

分类号 Q-3

密级 公开

UDC

编号 10486

武汉大学

硕 士 学 位 论 文

生命起源过程中原始细胞膜替代的
计算机模拟研究

研 究 生 姓 名 梁 瑜 祯

学 号 2019202040156

指 导 教 师
姓 名 、 职 称 马 文 涛 副 教 授

专 业 名 称 生 物 物 理

研 究 方 向 计 算 生 物 学

二〇二二年五月

Classified number Q-3

Secret grade open

U D C

Thesis number 10486

Wuhan University
Dissertation for master's Degree

**Computer Simulation of Protocells' Membrane
-Takeover During the Origin of Life**

Student Name : Yuzhen Liang
Supervisor : Associate Professor Wentao Ma
Name of Speciality : Biological Physics
Research Field : Computational Biology

May, 2022

生命起源过程中原始细胞膜替代的 计算机模拟研究

硕士研究生：梁瑜祯

导师：马文涛 副教授

(武汉大学生命科学学院，武汉，430072)

摘 要

细胞膜是组成原始细胞的关键部分，膜提供“隔间”将遗传物质和代谢机制紧密联系起来，构成生命最基本的单元。现存所有生物的细胞膜都是以磷脂作为主要成分的。目前主流的生命起源研究认为最原始的细胞膜可能主要是由存在于前生命环境中的脂肪酸构成的，而在原始细胞与现代细胞之间有一个过渡态。这个过渡态细胞的膜是由磷脂和脂肪酸组成的简单混合膜。由于生命起源仍然是一个缺乏化石证据的高度推测的领域，人们对前生命化学的了解和对生命本质的基本认识不足。在这种情况下，计算机模拟有望发挥重要作用，构建原始细胞模型有可能帮助我们进一步理解生命起源中的这种演化过程。本文介绍了用计算机模拟研究原始细胞的膜中两亲性分子从脂肪酸到磷脂的演化过程。模拟使用了基于概率蒙特卡洛模型的原始细胞模型模拟矿物表面上的 RNA 世界进化，通过基于确定规则的相对概率来模拟原始地球环境中的随机事件，以在计算机中实现生命起源这一罕见事件的发生及之后原始生命的演化。矿物表面通常被认为是生命起源的关键过程出现的场所，因为它有聚合、浓缩、组织和保护生物聚合物的作用，是模拟原始环境的第一选择。模拟结果显示磷脂合成核酶对系统中原始细胞的平均大小有显著影响。同时想要渗透代偿核酶代表的代偿机制获得达尔文进化的优势，需要磷脂合成核酶和渗透抑制机制同时存在，缺一不可，这代表磷脂合成酶大概率出现在渗透代偿酶之前。同时拥有核苷酸合成核酶、磷脂合成酶和渗透代偿酶的并以磷脂作为细胞膜主要成分的原始细胞在系统中被选择，成为优势种群，即达尔文进化的胜利者。

关键字：RNA 世界，生命起源，原始细胞，膜替代，元胞自动机

Computer Simulation of Protocells' Membrane -Takeover During the Origin of Life

Master Candidate: Yuzhen Liang

Supervisor: associate professor Wentao Ma

(College of Life Science, Wuhan University, Wuhan, 430072)

Abstract

The cell membrane is a key part of protocells. The membrane provides a "compartment" that closely links genetic material and metabolic machinery, and constitutes the most basic unit of life. The cell membranes of all living organisms are composed of phospholipids. The current mainstream research on the origin of life believes that the most protocells' membranes may be formed by single-chain amphiphilic molecules existing in prebiotic environments, while in protocells and modern cells There is a transition state in between, the main feature is that the membrane of transition state cells is a simple mixed membrane composed of phospholipids and fatty acids. As the origin of life remains a highly speculative field lacking fossil evidence, human knowledge of prebiotic chemistry and a basic understanding of the nature of life is insufficient. In this context, computer simulations are expected to play an important role, and the construction of protocells models has the potential to help us further understand the origin of life in the laboratory. This paper presents experiments to simulate in silico the evolution of protocells' membrane amphiphiles from fatty acids to phospholipids. This paper describes the use of computer simulations to study the evolution of amphiphilic molecules from fatty acids to phospholipids in the membranes of primitive cells. The simulation used a probabilistic Monte Carlo model-based protocells' model to simulate the evolution of the RNA world on mineral surfaces, by simulating random events in the primitive Earth environment by relative probabilities based on deterministic rules to achieve the rare event of the origin of life in a computer. Mineral surfaces are often considered to be the site of key processes in the origin of life, as they aggregate, concentrate, organize, and protect

biopolymers, and are the first choice for simulating pristine environments. Simulation results show that phospholipid synthesis ribozymes have a significant effect on the average size of protocells in the system. At the same time, if the compensatory mechanism represented by osmotic compensatory ribozymes is to gain the advantages of Darwinian evolution, both phospholipid synthesis ribozymes and osmotic inhibition mechanisms are required to exist at the same time, which means that phospholipid synthases have a high probability of appearing before osmotic compensatory enzymes. Protocells possessing nucleotide synthesis ribozymes, phospholipid synthases and osmotic compensatory enzymes at the same time and using phospholipids as the main components of the cell membrane were selected in the system and became the dominant populations, namely the winners of Darwinian evolution.

Key Words: RNA world, origin of life, protocells, membrane-takeover, cellular automata

目 录

学位论文原创性声明.....	I
摘 要.....	I
ABSTRACT	II
第一章 背景	1
1.1 RNA 世界的猜想.....	1
1.2 RNA 世界原始生命的存在形式	2
1.2.1 原始生命起源于生物大分子	2
1.2.2 原始细胞结构的猜想.....	4
1.3 RNA 世界原始细胞的膜成分	7
1.4 脂肪酸作为原始膜成分的可能性	9
1.5 原始细胞的膜两亲性分子从脂肪酸到磷脂的演变	10
1.5.1 磷脂膜的特性.....	10
1.5.2 脂肪酸-磷脂混合膜存在的可能.....	11
1.6 原始细胞的计算机模拟进展.....	13
第二章 研究方法.....	16
2.1 模型模拟的对象.....	16
2.2 模型的基本概念.....	17
2.3 模型使用的参数及其定义.....	19
2.4 模拟流程	26
第三章 实验结果.....	28
3.1 磷脂在系统中的出现	28
3.2 磷脂合成酶在系统中的出现.....	29
3.3 磷脂的渗透抑制效应在系统中的出现	32
3.4 代偿酶 MPR 在系统中的出现	33
第四章 讨论	35
4.1 模拟结果的综合分析	35
4.2 模拟中意料之外的结果	38
4.3 模拟的前提条件与模型的局限性	38
4.4 后续模拟研究的展望	39
参考文献.....	41

攻读硕士学位期间已发表的学术论文.....	43
附 录.....	44
致 谢.....	48

第一章 背景

1.1 RNA 世界的猜想

地球上生命的出现因其神秘性与根源性对科学家一直是最具有吸引力的事件之一。关于最早期的生命世界,研究者们提出了“RNA 世界”的猜想。RNA 世界学说认为地球上生命起源的最早时期是 RNA 占据主导地位的时期,这个时期中 RNA 分子同时作为功能物质和遗传物质,在生命活动中起到核心作用。而最早的生命始于可以自我复制的 RNA 分子,这些分子可以发生达尔文进化,可以进行新陈代谢甚至可以用于催化环境中的原材料合成核苷酸和膜结构分子。RNA 可以作为核酶催化化学反应的猜想已经部分获得了实验证明。可以说,以 RNA 为功能物质和信息载体的生命出现在生命起源时期的可能性非常高, RNA 世界假说正在得到坚实的实验支持。^[1,2]

关于以 RNA 为功能物质和信息载体并以脂肪酸双层膜作为物理边界的生命体形态已经有大量的计算机模拟证明其存在的可能性。^[3,4]因为细胞是公认的生命最基本单元,所以分子的前生命合成、矿物表面提供的环境以及核酶的出现只是生命起源之谜最开始的一部分,只有当生命体各部分包括且不限于遗传物质、物理边界、代谢物质组合起来相互作用,形成一个完整的具有达尔文进化潜力的系统时,才能认为我们已经发现了原始生命完整的形态。不管是哪一种猜想,膜的出现都被认为是生命起源中必不可少的第一步,甚至有研究认为膜的出现早于自我复制的 RNA 的出现。现存所有生物的细胞膜都是由磷脂组成的,目前主流的生命起源研究认为最原始的细胞膜可能是由存在于前生命环境中的脂肪酸自发组成的,而在原始细胞与现代细胞之间有一个过渡态时期。过渡态时期细胞的主要特征是细胞膜不是主要由磷脂组成的复杂双层膜,而是由磷脂和脂肪酸一起组成的简单混合膜。

RNA 世界的存在是科学家通过推理得出的,想要进一步证明它的存在,还需要更多的实验证据或者化石证据。然而,相关的实验研究证明构建一个完整的原始细胞是极其困难的,在实验室中展示类似过程的尝试距离成功依然非常遥远。所以生命起源仍然是一个缺乏化石证据的高度推测的领域,人们对前生命化学的了解和对生命本质的基本认识不足。并且生命的起源不仅是一个化学问题,还是一个演化问题。起源过程中任意一个不起眼的时间段在历史上真实消耗的时间都

是不可计量的，人们只能通过极少量的地质化石来对这个时间段发生的事件进行推理，毕竟人类文明经历的时间相较于生命起源经历的时间几乎是可以完全忽略的，任意一个演化事件的发生都可能消耗成百上千乃至过万年的时间，这使得研究者通过化学方法在实验室模拟起源过程极度困难。幸运的是，随着计算机技术的飞速发展和计算机算力的快速增长，我们可以通过计算机模拟来推演进化方面的问题并为化学实验提供帮助。

1.2 RNA 世界原始生命的存在形式

直至今天，关于对生命的定义依然有着极大的争议。有人认为只要可以自我复制就可以被称之为生命；另一些人认为一个严格定义上的“生命”除了自我复制还必须有达尔文进化的潜力，这些人里包括正在以非自然的方式重新排列自然基因，以尝试形成一个基因完全来自其他生命体却依然能够进行达尔文进化的人工细胞的合成生物学家们；还有一部分人选择关注新陈代谢。这些对生命定义的讨论常常使鉴定一部分特殊存在是否属于生命变得十分困难：病毒算生命吗？如果病毒算生命那寄生在病毒上的卫星病毒也算吗？朊病毒呢？很明显，“生命”不会是一个绝对的概念，在非生命与生命之间必然有一个很大的模糊地带，生命的起源就在这个地带发生。

1.2.1 原始生命起源于生物大分子

早在十七世纪，人们就有了猜想，认为生物体是由无生命物质逐渐转变来的。1929 年，英国生物学家 John Burdon Sanderson Haldane 发表了关于地球生命起源的假设，描述了原始地球上非生命物质逐步演化和生命出现的场景。Haldane 写道：“细胞由许多悬浮在水中并包裹在油膜中的半生命化学分子组成。当整个海洋是一个巨大的化学实验室时，这种薄膜的形成条件一定是相对有利的……” Goldacre 在 1958 年提出，第一个薄膜可能是由类脂表面活性剂的波浪作用干扰薄膜产生的。^[5] 而 Oparin 则将生命描述为一个多分子、多功能的代谢系统。Oparin 那个时候还不知道细胞具有膜边界，因此在他假设细胞是称为凝聚层的一个聚合聚集体。在那个时候流行的观点是细胞没有被膜包围。Haldane 和 Troland 的遗传理论则认为早期生命的特征是具有自我繁殖能力。伴随分子生物学的发展，人们对生命的认识已经发展到了生物大分子层面，对生命的定义引发了全新的问题：即几种生物大分子出现的先后顺序是什么？其后 RNA 世界理论出现，具有自我繁殖能力的

RNA 的发现更使得 RNA 世界成为生命起源领域的主流假设。RNA 世界理论认为达尔文进化可能始于分子水平，因为分子是自然的、简单的独立实体，可能会受到自然选择的影响。

Haldane 和 Goldacre 同时提出生命起源于有机化合物合成，他们认为简单的无机化学物质最初可以在一个原始海洋中合成，然后经历逐渐化学演化成为更加复杂的有机大分子，不需要生命的代谢过程。这一假设后来得到了著名的 Miller-Urey 实验的支持。1951 年秋天，芝加哥大学年轻的博士生 Stanley Miller 受诺贝尔奖获得者物理学家和化学家 Harold Urey 的关于可能存在于原始地球上的物理和化学条件，以及它们与生命出现的潜在相关性的演讲所吸引和启发，Miller 说服 Urey 让他尝试通过实验模拟这些条件。Miller 在他与 Urey 设计的玻璃装置中模拟了原始大气和海洋，在 1953 年春首次通过实验证明了模拟装置中有机分子的合成，主要是蛋白质的组成单体氨基酸。该实验表明氨基酸可以从水和诸如甲烷、氨和氢等无机物中合成，标志着前生命化学领域的开拓，此后脂肪酸也发现了非生物合成的途径，更复杂的分子包括核苷酸和磷脂的非生物合成也已得到部分证实。这些发现很快与生物化学和细胞生物学的发展相融合，并引发了将原生质的起源等同于生命起源的讨论。^[6]

之后，俄罗斯生物化学家 Alexander I. Oparin 提出将对第一个细胞出现的研究转变为可行的多学科研究计划。而且，随着分子生物学的发展，Troland、Muller 等研究者提出在分子水平上理解生物现象，单个分子或病毒代表原始生命系统。随着简单有机化合物被证明很容易在假定的原始条件下合成，在 Muller 实验之后的几年里，对 DNA 复制和蛋白质生物的合成研究加上 RNA 分子的催化多功能性的发现，导致了许多研究者对 Muller 的观点进行批判性评价。批评者认为生命并不一定需要 DNA 和蛋白质，这两类大分子出现大概率晚于 RNA。现存的生态圈可以被称作是广义的“RNA 世界”。除了极特殊且不能进化所以还不被认可为生命的朊病毒之外，几乎所有的生命都有 RNA。RNA 几乎参与了当代细胞除支撑形态以外所有的关键生命活动，不但参加编码和合成蛋白质，还能协助 DNA 复制，并且还可以逆向传递信息。从上个世纪中期发现活细胞比 DNA 含有更多的 RNA 时开始，就有了 RNA 在进化过程中先于 DNA 出现的猜想。蛋白质合成细节的发现揭示了大量的 RNA 分子参与了当代细胞内的各种生命活动，这促使人们猜测 RNA 既可以作为催化剂，也可以作为信息携带分子，在早期生命中有着不可替代的作用。RNA 本身的结构也表明其起源极大概率早于 DNA 和蛋白质。当然核酶的发现并不意味着前生命环境中被认为用作原始基因的自催化核酸分子漂浮在原始海

洋中，完全由原始环境中存在的简单前体组装而成。自核酶发现之后，研究者们普遍假设 RNA 世界在由 DNA 加 RNA 加蛋白质组成的生命出现之前存在，这在如今已经是生命起源的元命题。

1.2.2 原始细胞结构的猜想

支持原始地球上存在多种有机分子（包括成膜化合物）的证据表明，膜结合分子系统的进化先于地球上的细胞生命。^[7]这些前生命分子构成的第一批“细胞”会是什么样的形态？对此也有很多假说。自我复制大分子假说认为最早的生命是无膜的；“胶体”假说认为最早的细胞有区域分隔但是没有现代意义的脂质双层膜；原生质层假说和进阶的“脂肪酸世界”假说认为脂肪酸膜是原始细胞的支柱性组成成分。自我复制大分子假说假设了一种同时具备催化功能和复制模板功能以实现自我复制的大分子，这种假设在核酶发现之后被更广泛地传播。论点大多认为能够催化其自身复制的模板依赖性 RNA 合成酶核酶应该首先出现。这种核酶在进化中，生态位逐渐被 RNA 复制酶代替，它的出现可能代表了生命起源中的一个重要事件，标志着达尔文进化论的开始。现代生命新陈代谢活动的各个方面，从 RNA 的广泛存在到 RNA 在遗传信息翻译和 DNA 表达的中心作用，都有足够多令人信服的证据支持这样一个假设——即在几百亿年前，初始的生命使用 RNA 来存储遗传信息和新陈代谢。与 RNA 相比，其他生命大分子作用都更加单一，包括且不限于 DNA 和蛋白质。虽然功能单一意味着它们在一个方面性能较好，但也意味着它们对于原始生命来说功能不够全面。DNA 几乎只作为遗传物质发挥作用，它的稳定性有利于满足遗传信息传递的高保真要求，却也使它失去了结合金属离子发挥催化作用的潜力，而蛋白质在现代细胞中也只能发挥酶和结构的功能，不能作为遗传物质。有且仅有 RNA 同时具备“催化”和“遗传信息载体”这两种能力。所以长期以来，虽然现存的生命体中并没有发现原始自我复制酶存在的痕迹，但人们一直在努力通过化学合成实验和体外进化实验构建这种想象中的模板依赖性 RNA 合成酶核酶，然而 RNA 催化的 RNA 复制是一个非常复杂且难以控制的化学过程，整个反应包括多个细分步骤：按照 RNA 模板进行准确的底物选择、延伸、易位，然后迭代循环，现有的聚合酶核酶缺乏这种合成接近自身大小的 RNA 的能力。不过在科学家们的不懈努力下，James Attwater, Aniela Wochner, 和 Philipp Holliger 于 2014 年宣布发现了一种人工核酶，它能够催化一些 RNA 模板的复制，它的自身长度大约为 200nt^[8]。这种催化剂可以直接在体外环境甚至低温环境下的稳定介质中进化。尤其值得一提的是，它适应了零摄氏度以下这种在现代人看来

