

DOI:10.12405/j.issn.2097-1486.2025.03.005

遗传密码的历史溯源

——对“菱形密码”和“无逗点密码”的哲学思考

杨维恒

山西大学 科学技术哲学研究中心, 山西 太原 030006

摘要:“遗传密码”可以说是生物学中最常见的隐喻之一。本文首先对生物学中早期“密码”概念的引入进行了分析,展现了在遗传密码形成初期,“密码”隐喻在科学共同体中是如何实际使用并推广的。之后,从生物学史的角度,对遗传密码形成初期两个巧妙的理论假说——“菱形密码”与“无逗点密码”进行了具体的案例分析。通过案例分析,展现了那一时期的“密码”概念具有抽象性的特征,并对蛋白质合成问题的解决起到了描述性和解释性的作用。

关键词:遗传密码;菱形密码;无逗点密码

中图分类号:N031

文献标志码:A

文章编号:2097-1486(2025)03-0282-07

Historical origins of the genetic code: Philosophical reflections on "diamond code" and "comma-free code"

YANG Weiheng

Research Center for Philosophy of Science and Technology, Shanxi University, Taiyuan 030006, China

Abstract: "Genetic code" is arguably one of the most common metaphors in biology. This paper first analyzes the introduction of the early concept of the "code" in biology, demonstrating how the metaphor of "code" was actually used and promoted within the scientific community during the early stages of the genetic code's formation. Subsequently, from the perspective of the history of biology, it conducts a detailed case analysis of two ingenious theoretical hypotheses that emerged during the early formation of the genetic code—the "diamond code" and the "comma-free code". The case studies reveal that the "code" concept of that period is characterized by an abstract feature, and it also plays a descriptive and explanatory role in addressing the problem of protein synthesis.

Keywords: genetic code; diamond code; comma-free code

0 引言

“遗传密码”可以说是生物学中最常见的隐喻之一。如果说“信息”隐喻“诱导”了分子生物学的概念结构,那么反过来可以说是“密码”隐喻将信息概念正式嵌入了分子生物学的概念框架。虽然,“遗传密码”在分子生物学的语言中早已被普遍使用。但是,在生物学哲学中,“遗传密码”语义语言的使用一直存在着广泛的争议。例如,“遗传密码”明显的语义概念的使用是否合法;它所指的性质是什么;是否包

含某些特殊的语义特征;是否发挥了有价值的理论作用;抑或只是一个空的隐喻等,都一直是生物学哲学关注和讨论的问题。有的人认为它的使用是不恰当的、是具有误导性的;而有的人则通过一些好的动机来对它进行辩护。

本文主要是从生物学史的角度,对遗传密码形成初期两个巧妙的理论假说——“菱形密码”(diamond code)与“无逗点密码”(comma-free code)进行具体的案例分析。通过对早期“密码”隐喻具体使用的分析,试图展现,为什么在很短的时间内蛋白质合

基金项目:教育部人文社会科学研究青年基金项目(19YJC720041)

收稿日期:2025-08-20;**修回日期:**2025-09-12

成的问题能够被迅速地还原成抽象的符号操作问题;当DNA和蛋白质之间的关系迅速被构象为一种编码时,“密码”隐喻在科学共同体中又是如何被实际使用并推广的。当然,在这个过程中,有很多巧妙的“密码理论”被提出,然而其中最具有代表性的莫过于伽莫夫(George Gamow)的“菱形密码”和克里克(Francis Harry Compton Crick)的“无逗点密码”。尽管它们最终都失败了。但是,通过对它们具体内容的分析和对比,可以发现“密码”隐喻在当时实际使用中所发挥的具体理论作用。迈尔(Ernst Mayr)曾经指出,“学习一门学科的历史是理解其概念的最佳途径。只有仔细研究这些概念产生的艰难历程——即研究清楚早期的,必须逐个加以否定的一切错误假定,也就是说弄清楚过去的一切失误——才能有希望真正彻底而又正确地理解这些概念”^[1]。而文章的目的就是希望通过对遗传密码形成初期历史案例的分析,产生一些对当代遗传密码概念有益的哲学思考和结论。

