

# MATHEMATICAL MODELS OF PROTEIN FOLDING

DANIEL B. DIX

Department of Mathematics  
University of South Carolina

ABSTRACT. We present an elementary introduction to the protein folding problem directed toward applied mathematicians. We do not assume any prior knowledge of statistical mechanics or of protein chemistry. Our goal is to show that as well as being an extremely important physical, chemical, and biological problem, protein folding can also be precisely formulated as a set of mathematics problems. We state several problems related to protein folding and give precise definitions of the quantities involved. We also attempt to give physical derivations justifying our model.

## CONTENTS

- 1 The Protein Folding Problem
- 2 Classical Statistical Mechanics
- 3 The Born-Oppenheimer Potential
- 4 The Isobaric-Isothermal Ensemble
- References

## 1 THE PROTEIN FOLDING PROBLEM

Proteins are large polymeric molecules which form the predominate part of all living matter. They are polymers since they are formed by connecting small molecules, from a pool of twenty different types of amino acids, into long flexible linear chains. The twenty amino acids which are used in living organisms vary in size from 7 to 24 atoms, about half of which are Hydrogen atoms. This is what we mean by a small molecule. The only other atoms involved are Carbon, Nitrogen, Oxygen, and Sulfur. From these five types of atoms the amino acids are constructed in the general form:

In the above  $R$  stands for what is called the side chain. Each of the twenty amino acids has a different structure in its side chain. These differences give the amino acids a variety of chemical properties in a neutral solution of water. All cells synthesize proteins as long linear chains (consisting of between 50 and 1000 monomers) of the amino acid subunits. Two amino acids can be linked together by attaching the amino group of one to the carboxy group of the other, forming a new Carbon-Nitrogen bond called the peptide bond, with the removal of a water molecule. Thus in a chain of amino acids, there is only one exposed amino terminus and one exposed carboxy terminus. This linear arrangement makes it easier to synthesize polymers with a very specific order of the monomer subunits. After synthesis this long chain is released into the solvent (water), which begins to interact with the side chains according to their properties. This interaction does not involve the formation or breakage of covalent bonds, but does involve formation and breakage of much weaker hydrogen bonds, and other types of weaker chemical interactions. What ensues is a complex folding process where the protein chain forms a compact stable shape, pushing all the water molecules to its exterior. This shape is the active or native conformation of the protein.

Proteins can be thought of as little molecular machines, which process other molecules in very specific, repeatable and controllable ways. It is also remarkable that these machines assemble themselves automatically—as we have just described. The shape that the finally assembled protein takes is controlled by the specific sequence of amino acids used to synthesize the protein chain. Hence “machine design” is very compactly coded in the sequence of amino acids. The sequence of amino acids is coded in the DNA sequence of the gene for that protein. Hence if the DNA sequence is mutated, i.e. changed in some way, the resulting protein will be affected. The shape of the protein determines whether or not it will be able to do its job(s) properly. Hence organisms can be understood to evolve by discrete changes in the DNA and amino acid sequences, resulting in differently shaped proteins with different functionalities. The problem for understanding this evolution process in detail is that the code translating an amino acid sequence into an active protein machine is an analog code, implemented by a complicated physical process. If we know the mutated amino acid sequence, but cannot predict the shape of the resulting protein because of the complexity of the folding code, then we are hindered in making a theoretical assessment of the effect of the mutation on the organism. Biologists try to work around this problem by experimental observation of micro-evolutionary changes, where a surprising amount of information about the molecular details can be recovered. However, much macro-evolutionary speculation involves transformations occurring in hypothetical organisms, transformations of a type which have not been observed in the laboratory, nor are they likely to be observed. Organisms are integrated systems operating according to their own internal logic, which possesses a certain degree of plasticity. But the central assertion that organisms can be step-wise modified (through well understood DNA mutational events) so that at each step the organism is viable and competitive in its environment, and such that the overall result is that e.g. procaryotes (i.e. bacteria) evolve into eucaryotes (higher organisms capable of multicellular organization), or other such macro-evolutionary steps, has not been demonstrated. Hence a theoretical and detailed understanding of hypo-

thetical macro-evolutionary processes at the molecular level involves breaking the “folding code”.

Because of the central role of proteins as the “movers and shakers” of the cell, it is safe to say that many diseases are affected by some protein machine that is not doing its job properly. Most drugs are smaller molecules which attach to some protein and turn it on or off, or change its function in some way. In the past we did not know these details about how particular drugs work, but now this is changing. In the past drugs were found by trial and error. Furthermore these drug molecules have side effects: not only do they affect the problem protein, but they also affect many other good proteins, making them not work as well as before. However if we understand the exact shape of the problem protein, we can design drug molecules which will efficiently affect only the problem protein. This will result in highly effective drugs with no side effects. Thus we need to understand all the proteins in the organism and how they interact with one another. The easiest way to find all the proteins in an organism is to completely sequence all the DNA in that organism’s genome (its chromosomes). For humans this project is called the Human Genome Project. This information will tell us the amino acid sequences of all the proteins. But unless we can break the “folding code” we will not know the shape of the protein determined by a given amino acid sequence, hence we are greatly hindered in carrying out this program. The interactions of proteins in biochemical networks are controlled by many of the same physical interactions which are so important during the folding process. Hence understanding the folding process and being able to predict the folded conformation from the amino acid sequence is a fundamental scientific problem of great biological and medical importance.

