

遗传密码子和均一手性的起源问题

肖敬平

(华南农业大学生物技术学院, 广州, 510642)

摘要 针对生命起源研究领域, 遗传密码及手性均一性两大难题, 提出了新的思路。作者认为: 原始密码子应是“3 读 2”三联体。由于某种氨基酸-tRNA-寡核苷酸复合体之间的聚合反应, 导致原始密码子顺序同其编码二肽的同步起源。它的多态性则反映了进化轨迹, 是原始密码子第 3 位碱基, 分阶段分化的结果; 同时还认为, 手性均一性是周转在代谢系统中的生命分子的独特行为。它必然同代谢系统同步起源。原始类似糖生成途径可能起到关键性作用。

关键词 手性均一性; 遗传密码; 密码子; 生命起源; tRNA; 蛋白质

中图分类号 Q 493

50 年代初, Miller(1953)在模拟原始地球条件的试验系统中, 以含有 C, H, N 和 O 原子的小分子为原料, 意外地获得了氨基酸等多种生物分子, 从而激起了探索生命起源的高潮。此后, 关于生命起源的研究取得了巨大的进展。生命通过化学进化, 起源于原始地球上非生物世界的概念, 从此便牢固地树立起来。然而, 还有几个带基本性质的问题, 至今未能取得突破。它们已成为这个领域发展的巨大障碍。其中最令人感到困惑的难题, 有 2 个: 1 个是, 遗传密码的起源问题。究竟是一种什么样的神秘自然力, 使得 4 类碱基组成的三联体密码子, 去编码 20 种氨基酸? 另 1 个是, 生命分子的均一手性(homochirality)问题。为何在物理世界中, 氨基酸和糖分子的对映异构体是同时存在的, 而在生命系统中, 它们却是均一的? 除个别例外, 氨基酸都是 L 型, 而糖都是 D 型。在分子生物学已高度发展的情况下, 困难主要不可能来自技术方面, 而是由于缺乏富有突破性的思路, 以致迄今所提出的多种假说, 虽各有其特点, 但尚未达到可以确立的程度。为此, 本文将着重讨论有关的科学思路问题。

1 关于遗传密码的起源

遗传密码的起源是生命起源领域里的热点。学者一致认为, 这个在原核细胞, 真核细胞乃至病毒都通用的密码子系统, 反映了进化的轨迹。这个系统的主要特点有 2: (1)多态性。20 个被编码的氨基酸, 各自的密码数子极不相同。最多的, 如丝氨酸等, 有 6 个密码子, 而最少的, 如旦氨酸, 只有 1 个; (2)相关性。Woese(1967)首先发现, 氨基酸同其密码子之间有明显的相关。例如, 第 2 碱基为 U 的密码子, 它的编码氨基酸都是嫌水性的(丝氨酸, 脯氨酸, 苏氨酸, 丙氨酸); 而为 A 的密码子, 则包含仅有的 2 个为酸性氨基酸(天冬氨酸和谷氨酸); 另一个令人深思的特点是, 相似的氨基酸也有相似的密码子。Pelc 和 Welton(1967)发现, 氨基酸每改变 1 个原子或基团, 其密码子也改变 1 个碱基。例如, 丙氨酸(密码子: GCU)加 1 个 O 原子转为丝氨酸(密码子: UCC), 其密码子第 1 位碱基则由 G 转为 U; 减 1 个 -CH₂ 基团成为甘氨酸, 其第 2 位

碱基,则由 C 变为 G. 作者认为,多态性更能反映遗传密码进化过程的阶段性.

1.1 遗传密码子的多态性的分析

与人工密码不同,作为分阶段进化的产物,氨基酸的编码,并无既定之规. 丝氨酸(Ser),亮氨酸(Leu)和精氨酸(Arg)各有 6 个密码子;脯氨酸(Pro),苏氨酸(Thr),缬氨酸(Val),丙氨酸(Ala)和甘氨酸(Gly)各有 4 个;异亮氨酸(Ile)和终止密码(Stop)各有 3 个;苯丙氨酸(Phe),酪氨酸(Tyr),半胱氨酸(Cyt),组氨酸(His),谷氨酸(Glu),谷氨酰胺(Gln),天冬氨酸(Asp),天冬酰胺(Asn),赖氨酸(Lys),各有 2 个;而色氨酸(Trp)和蛋氨酸(Met)仅各有 1 个. 对这个奇特现象,前人归诸于偶然性. 然而,作者认为这正是研究密码子进化的切入点. Orgel(1989)推测,原始密码子是“3 读 2(two out of three)”三联体,只有头 2 位碱基编码. 为了便于分析,作者将编码同一氨基酸的密码子定义为同功密码子(isocodon);头 2 位碱基都相同的密码子,表明他们有同源(homologus)关系,如 Phe 的密码子:UUU, UUC 以及 Leu 的密码子:UUA, UUG. 头 2 位不完全相同的,则称为异源(heterologus). 现以第 3 位碱基 N 的分化水平为依据,将密码子的进化阶段列于表 1.

