

50 years since the double helix: Genetic Engineering is crude and old-fashioned 人类发现DNA双螺旋结构50周年： 基因工程是一项粗糙和过时的技术

Janet Cotter
国际绿色和平组织科学顾问
Chui stoph Thea
德国绿色和平组织项目主任

2003年4月

50年前当DNA的结构被确定的时候¹，人们高呼发现了“生命的秘密”。似乎DNA结构的确定就可以完全了解生命体，然而却忽略了关于DNA功能的基本问题，而仍是现在我们也还没有解决的。后来，在七八十年代，向生物体基因组中随机插入基因的技术发展起来，被称作“基因工程”（或转基因技术，译者注）。基因工程被誉为“生命”科学，因为它似乎可以按要求来规划和设计生命体。许多转基因作物被基因工程公司商业化生产，并故意释放到环境中。基因工程生物体，包括很多已经商业化生产的转基因作物，已经产生了很多非预期后果，却被解释为技术原因或者可以通过进一步的研究和采用更合适的技术来解决。但是，这些非预期后果的产生可能是由于基因工程技术本身的缺陷所致。DNA双螺旋结构发现后的50年里，科学显示，基因表达远远不像基因公司所希望的那么简单。

50年代提出的中心法则(The Central Dogma)，是基因工程的最重要的基础。然而在今天，中心法则在描述基因表达方面显得过于简单了。现在发现在高等生物（例如植物）中，哪个基因表达、何时表达是由很多因素（例如蛋白质、RNA）之间的反应(reactions)和互动关系(interactions)决定的。这些反应和互动关系在中心法则中被忽略了。在现今的科学界中，最初的中心法则已经让位于一个复杂的调控网络，而这个网络调控基因表达的方式还远远没有被阐述清楚。这样说来，基因工程是基于一个过于简单和过时的法则。由此，基因工程不可能生产出能被接受的释放到环境和食物链中的产品，因为没有考虑到生物体中现存的复杂的调控系统。

这篇文章详细阐明了科学界对DNA的本质和分子生物学的中心法则理解的变化是如何使基因工程的范例作废的。

两个定义：

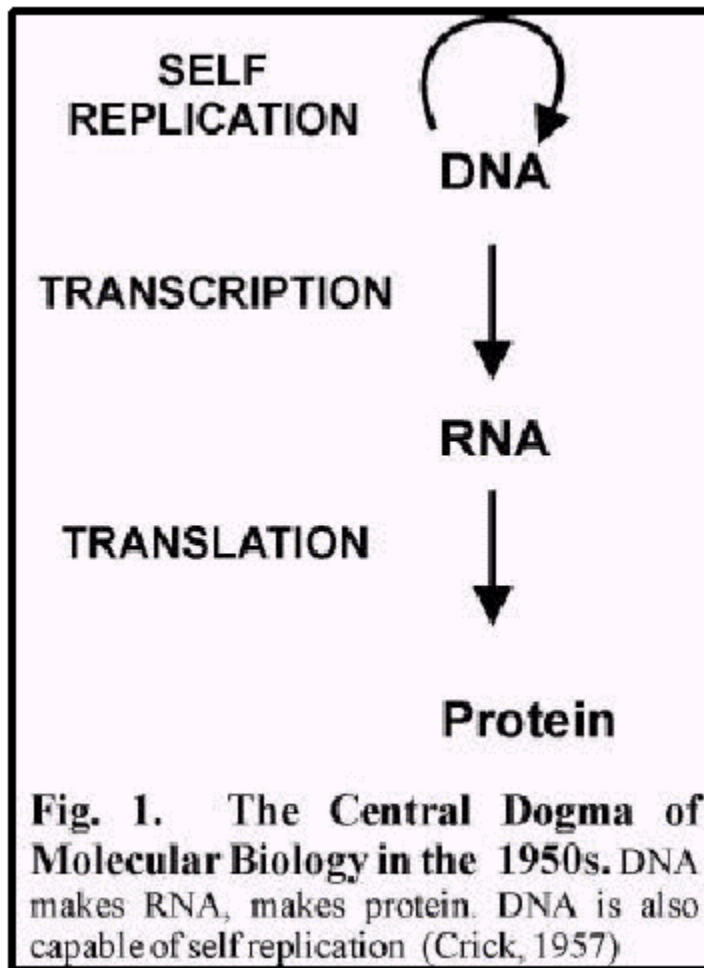
中心法则(The Central Dogma):1957年²，克里克提出了分子生物学中的重要法则，“中心法则”，认为DNA(遗传信息)生成RNA(中介物)，RNA生成蛋白质(它行使功能)，如图1所示。充当DNA和蛋白质中介的RNA叫信使RNA，mRNA。中心法则规定，没有信息从蛋白质传送到蛋白质，从蛋白质到RNA以及从蛋白质到DNA。然而，如同下面将要描述的，现在发现，许多基因调控是将信息传送给DNA，例如蛋白质传送信息到DNA。

