

当前位置: [首页](#)>[期刊文章](#)

[【小中大】](#) [【打印】](#) [【关闭窗口】](#) [【PDF版查看】](#)

转载需注明出处

《科学文化评论》第1卷 第2期（2004）：

学术前沿



科学文化评论

扩充遗传密码

——进化的重演还是未来的预演？

王磊^[1]

摘要 文章介绍了一种直接在活体细胞内掺入非天然氨基酸到蛋白质中的新方法。遗传编码非天然氨基酸有效地扩充了现存生物体通用的遗传密码。这对遗传密码的进化、蛋白质的结构、功能与应用、以及与蛋白质相关的生命过程等方面的研究提供了新的思路与方法。

关键词 遗传密码 非天然氨基酸 转移RNA 氨酰tRNA合成酶 定向进化 蛋白质工程

地球上所有的生命，无论是低等的细菌和植物，还是高等的哺乳动物以及人类，都是由64个密码子来编码20个氨基酸和3个终止信号。由这20个氨基酸所组成的蛋白质担负着生命过程的大部分重要功能，诸如光合作用、视觉、信号传导和免疫等。虽然蛋白质具有非凡的生物、物理和化学特性，20个氨基酸所含的功能团毕竟是有限的。有时蛋白质需借助额外的化学作用来完成其天然的功能。这些作用包括对蛋白质进行磷酸化、甲基化或羟基化等转译后的修饰和使用辅助因子。很多重要的肽中含有不寻常的氨基酸，但蛋白质的加工厂——核糖体——却不能合成这些肽，它们的制造需要复杂的额外酶来完成[Walsh et al., 2001]。在罕见的情形下，有的生物会用终止密码子来掺入新的氨基酸——硒代半胱氨酸和吡咯赖氨酸[Bock et al., 1991; Srinivasan et al., 2002]。这两个氨基酸的掺入机制与普通的20个氨基酸不同，需要额外的核酸和蛋白质来辅助。既然存在这些需要与不足，为什么只有20个氨基酸编码在生物的遗传密码中？遗传密码能否被扩充来包含新的氨基酸？扩充后的遗传密码是否会给宿主带来进化上的优势？虽然不少假设和理论已被提出[Crick, 1968; Wong, 1975; Di Giulio et al., 1999; Knight et al., 2001]，迄今还没有有说服力的直接证据来回答上述问题。另一方面，人们理性地操作蛋白质的结构与功能的能力还处于非常有限的水平。这种局限性直接影响对生命现象的理解、研究以及开发利用。如果能扩充遗传密码，也即能直接在活体细胞中掺入新氨基酸到蛋白质中，精心设计的功能团就可以根据结构和功能的需要量体裁衣地引入蛋白质。这种技术将极大地提高我们对蛋白质和相关生命过程的研究和掌控能力。本文将简单介绍一种添加非天然氨基酸到活体细胞蛋白质中的新方法[Wang et al., 2001a]。这种方法首次扩充了生物体的遗传密码，正逐步被应用于蛋白质工程以及活体细胞的研究中。

一 方法的原理与建立

在蛋白质的生物合成过程中，每个氨基酸都有一个对应的氨酰tRNA合成酶和至少一个转移RNA (tRNA)。氨酰tRNA合成酶负责把正确的氨基酸加载到对应的tRNA上。tRNA带着氨基酸被其它酶运送到核糖体，在那儿tRNA的反密码子与信使RNA的密码子匹配，氨基酸从tRNA转接到肽链中。藉此DNA的遗传信息被转译成蛋白质中的氨基酸序列。

增添一个非天然氨基酸到遗传密码中需要创建一套新的组分，包括tRNA、密码子和氨酰tRNA合成酶（图1）[Wang et al., 2002b]。具体地说，这个新tRNA（正交tRNA）不能被宿主的内源氨酰tRNA合成酶识别，但必须有效地参与蛋白质的翻译。这个

正交tRNA必须解码一个不编码任何20个普通氨基酸的密码子（独特密码子）。另外还需要构建一个新的氨酰tRNA合成酶（正交氨酰tRNA合成酶）来识别正交tRNA，但不识别任何内源tRNA。正交氨酰tRNA合成酶只能加载所需的非天然氨基酸但不加载任何普通氨基酸到正交tRNA上。同理，非天然氨基酸也不能是任何内源氨酰tRNA合成酶的底物。最后，非天然氨基酸在加入培养基后必须能高效地进入细胞质，或是设法由细胞自己生物合成。

