

朗之万方程及其在蛋白质折叠动力学中的应用*

蒋泽南¹⁾ 房超^{2)†} 孙立风²⁾

1) (西门菲莎大学物理系, 本拿比 V5A 1S6, 加拿大)

2) (清华大学核能与新能源技术研究院, 北京 100084)

(2010年9月10日收到; 2010年9月27日收到修改稿)

研究了朗之万方程的动力学性质,并用它模拟了蛋白质分子的折叠过程.首先在相空间中朗之万方程做连续映射,发现做布朗运动的粒子在位置坐标上存在明显的概率分布,这表明蛋白质折叠过程中分子空间构型是非遍历的.此外,本文还通过数值模拟得到了去折叠态蛋白质的紧密度指标,并验证了它与实验结果以及其他理论方法的一致性.本文还提出了一种利用重整化方法研究熔球体状态蛋白质的理论模型,并提供了考虑疏水基影响的蛋白质折叠过程的模拟思路.

关键词: 朗之万方程, 蛋白质折叠非遍历性, 紧密度指标, 重整化

PACS: 05.45.-a, 05.45.Df, 05.45.Pq

1. 引言

生物物理学是利用物理方法研究生命现象的学科,它把生命活动的各个层次有机的联系起来,在新的 heights 上揭示了生命的奥秘^[1,2].蛋白质的结构及其动力学是生物物理学中的一个重要领域^[3],其研究方法包括:非平衡态统计热力学^[4]、计算流体力学^[5]和蒙特卡罗^[6]等方法,这些方法在过去的20年中,取得很多成果^[7,8].然而由于上述方法计算量较大,效率较低,而且有时物理意义也不甚明确,因此,提出一个行之有效且物理内涵明确的方法来研究蛋白质折叠动力学是亟待解决的问题^[7].

蛋白质的折叠过程主要是其中的氨基酸分子在周围环境的影响下形成肽链,并伴随着折叠、螺旋,经历二、三级结构,最终构成蛋白质的过程.肽链中的每一个氨基酸分子,都在周围环境中的水分子撞击下做着无规则运动,这种撞击使肽链发生扭转与折叠,最终形成蛋白质.影响蛋白质形成的因素(如温度等)都可以归结成周围水分子撞击条件的差异——即不同的环境对应着不同的撞击情况,从而影响肽链的折叠过程^[9].

肽链折叠过程对应的物理模型是布朗运动,然

而与布朗运动中自由的花粉颗粒不同的是:肽链中的氨基酸分子是彼此相连的,有固定长度约束着两个相邻氨基酸分子的运动.鉴于这种相似性,我们希望用布朗运动的物理模型,来研究蛋白质折叠过程中各个形态的成因、特点.

以下首先介绍描述布朗运动的朗之万方程以及蛋白质折叠的有关概念和参量的选取;然后通过计算机数值模拟,利用朗之万方程、紧密度指标等概念描绘了去折叠态蛋白质折叠的动力学特性,并和已有的理论与实验结论进行了比较分析,论证了该方法的可行性;此外,我们还提出了一种考虑疏水基相互作用的熔球体状态的蛋白质折叠的动力学模型;文章的最后对于这种方法的应用前景作了展望.

2. 蛋白质折叠的动力学过程与朗之万方程

1827年,英国植物学家布朗在显微镜下观察到悬浮在水中的花粉与其他微小粒子在不停的做着无规则运动(又称为自回避随机游走),我们称这种运动为布朗运动^[10].为了研究布朗运动,朗之万提出:溶液中介质分子对布朗粒子的作用力可分为两个部分,一部分是流体对布朗粒子的黏滞阻力,来

* 清华大学骨干人才计划(批准号:413410002)资助的课题.

† 通讯联系人. E-mail:fangchao@tsinghua.edu.cn

自溶液中分子对布朗粒子的碰撞;另一部分是溶液分子无规则热运动施加在布朗粒子上的一种涨落不定的力,相当于溶液分子对静止的布朗粒子碰撞的作用力.在没有外场的作用下,质量为 m 的自由布朗粒子的运动方程可以写成^[11]

$$m\ddot{\theta} = -\gamma\dot{\theta} + F(t), \quad (1)$$

其中 $-\gamma\dot{\theta}$ 表征的是黏滞阻力,当粒子速度 v 很小时,黏滞阻力与 v 成正比, γ 为常数,由外界环境决定. $F(t)$ 表示随机力.

