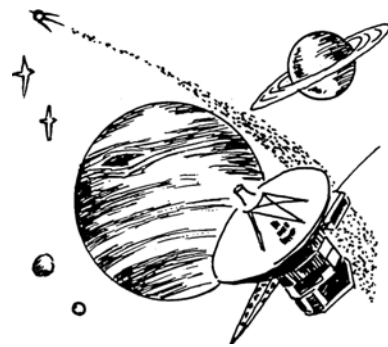


“宇宙与生命”系列讲座之五——

生命之难

章德海



莫大的宇宙中，居然有生命。物理规律的本质是对称的统一，物理世界的实现是对称的破缺。对称破缺的方式是多种多样的，因此物理世界的实现不会是唯一的。物理规律不会只产生一个宇宙。众多的荒芜宇宙中，能产生生命的宇宙是个奇迹。前四讲提到，我们的宇宙是为将来有生命而诞生的。由此可见生命对宇宙的重要性。但是我们只知道地球上生命。已经发现，地球上所有的生命的主要成分是由 DNA，RNA 和蛋白质构成。所有生命体的 DNA 都是由称为 A、G、C、T 的 4 种核苷酸靠磷酸二酯键连接构成的长链，两链之间碱基以氢键配对构成双螺旋。如果核糖不脱氧且 T 代换为 U，那么就构成 RNA。所有生命体的蛋白质都是由 20 种左旋氨基酸构成（极少古细菌略有例外）。我们称这 28 种分子为生命基本单元，它们都是 C、H、O、N、P、S 这六大生命元素构成，分子量在 100 到 300 左右。这就说明了地球上生命的统一性，也暗示了所有地球生命都有单一的生命起源系统。生命的中心法则由 DNA 储藏和复制信息，由 RNA 传递信息并制造蛋白质，由蛋白质实现功能。每三个碱基编码一个氨基酸。几乎每一步生化反应都需要不同的专用酶的催化。由各种成分的氨基酸的不同排列和构形可以形成大量不同种类的蛋白质。其中一部分蛋白质就是生命的至关重要要素——酶，另一些可以作为生命体的各种功能结构。各种功能的不同组合形成大量物种，地球物种在 100 万以上，包括历史灭绝的可能上亿种。生命的信息量是相当大的，酵母的 DNA 有 1000 万对碱基，人的 DNA 有 30 亿对碱基。也不是碱基对的数目越大，生物就越高级。玉米的 DNA 有 50 亿对碱基，蝶螈的 DNA 有 900 亿对碱基。这么庞大数量的碱基并不都对应蛋白质的编码，尽管基因的主要功能之一是编码蛋白质。已查明人的 DNA 只有约 3% 的序列编码了约 3 万个基因，算下来每个基因的编码部分有约 3000 个碱基，可以编码 1000 个氨基酸。基因中还包含前导区、尾区、许多功能不明的内含子、基因片断、拟基因。

20 卷第 5 期 (总 119 期)

加上这些以后，每个基因约有碱基 3 万个。基

因的功能高度复杂，一个基因可通过剪接 mRNA 实现多蛋白复制，基因本身和基因间形成复杂的调控机制，生命的实现依靠调控。我们深入不到这些问题。我们只列举一些简单情况，例如细胞中负责代谢的一个重要细胞器——线粒体，看看它对进化有什么启发性特征。线粒体有独立的 DNA 体系，是母系遗传的。动物的线粒体 DNA 有 1.7 万对碱基，编码 40 个基因蛋白质。其中一个研究得比较透彻的是细胞色素 c，它有约 104 个氨基酸，比起刚才提到的平均数 1000，是个很小的蛋白质分子了。人与猩猩的细胞色素 c 全同，与猴子的也仅仅只相差一个氨基酸。即使与酵母的细胞色素 c 相比，也仅有 44 个氨基酸不同，即要保持 60 个氨基酸的排列不得变动。在分子水平上其同异完全反映了生命体进化间的亲疏远近关系。