The “Protein Folding Problem” thus has two parts. The static problem is to predict the active conformation of the protein given only its amino acid sequence. The dynamic problem is to rationalize the folding process as a complex physical interaction. In some sense it is misleading to speak of “the” folded conformation. Many proteins undergo controlled conformational changes as an integral part of their activity, like a machine with moving parts. Some proteins have regions where the structure is random or amorphous, and yet for some purposes this is a help rather than a hindrance. Thus we wish to set up a mathematical model for a protein molecule in solution that will encompass both the static and dynamic aspects.

## 2 CLASSICAL STATISTICAL MECHANICS

Proteins typically contain tens of thousands of atoms, and their structure is strongly dependent on their interactions with thousands of water molecules. Physically realistic mathematical models must be able to predict the positions and velocities of all the atoms comprising the protein. Hence we will consider a many body problem. Atoms interact in complicated ways which require a quantum mechanical description for greatest accuracy. The stability of the atom as a system of charged particles was the puzzle which led to the development of quantum mechanics. The electrons cannot be treated as classical particles at all, but the nuclei are thousands of times heavier than electrons, and hence are more amenable to classical treatment. We will treat each atom as a single classical

particle, which we will think of as the nucleus of the atom. We will imagine these nuclei interacting via swarms of electrons which adjust so quickly to any change in the nuclear configuration that we can treat this adjustment time as instantaneous. Hence the effect of all the electrons is that each nucleus feels a classical conservative force determined only by the relative positions of all the nuclei. If  $r_1, \dots, r_N$  are the vector positions of the  $N$  nuclei (relative to some inertial reference frame) then this configuration possesses a potential energy  $V(r_1, \dots, r_N)$  due to the repulsion of the positively charged nuclei compensated by the swarm of negatively charged electrons. Exactly how this potential energy function should be determined in principle is explained in the next section. The physical interaction potential  $V(r_1, \dots, r_N)$  will be assumed to be a smooth function for  $r_i \neq r_j, 1 \leq i < j \leq N$ . (This is not physically obvious; if not true then we must use a smooth approximation to the physical  $V$  and hope for the best.) Also we may assume that  $V(r_1, \dots, r_N) \rightarrow \infty$  whenever  $r_i - r_j \rightarrow 0$ , since when close to each other the positively charged nuclei will strongly repel.

Suppose the particles have masses  $m_1, \dots, m_N$ , and momenta  $p_1, \dots, p_N$ . It is necessary that these particles be confined to a bounded spatial region. So suppose  $\mathcal{R}(1) \subset \mathbb{R}^3$  is a bounded connected open subset containing the origin, with unit volume, and with smooth boundary  $\partial\mathcal{R}(1)$ . We also suppose that  $\mathcal{R}(\lambda^3) \stackrel{\text{def}}{=} \lambda\mathcal{R}(1) = \{\lambda x \mid x \in \mathcal{R}(1)\}$  for all  $\lambda > 0$ . Therefore  $\mathcal{R}(\Omega)$  is a region with volume  $\Omega > 0$ . Suppose  $0 < \epsilon \ll 1$  and  $V_\epsilon: \mathbb{R}^3 \times (0, \infty) \rightarrow [0, \infty]$  is a function such that  $V_\epsilon(x, 1) = 0$  if  $x \in \mathcal{R}(1)$  is a distance greater than  $\epsilon$  from  $\partial\mathcal{R}(1)$ ,  $V_\epsilon(x, 1) = \infty$  if  $x \notin \mathcal{R}(1)$ ,  $V_\epsilon(x, 1)$  is a smooth function for  $x \in \mathcal{R}(1)$ , and  $V_\epsilon(x, 1) \rightarrow \infty$  as  $x \rightarrow \partial\mathcal{R}(1)$ . We also suppose that  $V_\epsilon(x, \lambda^3) = V_\epsilon(x/\lambda, 1)$  for all  $x \in \mathbb{R}^3$  and all  $\lambda > 0$ . Then the system has the total energy

$$H(r_1, \dots, r_N, p_1, \dots, p_N, \Omega) = \sum_{j=1}^N \frac{1}{2m_j} \|p_j\|^2 + V(r_1, \dots, r_N) + \sum_{j=1}^N V_\epsilon(r_j, \Omega).$$

Clearly  $H(r, p, \Omega) = \infty$  unless  $r_j \in \mathcal{R}(\Omega)$  for all  $1 \leq j \leq N$ ; we will be able to confine our attention to the case where  $H(r, p, \Omega) < \infty$ . These “containment” potential energy terms are unphysical, but very convenient. When the volume  $\Omega$  is independent of time we have an isolated system; the pressure could depend on the conformation of the protein. Proteins usually fold in an environment where the pressure is constant; we will model this refinement later by allowing the volume to “fluctuate”.