遗传密码(The genetic code):遗传密码是包含在DNA中的“字母表”，于60年代被破译³。DNA碱基的排列决定了构成蛋白质的氨基酸种类以及它们在蛋白质中的排列顺序。DNA或RNA中每三个碱基构成一个单位，称为密码子(triplet)，每个密码子编码一个氨基酸。

50年代如何看待DNA：

自从发现以来，双螺旋的简单结构不仅震惊了科学家也同样震惊了世人。这个结构不仅优美、稳定，而且易于复制，能分成两条单链并根据碱基互补形成两条新的单链。当时，物理学家和化学家发现很多物质都是由简单、重复的单元组成。生物学家也在寻找生命的最基本构成，DNA和它编码的基因看起来像是一个很好的结果。

遗传密码的破译使得DNA被认为是细胞的主宰、决定生命的高级实体(superior entity)。在遗传密码发现的同时，信息技术高速发展。DNA的序列贮存了所有的信息，被称为“生命之书”。DNA被认为是遗传信息的携带者，同时利用这些信息“构筑”细胞。在科学和非科学的描述中，DNA被认为是主动的(active)。这个主动的特性在通俗化描述中心法则的时候表现出来⁴：“DNA生成RNA，RNA生成蛋白质。”一般理解为DNA是发起者(initiator)，通过基因生成蛋白质。DNA作为主宰分子的概念构成了基因工程的理论基础。



DNA的基本结构和遗传密码至今并未受到质疑，但是它们的最初描述现在看来是过于简单化了。不幸的是，这个过于简单的理念阻碍了对DNA功能更全面的理解，更不幸的是，在几乎全无理解、根本没有考虑到基因调控的复杂性的情况下，转基因生物的商业化生产就已经开始了。一些重要的发现，由于它们不符合根深蒂固的简单模型，因此不幸地被忽视。例如，转座子(transposon)是移动的遗传物质，它们主动或被动的在基因组中移动。转座子是在50年代由Barbara McClintock提出的。但是遗传因子可以移动这个重要的发现在当时并没有被接受。因

为人们固执的认为DNA是一个固定的结构，基因在基因组的位置是固定的。直至1983年McClintock才因为她的发现获得了诺贝尔奖。

过去50年，出现了很多关于DNA特性和基因表达的重要发现。这些现象为我们展现了一个较50年代提出的中心法则更为复杂的图像。这从根本上动摇了基因工程的科学基础，因为它仍然建基于50年代提出的中心法则。

