研究生前沿教材书系

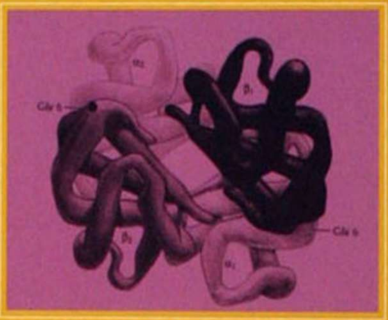
Lectures on Statistical Physics and Protein Folding

统计物理和蛋白质折叠讲义

（英文影印版）

Kerson Huang

(Massachusetts Institute of Technology, USA)



復旦大學出版社

研究生前沿教材书系

Lectures on Statistical Physics and Protein Folding

统计物理和蛋白质折叠讲义

(英文影印版)

Kerson Huang
(*Massachusetts Institute of Technology, USA*)

復旦大學出版社

图书在版编目(CIP)数据

统计物理和蛋白质折叠讲义 = Lectures on Statistical Physics and Protein Folding/〔美〕黄克逊著. —上海:复旦大学出版社,2006. 11
(研究生前沿教材书系)
ISBN 7-309-05203-X

Ⅰ. 统… Ⅱ. 黄… Ⅲ. ①统计物理学-应用-生物学-研究生-教学参考资料-英文②蛋白质-分子折叠-研究生-教学参考资料-英文
Ⅳ. ①Q-332②Q51

中国版本图书馆 CIP 数据核字(2006)第121628号

Lectures on Statistical Physics and Protein Folding
Kerson Huang

Copyright © 2005 by World Scientific Publishing Co. Pte. Ltd.
本书由(新加坡)世界科技出版公司 授权出版英文影印版

著作权合同登记号 图字:09-2006-682

统计物理和蛋白质折叠讲义

〔美〕黄克逊 著

出版发行 复旦大学出版社 上海市国权路579号 邮编 200433
86-21-65642857(门市零售)
86-21-65118853(团体订购) 86-21-65109143(外埠邮购)
fupnet@fudanpress.com http://www.fudanpress.com

责任编辑	梁 玲
总 编 辑	高若海
出 品 人	贺圣遂

印 刷	浙江省临安市曙光印务有限公司
开 本	787×960 1/16
印 张	10
版 次	2006年11月第一版第一次印刷
印 数	1—3 100

书 号	ISBN 7-309-05203-X/O · 375
定 价	18.00元

如有印装质量问题,请向复旦大学出版社发行部调换。

版权所有 侵权必究

Kerson Huang(黄克逊)

作者系美国麻省理工学院(Massachusetts Institute of Technology)荣誉退休教授。美籍华裔科学家。1928年出生于中国南宁市,先后于1950年和1953年获得麻省理工学院物理学学士和物理学博士学位,之后在普林斯顿大学(Princeton University)作短暂博士后研究,1957年回到麻省理工学院从事热力学和统计力学的教学和研究工作。他的*Statistical Mechanics*(Wiley, New York)多次重版,对大学物理教学产生过广泛影响。此外还有*Introduction to Statistical Physics*(Taylor & Francis, London)等著作出版。

内 容 简 介

本书是作者于2004年在清华大学周培源应用数学中心，给多种学科背景的学者讲述统计物理在生物学科的应用的讲义基础上形成的。全书分16章和一个附录。前10章简洁地归纳了生命科学中用得着的核心概念，它们分别是熵、麦克斯韦-玻尔兹曼分布、自由能、化学势、相变、相变动力学、关联函数、随机过程和朗之万方程。第11章开始，讲述的侧重点逐步转移到生命科学。其中第11章讲述蛋白质结构同生命过程的联系。第12章讲述自组装的生物学过程，第13章介绍蛋白质折叠的动力学机理，第14章讲述蛋白质折叠的指数律，第15章阐述自回避行走和湍流，第16章作为全书的结尾，提出了控制蛋白质一级、二级、三级结构的机制的假设，附录中介绍蛋白质分子中能量级联机制的物理学模型。

全书以简洁的语言，精辟地提出了可能的研究方向，对于从事生命科学研究的多学科读者都具有指导意义。

此为试读，需要完整PDF请访问：www.ertongbook.co

序　　言

本书是由作者在清华大学周培源应用数学中心,给包括统计物理知识在内的不同学科背景的听众,介绍生物学研究,尤其是介绍蛋白质结构研究的一系列讲稿组成的。这本书出版很及时,因为给人的感觉是,通过应用统计物理的原理,包括应用统计力学、动力理论及随机过程理论,生物学和生物物理学就经历了高速的发展。

1—10 章比较透彻地介绍了统计物理学。本书第二部分(11—16 章)的讲述逐步转向生物学的应用。本书的讲述风格是,一旦像自回避无规行走和湍流(15 章)等数学/物理原理建立起来,便能讲述生物物理的专题。