1 早期“密码”概念的引入

最早提出生物学中编码方案的是奥地利物理学家薛定谔(Erwin Rudolf Josef Alexander Schrödinger),他在《生命是什么》一书中提出了密码论并将遗传密码与莫尔斯电码进行了类比。在这本书中他提到“一个生物体及其所经历的所有生物学相关过程,必须具备一种极其‘多原子’的结构,才能避免因偶然的‘单原子’事件而受到太大的影响。只有这样,生物体才能遵循足够精确的物理学定律,从而执行让人叹为观止的有规律、有秩序的运作方式。”“许多小得令人难以置信的原子团,它们所包含的原子数少到无法服从精确的统计学定律,却真真切切地在生物体内有规律、有秩序的事件中发挥着关键作用。这些小小的原子团控制着生物体在生长发育过程中获得的可被观察到的宏观性状,还决定着生物体功能的重要特征。而所有这些性状和特征都遵循着精准而严格的生物学定律”^[2]。进而在书中他提出一个概念性的密码论,其核心观点是:高层次系统无法依据已知的传统科学理论直接还原为低层次系统。这两个层级之间存在着大量我们尚未完全理解的“密码关系”,必须通过某种形式的“翻译”或“解码”过程,才能与我们所熟知的物理化学规律相联系。在生物学领域,这种密码关系的最基础表现是遗传密码,而其最复杂、最高级的体现则是生物体及其精神活动。他在书中“压缩在微型密码中的丰富多彩

的内容”^[2],给出了遗传密码的概念性方案。为了把问题讲清楚,他提到“让我们以莫尔斯电码类比。莫尔斯电码中有‘点’和‘划’两种不同的符号,把这两种符号有序地排列组合,只需使用不超过4个符号就能表示30种不同的含义。如果除了‘点’‘划’再加上第三种符号,只需要不超过10个符号就能产生88 572种不同的‘字母’;如果总共有5种符号,每种组合的符号数量不超过25个,那么可能的排列组合方式有372 529 029 846 191 405种”^[2]。

然而,这只是一个概念性的编码方案。真正将编码方案具体带入生物学的是美籍俄裔理论物理学家伽莫夫。1953年,DNA双螺旋结构被提出之后,伽莫夫及其同事在很短的时间内便将蛋白质合成的问题还原成符号操作的问题,从而将编码的工作迅速带入了分子生物学。想要了解当时的工作,我们首先要认清他们当时所面临的研究问题——确定氨基酸序列的线性排列。

20世纪40年代,蛋白质的合成问题就已经进入了生物学家的研究视野。并且,确定氨基酸序列的线性排列在当时已经被视为是问题的关键。在DNA双螺旋结构被确立之前,已经有两组假设来用于解释这一问题:1)氨基酸的特异性序列是由酶促反应导致的,相邻氨基酸的连接是由酶的催化进行,不同的酶催化不同类型的氨基酸连接;2)氨基酸链是由游离的氨基酸和预先存在的蛋白质或核酸模板的暂时连接形成的。1953年,DNA结构被确立之后,伽莫夫在DNA双螺旋结构的基础上,将序列的问题发展成两种不同的分子间的映射。1953年10月,他在给鲍林(Linus Pauling)的信中写道“从我去年6月读了沃森和克里克的文章,我就一直尝试去计算,一个写在4位数系统(即核酸分子)中的长的数字如何能够特异性地确定一个相应长的基于20个字母的(即一个酶分子)单词”^[3]。不难发现,伽莫夫的表述中暗含着序列决定的思想,即蛋白质的氨基酸序列完全是由DNA双螺旋结构上的碱基序列决定的。尽管没有明确的证据支撑,但是这一假设很快在当时的科学共同体中达成共识。1953年,沃森(James Dewey Watson)在手稿中写道“我们相信二级结构不是相关的主要论点(二级结构对主要问题是不相干的)。这个‘二级结构’涉及核酸模板的构造(而不是碱基序列),主要论点是由核酸确定的氨基酸序列。”克里克也明确表示“氨基酸的顺序是由核酸的核苷酸顺序决定的。在蛋白质中通常可以发现20种天然存在的氨基酸,但是(通常)只有四