已经是极端的温度，甚至能够在低至负十九摄氏度的温度下在冰中合成 RNA。这种第一次发现的 RNA 聚合酶核酶可以在保证非常高的精确度的情况下，催化合成比自身更长的最多达到 206 个核苷酸的 RNA 序列。这是一个非常有意义的发现，很大程度上证实了 RNA 自我复制的化学潜力。通过低温适应性突变与上述在环境温度下操作的 5'末端结合，可以设计可能的有自我更新自我复制潜力的 RNA。此外，研究表明，在某些条件下，包括冰冻—融化循环，这种 RNA 聚合酶核酶可能由长度约为 30nt 的简单 RNA 寡聚体组装而成。我们完全可以假设这样的 RNA 复制酶在化学上是具有自我复制功能的并且通过非酶催化随机地与其他 RNA 分子出现在早期的 RNA 世界中，实现共同繁荣。共同繁荣的前提是忽略核酶的“寄生虫”问题，因为这类核酶不仅仅能促进自我复制，很有可能对其他 RNA 分子也有同样的助力，相当于其他分子寄生在了核酶上。因此单独的生命大分子起源假说逐渐被建立在这种核酶基础上的脂肪酸世界起源假说替代，因为脂肪酸囊泡提供了一个环境，帮助核酶与寄生分子分开。

在没有膜作为边界的情况下，细胞质内的遗传物质和能量物质会溢出到环境中并分散，而我们知道：在真实情况下细胞内这两类物质的浓度远远高于细胞外，所以膜的必要性是显而易见的。Haldane 提出了膜在生命起源中发挥作用的第一个猜想，他写道：“细胞由许多悬浮在水中并包裹在油膜中的半生命化学分子组成。当整个海洋是一个巨大的化学实验室时对这种薄膜的形成条件一定是相对有利的。该细胞由内环境和将内环境和外环境分割开的边界组成，内环境内相对外界更稳定，可能发生生命前化学反应，边界划定个体系统、限制细胞的表面渗透性并控制细胞与环境发生的化学物质交换。”^[5]Oparin 提出的生命起源情景则是根据生物化学领域的发展来构想的。在 20 世纪初，生物化学家主要关注酶的分离和研究，因此他们倾向于将生命视为“催化反应系统的自我调节动态平衡”。受限于当时使用的光学显微镜的能力，许多细胞成分被视为胶体凝胶：一种悬浮在液体中的聚合物小滴，而酶被当做“胶体催化剂”。作为一名生物化学家，Oparin 在他长期担任莫斯科著名的巴赫生物化学研究所所长期间将他的实验工作集中在酶学上。他认为由酶促反应主导的新陈代谢是生命的决定性特征，正是通过新陈代谢，细胞才能适应不断变化的环境。基于胶体化学，Oparin 提出原始海洋中的有机聚合物，例如蛋白质和糖，形成了一种胶体溶液，一旦这种胶体溶液达到一定浓度，封闭的聚合物形成的微小液滴就会从它们的水环境中分离出来。这些所谓的“凝结物”或“凝聚层”可以选择性地从溶液中吸收物质，进行原始酶促反应，除此之外还可以长大并最终分裂，内容物从父母转移到后代负责代谢能力的内部组织中。这样

的结构被 Oparin 视为可能的“原始细胞”。这种凝聚体形式的“原始细胞”通过原始分裂和遗传形成了原始的自然选择过程的基础，借此演化为更复杂和更有效的形态。因为 Oparin 开始假设的情景是异养的，所以他继续尝试解释自养生物的进一步进化，假设这些生物能够通过光合作用产生自己的有机物质。随着 1950 年代和 1960 年代分子生物学的建立，Oparin 还讨论了遗传装置的进化，声称这种进化可能只发生在可以代谢的原始细胞结构内。^[9]

伴随对细胞脂质膜的了解增加，不可避免地出现了原始细胞也具有基于脂质的包膜的想法，当然这种膜会比现代细胞简单得多，不一定具有蛋白质或磷脂。Herrera 通过对不同种类的物质进行实验，试图说明生命是物理化学的副产品。起初，Herrera 使用水和油的混合物来了解细胞状结构的形状、大小和运动，并在后来发展了他的“等离子体发生”理论，尝试解释原始光合作用原生质体的起源。1970 年代早期发现不饱和脂肪酸可以通过自组装成双层膜形成囊泡，这表明原始细胞也可能是由这种简单的单链两亲性组装而成的囊泡构成的。后来发现具有不同烷基链长度的脂肪酸也可以自组装成囊泡。在对岩石的地质分析中发现了生命起源早期存在的各类简单脂肪酸，这更加提高了脂肪酸囊泡作为原始细胞的一部分的可能性。

在没有模板指导的信息复制的情况下，第一个自我复制的核酶如何出现呢？为了让 RNA 世界存在，必须有一种能够催化其自身复制的核酶。Troland 试图通过物理化学手段来解释通过选择性催化和调节来实现自我繁殖和自我组织的生命的出现，为此他提出了一个抽象模型。模型专注于原始海洋中出现的具有两种显著催化能力的原始分子，这种原始酶分子应该既具有自催化作用，即催化其自身的形成，又具有催化膜分子合成的作用，催化其自身周围形成包膜。由于当时缺乏关于酶作用机制和遗传物质性质的知识，Troland 首先谈到了一种“酶”，然后谈到了一种“遗传酶”，将自催化和催化其他分子合成的能力归因于他的“遗传酶”。然后 Troland 将基因和酶结合在同一个分子中的概念称作“核酶”。“核酶”的概念从提出到被发现只经历了二十年，这个发现构成了 RNA 世界假说的基石，被大多数生命起源科学家视为生命出现的关键阶段^[10]。与当今 RNA 世界的支持者一样，Troland 将他的“遗传酶”的自催化方面视为其关键功能，并认为第一种基因酶的形成是一种极低概率事件，也就是说发生的时间很长，然而再低的概率在以百万年为单位计时的生命起源事件面前都绝非不可能，甚至可以说是必然事件。此外，他声称遗传酶的自我繁殖也为自然选择的突变和进化提供了基础。^[6]人们已经在体外发现了一系列性能不断提高的聚合酶核酶，尽管这些发现中没有一个可以说是

复杂系统，但这些自我复制核酶可以仅使用可能在地球上找到的无机化学资源来维持自身。毫无疑问，我们可以在研究中假定核酶是早期生命的组成成分。

1.3 RNA 世界原始细胞的膜成分

细胞形式生命的出现一直是生命起源领域的一项重大挑战，而膜又是维持细胞形态的关键部分。生命可能开始于受矿物表面限制的二维系统，其个体间的分离是通过封闭的脂质膜实现的。膜可以确保细胞的个体性和完整性，并介导它们与周围环境之间的相互作用。膜结合水环境中不同生化反应的划分，是生命自我维持的关键；膜组成的囊泡的分裂，是生命自我复制的关键。关于原始细胞从简单的自组装分子系统,缓慢演变为能够自我维持和自我复制的细胞实体的过程有很多猜想，但是不管哪种猜想，膜的出现都是最为关键的一步。膜提供“隔间”将遗传物质和代谢机制紧密联系，为生命提供了包括构造合适的代谢环境在内的多种便利，构成生命最基本的单元，是组成原始细胞的关键部分。大多数关于生命起源和早期进化的研究都集中在遗传系统的起源或代谢的出现上。先复制？还是先代谢？膜看起来对这两种机制都不一定是必须的，所以尽管膜既可以提供代谢和遗传物质存在的空间，又可以在遗传和代谢系统相互作用和共同进化方面发挥着至关重要的作用，膜的起源和进化受到的关注还是远少于遗传物质、能量和代谢的起源。不过在观测到具备分裂这种自我复制机制的脂肪酸囊泡之后，研究者们对膜的研究兴趣大大增加了。

膜结合水环境中不同生化反应的划分被认为是任何生命起源途径中必不可少的第一步。Oparin 首次提出生命起源于早期地球的有机材料自组装结构的概念。即“化学进化作为向生命的过渡”的概念。那时膜作为边界结构的作用尚未确立，相反，人们认为活细胞可以理解为聚集的胶体颗粒的集合。因此，Oparin 和他的同事从大分子成分，如阿拉伯树胶、明胶和组蛋白中制备了异质的球形聚集体，这些聚集体可以为酶促反应提供局部位点。Oparin 还讨论了他在工作中研究带相反电荷的离子之间的复杂凝聚产生的无膜液体微滴。从那时起，这种关联的液-液相分离已扩展到许多合成和天然聚合物，并且开始探索所得微滴作为原始细胞模型的可行性。重要的是，尽管它们缺乏物理边界，但凝聚液体微滴能够通过平衡分配自发地积累许多溶质。这种特殊的性质类似于囊泡隔室的封装和选择渗透性，可以解决早期地球上生物大分子前体可能被稀释导致原材料浓度低于反应所需浓度而代谢被阻断的问题。然而这些称为“凝聚体”的胶体聚集体本质上并不稳定，

它们的分子组成不稳定、没有渗透屏障、缺乏封装代谢和生物合成产物积累的能力。

随着人们更多地了解膜在细胞结构中的作用，一个现象被发现——很明显所有细胞都有脂质双层并将其作为主要的渗透屏障。磷脂是现代细胞双层膜中几乎无处不在的两亲性分子成分，并且磷脂可以自发组装形成大小接近细菌细胞的双层囊泡。脂质双层囊泡通常被称为脂质体，这种自组装的膜结构可以用作最早的细胞膜模型。人们需要解决的第一个问题涉及早期地球上可用的类脂质化合物的来源。一种可能性是磷脂是在前生命化学环境进化过程中合成的，早有实验表明磷脂可以在模拟的前生命化学条件下由脂肪酸、甘油和磷酸盐的混合物合成。然而，早期地球上同时存在这三种成分是高度推测性的，因此我们将注意力转向更简单的可成膜两亲性分子。有人提出，基于脂肪酸的囊泡可以被认为是一种模型原始细胞，具有通过达尔文进化而变为更复杂细胞的潜力。目前所有生物的细胞膜都是由磷脂组成的，细菌和真核细胞膜由二酰基磷脂组成，其中两个将一个极性头部基团和两个疏水尾部连接，古细胞的酯键被替换为醚键，但是组成部分依然是一致的。但早在二十世纪七十年代，人们就认识到磷脂可能不是早期细胞膜的最合适的组成部分，这主要是因为磷脂膜对极性和带电小分子的渗透性低。这种特性将严重限制原始细胞通过非生化过程吸收外部环境中合成的营养物质的能力。^[11]此外，目前的考古证据也未发现原始环境中磷脂的存在，人们推测原因和磷脂的合成材料有关。磷脂的合成需要高度氧化的磷酸盐，磷酸钙在水中是不溶的。在整个地球的历史上，几乎所有的磷元素都以不活泼的形式存在于矿物中，很难想象矿物中的无机磷酸盐能提供足够多的磷酸为原始生命提供进化所需的原材料。这些证据证明磷脂原料中的磷酸甘油不太可能在原始环境中有很高的丰度。**Murchison** 陨石中磷酸的发现表明，这些水溶性含磷分子有可能通过陨石或彗星运送到早期地球，这可能为化学演化的最初阶段提供了有机磷的供应。^[12,13]因此，最初的细胞膜被认为是由环境中合成的简单单链脂质即短链脂肪酸及其衍生物自组装而成。脂肪酸是磷脂的组成部分，相对于磷脂分子量更小、结构更简单，也就是说脂肪酸更有可能通过非生物合成获得。脂肪酸可以在适当的条件下自组装成双层膜，其中最重要的条件是浓度和 pH。^[14]另外一种猜想认为原始细胞膜的组成可能从其最早开始就是混合的而非单一的，因为磷脂前生命化学合成是可能存在的，并且脂肪酸可以作为磷脂的原材料，那么纯粹的脂肪酸膜可能并不存在，脂肪酸—磷脂混合膜在生命的最初就已经是主流了。这种猜想也是本文模拟的基础。

1.4 脂肪酸作为原始膜成分的可能性

主流观点认为脂肪酸和脂肪酸衍生物是早期的膜成分，主要原因有二，一是脂肪酸有实物证据和实验证据证明其在地球早期环境中存在；二是脂肪酸有可能形成类似现代细胞的双层膜结构。脂肪酸在一定条件下可以自发组成将亲水基团暴露于其双层边界的外部 and 内部的囊泡。这种脂肪酸囊泡能够在内部形成一个包含一定体积溶液与双层膜外环境区分开来的小环境，并且可以自发吸收游离脂肪酸，生长并分裂成两个结构一致的脂肪酸囊泡。

在模拟的前生命条件下通过 Fischer-Tropsch 型反应可以合成脂肪酸，对碳质陨石的研究也为脂肪酸在前生命条件下的存在提供了证据。这些研究表明使用在生命出现前地球上可用的分子进行非生物合成可以产生大量的脂肪酸，在特定环境下其浓度有可能达到形成囊泡的临界点。1973 年首次报道在实验室观察到脂肪酸可以形成结构和性能类似于由二酰基磷脂组成膜的双层膜结构囊泡，这里的脂肪酸指带有饱和或不饱和的长脂肪链的羧酸，脂肪酸囊泡指由封闭的脂肪酸双层形成的球形隔间，将膜内与环境分隔开来。^[15,16]从结构的角度来看，由脂肪酸和磷脂组成的膜的物化性质是相似的。陨石的有机提取物含有非生物合成的两性亲分子混合物，可以在水溶液中自组装成脂肪酸囊泡，与其他两亲性分子的混合物同样可以根据其周围环境自组装成不同的囊泡结构。^[17]虽然脂肪酸囊泡的不稳定性是这些单链两亲性分子的固有化学性质，但封装和有限的渗透性也是这一大类可成膜的分子形成的囊泡的一般特征。引人注目的是，两亲性分子也可以经历相分离过程，称为混浊现象，产生能够自发积累水溶性和有机化合物的无膜微滴。最近的研究报道显示，脂肪酸也可以在特定的抗衡离子存在下相分离成凝聚液滴，甚至通过 pH 的变化可逆地重组为囊泡。不过这些新的脂肪酸凝聚层能否可以替代原始细胞的细胞膜这点还不得而知。

由于脂肪酸的热力学特性，它们不断地在膜的内叶和外叶之间翻转。脂肪酸翻转可能在一些小分子进入囊泡的渗透性中起重要作用。在跨膜运输机制进步之前，脂肪酸囊泡的这种高渗透性在原始细胞中可能是必不可少的。由脂肪酸组成的最原始的原始细胞膜，最初是高渗透性的，随着它们的膜成分向磷脂含量增加的方向发展，它们的离子渗透性最终会降低。例如这些混合脂肪酸-磷脂膜选择性地保留 K^+ ，但允许 Na^+ 通过细胞外。这些混合脂肪酸-磷脂半透膜的 K^+/Na^+ 选择性表明，处于进化中间阶段的原始细胞在没有任何大分子运输机制或泵的情况下可能已经获得了 K^+/Na^+ 离子梯度。这可能促进基本的代谢，也可能对早期细胞有害。^[18]原始细胞中存在高膜通透性意味着它们可以防止由渗透压引起的潜在损害，但不

能保留和积累其内部原始代谢产生的小代谢物。原始细胞中的脂质膜从脂肪酸膜演变为混合脂肪酸-磷脂并最终演变为富含磷脂的膜——这可能是由环境压力驱动的，例如环境中 Mg^{2+} 离子的存在。已经有研究表明，二价阳离子可以选择性地与脂肪酸结合。在存在非酶合成的磷脂的情况下，混合脂质膜可能会在进化的中间阶段存在。随着磷脂含量的增加，原始细胞最终会变得不能渗透离子。在没有复杂的基于蛋白质的跨膜离子转运机制的情况下，膜中磷脂的增加对原始细胞来说可能是致命的，因为没有办法维持跨膜的渗透和电平衡。这可能成为推动离子转运膜蛋白产生的进化压力。基于脂肪酸的囊泡可以被认为是一种原始细胞模型，具有通过达尔文进化而发生变化的潜力，并最终演化为现代细胞。

1.5 原始细胞的膜两亲性分子从脂肪酸到磷脂的演变

目前人们发现的所有细胞的细胞膜都由磷脂作为骨架和主要成分，这种共性证明这种磷脂存在于现存所有生命的最后一个共同祖先中。然而，虽然现存的所有细胞，包括古细菌和细菌以及真核生物都有一样功能的磷脂脂质膜，但古菌的磷脂合成途径被认为是与细菌和真核生物的途径不同的，是一种独特的合成途径。细菌和古生菌之间的磷脂生物合成途径的不可忽视的区别导致了这样一种说法，即始祖菌是缺乏磷脂膜的。但是最近的系统发育研究表明所有现存生命的共同祖先很可能已经拥有了相对复杂的磷脂膜。^[19]。

1.5.1 磷脂膜的特性

过去对膜的起源的研究重点关注两亲性分子如何在前生命条件下形成囊泡并充当原始细胞的原始边界，这是一种是自下而上的研究思路。相比之下，一种自上而下的方法，允许通过比较当今的生物体来推断和检查边界的特征，是在发现古细菌及其独特的膜之后很久才出现的。根据细胞理论，细胞来自细胞分裂，而现代细胞被由相似的磷脂组成的脂质膜所包围，它们的祖先应当也是一样的磷脂膜。不过自然界中存在两种不同但结构相似的磷脂。细菌和真核生物具有相同的膜生物化学特性，具有基于 3-磷酸甘油的酯连接脂肪酸磷脂。这些磷脂被认为是通用的，但令人惊讶的是：古细菌生物化学的开创性研究表明，古细菌膜上的磷脂是由 1-磷酸甘油制成的，是与类异戊二烯链相连的醚。这种化学差异反映了古细菌和细菌中不同磷脂生物合成途径的使用，特别是使用独特的磷酸甘油脱氢酶合成 1-磷酸甘油。值得一提的是，虽然教科书和综述经常花费大量篇幅强调古菌

