

蛋白质分子网络的分形结构

王建平, 柴立和 (天津大学环境科学与工程学院, 天津300072)

摘要 在生命的新陈代谢和自我复制过程中, 蛋白质起着异乎寻常的作用。随着分形概念及理论的广泛应用和向分子生物学的渗透, 人们从单个蛋白质的研究上升到对蛋白质网络的研究。利用最大流原理从动力学角度探讨了蛋白质大分子网络的分形结构形成过程的机理。

关键词 蛋白质网络; 自组织; 最大流原理; 分形

中图分类号 Q502 文献标识码 A 文章编号 0517- 6611(2007)06 - 01580 - 02

Discussion on the Protein Macro-molecule Network Fractal Structure

WANG Jian ping et al (College of Environmental Science and Engineering, Tianjin University, Tianjin 300072)

Abstract In life metabolism and self-copy process the protein is playing the unusual role. Since the middle period of 19th century, the scholar always has not stopped to study the protein fractal characteristic rigid exploration. Along with the widespread of fractal concept and theory and its application in molecular biology, the research has been richened from the single protein to the protein network. In this article from the maximal-flux principle driven mechanics, the mechanism of the formed process of protein macro-molecule network fractal structure was discussed.

Key words Protein networks; Self-organization; Maximum flux principle; Fractal

蛋白质分子网络本质上是由大量子系统(残基或蛋白质分子)组成的开放复杂系统, 其子系统之间存在复杂的非线性相互作用, 因此可以利用最大流原理来研究。

1 最大流原理的引入

最大流原理可以表述为: 一个远离平衡的复杂系统总是努力寻找一种优化过程使得系统在给定的约束或代价下所获得的广义流最大^[2]。设任意一个复杂开放的蛋白质网络系统由 n 个子系统组成, 每个子系统演化的驱动力, 即从外界环境获得的广义流(物质流、能量流、信息流等)的动力(内禀增长率、需求率和获得率等), 其所有组元在空间内形成一个连续的单元, $dx = dx_1 dx_2 dx_n$ 为空间内的一个体积单元。设 t 时刻该系统处于体积单元时获得的广义流是 J , 则所有可能组元获得的平均广义流表示为:

$$\bar{J} = \int (x, t) J(x) dx \quad (1)$$

利用 Lagrange 乘算法, 使方程(1) 在下列约束(表示给定的代价) 下取极大值

$$(x_i) = f_1, (x_i x_j) = f_2, (x_i x_j x_k) = f_3, (x_i x_j x_k x_l) = f_4 \quad (2)$$

$$\text{经过相应的变换 } x_i = \sum_k k_i x_k \text{ 或 } i = \sum_k k_i x_k \quad (3)$$

可以得到取得最大广义流的概率密度函数^[2-3]:

$$= e^{\mu + \sum_i i x_i + \sum_{ij} ij x_i x_j + \sum_{ijk} ijk x_i x_j x_k + \sum_{ijkl} ijkl x_i x_j x_k x_l} + L \quad (4)$$

式(4) 的指数可看成一个势函数, 它表征了参考态的渐进稳定特性。最大流原理表述为以最小的代价获得最大的广义流, 实际上表现了蛋白质分子网络聚合成稳定结构的行为。

2 网络的选择和网络基序的形成

由于 x_i 为蛋白质分子, 那么由关系式(2) 可知, 就是代表一种稳定的模式- 蛋白质分子网络结构。关系式(2) 对应的网络结构, 在只有4 个组元组成的情况下可用图1 来表示:

由于 x_i 代表网络上的结点, 可以认为 i 代表子网络的结构。关系式(2) 表示网络是由网络结点通过一定的线性组合的方式来表示的, 不同的组合关系表征了不同的网络结构。系统中占优势的模, 决定了系统中特定的网络结构是占优势的结构。这种模式不仅能够稳定地遗传, 而且还能在一

定的条件下发生变异, 转化为新的模式。大分子结构或基因在一定的范围内保持不变, 能够稳定地遗传下去; 在某些特定条件下, 如环境的突变, 大分子结构或基因会发生突变(变异)。基因是否变异, 取决于值的大小, 在临界点上, 基因(基因型) 就会发生突变, 在宏观上表现为性状(表现型) 的改变。