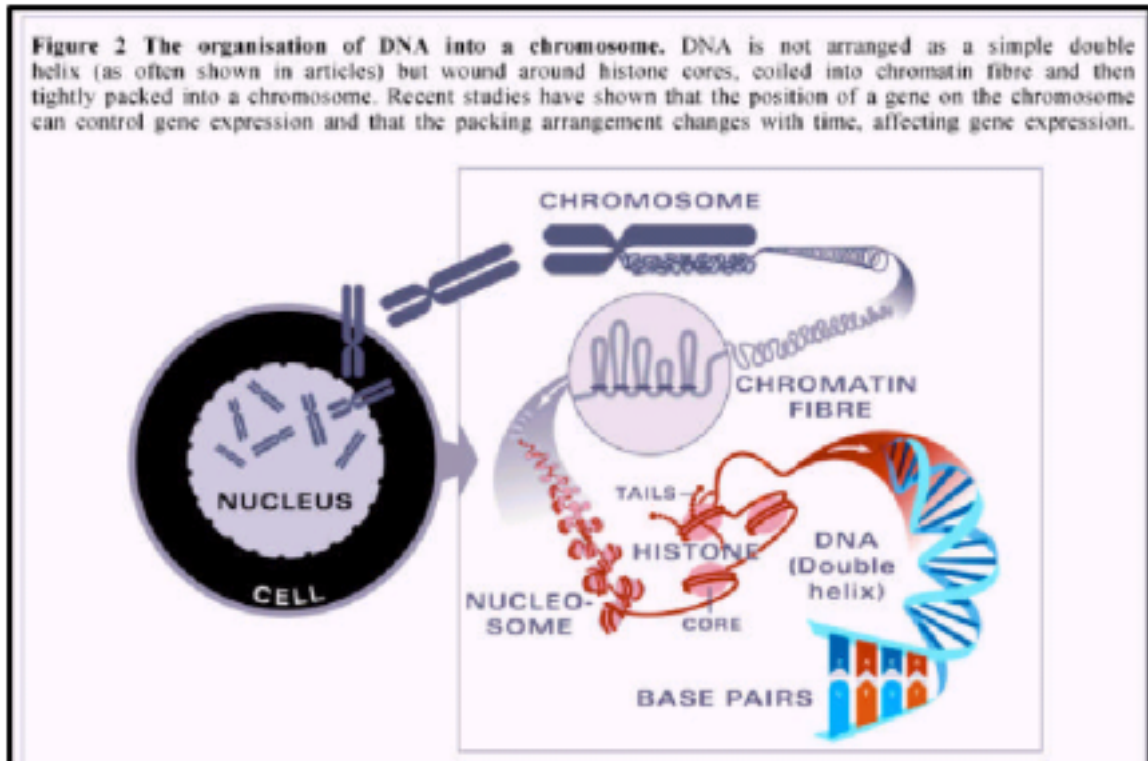
过去50年的发现：

回顾这50年，许多大大小小的发现动摇了最初的中心法则和DNA是固定的、不可改变的观点。这些新发现形成了现代分子生物学的基础，它们显示了虽然中心法则的核心部份仍然正确，但是很显然是一个过于简化的模型。基因表达受控于一个复杂的调控网络，而人类只是刚开始理解。⁵

1) DNA折叠(packi ng)影响基因表达

DNA 一般被描述为一个有直的轴的双螺旋结构。实际上，这个轴不仅不是直的，而且还不断地多级旋绕，直至形成一个复杂的三维结构。现在已经证实这种结构可以影响基因表达，而这并未在最初的中心法则中体现。

DNA是由6种不同的“建筑材料”组成：糖，磷酸和4种碱基。它们结合在一起形成一种线状的分子结构，与互补的链（4种碱基成对互补）形成双螺旋结构。理想的DNA结构是这个双螺旋围绕一条直的轴，DNA序列就是线性的碱基序列。实际上，这个轴又不断地旋绕，形成一个超螺旋结构，就是我们所称的染色质（chromatin），图2所示。染色质的基本亚单位是核小体（nucleosome），它的结构就是DNA盘绕在一组蛋白质（组蛋白，histones）上。每个核小体之间由一小段DNA连接，这个串珠（beaded）结构再旋绕成一个紧实的结构----染色质。在此过程中，DNA的长度又压缩了约50倍。染色质进一步缠旋成为染色体，所有的染色体集中在细胞核区形成了生物体的基因组。



DNA的螺旋结构是动态的，缠绕的状态乃至染色质的结构都是经常变化的，这种变化并不是随意的，而是体现某种调控，并对基因表达产生影响。⁶“进入细胞核的调节信号碰上了染色质而不是DNA，活化基因表达的有速度限制的生化反应在大多情况下牵涉到染色质结构的改变。”⁷

染色质的某些区域比其它区域的密度更高。染色质可以分为开放式的染色质和折叠密度高的、封闭式的染色质，后者似乎难以让调节蛋白和RNA渗入。在染色质中，经常被转录的基因往往位于染色质的边缘，有科学家估计这可能是因为这部分DNA比较容易接触到转录酶。⁸但究竟是基因的位置决定了它的活动或是其它，现在还不是太清楚。那些位于或接近紧密折叠染色质的基因可以是永久性沉默的基因，这些DNA的序列不改变，在遗传过程中作为特性被保留下来。⁹