膜和细菌膜磷脂的区别，但它们之间的差异实际上并没有那么明显。在细菌和真核细胞膜上发现过醚，同样在古细菌中检测到了脂肪酸。二者使用的前体即磷酸甘油的特性才是是两大类生命界限最分明的区分。这说明磷脂的起源可能比预期的更早。

磷脂膜相对于脂肪酸膜的核心优势是稳定。由于脂肪酸是单链两亲性分子，因此膜稳定所需的临界浓度远高于链长相当的磷脂。一旦确定简单的脂肪酸可以自组装成膜囊泡，很明显这种囊泡是早期原始细胞膜的良好模型。有趣的是，与由单独的纯脂肪酸制成的膜相比，包括脂肪酸、脂肪醇、脂肪酸甘油酯甚至多环芳烃在内的两亲物混合物产生的膜具有优异的热稳定性和渗透性。囊泡之间具备竞争关系，会发生相当程度的脂肪酸交换。因此，磷脂通过减少脂肪酸的流出来推动竞争性增长，在保持脂肪酸流入不变的同时离开膜。原则上，有两种效应会降低脂肪酸从膜上解吸的通量：首先，在磷脂存在的情况下，脂肪酸解吸率降低；其次，由于减少的部分减少，从膜的净流出减少，脂肪酸占据的膜表面积增大。这种从单链脂质到磷脂的转变必须是渐进的，这既是为了允许代谢和运输机制的共同进化，也是因为新生催化剂（例如核酶）的最初效率低下。因此，与磷脂合成相关的选择性优势必须适用于磷脂含量的微小差异，以推动这种转变。^[20]

1.5.2 脂肪酸-磷脂混合膜存在的可能

在现代细胞中，二酰基磷脂通过酶促酰化合成，由于参与二酰基磷脂生物合成的几种酶本身必须与膜结合才能发挥酶活性，因此尚不清楚在没有先进的酶促机制的情况下磷脂是否可以形成。但是已有研究表明：在使用小尺寸系统，加上热和紫外线驱动的蒸发循环，浓缩反应物以增加碰撞频率并克服出现和自组装的热力学和动力学障碍的情况下，一种原始磷脂可能从假定的原始环境中出现。这种磷脂聚集在一个碱性且相对较浅的火山热湖型系统中，该系统下有陨石、层状硅酸盐和富含铁的底栖生物。^[21]还原形式的磷可能与无数原始有机和矿物成分结合形成现代磷脂酰胆碱的原始还原磷磷脂类似物，具有自组装成原始膜结构的潜力。因此膜的形成可能发生在天然碱性的水中，例如热液海洋喷口。当形成的囊泡转移到酸性更强的溶液中时，电化学质子梯度会自发地建立和维持，这可能对原始细胞的代谢有作用。这种在水中天然磷脂的非酶合成为人工细胞中的脂质合成开辟了新的途径，并部分揭示了细胞膜的起源和进化。值得注意的是，磷脂前体，例如磷酸甘油或磷酸 2-氨基乙酯，也是在类似的条件下获得的。此外，一些磷脂也可以通过单酰基甘油和二酰基甘油的直接磷酸化获得。所以我们可以合理

推测，一种与现代细胞膜上磷脂类似的原始磷脂在原始环境下少量存在，甚至可能在早期就成为了脂肪酸以外的原始细胞膜的主要成分。

实验表明：随着磷脂含量的增加，细胞膜的脂肪酸流出量减少，从而导致细胞表面积增大，即从单链底物合成磷脂的能力对于早期细胞竞争有限的脂质供应非常有利。磷脂可以在竞争单链脂质的过程中推动原始细胞膜的生长，但是膜上磷脂含量的增加还将引发一系列新的选择压力，主要是磷脂比例升高导致的对营养物质的渗透性下降。所以早期细胞应当进化出用于代谢和转运机制的酶或结构，以克服磷脂膜的膜通透性降低。因此，磷脂膜的进化可能是内在物理过程的确定性结果，也是早期细胞进化的关键驱动力。^[20]简单脂肪酸细胞膜对单糖具有高渗透性，例如核苷酸及其前体，这意味着在没有任何大分子运输机制的情况下，以脂肪酸为主体的原始细胞可能已经从其环境中获得了复杂的营养物质。^[22]

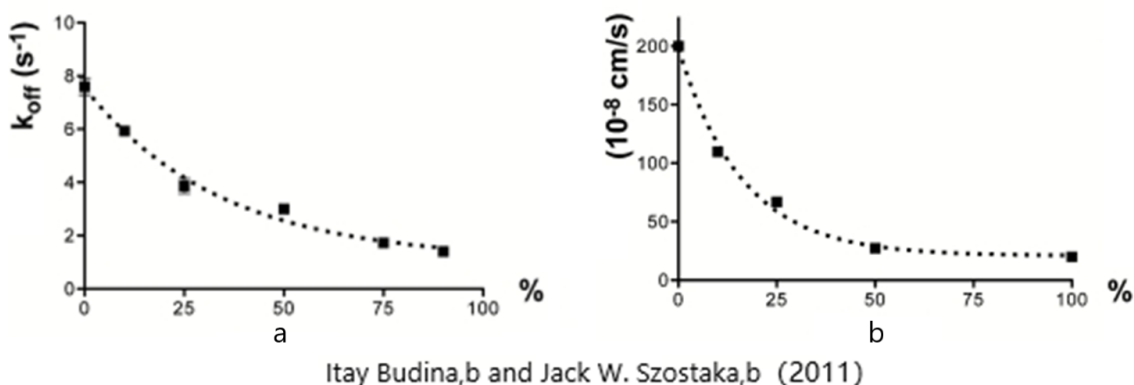


图 1 膜上磷脂比例对脂肪酸解离率和膜渗透率的影响

现代细胞的磷脂双层膜对离子、极性溶质和生物聚合物的渗透性非常低。在没有专门的蛋白质或者 RNA 分子的帮助下，核苷酸前体是很难通过扩散进入磷脂膜的。当然一部分简单的营养物质如甘油可以很容易通过磷脂膜。总之，无法依赖现代跨膜运输机制限制了磷脂双层膜实现脂肪酸双层膜的一些功能，例如大分子阳离子的扩散。^[23]由于磷脂膜的低渗透性，许多营养物质不能进入细胞内环境，而这是细胞维持自身所必需的。在现代细胞中，它们通过复杂跨膜蛋白机制来规避磷脂膜带来的渗透性问题，但显然地球上的第一个细胞不太可能将双层膜与特定的膜蛋白转运蛋白组装在一起。相反，一定存在中间进化步骤，这个阶段的细胞膜上的脂质分子应当具有当代膜的许多特征，而且结构必须相对简单，不依赖于先进的蛋白质机制。似乎有必要的是一种能够保留和释放特定分子的简单膜系统的出现。简而言之，原始细胞需要具有选择性渗透性，而选择性渗透性是脂肪

酸膜的固有特性，由脂肪酸组成的囊泡能够保留大分子聚合物和带电荷分子，同时允许更小、极性较小的溶质扩散。^[24] 因而，纯净的脂肪酸膜既有化石证据又有实验证据证明其在早期无生命的原始环境中就已经存在。当然，虽然缺乏化石证据，但磷脂的前生化合成已经在实验室实现了，而且脂肪酸可以作为磷脂的原料，所以早期的细胞膜成分可能已经混合，纯脂肪酸的膜组成的可能不存在，脂肪酸-磷脂混合膜从生命之初就伴随纯脂肪酸膜存在。

1.6 原始细胞的计算机模拟进展

常用的计算机模拟原始细胞的方法通常都基于元胞自动机，模拟矿物表面上的 RNA 世界进化。矿物表面通常被认为是生命起源的关键过程出现的场所，因为它有聚合、浓缩、组织和保护生物聚合物的作用。所以能高效模拟二维表面的元胞自动机是模拟原始环境的第一选择，例如 Tamás Czárán 等构建的代谢耦合复制子系统 (MCRS)。MCRS 包括一系列模型，实现了一个简单的、物理化学和生态上可行的场景，用于推演生命进化从原始环境发生的第一步。假设一组 n 个不同的、具有代谢活性的核酶复制子通过催化简单代谢的反应来竞争它们自己合作产生的单体，每个复制子占据一个站点，代表发生所有相互作用的矿物表面。一个站点的可能状态数为 $n + 1$ ，包括“空”状态和 n 种不同的占领状态。一次迭代由站点数量与点阵中的站点数量相等的基本更新组成。更新算法是随机的：每个站点的状态平均每单位时间更新一次，顺序也是随机的。一旦任何一个 RNA 分子通过从活化的单体进行模板定向自我复制而成为自催化的、非生物产生的 RNA，种群的进化就开始了，并开始呈指数增长。对有限的外部单体供应的竞争引发了有利于具有催化活性的 RNA 分子的选择，这些分子通过任何可能的方式帮助自我复制。提供这种自催化帮助的一种方法是成为复制酶核酶，另一种方法是通过促进来自外部资源的单体合成，即通过进化出代谢酶活性来增加单体供应。逆向进化可能会建立一个越来越自养、合作的代谢核酶群落，运行越来越复杂和越来越有效的新陈代谢。维持这样一个代谢复制子的合作社区会引发两个严重的生态问题：一个是尽管合作的复制子具有不同的可复制性，但要保持系统共存；另一个是分子寄生。表面结合的 MCRS 为这两个问题提供了自动解决方案：共存和产生对寄生分子抗性，是假设代谢相互作用的局部性质的结果。MCRS 模型框架自然地未来的膜进化建模提供了一个合适的起点，并将理论扩展到以自洽的方式涵盖第一个原始细胞的出现。假设代谢、遗传和膜功能的共同进化遵循渐进隔离的情景，

是原始细胞进化最早步骤的概念蓝图。从分子水平到细胞水平的转变可能在进化的早期自然发生，这个过程的进化机制很复杂。除了膜边界在保证分子合作和抵抗寄生分子方面的作用外，还应强调原始细胞内容物向外扩散和原始细胞的整体移动，获得原材料的更大机会。这可能在一定程度上弥补了膜对原材料的阻隔作用这个缺点。^[26]还有 Wentao Ma 等使用的蒙特卡罗模拟，这是一种计算机模拟，通过基于确定规则的相对概率来模拟现实中的随机事件。同样为了使用蒙特卡罗方法对系统进行建模，假设了一个二维表面。原因之一是二维系统更易于模拟，并且足以表示局部分子相互作用。另一个原因是模型系统中的一个基本反应，即从核苷酸形成多核苷酸已被证明能够在矿物表面发生，矿物离子具备催化合成功能。因此完全可以想象系统是一个极浅的，深度只够一个细胞存在的“池子”，池子底部有富含矿物离子的岩石，甚至有一层薄薄的流体层散布在岩石表面。模拟中的检查对象是可以是核苷酸、单个多核苷酸链或具有全部或部分互补链的多核苷酸链，此外，附加模板的复制酶也属于同类型的检查对象。二维系统表面被划分为 $N \times N$ 的网格单元，用于考虑分子相互作用的位置效应。一个网格单元代表一个区域，其中单元内的分子可以在一个蒙特卡罗步骤中相互交互，单元内可能存在一个原始细胞。纯元胞自动机方法通过检查网格单元的状态来处理自我复制 RNA 和复制酶复制的内在特征，而该模拟通过检查分子本身的状态来模拟分子自我复制。二维系统中的核苷酸原材料在每个蒙特卡罗步骤中，可能以一定的概率转化为一个核苷酸，并且每个类型核苷酸（即 A、G、C 和 U）的形成是随机的，并且是等概率的。在同一个蒙特卡罗步骤中，系统中的每个分子和细胞都被检查一次，并且可以执行包括移动在内的各种操作。^[27]

元胞自动机大量用于建模复杂动态系统，在包括生命科学在内的多个领域都有广泛应用。基本元胞自动机是一个有限二元格，其值为 0 和 1，并且不断更新。在完全自动机中，所有状态转换都是局部的，并且仅受重复更新的固定单元长度附近的数字状态 1 的影响。如果当前配置中的所有单元格同时更新，则机器称为同步元胞自动机。如果每次迭代只更新一个单元格，则自动机是异步的。另一个区别是自动机中包含的随机性程度。如果没有随机性，则自动机称为确定性自动机，如果要更新的单元格是随机选择的，则自动机是概率性的。概率元胞自动机通常用于研究原始生命模拟，使用基于每个单元的邻域的一些局部变换规则迭代地从一种配置转换到另一种配置。至于每次迭代允许更改的单元数决定了我们称它为同步还是异步自动机。如果转换规则以某种方式涉及随机性，我们称它们为概率自动机，否则称为确定性。对于任何类型的元胞自动机，主要的理论任务保

持不变：从任意初始配置开始并以高精度预测最终配置。如果自动机是确定性的，则结果将简化为两种状态之一，全为 0 或全为 1。如果自动机是概率性的，则结果为相应的正态分布。^[28]

蒙特卡洛模拟可以被认为是元胞自动机的一种，是一种是针对多参数和非线性的动态代谢系统的强大工具，蒙特卡洛的一般方法是创建多个平行的模拟世界。如果知道采样的实际参数，实验将使用实际参数来模拟对象。实验样本的随机性不仅需要估计模型的最优值，还需要估计参数的分布，检验关于这些参数值的统计假设，因而将分析模型应用于参数分布是完全不合适的，因为这些假设不适用于本质上的非线性回归。这就是我们使用蒙特卡洛方法的原因。

总之达尔文的自然选择进化论为设计一个可以自主增加复杂性的“进化”人为系统提供了坚实的理论基础。当竞争个体之间的利益冲突被部分抑制时，复杂性会增加。这之后进化生物学中的“选择水平”问题变得很重要，出现需了要多层次选择的复杂性的场景。在分析了人工细胞中竞争复制的动态行为后，分化催化剂的增殖或核酶催化效率的提高演变成合理设计的各种人工细胞。这些系统中的实验进化可以作为人工适应系统的一个很好的例子，为代谢复杂性进化的可能进化途径提供新的见解。由核酶的分化和或增加的催化效率催化的增殖可以发展成精心构建的人工细胞，并在一定程度上阻止不同类型复制之间的强烈内部竞争。这些系统是人工适应系统的好例子，它们为代谢复杂性进化的可能进化途径提供了新的见解。^[29]由核酶的分化或增加的催化效率催化的增殖可以发展成精心构建的人工细胞，并在某种程度上通过以相等的概率选择不同类型的复制品来防止它们之间的强烈内部竞争。Elias Zintzaras 等的蒙特卡洛模型主要模拟的就是这种机制下的模板竞争。^[30]

第二章 研究方法

2.1 模型模拟的对象

模拟将基于一个对起源过程的假设并对假设进行验证，如图 2 所示。最早大量脂肪酸和少量磷脂存在于原始环境中，在特殊条件下富集到一定浓度，从而组成双层的混合囊泡。在囊泡存在的条件下核苷酸前体自然形成了核苷酸，并组成了各种排列顺序的 RNA，其中一种特殊序列的 RNA 是具有催化功能的核苷酸合成核酶，我们将其称为 NSR (Nucleotide synthesis ribozyme)。囊泡内的稳定环境使得 NSR 能接触更多原材料，核苷酸降解概率变低，和 NSR 结合的囊泡获得进化优势，这在实验室之前的实验中是已经通过模拟证明了的。后来原始环境又出现了一种特殊序列的 RNA 能催化脂肪酸和磷酸甘油合成磷脂，我们将其称为 PSR (Phospholipid synthesis ribozyme)。PSR 使得膜上的磷脂比例变大，两种膜成分的两亲性分子离开膜的概率变低，同时膜的渗透性下降。作为 PSR 副作用的补偿，一种可以增加膜对极性小分子主要是核苷酸前体渗透性的 RNA 出现了，我们称之为代偿酶 MPR (Membrane permeation compensatory ribozyme)。

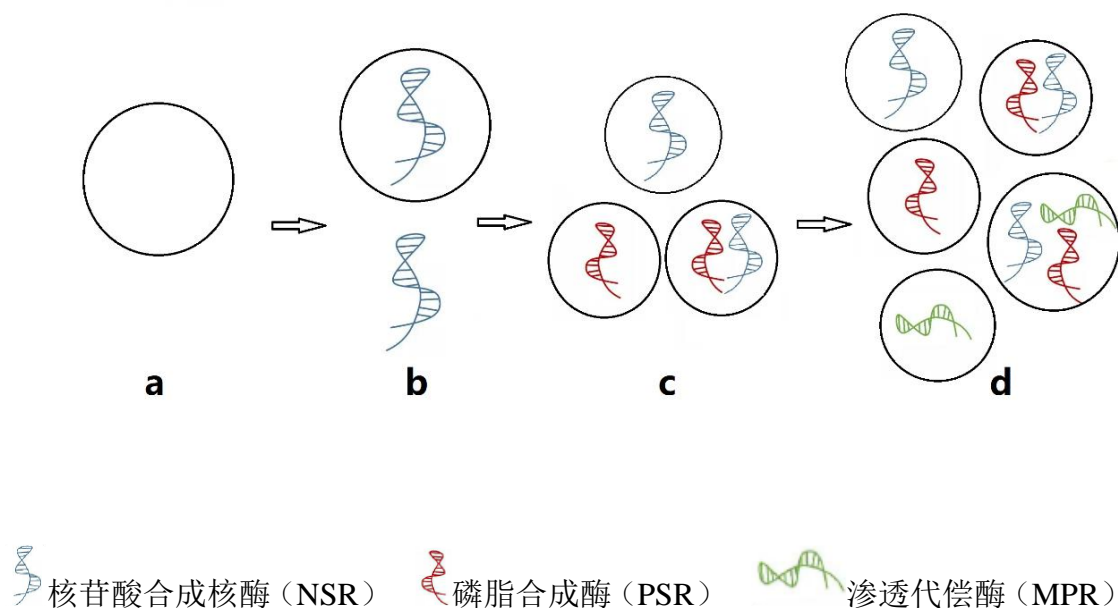


图 2 原始细胞从囊泡出现到三种功能酶出现的演化过程

图 2 展示了推定的原始细胞演化过程，从囊泡出现开始到三种功能酶出现总共经历四个阶段。蓝色 RNA 表示 NSR, 红色 RNA 表示 PSR, 绿色 RNA 表示 MPR。a 表示自然合成的脂肪酸和磷脂组成的混合囊泡，其中脂肪酸占绝对多数。b 表示核苷酸合成酶（NSR）出现，与自然合成的脂肪酸和磷脂组成的脂肪酸占绝对多数的混合囊泡结合成为具有自我复制功能的“原始细胞”。c 表示磷脂合成酶（PSR）出现，与磷脂占绝对多数的混合囊泡结合。d 表示磷脂渗透抑制代偿酶（MPR）出现。