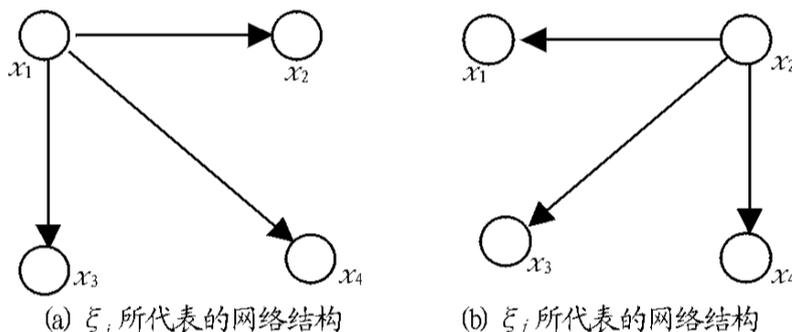


图1 网络结构选择的示意

通过以上分析, 结合已有理论, 可见这样一幅物理图象: 不同特征值的选择, 代表了不同的占优势的子网络结构。这些结构继续向下发展, 就形成了分形维数虽然相同, 但拓扑结构却发生了变化的网络结构; 或者分形维数和拓扑结构都不同的网络结构。笔者假设 i 所代表的网络结构是正常的网络结构, 在通常条件下, 应该最终形成这种网络结构。但由于外界条件的改变, 在一些比较强的外界流作用下, 系统中占优势的模发生变化, 网络结构从原来的 i 所代表的网络结构转变为 j 所代表的网络结构。这种网络结构的转变与具体条件的结合, 具有现实意义。比如说在生物网络结构中, 如果结点所代表的是氨基酸, 那么在蛋白质网络结构的排列方式将被改变。如果结点代表蛋白质, 那么必然会影响到蛋白质所形成的网络结构。蛋白质的结构与蛋白质的功能之间有明显的关系, 因此这种网络结点的变化最终会影响蛋白质网络的功能。蛋白质结构和 DNA 结构的联合作用, 最终会在生物的宏观表现上产生影响。

3 自复制和自组装

蛋白质分子网络结构的形成过程中, 从外界环境获得的广义流可分为稳定和 unstable 2 部分^[2]:

$$\bar{J} = J_u + J_s \quad (5)$$

在方程(5) 中, 脚标 u 代表了广义流中不稳定部分, s 则代表稳定部分。定义 μ 为控制参数, 在蛋白质分子网络结构形成

作者简介 王建平(1977-), 男, 重庆人, 硕士研究生, 研究方向: 非线性科学。

收稿日期 2006-11-30

过程中,当能量输入一定时,能量的转化效率可在一定程度上作为反应聚合成稳定结构的控制参数,可以得知 J_u 对 Ω 的依赖性很强,而 J_s 对 Ω 的依赖性很弱或近似为零。所以在不稳定点附近,有如下关系式^[2-5]:

$$\bar{J}(\Omega_1) - \bar{J}(\Omega_2) = J_u(\Omega_1) - J_u(\Omega_2) \quad (6)$$

一旦网络结构和网络基序确定以后,系统就会形成一种层次自组织/自组装的结构,从而形成一种复杂网络结构。

根据上面的推导,笔者已经初步阐明了网络演化的机制。下面,将从数学角度继续演绎自相似网络的构建过程。

首先,如图2所示,考虑一个有限尺寸为 Ω 的系统。首先选定一个尺寸为 Ω_1 的小单元,它含有 m_1 个子系统(单元的选择必须使得它含有潜在的不稳定子系统,即含有潜在的快模式)。为简单起见,笔者认为单元 Ω_1 是长方形的,至于具体形状如何并不影响问题的分析。

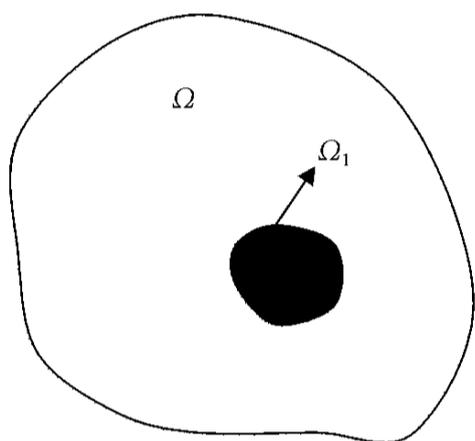


图2 任意开放系统

对称性告诉我们,起主导作用的活跃模式应放在单元 Ω_1 的中间,用标号 $(m_1 + 1)/2$ 表示。根据上面的分析,相变前和相变后,该单元流的变化量全部通过这一部分(图中用黑点表示,它也代表通过自然选择的作用保留下来的模式),即

$$J_1 = \bar{J}_1(\Omega_1) - \bar{J}_1(\Omega_2) = J_{u1}(\Omega_1) - J_{u1}(\Omega_2) \quad (7)$$