布朗运动是不连续的,因此需要将朗之万方程写成迭代差分方程的形式.设 $X^1 = \dot{\theta}$, $X^2 = \theta$, 并进行差分可得

$$\begin{aligned} X_{n+1}^1 &= (1 - \gamma\Delta t/m)X_n^1 + F_{\text{random}}, \\ X_{n+1}^2 &= X_n^2 + X_n^1\Delta t. \end{aligned} \quad (2)$$

由于布朗运动中微粒受的力是随机的,所以我们用 F_{random} 代替 $F(t)$. 其中 F_{random} 是服从以 0 为均值、 $C_0^2\Delta t^2/m^2$ 为方差的高斯型分布的随机变量(C_0 为常数,由环境中的水分子决定).随着映射次数的增加,若保证方程不发散,必有 $|1 - \gamma| \leq 1$, 则 $\gamma \in [0, 2]$. 后面我们把方程(2)称为朗之万方程,这也正是研究蛋白质折叠的基础.对于实际的物理问题,根据流体力学知识有^[12]

$$\gamma = 6\pi r\eta, \quad (3)$$

其中 r 为氨基酸分子的直径, η 为氨基酸所处环境溶液中的黏滞系数.另外,对于随机力,有^[13]

$$\frac{C_0^2}{2\gamma} = k_B T, \quad (4)$$

其中 $k_B T$ 为玻尔兹曼常数, T 为溶液的温度,这些参量均可以从实验数据得知.

布朗运动的分形维数是一个重要的概念^[14],这种维数和蛋白质折叠的空间维数联系有关.根据生物学的实验结果,可知蛋白质的折叠过程分为图 1 所示 3 个步骤^[9].

“去折叠态”是指未完全折叠成蛋白质的肽链,

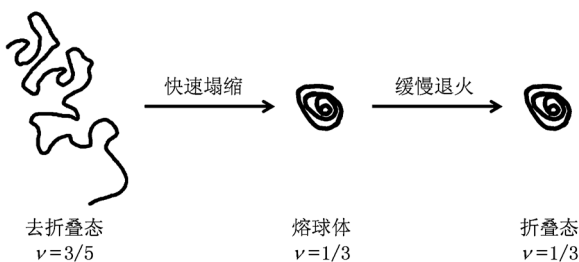


图 1 蛋白质折叠过程示意图

它上面的各个氨基酸分子自由分布在空间内,做有约束的布朗运动.按照布朗运动的相关理论,对于一条含有 N 个氨基酸的蛋白质分子,设其半径为 R , 则有

$$R \propto N^{3/5}, \quad (5)$$

其中 $3/5$ 我们称为分子的紧密度指标^[15],用 ν 表示.对于蛋白质的熔球体和折叠状态两种状态,根据非平衡态统计力学可以得到它们对应的紧密度指标 $\nu = 1/3$. 根据几何体空间维数的定义,可以发现紧密度指标的倒数就是物体的维数.可见,处于折叠过程中不同阶段的肽链,其维数是不同的,因此只要能求出朗之万方程对应的随机游走模型的分形维数,则可验证用朗之万方程研究蛋白质折叠动力学的可行性.

3. 朗之万方程的非遍历性

首先讨论普遍意义下方程的特点,从而论证其内禀性质和蛋白质折叠动力学过程相符合.在方程有意义的条件下任取 γ 和 C_0 的值, F_{random} 在 $(-\infty, +\infty)$ 范围内按高斯型分布 $N(0, C_0^2\Delta t^2/m^2)$ 概率随机取值.我们把方程(2)做连续映射(动量和坐标的初值均设为 0),结果如图 2 所示(图中纵轴为动量 X' ,横轴为坐标 X ,已做无量纲化处理).