2.2 模型的基本概念

本文使用的原始细胞模型代码基于 Wentao Ma, Jiming Hu 在 2012 年发表的模拟^[22]。模型代码使用的是 C 语言，在支持 C 的平台均可运行。模拟基于蒙特卡洛模型，以一个 $N \times N$ 网格的二维表面代表原始细胞所在的环境，采用环形拓扑以避免边缘效应，其中系统中的每个事件都可能在一个时间蒙特卡洛步骤中以一定的数值概率发生。“网格”代表网格中的一个正方形。它在传统的随机元胞自动机中被称为“网格单元”。在这里，为了避免它与模型中的原始细胞混淆，并强调它是分子驻留的空间，称之为“网格室”。在该系统中，“网格室”可以容纳一些被认为是足够相邻以相互作用的分子。不同于随机元胞自动机，一个细胞占据一个“网格室”并且相邻“网格室”之间发生相互作用，这种处理将节省计算成本并有利于考虑到数字的复杂相互作用的更大系统的模拟分子。

假定一种含有大量可以合成 β -d-核苷酸，核苷酸也可能衰变成原材料的原始矿物表面，构成一个核苷酸库为原始细胞提供遗传物质和酶。为简化起见，系统中的总原材料量设定为常数。而原始环境中所有可能存在的分子被我们分为三大类：RNA、成膜分子和其它分子。这里考虑的是，我们模拟专注的是核酶以及核酶在原始环境中的影响，所以 RNA 需要有较为复杂的行为，因此 RNA 专门用一个结构体来表示。结构体内包含了 RNA 两条链的链长和 RNA 的碱基序列信息，系统内的每一个 RNA 都分配了这样的一个内存空间。而成膜分子指脂肪酸和磷脂，这种分子在系统中有两种存在形式，包括游离和在膜上，所以由两个数组用于表示它们。而其他分子即核苷酸前体、脂肪酸前体和磷酸甘油只有一种状态，这个状态用 0 和 1 来指示，存在即为 1，不存在了即为 0。它们的存在被记录在和网格室关联的几个三维数组中，这些三维数组用于指证这些小分子在模拟的原始环境中的分布。

已经有化学实验证明在一个推定的起始核苷酸库中，通过矿物离子的催化机制可以形成不同长度和序列的多核苷酸，其后多核苷酸可以作为形成其互补链的模板，包括非酶模板定向连接和 RNA 催化下的模板定向连接，被模板吸引的相邻原料按照模板序列加入。我们假设互补链的合成倾向于更多地基于连接而不是聚合，聚合仅利用单体作为底物，这是在许多短寡核苷酸已经通过矿物催化在池中形成的背景下。因此，这里的复制酶被认为是连接酶而不是聚合酶，聚合酶被认为是未来进化中出现的東西。考虑到互补链合成的保真度，可能会发生不正确的碱基配对包含假定的“复制酶特征序列”作为子序列的多核苷酸在这里被假定为原始复制酶。这种简化是可以接受的，因为核酶的功能是由它的结构和它的序列决定的，而背景是我们对可能的原始 RNA 复制酶的关键结构和关键序列知之甚少。一旦复制酶结合到模板上，它就可以在未来的步骤中充当核酶。结合核酶可能在三个方面发挥作用：帮助模板吸引底物；在结合位点可以保护被吸附的底物，使之不掉落；催化模板上相邻排列的底物的连接。能量问题在模拟中得到了简化。假设每个分子一旦形成就会在其 5' 端被激活，并准备与其他分子反应形成磷酸二酯键。例如，当一个核苷酸从原材料中形成时，它会立即被 5' 端激活。当多核苷酸链降解时，掉落的 3' 末端产物也被假定为立即被 5' 末端激活。

系统内任意一个“网格室”形成了囊泡，我们就称之为一个原始细胞，这里原始细胞大小只会影响细胞对小分子的渗透率，而不影响细胞占据的空间，无论细胞的大小如何，它都只占据一格，处于这一“网格室”的所有分子都被认为处于膜内，只与膜内的分子交互，并随着细胞的移动被带着一起移动。在“网格室”内的分子想要移动，将同时受到所在“网格室”和想要前往的“网格室”，两个格子的影响。任意一个“网格室”被细胞占据，都会影响移动的成功率。我们通过这种机制模拟原始细胞的渗透。渗透概率本身是一个定值，但是在实际模拟中会受到代偿酶和膜磷脂比例的影响。原始细胞的形成和破裂都由随机概率决定，但是受到成囊泡所需的最低两亲性分子数量的影响，需要检测到“网格室”内有超过下限的两亲性分子数量才有可能形成原始细胞，同样地一旦发现“网格室”内的原始细胞的两亲性分子数量低于下限，细胞会立刻破裂。而高于下限的细胞才是概率性地破裂，并且网格内细胞的大小高于两倍最低两亲性分子数量之后才有可能发生细胞分裂。这种机制保证了细胞大小不会低于一个定值，它是为了模拟现实中囊泡的最低浓度临界值。

由于对原始核酶的了解不足，模拟中三种酶都是高度抽象的，三种 6nt 长度的序列被认为分别是具有三种功能域的核苷酸排列。序列是随机设置的，实验中

我们设置了 G, C, A, C, G, U 拥有 NSR 也就是 RNA 合成的能力; C, C, U, G, U, A 则行驶 PSR 也就是催化磷酸甘油和脂肪酸合成磷脂分子的能力; 而 A, U, U, G, G, U 发挥渗透代偿功能, 系统在每一个步长中都会检测网格室内的 RNA 是否有这些序列。改变序列 A, U, C, G 的排列顺序不应当影响模拟结果, 在满足条件之后发挥作用, 我们实验中发现将这里的长度 5-7nt 都不会影响最终结果, 设定超过 7nt 的序列为有功能的结构域虽然会使序列的繁荣变困难, 但是这个困难可以通过原材料浓度的增加或者 RNA 自然合成率来克服。每一步我们都会检测该“网格室”内是否有 RNA, 并按链表检测 RNA 是否有某一结构域, 一旦结构域存在我们即认为有一到数次机会为原材料提供合成位点或者发挥其他功能, 同一类核酶能提供的酶催化次数应当是相同的。

2.3 模型使用的参数及其定义

各种小分子随机分布在 $N \times N$ 网格中的任意一格, 具备移动、形成囊泡、与其他分子发生相互作用等基本行为, 其中只有同一个“网格”内的分子才能相互作用, 在一个时间步长中, 以下每个事件都可能以某种概率发生:

表 1 模型参数表

参数名	参数值	参数意义
STEPNUM	1000000	蒙特卡罗模拟的总时间步长
MECH_CHANGE_STEP	100000	接种 psr 和 MPR'函数的步骤
STAREC	0	开始记录的步骤
REC RECINT	1000	记录的间隔步骤
MAX_RNA_LENGTH	100	模拟中允许的最大 RNA 长度
NRSEQ	G,C,A,C,G,U	核苷酸合成核酶的假定活性结构域序列
ARSEQ	C,G,U,C,A,A	脂肪酸合成核酶的假定活性结构域序列
PRSEQ	C,C,U,G,U,A	磷脂合成核酶的假定活性结构域序列
MRSEQ	A,U,U,G,G,U	渗透代偿核酶的假定活性结构域序列
CONTRSEQ1	G,U,G,A,U,C	对照(无功能)序列 1
CONTRSEQ2	U,C,A,A,G,U	对照(无功能)序列 2
CONTRSEQ3	A,G,C,U,G,C	对照(无功能)序列 3
INOCUSTEP	1000	接种的步骤
INOCU_CELL_NUM	5	接种细胞数
INOCU_SEQ_NUM	5	接种 CONTRSEQ1 和 NSR 的数量

参数名	参数值	参数意义
INOCUNUM_PSR	5	接种 PSR 数
INOCUNUM_MPR	5	接种 MPR 数
SD	25	随机数种子
N	40	系统表面定义为 $N \times N$ 网格
ROOMNUM	(N*N)	网格室数量
TNPB	150000	开始时接种到系统中的总核苷酸前体（核苷酸测量中的商）
TAPB	80000	开始时接种到系统中的总两亲性分子前体数量
PND	0.2	核苷酸衰变的概率
PRL	0.000001	核苷酸和 RNA 随机连接的概率
PBB	0.00005	原始细胞中磷酸二酯键断裂的概率
FBO	5	原始细胞中磷酸二酯键断裂的因素
PMV	0.9	分子移动的概率
LAM	200	形成膜的两亲性分子的下限数量
FSI	1	内部不透水分子抑制替代离子进入的因素
PPDM	0.002	磷脂在膜内裂解的概率
TGB	40000	初始接种到系统中的磷酸甘油数量
FTR	4	代偿酶效用因子
FPM	0	磷脂对脂肪酸前体分子及磷酸甘油透膜的影响因素
FPP	45	磷脂对核苷酸前体过膜的影响因素
PMF	0.5	在 LAM 处成膜的概率
PCB	0.00007	原始细胞破裂的概率
PCD	0.3	原始细胞分裂的概率
PALM	0.008	两亲性分子离开膜的概率
PAJM	0.9	两亲性分子加入膜的概率
PPJM	0.9	磷脂加入膜的概率
PCF	0.001	相邻时原始细胞融合的概率
PPD	0.2	磷脂衰变的概率（膜外）
PAF	0.003	脂肪酸合成核酶催化其前体形成脂肪酸的概率
PPF	0.008	从前体（非催化）形成磷脂的概率

参数名	参数值	参数意义
PPFR	0.8	磷脂合成酶核酶催化其前体形成磷脂的概率
PNF	0.0005	从前体（非催化）形成核苷酸的概率
PNFR	0.9	核苷酸合成核酶催化其前体形成核苷酸的概率
PAT	0.9	模板吸引底物的概率
PFP	0.01	错误碱基配对的概率（与 RNA 的复制保真度有关）
PTL	0.007	非酶模板定向连接的概率
PSP	0.5	碱基对分离的概率
PNPP	0.05	核苷酸前体渗透膜的概率
PAPP	0.09	脂肪酸前体渗透膜的概率
PGPP	0.9	磷酸甘油渗透膜的概率
PMC	0.09	原始细胞运动的概率
PAD	0.08	脂肪酸衰变的概率（膜外）
PADM	0.002	脂肪酸在膜内衰变的概率

模型内核苷酸及其前体以及 RNA 的行为类似现实中的核苷酸、RNA 与核苷酸前体。系统中核苷酸前体可以非酶转化或者在酶催化下等概率转换为 A、U、C、G，概率为 PNF、PNFR（参数含义和值见表格 1）。一个核苷酸可能会转换成一个带有 PND 的核苷酸前体。一个核苷酸可以与另一个核苷酸或 RNA 链在其末端与 PRL 连接。RNA 链内的磷酸二酯键可能会因 PBB 断裂，对于原始细胞外的 RNA，概率将乘以 FBO。RNA 可以变成模板并用 PAT 吸引核苷酸或寡聚体通过碱基配对，每个残基位点发生错误碱基配对的概率为 PFP，对于寡聚体，碱基配对测试适用于其所有残基。在 RNA 模板上相邻排列的核苷酸和寡聚体可以通过 PTL 模板指导相互连接。每个碱基对有 PSP 的概率分开，如果它们之间的碱基对可以分开，则 RNA 模板上的底物或产物可以解离。包含特征结构域的 RNA 可以作为相应的核酶发挥作用：Rep、PSR、NSR、MPR，催化相应的分子间反应或者发挥特定功能，NSR 分子有 PNFR 的可能可以催化“网格室”的核苷酸前体形成核苷酸，这个概率如同现实中的酶促反应，远超无酶催化的概率 PNF。同样的 PSR 分子可以有 PPFR 的概率催化膜上的脂肪酸前体和磷酸甘油形成磷脂，MPR 分子可以代偿磷脂对核苷酸前体进入膜概率的抑制。但是，如果 RNA 的长度等于或超过特征域的两倍，则认为“正确”的结构会受到冗余残基的干扰，则 RNA 不会发挥核酶的功能。

一个 Rep 分子可以有 PRB 的概率结合到一个 RNA 模板上，并有 PRD 的概率从模板上脱落。如果模板上有 Rep，则模板导向的连接可能与 PTLR 发生，如果连接位点两侧的一个或两个碱基对都是假的，则连接不会发生，除非满足另一个概率 PFLR。在模型中，PFP 和 PFLR 是与复制保真度相关的两个参数。当 PFP 和 PFLR 较小时，复制保真度会增加。

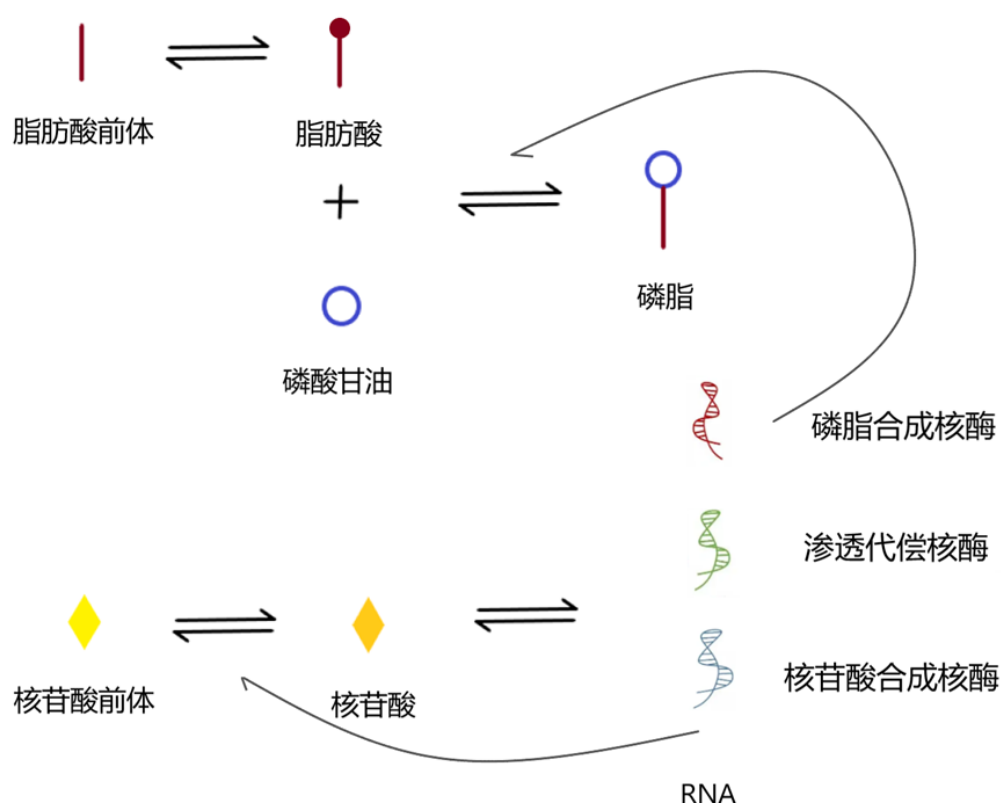


图 3 模型内分子与分子间的反应模式图

模型内两亲性分子及其前体的行为类似现实中的脂肪酸与磷脂。脂肪酸分子前体有 PAF 的概率转换为脂肪酸分子。脂肪酸分子所在的“网格室”内如果存在原始细胞，那么膜上的脂肪酸分子和细胞内的甘油磷脂有 PPF 的概率反应生成磷脂，在“网格室”内有 PSR 的情况下，这个概率是 PPFR，模拟为了简化计算，在实验中将所有链长的脂肪酸分子视为等价，即只有一种长度的原材料，磷脂一律视作单尾磷脂，不考虑更复杂的双尾磷脂的存在。游离的脂肪酸分子有 PAD 的概率会降解成脂肪酸分子前体，而在原始细胞膜上的脂肪酸分子降解的概率是 PADM。磷脂有 PPD 的概率降解为甘油磷脂和脂肪酸分子。一个游离的脂肪酸分子有 PAJM 的概率加入“网格室”的和邻近“网格室”的原始细胞膜，而在原始细胞膜上的脂肪酸分子有 PALM 的概率离开膜，概率受到膜上磷脂和脂肪酸的比例影