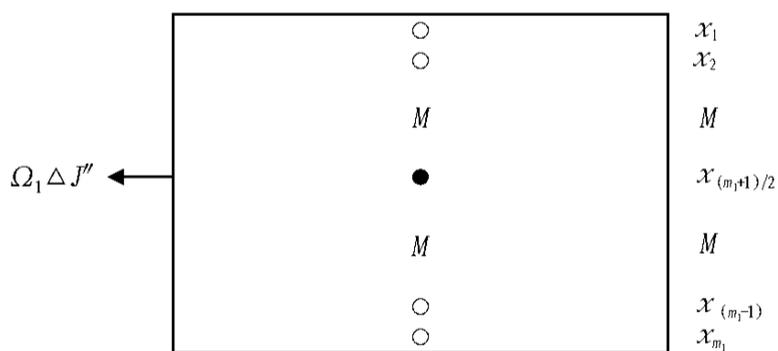


图3 长大的单元 Ω_1

式中 J 相变前和相变后流率的变化,这里形成的结构是系统最低层的结构。随着系统的长大,最小的单元会不断长大直至覆盖整个单元(图2)。显然,我们可以多次重复上面的过程,直至达到整个单元 Ω 。

类似地,随着时间的推移, Ω_1 将在更大的空间中发展,它可能含有几个 Ω_1 ,相应地就有了第二个单元 Ω_2 ,它含有 m_2 个子系统,如图4所示。对称性再次告诉我们,起主导作用的活跃模式应放在单元 Ω_2 的中间,可用标号 $(m_2 + m_1 + 1)/2$ 来表示。根据上面的分析,相变前和相变后,该单元的流的变化量全部通过这一部分,即

$$J_2 = \bar{J}_2(\Omega_1) - \bar{J}_2(\Omega_2) = J_{u2}(\Omega_1) - J_{u2}(\Omega_2) \quad (8)$$

这里形成的结构是系统第二层的结构。

类似地,对 n 级单元,可以获得如图5所示,则有

$$J_n = \bar{J}_n(\Omega_1) - \bar{J}_n(\Omega_2) = J_{un}(\Omega_1) - J_{un}(\Omega_2) \quad (9)$$

这里形成的结构是系统第 n 层的结构。

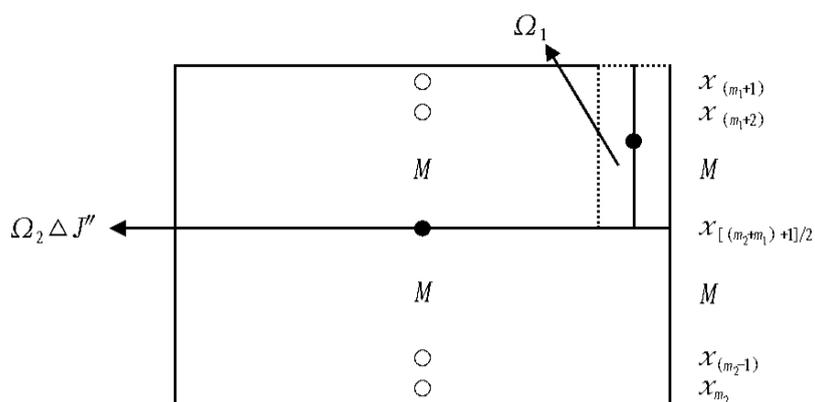


图4 长大的单元 Ω_2

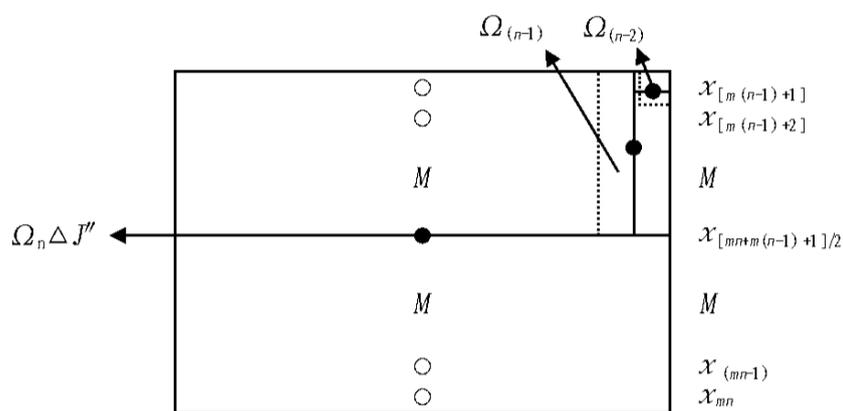


图5 n 级生长后的结构

显然,可得下式^[3-5]