“生命过程”从 11 章开始讨论,这里包括一级、二级、三级结构的基本课题。第 16 章作为结尾,提出了掌控二级、三级结构的形式和相互作用的基本原理的有用假设。作者尽量回避对经验信息的详细讨论,代之以给出标准出版物的参考目录。有兴趣的读者可以沿着文献指引的方向深入探索,推荐他们读读由 Roger H. Pain 主编的 *Mechanisms of Protein Folding*(Oxford, 2000)一书。传统上讲,从蛋白质的氨基酸顺序来预测它的结构是研究蛋白质结构的关键手段。不过近来侧重点则开始向研究机制的方向倾斜了。如果读者为了更好地理解这本书而对一般的背景信息感兴趣的话,那就推荐你读读由 Carl Branden 和 John Tooze 的 *Introduction to Protein Structure*(Garland,1999)一书。本书的另一特点是从上述两本书中复制了大量关键性的图。

蛋白质结构问题很复杂,的确所有的问题都很复杂,对它的研究需要采用几种不同的平行方法,这些方法彼此补充才行。因此可想而知,在很长一段时间里,更好地弄懂折叠的机制将有利于推动更好的预估方法的发展。我希望在不多的几年里就能出版本书的

第二版,那就可以把所有新的进展,详尽地充实到书中来。的确,上述提到的两本很有影响的书,已是第二版了。我希望本书在生物物理学科的发展中,也会起到相似的作用。

林家翘
2004 年 6 月
于北京清华大学周培源应用数学中心

Introduction

There is now a rich store of information on protein structure in various protein data banks. There is consensus that protein folding is driven mainly by the hydrophobic effect. What is lacking, however, is an understanding of specific physical principles governing the folding process. It is the purpose of these lectures to address this problem from the point of view of statistical physics. For background, the first part of these lectures provides a concise but relatively complete review of classical statistical mechanics and kinetic theory. The second part deals with the main topic.

It is an empirical fact that proteins of very different amino acid sequences share the same folded structure, a circumstance referred to as "convergent evolution." It other words, different initial states evolve towards the same dynamical equilibrium. Such a phenomenon is common in dissipative stochastic processes, as noted by C.C. Lin.[1] Some examples are the establishment of homogeneous turbulence, and the spiral structure of galaxies, which lead to the study of protein folding as a dissipative stochastic processes, an approach developed over the past year by the author in collaboration with Lin.

In our approach, we consider the energy balance that maintains the folded state in a dynamical equilibrium. For a system with few degrees of freedom, such as a Brownian particle, the balance between energy input and dissipation is relatively simple, namely, they are related through the fluctuation–dissipation theorem. In a system with many length scales, as a protein molecule, the situation is more complicated, and the input energy is dispersed among modes with different length scales, before being dissipated. Thus, energy

[1] C.C. Lin (2003). On the evolution of applied mathematics, *Acta Mech. Sin.* **19** (2), 97–102.

flows through the system along many different possible paths. The dynamical equilibrium is characterized by the most probable path.

- What is the source of the input energy?

The protein molecule folds in an aqueous solution, because of the hydrophobic effect. It is "squeezed" into shape by a fluctuating network of water molecules. If the water content is reduced, or if the temperature is raised, the molecule would become a random coil. The maintenance of the folded structure therefore requires constant interaction between the protein molecule and the water net. Water nets have vibrational frequencies of the order of 10 GHz. This lies in the same range as those of the low vibrational modes of the protein molecule. Therefore, there is resonant transfer of energy from the water network to the protein, in addition to the energy exchange due to random impacts. When the temperature is sufficiently low, the resonant transfer dominates over random energy exchange.

- How is the input energy dissipated?

The resonant energy transfer involves shape vibrations, and therefore occurs at the largest length scales of the protein molecule. It is then transferred to intermediate length scales through nonlinear couplings of the vibrational modes, most of which are associated with internal structures not exposed to the surface. There is thus little dissipation, until the energy is further dispersed down the ladder of length scales, until it reaches the surface modes associated with loops, at the smaller length scales of the molecule. Thus, there is energy cascade, reminiscent of that in the Kolmogorov theory of fully developed turbulence.

The energy cascade depends on the geometrical shape of the system, and the cascade time changes during the folding process. We conjecture that

> *The most probable folding path is that which minimizes the cascade time.*

This principle may not uniquely determine the folded structure, but it would drive it towards a sort of "basin of attraction." This would provide a basis for convergent evolution, for the energy cascade blots out memory of the initial configuration after a few steps. A simple model in the Appendix illustrates this principle.

We shall begin with introductions to statistical methods, and basic facts concerning protein folding. The energy cascade will be discussed in the last two chapters.