种不同的核苷酸。四种事物(核苷酸)的序列如何能够决定20种事物的序列的问题被称为‘编码’的问题^①。1954年—1958年之间,公开发表的有关蛋白质合成的文章中也都直接或间接表明对这一假设的支持。也正是在这一思想的基础上,克里克在1958年提出了著名的序列假说,即蛋白质的三维结构完全是由其氨基酸的序列导致的,而氨基酸的序列则是由DNA核苷酸的序列决定的。

可以看出,起初科学共同体内对“序列决定”思想的迅速认同,使得DNA与蛋白质结构之间的对应关系,被迅速转化为一个关于符号运算的抽象问题。分子层面的复杂细节因此被暂时搁置,研究目标最终被简化为两组不同符号之间的数学映射问题。

尽管此时已经将解决方案确定为映射或编码方案,但是“密码”并不是所涉及的唯一术语。除此之外还有“方案(scheme)”“对应(correspondence)”“分配(assignment)”等。例如,最初伽莫夫所使用的表述有“建议的方案(the proposed scheme)”“独特的对应(unique correspondence)”等;克里克在这一时期(1955年)经常会使用“方案(scheme)”这一术语等。然而,通过对这一时期他们的私人信件和个人记录手稿的分析,可以发现,尽管在公开发表的文章中会有很多不同的术语表述,但是在一些非正式或半正式的研究和交流中,如私人信件的往来、个人研究成果的交流、内部会议的讨论等,“密码”是他们经常使用的表述。他们将“密码”与一个前缀结合使用,表示某一种特定的假设方案,以区别于不同的假设,如菱形密码(Diamond code)、无逗点密码(Comma-free code)、三角密码(Triangular code)、组合密码(Combination code)、连续密码(Sequential code)等^②。

显然,在最初表达映射关系时,“密码”隐喻使用的选择具有一定的任意性。针对这一问题,有人从社会政治的角度提出,“‘密码’的使用很大程度上是历史的偶然事件,它是由二战期间对密码分析大范围的使用和研究者个人之间的联系——如伽莫夫和美国军队的关系,促进的”^[5]。从这个视角看,密码隐喻只是提供了一个备用的符号。也有人从因果关系重新解释的角度指出,遗传密码可以被证明是一个多余的概念。例如,韦伯(Marcel Weber)对位置

信息概念的有用性提出质疑,指出用纯粹的因果术语可以对信息进行清晰的表达,当然,他也谨慎地指出了,对因果关系重构的可能性本身并不意味着“密码”概念就完全不重要了^[6]。相同的,苏斯塔(Peter Sustar)也通过因果术语对中心法则重新进行了解释,强调可以用纯粹的因果关系对信息进行解释^[7]。

但是,通过对“密码”术语的分析可以发现,密码这一人类符号的术语,在分子生物学中的应用不仅反映了序列间的映射关系,更是表达了一种从一类序列到另一类序列的映射思想。相比于其他术语,这对解决蛋白质合成中编码问题的映射是更为恰当的。其次,通过对“密码”术语在特定科学语境下实际使用的分析,可以发现,20世纪60年代,科学共同体开始在蛋白质合成早期的研究语境中使用“密码”术语,尤其是在目的是用于内部交流和个人记录时。

同时,我们认为,“密码”隐喻最初在科学共同体中的使用只是为了方便式的表达。出现这种方便式的表达是因为他们具有专业的技能,能够对“密码”的内涵进行特定的区分。尽管在当时,“密码”并没有一个准确的概念,甚至是一个可变的内容,但是在科学话语中它的内容始终是明确的。序列问题的明确引出了编码方案在科学共同体内部的实际使用,进而逐步巩固了密码隐喻。也就是说,在20世纪60年代初遗传密码理论被阐明之前,“密码”隐喻早已被实际使用,并已经发挥了其特定的作用。而随着科学认知的不断深化以及“密码”这一隐喻被反复运用,它所促成的语义共识,最终固化为界定遗传密码理论实体及其关联的规范性框架。尽管作为引入概念的原始隐喻过程已然完成,但该隐喻语词的内涵在认知重构中被自然化,并由此获得了新的理论阐释与说明。