为建立此方法，大肠杆菌被首先选用为宿主，因为它易于遗传操作。在遗传密码中有三个终止密码子，它们不编码任何氨基酸而是终止蛋白质的翻译。因为原则上一个终止子即可满足翻译终止，另外两个便可改用于编码非天然氨基酸。琥珀终止子（TAG）是大肠杆菌中最少用的终止子，因而被选作独特的密码子来编码非天然氨基酸，以把对宿主的干扰降到最小。

下一步是如何构建一对正交的tRNA/合成酶来解码琥珀终止子。大肠杆菌中有21个氨酰tRNA合成酶和86个tRNA。它们之间的相互作用进化得如此之精细，每个氨基酸对应的tRNA/合成酶都对高度专一地识别自己的对应物，不同氨基酸之间的组分则互不干扰。因而从头开始设计一对正交于所有内源对应物的tRNA/合成酶几乎是不可能的。利用物种之间的差异巧妙地解决了这个难题。交叉测试不同生物界（例如原核与真核生物）的tRNA/合成酶，例如用酵母的酪氨酸合成酶来氨酰化大肠杆菌的酪氨酸tRNA，发现活性通常都很低。因而引入其它生物的tRNA/合成酶到大肠杆菌中或可构成这样的正交tRNA/合成酶对。古细菌*Methanococcus jannaschii* (*Mj*)的酪氨酸tRNA的反密码子被突变为CUA以解码琥珀终止子TAG。经测试表明这个突变后的琥珀抑制tRNA (*Mj*) 与其同源的酪氨酸tRNA合成酶 (*MjTyrRS*) 可以在大肠杆菌中高效表达并构成了一对正交对[Wang et al., 2000]。但*Mj* 还能在很低程度上被某些大肠杆菌内源合成酶识别。这种即使是非常微小的错误识别都有可能造成少量普通氨基酸被掺入到TAG指定的位置，从而造成不纯的蛋白质混合物。为进一步降低*Mj* 对大肠杆菌内源合成酶的活性，同时保持其对*MjTyrRS*的活性，它的11个核苷被饱和和突变构建出一个tRNA文库。通过精心设计的负向选择和正向选择，从此文库中找到一个突变体，此突变体不再被大肠杆菌的内源合成酶识别，而*MjTyrRS*仍高活性地识别它[Wang et al., 2001b]。因而*Mj/MjTyrRS*在大肠杆菌中构成了一对完美的正交tRNA/合成酶对。

*MjTyrRS*加载酪氨酸到*Mj*上。为使正交的合成酶选择地加载一个非天然氨基酸到正交的*Mj*上，*MjTyrRS*的底物特异性必须改变。合成酶对其对应的氨基酸和tRNA都具有高度的专一性以确保转译的忠实性，试图改变*MjTyrRS*的氨基酸特异性的同时不能削弱它的tRNA特异性。一个系统的组合方法被创建出来解决此问题[Wang et al., 2001a]；突变被引入到野生型合成酶中的氨基酸结合位点以构建一个合成酶文库，正负向选择被依次用来筛选此文库，以寻找那些对非天然氨基酸具有高特异性但不识别普通氨基酸的突变体合成酶。更多的突变可以用无规突变或DNA shuffling引入到初始选择的合成酶中以构建第二代文库。这个过程可以被重复直到具有所需性质的合成酶被定向进化出来。