表 1 遗传密码子的多态性及进化水平

密码子(XYN)中,N 碱基分化水平	密码子进化水平
I N 未分化,有 4 个同源同功密码子 ^a UCN(Ser); ^b CUN(Leu);CCN(Pro); ^c CGN(Arg); ACN(Thr); GUN(Val); GCN(Ala); GUN(Gly)	由于语义确切,N 未分化,第 3 位碱基可有 4 种选择,故有 4 个同功密码子. 实质上保留了“3 读 2”原始形态. 属于原始密码子族.
II N 初步分化,有 2 个同源同功密码子 CAP _y (His), CAP _u (Gln); UUP _y (Phe), ^b UUP _u (Leu); AAP _y (Apn), ^c AAP _u (Lys); ^a ACP _y (Ser), ACP _u (Arg); UAP _y (Tyr), UAP _u (Stop); UGP _y (Cys), UGP _u (Stop 或 Trp)*; ^d AUP _y (Ile), AUP _u (Ile 或 Met)*	其余语义不清的原始密码子,N 初步分化为 P _y (吡啶)和 P _u (嘌呤),开始出现“3 读 3”密码子. 由于第 3 位碱基具有变偶性,有 2 个可能碱基,它们仍各有 2 个同源同功密码子. 属于中期的密码子族. 至此,只剩下 2 组仍有共用现象的密码子(*)是语义不清的,有待最后分化
III N 已最后分化,再没有同源同功密码子 ^d AUA(Ile);AUG(Met);UGA(Stop);UGG(Trp)	N 已最后分化为 A 和 G,不再有同功密码子. 属后期密码子族. 至此语义不清现象完全消除

1) a, b, c 尚有 2 个属于 II 类的异源同功密码子,共有 6 个同功密码子; 2) d 尚有 1 个属于 III 类的异源同功密码子,共有 3 个同功密码子

表 1 的分析,清楚地表明了“3 读 2 原始密码子”分阶段进化的过程. 由于用 4 类碱基组成的三联体有 64 种组合,对 20 种氨基酸必然会有大量的同功密码子. 然而,原始密码子只有头 2 个碱基编码,因此只能有 16 种组合,就必然有部分密码子被共用而出现语义不清现象(ambiguity). 其中非共用的,语义确切的一族,始终保持着 3 读 2 密码子的原始形态. 当第 3 位碱基 N,分 2 个阶段最后分化为 A 和 G 后,语义不清的原始密码子共用现象便最后消除了.

1.2 关于原始密码子起源的思路

60 年代中期, Woese(1967)提出氨基酸与其密码子立体化学配对假说(简称立体化学假说). 除了发现第 1, 2 位碱基与编码氨基酸的性质有关,他们还发现,相似的氨基酸也具有相似的密码子,即:氨基酸每增加 1 个原子或基团,其密码子也相应改变 1 个碱基. 例如,Ala 增加一个 O 原子转为 Ser,其密码子也由 GCN 变成 UCN,如再增加 1 个-CH₂ 基团转化为 Thr 后,密

密码子也相应变为 CAN, 等等 (Pelc et al, 1966). 据此, 他们相信密码子与其编码氨基酸是全面配对 (solid pairing) 的. 问题是, 虽几经尝试, 却拿不出经得起批评的配对模型来. 随后, DNA 双螺旋结构发现者, Crick (1968) 又另提出, 原始密码子是由于原始 tRNA 分子中, 反密码子结构与该分子识别特定氨基酸能力偶然对应, 而后在进化过程中冻结了下来, 至于何似相似的氨基酸一定具有相似的密码子, 他推测, 由于 tRNA 在其反密码子某个碱基发生变化的同时, 其识别氨基酸的能力也发生了识别相似的氨基酸变化 (Crick, 1968). 然而, 对原始密码如何起源的种种疑难, 统统以偶然事件来解释, 就连 Crick 本人也承认“未免太过于迁就 (too accommodative)”, 故而是该假说的主要缺点. 以后提出的各种假说, 基本上属于这两大类. 其实, 从基本点看, 两派的观点是互补的. 对原始密码子的起源, Crick 也认为, 在进化初期氨基酸与核苷酸之间必然存在着某种并非编码关系的立体化学或物理化学性质的联系. 作者参考众家之说, 将原始密码子的起源, 初步分为以下 3 个阶段:

(1) 阶段 1: 氨基酸-寡核苷酸对接体的形成. 虽多数学者都承认, 原始密码子的前体与其编码氨基酸之间存在某种关系. 然而, 究竟是何种关系, 就只有立体化学派提出过全面配对模型 (solid pairing). Pelc 和 Welton (1966) 宣称, 以密码子的第 1 位碱基与编码氨基酸的羧基配对, 第 2 和第 3 位碱基依次与侧链配对, 成功地建立起了一套氨基酸与其密码子的立体化学配对模型. 但是, Crick (1966) 随即指出, 虽然该模型构建方法大体上遵循了立体化学原则, 但配对后的三联体的阅读方向完全颠倒了. 例如, 模型将与赖氨酸 (Lys) 配对的三联体 GAA, 倒向错读为它的密码子 AAG. 而 GAA 其实是另一氨基酸, 谷氨酸 (Glu) 的密码子. 尽管有不少支持此观点的学者认为, 这种原始的配对是可信的, 倒向现象发生在进化过程的某个环节上. 但由于无法给出具体解释, 该模型就被完全否定了. 作者认为, 该学派的主要错误在于: (1) 寻求建立将密码子的 3 个碱基与其编码氨基酸完全配对模型, 走了极端. 以后有不少学者作出原始密码子可能是“3 读 2”的, 合理推测; (2) 误认为寡核苷酸与氨基酸在早期结合时, 就已经具有编码功能. 这就把编码功能的复杂进化过程简单化了. 针对这些特点, 作者推测, 在原始浓汤中的寡核苷酸倾向于同氨基酸结成各种形式的复合体, 以取得比较稳定的生存空间. 其中有少数复合体偶然采取了, 以顶端二联体 N_1N_2 与决定氨基酸特点的侧链 R, 相对接的形式, 为 3 读 2-原始密码子的起源奠定了基础, 示意如下:



(2) 阶段 2: 原始 tRNA 分子的起源. 持进化观点的学者, 大都强调原始 tRNA 分子在遗传密码的起源上起着关键作用. 然而, 一如前述, 最大的困难乃是, 什么机理使得 tRNA 分子的反密码子与所识别的氨基酸之间, 发生某种结构上的联系? 作者认为, 如果把 tRNA 分子的起源, 视为 tRNA 与氨基酸-寡核苷酸对接体相互选择的结果, 问题就不难想象了. Orgel (1989) 推测, 在进化的某个阶段上, 可能出现了某种发夹式结构的聚核苷酸分子, 能够与特定的氨基酸相结合. 作者进一步推测, 在这类分子群体中, 偶然存在一类特殊的分子, 其两端可分别与对接体的氨基酸和寡核苷酸结合, 形成某种“氨基酸-tRNA-寡核苷酸复合体” (tRNA 复合体) (肖敬平, 1998), 如图 1 所示. 这样一来, tRNA 分子中的反密码子, 与识别的氨基酸的对应关系, 便建立起来了. 不过, 应强调的是, 由于寡核苷酸并非必然是三联体, 因此还算不上是密码子. 至于 tRNA 复合体的功能, 也仅仅是形成了某种为稳定的结构而已.

(3) 阶段3:原始密码子及多肽的同步起源. Reanny 和 Ralph(1967)曾推测,氨基酸和密码子的起源必然是难解难分地交织在一起的. 最近,作者提出了蛋白质与基因同步起源的观点(肖敬平,1998). 以2个 tRNA 复合体中的 tRNA 核心为亚基,构成一种聚合核酶,同步催化氨基酸之间,以及核苷酸之间的聚合. 由于三联体较之其它寡核苷酸,在空间上占据较为有利的位置,最终成为优胜者. “3 读 2-密码子顺序便与二肽同步现了(图1).

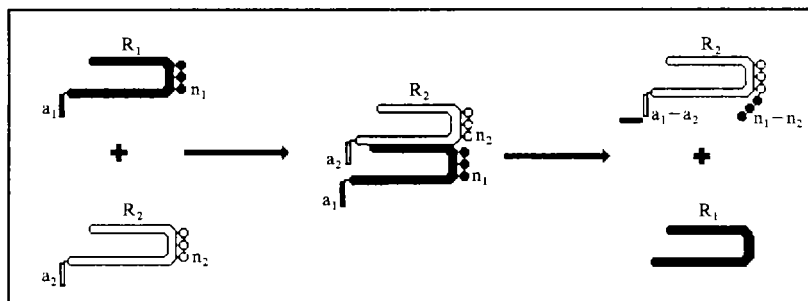
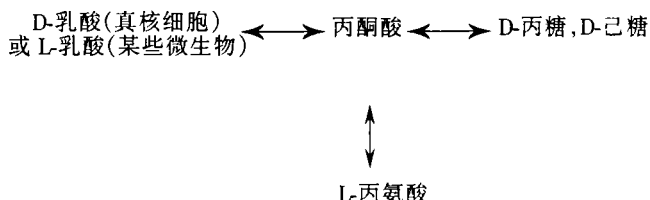