$$\frac{\bar{J}_n(\Omega_1) - \bar{J}_n(\Omega_2)}{\bar{J}_{(n-1)}(\Omega_1) - \bar{J}_{(n-1)}(\Omega_2)} = \frac{J_{un}(\Omega_1) - J_{un}(\Omega_2)}{J_{u(n-1)}(\Omega_1) - J_{u(n-1)}(\Omega_2)} = \frac{n}{n-1} \quad (10)$$

式中 \bar{J} 和 J_u 分别是整个系统的流和不稳定子系统的流, Ω 是控制参数, Ω_1 是生长单元。方程(10)明显地表明了分形自相似层次结构的特征,也在一定程度上说明了蛋白质分子网络分形结构的形成机理。蛋白质分子网络以最小的代价获得最大的广义流为驱动力,在小尺度范围里,通过个体之间的相互非线性作用形成自组织基本单元,然后以这些低层次的个体自组织单元为单位,在逐渐变大的尺度上进行自组织,形成高一级的自组织单元,最后形成一个稳定的蛋白质分子网络结构。

显然随着给定条件的完善,由给出的纯理论而导出的分形结构会越来越接近实际。这种过程就是一种自组装过程,同时也是一种“涌现”过程,在这个过程中不断出现低层次所没有的结构,并表现出了新的功能。这种网络结构具有自适应性,随外界环境的变化而变化。由于自适应性,当环境变化或非均匀性引起网络中内在演化途径的多样性时,系统形态会发生动态变化,这种动态变化使得整个复杂系统形成了新的不变流形,从而产生新的斑图。这种新的不断变化的不变流形就是斑图的涌现^[6]。这样就出现了一种时空有序的蛋白质大分子网络结构,形成了一种时空斑图结构,这种斑图结构具有自识别、自适应、自组装等特性。

自识别对应于特定(一种大分子结构模板对应一个)的选择识别,不同的对应不同的构型。因此可以通过特定的技术手段,产生特定的 Ω 值,从而能够产生特定的生物大

功能,因而将来我们有可能按要求人工设计出特定功能的蛋白质结构。自适应对应于对环境的适应性,既可是被动的也可是主动的,体系是智能的、可训练的,具有响应外界刺激的自修改能力和适应性。在外界环境压力下,可以发生变异,选择不同的,从而产生适应外界环境的模板。在模板的作用下,体系可进行自复制和自组装。

上面分析的网络结构的选择,可看作是一种分子进化机制,一旦选择了某种结构,就会形成一种特定的连接结构。这种特定的连接方式种类在网络中反复出现,而在不同的网络中这些连接方式种类的出现频率是不一样的。也就是说不同类型的网络具有不同的典型连接方式。这些特定的连接方式种类称为网络基序(Network Motif)^[7]。在模型中,它们是以占优势的模_i来表示的。它们被认为是构造网络的基本砖块。这些网络基序一旦出现,就会遗传下去,若干个网络基序形成具有明显空间结构和功能特征的模块(Motifs)。这样,网络常常体现出另外一个构造特点:网络的模块化。模块有2个显著特征:模块内的分子与分子间有直接的相互作用;模块与模块或模块与非模块之间有清晰的边界。如动物、植物、微生物虽然代谢方面有较大的差异,但它们的代谢网络都有着类似的构成方式。即依照一种等级的(Hierarchical)方式构建。首先形成许多紧密联接的小型模块,然后再形成相对松散的网络。也许可以这样认为,网络的构成遵循着:从底到顶(bottom up)的原则。首先形成网络基序,作为网络的基本元素,在此基础上形成较高级的结构-模块,最后形成完整的网络。这也是蛋白质分子网络经过长期进化而得到的最佳结构层次或分形结构。