响，原理和参数来自 Itay Budin 和 Jack W. Szostak 在 2011 发表的文章，影响系数设为 $9.3 / (1 + (\text{磷脂比例})^{0.7} / 36 - 1.2) / 10$ 。一个游离的磷脂分子有 PPJM 的概率加入“网格室”的和邻近“网格室”的原始细胞膜，而在原始细胞膜上的磷脂分子有 PPLM 的概率离开膜，概率同样受到膜上磷脂和脂肪酸的比例影响。脂肪酸和磷脂作为两亲性分子可能会在“网格室”的边界聚集并形成膜，然后“网格室”被原始细胞占据。两亲分子也可能离开膜。如果在移动方向上没有膜，分子可能会移动到相邻的房间，如果在移动方向上有膜，两亲性分子可能会加入膜，核苷酸和 RNA 是不可渗透原始细胞膜的，而核苷酸前体和两亲性分子前体分别有 PNPP 和 PAPP 的概率渗透膜。核苷酸前体渗透膜的概率受到磷脂比例的影响，比例越高，过膜概率越低，所以我们使用系数 $(1 + (FPP / (\text{MPR 的数量}) * \text{mr_turn}) * (\text{磷脂比例}))^{(-1)}$ 来模拟对 PNPP 的影响，FPP 越大，磷脂对渗透率的影响越大，而 FTR 决定 mr_turn 的大小，FTR 越大，代偿酶 MPR 的功能越强。

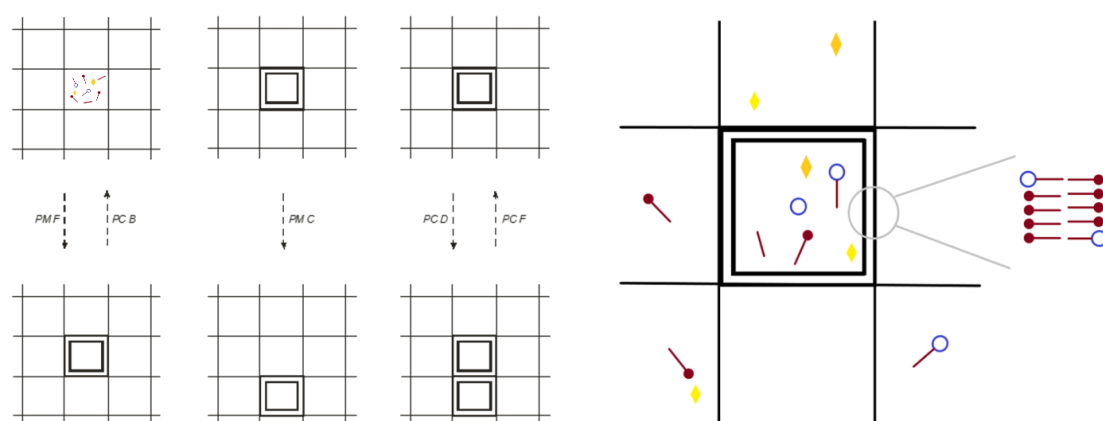


图 4 模型内原始细胞结构与行为模式图

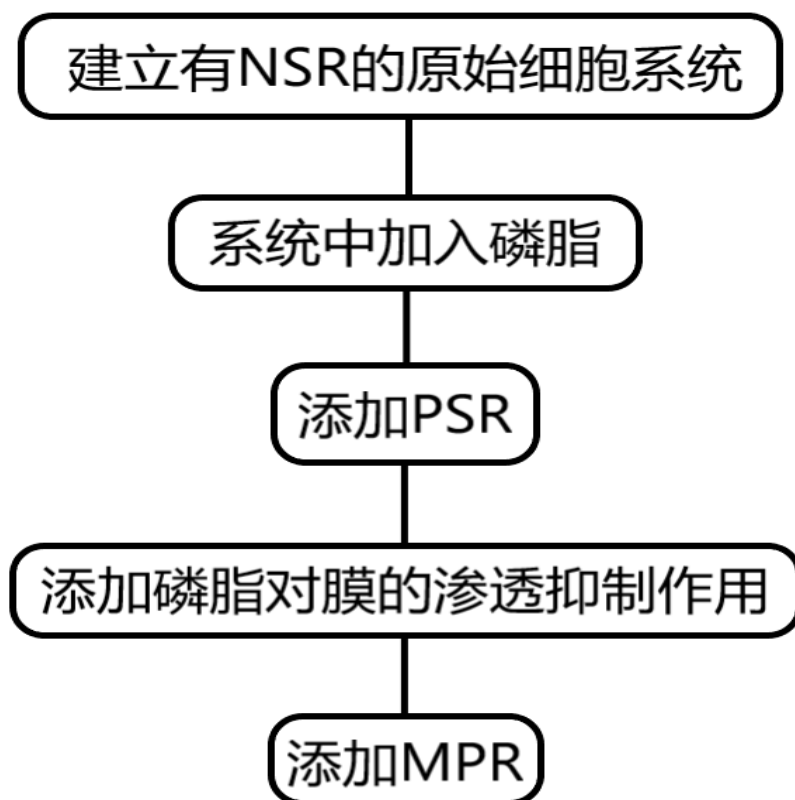
模型内原始细胞的行为类似现实中的细胞。产生囊泡也需要最低浓度的两亲性分子，称为临界囊泡浓度，由于系统内每个“网格室”的面积是相同的，所以我们认为一个“网格室”内最低临界两亲性分子数量也是一个定值，称之为 LAM。一个没有细胞存在的“网格室”的两亲性分子数量在达到下限 LAM 之后有 PMF 的概率在“网格室”边缘组装成膜，将分子包围在其中并形成“原始细胞”。一个原始细胞有 PCD 的概率分成两个，两个相邻的原始细胞有 PCF 的概率融合成一个。原始细胞在分裂成为两个原始细胞时，会优先保证分裂后的两个细胞细胞膜上的两亲性分子都不低于 LAM，然后多余的两亲性分子再等概率成为任意一个原始细胞膜上的膜分子，而分裂的原始细胞内容物即各种 RNA 和原材料及散落的脂肪酸和磷

脂也将等概率成为分裂后任一新的原始细胞的内容物。这个思路建立在原始细胞不太有可能具备类似有丝分裂的等量分裂机制的假设上。生命体的达尔文进化应当是循序渐进、由简到繁的，想要将保证分裂的两个细胞具有同样的 RNA 显然是需要一种非常复杂的机制，常规的早期模拟都认为这种机制发生在蛋白质出现之后。而分裂前原始细胞有 PCB 的概率破裂，膜上的两亲性分子变成游离的两亲性分子，前提是“网格室”内的两亲性分子数量在高于下限 LAM 的水平上，一旦低于 LAM 细胞会无条件破裂，对应实验室内测得的囊泡最低浓度临界值，囊泡过小时不应当还能维持细胞形态。在下一个时间段之前，网格房间中的分子和原始细胞可能会移动到相邻的“网格室”，在原始细胞外的核苷酸、核苷酸前体、两亲性分子或两亲性分子前体前体有 PMV 的概率移动到周围未被原始细胞占据的“网格室”。而原始细胞有 PMC 的概率向周围未被原始细胞占据的“网格室”移动，原始细胞内的核苷酸、核苷酸前体、RNA、两亲性分子或两亲性分子前体将跟随原始细胞一起移动，如同现实世界囊泡的行为。

模拟系统内的概率值应满足现实中的约束条件。核酶反应应该比相应的非酶反应更有效，因此 PTLR 应当大于 PTL、PNFR 应该远大于 PNF，还有 PNFR 也应该远大于 PNF，至少要有两个数量级的差异。模板指导的连接应该比随机连接更有效，所以 $PTL \gg PRL$ ，以保护 RNA 链中的核苷酸残基。这里，假设链内的核苷酸不能降解，而链末端的核苷酸降解率明显低于游离核苷酸，即 $PNDE \ll PND$ 。膜内的脂肪酸应受到保护，因此 $PADM \ll PAD$ 。由于两亲性分子的自组装特性， $PMF \gg PCB$ ， $PAJM \gg PALM, PPJM \gg PPLM$ 。由于磷脂膜比脂肪酸膜更加稳定， $PPLM \gg PALM$ 。游离分子的运动应该比原始细胞更容易，所以 $PMV > PMC$ 。PBB 可能与 PRL 处于相同的顺序，因此 $PND > PNF$ 。膜上磷脂比例高的细胞，两亲性分子离开膜的概率更低且满足一个非线性关系。^[18]膜上磷脂比例低的细胞，核苷酸前体应当更容易渗透膜。而磷酸和脂肪酸前体由于性质和膜分子类似，根据相似相溶原理，渗透率不受磷脂比例的影响。

原始生命的发生应该是特定环境下的一个随机事件，因此在模拟案例的初始步骤中引入 TNPB 数量的核苷酸前体，TAPB 数量的脂肪酸前体，TGPB 数量的磷脂前体之一的磷酸甘油，INOCUSTEP 时接种原始细胞、功能酶和无功能 RNA。INOCU_CELL_NUM 是接种原始细胞的总数；INOCU_SEQ_NUM 接种的无功能 RNA 数量和 NSR 的量，分别用于对照；INOCUNUM_PSR 和 INOCUNUM_MPR 则是接种步骤中加入的 PSR 和 MPR 的数量。在这里我们将接种的 RNA 数量统一为 25，因为同时出现过高数量的序列一致的 RNA 是不符合逻辑的，但是过低的量

又很容易受到环境波动和小概率事件的影响降解，接种失去意义，因此折中选择了这个数字。将各类 RNA 的数量统一则可以为我们提供解读结果时的便利，在真实原始生命早期环境中这种情况概率不高但绝非不可能事件。接种的步骤代表此时生命事件已发生，最早的功能酶已经出现，如果酶的数量和细胞数量能维持一定水平，那么我们就可以认为模拟是成功的。



NSR（核苷酸合成核酶）、PSR（磷脂合成酶）、MPR（渗透代偿酶）

图5 实验流程图

2.4 模拟流程

模拟磷脂合成酶 **PSR** 出现前的演化时期时系统内只有 **NSR** 是有功能的，为了确定磷脂对膜稳定性的影响，我们设定一个磷脂既没有酶催化，也没有功能的系统作为对照组，与其他条件一致但是磷脂对膜稳定性有增益的系统作对比，观察系统的行为。重点是记录 **NSR** 和原始细胞的数量，以及原始细胞的大小。如果实验组 **NSR** 和细胞的数量能稳定超过接种的水平，而对照组不行。那么磷脂对原始细胞的益处就可以证实了。而原始细胞的其他行为则可以解释如下原理：磷脂膜的稳定性是通过掠夺其他原始细胞膜上的脂肪酸来实现的。

模拟磷脂合成酶 **PSR** 出现的演化时期时系统内除了 **NSR**, **PSR** 也是有功能的，**PSR** 催化磷脂合成反应，使膜上的脂肪酸和膜内的磷酸甘油发生反应成为磷脂。具体实现见附录代码标红处。这里为了使系统内环境更加接近真实环境，磷脂的酶催化合成只在膜上发生，因为磷脂的合成相当复杂，与核苷酸以及脂肪酸不同，需要两种原材料和酶同时接触。为了使 **PSR** 催化合理，最佳选择是设定反应只在膜上，现代细胞大量的酶都与膜结合也可以说明这一点。实验组设为 **PSR** 有功能，对照组设为 **PSR** 没有功能，两个组的磷脂都不影响原始细胞的渗透率，其他条件也相同。同时记录 **NSR**, **PSR** 和原始细胞的数量，以及原始细胞的大小。如果实验组 **NSR**, **PSR** 和原始细胞的数量能稳定繁荣，对照组不能，那可以证明 **PSR** 在 **NSR** 之后出现是可能的进化方向。

模拟磷脂合成酶 **PSR** 出现且磷脂膜影响渗透性的演化时期时系统内除了 **NSR**, **PSR** 有功能，磷脂可以降低两亲性分子离开膜的概率以外，磷脂对膜渗透性的抑制也被考虑进去了。磷脂比例越高，核苷酸前体进入膜的概率越低，以用于模拟真实环境中磷脂膜的低渗透性。这点也是人们之所以认为原始细胞膜是脂肪酸而非磷脂的重要原因。实验组设为磷脂可以降低两亲性分子离开膜的概率，磷脂不影响两亲性分子离开膜的概率两种情况，其中概率为定值，其他条件不改。定期追踪 **NSR**, **PSR** 和原始细胞的数量，以及原始细胞的大小。如果实验组 **NSR**, **PSR** 和原始细胞的数量能稳定繁荣，对照组的数量低于实验组，但是依然可以繁荣，那除了证明渗透抑制是副作用之外，还可以进一步证明 **PSR** 在 **NSR** 之后出现是可能的进化方向，即 **PSR** 的副作用可以被其对原始细胞的正向作用克服。

模拟磷脂合成酶 **PSR** 和代偿磷脂酶 **MPR** 都出现的演化时期是模拟系统同时有三种功能酶 **NSR**, **PSR**, **MPR** 的阶段，也是本文对原始细胞膜替代最终阶段的模拟。这个阶段拥有 **PSR** 的原始细胞的膜已经可以被视作磷脂膜了。相对于现代细

胞膜复杂的过膜机制，这里把最原始的过膜辅助酶称为代偿酶 **MPR**。具体实现见附录代码标红处。这种假定中最早出现的结构未知的酶功能应该是针对原始细胞的需要的，显然此时蛋白质未必已经出现，原始细胞也不见得能利用质子梯度，那么最原始的过膜辅助酶不应该是辅助氨基酸和质子等过膜的，而应该是帮助核苷酸前体渗透膜，用于补偿磷脂比例过高对细胞获取核苷酸前体的困难。就如同脂肪酸膜被认为比磷脂膜出现更早的原因，磷脂膜获取某些原材料太过困难。磷脂在这里是一把双刃剑，同时对原始细胞有增益和副作用，降低两亲性分子离开膜的概率有利于原始细胞的存在，而降低膜渗透性则使原始细胞更难获得自我复制需要的核苷酸前体，一旦 **MPR** 出现并发挥作用，磷脂膜的优势应该被放大。为此实验组 **MPR** 有功能，对照组设为 **MPR** 无功能，其他参数一致。每隔一段时间统计系统内 **NSR**，**PSR**，**MPR** 和原始细胞的数量。如果实验组 **NSR**，**PSR**，**MPR** 和原始细胞的数量能稳定繁荣，对照组的繁荣度低于实验组或者 **MPR** 不能稳定存在，**MPR** 的起源可以视作模拟成功。如果 **MPR** 不能在缺少 **PSR** 的情况下繁荣，以及 **MPR** 和 **PSR** 出现的先后顺序也就得到了确认。

第三章 实验结果

3.1 磷脂在系统中的出现

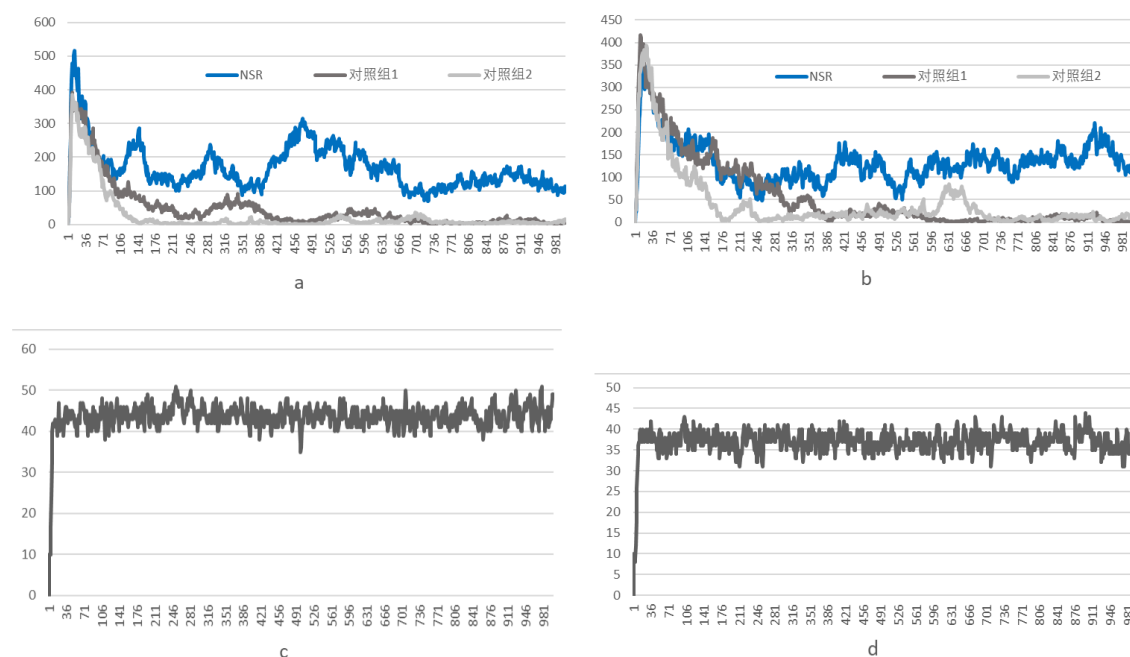


图 6 磷脂在系统中有正面作用的模拟结果图

图 6 中 (a) (b) 纵坐标代表该序列 RNA 在系统中的总数量，横坐标代表系统运行步数即时间。NSR 是核苷酸合成核酶，对照组 1、2 分别指两种没有功能的序列。(c) (d) 纵坐标指示系统中原始细胞的总数量，横坐标代表系统运行步数即时间。(b) (d) 此模拟磷脂出现但没有功能，NSR 有功能。(a) (c) 此模拟磷脂出现且对膜起稳定作用，NSR 有功能。

如图 6a,6b 所示,不论磷脂是否具备正面的功能,系统的 NSR 总量都可以维持 100 左右的规模,相对于我们初始接种 NSR 的个数增加了三倍,这个结果不但符合实验室之前的模拟结果,而且说明磷脂不管有无对原始细胞的正面作用都不会影响 NSR 在系统中的繁荣,换句话说,NSR 高度适应本系统。如同我们对原始环境的猜想,最早出现的核酶极大可能就是 NSR,尽管 NSR 不止有利己作用还有利他作用。对照组从接种之后的短暂繁荣数量逐渐下降到了个位数水平,这点佐证了 NSR 的繁荣是因为其功能而非某种偶然性的成功,不具备功能的 RNA 分子最初可以在 NSR 的帮助下扩增,但逐渐因为缺乏功能导致拥有它的原始细胞失去竞争力而被