位置的影响超越了染色质的三维结构。至少在一些细胞中，每条染色体在细胞核中都倾向于保持自己在细胞核中的原有位置，在细胞分裂后回到本来的位置。¹⁰含有活跃基因较多的染色体往往位于核的中间位置，而含活跃基因较少的染色体则位于核的边缘。^{11, 12}所有这些导致在《自然》杂志DNA双螺旋结构发现50周年纪念刊的时候提出：“自从双螺旋发现后，经过了50年，多少的庆祝声音会坦承我们仍然不能详尽了解DNA在细胞中是什么样子？”¹³

组蛋白（DNA所缠绕的蛋白质）在调控基因表达的过程中也起到很大的作用。组蛋白在它们的“尾”部区域可以被化学修饰，而这个过程现在被认为是调节基因活化或者沉默的主要机制。甚至有意见认为组蛋白尾巴的修饰模式可能构成组蛋白密码，与DNA遗传密码一起决定生物的基因活动。¹⁴

简言之，染色体内的DNA结构影响了基因表达。DNA所缠绕的组蛋白和染色质的超螺旋结构，都在调节基因表达上有重要作用。

2) 基因表达 “开”和“关”的转换

DNA包含遗传信息。然而现在我们已经明确这些遗传信息需要被启动才能进行蛋白质的表达。这个激活的过程需要转录因子，即结合到特定DNA序列(例如负责启动基因表达的启动子(promoters)和增强子(enhancers))的蛋白质。染色质的结构必须允许转录因子与之产生作用(即是开放而非封闭的)。根据中心法则，遗传信息不能从蛋白质传递给DNA，但是很明显地，重要的基因调控信息是由蛋白质以转录因子的形式将信息传递给DNA，基因表达才能得以激活。

基因可以被关闭，或者说沉默。小的碳氢化合物分子(hydrocarbon molecules)或甲基原子团(methyl groups)可以加到基因组的DNA上，这些分子可以阻止转录，“关闭”基因表达作用。科学家至今仍弄不清甲基化的原因及在分子层次上基因沉默化如何发生，但有意见认为这个过程可能牵涉到组蛋白的甲基化修饰。¹⁵

3) 从RNA到DNA “逆转录”

中心法则认为遗传信息从RNA到DNA是“罕见”¹⁶的。然而，一些病毒将遗传物质带入宿主细胞，并整合到宿主的DNA上。某些RNA病毒就能用自己的RNA为模板在宿主细胞内合成DNA，并插入到宿主的DNA上¹⁷，这被称为“逆转录”。

4) RNA的角色----越来越受到重视

中心法则(图1)视RNA¹⁸仅仅为核中的DNA制造蛋白质的信使，这从它的名字就可以看出来----信使RNA(mRNA)。然而现在已经发现了很多种类的不编码蛋白质的RNA，称为不编码RNA(noncoding RNA, ncRNA)¹⁹。这些ncRNA 大部分都是最近才发现的。如此长时间的未被发现是由于它们不编码蛋白质所以不适合中心法则描述的：“几乎所有基因的辨别都假设这些基因编码蛋白，所以即使在全基因组时代，ncRNA还是被忽略了。”²⁰“有没有一种可能性，就是一大类基因由于它们不编码蛋白质，所以一直不被发现？”²¹

关于ncRNA最新的发现是干扰RNA(RNA interference, RNAi)，在顶尖的科学杂志《自然》²²被誉为2002年最重要的发现。RNAi是一些小的RNA片断，这些片断结合到互补的mRNA上，帮助其形成双链结构，后来再被酶分解掉。这样，RNAi就可以干扰mRNA，即通过作用于中间产物RNA，停止蛋白质的生产。推测这种行为可能是作为细胞内的免疫系统，对RNA病毒采取预防的一种措施，但是它们也可以关闭基因²³。其它一些ncRNA似乎起酶的作用。不编码蛋白但编码ncRNA的基因位于DNA非编码区。这些DNA非编码片断曾经被认为是基因组序列的“垃圾DNA”(junk DNA)²⁴。现在，这些所谓的“垃圾DNA”不仅不是垃圾，而且可能是基因调控网络的重要部分。