图1 二密码子顺序及其编码二肽的同步起源

2 关于手性的来源

生命分子的手性问题,是化学进化领域的热门课题,并已经取得了重要进展,证明了在一定条件下,β射线能对氨基酸分子进行非对称分解. 王文清等(1993)指出,这种分解只能在低温和远离平衡点的开放条件下,才可能观察到. 此论断对探索生命系统中生命分子手性均一行为有重要意义. 因为,生命系统正是与环境有物质和能量交换,且具有远离平衡点的性质. 在作进一步探索之间,先要明确生命分子与一般生物化学分子的区别. 作者曾强调指出:只有周转在代谢系统中的生物化学分子才能表现出生命行为,才是生命分子. 因此,它们也只能与代谢同步起源(肖敬平,1998). 既然均一手性是生命分子的特点,它也只能与代谢同步起源. 用这种新的思路来观察细胞内糖生成途径中的手性变化,就会发现,无论D型还是L型的乳酸,在通过非手性的丙酮酸后,最终都转化为D型丙糖和己糖. 如果再检查一下 Miller(1953)模拟原始大气层的试验结果,就会注意到,在它的多种多样的产物中确实存在着糖生成途径的底物—乳酸. 这启示,在原始浓汤中极可能存在过某种类似糖生成途径的反应. 它对代谢的起源以及均一手性的生命分子的起源,都起到关键性作用. 这也为生命分子与代谢同步起源的观点,提供了有力的证据. 更值得注意的,当丙酮酸进一步胺基化形成丙氨酸时,产物就不再是D型分子,而是L型丙氨酸,从而实现了生命分子从D型向L型的转化:



看来,代谢网络中的丙酮酸,正处在均一手性的生命分子的起源,以及不同手性生命分子

相互转化的十字路口上. 非手性的丙酮酸, 在不同驱动力作用下向不同方向转化时, 表现出不同的手性行为. 具体地说, α -羰基在转化为手性基团时, 在磷酸化和水合作用下, 产生 D 型化合物; 在胺化作用下, 则产生 L 型; 在还原为乳酸时, D 型和 L 型 2 种对映异构体都可能产生. 这很可能是解开生命分子手性均一性之谜的突破口.

参 考 文 献

- 肖敬平. 1998. 细胞和病毒起源问题研究的进展. 华南农业大学学报, 19(2): 117 ~ 122
- Crick F H C. 1968. The origin of the genetic code. J Mol Biol, 38: 367 ~ 378
- Giulio M D. 1998. Regulations on the origin of genetic code: a hypothesis. J Theol Biol, 191: 190 ~ 196
- Miller S L. 1953. A production of amino acids under possible primitive earth condition. Science, 11: 528 ~ 529
- Orgel L E. 1989. The origin of polynucleotide-directed protein syntheses. J Mol Evolution, 29: 65 ~ 74
- Pelc S R, Welton M G. 1966. Origin of genetic code. Nature, 209: 868 ~ 872
- Reanny D C, Ralph R K. 1967. Origin of code. J Theoret Biol, 15: 41 ~ 45
- Wang W Q, Wu J L, Jiang J. 1993. Differences in radiolysis behavior of D, L-amino acid in primary stage and thermodynamic equilibrium state. In: Ponnamperuma C, Chela-Flore J, eds. Chemical evolution: Origin of life. Hampton: A. Deepak Publishing, 249 ~ 260
- Woese C R. 1967. The genetic code: The molecular basis for genetic expression. New York: Harper. 1 ~ 75
- Wong J T. 1975. A co-evolution theory of genetic code. Proc Natl Acad Sci U S A, 72: 1 909 ~ 1 912

Origin of Genetic Codon and Homochirality

Xiao Jingping

(College of Biotechnology, South China Agric. Univ., Guangzhou, 510642)

Abstract A new way of thinking about genetic code and homochirality, which remain the major difficulties in the field of the origin of life, is suggested. The author reasons that procodons must have been two-out-of-three triplets. They originated synchronously with the dipeptides coded thereby via polymerization of certain amino acid-tRNA-oligopeptides complexes. The polymorphism of genetic code reflecting the evolutionary track is the outcome of differentiation of the third position bases of the procodon, step by step; and that homochirality, as an unique behavior of life molecules turning over in metabolic system, must have originated synchronously with the latter. A key role could have been played by the primitive glycogenesis-like pathway.

Key words homochirality; genetic code; codon; origin of life; tRNA; protein

【责任编辑 李 玲】