淘汰。纯脂肪酸膜和 NSR 组成的细胞是最原始的细胞的可能也上升了。

当然图 6c 和 6d 记录的原始细胞总个数在一百万步中的变化说明，磷脂对细胞个数的增加是有帮助的，磷脂出现功能前原始细胞的个数维持在 35 个并上下波动，而磷脂出现功能后原始细胞的个数维持在 45 附近。我们还统计了两个系统的细胞平均大小，结论是没有统计意义上的差异，等价为磷脂对膜的稳定作用导致了两亲性分子在膜上的富集。这个结果说明磷脂和脂肪酸混合膜作为早期原始细胞膜也不无可能，磷脂稳定细胞的作用确实可以影响到整个系统内的原始细胞群落的状态。

3.2 磷脂合成酶在系统中的出现

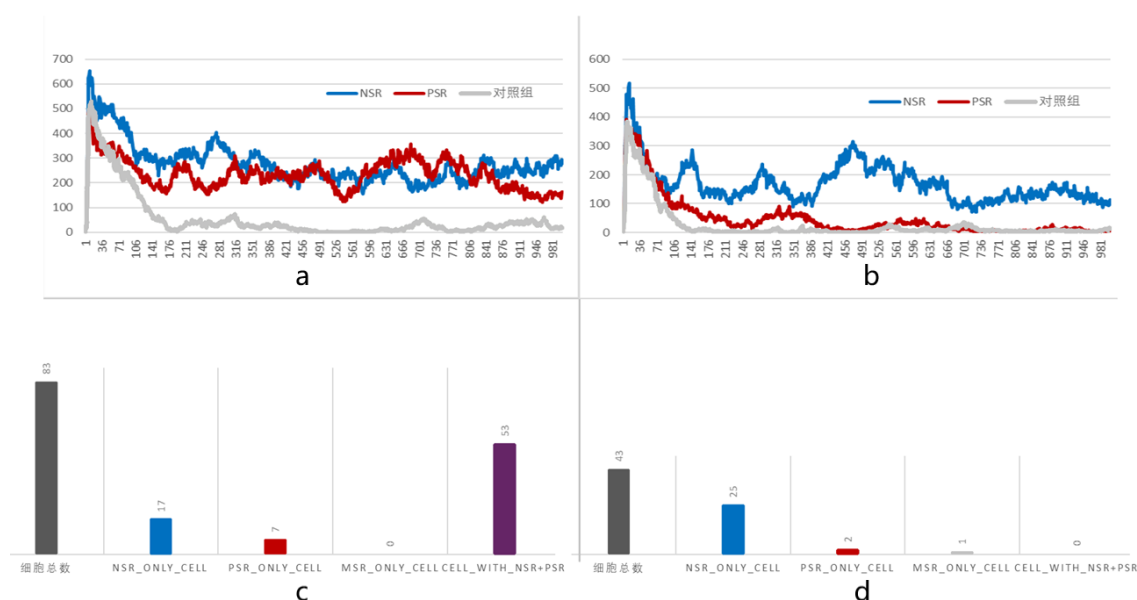


图 7 磷脂合成酶在系统中出现的模拟结果图

图 7 中 (a) (b) 纵坐标代表该序列 RNA 在系统中的总数量，横坐标代表系统运行步数，总计一百万步。NSR 指核苷酸合成酶，PSR 是磷脂合成核酶，对照组是一个随机选取的没有功能的序列。(c) (d) 纵坐标代表系统中该类原始细胞的数量，从左到右依次表示了所有原始细胞数量之和、只有 NSR 没有 PSR 和 MPR 的细胞数量、只有 PSR 没有 NSR 和 MPR 的细胞数量、只有 MPR 没有 PSR 和 MPR 的细胞数量、同时拥有 NSR 和 PSR 的原始细胞的数量。(a) 是 PSR 具备催化功能的实验组，(b) 是 PSR 没有催化功能的对照组。(c) (d) 中实验组与对照组也是相同的含义，值得一提的是这里的 MPR 是没有功能的，仅仅作为对照使用。

如图 7a 所示, 在磷脂可以稳定膜使原始细胞膜上两亲性分子离膜概率降低的情况下, NSR 和 PSR 都能在系统中维持约 200 个的繁荣状态, 无功能的对照组序列在早期借助 NSR 和 PSR 的便利大量复制后逐渐被进化淘汰降低到个位数水准。

图 7b 作为 PSR 无催化作用的对照组, PSR 和其它无效用 RNA 序列一样, 不能在系统中长期存在, 偶尔出现也会很快归零, 这说明 PSR 的繁荣并非偶然事件。对照组系统中 NSR 的总量保持在 100 左右, 虽然远高于接种的 NSR 数量, 但是也远不如 PSR 发挥效用的实验组, 只有实验组的一半不到。PSR 自身的兴旺和对 NSR 发展的助益, 将其作为催化原始细胞膜向磷脂比例增多方向发展的核酶, 这在达尔文进化中具备优势的论点得到了诠释。

为了更进一步了解 PSR 的进化优势, 我们抽样调查了系统内一个时间步上各种 RNA 序列在原始细胞中的分布情况, 随机选取了演化中后期的一个步长统计, 如图 7c 所示, 此时系统趋于稳定, 邻近步数的数值标准差很小, 随意一个步数都可以代表整体情况。可以看到, PSR 的发生使系统的原始细胞总量从 40 变为 80, 几乎翻倍, 这在原始环境中原材料浓度并未增高的前提下是很可观的影响。

结合图 7a、7b 可以得出结论: 不考虑渗透抑制只考虑稳定膜功效的情况下, PSR 对 NSR 和原始细胞种群同时有极大的益处。而对含有各种不同类型排列顺序 RNA 的细胞的统计分析表明: NSR 和 PSR 可能是协同进化的。实验组 NSR 和 PSR 同时具备催化能力, 所以系统内超过六成的原始细胞是同时拥有这两种功能酶而没有在这次模拟中无功能且被视作寄生分子与功能酶 NSR、PSR 同时接种的 MPR 的。剩下的原始细胞几乎都拥有 NSR 或 PSR 其中一种, 这很可能和我们模型中的分裂机制有关——我们假设原始细胞繁殖时遗传信息传递的保真度很低, 一个同时拥有两个 NSR、两个 PSR 的原始细胞分裂后的子代细胞只有十六分之九的可能依然同时拥有这两种 RNA, 原因是没有设定 RNA 均等分配到两个子代细胞的机制, 分配都是随机发生的, 任意一个分子进入哪个子代都是完全相同的概率, 毕竟目前这种功能需要大量的蛋白质辅助, 在 RNA 世界显得过于奢侈了。但是实验中两种 RNA 都有的原始细胞分布的比例已经超过了十六分之九, 可以说明, 这种细胞具备极高的竞争力。图 7c 同时展示了 PSR 无功能对照组的酶分布情况, 只有 NSR 的细胞占据主要生态位, 符合我们的推演: 拥有 NSR 的细胞拥有达尔文进化优势。

除了图 7 展示的数据之外我们还统计了系统内细胞的膜上磷脂比例和平均大小, 如表 2 所示。根据表上数据我们可以知道: 在 PSR 无功能的系统内 PSR 的存在与否, 对细胞大小和膜磷脂和脂肪酸的比例没有影响, 磷脂占比均为 26% 左右,

这个值是合理的。因为系统设定了低于脂肪酸自然合成率的磷脂自然合成率。我们发现磷脂自然合成率 **PPF** 是决定这个比例的核心参数。细胞平均大小也都是在 330—340 之间。而一旦 **PSR** 有了功能，不论细胞是否拥有 **PSR**，系统内的原始细胞膜都将出现磷脂占两亲性分子的比例大幅度上升，无 **PSR** 的细胞这个比值会升到 45%，拥有 **PSR** 的细胞的比值更是会增加到惊人的 94%。这个结果很好解释，**PSR** 的催化效果将会推动系统内脂肪酸和磷酸甘油转化为磷脂的化学平衡向右移动，整个系统内的磷脂都增加了，而拥有 **PSR** 的原始细胞由于膜上脂肪酸分子是直接受到 **PSR** 的影响的，磷脂比例提升极大，没有 **PSR** 的细胞接收了一部分游离的磷脂分子，间接地增加了膜上磷脂的比例，所以比例增加得较少。很有趣的一点：在 **PSR** 出现后的系统内，没有 **PSR** 的细胞大小显著小于有 **PSR** 的细胞，而有 **PSR** 的细胞的大小不会超过 **PSR** 出现之前。这说明“掠夺膜分子”的效应确实存在，通过磷脂对离膜概率 **PALM** 和 **PPLM** 的影响，磷脂不是通过稳定性而将自身所在的囊泡变大获得进化先进性，而是通过掠夺脂肪酸让磷脂比例低的细胞变小得到的进化优势。

为了进一步确认这种掠夺效应的详细机理，实验还追踪了拥有 **PSR** 的原始细胞破裂的情况，结果非常显著。拥有 **PSR** 的细胞几乎不存在因为太小而破裂的情况，细胞内缺失 **PSR** 的细胞则有 10% 的破裂事件是因为细胞膜上两亲性分子的数量低于临界值 **LAM** 而发生的。这个结果很好地解释了“掠夺膜分子”的现象是如何影响到原始细胞的存在。

表 2 拥有 **PSR** 的原始细胞破裂追踪结果

		膜上磷脂比例	平均细胞大小
PSR 无功能系统	有 PSR 的细胞	26.54%	337
	无 PSR 的细胞	26.90%	333
PSR 有功能系统	有 PSR 的细胞	93.89%	314
	无 PSR 的细胞	45.35%	246

3.3 磷脂的渗透抑制效应在系统中的出现

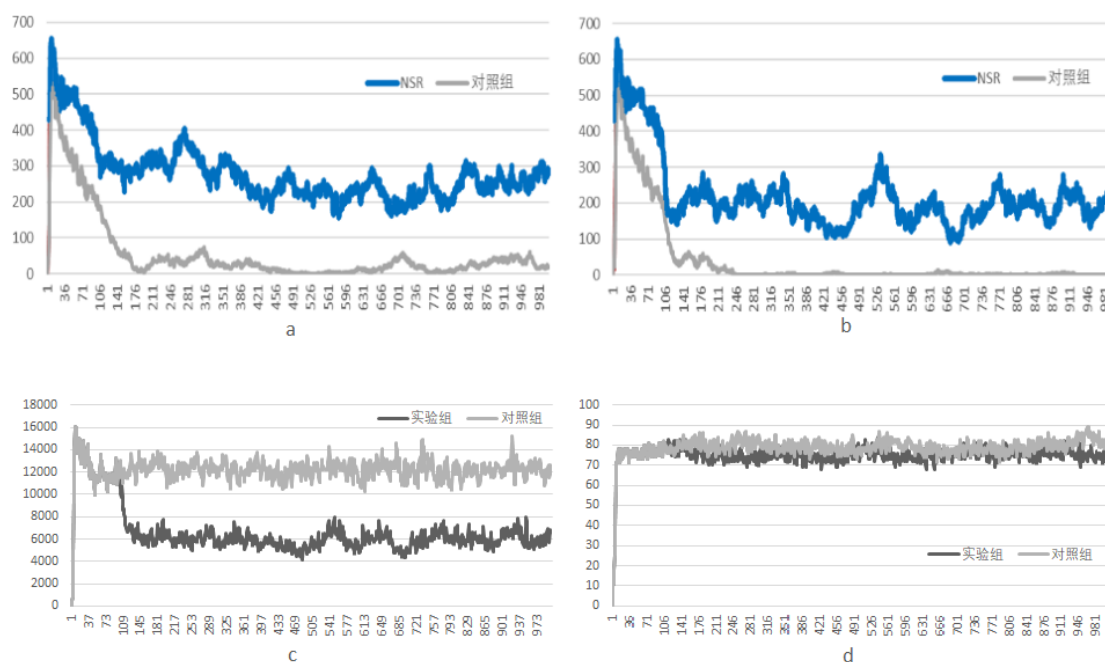


图 8 磷脂的渗透抑制模拟结果图

图 8 中 (a) (b) 纵坐标代表该序列 RNA 在系统中的总数量, 横坐标代表系统运行时间, 总共记录了一百万步。NSR 指核苷酸合成酶, 对照组是一个随机的序列。(a) 是磷脂不会产生对渗透的抑制的系统, (b) 是磷脂对渗透的抑制被激活了的系统。(c) 纵坐标代表系统中所有 RNA 的个数, 不论序列如何, 横坐标代表系统运行时间, 同样是记录了一百万步。(d) 纵坐标指示系统中原始细胞的总量, 横坐标代表系统运行步数。(c) (d) 中的实验组均指代 PSR 和 NSR 均有催化能力, 磷脂同时拥有对核苷酸前体渗透抑制能力和稳定原始细胞膜效力的系统; 对照组则是 PSR 和 NSR 均有功能, 磷脂只拥有稳定原始细胞膜的能力而对核苷酸前体不存在渗透抑制能力的系统。

实验引入了参数 FPP 来模仿现实世界磷脂膜对早期生命获取营养物质的负面效果。这个效果直接作用于核苷酸前体渗透原始细胞的概率, 使磷脂比例越高的细胞与环境的核苷酸交换频率变低。从图 8a、8b 的对比可以看到: 与对照组比较, 实验组的 NSR 总量有一定程度的下滑, 虽然 NSR 在系统内依然是高度繁荣的, 对照组也证明这种繁荣来自其功能而非某个偶然事件。然而实验组约 200 的 NSR 水平和对照组大约 250 的 NSR 水平并不能算是很显著且具备说服力的结果。可以看到图 8a 和 8b 中 NSR 的数量波动都很剧烈, 很多时候实验组甚至比对照组的值更高。换个说法表示, 引入对核苷酸前体渗透抑制能力后, 对 NSR 的影响不太显著。

我们在后续试验里发现 PSR 受到的影响比 NSR 更低，这当然与我们设计试验时预期的结果相距甚远。

为了探究磷脂膜对核苷酸渗透抑制机制对模型代表的原始生命体系是否真的具备明显影响，我们尝试了重复试验并追踪模型其他指标的变化，如图 5c、5d 表示的系统群落细胞总量和体系内 RNA 总量。渗透抑制功能对细胞总量的影响同样不是很明显，平均值下降了不到十个细胞。这个数据我们解释起来并不难，因为两亲性分子受到的影响是间接的。而 RNA 的总量下滑得非常快，这说明磷脂将核苷酸前体拦在膜外确实实地影响了 NSR 获取原材料，虽然 NSR 总量没有太多下降，但是核苷酸前体在原始细胞内的浓度降低了。RNA 总量下降到只有原来的一半左右，是腰斩水平的影响。RNA 总量值也为我们提供了一种解释结果的思路，即 NSR 和 PSR 的效益使拥有它们的原始细胞具备太高的竞争力，以至于它们也可以“掠夺”其他类型 RNA 的原材料。磷脂的副作用大部分都由其它序列的 RNA 承担了。作为受到原材料影响最直接的 NSR，则有一小部分因为 NSR 的进化优势建立在核苷酸供应上；而 PSR 完全不必担心自身竞争力下滑。

3.4 代偿酶 MPR 在系统中的出现

如图 9b 所示，在磷脂对渗透有抑制作用且 PSR、NSR、MPR 都是有功能的前提下，三种功能酶可以在系统内共同繁荣。这意味着同时拥有三种酶的原始生命体是很有可能存在的。为了排除 MPR 的功能无意义只是因为序列或者某种偶然实现了繁荣，我们将它与系统内的对照组交换了序列。也就是认定原来的对照组现在有代偿功能，而原本的 MPR 序列现在只是一个无功能的普通序列，结果见图 9d。换了新序列的 MPR 可以在系统内繁荣，失去了能力的原 MPR 序列则不能长期存在。换言之，被环境选择的是拥有这种能力的序列，重点是序列作为核酶的能力。

为了探究 MPR 被选择是否受其他条件的影响，又分别更改了两个条件作对比。MPR 是针对性地为了缓解磷脂对原始细胞获取核苷酸前体设计的，所以我们称之为“代偿酶”，因此系统内如果磷脂对膜的渗透抑制作用消失的话会如何呢？结果见图 9a：三种功能酶 PSR、NSR、MPR 都有功能的前提下，只有 NSR 和 PSR 在系统里被保留了，MPR 则随着时间进行被淘汰。

确定了渗透代偿功能要在渗透抑制效应存在的情况下发挥作用之后，我们还尝试了虽然磷脂对渗透存在抑制作用，但是 PSR 没有功能的系统。在这个系统内 NSR、MPR 是有能力的，结果如图 9c 所示。随着步长增加，只有 NSR 在系统中

能维持一定水平的数量，PSR 和 MPR 逐渐趋近于消失，即使偶尔出现超过接种数的情况，也会很快回到零点附近。PSR 因为功能消失自然无法在系统内存在，但是有功能的 MPR 竟然也同时失去了在系统内维持繁荣的能力。