如果这 60 个氨基酸的排列是任意的，那么就会有 $20^{60}=10^{78}$ 种组合构形，要从这么多个组合中选出刚才所说的那个唯一正确的组合，真是太不容易了。要知道我们当今可见宇宙内的总核子数也才 10^{78} 个！我们来算算全宇宙最多能有多少个蛋白质分子。即使我们宇宙全视界内有 1 千亿个星系，每个星系有 1 千亿个恒星，100 每个恒星有一个地球（地球是罕见的），每个地球表面有 1 米厚的 100 肽蛋白质，那么总共才有 10^{60} 个蛋白质分子。让每个蛋白质分子每秒钟变异一个组合构形，那么宇宙 100 亿年龄里，所有这些蛋白质分子也才只能有 10^{78} 个组合构形。这样的算法，勉强把唯一正确的组合构形找了出来。当然其中每个环节的考虑变动可能很大，所以算法并不精确。但是它大体给我们一个印象，为什么宇宙要这么大，为什么宇宙要这么老，为什么生命进化时间要这么长。足由此可见，生命之难。这里有个例子，动员全世界的业余计算机能力去寻找梅森素数，最近也要好几年才能找到一个。

生命进化中找到蛋白质的正确氨基酸组合既不是完全随机的，也不是完全遗传的。完全遗传就没有了进化，没有继承完全随机就不成其生命。生命进化的速度一定被遗传和随机变异的对立统一所控制，但是目前并没有被真正研究清楚。从生命起源到文明诞生，究竟需要多长时间，很难回答。我们仅知一个例子，从地球生命起源到人类出现大约需要 36 亿年。也许有人觉得，生命的起源和进化不一定要这么难。也许例如少 10%，32 亿年就够了。如果这样，我们马上就要遇到一个费米悖论

费米有个关于外星文明的很有名的悖论。费米认为，如果生命进化到高级文明是容易的，那么为什么没有外星高级文明来访问我们？费米问题不是个玩笑，但对费米问题的回答却往往是些玩笑。费米问得很有理。我们人类文明才短短 5000 年，就已经上知宇宙结构和起源，下知顶夸克 τ 中微子。人类 300 年的技术已经进步到能够人类登月、核能发电、核能潜艇、全球联网、电脑控制、精确制导。再过不到 100 年，石油告罄、人口过剩、粮食短缺、全球变暖、环境恶化、核武临空，人类必然面临严重危机和发展瓶颈。如果这道道难关能被人类跨越，将来 1000 年人类技术会发达到怎样一个高度？是受控核聚变、月球氦三矿、量子计算机？谁也无法预测。那时是否已使用光子火箭来去自由，或者进化到冬眠人、克隆人、机器人再向外星移“民”？一个银河直径不过 10 万光年。一旦高级文明有能力向星际扩张，其向全银河扩展所需的时间不会超过 1 亿年。这 1 亿年与我们生命进化的 36 亿年相比，不过是个小数字。我们银河的年龄在 100 亿年以上，我们的太阳只是第三代恒星，因此比我们太阳早诞生 1 亿年的恒星肯定有的是。事实是第二代恒星要比太阳早几十亿年。假如外星类地球并不希罕，生命的起源和进化顽强而普遍，比地球生命效率高级一些，那么有的外星类地球生命进化就用不了 36 亿年，而是只需例如 32 亿年。假如这样的外星地球又比我们早诞生 4 亿年，那么它的文明就足足比我们早了 8 亿年！这个时间差足以让它们的文明扩展到全银河。那么为什么没有外星文明来访问我们？因此费米这个问题问得很好，问得很科学。

