

遗传密码是怎样破译的^{*}

向义和

(清华大学物理系, 北京 100084)

(收稿日期: 2006-11-10)

摘要 文章具体介绍了遗传密码破译的过程: 三联体密码的提出; 克里克对遗传密码性质的阐述和实验证明; 尼伦伯格关于同聚核苷酸对应的氨基酸的发现; 奥乔亚关于异聚核苷酸对应的氨基酸的发现; 科拉纳关于密码子系列次序的确定; 全部密码子的破译及其特征。

关键词 三联体密码; 遗传密码简并; 基因突变; 无细胞蛋白质的合成; 多聚核苷酸

HOW THE GENETIC CODE WERE DECIPHERED

Xiang Yihe

(Physics Department of Tsinghua University, Beijing 100084)

Abstract The process of decoding the genetic codes is introduced. The key events include the putting forward of triplet code, the expatiation and experimental proof of the nature of genetic code by F. H. C. Crick, the discovery of amino acid corresponding to homopolynucleotide by Marshall W. Nirenberg, the discovery of amino acid corresponding to hetero-polynucleotide by S. Ochoa, the determination of the sequence of codon series by H. G. Khorana, and the decipherment and characteristics of all codons.

Key Words triplet code; genetic code degeneration; gene mutation; synthesis of cell-free protein; polynucleotide

1953 年 5 月英国物理学家克里克 (Crick, France, 1916—) 和美国生物学家沃森 (Watson, James. D, 1926—) 在他们发表的《DNA 结构的遗传学意义》一文中提出: “碱基的排列顺序就是携带遗传信息的密码”的论断。这就是说, 在碱基(或核苷酸)的顺序中含有特定的遗传信息, 这种遗传信息能翻译成蛋白质上确定的氨基酸顺序。

要解答遗传密码和其中所含的信息问题, 需要从两方面思考: (1) 克里克和沃森倡导的密码问题其解决方式如何? 密码是怎样破译的? (2) 细胞使用怎样的机制从遗传信息合成蛋白质? 本文要讨论的是第 1 个问题。

从 1953 年提出遗传密码问题到 1966 年破译全部遗传密码, 其间经历了从抽象的理论分析到

具体的实验研究的过程。20 世纪 50 年代, 物理学家做了大量的理论分析工作, 伽莫夫 (Gamow, George) 首先提出了遗传密码的概念, 并提出了氨基酸是有 3 个一组的碱基编码的天才猜测, 其后克里克通过对生物化学家所做的基因实验结果的分析, 确定了遗传密码的一般性质, 为遗传密码的破译奠定了理论基础。20 世纪 60 年代, 生物学家做了大量的实验工作。在生化实验研究中, 经历了从简单到复杂的过程。从以均聚物为模板指导蛋白质的合成, 再以随机共聚物为模板指导蛋白质的合成, 到以特定共聚物为模板指导蛋白质的合成, 最后运用多种实验手段和多种实验方法, 确定了全部密码所对应的氨基酸, 揭示了生命的秘密。通过对这一探索过程的分析与讨论, 我们从中可以

* 本文研究课题得到清华大学工科物理基地教改项目资助。

得到丰富的教益和启迪。

三联体密码的提出

1953 年克里克和沃森的文章发表后, 首先引起了物理学家伽莫夫的注意. 这时他正在访问加利福尼亚大学伯克利分校, 正如他后来回忆的: “我走在放射实验室的走廊上, 碰上了阿尔瓦雷斯 (Alvarez, Luis), 他手上拿着一份《自然》杂志……他说: ‘看, 沃森和克里克写了一篇多棒的文章.’ 这是我第一次见到这篇文章. 后来我便回到了华盛顿, 开始思考这篇文章.”^[1]

1954 年 2 月伽莫夫把他对于密码问题思考的最初成果, 以题为“DNA 和蛋白质结构之间可能的联系”的文章发表在《自然》杂志上. 他在文章中首先表述了关于遗传密码的概念, 他写道: “任何生物体的遗传性状都可以突出地表现为用 4 个字母写成的核酸语言. 另一方面, 这些蛋白质是由大约 20 种不同的氨基酸组成的长肽链, 可以认为是用 20 个字母写成的蛋白质语言. 这样便出现了 4 个字母写成的核酸语言如何翻译成 20 个字母写成的蛋白质语言的问题.”^[2]

他在这篇文章中提出了一种设想, 认为蛋白质分子直接建立在 DNA 分子上, 而 DNA 分子则充当了使氨基酸按正确顺序排列的模板. 这一想法基于这样一种推测, 即碱基在双螺旋上的排列会产生一系列 20 种形状略微不同的菱形孔洞 (rhombshaped holes), 它与构成蛋白质的 20 种形状不同的氨基酸的关系, 犹如锁与钥匙的关系. 于是他大胆地假设, 每种氨基酸刚好对应一种孔洞. 因为相邻的孔洞间有两个相同的核苷酸, 所以每个孔洞由形成它的 4 个核苷酸中的 3 个确定. 当周围媒质中自由的氨基酸遇到了形状相同的空洞时, 就会插入这些洞内, 互相连接形成相应的肽链, 然后从洞中脱出.