When  $\Omega$  is constant in time, these particles will move according to the evolution equations of classical Hamiltonian mechanics, namely

$$\frac{dr_j}{dt} = \nabla_{p_j} H, \quad \frac{dp_j}{dt} = -\nabla_{r_j} H, \quad j = 1, \dots, N.$$

We consider this evolution on the phase space  $\Phi_\Omega = T^*[\mathcal{R}(\Omega)^N] = \mathcal{R}(\Omega)^N \times (\mathbb{R}^{3N})^*$ . By Sard’s theorem almost every finite value  $E$  of  $H$  is a regular value, and hence  $\{(r, p) \in \Phi_\Omega \mid H(r, p, \Omega) = E\}$  is a compact smooth submanifold of  $\Phi_\Omega$  without boundary on which the Hamiltonian vector field lives and is complete. Hence for  $(r, p)$  in such a submanifold, a global in time solution  $\phi_t(r, p)$  of the above system of evolution equations exists.

The phase space and flow described above is extremely general, and can be used to model any chemical transformation of the set of  $N$  atoms under consideration. However, at the outset we want to specialize to the particular type of chemical phenomenon of interest, namely protein folding. Thus our system of  $N$  atoms consists of a protein molecule (in its dominant ionized form at neutral pH) and some number of (non-ionized) water molecules. In a water molecule we know experimentally the average hydrogen-oxygen bond length and the H-O-H bond angle. In a protein we also know the average bond lengths and the average bond angles of all the covalently bonded pairs and triples in the molecule. We also know some dihedral angles, such as those which determine the chirality at asymmetric Carbons (if A-B-C is a covalently bonded triple and D is covalently bonded to either B or C then the dihedral angle is the angle between the planes ABC and BCD). These quantities do not change appreciably during protein folding. We do not know the *all* the dihedral angles and consequently we do not know the protein conformation (which changes during protein folding). Let  $f_j(r)$  for  $j = 1, \dots, J$  be all the functions, either of bond length type

$$f(r_A, r_B) = \|r_A - r_B\|,$$

of bond angle type

$$f(r_A, r_B, r_C) = \cos^{-1} \left[ \frac{(r_A - r_B) \cdot (r_C - r_B)}{\|r_A - r_B\| \|r_C - r_B\|} \right],$$

or of dihedral angle type (if  $u \in \mathbb{R}^3$  is a column vector then  $u^T$  is its transpose)

$$f(r_A, r_B, r_C, r_D) = \theta, \quad \text{where } v^T w = \|v\| \|w\| \cos(\theta), v \times w = \|v\| \|w\| \sin(\theta)u,$$

$$u = \frac{r_C - r_B}{\|r_C - r_B\|}, v = (1 - uu^T)(r_A - r_B), w = (1 - uu^T)(r_D - r_C),$$

needed to specify these known degrees of freedom for the system of protein plus water molecules. Let  $\bar{f}_j$  denote their experimentally known values. Consider the subset  $\mathcal{E}$  of  $\mathcal{R}(\Omega)^N$  of all  $r = (r_1, \dots, r_N)$  such that  $f_j(r) = \bar{f}_j$  for all  $1 \leq j \leq J$  and no steric conflict exists between two non-bonded atoms (a steric conflict is where the distance between two atoms is less than the sum of their van der Waals radii; a pair of atoms is considered non-bonded if we have not specified a bond length for that pair). The interesting region of configuration space certainly includes  $\mathcal{E}$ . However, under normal protein folding conditions, even if we start a trajectory of the equations of motion on or near  $\mathcal{E}$  (e.g. with zero momentum) this trajectory (in its position component) will not in general stay near  $\mathcal{E}$ . The reasons for this are varied, depending on the (conserved) energy of the trajectory. At reasonably low energies there is the phenomenon of hydrogen exchange. When a hydrogen bond is formed between two water molecules, or between a water molecule and a partially charged site on the protein, the energy barrier for the proton to migrate along the hydrogen bond so that the covalent bonding network changes is not very high. Since the network of hydrogen bonds in the liquid solvent is constantly changing, a particular hydrogen ion can be passed around quite a bit. After a short time we cannot tell which atom the hydrogen

will be bonded to. Since the subset  $\mathcal{E}$  describes situations where we know exactly which atom each hydrogen is bonded to, this subset is only a good starting point for our phase trajectories. At much higher energies we can obtain various reactions which degrade the protein in some way. For example we can get chirality switching isomerizations at the asymmetric Carbons; also Sulfur atoms can be oxidized, and Glutamine and Asparagine residues can become deamidated. At extremely high energies peptide bonds might be hydrolyzed. Since these degrading reactions can produce anomalous conformations of the protein, we would like to exclude them from our consideration. Hence we introduce an energy cutoff  $E_c$  which is as large as possible so that phase points  $(r, p)$  with  $H(r, p, \Omega) < E_c$  do not contain degraded configurations  $r$ . Our phase space  $\Phi_{\Omega, E_c}$  will then be the union of the connected components of  $\{(r, p) \in \Phi_{\Omega} \mid H(r, p, \Omega) < E_c\}$  non-trivially intersecting  $\mathcal{E} \times \{0\}$ . See section 3 for more discussion.