当然否定的答案可以很多。这里五花八门，我们仅列举一些以搏一笑。1) 任何文明发展都过不了能源、环境、核战争的道道难关，早早灭亡了。此

回答叫“文明短命论”。2) 移民外星困难太大，根本无法实现。这有一定的可能，裂变核能难以飞船化，有开采意义的裂变核矿源毕竟很有限，且开矿和核废料对环境危害重大。受控聚变核能由于技术难点也许永远是个梦想。此回答叫“技术有限论”。3) 地球过于特殊来之不易而罕见，生命的长期进化条件苛刻，不要指望有比我们更早的文明。这相当有可能的，这说明有人动不动幻想 UFO 外星人访问地球毫无根据。此回答叫“地球生命最高级论”。4) 其实外星人早就在访问地球了，人类不知道它们的到访是因为外星人不愿打搅人类的自然状态，它们用高技术把自己极好地隐藏起来，在一旁观察人类的内斗，以此作个典型生物实验。这叫“动物园理论”，我们地球只是它们进行观察的动物园。甚至有人幻想，我们的火箭燃烧的是化学燃料，外星人燃烧的是反物质。外星人用的是中微子通讯、引力波通讯，人类用的是无线电或光波通讯，比别人差了几大档次。我们远未掌握中微子或引力波，所以你根本发觉不到它们的存在。必须指出，外星人想愚弄人类毕竟没有人愚弄蚂蚁那么容易。人已经很了不起了，不要自卑地把自己看作蚂蚁。人类已对物理世界有深刻认识，这蚂蚁怎么比得了。人类办不到的事情，外星人想办到恐怕也不容易，想绝对逃脱人类的探测可能更困难。个别人愿当傻瓜，全人类绝不愿当傻瓜。外星人没来过地球也许还说得过去，来过不被我们发现的可能性就比较小了。5) 外星人已经过来了，也有古今证据，但就是政府故意隐瞒和死板的科学家不愿承认。但是政府和科学家一致作弊可能性很小。这叫“隐瞒论”。许多人在疯狂收集外星人证据，遗憾的是他们的证据那么脆弱。6) 地球生命是外星高级生命播的种，叫“播种论”。外星生命亲自来地球有困难，但它采用高技术播种生命的较原始形式确实可以大大减轻技术难度。总不能说，外星生命自己的生命系统最高效，却要另外制造一套低效的生命系统用以播种吧。那么至少也说明，要播的这个种肯定是先进的，于是地球生命也是先进的，进化时标也是最短的。但地球有真细菌、古细菌和真核生物三大界生命，似乎地球生命不像是播种的而是原生的。

这里 2) 与 4) 中我们看到对未来技术的截然相反的两种估计。2) 觉得人类技术尤其是能源技术已经接近最高极限，而 4) 则认为未来人类的技术能

力将远远高出目前的水平的千万倍，外星人或未来人类将无所不能。人们把以前对神仙的向往转移到外星人的身上，大千世界无奇不有。但是不要认为任何幻想都迟早能变成现实。点石成金曾是人们的幻想。后来从物理上知道了，只要把汞核减少一个质子，就真的可以把汞变成黄金。但是，再后来发现，这么改变核内质子数去制备黄金，成本远远高于淘淘金砂矿，人们于是再也不想点石成金了。物理原理上能办到的事情，技术上未必办得到。受控热核的确是个烫手山芋，它要解决了，人类的确就获得了取之不尽用之不竭的能源，外星人也才可以来去自由。此技术何等重要，其原理非常明确，遗憾的是虽然时时在声称取得进展，但突破却遥遥无期。五十年代时两大阵营都以为受控热核技术马上就会获得突破，曾严加保密激烈竞争。但是时间已过去快六十年，两大阵营早已消失，技术已进步到何等程度，全球科学家即使现在携手攻关，也仍未奏效。这是人们未能料到的硬骨头。没有外星人的到访，也许暗示受控热核对外星高级文明也是个难题。对费米悖论解答的种种猜测是不会有答案的，科学没有发展到这样的程度足以肯定回答外星文明问题。

我更愿接受一个比较理性的猜测：生命的起源和文明的出现具有最大限度的艰难性和漫长性。我们的出现已经很不容易，比我们出现得更早就几乎不可能了。如果有比我们文明更早的文明，那么那个高级文明肯定也不满意自己就是最高级的文明，必定还要怀疑是否有比它还要更高级文明的存在。那么何处是个头？总不能宇宙刚一诞生，高级文明就已诞生？一定有个宇宙年龄，第一文明在这个时段诞生。那么第一文明诞生的宇宙年龄是什么？它只能往后推，推到不能再推为止。那么这个“不能再往后推的年龄”就是我们宇宙的现在年龄，137亿年。比它早了，不符合“最大艰难漫长性猜测”，且会出费米悖论；比它晚了与事实不符。简单地说法是我们宇宙的头一批文明。虽然说到生命的起源和文明的出现具有最大限度的艰难性和漫长性，但是从地球生命发展史看，似乎并不如此。从地球生命发展史看，人的出现却是显得那么轻松自如和必然，环环紧扣，这给了我们极大讽刺。仅仅恐龙灭绝给了我们一些遐想。恐龙从3.5亿年前石炭纪初出现到0.67亿年前白垩纪末灭绝，已经统治地球