图 2(a), (b), (c), (d) 4 种情况下的相空间轨迹坐标范围分别是 $(-18, 120)$, $(0, 60)$, $(-680, 190)$ 和 $(-28, 80)$. 显而易见,对于各种情况,动力学轨迹的坐标范围是不对称的.随着 γ 的增大,坐标的活动范围变小,并有向中间压缩的趋势.这说明当周围水环境黏滞阻力变大时,布朗粒子的运动会更加局域化;此外, C_0 取值对于相空间轨迹坐标的活动范围影响较大,且随着 C_0 的增大,相空间轨迹的中心向 X 负方向偏移(如图 2(c)),这说明水分子的随机碰撞是影响布朗运动的关键性因素.另外,通过比较图 2(a) 和 (d),发现迭代次数的增加对相空间轨迹的坐标范围跨度几乎没有影响,仅仅是数值上有所变化.4 种情况下的相空间轨迹速度分布范围分别是 $(-0.4, 0.4)$, $(-0.3, 0.3)$, $(-3, 3)$, $(-0.4, 0.4)$, 即都是以 0 为中心的对称分布.可以看出, γ 的大小对于速度分布几乎无影响,迭代的次数也不影响布朗粒子的速度分布,而真正起到决定作用的是 C_0 的数值.从坐标和速度两个方面来观察,可以发现随机力的大小对于布朗运动起着决

定性的作用,且这个随机力的方差越大,布朗运动的局域性越差.

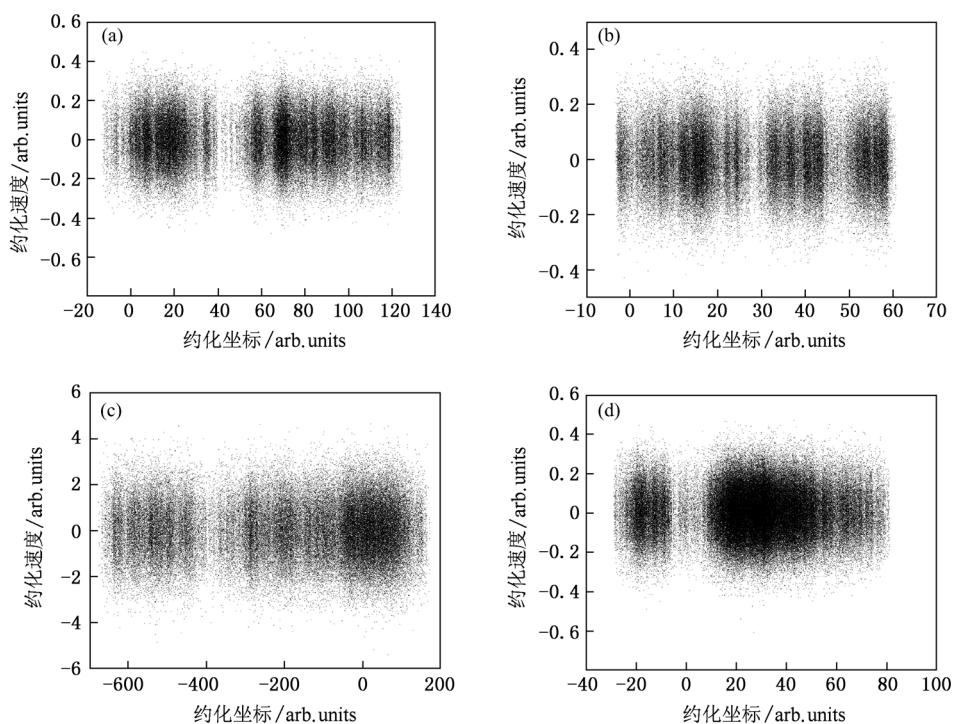


图2 朗之万方程的相空间轨迹 (a) $\gamma=0.5, C_0=0.1$ 映射 100,000 次; (b) $\gamma=1.0, C_0=0.1$ 映射 100,000 次; (c) $\gamma=0.5, C_0=1.0$ 映射 100,000 次; (d) $\gamma=0.5, C=0.1$ 映射 200,000 次

此外,我们还发现布朗粒子在相空间的轨迹始终呈现为一种条纹状分布,这种分布不随参数 γ, C_0 的取值改变,也不受映射次数变化的影响. 这就是说,粒子在运动过程中坐标方向的取值有明显的概率分布,即粒子显然更偏向于在某处运动,而很少在另一处停留. 与之相对的,布朗粒子在速度方向的取值基本呈对称,即反映了布朗运动的无规性.