不久以后, 克里克和沃森接到伽莫夫的信, 他们对信的内容感到很新奇, 从而促使他们去调查研究蛋白质中到底有多少种氨基酸. 他们发现伽莫夫列出的 20 种氨基酸中有几种不大可能, 有几种明显存在的, 如天冬酰胺和谷氨酰胺又未列入. 他们将蛋白质中的氨基酸分类, 将在许多不同的蛋白质中都存在的氨基酸归入标准类, 将在少数几种不常见的蛋白质中存在的氨基酸归入异常

类. 结果他们惊讶地发现标准的氨基酸正是 20 种^[3].

在此前提下, 他们思考了遗传密码的问题, 即用 4 个字母表示的核苷酸语言和用 20 个字母表示的氨基酸语言之间的关系问题. 他们用排列组合的方法分析, 在 4 种碱基中, 如果一次只取 1 个碱基, 你只能有 4 种不同的组合; 如果一次取 2 个一一对的碱基, 你可以有 16 种不同的碱基对 (4×4), 因为无论 4 种碱基中你取哪个作为碱基对中的头一个, 第二个碱基都有 4 种选择. 然而 16 种组合对于所有生物体中 20 种不同的氨基酸基本成分的编码来说仍是不够的. 如果一次取 3 个一组的碱基, 你可以得到 64 种不同的组合 ($4 \times 4 \times 4$). 这样对全部 20 种氨基酸进行编码就足够了^[3].

克里克和沃森肯定了伽莫夫的贡献, 他们认为伽莫夫关于每种氨基酸由 3 个一组的碱基 (即三联体) 编码是正确的. 而且从伽莫夫的研究中使他们认识到, 可将密码问题与实际的生化细节分开而作为一个抽象的问题来研究. 他们预期当人们能够得到各种不同的氨基酸序列时, 人们或许可以通过研究它们所受到的限制以及突变对某一特定序列的影响, 不需要搞清楚所有相关的生化步骤就可揭示出遗传密码, 这样的方法对于面临复杂的化学和生物化学的物理学家就更为自然^[3].

克里克和沃森也指出了伽莫夫在理论上的错误, 不同意他关于重叠密码的见解. 克里克在他的《狂热的追求——科学发现之我见》一书中写道: “从几个方面来看, 伽莫夫的密码是不同于一般的, 代表下一个氨基酸的三联体与前一个是重叠的. 例如, 如果一段碱基是 …GGAC…, 那么 GGA 代表一种氨基酸, GAC 代表下一个氨基酸. 这自然使氨基酸的顺序受到了某些限制, 有些序列是无法由伽莫夫的密码编码的.” 除此之外, 他们还不同意伽莫夫关于 DNA 直接编码蛋白质的想法, 并且非常怀疑伽莫夫关于 DNA 上的孔洞会有如此功能, 能使不同的氨基酸插入不同的洞内^[3].

1955 年 12 月伽莫夫和于卡斯 (Ycas, M) 在《美国国家科学院会报》上又发表了一篇关于遗传密码的文章, 题为“蛋白质和 RNA 组分的统计关系”. 提出了一种不重叠的三联码, 不重叠可使蛋白质中氨基酸次序上之限制予以清除; 最重要的

改变是将密码视作 RNA 模板之功能,而非 DNA 之功能.他在该文中说:“经验证据越来越清楚地表明,在蛋白质的合成中 RNA 起着一种活体(vital)的作用,假设表示给定的蛋白质特征的氨基酸序列是由 RNA 分子中核苷酸序列惟一确定的似乎是合理的。”^[4]

遗传密码的性质

1961 年 12 月 30 日克里克等人在《自然》杂志上发表了一篇题为“蛋白质遗传密码的一般性质”的论文.他们在论文一开始就说:“现在大量的证据说明沿着蛋白质的多肽链上的氨基酸序列是由沿着遗传物质的核酸的某些特定部分的碱基序列决定的.因为在整个自然界中发现全部蛋白质是由 20 种氨基酸组成的,但是只有 4 种共通的碱基,人们通常推测 4 种碱基的序列以某种方式为氨基酸序列编码。”^[5]在这篇文章中他们介绍了他们做的基因实验连同其他人的工作,说明遗传密码具有下列的一般性质:

(1) 3 个碱基一组编码一个氨基酸.碱基序列上这 3 个相邻的碱基称为三联体(triplet),又叫密码子(codon).

(2) 遗传密码互不重叠.

(3) 碱基序列是从一个固定的起始点读取的.这就确定了这条长的碱基序列如何正确地用三联体读出.没有特别的逗号来显示如何正确地选择三联体.如果开始点被一个碱基置换了,那么读入的三联体就被置换了,读出的密码信息就成为不正确的了.