4 蛋白质分子网络的分形维数计算

层次间广义流耦合的平衡式为^[4]:

$$s \frac{d}{d} \int_0^y u dy - a \left(\frac{d}{y} \right) \Big|_w = \frac{d}{d} \int_0^y u dy \quad (11)$$

式中, s 和 a 为常数, u 和 y 为层次耦合出的广义流的浓度和速度分布,可由相应的试验测得。对一般的 u 和 y 的分布,根据上式我们可求得广义流传递速率为 $h_n \cdot l^{-p}$, p 值介于0和1之间。与前面的分层结构相结合,可得到如下关系式:

$$\frac{\bar{J}_n(n-1) - \bar{J}_n(n-2)}{\bar{J}_{(n-1)}(n-1) - \bar{J}_{(n-1)}(n-2)} = \frac{\bar{J}_{nm}(n-1) - \bar{J}_{um}(n-2)}{\bar{J}_{u(n-1)}(n-1) - \bar{J}_{u(n-1)}(n-2)} = \frac{n}{n-1} \int_0^1 \frac{1}{1-x} J^{(d-p)} \quad (12)$$

式中: $d=1, 2, 3$; $d-p$ 为分形维数。从式(11)可看出,一旦

得到了 u 和 y 的分布,就可确定蛋白质分子网络的分形维数。

5 蛋白质分子网络复杂性的增加

随着蛋白质分子进入系统内部后的分子自组装,复杂性从大分子网络系统的外部进入了内部。系统外部的复杂性减少,而内部的复杂性增加。在系统外部,各种分子处于杂乱无章的状态,其复杂性非常小;进入系统内部,自组织自组装成为特定的大分子网络,其复杂性大大增加。

随着生物细胞的进化,细胞从外界抓取广义流的能力也在增强,因此广义流的形式,也在不断演化,从随机运动—层流—湍流—极度湍流。伴随着流形的演化,其分形维数也在增加。

6 小结

蛋白质大分子网络结构本质上就是生物大分子经过复杂的相互作用、竞争与合作的淘汰机制,在时空上形成的一种有序结构,即是一种时空斑图。笔者在最大流原理的基础上,重点分析了蛋白质大分子网络结构的选择和进化,以及遗传、复制和变异等特征。

对蛋白质分子网络的研究将帮助我们理解单个蛋白质的功能并为对细胞功能的分子基础提供新认识。如人们已知道所发现的NMDA受体复合物的77个蛋白质中有大约1/3是与长时程增强或抑制的诱导相关。这提示该复合物的主要功能是参予了突触可塑性的诱导。同时,因为相互作用的蛋白质在功能上可能是协同的,所以某些蛋白质的功能可根据其相关蛋白的功能来推测。因此,在一个蛋白质分子网络中将可更加全面、系统和方便地了解单个蛋白质的功能。

参考文献

- [1] 吴家睿. 生物系统的分子构建[J]. 科学, 2003, 55(3): 27-28.
- [2] CHAI L H, SHOU M. Self-organization and self-similarity in boiling systems[J]. ASME Journal of Heat Transfer, 2002, 124: 505-517.
- [3] CHAI L H, WEN D S. Hierarchical self-organization of complex systems[J]. Chemical Research of Chinese University, 2004, 20(4): 440-445.
- [4] CHAI L H. Theoretical analysis of bubble interaction in boiling systems[J]. International Journal of Thermal Science, 2004, 43: 1067-1073.
- [5] 柴立和, 马德刚. 分形生长的新模型[J]. 天津大学学报: 自然科学版, 2004, 37(4): 326-330.
- [6] 刘曾荣. 同步、自适应性、网络和复杂系统中的涌现—神经和经济系统中的复杂性方法[C]//方锦清. 第二届全国复杂动态网络学术论坛论文集. 北京: 中国高等科学技术中心, 2005.
- [7] MILO R, SHEN ORR S, ITZKOVITZ S, et al. Network motifs: simple building blocks of complex networks[J]. Science, 2002, 298: 824-827.
- [8] 阎隆飞, 孙之荣. 蛋白质分子结构[M]. 北京: 清华大学出版社, 1999.
- [9] 章文, 伍龙军, 徐天乐. 蛋白质组学与神经科学: 从蛋白质到网络[J]. 医学与哲学, 2002, 23(3): 1-4.
- [10] TUCKER C L, GERA J F, UETZ P. Toward an understanding of eukaryotic protein networks[J]. Trends Cell Biol, 2001, 11: 102-106.