The classical phase space  $\Phi = T^*\mathbb{R}^{3N} = \mathbb{R}^{3N} \times (\mathbb{R}^{3N})^*$  has a natural measure, called the Liouville measure, which is simply

$$d\mu = \prod_{j=1}^N \frac{d^3 r_j d^3 p_j}{h^3}.$$

In the above  $h$  is Planck's constant, which has dimensions of length times momentum. A set of Liouville measure 1 is considered a single *microstate*; to know that the system phase point lies in such a set is the maximal information which can be known. (This assertion is not supposed to be obvious; its justification comes from quantum mechanics.) The set  $\{(r, p) \in \Phi_{\Omega, E_c} \mid H(r, p, \Omega) \text{ is a regular value of } H\}$  is a Liouville conull set, i.e. its complement in  $\Phi_{\Omega, E_c}$  has Liouville measure zero. Hence  $\phi_t(r, p)$  is defined for all  $t \in \mathbb{R}$  for Liouville-almost all  $(r, p) \in \Phi_{\Omega, E_c} \subset \Phi$ . It is a theorem of classical mechanics that the Hamiltonian flow  $\{\phi_t\}$  preserves this measure, i.e. if  $A \subset \Phi_{\Omega, E_c}$  is Borel measurable, then  $\mu(\phi_{-t}(A)) = \mu(A)$  for all  $t \in \mathbb{R}$ . The flow is in this sense "incompressible".

If the initial phase point  $(r, p)$  were known with exactitude, then we could in principle find its entire past and future behavior by simply solving the initial value problem for the Hamiltonian equations of motion. But in practice we know that exact solutions are unknown and numerical estimates will introduce errors whose size will grow rapidly due to the chaotic nature of the dynamics; after a relatively short time the numerical solutions cannot be expected to be quantitatively reliable. Furthermore, an exact knowledge of the initial positions and momenta of all the particles in the system is very unrealistic. Hence we describe our incomplete knowledge about the state of the system with a probability measure on  $\Phi_{\Omega, E_c}$ . If this measure is absolutely continuous with respect to the Liouville measure then it is given by a probability density function  $\rho(r, p, t)$ . Since we know how an exactly known phase point should evolve in time (along the incompressible flow), this probability density should be transported along with the flow, i.e.  $\rho(\phi_t(r, p), t) = \rho(r, p, 0)$ . Hence  $\rho$  satisfies the *Liouville equation*

$$\frac{\partial \rho}{\partial t} = \{H, \rho\} \stackrel{\text{def}}{=} \sum_{j=1}^N (\nabla_{r_j} H \cdot \nabla_{p_j} \rho - \nabla_{p_j} H \cdot \nabla_{r_j} \rho).$$

This partial differential equation is the fundamental equation of statistical mechanics for isolated systems; it is solvable in principle by the method of characteristics, which would however lead us right back to the Hamiltonian evolution equations.

This is a mathematical framework capable in principle of describing the protein folding process. First one must specify the initial probability density  $\rho(r, p, 0)$ , say in such a way that is consistent only with an unfolded conformation of the protein. (We will discuss the proper way to do this presently.) Then the solution  $\rho(r, p, t)$  of the Liouville equation describes for each time  $t$  the probability distribution in phase space of the system. However this description does not give us information about particular trajectories. Certain folding intermediates may appear during the folding process as regions of phase space where  $\rho$  is relatively concentrated. As folding progresses we expect that the bulk of the probability density will shift from the earlier intermediates to the later intermediates. Finally, a steady state probability distribution which is heavily concentrated around the folded conformation will be observed, in the typical time required for folding (usually less than 100 seconds). The final “equilibrium” distribution  $\lim_{t \rightarrow \infty} \rho(r, p, t) = \rho_\infty(r, p)$  will still have some probability density in unfolded conformations. This means that even when the system is described by the density  $\rho_\infty(r, p)$  there is a small probability that at any particular time the protein is unfolded. The time independence of  $\rho_\infty$  does not imply that the protein itself is not moving.

We have been describing the “dynamic” picture of protein folding because  $\rho(r, p, 0)$  is not completely in equilibrium. The “static” picture is given by the distribution  $\rho_\infty(r, p)$ . The static picture however also yields a great deal of important information. The folded conformation is where  $\rho_\infty$  is greatest. From the dependence on  $p$  information about the typical motions experienced by the protein can be discerned. If there are several interconvertible conformations, that should be also evident. The folding intermediates should have a trace presence in  $\rho_\infty$ , since the later intermediates are probably also on the most likely unfolding pathways. However this information would be much more difficult to separate from noise. The dynamic picture presumably involves techniques from non-equilibrium statistical mechanics, a discipline which is not nearly as well understood or settled as its equilibrium counterpart.