(4) 密码可能是“简并”(degenerate)的,因为有 64 组三联体,而需要编码的氨基酸只有 20 种,所以有些氨基酸可能对应多个三联体^[5].

克里克在论文中首先依据基因的碱基置换(base substitution)突变的实验证明了遗传密码互不重叠的性质.碱基置换指的是通过化学反应在基因内部一个碱基直接变成另一个碱基.常见的一类诱变剂是能直接一个或多个碱基发生作用而改变它们配对性质的化合物.例如,亚硝酸的主要作用是引起胞嘧啶 C (cytosine) 氧化脱氨,使胞嘧啶转换成尿嘧啶 U (uracil).在下次复制周期中, U 与 A (adenine, 腺嘌呤) 配对代替了原来的 G (guanine, 鸟嘌呤) 与 C 的配对,这样再下一

次复制周期中,当 A 与 T (thymine, 胸腺嘧啶) 配对时, T—A 对就代替了 C—G 对.

这个证据不是来自于克里克本人的工作,而是来自于其他生物化学家所做的由亚硝酸产生的烟草花叶病毒突变的实验.克里克认为在一个相互重叠的三联体中,一般来说,一个碱基的改变,在多肽链上将引起 3 个相邻的氨基酸的改变,如图 1 所示.图中显示出重叠密码和非重叠密码的区别,图中的短竖直线表示核酸的碱基.实验表明,用亚硝酸处理过的烟草花叶病毒的 RNA,在该病毒的蛋白质内产生的变更,通常一次只改变一个氨基酸.在很少的情况下改变了两个氨基酸,但是它们也并不处在多肽链上相邻的位置上,这是由于在一条 RNA 链上大概有亚硝酸的两个独自的脱氨基作用,分别置换了两个碱基,产生了双突变的结果^[5].

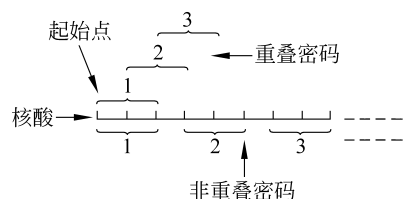


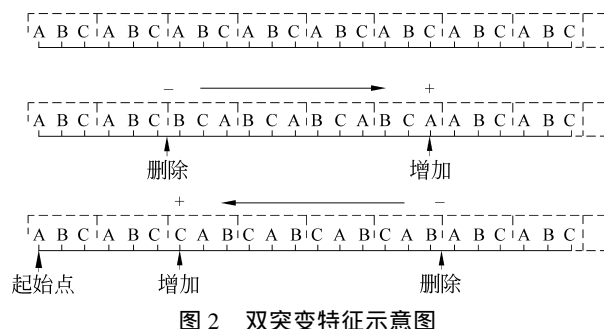
图 1 重叠密码与非重叠密码的区别

克里克指出,如果遗传密码是普遍存在的,也就是说,各类生物都使用相同的遗传语言,那么所有关于相互重叠密码的论断就不能成立.而且,他还指出,从已经详细研究过的非正常人的血红蛋白表明,这是由于在正常血红蛋白基因上产生了碱基置换突变,碱基 U 取代了碱基 G,使得三联体 GUA 代替了三联体 GGA,从而导致了单个氨基酸的改变,在正常血红蛋白多肽链上的甘氨酸被一个缬氨酸取代.这个新的实验结果充分证实了遗传密码互不重叠的性质.

克里克关于三联体密码的证据是通过基因的移码突变(frameshift mutation)实验分析发现的.移码突变指的是通过丫啶类(acridines)化合物的作用,在基因内部插入或删除一个碱基对.丫啶类化合物与 DNA 结合使双螺旋结构变形,其结果使 DNA 在复制过程中可能掺入一个碱基或删除一个碱基.如果添加一个碱基用 + 号表示,删除一个碱基用 - 号表示,则上述丫啶类突变可归类为 (+)型和 (-)型.如果丫啶类突变增加了一个核苷酸,又丢失了这个核苷酸,则这类双突变归类为

(十一)型或(一十)型,经过这种双突变的物种必定会回复到正常型,也称为野生型(wild type),即自然界中野生状态的生物类型.但是如果在添加核苷酸的位点的附近缺失一个不同的碱基,也能使这两个突变之外的区域恢复到野生型.这里的第二次突变称为抑制基因(suppressor)突变,这类双突变的表型称为假野生型(pseudo-wild).

克里克认为遗传密码必须从固定起始点开始按三联体方式阅读,如图2所示,由左端开始向右阅读.图中字母A、B、C表示不同的碱基.为简便起见,用重复的ABC表示一个由相同三联体组成的碱基序列,这个序列将编码一条由相同氨基酸组成的多肽链.图中的点线表示设想的阅读框架.符号+表示添加一个碱基,符号-表示除去一个碱基.图中的箭头总是由一指指向+,表示阅读框架移动的方向和移动一个碱基的距离.



