在,对上述分配原则和前提,作些扼要的说明。

我们知道,密码子与反密码子在相互识别时,二者所含碱基(或核苷)的对应关系是重要的因素。所以每组密码子的分配,必须与这种对应关系相协调;换句话说,在天然情况下,分配原则受到这种对应关系的制约,否则将造成密码语言的误译。

在线粒体内和线粒体外的密码表中,都存在 XYN 编码同一种氨基酸的族性密码子。识别线粒体外的这类密码子,其反密码子的 5' 位有 G、U、C、I (I 取代 A,故反密码子的 5' 位极少出现 A); 而在线粒体内,5' 位仅有不被修饰的 U^[8],极个别时出现 A^[9],它们靠三配二原则(主要在线粒体)与 XYN 4 个密码子相识别。这就说明,识别族性密码子时,通常不易发生错读现象。

当一组内 4 个密码子表示 3 种涵义时,仅有一种分配方式,即 XY_C^U 代表一种氨基酸,其反密码子的 5' 位是 G; XYG 代表另一种氨基酸(若与 XY_C^U 涵义相同,只代表两种涵义,此种情况在密码表中尚未找到),其反密码子 5' 位为 C; 而 XYA 只能代表终止信号^[10],否则将出现错配现象。

同组内 4 个密码子是否可以指令 3 种不同的氨基酸呢? 在天然情况下,一般是不允许的。可以想象到,把 3 个不同氨基酸的密码子编为一组,能排出 12 种不同的分配方式,如表 2 所示。在通常情况下,它们与反密码子相识别时,都有不可解决的矛盾。例如,表中左上角被粗线框起来的一组,识别氨基酸 2 密码子 XYA 的

表 2 假设同组内含三个氨基酸的密码子分配

X	Y				U C A G
	1 1 2 3	1 1 3 2	1 2 1 3	1 2 3 1	
	1 3 1 2	1 3 2 1	2 1 1 3	2 1 3 1	
	2 3 1 1	3 1 1 2	3 1 2 1	3 2 1 1	

注：表中 1、2、3 代表任意三种不同的氨基酸。

反密码子 5' 位为 U，按照变偶原则，它又能与 XYG 配对，而 XY_G^A 在此编码不同的氨基酸，势必引起错读。同理，在右下角粗线框起来的一组内，密码子 XY_C^U 代表不同的氨基酸，识别其中 XYC 的反密码子 5' 位是 G，由于变偶配对，该反密码子也能识别 XYU，一般也要造成错读。表中其他各组都存在上述类似现象，有的是两种情况并存。所以，在天然密码表中没有发现 3 种不同氨基酸的密码子共处在同一组内；换言之，以上 12 种分配方式一般都不成立。至于同组内代表 4 种或 4 种以上不同涵义者，则更不应存在。

假如同组内的 4 个密码子代表两种涵义，除前面提到的一种情况外，还可从两个方面去分析：

第一、由 XY_C^U 表示一种涵义，它们共用 5' 位为 G 的反密码子。另一种涵义则由 XY_G^A 表示，此时，在线粒体外有 5' 位为 U 的共用的反密码子，还有一个 5' 位为 C 的仅与 XYG 相识别的反密码子，但 5' 位不会出现次黄苷^[14]；而在在线粒体内，对于 XY_G^A 来说，往往只有一种反密码子，它的 5' 位是被修饰了的 U（也可记作 U*）。U* 保证了仅与 XY_G^A 的正确识别，而不与 XY_C^U 识别，这与族性密码子和反密码子的

识别不同^[12]。

第二、在同组内表示两个涵义时，还可以由 XYG 代表一种涵义，其反密码子的 5' 位只能为 C，这一点适用于线粒体内和外。另一种涵义则由 XY_C^U 所表示，它们的反密码子是：在