For references on statistical physics, the reader may consult the following textbooks by the author:

K. Huang, *Introduction to Statistical Physics* (Taylor & Francis, London, 2001).

K. Huang, *Statistical Mechanics*, 2nd ed. (John Wiley & Sons, New York, 1987).

Contents

此为试读，需要完整PDF请访问：www.ertongbook.co

Chapter 1

Entropy

1.1. Statistical Ensembles

The purpose of statistical methods is to calculate the probabilities of occurrences of possible outcomes in a given process. We imagine that the process is repeated a large number of times K. If a specific outcome occurs p number of times, then its probability of occurrence is defined as the limit of p/K, when K tends to infinity. In such an experiment, the outcomes are typically distributed in the qualitative manner shown in Fig. 1.1, where the probability is peaked at some average value, with a spread characterized by the width of the distribution.

In statistical physics, our goal is to calculate the average values of physical properties of a system, such as correlation functions. The statistical approach is valid when fluctuations from average behavior are small. For most physical systems encountered in daily life,

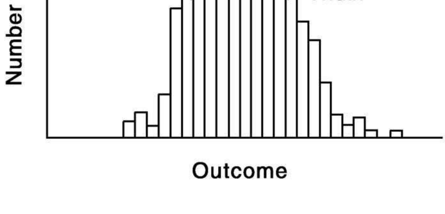


Fig. 1.1. Relative probability distribution in an experiment.

fluctuations about average behavior are in fact small, due to the large number of atoms involved. This accounts for the usefulness of statistical methods in physics.

We calculate averages of physical quantities over a *statistical ensemble*, which consists of states of the system with assigned probabilities, chosen to best represent physical situations. By implementing such methods, we are able to derive the law of thermodynamics, and calculate thermodynamic properties, starting with an atomic description of matter. Historically, our theories fall into the following designations:

- *Statistical mechanics*, which deals with ensembles corresponding to equilibrium conditions;
- *Kinetic theory*, which deals with time-dependent ensembles that describe the approach to equilibrium.

Let us denote a possible state of a classical system by s. For definiteness, think of a classical gas of N atoms, where the state of each atom is specified by the set of momentum and position vectors $\{\mathbf{p}, \mathbf{r}\}$. For the entire gas, s stand for all the momenta and positions of all the N atoms, and the phase space is $6N$-dimensional. The dynamical evolution is governed by the Hamiltonian $H(s)$, and may be represented by a trajectory in phase space, as illustrated symbolically in Fig. 1.2. The trajectory never intersects itself, since the solution to the equations of motion is unique, given initial conditions.

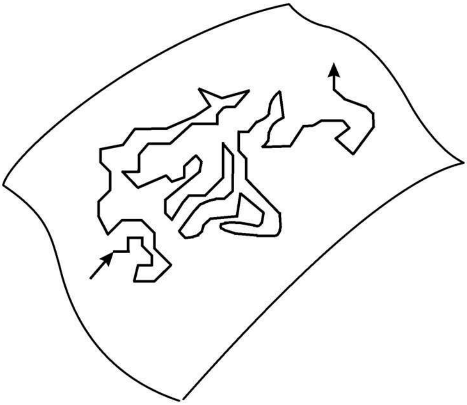


Fig. 1.2. Symbolic representation of a trajectory in phase space.

It is exceedingly sensitive to initial conditions due to interactions. Two points near each other will initially diverge from each other exponentially in time, and the trajectory exhibits ergodic behavior: Given sufficient time, it will come arbitrarily close to any accessible point. After a short time, the trajectory becomes a spacing-filling tangle, and we can consider this as a distribution of points. This distribution corresponds to a statistical ensemble, which will continue to evolve towards an equilibrium ensemble.

There is a hierarchy of time scales, the shortest of which is set by the collision time, the average time interval between two successive atomic collisions, which is of the order of 10^{-10} s under standard conditions. Longer time scales are set by transport coefficients such as viscosity. Thus, a gas with arbitrary initial condition is expected to settle down to a state of local equilibrium in the order of 10^{-10} s, at which point a hydrodynamic description becomes valid. After a longer time, depending on initial conditions, the gas finally approaches a uniform equilibrium.

In the ensemble approach, we describe the distribution of points in phase space by a density function $\rho(s,t)$, which gives the relative probability of finding the state s in the ensemble at time t. The ensemble average of a physical quantity $O(s)$ is then given by

$$\langle O\rangle = \frac{\sum_s O(s)\rho(s,t)}{\sum_s \rho(s,t)} \tag{1.1}$$

where the sum over states s means integration over continuous variables. The equilibrium ensemble is characterized by a time-independent density function $\rho_{\mathrm{eq}}(s) = \lim_{t\to\infty} \rho(s,t)$. Generally we assume that $\rho_{\mathrm{eq}}(s)$ depends on s only through the Hamiltonian: $\rho_{\mathrm{eq}}(s) = \rho(H(s))$.

1.2. Microcanonical Ensemble and Entropy

The simplest equilibrium ensemble is a collection of equally weighted states, called the *microcanonical ensemble*. To be specific, consider an isolated macroscopic system with conserved energy. We assume that all states with the same energy E occur with equal probability. Other parameters not explicitly mentioned, such as the number of particles and volume, are considered fixed properties. The phase-space volume

此为试读，需要完整PDF请访问：www.ertongbook.co