2 从术语引入到理论假设

在20世纪60年代,曾经有各种各样的“密码”理论假设。所有的方案都拥有一个核心的特征,即氨基酸链的序列是由核苷酸链的序列决定的。它们大致可以分为“重叠性”和“非重叠性”两类。而“菱形密码”与“无逗点密码”莫过于分别是这两类方案

① 当时的编码问题指的是,蛋白质中的氨基酸有20多种,但是DNA中的碱基只有4种。因此,从碱基到氨基酸无法形成任何一个一对一的映射。即便使用两个一组的碱基组,也只有16组的碱基,同样无法与氨基酸形成对应映射。因此,遗传密码中信息的碱基单位似乎是不能少于三个一组的碱基组。但是这样会有64组三个一组的碱基组,超过了需要数目20的三倍。如何解释这些多余的碱基组成为当时密码理论家的一个当务之急。

② 可以通过维尔康姆网上图书馆(wellcom library)查看当时电子版的手稿和信件档案。

的典型代表。

2.1 菱形密码

1954年伽莫夫在*Nature*上发表了题为《脱氧核糖核酸和蛋白质结构之间可能的关系》(*Possible Relation Between Deoxyribonucleic Acid and Pro-*

tein Structures)的文章。在这篇文章中他提出了“菱形密码”的理论假设。尽管在整篇文章中并没有出现“code”一词。但是,伽莫夫在当时的手稿中明确使用了“菱形密码”这一术语(伽莫夫关于菱形密码手稿部分内容见图1)。

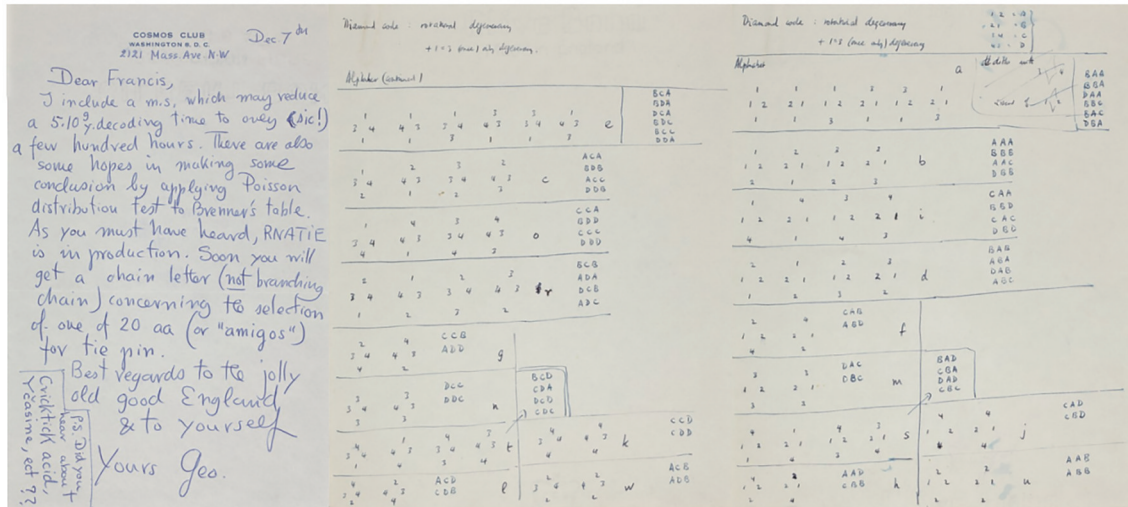


图1 伽莫夫给克里克的信以及部分菱形密码手稿

Figure 1 Gamow's letter to Crick and a portion of Gamow's manuscript on the "diamond code"

“菱形密码”假说认为,双链DNA本身即可充当合成蛋白质的直接模板。其机制是:沿着双螺旋的凹槽,不同的碱基组合会形成形状各异的“模腔”(hole),特定的氨基酸侧链会因空间匹配而嵌入其中。每个模腔会选择性地“捕获”与其构型互补的氨基酸。当这些氨基酸沿着凹槽按序排列后,一种酶会将它们连接成肽链。该模型的核心在于,每个模腔的三维结构由一个“菱形”区域的四个碱基决定。若将DNA螺旋视为垂直方向,则菱形的上下两个顶点由同一条链上的碱基构成,中间隔着一个碱基;而左右两个顶点则由另一条链上的互补碱基对占据(菱形密码图见图2)。