长达2.8亿年。是那颗尤卡坦小行星偶然撞击墨西哥湾，引起地球生命特别是恐龙大灭绝，才让哺乳动物有崭露头角的机会获得发展最终出现人类。如果这颗小行星晚来1亿年，没有理由怀疑恐龙的能耐差不能多统治地球1亿年，而让其统治期变得可以长达3.8亿年？那么从生命起源到人类出现就不是36亿年，而是37亿年了？如果小行星不来撞地球，那么人就不能及时出现了吗？这里是否有偶然性存在？

我们是宇宙的头一批文明。是否以后宇宙中的文明会不断地涌现呢？也不是。随着宇宙的膨胀，宇宙中恒星诞生的高潮已过。因此宇宙中类地球的诞生会越来越稀少，总有停止诞生文明的时刻。我们不知道文明诞生延续期的时间有多长。但是137亿年这个特殊时间又显得更加突出了。更重要的是，早在宇宙年龄100亿年的时候，我们的宇宙已进入了加速膨胀时期，这正是宇宙大大降低了恒星生成率的主要原因。宇宙安排得太巧了，宇宙诞生了第一文明以后，就安排宇宙基本截断文明的诞生。似乎宇宙的演化就是为这几乎唯一的文明而特意安排的，暗能量数值的安排似乎具有极大的目的性。如果你设想宇宙为我们文明的诞生而安排了大量几乎同时出现的其它文明，那反到是一件不可思议的事情了。如果有大量文明同时出现，说明文明出现并不艰难。既然不艰难，为什么非要集中在一个短时段出现？完全可能分散在一个比较长期的时段中陆续出现，那又违背费米悖论了。于是只能得出结论，这个特殊的宇宙年龄仅仅是为我们这些为数不多的文明的出现特意安排的。暗能量数值的巧合安排，如果倒过来看，其实就是因为我们在宇宙中诞生了，我们才能看到这个宇宙因为有我们存在所特有的参数。这就是所谓“人择原理”，不是人去选择了宇宙的参数，是因为宇宙必须有这样的参数才有人，“人择原理”就是“择而有人原理”，“依人而择原理”，不要从字面上误解了其中的因果关系。

我们来看看宇宙生命安排之巧妙。为什么用氨基酸来构成生命体的主要成分蛋白质？因为 $-COOH$ 是最普通的酸性基，而 $-NH_2$ 是最普通的碱性基，它们的结合才可能是中性的。一个氨基酸的羧基与另一个氨基酸的氨基缩水成为肽键而把两个氨基酸连接起来，这种连接可以继续下去生成多肽。为什么地球生命会只有20种氨基酸？要知道氨

氨基酸的侧链是可以随意更换的，能作侧链的分子多得难以胜数。简单结构的氨基酸绝不下千种，而且左右旋的都有。20 这个数目的安排是因为生命的需要，数目少了将来蛋白质的多样性不够，多了将来生命运行的管理成本太高。蛋白质如果有 100 位氨基酸，每位有 20 种氨基酸，那么总构形量算出高达 10^{130} ，这本身就说明，生命需要从大量的一般是无用的可能构形中选择出有生命意义的活性构形，说明生命建立在复杂性基础之上，因而进化必然是艰难漫长的。为什么蛋白质不能只由 10 种氨基酸构成？要知道形成 10^{130} 个构形也只不过需要 130 位 10 种氨基酸，肽链位数增加只不过 30%，负担似乎并不过分。但是，10 种氨基酸功能不够齐全呀。我们的 20 种氨基酸中，侧链有长有短，有酸性碱性，有亲水疏水，有含与不含硫、氮、氧、苯等原子和基团，氢键有各种安排，等等，将来花样就多得多，可形成极为复杂的二级和三级结构。10 个氨基酸品种花色不全怎么能完成复杂性使命！但是如果有 30 种氨基酸不更好吗？不，30 种氨基酸，编码负担更重，中性进化的富余度太小。需要的 tRNA 类型也更多。有大量蛋白质需要制造，因此需要在细胞中存储更多的 tRNA 成分。对这些更多的 tRNA 的基因管理成本也就更高。品种一多，正确结合的随机碰撞概率就降低。各种因素都可能影响整个生命进化的速度，很不合算。如果生命进化速度慢一倍，那么太阳寿命可能就等不及我们文明的诞生了。太阳的寿命是与生命的进化时标相匹配的。可见，上帝安排或淘汰出一套必要氨基酸时，煞费了一番苦心。20 绝不是个偶然之数，其成员想必各有要害，或许它是使生命进化速度达到最快的方案，真是一个值得研究的课题。