Any reasonable choice of the initial probability density  $\rho(r, p, 0)$  should entail only the knowledge that we actually possess about the initial state. Suppose particles 1 through  $M$  are the non-hydrogen atoms of the protein. One measurable quantity that relates directly to the folded/unfolded state of the protein is the *radius of gyration*

$$R_G(r) = \left\{ \frac{1}{M} \sum_{i=1}^M \left\| r_i - \left( \sum_{j=1}^M m_j \right)^{-1} \left( \sum_{j=1}^M m_j r_j \right) \right\|^2 \right\}^{1/2}.$$

This quantity can be inferred from measurements of solution viscosity. One very small and well studied protein is Bovine Pancreatic Trypsin Inhibitor (BPTI), which has 58 amino acids. When this chain is fully extended it is 211 Å long. If the peptide chain is allowed to assume a random conformation the radius of gyration is about 35 Å. However the folded conformation is much more compact, with a maximum end to end length of 29 Å. Clearly

the radius of gyration shrinks very significantly during folding. If we knew that

$$\overline{R_G} = \int_{\Phi_{\Omega, E_c}} R_G(r) \rho(r, p, 0) d\mu$$

was equal to 35 Å then we could be certain that the protein BPTI was significantly unfolded. The physical interpretation of this constraint is that during protein synthesis there is an interaction with the ribosome which produces the polypeptide chain in an unfolded state. Our model applies after this interaction with the ribosome has been turned off, and the conditions are favorable for folding.

Other quantities that can be known are the macroscopic thermodynamical variables, such as internal energy. In our case, since we have a closed isolated system with a fixed volume and number of particles, in order to fix a particular state of thermal equilibrium it is customary to suppose that the internal energy is also fixed and given. The (initial) internal energy is defined by

$$U = \overline{H} = \int_{\Phi_{\Omega, E_c}} H(r, p, \Omega) \rho(r, p, 0) d\mu.$$

The given value of  $U$ , as well as that of  $N$  and  $\Omega$ , had better be consistent with room temperature or body temperature, otherwise covalent bonds that we have specified above would not be stable, and again we would not be discussing protein folding.

In living cells proteins are fabricated by the ribosome. This is an extremely large (by molecular standards) and complex machine which can manufacture any specific sequence of amino acids according to an informational template, the messenger RNA. The fabrication process involves the formation of new chemical bonds (specifically peptide bonds) and relative internal motions of the parts of the ribosome. These require the consumption of chemical energy, and the excess energy is dissipated into the surrounding cytosol (the cellular fluid, mostly water). The procaryotic ribosome adds on the average 15 new amino acids to the growing polypeptide chain every second. The newly added amino acids pass through a narrow channel in the body of the ribosome before they emerge into the cytosol. This channel is about 30 amino acids in length. The chain emerges from this hole in the ribosome on the opposite side from where all the energy is being consumed and dissipated. Hence when the new polypeptide chain first encounters an environment suitable for folding, it is close to thermal equilibrium. The amino terminus emerges first, and so this end of the chain has early exposure to folding conditions. In large multi-domain proteins, the individual domains may fold as they emerge—before the entire chain is released from the ribosome. In smaller single domain proteins, it is not known if significant progress toward folding is made before the entire chain is released from the ribosome and exposed to solvent. For an 180 amino acid single-domain protein, synthesis is complete in 12 seconds, whereas folding is usually complete in less than 100 seconds. In our simple model we will assume the entire chain is free from the ribosome and is surrounded by water near thermal equilibrium.

For simplicity, let us suppose we know at time 0 only the two numbers,  $\overline{R_G}$  and  $U$ . We wish to find the probability density  $\rho(r, p, 0)$  which is as unrestrictive as possible



subject to these constraints. To assume that the initial distribution is as information deficient as possible subject to these constraints is to acknowledge that thermal fluctuations destroy information, and to assume that this process is very close to completion even at the beginning of folding. This makes sense for folding off a ribosome, since most of the chain has been exposed to solvent for some time before the carboxy terminus is released. As we have hypothesized, the effect of the ribosome is to cause the initial radius of gyration to be too large.