克里克假定这个基因突变是在一个野生型的碱基序列中插入一个附加的碱基引起的.这意味着该处右边的所有三联体的阅读都将向右移动一个碱基,于是由这个基因产生的蛋白质的氨基酸序列,从那点以后将完全改变,从而失去了该基因的功能.接着克里克又假定该基因的抑制基因是由删去一个碱基形成的.当后一个突变单独存在时,所有它右边的三联体仍将不能正确地读出,从而失去这个基因的功能.然而当两个(十一)型或(一十)型双突变同时在一个基因或两个相邻的基因上存在时,在两个突变部位之外将恢复到原来正确的阅读,而在两个突变部位之间仍将造成不正确的阅读.这就说明经过这种双突变后的物种不是真正的野生表型,而通常是一个假野生型,因为在它的氨基酸序列中有一小段长度是不同于野生型的^[5].

接着克里克预计,如果在基因内部两次突变都是添加一个碱基的(++)型双重突变,或者两

次突变都是减少一个碱基的(--)型双重突变,则将给出突变体的特征.他分析研究了这种情形下的14对双突变,都证明了这个预计的正确性.

上述的双突变还不能揭示出多少个核苷酸组成一个密码子.然而,当克里克构建三重突变体时,发现只有在基因内部三次突变都是添加一个碱基的(+++)型,以及都是减少一个碱基的(---)型,才会出现野生型或假野生型的特征,而其他的结合方式仍是突变体.如果我们将添加3个或缺少3个碱基分别对应于在它的氨基酸序列上增加或删除一个氨基酸,这将会不会太大地干扰这个基因产物的功能,这就意味着密码子是以三联体方式阅读的.三重突变体也与补偿的双重突变体一样,在两个外侧突变部分之间有不正确的氨基酸序列,但是在两个外侧突变部分之外的氨基酸序列仍是野生型^[5].

同聚核苷酸对应的氨基酸的发现

早在1955年在纽约大学医学院工作的法国生物学家马纳戈(Manago M. Grunberg)和西班牙裔的美国生物学家奥乔亚(Ochoa, S)发现在放有多核苷酸磷酸化酶(polynucleotide phosphorylase)的试管里置入作为生物能量载体的腺苷三磷酸(adenosine triphosphate缩写为ATP)后,该酶摄取了ATP,抛弃了核苷酸末端的两个额外的磷酸基,而把一个个核苷酸连在一起形成一条长核苷酸链.与细胞中通常发现的RNA相比,它是非常奇怪的核酸.与通常一样,其主链是一串糖和磷酸根连在一起,但每个糖都带着一个同样的碱基,呈一长串没完没了的信息,例如AAAAAA……这个多核苷酸名叫多聚腺苷酸(polyadenylic acid缩写为poly A).不久又发现,这种酶也同样会利用等价的含有鸟嘌呤的鸟苷三磷酸(GTP),含有尿嘧啶的尿苷三磷酸(UTP)和含有胞苷的胞苷三磷酸(CTP)的分子,分别制造出多聚鸟苷酸(poly G)、多聚尿苷酸(poly U)和多聚胞苷酸(poly C),他们甚至还利用包含腺嘌呤和尿嘧啶的磷酸混合物制造出poly AU,它在链上以不确定的顺序排列着不同的碱基^[1].

20世纪50年代末在华盛顿卫生研究院(NIH)工作的美国生物化学家尼伦伯格(Nirenberg, M)和德国化学家马特伊(Matthaei, J)对在试管里创造蛋白

质合成的适合条件这个问题产生了兴趣,在受控条件下可以测到细胞各部分的活动,识别它们在细胞活动过程中的作用.这样的一个系统由于不含整个细胞(当然它要包含某些细胞器)而被称为“无细胞”系统.1961年5月他们在实验中发现,在无细胞蛋白质合成系统中加入多聚尿苷酸(poly U)后,合成出了多聚苯丙氨酸(poly-Phe).显然 poly U 被用来作为遗传信息的携带者了.按照密码的三联体性质,应该是 UUU 为苯丙氨酸(Phe)编码.这是一个带有里程碑性的实验,它指出了完全阐明遗传密码的途径^[1].

1961年8月3日他们以题为“大肠杆菌的无细胞蛋白质合成体系取决于自然的或合成的多核苷酸”的论文投寄《美国国家科学院会报》.指出这个实验的两个必要的组成部分是无细胞蛋白质合成体系和人工合成的多核苷酸.无细胞蛋白质合成体系是从大肠杆菌中得到的,其中含有DNA、RNA、核糖体、酶和其他细胞成分,当加入ATP、GTP和氨基酸时,此无细胞体系就可合成出蛋白质来.在加入的氨基酸中至少有一种是具有放射性的,从中可测出其掺入蛋白质的情况.将此混合体系放在35°C保温培育1小时,然后加入DNA酶和三氯醋酸终止反应并沉淀蛋白质,游历的氨基酸则在上清液中.沉淀物经洗涤后测定其放射性,以确定掺入新合成的蛋白质中的有放射性标记的氨基酸的含量.