此外我们统计了坐标空间中每个坐标点对应的随机点的数量,从统计学的角度研究朗之万方程的动力学过程,如图 3.

从图 3 中,我们可以发现朗之万方程中的一些动力学行为特点:1) 随着迭代次数的增加,相空间的轨迹向着坐标数值较大的范围扩散,这说明在自回避随机游走过程中,布朗粒子首先在其原点附近运动较长时间,随着迭代次数的增加,才慢慢的开始运动到较远位置;2) 在各种迭代次数下,布朗粒子随机运动时落入某些特殊的坐标空间区(如 $X=20.2$ 附近, $X=51.3$ 附近等)的概率都较小,且相对比例并不随着映射次数的增加而改变,这说明在此随机过程中的空间非遍历性是一种内禀属性,并不因为随机游走的时间增长而发生变化.

对于真实的蛋白质体系,我们只需要找到适当

的 γ 与 C_0 , 确定粒子间的间距,并将维数扩充到三维,就可以得到蛋白质折叠时氨基酸分子的运动特点. 实验表明:在蛋白质折叠过程中,氨基酸分子的运动并不是完全随机的,而是在一些特定方向上有偏好,这也就说明了蛋白质折叠过程中,并不会把所有的构象都“搜索”一遍,而是只向某些特定的构象折叠. 虽然肽链分子周围环境整体各向同性,肽链中的氨基酸分子中做布朗运动的方向可以是任意的,但是蛋白质最后会折叠成特定的形状并不是偶然的. 普利高津等统计物理学家认为,在蛋白质折叠过程中,要满足自由能最小、熵最大、每步熵增量最小^[16] 3 个要求,最后的蛋白质形态,是所有可能的折叠构型中上述 3 个条件综合作用的结果. 但事实上,蛋白质不可能进行无偏差的构象的空间搜索来折叠,因为如果通过逐个筛选所有不同的空间构象将需要数十亿年的时间,这与实验上发现的毫秒量级的折叠时间相悖. 而用朗之万方程描述的布朗粒子随机游走模型很好的描述了肽链中氨基酸分子的动力学过程特点. 以上结论说明:朗之万方程的本质和蛋白质折叠的物理图像相契合,它可以解释蛋白质折叠过程的非遍历性,因此用该方程描述蛋白质折叠过程是可行的.