再来看 DNA，它的 4 种碱基 AGCT 的结构其实是很奇妙的，不想不知道，一想吓一跳。我们知道，遗传信息的关键任务是信息的保存和复制。保存和复制信息是一对矛盾。复制信息必须有所拆散，保存信息必须难以拆散。第一步，怎么解决这一分工？共价键和氢键的相似与区别性为遗传提供了绝妙的解决方案。遗传物质内部必须以共价键连接以保证信息的保守性，信息间的复制必须以氢键结合以保证复制的可分拆性。如果模板与复制品也同样用共价键连接，那么当我们把模板与复制品拆开时，且不容易把模板也撕裂破坏了吗？可见氢键成了生

命最关键的角色，没有氢键就没有生命，要在这个高度上认识氢键的至关作用。而水与氢键最相关，这就决定了水在生命中至高无上的地位。第二步，我们要找出复制信息的基本方式，也就是说，复制方式是线性的，还是平面的，还是立体的？平面信息复制我们见得多了，我们的复印机就是平面型信息复制。但是这样的复制方式要让大自然去自动完成，恐怕太难为它老人家了。这很不便管理，成本太高。大自然喜欢简单、节省、可靠、高效、低智商的解决方案。因此只能是线性信息复制，是链状储存信息。那么立体信息复制就不要提了，连人都觉得这种方法麻烦而少用。那么第三步，信息单位是什么？信息单位怎么构成？怎么连接？首先信息单位必须带有“基本部分”、“区别部分”和基本部分间的“连接部分”。没有区别部分哪能储存信息？这个道理很简单。基本部分作为信息基座，必须普及、结实、可靠，大概没有比环戊糖更优越的候选者了。它是碳水化合物 $C_5(H_2O)_5$ ，生命的能源分子，的确普及且用途广泛。它呈环状，结实。为什么不是环丁糖？链太短，首尾环接弯曲不易。为什么不是环己糖？浪费点了，不经济。好，基座环戊糖，对它的六个碳原子编号。如果二号碳原子不脱氧，将来就生成 RNA。如果二号碳原子脱了氧，将来就生成 DNA。这点对于 RNA 和 DNA 性质的区别至为关键，容后叙述。那么第四步，区别部分要用什么类型的分子呢？区别部分就是信息部分，将来一定涉及信息复制，因此必须带有氢键，于是含有氮或氧原子，最好富氮。我们一会儿再探讨其详细结构。那么基座部分之间怎么连接呢？看来不能再用碳碳键、碳氧键或碳氮键去连接了，它们在基座部分和信息部分已经用过了，要避免混淆。如果还用这些键连接基础部分，那么当我们拆散和组装基本单元时，容易破坏基本单元。那么剩下可用的元素只能是磷和硫。硫键不够结实，于是只有磷。幸运的是，磷酸键是高能键，连接遗传单元会出奇效，构成生命能量消费分子 ATP，容后叙述。

第五步是个漫长分析的关键点。我们到底需要多少个不同的信息部分？蛋白质信息编码的位数又应是多少？首先，信息部分的个数必须成双成对，这是信息复制时氢键配对的需要。这时我们来讨论信息部分的种类数目分别是 2、4、6 的情况。先看 2，考察编码位是 4 到 5 时能编码的数目，分别是

$2^4=16$ 和 $2^5=32$ 。可见，编码位为 4 不足以应付 20 个氨基酸的编码任务，只能淘汰。编码位为 5 足以应付 20 个氨基酸的编码任务了，但是 5 位似乎太长了，太不节省遗传物质资源了，要知道每个信息单元也要几十个原子才能装配出来。再看信息部分种类数目 4 情况，考察编码位是 2 到 3 时，能编码的数目分别是 $4^2=16$ 和 $4^3=64$ 。可见，编码位为 2 不足以应付编码任务，只能淘汰。编码位为 3 足以应付 20 个氨基酸的编码任务了，还有 $64/20=3.2$ 的编码富余量，非常适合中性进化。而且比刚才的 5 位减少 40%，节约性相当可观。再看信息部分种类数目 6 情况，考察编码位是 2 时，能编码的数目是 $6^2=32$ ，似乎也能编码 20 种氨基酸，但编码富余量 $32/20=1.6$ 显然比不上刚才的 3.2。似乎在编码位上大有节省，但重要的是，要找出 6 个不同的信息单元居然是件难事。