在该文的题为“由合成的多核苷酸激励氨基酸的合成”一节中,他们写道:“如图3所示,在反应混合物中每ml增加多尿苷酸(poly U)10⁴g就会明显地激励C¹⁴-L-苯丙氨酸的结合,苯丙氨酸的结合几乎完全取决于多尿苷酸的增加.在稍微滞后一点时间后,合成以线性比例继续进行约30分钟.”接着,他们又在无细胞蛋白质体系中加入不同的多核苷酸,分别测定用C¹⁴标记的苯丙氨酸掺入蛋白质的量,实验表明poly U与poly A和poly C比较起来其作用最为显著^[1].

后来他们又在加入poly U的无细胞蛋白质合成体系中加入20种氨基酸的混合物,每种混合物中分别含有一种用C¹⁴标记的氨基酸.实验结果表明只有C¹⁴标记的苯丙氨酸掺入到蛋白质合成物中,产物为多聚苯丙氨酸.说明poly U只激励了苯丙氨酸的合成,对激励其他放射性氨基酸的合成影响很小.当时还未能从实验上确定密码的

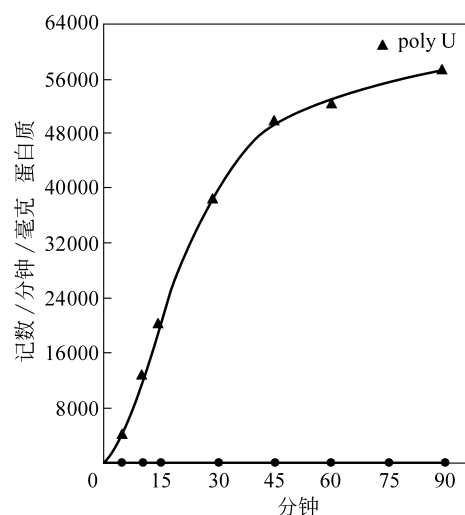


图3 由多尿苷酸激励的C¹⁴-L-苯丙氨酸的合成

三联体性质,正如他们在该文的“结束语”中说的话:“在最近的实验中需要的多聚物的尿苷酸单元和结合的L-苯丙氨酸之间的比率已经接近1:1的比值,但是对于形成苯丙氨酸密码的尿苷酸残基的数目的直接的证据还未确定.”^[1]

1961年夏天在莫斯科召开了第五届国际生物化学大会.在会议日程上尼伦伯格被安排在一个较小的房间里宣读10分钟的论文,只有极少的几个人意识到这项研究工作的重要性.其中一位自言“被这篇论文惊倒”的生化学家梅塞尔森马上去找身为会议主席的克里克,把这个消息告诉了他,克里克重新安排尼伦伯格在主报告厅面对全体与会听众再次宣读他的论文.梅塞尔森回忆道:“讲完后,我跑到尼伦伯格面前拥抱了他,向他祝贺……一切都是那么富有戏剧性.”^[1]1961年12月克里克在给《自然》杂志的论文中写道:“听到尼伦伯格宣告的结果使我大吃一惊,他和马特伊通过把多聚尿苷酸加入到无细胞蛋白质合成体系中产生了多聚苯丙氨酸.这意味着尿嘧啶序列在为苯丙氨酸编码,我们的工作说明这个密码大概是尿嘧啶的三联体.”^[5]

后来用同样的方法发现多聚腺苷酸{poly A}和多聚胞苷酸(poly C)分别导致多聚赖氨酸和多聚脯氨酸的合成.

异聚核苷酸对应的氨基酸的发现

尼伦伯格研究小组在发现多聚尿苷酸(poly U)控制了苯丙氨酸的合成,从而证实了多聚尿苷

酸在为苯丙氨酸编码后,又研究了一系列随机混合的多核苷酸控制的氨基酸的合成.1961年12月4日他们以题为“遗传密码的核苷酸组成”的论文寄给《生物化学、生物物理研究通讯》.他们在文章中介绍了当把人工合成的多聚核苷酸 poly UA、poly UC、poly UG、poly UAC、poly UGC、poly UGA 制备出来后,分别放入无细胞蛋白质合成体系中,测定 C^{14} 标记的各种氨基酸的掺入,列出了对于各种多聚核苷酸每种氨基酸相对于苯丙氨酸掺入蛋白质的百分比.他们从这些实验数据中发现了遗传密码的三联体性质和简并性质.他们发现每种氨基酸的最小密码单位包含了3个碱基,这是由下列数据获得的.有两种氨基酸——精氨酸(Arg)和丙氨酸(Ala)的合成都是由 poly UGC 激励的,它们的编码单位每个都包含3个不同的核苷酸.这个数据把单密码和双密码的可能性都排除在外.假定所有氨基酸有相同的密码率(coding ratio 即核苷酸数目与编码的氨基酸数的比率),这个最小的密码率必定是3,有可能它是更大的.此外,这些实验数据还说明了对应于亮氨酸(Leu)的密码单位,既包含了 UC 又包含了 UG,它们都能激励亮氨酸,使其掺入新合成的蛋白质中.由于两个不同的密码单位都可以确定同一个氨基酸,显示遗传密码至少有一部分含有数个密码语言以导引单一的氨基酸,这种现象称为遗传密码简并^[7].