*MjTyrRS*的酪氨酸结合位点的5个氨基酸残基被饱和和突变用来构建一个“*MjTyrRS*对位文库”。经过两轮筛选从此文库找到了一个合成酶突变体，这个突变体和*Mj*被表达在大肠杆菌中后，高度选择地掺入一个非天然氨基酸，O-甲基酪氨酸，到TAG指定的蛋白质位点。其掺入效率和忠实性均可与普通氨基酸媲美。这样，遗传密码首次被人工的方法扩充到包含第21个氨基酸[Wang et al., 2001a]。O-甲基酪氨酸在结构上接近酪氨酸与苯丙氨酸（图2），如此相似的氨基酸能被定向进化的合成酶准确地识别和区分，证明上述组合方法是强大而有效的。紧接着另一个合成酶突变体也很快被找到，它能同*Mj*一起掺入第二个非天然氨基酸，萘基丙氨酸。与第一个非天然氨基酸恰恰相反，萘基丙氨酸在结构上极大地有别于酪氨酸。此结果证明了上述方法的通用性，可被运用于各种不同的氨基酸[Wang et al., 2002a]。

二 对遗传密码的思考和展望

上述结果虽不能回答文首提到的所有问题，但至少表明遗传密码确实可以被扩充，而且这种扩充是通过模拟自然界掺入普通氨基酸到蛋白质中去的办法来实现的，即用氨酰tRNA合成酶加载氨基酸到同源tRNA上。既然自然界用这种方法掺入20个普通氨基酸，为什么它不简单地重复此过程来掺入第21和第22个呢？一些生物在掺入硒代半胱氨酸或吡咯赖氨酸时并不使用氨酰tRNA合成酶直接装载这两个氨基酸到tRNA上，而是借助额外的酶来转换已装载在tRNA上的别的氨基酸来生成所需的氨酰tRNA。值得注意的是，这两个氨基酸都是通过终止密码子来掺入的，而不是试图打破简并的有义密码子。也就是说，自然界在选择压力下为了掺入额外的氨基酸，其采取的策略虽不完全同于上述人工方法，但确是朝此方向——它熟悉的方向——发展的。这种扩充是否意味着一些普通的氨基酸在进化的过程中就是通过类似的途径而被添加到遗传密码中，使之成为现在的状态？倘若此猜测成立，即遗传密码的形成经过了扩充的过程而不是一个突然的冻结[Crick, 1968]，那它也没有理由停留于现存阶段。或许硒代半胱氨酸和吡咯赖氨酸便是它继续进化的蛛丝马迹。另外一个实验现象是，由琥珀终止子来编码非天然氨基酸并没有对宿主大肠杆菌造成明显的不便，诸如生长缓慢和毒副作用[Wang, 2002]。这在一定程度上意味着进化过程中密码子的重新编排存在其可能性。

为了探讨往遗传密码中添加新的氨基酸会带来何种进化上的后果，一种能自我生物合成并利用21个氨基酸的大肠杆菌已被制造出来。运用上文提到的方法，一对*Mj*/突变合成酶被用来掺入非天然氨基酸，对氨基苯丙氨酸，到TAG指定的蛋白质位点。另外，额外的酶被表达到大肠杆菌中，用以从简单的碳源来生物合成对氨基苯丙氨酸[Mehl et al., 2003]。应用类似的方法，许多非天然氨基酸都应能被活体细胞生物合成和编码到遗传密码中。这种能自发生产并运用非天然氨基酸的生物体提供了极好的机会来应用活体非天然氨基酸诱变研究定向蛋白质和定向有机体的进化。一种“Amber Walking”方法可以用来构建把非天然氨基酸掺入到目标蛋白质各个可能位置的蛋白质文库[Wang, 2002]。由于非天然氨基酸可以引入新的更丰富的功能团，这些文库预计可以提供具有增强或新颖功能的蛋白质。随机基因组诱变与基因替换技术则可结合用于寻找新的、依赖于非天然氨基酸的大肠杆菌表型[Wang, 2002]。这些技术正用于研究大肠杆菌在不同选择压力下的进化情形，用以测试非天然氨基酸是否可以提供生存与进化优势。

三 应用与延伸

自2001年上述方法创立后，正交的*Mj/MjTyrRS*对和*MjTyrRS*的对位文库已分发给不同的研究者。应用这个系统和方法，迄今为止超过20个非天然氨基酸已经被掺入到遗传密码中。其中许多氨基酸具有普通氨基酸所不具备的独特性质，遗传编码它们进蛋白质使随意地在活体细胞中改变蛋白质成为可能。