伽莫夫指出,沿着水平对角线的碱基是受1—2以及3—4的绑定条件限制的。因此,DNA双螺旋的每个“腔”都是由菱形4个角上的3个所决定的。这样就很容易会得到20个不同类型的“腔”。前8个比较简单,上下对称。剩下12个以右转或左转的两种变换形式存在。而这20个“腔”,每一个都对应一个特定的氨基酸。这意味着“菱形密码”具有双向对称性:无论从哪一端读取,或将其翻转(镜像),其对应的氨基酸信息保持不变。例如,序列CAG反向读取为GAC,二者编码同一种氨基酸。若将CAG镜像翻转,根据碱基互补配对原则,中间的A变为T,

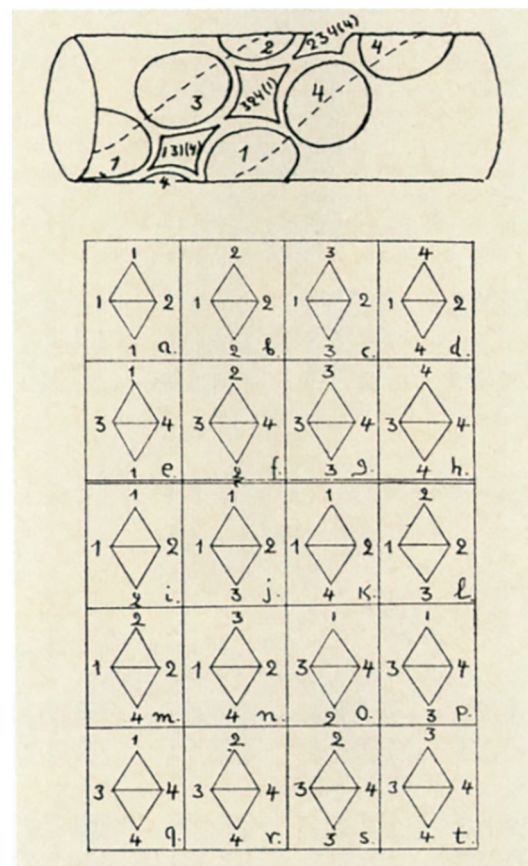


图2 伽莫夫的菱形密码组

Figure 2 Gamow's diamond code

得到序列CTG,其反向互补序列GTC也是与CAG编码相同的氨基酸。这样一来,由菱形4个角上的3个碱基决定的密码子共有64个,如果将以上全部的对称性都考虑进去时,这64个密码子恰好可以分为20组,对应20个氨基酸。而20也正是伽莫夫一直所要寻找的特殊数字。

若干年后,克里克对伽莫夫关于菱形密码的工作是这样评价的:“伽莫夫工作的重要性在于它确实是一个抽象的编码理论,不会被很多不必要的化学细节所干扰……”^[8]。客观而言,与后来提出的诸多密码理论相比,伽莫夫对“菱形密码”的描述在化学层面上确有其模糊之处。尽管如此,该模型所展现出的高度抽象性与理论洞察力,至今仍令人叹服,并持续产生着深远影响。特别是他对“不匹配字母”问题的开创性处理,至今仍是现代遗传学教材中阐释遗传密码起源的经典切入点。

在当下的生物学语境下对菱形密码进行分析,可以发现其具有以下特征:(1)模板是由双链构成的,是由DNA的两条链在螺旋的过程中形成的“腔”所决定的,因此,菱形密码的表达是立体的;(2)由3个碱基所决定;(3)表达的过程中无连接物;(4)具有双向可阅读性;(5)中间碱基互补时意义相同;(6)具有简并性。