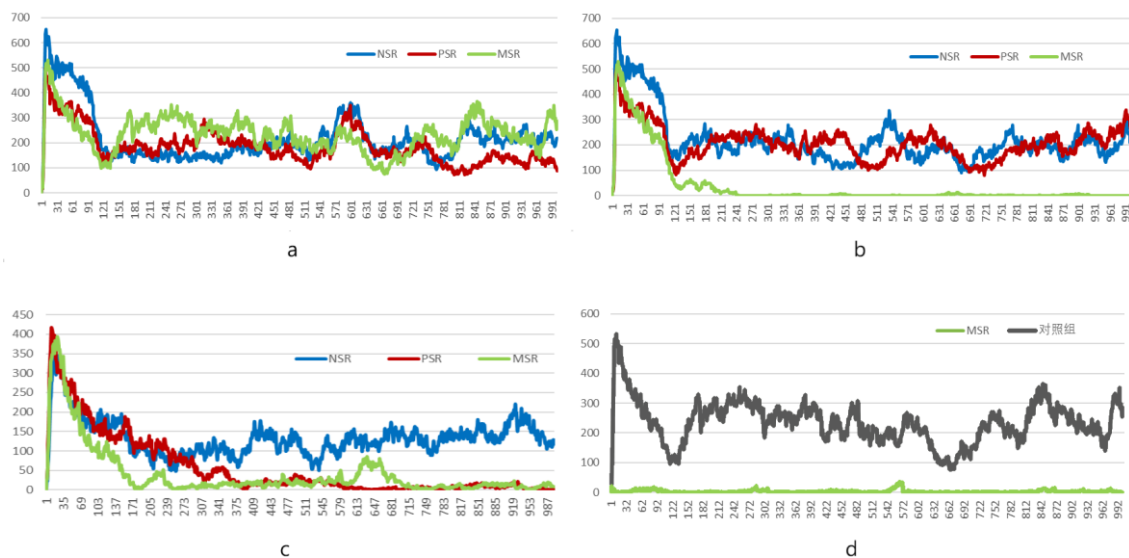


图 9 代偿酶 MPR 的模拟结果图

(a) (b) (c) (d) 纵坐标代表该序列 RNA 在系统中的总数量，横坐标代表系统运行时间，总共记录了一百万步。(a) 是 MPR 无功能的对照组的运行结果。(b) 是 MPR 有功能的实验组的运行结果。(c) 是 MPR 有功能但 PSR 无功能的运行结果。(d) 是在 (a) 的环境下将 MPR 的序列与对照 RNA 序列交换后的结果。

结合图 9a、图 9b、图 9c 我们可以看到，想要 MPR 代表的代偿机制获得达尔文进化的优势，需要 PSR 和渗透抑制机制同时存在，缺一不可。这和人们按照常理推定的结果相符合。MPR 代表的渗透代偿机制出现在 NSR 代表的核苷酸自我合成能力以及 PSR 代表的现代膜分子自我合成能力之后。在生命的最早期,这种功能显得无关紧要，而在磷脂成为膜的主要成分之后，MPR 才开始大放异彩。

第四章 讨论

4.1 模拟结果的综合分析

模型采用了最初始的细胞膜就拥有磷脂成分的观点,并根据之前的实验认为核苷酸合成核酶是最早出现的核酶,其它功能的核酶的出现都建立在核苷酸合成核酶存在的前提下。预实验中我们尝试了单独的磷脂合成酶和单独的渗透代偿酶,它们在核苷酸合成核酶未出现的情况下,都无法出现可以被称作繁荣的数量水平,因而实验流程从核苷酸合成核酶和脂肪酸膜组成的最早期原始细胞开始。然后实验验证了磷脂合成酶和渗透代偿酶在原始细胞中的出现是否能给原始细胞带来进化优势。结果表明:1)拥有核苷酸合成核酶的细胞具有达尔文进化优势,最早出现的核酶极大可能就是核苷酸合成核酶;2)磷脂合成核酶对系统中原始细胞的平均大小有显著影响。它不会使膜磷脂比例高的细胞变大,而是使膜磷脂比例低的细胞变小。3)磷脂膜对早期营养获取的不利影响似乎对核苷酸合成核酶总量和磷脂合成核酶总量影响不大。但是后续试验中代偿酶的共同繁荣说明影响并非不存在,只是以别的形式表现。4)想要渗透代偿核酶代表的代偿机制获得达尔文进化的优势,需要磷脂合成核酶和渗透抑制机制同时存在,缺一不可。最后得出的结论是:磷脂合成酶出现在渗透代偿酶之前。同时拥有核苷酸合成核酶、磷脂合成酶和渗透代偿酶的并以磷脂作为细胞膜主要成分的原始细胞在系统中被选择,成为优势种群,即达尔文进化的胜利者。

虽然只有脂肪酸被证实在原始地球环境中存在,但是脂肪酸和磷酸甘油合成的磷脂也存在无酶催化合成的可能,毕竟酶只是起到催化作用,必须先有非酶合成途径才有酶发挥功能的地方。这种思路下我们设定了一个非单纯的脂肪酸世界作为模拟的第一步。最早大量脂肪酸和少量磷脂存在于原始环境中,在特殊条件下富集到一定浓度,从而组成双层的混合囊泡,此时脂肪酸是膜的主体成分。在囊泡存在的条件下,核苷酸前体自然形成了核苷酸,催化脂肪酸和磷酸甘油合成磷脂使得膜上的磷脂比例变大,膜变得更稳定,但是渗透率也随之下降了。作为磷脂合成核酶副作用的补偿,一种可以增加膜对核苷酸前体渗透性的RNA应当可以增加原始细胞的竞争能力。模拟结果表明,不论磷脂是否具备正面的功能,系统的核苷酸合成核酶总量都可以维持远超接种数量的规模,相对于我们初始接种核苷酸合成核酶的个数增加。这一切证明:磷脂不管有无对膜的稳定作用,都不会

影响核苷酸合成核酶在系统中的繁荣。核苷酸合成核酶的细胞占据主要生态位，说明拥有核苷酸合成核酶的细胞拥有达尔文进化优势。最早出现的核酶极大可能就是核苷酸合成核酶，尽管核苷酸合成核酶不止有利己作用还有利他作用。而对对照组从接种之后的短暂繁荣数量逐渐下降到了个位数水平，则进一步佐证了：核苷酸合成核酶的繁荣是因为其功能而非某种偶然性。不具备功能的 RNA 分子最初可以在核苷酸合成核酶的帮助下扩增，但逐渐因为缺乏功能导致拥有它的原始细胞失去竞争力从而被淘汰。因此，纯脂肪酸膜和核苷酸合成核酶组成的细胞是最原始的细胞的可能性上升了。

磷脂对系统中原始细胞的数量有一定的影响。功能性磷脂出现后，系统内原始细胞的总量显著增加。但我们计算了两个系统的平均细胞大小，得出结论认为：它们没有统计学意义上的显著差异。这对应于磷脂对膜的稳定作用，导致膜上两亲分子的总量上升，游离的两亲性分子数量降低。这个结果表明磷脂和脂肪酸的混合膜作为最早期的原始细胞膜并非不可能。核苷酸合成核酶、磷脂合成核酶在系统内可以做到维持大约 200 个同类分子的繁荣度，前提是磷脂可以稳定膜并降低原始细胞膜上两亲分子脱离的可能性。否则磷脂合成酶将逐渐被淘汰，降到个位数的水平。在磷脂合成核酶无功能的对照组中，磷脂合成核酶与其他无效 RNA 序列一样，不能在系统中长期存在，偶尔出现就归零。这说明这里磷脂合成核酶的繁荣是偶然的事件。对照系统中，核苷酸合成核酶的总量远高于接种核苷酸合成核酶的数量，但不如磷脂合成核酶。磷脂合成核酶本身的繁荣及其对核苷酸合成核酶发展的贡献足以解释：达尔文进化选择了这种核酶，该核酶可催化质膜向增加磷脂比例的方向发展。

磷脂合成核酶对系统中原始细胞的平均大小有显著影响。它不会使膜磷脂比例高的细胞变大，而是使膜磷脂比例低的细胞变小。因为这种酶会在膜上行使催化功能，使无磷脂合成核酶的细胞膜上磷脂的比例相对低很多。在磷脂合成核酶的直接影响下，可以看到膜替代的发生。当磷脂合成核酶起作用时，受磷脂的合成速率限制，无磷脂合成核酶的细胞膜上磷脂的比例大约是一半；而有磷脂合成核酶的细胞的膜几乎是完全的磷脂膜了。在磷脂合成核酶已经出现的系统中，没有磷脂合成核酶的单元的大小，明显小于有磷脂合成核酶的单元的大小；而有磷脂合成核酶的单元的大小，不超过磷脂合成核酶出现之前的大小。这是很有趣的：磷脂对分离概率的影响，是通过掠夺其他磷脂比例低的细胞的脂肪酸获得的，而不是通过稳定性扩大其囊泡来获得进化优势。为了进一步证实这种效应的详细机制，实验还跟踪了原始细胞的破裂次数及原因。具有磷脂合成核酶的细胞很少破

裂,即使破裂也是因为随机概率破裂,而非因为它的大小低于临界值。而缺乏磷脂合成核酶的细胞,有较大的可能因为细胞膜上的两亲分子数量低于临界值而破裂。进一步观察磷脂合成核酶可以看到,磷脂合成核酶的发生使系统的原始细胞总量翻倍,这在原始环境中原材料浓度并未增高的前提下是很可观的影响。不考虑渗透抑制只考虑稳定膜功效的情况下,磷脂合成核酶对核苷酸合成核酶和原始细胞种群同时有极大的益处。而对含有各种不同类型排列顺序 RNA 的细胞的统计分析表明:核苷酸合成核酶和磷脂合成核酶可能是协同进化的。实验组核苷酸合成核酶和磷脂合成核酶同时具备催化能力,所以系统内超过六成的原始细胞是同时拥有这两种功能酶。而在这次模拟中,无功能且被视作寄生分子与功能酶核苷酸合成核酶、磷脂合成核酶同时接种的渗透代偿核酶则不能繁荣。

渗透代偿核酶无法脱离另外两种合成酶独立繁荣。渗透代偿核酶是针对性地为了缓解磷脂对原始细胞获取核苷酸前体的抑制作用而设计的。如果磷脂对膜的渗透抑制作用消失的话,只有核苷酸合成核酶和磷脂合成核酶在系统里被保留了,渗透代偿核酶则随着时间进行被淘汰。所以渗透代偿功能要在渗透抑制效应存在的情况下发挥作用。之后,我们还尝试了磷脂对渗透存在抑制作用,但是磷脂合成核酶没有功能的系统。在这个系统内,核苷酸合成核酶,渗透代偿核酶是有能力的。随着步长增加,只有核苷酸合成核酶在系统中能维持一定水平的数量,磷脂合成核酶和渗透代偿核酶逐渐趋近于消失,即使偶尔出现超过接种数的情况,也会很快回到零点附近。磷脂合成核酶因为功能消失自然无法在系统内存在,但是有功能的渗透代偿核酶竟然也同时失去了在系统内维持繁荣的能力。我们可以看到,想要渗透代偿核酶代表的代偿机制获得达尔文进化的优势,需要磷脂合成核酶和渗透抑制机制同时存在,缺一不可。在磷脂具有抑制渗透的条件下,如果磷脂合成核酶、核苷酸合成核酶和渗透补偿核酶都是功能性的,含有三种酶的原始细胞会因为具备进化优势而成为系统内的主要优势种。我们与系统中的对照组交换了序列,排除了渗透补偿核酶的繁荣是偶然的可能。综上所述,渗透代偿核酶代表的渗透代偿机制,出现在核苷酸合成核酶代表的核苷酸自我合成能力以及磷脂合成核酶代表的现代膜分子自我合成能力之后。在生命的最早期,这种代偿功能显得无关紧要,而在磷脂成为膜的主要成分之后,原始细胞对渗透代偿核酶才有了需要。

4.2 模拟中意料之外的结果

磷脂膜对早期营养获取的不利影响似乎对核苷酸合成核酶总量和磷脂合成核酶总量影响不大，这和人们按照常理推定的结果不符。但是后续试验中代偿酶的繁荣说明影响并非不存在，只是以别的形式表现。实验跟踪模型的另一个指标——RNA 总量的变化，判断磷脂膜的核苷酸渗透抑制机制是否真的对模型所代表的原始细胞系统没有影响。结果显示虽然系统群落中的 RNA 总量、渗透抑制对总细胞量的影响不明显，但 RNA 的总量下降得非常快，原始细胞中核苷酸前体的浓度降低，RNA 总量减少到原始量的一半左右。这意味着虽然核苷酸合成核酶的总量没有明显减少，但磷脂将核苷酸前体从膜上阻断，实际上影响了核苷酸合成核酶获得原料。一个解释是核苷酸合成核酶和磷脂合成核酶可以使拥有它们的原始细胞非常具有竞争力和“掠夺性”，使其他类型 RNA 的原材料被抢走，也就是说渗透率降低的代价被转移给了其他的无功能 RNA。

在模拟渗透代偿核酶时，同时拥有渗透代偿核酶和磷脂合成酶而缺少核苷酸合成酶的原始细胞数量，与同时拥有核苷酸合成酶和磷脂合成酶而缺少渗透代偿核酶的原始细胞数量没有显著差异。这显然不合常理，因为渗透代偿核酶和磷脂合成酶在系统缺乏核苷酸合成酶时都是不能繁荣的，而核苷酸合成酶和磷脂合成酶在系统缺少渗透代偿核酶的情况下可以共同繁荣，也就是说后者类型的细胞理论上应比前者更有优势。不过意料之中的是系统内同时拥有三种功能酶的细胞大约占 40%，是最具有优势的种群。我们推测这个结果和我们模型中的分裂机制有关。随机分配内容物的分裂机制，使具备优势的同时拥有三种功能酶的细胞不能稳定地将完整遗传信息传递给后代，这使得即使系统中只有这种细胞能分裂，丢失关键酶的细胞还是会一直产生，并在系统中占据一定比例。

4.3 模拟的前提条件与模型的局限性

关于模型使用的参数，我们在设置了逻辑约束下参数间的大小关系和取值区间之后，使用了机器学习里的梯度法学习，得出了一套具体的参数值，方法原理见附录 1 论文。这套参数应该是模型的最优解之一，最利于我们系统中原始细胞种群的繁荣。我们还对这套参数进行了微调，可以得到类似的结果，证明了系统的鲁棒性。由于使用的概率性参数过多，我们只使用了机器学习三十轮次的结果，类似的最优参数不止我们在表 1 中展示的这一套。我们也不能保证没有合乎逻辑的更优解存在，但是这套参数下的模拟已经足够证明我们的猜想。

在我们参数设置的原始环境下，原始细胞是可以发生进化的。之前的实验中，已经通过模拟证明了脂肪酸囊泡内的稳定环境，使得核苷酸合成核酶能接触更多原材料，核苷酸降解概率变低，和核苷酸合成核酶结合的囊泡获得进化优势。这个结论换到脂肪酸及磷脂混合囊泡系统中也是成立的。

关于膜和核酶谁最先出现的这个争议在系统中是被搁置了的，脂肪酸膜和磷脂膜谁先出现的这个问题也是一样。我们假定最早的膜是脂肪酸占绝对主体的混合膜，逐渐演变成了磷脂占主体的混合膜，模拟结果证明了几乎完全以磷脂分子作为膜成分的细胞在我们的系统内逐渐替代了纯粹以脂肪酸作为膜分子的原始细胞。它们是在核苷酸合成核酶、磷脂合成酶和渗透代偿酶的帮助之下实现的。这说明膜替代极有可能是出现在生命起源早期的现象，从磷脂合成核酶出现起磷脂膜就成为了主流。

模型中的细胞分裂机制是高度简化的。真核生物和原核生物都是通过染色体的机制，将基因连在一起，使分裂的两个子细胞能最大可能地获得一样的遗传信息。但是与染色体配套的转录 RNA 剪切机制对于 RNA 世界中的原始细胞太过复杂，一般认为它发生在基因编码翻译蛋白质机制之后，所以我们使用了简化的分裂机制。这个简化机制中 RNA 分配和其它分子一样都是随机发生的。任意一个分子进入哪个子代都是完全相同的概率，毕竟现代细胞的分裂需要大量的蛋白质辅助，在 RNA 世界显得过于奢侈了。所以我们的系统中，一个同时拥有多种核酶的原始细胞分裂后的子代细胞，很难依然同时拥有这几种 RNA。这使我们系统内原始细胞繁殖时遗传信息传递的保真度过低，即使是进化优势很高的细胞，也不能完全占领整个原始环境。因为它的后代有很大概率丢失功能酶中的一种。我们认为：自发组装的囊泡内随机获得复制分子是最初的细胞遗传方式。

4.4 后续模拟研究的展望

关于我们对渗透代偿酶的定义过于抽象这一点，未来的实验可以通过引入蛋白质来解决。模拟中设定的这种渗透代偿核酶只有 6nt 长且专一协助核苷酸前体过膜，这是因为在所有现代细胞中见到的过膜辅助机制几乎全部依赖蛋白质。不直接依赖蛋白质的机制例如胞吞和胞吐，在现代细胞中一样和蛋白质息息相关。然而具备特殊功能的蛋白质不太可能在早期的 RNA 世界大量存在。越早的生命应该是越简单的。而蛋白质的单体有二十多种，期待非酶催化下能大量合成某种序列的蛋白质几乎是天方夜谭。即使有了氨基酸连接酶，没有模板指导的蛋白质复制

效率一样低得可怕。另外一个原因是蛋白质翻译相关的基因保守度高到了极点，密码子翻译表的保守度也是一样。因而我们很难判断成规模的翻译机制是在什么时候出现的。所以这里我们设置了一种过渡的核酶代替蛋白质发挥作用。现在有了一种学说假设了一种最简单的翻译机制，这种机制下翻译所需的结构复杂度远低于现在细胞的翻译系统，未来我们可以改良现有模型，验证蛋白质组成的渗透代偿酶是否可以繁荣。