许多关于ncRNAs的问题将会进一步的研究：“有多少RNA基因？它们的重要性是什么？细胞除了赋予蛋白质作用以外，还赋予了RNA何种作用，为什么？”²⁵

5) 人类基因组的DNA序列

2001年人类基因组序列的公布^{26,27}被誉为是开创了新的纪元。这其中最使人震惊的是人类基因组的大小。人类基因组一共有大约30,000到40,000个基因，这远远小于预测的100,000个基因。这个发现已经深深的动摇了对现有基因调控的理解。以往过于简单的中心法则再不能解释基因的表达了。较之一些简单的生物比如一种蛔虫(*Caenorhabditis elegans*, 将近19,000个基因)，果蝇(*Drosophila malogaster*, 将近14,000个基因)，或者一种简单植物鼠耳芹(*Arabidopsis*

thaliana，将近25,500个基因)，为什么像人类这种高度进化及复杂的物种只相对多出如此少量的基因？如果按照一个基因对应一个蛋白质的说法来算，人类基因组产生的蛋白种类远远多于基因的数量。理论上，一个基因必定能够通过复杂的DNA - RNA- 蛋白质途径生产多于一种的蛋白质，这个途径被认为是通过拼接机制(splicing mechanisms)实现的。

基因和拼接机制

什么是基因？基因的定义仍然没有完全界定清楚。它于1909年被提出作为一个遗传单位来理解²⁸。随着1953年DNA的结构确定以及60年代遗传密码子的发现，基因被定义为一段编码蛋白质的DNA序列，定义自此开始模糊。从DNA生成蛋白质需要很多的信息和调控而不只是由DNA序列决定的组成蛋白质的氨基酸顺序（按照遗传密码子），例如，需要“激活”和“停止”的指示。另外，一个“基因”还应该能够表达不只一种蛋白质。

在真核生物中，由DNA转录成的原初RNA需经过多次的加工才成为mRNA，然后才被翻译成蛋白质。一个基因可以被看作是一系列的小单元、外显子和内含子的组合，携带部分蛋白质信息的外显子和不编码的内含子交替出现。真核生物的拼接过程可以使编码蛋白质的外显子连接成成熟的mRNA。内含子不包含蛋白质的信息，但是含有调节序列和拼接信息。外显子的不同组合可以生成不同的mRNA，这个过程称为“选择性拼接”。这样，mRNA 被剪切形成合适的蛋白质模板²⁹。蛋白质和小的ncRNAs都是拼接的控制因素³⁰。据估计40%-60%的人类基因存在不同的拼接形式³¹。“明显地，增加蛋白质的种类并不简单只是关系于增加基因的数目，而是取决于基因组的基因数目及对这些基因的选择性拼接。”³²

切割机制

不同的机制也作用在蛋白质的形成上。一些蛋白质形成的时候是较大的前体形式，然后被剪切成单个或多个蛋白³³，起这个作用的是特定的酶（蛋白酶）。

关于“基因是什么”的问题现在仍然没有找到答案。在很多情况下，DNA序列、“基因”不能再被认为是编码一种蛋白质了，最终的蛋白质产物是所有不同的RNA相互作用的结果，包括ncRNA和调控蛋白。DNA的结构和遗传密码都不能满意的解释从如此之少的人类基因组中如何产生如此之多种类的蛋白质。

Stephen Jay Gould 在人类基因组序列发布的时候评论说：“复杂性的第一要素不是更多的基因，而是由这些较少编码单元产生的较多拼接和相互作用 --- 这些相互作用（术语上叫聚现特征，emergent properties）必须在存在性的水平上解释，因为它们不能单独在独立的部分被预测。”³⁴

DNA 研究的现状：

50年来的发现使我们的认识从DNA是贮存信息、控制生成蛋白质的主宰分子转变为DNA是一个多层次复杂调控网络中的信息贮存库(store of information)。中心法则的语言将被修改，DNA不再是主动的，而变成了被动的。DNA不是自我复制，是由一系列相互作用的酶复制的。DNA不能复制DNA，它只是模板。在新的科学语言中，DNA的角色变得更被动。DNA的物理化学性质解释了它的被动性。尽管分子量很大，DNA仍是稳定的分子，表现出不活跃、被动的性质。

分子生物学的中心法则是基于DNA是活跃的，及可以控制基因表达的基础上提出“DNA生成RNA，RNA生成蛋白质”。50年来的研究已经改变了这个图像。事实上，新的发现不断修改了基因表达的范例。例如，两份最具影响力的科学杂志《自然》和《科学》几乎在2002年每个月都发表了一些关于基因表达的新方面和评论。这都给我们展现了细胞内基因表达是一个由各种因子和调控机制组成的复杂网络。