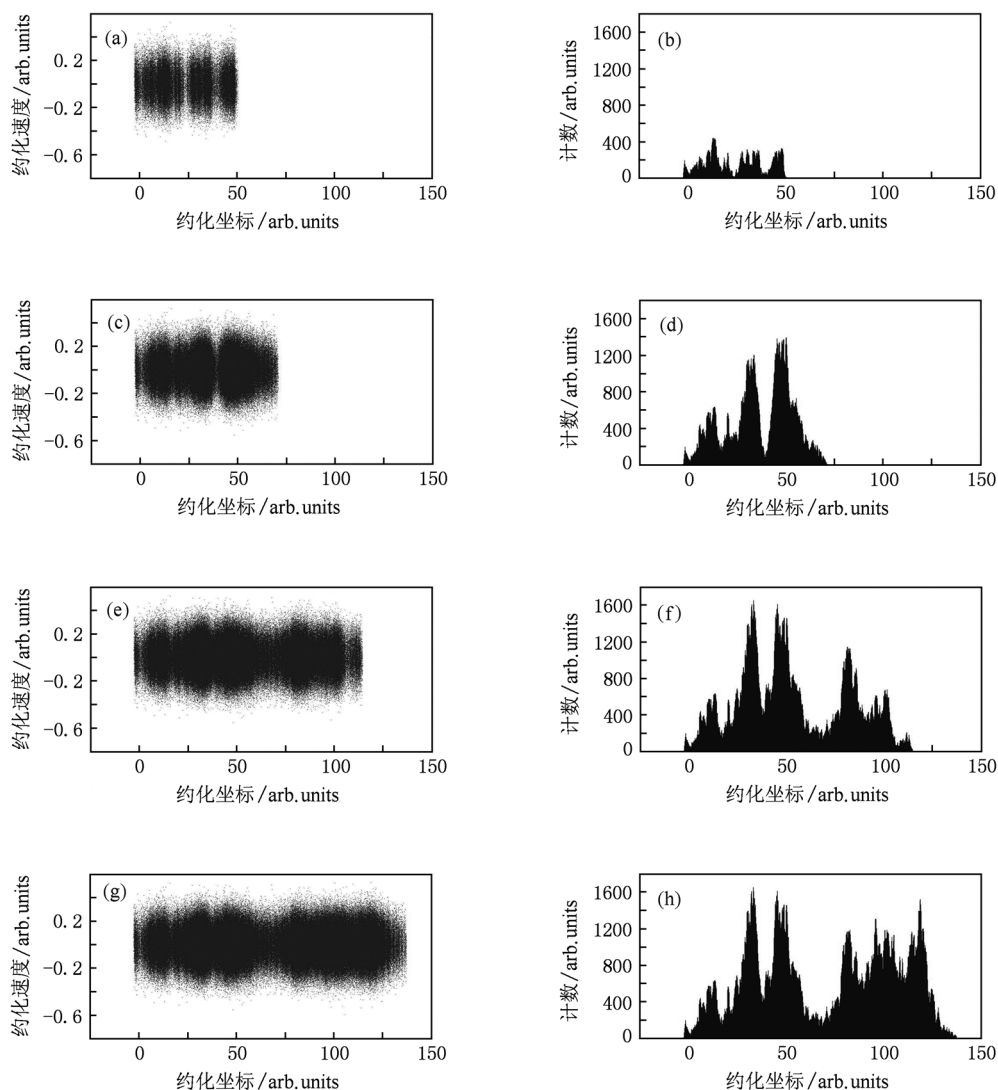


图3 $\gamma = 1.5, C_0 = 0.1$ 按照朗之万方程映射 10^6 步的相空间轨迹在不同步数的相空间图和坐标方向概率分布图
(a)和(b) 5×10^4 步; (c)和(d) 10^5 步; (e)和(f) 5×10^5 步; (g)和(h) 10^6 步

4. 利用朗之万方程研究蛋白质折叠动力学

4.1. 去折叠态蛋白质折叠过程的维数研究

我们考虑有 100 个氨基酸分子的肽链. 在三维空间中, 自某一定点 (设为坐标原点) 开始, 使粒子固定步长 d 做回避行走 100 步, 可以生成一条含 100 个节点的链. 根据各态历经假设, 这种条件下生成的链的动力学行为与 100 个相互连接的粒子的动力学行为完全一致^[17]. 行走过程中, 粒子在 X, Y 两方向遵循朗之万方程运动, Z 方向的位置坐标由步长 d 控制, 即使每相邻两节点间的距离恒定为 d . 粒

子在 X, Y 两方向所受的随机力相互独立, 都服从以 0 为均值, $C_0^2 \Delta t^2 / m^2$ 为方差的高斯型分布. 蛋白质是通过氨基酸的 α 碳原子上相邻氨基和羧基间形成的酰胺键连接而成的, 假设 3 键在同一直线上, 对键长求和近似作为残基间的距离, 为 4.32 Å. 蛋白质分子二级结构下, 计算得相邻两残基间的直线距离为 4.39 Å^[18]. 综合以上两种方法的结果, 在实际对蛋白质分子链生成的模拟中, 我们采取 4.5 Å 作为步长. 考虑如下条件: 水的黏度 $\eta = 1.35 \times 10^{-2} \text{ N} \cdot \text{s} / \text{m}^2$, 温度 $T = 310 \text{ K}$, 20 种氨基酸分子的平均半径为 $r = 2.25 \times 10^{-10} \text{ m}$, 平均质量为 $m = 2.1583 \times 10^{-25} \text{ kg}$ ^[18]. 通过计算可得 $\gamma = 1.5, C_0 = 1.2$, 步长 $d = 4.5$ (已作无量纲化处理). 由上述参数, 模拟计算

可得 3 维空间中的一条随机链如图 4 所示.