第六步，我们用什么构形去构造可靠的信息部分？环状结构最结实，成为首选。靠什么去区分信息部分的不同结构？采用不同的环数和不同的氢键数。那么不同环数的配搭应如何取舍？最佳答案其实就是 1:2。因为 1:3 总环数为 4 太浪费，2:3 更浪费不说且区别性反而下降。氢键数的搭配应如何取呢？最好是 2:3。因为 1:2 和 1:3 都有 1 个连接是仅仅 1 个氢键，太易断裂且没有备份，容易造成变异。2:3 在强度上相差不大，考虑到氢键的方向性，整体稳定性均匀。双键连接断开一个还有另一个连着，有保险性。于是最终方案已经出台，氢键搭配 2:3，环数搭配 1:2，一共 4 种可能，所以要找出 6 种信息单元确实不易。四种碱基恰恰具有这种搭配，它们就是 DNA 的 A、G、C、T 或 RNA 的 AGCU。以 A 为例，它恰恰是宇宙空间中最常见的氢氰酸分子 HCN 的五合体，双环双键， $(HCN)_5 \rightarrow H_5C_5N_5=A$ 。由 A 适当变换可以得到其余的 GCTU。这里第二环少个原子，基本格局不变，甚至更为结实。双环双键腺嘌呤 A 与单环双键胸腺嘧啶 T 配，双环三键鸟嘌呤 G 与单环三键胞嘧啶 C 配。由腺嘌呤加三个磷酸构成 ATP，是生命的基础能源分子，能源消费的通货。可见大自然在安排遗传物质时的极为高明之处。也可以推想我们地球的生命结构就是宇宙生命的最优结构，我们就是宇宙首先演化出的文明。

因为 DNA 比 RNA 在第二碳位少个氧，于是 DNA 比 RNA 更紧凑，DNA 长链分子更结实、笔直和难以弯曲，于是更适合承担信息储存的任务。而 RNA 长链分子较易弯曲，可以自行折叠，其中部分碱基内部配对，于是它既可以充当信息传递的角色，又因复杂折叠而形成初步的催化功能，充当某些酶的作用，例如 rRNA。它的多重角色使人们猜想，生命起源的第一系统可能仅有 RNA，称为 RNA 世界。生命起源的第一步或许就是随机演化出自我复制自我催化的某个特定 RNA 分子。当然这同一分子必须处在模板和催化两种状态上才能实现自我催化自我复制，这里又必有玄机。这样的特定原始生命 RNA 的碱基总数是多少？它们的碱基序列是什么？很值得在实验室中找一找。也有一种假说，认为磷酸化氨基酸可以促进寡肽和寡聚核苷酸合成，因而可能与生命起源有关。

这些解释看上去有点“马后炮”。我们在大自然面前，能当“马后炮”已是很荣幸了。难道我们还能证明我们比大自然更聪明？自从有了人类，大自然是把部分发明权交给了人类，但人类能自大的起来吗？在人类生命、人类出现和人力不及的每个自然环节，人类怎敢与大自然比肩。有些自然事物，我们虽然发现了它，但却不一定理解了它。当初元素周期表就是一例。今天，我们已发现但并没有被深刻理解的东西比比皆是。古人就“发现”了太阳发光，直到今天，我们对此虽已有层层深入认识，但是还有更深层次的内因需要我们去理解。DNA、RNA 和蛋白质的结构又是一例。我们只看到大自然从众多的分子中挑选了 28 种作为生命的基础元件，但不知道它为什么要这样挑选，不知道它是不是最优选择。今天的分析就是一次可喜的尝试，这是值得研究的好问题。生命构建在复杂性之上，复杂性构建于多样性之上，多样性决定了进化选择性，选择性决定了进化的艰难性。要实现生命的复杂性，物质世界必须有多层次复杂结构。这必须使物理世界发生巧合性多因素破缺。宇宙创造了如此艰难的生命，这样的宇宙就不是一般的宇宙，宇宙与生命之间的关系就非同寻常，这就是本讲座要强调的主题。

（中国科学院研究生院 100049）