1961年10月25日在美国纽约大学医学院生物化学系任教的生化学家奥柯亚给《美国国家科学院会报》寄去题为“合成的多核苷酸和氨基酸密码”的文章.该文对尼伦伯格研究小组之成果予以肯定,并且进一步提出有关其他氨基酸的密码证据,从实验上验证了各种三联体密码的相对出现率.例如由 poly UC ($U:C=5:1$) 激励的氨基酸的合成,在一个由 U 和 C 随机排列组成的共聚体中,就可能有8种不同的三联体:UUU、UUC、UCU、CUU、UCC、CUC、CCU 和 CCC.因为各个三联体出现的几率等于三联体中各个碱基出现的几率的乘积,显然 UUU 对 UUC(或 UCU 或 CUU)三联体的比率将与聚合物的 $U:C$ 的比率是相同的,也就是 $5:1$.这个 UUU 对 UCC(或 CUC 或 CCU)三联体的比率是 $25:1$.从表1中实测数据他得出由 poly UC 合成的苯丙氨酸对丝氨酸的比率是 $4.4:1$,由此他推断出丝氨酸的三

联体密码可能是 UUC、UCU 或 CUU.对于由 poly UA ($U:A=5:1$) 合成的苯丙氨酸对酪氨酸的比率是 $4:1$,从而得出 UUA、UAU 或 AUU 很可能是酪氨酸的三联体密码的结论.这三种密码子虽然可能具有相同的理论频率,它们之间仍有区别存在.在这两组三联体中,是哪个编码了丝氨酸,哪个编码了酪氨酸还不能确定,这时还只知道氨基酸密码子包含的碱基种类,尚不能确定碱基的次序^[8].

1962年奥柯亚等人又完成题为“合成的多核苷酸和氨基酸密码v”的论文,发表于该年《科学》杂志48卷上.在这篇文章中遗传密码的证据是从遗传突变获得的.由于它是从完整细胞所得的结果,故特别有价值.现用人工诱变的烟草花叶病毒(TMV)突变体所得的结果予以说明.用亚硝酸处理烟草花叶病毒的RNA会导致单个碱基氧化脱氧,这种脱氧作用会发生两种变化,使尿嘧啶取代了胞嘧啶,即导致 $C \rightarrow U$ 的置换;使鸟嘌呤取代了腺嘌呤,即导致 $A \rightarrow G$ 的置换.由于核苷酸链上的碱基置换,改变了它们的核苷酸序列,从而造成了在这个蛋白质多肽中在某个特定位置上一个氨基酸被另一个氨基酸置换^[9].

接着奥柯亚介绍了用亚硝酸处理过的多聚物的实验,他把具有 poly UA、poly UG 和 poly UC 的实验结果列于表1中,实验结果表明用亚硝酸处理后的多聚物在活性上普遍的明显的减小.对于 poly UA ($5:1$),由于脱氨基的作用,A转换为G, poly UA 编码特征的改变是容易明白的.在脱氨基之前 poly UA 激励了包含 2U1A 成分的异亮氨酸、亮氨酸和酪氨酸的合成,但是没有激励半胱氨酸和缬氨酸的合成.脱氨基的作用大大地减少了激励异亮氨酸和酪氨酸合成的能力,促进了激励半胱氨酸和缬氨酸合成的能力.这一实验结果可以从脱氨基作用 $A \rightarrow G$ 的置换加以说明,亮氨酸极大地保持了它用 2U1A 和 2U1G 编码的一致性,而半胱氨酸和缬氨酸可能是由 2U1G 编码的^[9].

密码子排列次序的确定

1963年春在美国威斯康星大学酶研究所工作的巴基斯坦裔美籍生物化学家科拉纳(Khorana, H. G 1922—)开始研究具有重复序列

(repeating sequence)的多核苷酸在无细胞系统中合成多肽.对于具有重复二核苷酸序列的多聚体,例如 $(AG)_n$ 包含了两种三联体AGA和GAG.假设3个字母没有密码重叠的性质,则这种多聚体应该导致精氨酸(Arg)和谷氨酸(Glu)精确地在交替的序列 $(Arg-Glu)_n$ 中合成共聚肽(copoly peptide)^[10].

对于重复的三核苷酸序列的多聚体,例如 $(AAG)_n$ 包含了取决于起始点的3个重复的三联体,有AAG、AGA和GAA.人们可以预计一次应该合成其中一种氨基酸形成的同型多肽(homopoly peptide)链,最多可能产生3条这样的链.其中有由单一赖氨酸(Lys)组成的链;单一精氨酸(Arg)组成的链和单一谷氨酸(Glu)组成的链^[10].