除此之外,菱形密码还具有一个重要的特征——可重叠性。具体而言,序列中的单个碱基并非专属某个密码子,而是可以被前后相邻的密码子所共享。以碱基序列GATTACA为例,它可被解析为GAT、ATT、TTA、TAC和ACA这五个相互重叠的三联体组合。现在回过头来看,这一特征对遗传密码的发现产生了很大的误导。但是在当时,针对这一特征却具有很强的论据。首先,从立体化学方面的论据来看,当时发现蛋白质氨基酸之间的距离与DNA碱基之间的距离非常类似,这样一来,重叠的密码子就可以使得蛋白质氨基酸与DNA碱基的亚基在一对一匹配时,它们的聚合物能够更好地吻合。其次,从氨基酸与碱基数量的论据来看,当时发现氨基酸与碱基数量的总比接近1:1,这样一来,重叠的密码子就尽可能地避免了由3个碱基对应1个氨基酸的数量问题。同时,重叠的密码子还可以最大限度地提高信息储存的密度。当时的人们可能无论如何也不会想到通过自然进化的基因,中间竟然存在着许多“垃圾基因”。最终也正是重叠这一特征,使得菱形密码假说很快就被结束了。因为,其根本矛盾在于信息容量的不匹配。若采用1:1的编码规则(即一个碱基对应一个氨基酸),那么一个长度为N的

核苷酸序列,最多只能编码 4^N 种不同组合。然而,对于同样长度为N的氨基酸序列,其可能的组合数却高达 20^N 种。由于 4^N 远小于 20^N ,这就从数学上决定了必然存在大量氨基酸序列无法被任何核苷酸序列所编码。以最简单的二肽(由两个氨基酸组成)为例,氨基酸组合共有 $20^2=400$ 个二肽的可能,但两个重叠的密码子仅包含4个碱基,密码子组合只有 $4^4=256$ 种。由此可见,在重叠密码子的编码体系中,有 $400-256=144$ 种二肽是不可能出现的。即存在菱形密码不能产生的氨基酸复制方式。而这些在当时仅有的蛋白质数据中,也能够被实验所证明。之后,伽莫夫又提出了可重叠的“三角密码”等,但都以失败告终。

可以发现,重叠密码这种方案在当时由于“有很强的理论支撑”,导致重叠密码在遗传密码的理论发展过程中起到了很强的误导性。之后,克里克提出了不重叠的“无逗点密码”,被有些人称其为20世纪科学思想中最漂亮的错误想法^[9]。

2.2 无逗点密码

自20世纪50年代开始,科学界涌现出日益增多的证据与假说,共同指向并支持信使RNA,认为DNA所承载的遗传信息需经由一种单链的中介分子——信使RNA(mRNA)——才能传递至蛋白质合成系统。几乎在同一时期,克里克也在构想一个全新的理论,他称之为“连接物假说”。他推断,蛋白质的基本单位——氨基酸,并不具备直接“阅读”或“理解”mRNA分子上遗传指令(即密码子)的能力。因此,必然存在一种尚未被发现的、分子量较小的“桥梁”分子。这种分子的一端能够精确地识别并结合到mRNA的特定密码子上,另一端则负责携带对应的氨基酸。这个巧妙的“中间人”角色,后来被证实就是我们今天所熟知的转运RNA(tRNA)。

克里克将基因表达的过程想象为:首先,适当的DNA片段被转录成信使RNA,这个过程就像复制一样,并且它是一种盲目复制,并不考虑序列的含义;然后,信使RNA在细胞质中伸出其很长的密码子排列,就像母猪暴露的乳头一样,每一个连接物分子都与正确的氨基酸相结合,它们在周围不断地试触,直到锁定正确的密码子;之后,当所有的密码子被占用后,氨基酸被链接在一起;最后,完全合成的蛋白质与模板相剥离^[4]。

在当时,这一设想已展现出高度的合理性。即便以今日的视角审视,其核心逻辑仍与生命体中的典型分子识别与催化过程高度契合。该机制可类比于两

种经典的生物化学模型:酶—底物模型,如同酶分子通过其活性位点精准识别并结合特定底物,催化其转化;抗体—抗原模型,又或如抗体通过其可变区与特定抗原表位进行高特异性的“锁钥式”结合。同理,该假说认为,mRNA链上的密码子与相应的适配器分子(tRNA)之间,也通过这种高度特异的相互作用,被逐一识别并串联起来,从而实现遗传信息的解码。