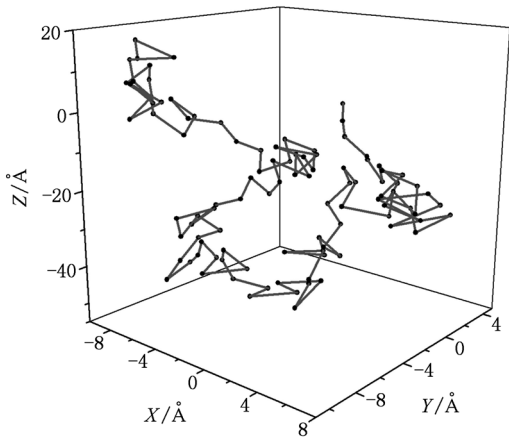


图 4 模拟去折叠态蛋白质分子链

由紧密度指标的定义,有

$$\ln N = \frac{1}{\nu} \ln R + C, \quad (6)$$

因此只需对半径 R 中包含的随机点的数量 N 做线性拟合,即可求得紧密度指标,如图 5. 可得

$$\ln N = 1.663 \ln r - 0.777, \quad (7)$$

即 $\nu = 1/1.663 = 0.601 \approx 3/5$, 这正符合实验结论. 这一结果说明用简单的朗之万方程可以很好的模拟去折叠态蛋白质的折叠过程.

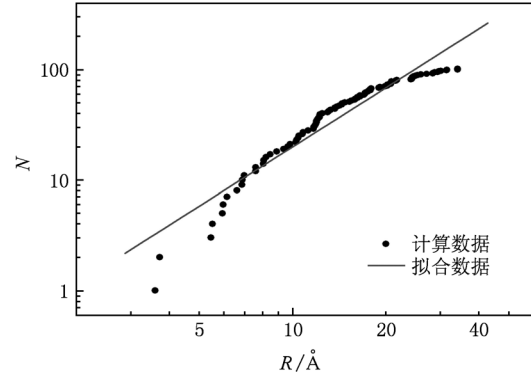


图 5 去折叠态蛋白质分子链中固定半径与其中所包含的随机点数量关系图及其直线拟和(双对数图)

4.2. 对于熔球体状态的蛋白质折叠模型的初步探讨

处于熔球体状态的蛋白质的折叠主要依靠局部序列的近程或中程(3—4 个残基)相互作用来维系. 氨基酸以非特异性布朗运动的方式扩散、碰撞、相互黏附,形成具有疏水核心和二级结构的球状结构. 基于以上机理,我们选用描述扩散-碰撞的朗之万方程作为折叠过程中节点运动的基本动力学方程,同时考虑氨基酸残基疏水相互作用,所建立模型方程为

$$\begin{aligned} X_{n+1}^1(i) &= (1 - \gamma \Delta t/m) X_n^1(i) + F_{\text{random}}, \\ X_{n+1}^2(i) &= X_n^2(i) + X_n^1 \Delta t(i), \\ Y_{n+1}^1(i) &= (1 - \gamma \Delta t/m) Y_n^1(i) + F'_{\text{random}}, \\ Y_{n+1}^2(i) &= Y_n^2(i) + Y_n^1 \Delta t(i), \\ Z_{n+1}^2(i) &= Z_{n+1}^2(i-1) \pm \sqrt{d^2 - (X_{n+1}^2(i) - X_{n+1}^2(i-1))^2 - (Y_{n+1}^2(i) - Y_{n+1}^2(i-1))^2}, \\ Z_n^1(i) \Delta t &= Z_{n+1}^2(i) - Z_n^2(i), \end{aligned} \quad (8)$$

其中 i 表示氨基酸节点的编号($i=1,2,\dots,100$). 此方程加入了对相邻氨基酸之间的相对距离的考量. 将任意两氨基酸分子间的肽键看做硬杆,随机链上第 i 个节点在 X, Y 方向可做自由的自回避布朗运动,而在 Z 方向的坐标由其与第 $i-1$ 节点的距离确定. 可以发现:已知随机链上第 i 个节点在 t 时刻的位置和动量,按(8)式做一次映射,即可得到第 i 点在 $t + \Delta t$ 时刻的位置和动量.