然而,对于密码子测定这个问题,除了人们做了大量的多肽合成外,上述的这些实验还不能单独地提供对密码子的惟一的测定.例如用具有重复二核苷酸序列的多聚体 $(AG)_n$ 合成的具有交替的序列的共聚肽 $(Arg-Glu)_n$,但是不能确定AGA和GAG哪个密码是精氨酸(Arg),哪个密码是谷氨酸(Glu).事实上现在这个密码已经通过使用尼伦伯格和李德(Lede, Philip)研究出的结合技术获得的结果结合上述的重复多聚体的工作而推导出来了.

1964年9月尼伦伯格和李德在《科学》杂志上发表题为“RNA 密码子和蛋白质的合成”的论文.他们发现在无蛋白质系统合成的情况下,特殊的三核苷酸能促进特殊的 C^{14} 氨酰 tRNA(即转移RNA)分子与核糖体结合.例如,加入三核苷酸UUU促进苯丙氨酸 tRNA 与核糖体结合;加入三核苷酸AAA促进赖氨酸 tRNA 与核糖体结合;加入三核苷酸CCC促进脯氨酸 tRNA 与核糖体结合.二核苷酸并不促进 tRNA 与核糖体结合.此研究结果表明三核苷酸能促进相应的 C^{14} 氨酰 tRNA 结合到核糖体上,而这个三核苷酸就是为相应的氨基酸编码的密码子.尼伦伯格设计了一种简便快速的结合分析法,利用核糖体与硝酸纤维素滤膜结合的能力,可以分离出三核苷酸、氨酰 tRNA 和核糖体的三元复合体.当结合在核糖体上的 tRNA 分子在用硝酸纤维素滤膜过滤时,能被截留下来,而未结合的 tRNA 能通过滤膜.通过测试截留在滤膜上的标记了 C^{14} 氨酰 tRNA,就可以确定这个三联体所表示的氨基酸,从而破解这

个密码子^[11].

后来科拉纳应用尼伦伯格的研究结果解决了密码子排列次序的问题.例如上述的由重复的三核苷酸序列 $(AAG)_n$ 合成的同型多肽:由单一赖氨酸(Lys)组成的链;单一精氨酸(Arg)组成的链和单一谷氨酸(Glu)组成的链.但是不能确定在3个密码子AAG、AGA、GAA中,哪个是赖氨酸?哪个是精氨酸?哪个是谷氨酸?以分析哪个是赖氨酸为例,他们在分别增加这些三核苷酸量的情况下,测量了 C^{14} 赖氨酸结合到核糖体上的量.发现三核苷酸AAG特别促进结合,从而确定三联体AAG是赖氨酸密码子.用同样的方法他们确定了AGA是精氨酸的密码子,GAA是谷氨酸的密码子.结合上述重复二核苷酸序列 $(AG)_n$ 合成的共聚肽的结果,很快就可以判断出GAA也是谷氨酸的密码子^[10].

全部密码子的破译及其特征

从1961年尼伦伯格用同聚核苷酸poly U合成出了苯丙氨酸开始,经历了用随机排列的异聚核苷酸合成多肽,直到用特定序列的多核苷酸合成多肽,最后通过尼伦伯格用三核苷酸促进 C^{14} 氨酰 tRNA 和核糖体结合的研究,终于在1966年破译了全部64个密码子(表1).其中61个是氨基酸的密码子,另外3个密码子是UAA、UGA和UAG,它们并不编码任何氨基酸,是终止密码子.这些密码子表示肽链的终止,是每个基因终止的标志.还有一个起始密码子AUG,它既是肽链的起始信号又是甲硫氨酸的密码子.

因为20种氨基酸有61种密码子,所以许多氨基酸有多个密码子,这称为密码子的简并性.实际上除色氨酸(Trp)和甲硫氨酸(Met)只有一个密码子外,其他氨基酸都有一个以上的密码子.其中有2个密码子的9种氨基酸是:苯丙氨酸(Phe)、酪氨酸(Tyr)、半胱氨酸(Cys)、组氨酸(His)、谷氨酰胺(Gln)、天冬氨酸(Asp)、天冬酰胺(Asn)、赖氨酸(Lys)、谷氨酸(Glu);有3个密码的1种氨基酸是异亮氨酸(Ile);有4个密码子的5种氨基酸是:脯氨酸(Pro)、苏氨酸(Thr)、缬氨酸(Val)、丙氨酸(Ala)、甘氨酸(Gly);有6个密码子的3种氨基酸是:亮氨酸(Leu)、丝氨酸(Ser)、精氨酸(Arg).