然而,这种“小猪哺乳”式的模型存在一个致命缺陷,那就是“小猪”(即连接物分子)可能因识别错误而“卡在”两个“乳头”(即碱基)之间。例如,对于一段mRNA序列...UGUCGUAAG...,理想的读取方式是...UGU、CGU、AAG...。但RNA序列本身并没有标点符号来界定密码子的边界,因此也可能被误读为...UG、UCG、UAA、G...或...U、GUC、GUA、AG...等多种形式。这种不同的“读法”(即不同的读码框架)会产生完全不同的蛋白质序列。在这种模型下,这些在不同框架上结合的连接物分子会相互干扰,最终可能导致蛋白质合成的全面失败。这就是分子生物学中著名的“移码问题”。当然,在之前提到的重叠密码子模型中,由于所有可能的三联体读码框都被认为是有效的,因此反而不存在这个“移码问题”。

为解决遗传信息的正确读取问题,克里克于1957年提出了“无逗点密码”的理论模型。这个理论模型主要包含了三方面的特性。一是密码子的选择性。并非全部64种可能的三联体密码子都用于编码氨基酸。连接物分子可能是64个密码子的一个子集,而且,只有这个子集是有意义的,其余则为“无义密码子”。二是框架锁定机制。这个“有意义”的密码子集被设计成一个特殊系统,任意两个有意义的密码子拼接后,在其连接处因移码而产生的任何三联体,都必然是无义密码子。比如,假设CGU和AAG是有意义的密码子。当它们拼接成CGUAAG时,连接处可能产生的移码三联是GUA和UAA则是无义密码子。三是唯一解读结果。由于所有读码框架外的三联体都是无义的,这就从根本上杜绝了从错误起点开始读取的可能性,从而保证了遗传信息解读的唯一性。这个模型的精妙之处在于,它无需任何物理分隔符(即“逗号”),就能确保信息被正确读取。例如,无空格的Iamastudent,我们也能唯一地理解其含义。

那么这样的密码子是否存在呢?一个不包含分隔符的遗传密码,究竟能容纳多少个独立的密码子?针对这一问题,克里克与他在剑桥的两位同事——约翰·格里菲斯(John Griffith)和莱斯利·奥格尔

(Leslie Orgel)——以RNA为研究对象,进行了初步的理论推演。他们指出,首先密码子AAA,CCC,GGG和UUU这样由单一碱基重复构成的密码子不能出现在任何无逗点密码中,因为它们在拼接时无法提供足够的差异性来界定读码框架,它们和自身结合在一起都不能产生明确的阅读框。其次,剩余的60个密码子根据其循环排列关系被分成3个一组。例如,密码子AGU、GUA和UAG这三个密码子,彼此之间可以通过循环移位得到,因此它们同属一组。第三,为了维持“无逗点”的特性,在构建密码表时,最多只能从每个循环小组中选择一个密码子。那么,一共有多少种呢?将剩余的60个密码子分成三个一组,正好可以分为20组。由于每个小组最多只能贡献1个密码子,因此一个无逗点密码表的理论最大容量就是20个密码子^[4]。

虽然,克里克和他的同事们明确指出,他们没有关于无逗点密码的实验证据。但他们的理论计算结果20,为该理论提供了很强的吸引力。他们写道:“我们推断密码子所使用的论据和假设不能使我们有足够的信心认为它是一个纯粹的理论。……我们提出它是因为它以一种整齐的方式给出了一个神奇的数字——20,并且还有合理的物理假设”^[4]。然而在当时仅仅凭借这个神奇的数字就足以说服生物学家和广大的公众。后来,生物哲学家卡尔·伍兹(Carl Woese)写道:“无逗点密码立刻就被接纳,并且几乎被普遍接受……它们成为编码领域的重点,仅仅是因为它们巧妙的推理以及数字命理的帮助……在之后为期5年的时间中,这一领域大多数的思想都要么是来自它们的无逗点密码,要么是需要与它们相容性的基础来判断”^[10]。

无逗点密码并不是遗传密码构建过程中的最后一个假设。直到1961年,美国国立卫生研究院的马歇尔·尼伦伯格(Marshall W. Nirenberg)和海因里希·马特哈伊(J. Heinrich Matthaei)的实验带来了关键突破。他们发现人工合成的RNA(如poly-U)能在无细胞系统中诱导蛋白质合成,并首次将UUU密码子与苯丙氨酸对应起来^[11]。这一发现直接挑战了“无逗点密码”中UUU为无义密码子的假设。随后的两年里,更多密码子被陆续破解。最终被证实的密码,没有与包括“无逗点密码”在内的之前的任何一个理论概念相类似。