未来我们还可以尝试引入最简单的剪切酶，看其对遗传的作用能否克服其消耗资源的劣势。实验室之前的工作使用类似的蒙特卡洛模型用计算机模拟了围绕标签驱动开发的进化动力学。标签机制是一种假说，这个机制使复制酶可以抵抗和繁殖寄生分子。标签机制与空间限制机制分别为复制酶逃离寄生分子提供时间和空间，二者还有协同作用。细胞膜带来的空间限制使原始细胞中的酶远离很多寄生，同时细胞的随机分裂机制使寄生分子有不低的可能和酶分离，这保护了复制子，减少了复制子受竞争者和寄生序列的影响。但是随机分裂机制可能会使细胞的选择优势消失。染色体配合原始剪切酶可以保护多基因细胞的性状保留，但是会一定程度降低对寄生分子的抗性。

参考文献

- [1] Meng Wu, Paul G Higgs. 2009. Origin of self-replicating biopolymers: autocatalytic feedback can jump-start the RNA world[J]. *Journal of Molecular Evolution*: 69:541-554.
- [2] Meng Wu, Paul G Higgs. 2011. Comparison of the roles of nucleotide synthesis, polymerization and recombination in the origin of autocatalytic sets of RNAs[J]. *Astrobiology*: 11:895-906.
- [3] Julie A, Shay Christopher Huynh, Paul G. Higgs. 2015. The origin and spread of a cooperative replicase in a prebiotic chemical system[J]. *Journal of Theoretical Biology*: 364: 249-259.
- [4] Wentao Ma, Chunwu Yu, Wentao Zhang, Jiming Hu. 2010. A simple template-dependent ligase ribozyme as the RNA replicase emerging first in the RNA world[J]. *Astrobiology*: 10(4):437-447.
- [5] Stéphane Tirard. 2017. J. B. S. Haldane and the origin of life [J]. *Journal of Genetics*: 96(5):735-739.
- [6] Iris Fry. 2006. The origins of research into the origins of life[J]. *Endeavour*: 30(1):24-28.
- [7] Antonio Lazcano. 2010. Historical development of origins research[J]. *Cold Spring Harbor Perspect Biol*: 2(11): a002089.
- [8] James Attwater, Aniela Wochner, Philipp Holliger. 2013. In-ice evolution of RNA polymerase ribozyme activity[J]. *Nature Chemistry*: 5(12):1011-1018.
- [9] Alexandr I. Oparin. 1953. *Origin of Life* [M]: 173-175.
- [10] Walter Gilbert. 1986. Evolution of antibodies. The road not taken [J]. *Nature*: 320(6062):485-6.
- [11] Jack W. Szostak. 2011. An optimal degree of physical and chemical heterogeneity for the origin of life[J]? *Biological Science*: 366(1580):2894-2901.
- [12] R.M.de Graaf, J. Visscher, Alan W. Schwartz. 1995. A plausibly prebiotic synthesis of phosphonic acids[J]. *Nature*:378(6556):474-477.
- [13] George Cooper, Novelle Kimmich, Warren Belisle, Josh Sarinana, Katrina Brabham, Laurence Garrel. 2001. Carbonaceous meteorites as a source of sugar-related organic compounds for the early Earth[J]. *Nature*: 20-27.
- [14] JM Gebicki, M Hicks. 1973. Ufasomes are stable particles surrounded by unsaturated fatty acid membranes[J]. *Nature*: 243(5404):232-234.
- [15] Gebicki JM, Hicks M. 1976. Preparation and properties of vesicles enclosed by fatty acid membranes[J]. *Chemistry and Physics of Lipids*:16: 142-146.
- [16] Willia R Hargreaves, David W Deamer. 1978. Liposomes from ionic single-chain amphiphiles[J]. *Biochemistry*: 17: 3759-3768.
- [17] Anne-Laure Fameau, Audrey Arnould, Arnaud Saint-Jalmes. 2014. Responsive self-assemblies based on fatty acids[J]. *Chemical Communications*: 19:471-479.
- [18] Xianfeng Zhou, Punam Dalai, Nita Sahai. 2020. Semipermeable Mixed Phospholipid-Fatty Acid Membranes Exhibit K⁺/Na⁺ Selectivity in the Absence of Proteins[J]. *Life (Basel)*:

10(4):39.

- [19] Jonathan Lombard, Purificación López-García, David Moreira. 2012. The early evolution of lipid membranes and the three domains of life[J]. *Microbiol*: 10(7):507-515.
- [20] Itay Budin, Jack W Szostak. 2011. Physical effects underlying the transition from primitive to modern cell membranes[J]. *Proc Natl Acad Sci U S A*:108(13):5249-5254.
- [21] Michael O Gaylor. 2021. Plausible Emergence and Self Assembly of a Primitive Phospholipid from Reduced Phosphorus on the Primordial Earth[J]. *Orig Life Evol Biosph*: 51(3):185-213.
- [22] Sheref S Mansy, Jason P Schrum, Mathangi Krishnamurthy, Sylvia Tobé Douglas A Treco, Jack W Szostak. 2008. Template-directed synthesis of a genetic polymer in a model protocell[J]. *Nature*: 454:122-125.
- [23] Irene A. Chen, Martin M. Hanczyc, Peter L. Sazani, Jack W. Szostak. 2006. Protocells: Genetic polymers inside membrane vesicles[J]. *Cold Spring Harbor Monograph Archive*:43.
- [24] Sheref S. Mansy. 2010. Membrane Transport in Primitive Cells[J]. *Cold Spring Harb Perspect Biol*:2(8): a002188.
- [25] Wentao Ma, Jiming Hu. 2012. Computer simulation on the cooperation of functional molecules during the early stages of evolution[J]. *PLoS One*:7(4): e35454.
- [26] Tamás Czárán, Balázs Könnyű, Eörs Szathmáry. 2015. Metabolically Coupled Replicator Systems: Overview of an RNA-world model concept of prebiotic evolution on mineral surfaces[J]. *Journal of Theoretical Biology*: 381:39-54.
- [27] Wentao Ma, Chunwu Yu, Wentao Zhang. 2007. Monte Carlo simulation of early molecular evolution in the RNA World[J]. *Biosystems*: 90(1):28-39.
- [28] Alexandru Agapie, Marius Giuclea, Anca Andreica. 2014. Probabilistic Cellular Automata[J]. *Journal of Computational Biology*: 21(9): 699-708.
- [29] Juli Peretó, Purificación López-García, David Moreira. 2004. Ancestral lipid biosynthesis and early membrane evolution[J]. *Trends in Biochemical Sciences*: 29(9):469-477.
- [30] Elias Zintzaras, Mauro Santos, Eörs Szathmáry. 2010. Selfishness versus functional cooperation in a stochastic protocell model[J]. *Journal of Theoretical Biology*: 267(4):605-613.

攻读硕士学位期间已发表的学术论文

1、Yuzhen Liang, Chunwu Yu, Wentao Ma. 2021. The automatic parameter-exploration with a machine-learning-like approach: Powering the evolutionary modeling on the origin of life. PLOS COMPUTATIONAL BIOLOGY: 17(12):e1009761.

附录

本论文主题相关的关键代码，主要为膜上磷脂比例影响脂肪酸解离率和核苷酸前体渗透率的实现代码。

```
//----- Amphiphilic molecule moving    脂肪酸的移动

for (y = 0; y<N; y++)
{
    for (x = 0; x<N; x++)
    {
        if (m_arr[0][y][x] == 0 || m_arr[0][y][x] - m_arr[1][y][x] == 0)continue;
// No Amphiphilic molecules on membrane
        tmr_xy = 0;
        for (p = room_head[0][y][x]->next; p != room_head[0][y][x]; p = p->next)
        {
            tmr_xy += p->length1 + p->length2;
        }
        // To consider the effect of osmotic pressure caused by the impermeable materials in the protocell
        m_bef = m_arr[0][y][x] - m_arr[1][y][x];
        palm = PALM * (9.3 / (1 + pow((m_arr[1][y][x] / m_arr[0][y][x]) / 36, 0.7)) - 1.2) / 10;
        // 膜上磷脂比例影响解离率
        for (k = 0; k<m_bef; k++)
        {
            if (randd()<palm / (1 + tmr_xy / pow(m_arr[0][y][x] / 2.0, 1.5)))
            {
                m_arr[0][y][x]--; // Leaving
                if (randd()<0.5) a_arr[1][y][x]++; // Going into the protocell
                else // Going out of the protocell
                {
                    randcase = randl(4); // Four possible directions
                    switch (randcase)
                    {
                        case 0:
                            if (m_arr[0][y][(N + x - 1) % N] == 0) a_arr[1][y][(N + x - 1) % N]++;
                            else m_arr[0][y][x]++;
                            break;
                        case 1:
                            if (m_arr[0][y][(x + 1) % N] == 0) a_arr[1][y][(x + 1) % N]++;
                            else m_arr[0][y][x]++;
                            break;
                        case 2:
                            if (m_arr[0][(N + y - 1) % N][x] == 0) a_arr[1][(N + y - 1) % N][x]++;
                            else m_arr[0][y][x]++;
                            break;
                        case 3:
                            if (m_arr[0][(y + 1) % N][x] == 0) a_arr[1][(y + 1) % N][x]++;
                            else m_arr[0][y][x]++;
                            break;
                        default: printf("amphiphilic molecule leaving error");
                    }
                }
            }
        }
    }
}
}
```

```

//----- Nucleotide precursors moving 核苷酸前体的移动
for (y = 0; y<N; y++)
{
    for (x = 0; x<N; x++)
    {
        if (i<MECH_CHANGE_STEP) fpp = 0, ftr = 0;
        else fpp = FPP, ftr = FTR;
        if (m_arr[0][y][x]>0) // Membrane existing
        {
            // Possibly moving and permeating out
            pn_bef = pn_arr[0][y][x];
            pn_arr[0][y][x] = 0;

            mr_turn = 1;
            for (p = room_head[0][y][x]->next; p != room_head[0][y][x]; p = p->next)
            {
                // Finding membrane permeating synthetase ribozymes in the room
                flag1 = findseq(mrseq, mrlength, p);
                if (flag1 == 0 && p->length2 == 0) mr_turn += ftr;
            }
            // Calculating possible catalysis turns in the room
            for (k = 0; k<pn_bef; k++)
            {
                // 渗透代偿酶和膜上磷脂比例影响渗透率
                if (randd()<PMV&&randd()<PNPP / (1 + (fpp / mr_turn)* (m_arr[1][y][x] / m_arr[0][y][x])))
                {
                    randcase = randl(4); // Four possible directions
                    switch (randcase)
                    {
                        case 0:
                            if (m_arr[0][y][(N + x - 1) % N] == 0) pn_arr[1][y][(N + x - 1) % N]++;
                            else pn_arr[1][y][x]++;
                            break;
                        case 1:
                            if (m_arr[0][y][(x + 1) % N] == 0) pn_arr[1][y][(x + 1) % N]++;
                            else pn_arr[1][y][x]++;
                            break;
                        case 2:
                            if (m_arr[0][(N + y - 1) % N][x] == 0) pn_arr[1][(N + y - 1) % N][x]++;
                            else pn_arr[1][y][x]++;
                            break;
                        case 3:
                            if (m_arr[0][(y + 1) % N][x] == 0) pn_arr[1][(y + 1) % N][x]++;
                            else pn_arr[1][y][x]++;
                            break;
                        default: printf("pn moving error");
                    }
                }
                else pn_arr[1][y][x]++;
            }
        }
        else // no membrane
        {
            // Possible moving, possibly permeating into a protocell at an adjacent room
            pn_bef = pn_arr[0][y][x];
            pn_arr[0][y][x] = 0;
            for (k = 0; k<pn_bef; k++)
            {
                if (randd()<PMV)
                {
                    randcase = randl(4); // Four possible directions

```

```

switch (randcase)
{
    case 0:
        if (m_arr[0][y][(N + x - 1) % N] == 0) pn_arr[1][y][(N + x - 1) % N]++
// Moving
        else{
            tmr_xy = 0;
            for (p = room_head[1][y][(N + x - 1) % N]->next; p !=
room_head[1][y][(N + x - 1) % N]; p = p->next)
                tmr_xy += p->length1 + p->length2;
// Caculating total materials of nucleotides and RNA in the target protocell

// To consider the effect of Donnan's equilibrium

            mr_turn = 1;
            for (p = room_head[1][y][(N + x - 1) % N]->next; p !=
room_head[1][y][(N + x - 1) % N]; p = p->next)
            {
// Finding membrane permeating synthetase ribozymes in the room
                flag1 = findseq(mrseq, mrlength, p);
                if (flag1 == 0 && p->length2 == 0) mr_turn += ftr;
// Caculating possible catalysis turns in the room
            }

            if (randd() < (PNPP / (1 + (fpp / mr_turn) * (m_arr[1][y][(N + x -
1) % N] / m_arr[0][y][(N + x - 1) % N])) * (m_arr[0][y][(N + x - 1) % N] / LAM) / (1.0 + FSI * tmr_xy / pow(m_arr[0][y][(N
+ x - 1) % N] / 2.0, 1.5)))
                pn_arr[1][y][(N + x - 1) % N]++;
            else pn_arr[1][y][x]++;
        }
        break;
    case 1:
        if (m_arr[0][y][(x + 1) % N] == 0) pn_arr[1][y][(x + 1) % N]++;
        else
        {
            tmr_xy = 0;
            for (p = room_head[1][y][(x + 1) % N]->next; p !=
room_head[1][y][(x + 1) % N]; p = p->next)
                tmr_xy += p->length1 + p->length2;

            mr_turn = 1;
            for (p = room_head[1][y][(x + 1) % N]->next; p !=
room_head[1][y][(x + 1) % N]; p = p->next)
            {
// Finding membrane permeating synthetase ribozymes in the room
                flag1 = findseq(mrseq, mrlength, p);
                if (flag1 == 0 && p->length2 == 0) mr_turn += ftr;
// Caculating possible catalysis turns in the room
            }

            if (randd() < (PNPP / (1 + (fpp / mr_turn) * (m_arr[1][y][(x + 1) %
N] / m_arr[0][y][(x + 1) % N])) * (m_arr[0][y][(x + 1) % N] / LAM) / (1.0 + FSI * tmr_xy / pow(m_arr[0][y][(x + 1) %
N] / 2.0, 1.5)))
                pn_arr[1][y][(x + 1) % N]++;
            else pn_arr[1][y][x]++;
        }
        break;
    case 2:
        if (m_arr[0][(N + y - 1) % N][x] == 0) pn_arr[1][(N + y - 1) % N][x]++;
        else
        {

```

```

tmr_xy = 0;
for (p = room_head[1][(N + y - 1) % N][x]->next; p !=
room_head[1][(N + y - 1) % N][x]; p = p->next)

    tmr_xy += p->length1 + p->length2;

mr_turn = 1;
for (p = room_head[1][(N + y - 1) % N][x]->next; p !=
room_head[1][(N + y - 1) % N][x]; p = p->next)
{
// Finding membrane permeating synthetase ribozymes in the room
flag1 = findseq(mrseq, mrlength, p);
if (flag1 == 0 && p->length2 == 0) mr_turn += ftr;
// Caculating possible catalysis turns in the room
}

if (randd() < (PNPP / (1 + (fpp / mr_turn) * (m_arr[1][(N + y - 1) %
N][x] / m_arr[0][(N + y - 1) % N][x])) * (m_arr[0][(N + y - 1) % N][x] / LAM) / (1.0 + FSI * tmr_xy / pow(m_arr[0][(N
+ y - 1) % N][x] / 2.0, 1.5)))

    pn_arr[1][(N + y - 1) % N][x]++;
else pn_arr[1][y][x]++;
}
break;
case 3:
if (m_arr[0][(y + 1) % N][x] == 0) pn_arr[1][(y + 1) % N][x]++;
else
{
tmr_xy = 0;
for (p = room_head[1][(y + 1) % N][x]->next; p != room_head[1][(y
+ 1) % N][x]; p = p->next)

    tmr_xy += p->length1 + p->length2;

mr_turn = 1;
for (p = room_head[1][(y + 1) % N][x]->next; p != room_head[1][(y
+ 1) % N][x]; p = p->next)
{
// Finding membrane permeating synthetase ribozymes in the room
flag1 = findseq(mrseq, mrlength, p);
if (flag1 == 0 && p->length2 == 0) mr_turn += ftr;
// Caculating possible catalysis turns in the room
}

if (randd() < (PNPP / (1 + (fpp / mr_turn) * (m_arr[1][(y + 1) %
N][x] / m_arr[0][(y + 1) % N][x])) * (m_arr[0][(y + 1) % N][x] / LAM) / (1.0 + FSI * tmr_xy / pow(m_arr[0][(y + 1) %
N][x] / 2.0, 1.5)))

    pn_arr[1][(y + 1) % N][x]++;
else pn_arr[1][y][x]++;
}
break;
default: printf("pn moving error");
}
}
else pn_arr[1][y][x]++;
}
}
}

```

致 谢

在撰写论文的过程中,我得到了马文涛老师的大力支持和协助,在此谨向马文涛老师表达深深的谢意。我本科专业是生命科学,在计算机领域的知识只够入门所需,因此在尝试踏入计算生物学领域时遇到了许多的阻碍。非常幸运的是,在马文涛老师亲自教导下,我从掌握基本的计算机术语到学会更复杂的编程。这个过程耗时甚久,花费了老师大量的时间与精力。

在同一个实验室学习的师兄们留下了大量的实验记录,这些记录为我构思论文与分析结果提供了许多参考。有深厚的计算机基础的师弟也为模型运行提供了很多帮助,在此我想表达我对大家的谢意。

正式开始进行实验的时间正好是新冠疫情在武汉出现的时间,疫情导致了在长达半年的时间里课程的学习与实验的准备都无法在学校进行,学生只能进行线上教学。这段时间我在课程老师与班级同学的帮助下才得以完成培养计划。

我十分感谢我的同学对我工作的支持,也特别感谢我的家人对我学习与研究的支持。

梁瑜祯

2022 年 5 月于武汉