表 1 遗传密码子

第一位核苷酸	第二位(中间)核苷酸				第三位核苷酸
U	U	C	A	G	U C A G
	Phe	Ser	Tyr	Cys	
	Phe	Ser	Tyr	Cys	
	Leu	Ser	Stop	Stop	
C	Leu	Ser	Stop	Trp	U C A G
	Leu	Pro	His	Arg	
	Leu	Pro	His	Arg	
	Leu	Pro	Gln	Grg	
A	Leu	Pro	Gln	Arg	U C A G
	Ile	Thr	Asn	Ser	
	Ile	Thr	Asn	Ser	
	Ile	Thr	Lys	Arg	
G	Met *	Thr	Lys	Arg	U C A G
	Val	Ala	Asp	Gly	
	Val	Ala	Asp	Gly	
	Val	Ala	Glu	Gly	
G	Val	Ala	Glu	Gly	U C A G

sotp 为终止密码子, *常作为起始信号.

对应于同一氨基酸的密码子称为同义密码子(synonym codon). 同义密码子一般都不是随机分布的. 编码同一氨基酸的几种同义密码子列在同一格中(除非同义密码子多于 4 个). 同一格中氨基酸的密码子的第一、第二位核苷酸都是相同的, 只有第三位核苷酸不同. 例如, 缬氨酸(Val)的 4 个同义密码子 GUU、GUC、GUA 和 GUG, 前两个核苷酸都是 GU, 只有第三位可以是 U、C、A、G 中任一种核苷酸. 因此, 大多数同义密码子的不同仅在于三联体的最后一个核苷酸. 1968 年 12 月尼伦伯格在他的题为《遗传密码》的诺贝尔演讲中说:“密码子简并性的有规则的形式好像是因为碱基 U 通常等效于碱基 C, 而碱基 A 等效于碱基 G.”^[12] 检查一下密码子, 如果以 X 和 Y 代替任一碱基的话, 可发现 XYZ 和 XYU 总是编码同一氨基酸, 而 XYG 和 XYA 也总是编码同一氨基酸. 例如谷氨酰胺(Gln)的同义密码子就是 CAA 和 CAG, 而组氨酸(His)的同义密码子是 CAU 和 CAC.

遗传密码是否具有普遍性? 是否在所有生物中都相同? 尼伦伯格等人通过研究细菌、两栖类和哺乳类动物细胞中的氨酰 tRNA 对三核苷酸密码子的响应, 发现三核苷酸几乎是等同地翻译成氨基酸. 他指出按照古生物学家研究的结果, 30

亿年前, 细菌出现于地球上, 此即为机能性遗传密码之开始; 脊椎动物大约出现在 5 亿年前, 几乎可以肯定遗传密码的执行功能至少已达 5 亿年以上. 他在题为《遗传密码》的诺贝尔演讲中说:“虽然在密码子的翻译中出现了一些改变, 但是在用细菌、两栖类和哺乳类动物的氨酰 tRNA 识别密码子的碱基序列中明显的类似性显示出, 地球上大多数或许全部生物基本上使用相同的遗传语言, 并且按照普遍的规则翻译这种语言.”^[12]

参 考 文 献

[1] [英] 格里宾(Gribbin J). 双螺旋探秘——量子物理学与生命. 方玉珍等译. 上海: 上海科技教育出版社, 2001. 236, 256~259

[2] Gamow G. Possible Relation between Deoxyribonucleic Acid and Protein Structures. *Nature* 1954, 173: 318

[3] [英] 克里克(Crick F). 狂热的追求——科学发现之我见. 吕向东等译. 合肥: 中国科学技术大学出版社, 1994. 100~105

[4] Gamow G, Ycas M. Statistical Correlation of Protein and Ribonucleic Acid Composition. *Proceedings of the National Academy of Sciences of the United States of America*. 1955. 41: 1011

[5] Crick F, Barnett L, Brenner S, Tobin R. General Nature of the Genetic Code for Proteins. *Nature* 1961, 192: 1227~1232

[6] Nirenberg M, Matthaei J. The Dependence of Cell-Free Protein Synthesis in E. Coli upon Naturally Occurring or Synthetic Polynucleotides. *Proceedings of the National Academy of Sciences of the United States of America*. 1961. 47: 1588~1601

[7] Nirenberg M, Martin R, Matthaei J, Jones O. Ribonucleotide Composition of the Genetic Code. *Biochemical and Biophysical Research Communications*. 1962, 6: 410~414

[8] Ochoa S, Lengyel P, Speyer J. Synthetic Polynucleotides and the Amino Acid Code. *Proceedings of the National Academy of Sciences of the United States of America*. 1961. 47: 1936~1942

[9] Ochoa S, Basilio C, Wahea A, Lengyel P, Speyer J. Synthetic Polynucleotides and the Amino Acid Code. V. *Science* 1962, 48: 613~616

[10] Khorana H. Nucleic Acid Synthesis in the Study of the Genetic Code. *Nobel Lectures: Physiology or Medicine (1963~1970)*. 350~356

[11] Nirenberg M, Leder P. RNA Codewords and Protein Synthesis—the Effect of Trinucleotides upon the Binding of Srna to Ribosomes. *Science* 1964, 145: 1399~1407

[12] Nirenberg M. The Genetic Code. *Nobel Lectures: Physiology or Medicine(1963~1970)*. 375~392