3 结语

在遗传密码破译的过程中,一个尤其引人注目

的问题是,DNA与蛋白质之间的复杂关系,在极短时间内就被还原成一个符号操作的抽象问题上。短短数月,分子层面的所有繁琐细节似乎都被暂时搁置一旁,整个生物学难题被重新定义为两种不同符号系统(即DNA的碱基与蛋白质的氨基酸)之间的数学映射关系。这便是早期“密码”概念的抽象性特征。

这一抽象性特征为后续基于排列组合的密码子规则构建奠定了基础。伽莫夫的“菱形密码”和克里克的“无逗点密码”,均基于此框架。菱形密码假设密码子具有对称性(如GAC与CAG等同),且为了提高信息密度,密码子是重叠的。无逗点密码中的无逗点性决定了AAA、CCC、GGG、UUU为无义密码子,并将其余密码子按循环规则分组。同时,它引入了RNA作为中介,并假设密码子不重叠。最终菱形密码与无逗点密码分别形成了各自的密码表。

这些理论中的“密码”隐喻,为理解蛋白质合成提供了有效的描述工具,但其局限性也很明显。隐喻语言的使用在当代科学中是非常常见的。尤其是当理解对象超出了人类经验范围时,隐喻语言的使用几乎成为科学工作不可避免的选择。它们可以使用简单的事物去描述复杂的概念,可以有助于交流和学习,但同样也容易使人误入歧途。如果不使用隐喻去理解自然,就会错失一些真正需要研究的事物;而如果过分强化最初的隐喻图景,又会忽略那些不适于这个隐喻系统的方面。尤其是当一些隐喻仅仅是为不存在的科技概念充当替代品时,它们的误导性更会超于它们的作用。

参考文献:

[1] 恩斯特·迈尔.生物学思想发展的历史[M].涂长晟,

译.成都:四川教育出版社,2010:14.

[2] 埃尔温·薛定谔.生命是什么[M].梁震宇,译.成都:四川文艺出版社,2022,10.

[3] GAMOW G. Letter to Linus Pauling[Z/OL]. 22 October 1953. ID: sci9.001.43-gamowlp-19531022. Oregon State University Special Collections[2025-08-15]. <http://osulibrary.oregonstate.edu/specialcollections/coll/pauling/dna/corr/sci9.001.43-gamow-lp-19531022.html>.

[4] CRICK F H, GRIFFITH J S, ORGEL L E. Codes without commas[J]. *Proceedings of the National Academy of Sciences of the U S A*, 1957, 43: 416-421.

[5] KAY L E. Who Wrote the Book of Life? A History of the Genetic Code[M]. Stanford: Stanford University Press, 2000, 2: 329.

[6] WEBER M. Philosophy of Experimental Biology[M]. Cambridge: Cambridge University Press, 2005.

[7] SUSTAR P. Crick's Notion of Genetic Information and the "Central Dogma" of Molecular Biology[J]. *The British Journal of Philosophy of Science*, 2007, 58: 13-24.

[8] CRICK F H. The genetic code: yesterday, today and tomorrow[C]//The Genetic Code, Proceedings of the XXXI Cold Spring Harbor Symposium on Quantitative Biology. Cold Spring Harbor, N Y: Cold Spring Harbor Laboratory of Quantitative Biology, 1966: 3-9.

[9] CRICK F H. What Mad Pursuit: A Personal View of Scientific Discovery[M]. New York: Basic Books, 1988: 23.

[10] WOESE C R. The Genetic Code: The Molecular Basis for Genetic Expression[M]. New York: Harper and Row, 1967.

[11] NIRENBERG M W, MATTHAEI J H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1961, 47(10):1588-1602.



杨维恒,男,副教授,博士,硕士生导师。主要研究方向为科学哲学、生物学哲学,先后主持和参与多项国家级课题研究,于核心期刊发表多篇学术论文。发表论文和出版著作分别曾获山西省第九次社会科学研究优秀成果奖、2017年度“百部(篇)工程”二等奖。

(责任编辑:赵云波)