线粒体外 5' 位为 G（识别 XY_C^U ）和 I；而在在线粒体内，因至今尚未发现次黄苷 I，则可能为 G，有人报道已在某些酵母中找到^[9]。对于这第二种分配方式，在线粒体内和线粒体外，一般均无 5' 位为 U 的反密码子，因为它能同时识别代表不同涵义的密码子 XYA 和 XYG，这样，在蛋白质生物合成时，可能引起混乱。然而，5' 位是某种被修饰的 U* 时，后者又专一性地与 A 相识别，则可能是允许的^[10]。

上述各点，着重从密码子与反密码子的碱基（或核苷）按照标准配对、变偶配对和三配二原则加以阐述的。当然，这是主要的，但也不能否定其他因素的作用。例如反密码环的整个序列，tRNA 在核糖体上的正确就位以及识别时某些碱基（或核苷）的特殊贡献，等等。

四、某些密码表

迄今除“通用”密码表外，已陆续发现了一些线粒体内遗传密码表，它们与所述的分配原则十分符合，这为密码子的分配原则提供了佐证，现将部分内容简介如下：

表 3 是常见的、最早发现的、从原核生物到真核生物“通用”的遗传密码表，主要适用于线粒体外。该表中，不存在同组密码子代表 4 种或 4 种以上不同涵义的现象。同组 4 个密码子代表 3 种涵义的仅有第 13 组，但不是指令 3 种不同的氨基酸。代表两种涵义的，一种情况如第 1、9、10、11、12、15 组；另一种情况是第 3 组。代表一种涵义的包括第 2、4、5、6、7、8、14、16 组，即族性密码子。现有资料表明，适用于该表的实例颇多，不必一一介绍，本文仅给出少量文献^[13-16]，以资参考。

进入八十年代以来，报道了一些线粒体遗

表 3 “通用”密码表

	U	C	A	G	
U	F 1----- L	5 S	Y 9----- 终	C 13----- 终 W	U C A G
C	2 L	6 P	H 10----- Q	14 R	U C A G
A	I 3----- M	7 T	N 11----- K	S 15----- R	U C A G
G	4 V	8 A	D 12----- E	16 G	U C A G

注：氨基酸以单字缩写符号表示。“终”为终止信号。

表 4 某些哺乳类线粒体密码表

				C ----- W	
	I ----- M			S ----- 终	

注：表 4—8 空白处的涵义与表 3 同。

传密码表,现把它们列出来,并略加分析。

表 4 是已经找到的某些哺乳类如小鼠^[17]、牛^[18]和人^[12,19,20]的线粒体密码表,它仅包括族性密码子和具有两个涵义的密码子组(属 XY_C^U 和 XY_G^A 各表示一个涵义者)。该表中,有 3 组密码子与表 3 不能完全通用。长臂猿、猩猩、大猩猩和黑猩猩的线粒体密码表,可能也与表 4 相同^[21]。

表 5 为酵母线粒体密码表^[19,22,23],它与表 3、表 4 都有不同之处。值得注意的是,在其他密码表中,CUN 是指令 Leu 的族性密码子,而

在酵母线粒体内则代表 Thr。产生这种现象的原因是,与 CUN 对应的 tRNA 的反密码环由 8 个碱基(或核苷)组成,而通常为 7 个,结果氨酰-tRNA 合成酶使 Thr (而不是 Leu) 与该 tRNA 结合^[24]。尽管如此,它仍符合上述密码子的分配原则。

表 6 是玉米 (maize) 的线粒体密码表^[19],它与表 4、表 5 也不完全相同。

表 5 酵母线粒体密码表

				C ----- W	

表 6 玉米线粒体密码表

				C ----- W	
				R ----- W	
	I ----- M			S ----- 终	

表 7 为黄猩猩果蝇 (*Drosophila melanogaster*) 的线粒体密码表^[25],与表 4 比较,密码子 AG_G^A 所代表的涵义可能不同。