Fortunately we can quantify the amount of information that the state description  $\rho$  is lacking in comparison with the optimal possible information, i.e. that the system phase point lies in a set of Liouville measure 1. This *missing information* is given by the entropy

$$S(\rho) = -k \int_{\Phi_{\Omega, E_c}} \rho(r, p) \ln \rho(r, p) d\mu.$$

In the above,  $k$  is Boltzmann's constant. Hence we wish to choose  $\rho(r, p, 0)$  so that  $S(\rho)/k$  is maximized subject to the constraints we have already discussed. This method for choosing a measure on phase space is called *Jaynes' principle*. This optimization problem can be solved by introducing Lagrange multipliers  $\beta$ ,  $\lambda$ , and  $\ln(Z/e)$ , with the result that

$$\rho(r, p, 0) \stackrel{\text{def}}{=} \exp[-\beta H(r, p, \Omega) - \lambda R_G(r)]/Z,$$

where the Lagrange multipliers are determined by the constraints

$$\begin{aligned} Z &= \int_{\Phi_{\Omega, E_c}} \exp[-\beta H(r, p, \Omega) - \lambda R_G(r)] d\mu, \\ U &= \frac{1}{Z} \int_{\Phi_{\Omega, E_c}} H(r, p, \Omega) \exp[-\beta H(r, p, \Omega) - \lambda R_G(r)] d\mu, \\ \overline{R_G} &= \frac{1}{Z} \int_{\Phi_{\Omega, E_c}} R_G(r) \exp[-\beta H(r, p, \Omega) - \lambda R_G(r)] d\mu, \end{aligned}$$

Because  $Z$  is a function of the other two multipliers, they can be found by solving simultaneously the last two equations.

**Existence Problem.** *Does a simultaneous solution  $(\beta, \lambda)$  of this pair of equations exist?*

The Lagrange multiplier  $\beta$  usually has the interpretation of  $\beta = 1/(kT)$ , where  $T$  is the absolute temperature. This is not exactly true in our situation. In all the standard equilibrium ensembles the following integral has the value  $3kT$ , which suggests that at equilibrium temperature should be related to the average kinetic energy of each of the particles. By integration by parts we can prove for all  $1 \leq j \leq N$  that

$$\int_{\Phi_{\Omega, E_c}} p_j \cdot \nabla_{p_j} H \rho(r, p, 0) d\mu = \frac{3}{\beta} \left[ 1 - \frac{1}{Z} \int_{\Phi_{\Omega, E_c}} \exp[-\beta E_c - \lambda R_G(r)] d\mu \right].$$

Thus the energy cutoff causes a reinterpretation of  $\beta$ , but this is a relatively small change if  $E_c - \min H$  is large.

It can be shown that if a simultaneous solution  $(\beta, \lambda)$  exists then it is unique. The existence of a solution indicates that the assumed values of  $\overline{R_G}$ , as well as  $N$ ,  $\Omega$  and  $T$  are properly “tuned”. Our physical intuition suggests that a solution will exist for a range of  $\overline{R_G}$  between the folded value and the completely extended value.

**Dynamic Problem.** *If  $\rho(r, p, t)$  is the solution of Liouville’s equation with the above choice of the initial condition  $\rho(r, p, 0)$  then what is the large time behavior of  $\rho(r, p, t)$ ? Does  $\rho(r, p, t)$  stabilize to a distribution  $\rho_\infty(r, p)$  as  $t \rightarrow \infty$ , and if so what are its characteristics?*

This scenario could be called, *protein folding in the micro-canonical ensemble*. It is perhaps the simplest physically realistic model of the protein folding process. The reader might think that we really mean the canonical ensemble, rather than the micro-canonical ensemble, since our initial distribution is very close to that for the canonical ensemble. However, our system is evolving in a strictly isolated manner—there is no heat flowing into or out of the system from the environment. However we define temperature during this non-equilibrium process, we do not expect it to remain precisely constant.

Our initial condition is consistent with a thermally non-isolated system in equilibrium (except for the energy cutoff). But then we evolve this initial probability distribution as if it were thermally (and mechanically) isolated. This isolation is not physically realistic. But if there are enough water molecules, the protein molecule will probably not be able to tell the difference. The main point is that the protein not be isolated from the solvent at approximately constant temperature and pressure conditions. Thus we will need  $N$  to be fairly large to simulate this. But it is not necessary that we pass to the thermodynamic limit of extremely large  $N$ ,  $\Omega$ , and  $U$ .

Because of the complexity of the Hamiltonian flow it would seem hopeless to address the dynamic problem as we have formulated it. However, the paucity of information in the initial condition, and its relation to the Hamiltonian suggests that all the intricacies of the Hamiltonian flow may not have to be considered in order to accurately approximate this particular solution of Liouville’s equation.

As we mentioned earlier, if we try to follow the dynamics under conditions of constant temperature and pressure things become even more complicated. However, we will be able to formulate the static problem under constant temperature and pressure conditions (see section 4). But first we should note that much of the physical information that makes this a protein folding problem as opposed to a general problem in statistical mechanics is in the potential  $V(r)$ . It is this subject that we now take up.