例如, 酮基是能参与许多有机化学反应的一个功能团, 但普通氨基酸恰恰缺失它。含有酮基的对乙酰苯丙氨酸已被编码进蛋白质中[Wang et al., 2003]。利用酮基的独特反应性能, 含有这个氨基酸的蛋白质能够选择性地被荧光标记用以跟踪和成像, 或是被生物素标记用作高灵敏度检出(图3)。其它试剂都可以类似地通过酮基这个化学手柄选择性地结合到目标蛋白上, 例如: 连接自旋标记可用于电子顺磁共振测量, 连接金属离子螯合剂可用于产生新的催化功能, 连接交联剂可用于研究蛋白与蛋白或蛋白与DNA的相互作用等等。在生物制药方面, 聚乙二醇修饰可以延长蛋白质药物在血清中的寿命从而提高其效能。但常规的修饰方法常常造成非选择性和不完全修饰, 影响药物的纯度与批准上市的机会。应用对乙酰苯丙氨酸的酮基来进行聚乙二醇修饰, 选择性和反应效率都极大提高。另外一例是蛋白质的糖基化, 它能够调节蛋白质的折叠、生物活性、血清寿命和定位。很多蛋白质药物或候选药物其天然形式都是糖基化的。细胞生产的糖基化蛋白往往是不同多糖修饰的混合物。获得均一的糖基化蛋白对糖基化机理研究和生物应用都是非常重要的。虽然很多方法已经被建立用来制造纯的糖蛋白, 它们经常受限于蛋白质的大小, 或是产物的产量和质量不理想。同样地, 利用对乙酰苯丙氨酸的酮基, 单糖和多糖均可定点地直接连上蛋白质来制造均一的糖蛋白类似物。初始连上的单糖还可以用糖基转移酶来嫁接另外的单糖, 生成具有特定组成的多糖[Liu et al., 2003]。所有这些步骤的产率和纯度都高于其它方法。应用正交的Mj/突变TyrRS对, 糖基化的氨基酸也可能直接被编码入蛋白质中, 从而在大肠杆菌中生产纯的糖蛋白。又如, 定位地掺入含有重原子的非天然氨基酸到蛋白质中, 将加速解决蛋白质X-射线晶体结构的位相问题; 电化学活性或光活性的非天然氨基酸选择性地掺入蛋白质中后, 体外讯号便可用于控制体内的蛋白等等。

除了适用于不同的非天然氨基酸, 上述扩充遗传密码的方法也可被用于不同的细胞和生物体中。正交的Mj/MjTyrRS对是第一对成功地用来掺入非天然氨基酸的tRNA/合成酶对, 并且在短时间内许多非天然氨基酸特异的合成酶均可自MjTyrRS的对位文库中筛选出, 显示了酪氨酸合成酶的高度可塑性与定向进化性。因此, 酪氨酸的tRNA/合成酶对成为延伸此方法到其他生物体的首选。在接下的工作中, 细菌的酪氨酸tRNA/合成酶被用于掺入非天然氨基酸到哺乳动物细胞[Sakamoto et al., 2002]和酵母中。MjTyrRS对位文库中有5个氨基酸残基被饱和和突变, 细菌TyrRS中对应的残基也同样地被突变来建立相似的文库。应用类似的正负向选择筛选到了合成酶的突变体, 一些原来被编码到大肠杆菌中的非天然氨基酸现在已可以被定位地掺入到酵母蛋白质中去[Chin et al., 2003]。产率方面, 在大肠杆菌中, 用一升培养基通常可以获得2-10毫克含有非天然氨基酸的蛋白质, 哺乳动物细胞和酵母中的产量常为几十微克。进一步优化各个组分的活性和表达可望能使它们都得以提高。

可以预计, 除了酪氨酸的tRNA/合成酶对外, 其它氨基酸的tRNA/合成酶对也应该可以相似地被建立为正交对。这些专一于不同氨基酸的tRNA/合成酶对应该可以用来掺入更多不同种类的非天然氨基酸。另外, 通常的密码子包括终止子都是三联体密码。通过对tRNA和合成酶进行适当的改变, 它们也可以被用于解码四联体的密码子[2]。运用四联体密码子来编码非天然氨基酸可大大增加遗传密码的容量, 并把多个非天然氨基酸同时掺入一个蛋白质中。