此外,还需要考虑疏水基对折叠过程的影响. 在蛋白质分子链上随机选取 n 个节点,设为疏水基,

并记下这些节点的序. 模拟折叠过程时,对于任意节点我们先生成该点 X, Y 方向的位置和动量,然后对 Z 方向的位置做检验. 如果此节点为疏水基,那么在折叠过程中,由于受到极性分子水的排斥作用,这个基团会受到远离水分子的作用力,相当于在折叠过程中,这个节点会被亲水基团包裹在内部. 因此,当节点为疏水基时,要求节点必须向更靠近随机链质心的方向移动;若节点包含亲水基,则要求节点每次运动都远离质心方向;若节点不包含亲水疏水基团,则节点仅受到方程(8)的约束.

以往的理论和实验研究表明,蛋白质分子完成从去折叠到熔球体态所需时间在 10^{-3} s 量级. 假设蛋白质分子在水介质中折叠,用水的黏度以及 20 种氨基酸分子的平均半径及质量作为参数代入方程 (8) 中做计算,则需要迭代 10^{14} 次. 显然,迭代过程中方程很容易发散,误差也很高. 因此,单纯使用方程 (8) 计算是不合理的.

我们可以利用重整化方法解决该问题. 重整化方法原本是用于粒子物理学中处理无穷大问题的,这里的重整化含义是:通过将多次迭代过程视为一个整体,以扩大积分时间,从而使积分时间和蛋白质折叠过程的时间可比拟,以达到降低误差的目的. 这和理论物理中量子场论重整化的意义并不相同,量子场论的重整化是指将无穷大避开而关注其主要部分,其基本思想是将原来的无穷大项经过一些处理后使之不影响最后的结果,而此处的重整化意义是将积分单元重构,以消除积分发散,从根本上解决问题. 需要注意的是:这个积分单元重构的过程并不是简单的增大积分精度,而是通过拓展其动力学维度才引入的. 为简明起见,这里只写出 X 方向的动量和位置的重整化方程. 将方程 (8) X 方向两个方程迭代 p 次,得到

$$\begin{aligned} X_{n+p}^1 &= (1 - \gamma\Delta t/m)^0 X_n^1 + F_1, \\ X_{n+p}^2 &= X_n^2 + \frac{m}{\gamma} [1 - (1 - \gamma\Delta t/m)^p] X_n^1 + F_2, \end{aligned} \quad (9)$$

其中

$$\begin{aligned} F_1 &= \sum_{i=1}^p (1 - \gamma\Delta t/m)^{p-i} F_{n+i-1, \text{random}}, \\ F_2 &= \sum_{i=1}^{p-1} \frac{1 - (1 - \gamma\Delta t/m)^{p-i}}{\gamma/m} F_{n+i+1, \text{random}}. \end{aligned} \quad (10)$$

在多次迭代的过程中,随机力在 t 时刻对节点动量的作用将影响 $t + \Delta t$ 时刻节点的位置,因此在坐标方程中也出现了随机力. 经过多次迭代后,随机力在两个方程中分别表现为 F_1 和 F_2 ,二者也是满足高斯型分布,其均值为 0,方差分别为

$$\begin{aligned} C_1^2 &= \frac{C_0^2 \Delta t^2}{m^2} \frac{1 - (1 - \gamma\Delta t/m)^{2p}}{1 - (1 - \gamma\Delta t/m)^2}, \\ C_2^2 &= \frac{C_0^2}{m^2} \sum_{i=1}^{p-1} \left(\frac{1 - (1 - \gamma\Delta t/m)^{p-i}}{\gamma/m\Delta t} \right)^2. \end{aligned} \quad (11)$$

用重整化之后的方程计算,可以将迭代的次数降低为单纯使用方程 (8) 的 $1/p$ 倍,在保留原有计算精度的情况下,避免积分发散的问题.

利用重整化之后的方程加上相应的实验参数,就能用这个方法模拟出熔球体状态的蛋白质折叠动力学过程.