“如果所有这些新发现破坏了沃森(Watson)和克里克(Crick)的经典模型所带来的美丽幻像，但却给我们提供一个更加丰富的全景。信息传递最基本的机制，碱基互补，是如此的经典，以至可能蒙闭了大家的眼睛，看不见信息传递的总过程是令人敬畏的复杂。这些分子不是随意同别的分子结合、连系。它们的功能是预先确定的，然后在实现功能前还需要向更高层次提出申请并被允许，形成一个复杂的调控网络。”³⁵

关于转基因作物的结论

基因工程是基于那个过于简化的中心法则(见图1)。基因工程假定一个基因在系统内扮演着独立的单位。事实不是这样，基因在基因组中的位置很大程度上决定了它如何受控于严格的调控。将基因随机性地插入到基因组中意味着它不受控于调节系统，所以基因工程违背了现有的对基因组复杂性状的理解。基因工程也没有考虑到RNA、蛋白质和其它细胞组分的调控作用。转入的基因在构建时通常包括一个启动基因表达的启动子，然而这些启动子不是原本细胞调节系统的一部分，大多转基因生物都采用了花椰菜花叶病毒启动子(CaMV35S)，它们会使插入的基因时刻都在所有细胞表达。

DNA插入基因组是随机性的，它可以导致插入位点的原DNA被删除和重新组合³⁶。而且DNA的插入往往是强迫性的，例如，“基因枪”(Gene gun)方法就是将沾有DNA的粒子打进细胞。这个方法由于是随意的插入基因片段，导致细胞原有的基因被删除和重组而被广泛批评。基因枪技术是孟山都公司用来研制转基因作物的方法。现在发现在多种转基因作物中都存在不规则变异，例如抗草甘膦(Roundup Ready)玉米(GA21, NK603)和抗草甘膦大豆。

DNA研究的新进展认为DNA只是活细胞中基因调控网络的一部分，这一点就足以彻底地削弱基因工程所依据的基础。调控网络的复杂性使得高等生物的转基因技术显得粗糙和过时。

人们已经认识到转基因生物的问题，如基因的沉默现象(events of gene silencing)，但基因工程发展者只把它视为可被解决的技术问题。随机的、强迫性的将DNA插入基因组必定会影响染色质的三维高级结构，但这些都未被引起重视。同样，也没有研究DNA的插入对调控网络的影响。但非常可能的是，至少一部分非预期的效果是由于破坏了调控网络所引起的。

在转基因植物的培育过程中，有很多植株由于出现不合适变异而被筛除。亦有许多关于商品化转基因植物出现非预期性状的例子，例如孟山都公司的抗草甘膦大豆在干热环境下，茎杆会爆裂，可能是由于木质素含量的增加³⁷。抗草甘膦棉花的棉桃也会不明原因的掉落³⁸。但是这些都没有被调查原因(或者是调查了而结果没有公布)和给予完满解释。一个最惊人的非预期的性状是发生在抗草甘膦大豆的减产现象。在对比试验中，抗草甘膦大豆要比同一品系的非转基因大豆产量平均低10%，研究人员认为5%是由于插入的外来基因或改造过程所致³⁹

转基因生物的拥护者会说，许多对代谢有干扰的情况会在培育过程中按照标准被剔除。可是不是所有的性状都容易被检测到，有些性状要在几个世代⁴⁰以后或极性环境条件下(干旱、高温)才会显现。

基因工程的监管

尽管监管机构(如一些欧美国家的)已经认识到非预期后果会发生,但是评估转基因生物对环境以及食物/饲料的风险仍然基于那个过于简化的中心法则:一个基因对应一个功能的线性法则(图1)。如果要检测插入基因与基因组的相互作用,不能简单的从插入基因的功能推测,而牵涉到寻找非预期后果,这是一项艰巨的任务。