### 3 THE BORN-OPPENHEIMER POTENTIAL

We suppose that the  $N$  nuclei are fixed in space at positions  $r_1, \dots, r_N$ . Now we will describe without a lot of commentary how the potential energy  $V(r_1, \dots, r_N)$  should be defined. This will be using what in quantum chemistry is called the *Born-Oppenheimer approximation*, namely we suppose the nuclei are so much heavier than the electrons that

we can assume they are motionless. Suppose the nuclei have charges  $Z_1e, \dots, Z_Ne$ , where  $Z_1, \dots, Z_N$  are positive integers and  $e$  is the elementary charge. The electron has charge  $-e$ . Suppose these nuclei are surrounded by  $N_e$  electrons, with positions  $x_1, \dots, x_{N_e}$ , all ranging over  $\mathbb{R}^3$ . Let  $m_e$  be the mass of the electron. Let  $\sigma_1, \dots, \sigma_{N_e}$  be the discrete spin variables of the electrons:  $\sigma_j \in \{0, 1\}$  for each  $j$ . We abbreviate  $x$  for  $(x_1, \dots, x_{N_e})$  and  $\sigma$  for  $(\sigma_1, \dots, \sigma_{N_e})$ . We define a Hilbert space  $\mathfrak{H}$  consisting of measurable complex-valued functions  $\psi(x, \sigma)$  satisfying

$$\psi(\dots, (x_k, \sigma_k), \dots, (x_j, \sigma_j), \dots) = -\psi(\dots, (x_j, \sigma_j), \dots, (x_k, \sigma_k), \dots).$$

These conditions express the Pauli exclusion principle. In the above we use the notation  $\psi(x, \sigma) = \psi((x_1, \sigma_1), \dots, (x_{N_e}, \sigma_{N_e}))$ . We require that our functions  $\psi$  have finite norm with respect to the inner product  $\langle\langle \cdot, \cdot \rangle\rangle$  defined as:

$$\begin{aligned} \langle\langle \psi, \phi \rangle\rangle &= \int_{x \in \mathbb{R}^{3N_e}} \langle \psi(x, \cdot), \phi(x, \cdot) \rangle d^3x_1 \dots d^3x_{N_e}, \\ \langle \psi(x, \cdot), \phi(x, \cdot) \rangle &= \sum_{\sigma_1=0}^1 \dots \sum_{\sigma_{N_e}=0}^1 \overline{\psi(x, \sigma)} \phi(x, \sigma). \end{aligned}$$

On functions  $\psi$  which are smooth in the  $x$  variables and with compact support we define a Hamiltonian operator

$$\begin{aligned} H\psi \stackrel{\text{def}}{=} & -\frac{\hbar^2}{2m_e} \sum_{j=1}^{N_e} \Delta_j \psi + ke^2 \sum_{j=2}^N \sum_{k=1}^{j-1} \frac{Z_j Z_k}{\|r_j - r_k\|} \psi + ke^2 \sum_{j=2}^{N_e} \sum_{k=1}^{j-1} \frac{1}{\|x_j - x_k\|} \psi \\ & - ke^2 \sum_{j=1}^N \sum_{k=1}^{N_e} \frac{Z_j}{\|r_j - x_k\|} \psi. \end{aligned}$$

$\Delta_j$  denotes the Laplacian in the  $x_j$  variables.  $\hbar = h/(2\pi)$  in terms of Planck's constant.  $k$  denotes the Coulomb force constant. In order of occurrence on the right we have the kinetic energy of the electrons, the repulsive energy of the nuclei, the repulsive energy of the electrons, and the mutual attraction of the electrons for the nuclei. The operator  $H$  depends on  $(r_1, \dots, r_N)$  as parameters. Finally we define

$$V(r_1, \dots, r_N) \stackrel{\text{def}}{=} \inf \frac{\langle\langle \psi, H\psi \rangle\rangle}{\langle\langle \psi, \psi \rangle\rangle}, \quad 0 \neq \psi \in \mathfrak{H} \cap C_c^\infty(\mathbb{R}^{3N_e}).$$

**Smoothness Problem.** *Is  $V$  a smooth function of its arguments, provided of course that none of the nuclei occupy the same point of space?*

In the previous section we introduced an energy cutoff  $E_c$  with the idea that trajectories with energies below the cutoff would not experience degradation. This idea really entails an assumption about the behavior of  $V(r)$ , namely that there is a reasonably large value  $E_c$  with this property. This could fail if for example these degradatory reactions happen slowly not because of the high energy they require but because the saddle point representing the transition state for one of these reactions is in an extremely narrow “mountain pass”.

**Energy Cutoff Problem.** *Show that for the above function  $V(r)$  there exists an energy cutoff  $E_c$  such that whenever  $(r, p) \in \Phi_{\Omega, E_c}$  the configuration  $r$  contains no degradation of the protein. Does  $E_c$  depend on  $\Omega$ ?*

The potential  $V(r)$  we have defined in this section is obviously translationally and rotationally invariant. Furthermore, when restricted to the configurational projection of the phase space  $\Phi_{\Omega, E_c}$ , certain configurational degrees of freedom are clearly strongly constrained, whereas others are much less so. These “constrained” degrees of freedom essentially execute vibrational motion.