四 小结

通过为天然氨基酸制备一套独特的tRNA、氨酰tRNA合成酶与密码子, 非天然氨基酸也可以像普通的20个氨基酸一样被遗传编码到蛋白质中。这表明遗传密码虽然通用于现有的所有生物体, 但它可以被人工方法扩充。此扩充方法对遗传密码自身进化提供了一个新的思考和研究视角。非天然氨基酸的掺入对生物体进化的影响也正在研究之中。同时, 由于这个扩充方法使研究者可以直接在活体细胞内定点地掺入非天然氨基酸到蛋白质中, 各种各样的非天然氨基酸将成为非常有用的工具用在体外和体内研究蛋白质的结构与功能, 从而更精确了解和控制这类重要的生物大分子及其参与的许多生命过程。被扩充的遗传密码也可能用来定向进化具有新颖性质的蛋白质甚至整个生物体。

参考文献

- Bock, A., et al. (1991). *Mol. Microbiol.* **5**: 515-520.
- Chin, J. W., et al. (2003). *Science* **301**: 964-967.
- Crick, F. H. (1968). *J. Mol. Biol.* **38**: 367-379.
- Di Giulio, M., and Medugno, M. (1999). *J. Mol. Evol.* **49**: 1-10.
- Knight, R. D., Freeland, S. J., and Landweber, L. F. (2001). *Nat. Rev. Genet.* **2**: 49-58.
- Liu, H., Wang, L., Brock, A., Wong, C.-H., and Schultz, P. G. (2003). *J. Am. Chem. Soc.* **125**: 1702-1703.
- Mehl, R. A., et al. (2003). *J. Am. Chem. Soc.* **125**: 935-939.
- Sakamoto, K., et al. (2002). *Nucleic Acids Res.* **30**: 4692-4699.
- Srinivasan, G., James, C. M., and Krzycki, J. A. (2002). *Science* **296**: 1459-1462.
- Walsh, C. T., et al. (2001). *Curr. Opin. Chem. Biol.* **5**: 525-534.
- Wang, L., Magliery, T. J., Liu, D. R., and Schultz, P. G. (2000). *J. Am. Chem. Soc.* **122**: 5010-5011.
- Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001a). *Science* **292**: 498-500.

Wang, L., and Schultz, P. G. (2001b). *Chem. Biol.* **8**: 883-890.

Wang, L. (2002) Expanding the Genetic Code of Escherichia coli, Ph.D. Dissertation, University of California at Berkeley, Berkeley.

Wang, L., Brock, A., and Schultz, P. G. (2002a). *J. Am. Chem. Soc.* **124**: 1836-1837.

Wang, L., and Schultz, P. G. (2002b). *Chem. Commun.*, 1-10.

Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003). *Proc. Natl. Acad. Sci. U S A* **100**: 56-61.

Wong, J. T. (1975). *Proc. Natl. Acad. Sci. U. S. A.* **72**: 1909-1912.

图1 一种通用的遗传编码非天然氨基酸到蛋白质中的新方法。正交氨酰tRNA合成酶装载非天然氨基酸到正交tRNA上，后者解码一个独特密码子，把非天然氨基酸定位地掺入到蛋白质中。

图2 文中提到的几个氨基酸的结构。

图3 通过遗传编码对乙酰苯丙氨酸入蛋白质中，其所含酮基可用于选择性地连接各种试剂到蛋白质上。图中示意了荧光标记和糖基化。

① 王磊，1972年生，加州大学圣迭哥分校博士后，Merck Fellow of the Damon Runyon Cancer Research Foundation，2003年Amersham Biosciences / *Science* “青年科学家”最高奖获得者。

② 四联体密码子中的前三位所构成的三联体密码子有可能被内源tRNA解码为一个氨基酸，所以在创建四联体密码子的时候，前三位通常选为罕见密码子或终止密码子，以降低内源tRNA的竞争。而四联体密码子的后三位通常不用担心，因为在核糖体中，mRNA是顺次被解码的，跳过一个核苷酸而直接解码后三位的几率很低。当然，从扩充密码子的角度考虑，最好的办法是用非天然碱基来构建新的核苷酸，即除了AT对和GC对外，构建新的XY对。这方面的研究工作已有了一定进展。