转基因作物的风险评估是基于DNA插入序列及其表达,但不包括那些因插入序列而受影响的其它基因(无论它们编码蛋白质与否),也没有包括插入序列对基因组整体结构的影响。风险评估(risk assessments)疏漏了这些方面,所以得到的结果显然也是不全面的。预防原则(Precautionary Principle)就比风险评估对于基因工程生物合适的多,预防原则是针对那些潜在的危险采取一定的防范措施而定的,即使这些危害还没有表现出来。任何转基因生物的释放会是不可逆的、带有潜在危害的。所以说,为防止转基因生物体释放到环境中需要施行预防原则。

生物技术不等同于基因工程

多年来生物技术一直把基因工程放在研究的重点,现在已转向基础研究,去了解DNA的实际功能。如果早一些将重点转移到这方面来,或许我们早就发现了生命的复杂性不是基因工程所能操作的。生物学并不存在单一种的“积木”,就像物理学也没有。我们面对的生物是一个网络的、调控的、没有轻松解决办法的世界。所以,如果转基因技术注定要失败,遗传学研究如何能够提高我们的生活质量呢?

生物技术经常被误会等同于基因工程,这是错误的。生物技术是有关生物的,一般是指分子水平技术的名词。我们对植物、动物基因组和基因调控的了解可以为我们做很多事。例如遗传标记(genetic markers,特定的DNA片段)会对应一些特定的性质,这一特点可被利用作遗传标记辅助育种(Marker-assisted selection, MAS)。遗传标记的选择和观察需要遗传学的概念和技术。遗传标记辅助育种不能通过基因操作改变植物的DNA,而是在种子发芽和培养过程中,它可以根据是否含有遗传标记,来选择继续培养的植株。比起基因工程, MAS是一项与基因表达新发现相一致的生物技术,它认识到有复杂的网络调控着基因表达。MAS在培育新植物品种方面有着极大的空间,因为它允许复杂的特征(如干旱抗性),甚至多重复杂特征参与育种。遗传标记辅助育种一般被认为是加快了传统选育的过程。所以,通过这种方法选育的作物不会像转基因作物那样对环境带来不可预测的后果。

遗传标记辅助育种被越来越多的应用,而且随着新标记物的发现将会更加广泛。例如,大米抗旱、玉米维生素E的遗传标记物已经被鉴定出来⁴¹。然而问题是支持这方面研究的经费需要增加,但是现在可用的经费都被用于基因工程了。⁴²

结论

发现DNA双螺旋结构以来的50年里,尤其是近几年,许多关于DNA性质的发现都在冲击中心法则。DNA螺旋结构的作用,基因的沉默现象,反转录以及多种RNA的发现,将基因表达的法则从一个简单的线性关系(DNA生成RNA,再生成蛋白质)改变成一个复杂的网络,基因表

达在其中受到许多非DNA因素的严密调控。

基因工程技术是粗糙和过时的技术。它基于过于简化的中心法则和一个基因生产一个蛋白质的线性关系。基因工程与现在发现的生物具有复杂的调控网络不相吻合，是注定要失败的。基因工程不是唯一的生物技术。其它的现代生物技术手段（例如遗传标记辅助育种）将会得到广泛的应用，而不会有转基因生物所具有的潜在风险。

DNA是重要的，但并不是生命奥妙的全部。

References:

- 1 Watson JD & Crick FHC (1953) Molecular structure of nucleic acids. *Nature*, 171, 737-738.
- 2 Crick FHC (1957) On protein synthesis. *Symp. Soc. Exp. Biol.*, 12, 138-163.
- 3 Crick FHC, Barnett L, Brenner S, Watts-Tobin RJ (1961) General nature of the genetic code for proteins. *Nature* 192, 1227-1232.
- 4 Crick FHC (1957) On protein synthesis. *Symp. Soc. Exp. Biol.*, 12, 138-163.
- 5 Crick FHC (1970) Central Dogma of Molecular Biology. *Nature*, 227, 561-563.
- 6 Lee T et al. (2002) Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science*, 298, 799-804.
- 7 Felsenfeld G & Groudine M (2003) Controlling the double helix. *Nature*, 421, 448-453.
- 8 Felsenfeld G & Groudine M (2003) Controlling the double helix. *Nature*, 421, 448-453.
- 9 Pearson H (2003) Beyond the Double Helix. *Nature*, 421, 310-312.
- 10 Holmes B (2003) Location, location, location. *New Scientist*, 15 March 2003, 48-49.
- 11 Holmes B (2003) Location, location, location. *New Scientist*, 15 March 2003, 48-49.
- 12 Holmes B (2003) Location, location, location. *New Scientist*, 15 March 2003, 48-49.
- 13 Pearson H (2003) Beyond the double helix. *Nature*, 421, 310-312.
- 14 Tanabe H et al. (2002) Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates *Proceedings of the National Academy of Sciences USA*, 99, 4424-4429.
- 15 Ball P (2003) Portrait of a molecule. *Nature*, 421, 421-422.
- 16 Felsenfeld G & Groudine M (2003) Controlling the double helix. *Nature*, 421, 448-453.
- 17 Felsenfeld G & Groudine M (2003) Controlling the double helix. *Nature*, 421, 448-453.
- 18 Crick FHC (1970) Central Dogma of Molecular Biology. *Nature*, 227, 561-563.
- 19 Watson JD et al. (1992) *Recombinant DNA*. 2. edn. Scientific American Books, New York, USA.
- 20 Messenger RNA (mRNA), transcribed as complimentary strand to a DNA sequence and translated into a protein, is considered coding RNA, while all other classes of RNA, especially transfer and ribosomal RNA are non-coding.
- 21 Henikoff S (2002) Beyond the Central Dogma. *Bioinformatics*, 18, 223-225.
- 22 Eddy SR (2001) Non-coding RNA and the modern RNA world. *Nature Reviews Genetics*, 2, 919-929.
- 23 Eddy SR (2001) Non-coding RNA and the modern RNA world. *Nature Reviews Genetics*, 2, 919-929.
- 24 Dennis C (2002) 2002 in context: the genome's guiding hand? *Nature*, 420, 732.
- 25 Henikoff S (2002) Beyond the Central Dogma. *Bioinformatics*, 18, 223-225.
- 26 Keller EF (2000) *The century of the gene*. 186p. Harvard University Press, USA.
- 27 Eddy SR (2001) Non-coding RNA and the modern RNA world. *Nature Reviews - Genetics*, 2, 919-929.
- 28 International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921.
- 29 Venter JC et al. (2001) The sequence of the humane genome. *Science*, 291, 1304.
- 30 Keller EF (2000) *The century of the gene*. Harvard University Press, USA.
- 31 Sorek R & Amitai M (2001) Piecing together the significance of splicing. *Nature Biotechnology*, 19, 196.
- 32 Lewin B. (2000) *Genes VII*. Oxford University Press, Oxford.
- 33 Modrek B & Lee C (2002) A genomic view of alternative splicing. *Nature Genetics*, 30, 13-19.
- 34 Sorek R & Amitai M (2001) Piecing together the significance of splicing. *Nature Biotechnology*, 19, 196.
- 35 see, e.g. Lewin B. (2000) *Genes VII*. Oxford University Press, Oxford.
- 36 Gould SJ (2001) *The New York Times*, 19 February 2001.
- 37 Ball P (2003) Portrait of a molecule. *Nature*, 421, 421-422.

- 36 Svitashv, SK & Somers DA (2001) Genomic interspersions determine the size and complexity of transgene loci in transgenic plants produced by microprojectile bombardment. *Genome*, 44, 691–697.
- 37 Coghlan A (1999) *New Scientist*, 20th November, p. 25.
- 38 Fox JL (1997) Farmers say Monsanto 's engineered cotton drops bolls. *Nature Biotechnology*, 15, 1233.
- 39 Elmore RW, Roeth FW Nelson LA Shapiro CA Klein RN Knezevic SZ & Martin A (2001) Glyphosate-resistant soybean cultivar yields compared with sister lines. *Agronomy Journal*, 93, 408–412.
- 40 Riha K McKnight TD Griffing LR & Shippen DE (2001) Living with instability: plant responses to telomere dysfunction. *Science*, 291, 797-1800.
- 41 Price AH, Townend J Jones MP Audebert A & Courtois B (2002) Mapping QTLs associated with drought avoidance in upland rice grown in the Philippines and West Africa. *Plant Molecular Biology*, 48, 683–695.
- Rocheford TR, Wong JC, Egesel CO & Lambert RJ (2002) Enhancement of Vitamin E Levels in Corn. *Journal of the American College of Nutrition*, 21, 191S–198S.
- 42 Knight J (2003) A dying breed. *Nature*, 421, 568-569.