表 8 是链孢霉类的线粒体密码表^[26],它与上述各表又有不同之处。

表 7 黄猩猩果蝇线粒体密码表

				C ----- W
	I ----- M			S

表 8 链孢霉线粒体密码表

				C ----- W

对线粒体翻译系统的研究虽已取得了可喜的进展,但主要还是近几年的事,在某些问题上不同的作者仍有各自的看法。不过,从现有的资料可以看出,在线粒体内涵义容易发生变化的密码子是第 13 组,其次是第 2、3、14、15 组等。尽管在各个密码表之间都有不同之处,不能完全通用,但它们的密码子分配都与本文所叙原则相一致,在与反密码子相识别时,能保证翻译的忠实性。关于叶绿体内遗传密码表和密码子的分配,就笔者所见资料均和“通用”密码表相同^[27-29],是否存在某种特殊情况,尚需根据更多的实例再作判定。

* * *

基于标准配对、变偶配对和三配二原则,本文提出了密码子在线粒体内遗传密码表中的分配原则。讨论这一问题,有助于加深认识密码子与反密码子相互作用的规律;能够大体上预测新的密码表,它们也应符合这些原则,如果相悖,则暗示可能存在某种新的值得探讨的问题。了解密码子的分配规律,对有关的科研和教学等工作都有裨益。

承杨志铭教授审阅全文,谨致谢意。

参 考 文 献

- [1] Sprinzl, M. et al.: *Nucl. Acids Res.*, 13 (Supplement), r1, 1985.
- [2] 王德宝等:《中国科学》B 辑, 5, 387, 1983.
- [3] 曲善乐:《生物化学与生物物理进展》, 6, 2, 1985.
- [4] Clayton, D. A.: *Annu. Rev. Biochem.*, 53, 573, 1984.
- [5] 杨雨善等:《国外医学分子生物学分册》, 4, 156, 1985.
- [6] Lagerkvist, U.: *Proc. Natl. Acad. Sci. USA*, 75, 1759, 1978.
- [7] Lagerkvist, U.: *Cell*, 28, 305, 1981.
- [8] Heckman, J. et al.: *Proc. Natl. Acad. Sci. USA*, 77, 3159, 1980.
- [9] Bonitz, S. G. et al.: *ibid.*, 77, 3167, 1980.
- [10] 杨雨善等:《国外医学分子生物学分册》, 3, 111, 1985.
- [11] Jukes, T. H.: *Science*, 198, 319, 1977.
- [12] Barrell, B. G.: *Proc. Natl. Acad. Sci. USA*, 77, 3164, 1980.
- [13] Bedwell, D. et al.: *Nucl. Acids Res.*, 13, 3891, 1985.
- [14] Reynolds, P. et al.: *ibid.*, 13, 2357, 1985.
- [15] Choe, J. et al.: *Mol. Cell. Biol.*, 5, 3261, 1985.
- [16] Youngblom, J. et al.: *ibid.*, 4, 2686, 1984.
- [17] Bibb, M. J. et al.: *Cell*, 26, 167, 1981.
- [18] Anderson, S. et al.: *J. Mol. Biol.*, 156, 683, 1982.
- [19] Jukes, T. H.: *J. Mol. Evol.*, 18, 15, 1981.
- [20] Anderson, S. et al.: *ibid.*, 290, 457, 1981.
- [21] Brown, W. M. et al.: *J. Mol. Evol.*, 18, 225, 1982.
- [22] Hudspeth, M. E. S. et al.: *Cell*, 30, 617, 1982.
- [23] Coruzzi, G. et al.: *J. Biol. Chem.*, 256, 12780, 1981.
- [24] Sibley, A. P. et al.: *FEBS Letter*, 132, 348, 1981.
- [25] de Bruijn, M. H. L.: *Nature*, 304, 234, 1983.
- [26] Mainwaring, W. I. P. et al.: *Nucleic Acid Biochemistry and Molecular Biology*, Black-well Scientific Publications, Oxford London, p 475, 1982.
- [27] Dron, M. et al.: *J. Mol. Biol.*, 162, 775, 1982.
- [28] Montandon, P. E. et al.: *Nucl. Acids Res.*, 11, 5877, 1983.
- [29] Karabin, G. D. et al.: *ibid.*, 12, 5801, 1984.

[本文于 1986 年 5 月 6 日收到]