5. 结 论

利用朗之万方程来描述蛋白质折叠的动力学过程. 通过对相关参量的计算,发现了蛋白质折叠过程中的非遍历性以及动量分布的对称性. 进一步用朗之万方程刻画了去折叠态蛋白质折叠的动力学过程,发现用这种方法模拟出来的蛋白质链的紧密度指标为 $3/5$,这恰恰说明该方法在模拟蛋白质折叠动力学的可行性. 我们还给出了处理熔球体状态的蛋白质折叠的动力学方程以及疏水基在其中的作用. 研究表明,通过变形以及重整化方法,朗之万方程也可以用来处理熔球体状态的蛋白质折叠动力学问题,这为蛋白质折叠动力学的研究提供了一种新的思路.

[1] Meirovitch H 2010 *J. Mol. Recog.* **23** 153
 [2] Fragneto G 2009 *Eur. Phys. J. E* **30** 239
 [3] Brown S P, Muchmore S W 2009 *J. Med. Chem.* **52** 4549
 [4] Yan X S 2004 *Acta Phys. Sin.* **53** 2852 (in Chinese) [邢修三 2004 *物理学报* **53** 2852]
 [5] Papastavrou A, Steinmann P 2010 *Arch. Appl. Mech.* **80** 969
 [6] Ma Z F, Zhuang Y Q, Du L, Wei S 2005 *Chin. Phys.* **14** 808
 [7] Lobkovsky A E, Wolf Y I, Koonin E V 2009 *Proc. Nat. Acad. Sci. U. S. A.* **107** 2983
 [8] Saoud M, Blazzykowski C, Ballantyne S M, Thompson M 2009 *Analyst* **134** 835

[9] Huang K 2006 *Lectures on Statistical Physics and Protein Folding* (Shanghai: Fudan University Press)
 [10] Li H Q, Wang F Q 1993 *Fractal Theory and its Application in Molecular Science* (Beijing: Science Press)
 [11] Creighton T E 1994 *Mechanism of Protein Folding* (Oxford: Oxford University Press)
 [12] Clayton T C, Donald F E, John A R 2001 *Engineering Fluid Mechanics* (New York: John Wiley & Sons)
 [13] Huang K 2001 *Introduction to Statistical Physics* (London: Taylor & Francis)
 [14] Ott E 1994 *Chaos in Dynamical System* (Cambridge: Cambridge University Press)

- [15] Roger H P 2000 *Mechanisms of Protein Folding* (Oxford: Oxford University Press)
- [16] Prigogine I, Wiame J M 1946 *Experientia* **2** 451
- [17] Komorowski T, Peszat S, Szarek T 2010 *Anna. Prob.* **38** 1401
- [18] Branden C, Tooze J 1999 *Introduction to Protein Structure* (2nd Ed.) (New York: Garland Publishing)

Langevin equation and its applications in the simulation of protein folding process *

Jiang Ze-Nan¹⁾ Fang Chao^{2)†} Sun Li-Feng²⁾

1) (Department of Physics, Simon Fraser University, Burnaby, BC V5A 1S6, Canada)

2) (Institute of Nuclear and New Energy Technology, Tsinghua University, Beijing 100084, China)

(Received 10 September 2010; revised manuscript received 27 September 2010)

Abstract

Some ideas concerning the properties of Langevin equation and its applications to the protein folding process are introduced in this paper. Continuous mapping of Langevin equation is constructed in the phase space and a clear inhomogeneous distribution at position of the Brownian particles is discovered, which is consistent with the hypothesis of the nonergodicity of space structure of protein molecule in the folding process. Besides, a random coil with the same compactness index as proteins in the denatured state is obtained by simulation under the self-avoiding walk condition. Taking the hydrophobic effects as well as the renormalization model into consideration, we present the method to simulate the process of protein shrinking from a random coil to a molten globule.

Keywords: Langevin equation, nonergodicity of protein folding, compactness index, renormalization

PACS: 05.45.-a, 05.45.Df, 05.45.Pq

* Project supported by the Key Scientists Program of Tsinghua University of China (Grant No. 413410002).

† Corresponding author. E-mail: fangchao@tsinghua.edu.cn