**Separation of Variables Problem.** *Introduce new coordinates in configuration space so that the position of the center of mass, the overall rigid body orientation, and the vibrational normal modes are separated from the other degrees of freedom in the sense that tangent vectors to coordinate curves associated to a pair of separated coordinates are perpendicular with respect to the kinetic energy metric, and the Hessian of  $V$  is block diagonal in these new coordinates.*

#### 4 THE ISOBARIC-ISOTHERMAL ENSEMBLE

As we saw in section 2 the state of the system consisting of the protein and some water molecules under constant volume conditions is described by a probability measure on  $\Phi_{\Omega}$ . If we are content to understand merely the final equilibrium state  $\rho_{\infty}$  then we can take advantage of the better understood theory of equilibrium statistical mechanics. Knowing the volume  $\Omega$  exactly is an idealization of course, akin to knowing the exact value of the total energy  $H$ . When a fluid wants to expand it takes tremendous pressures to frustrate that desire. It is hard to know how to adjust the pressure to maintain constant volume; it is much easier to maintain constant pressure and allow the volume to fluctuate as it wants to. At equilibrium however we expect these volume fluctuations to take place about some average value which is fixed in time. Thus we will describe the equilibrium state as a probability measure on  $\Psi = \Phi \times (0, \infty)$ , which will be absolutely continuous with respect to the measure  $d\nu = d\mu d\Omega/\Omega^0$  with density  $\rho(r, p, \Omega)$ .  $\Omega^0$  is some standard volume, which we can take to be a unit volume in whatever system of units we are using. This normalization insures that the measure on  $\Psi$  is dimensionless, as is  $\rho$ . We will drop the  $\infty$  subscript, even though this equilibrium state should be the end state of some (undiscussed) time dependent process. Notice that we are using  $\Phi$  and not  $\Phi_{\Omega, E_c}$ ; hence we are not employing the energy cutoff that was introduced in section 2.

Equilibrium statistical mechanics has an independent method of deriving (meta-stable) equilibrium states depending only on the values of certain macroscopic conserved quantities. In our case we want the average values of  $H$ , and  $\Omega$  to be given. The equilibrium distribution satisfying these constraints should be found by employing Jaynes’ principle, i.e. we choose  $\rho$  so that the entropy

$$S(\rho) = -k \int_{\Psi} \rho \ln \rho d\nu$$

is maximized subject to the given constraints. As we saw in section 2 this maximization problem is solved using Lagrange multipliers  $\beta$ ,  $\eta$ , and  $\ln(Z/e)$ , with the solution

$$\rho(r, p, \Omega) = \exp[-\beta H(r, p, \Omega) - \eta\Omega]/Z,$$

where the Lagrange multipliers are determined by the constraints

$$\begin{aligned} Z &= \int_{\Psi} \exp[-\beta H(r, p, \Omega) - \eta\Omega] d\nu, \\ U &= \frac{1}{Z} \int_{\Psi} H(r, p, \Omega) \exp[-\beta H(r, p, \Omega) - \eta\Omega] d\nu, \\ \bar{\Omega} &= \frac{1}{Z} \int_{\Psi} \Omega \exp[-\beta H(r, p, \Omega) - \eta\Omega] d\nu. \end{aligned}$$

The Lagrange multipliers have macroscopic interpretations:

$$\beta = \frac{1}{kT}, \quad \eta = \beta\mathcal{P}, \quad \mathcal{P} = - \int_{\Psi} \frac{\partial H}{\partial \Omega} \rho d\nu$$

where  $T$  is the absolute temperature, and  $\mathcal{P}$  is the pressure. When we fix  $T$  and  $\mathcal{P}$  and determine  $U$  and  $\bar{\Omega}$  in terms of them, then we are said to be in an *isothermal-isobaric ensemble*.

This probability distribution describes a system which is neither thermally nor mechanically isolated from its environment. In this situation there is no obvious simple way to trace the trajectories of the system. The probability density decays exponentially as the micro-enthalpy  $H + \mathcal{P}\Omega$  increases, but there is no cutoff, i.e. there are regions of small positive probability which have arbitrarily large micro-enthalpy. Also note that there is no restriction to a region of phase space which concerns a protein; actually any collection of chemical species which can be formed from the set of  $N$  atoms is allowed. However, we can conditionalize this probability to the set

$$\Psi_{E_c} = \{(r, p, \Omega) \in \Psi \mid (r, p) \in \Phi_{\Omega, E_c}\},$$

where  $E_c$  is the cutoff energy introduced in section 2. The conditional density  $\rho_{E_c}$  is defined exactly like  $\rho$  except  $Z$  should be replaced by

$$Z_{E_c} = \int_{\Psi_{E_c}} \exp[-\beta H(r, p, \Omega) - \eta\Omega] d\nu.$$

Thus  $\rho$  already (before conditionalization) gives us the information about the *relative* importance of different regions of  $\Psi_{E_c}$  in regard to the folded state.

In elementary chemistry one considers the reactants and products of a chemical reaction and expresses the ratio of equilibrium concentrations of products to concentrations of reactants to the difference of the Gibbs free energies of the products and the reactants. In

our scheme, a particular configuration of the chemical elements involved can be identified with a measurable subset of  $\Psi$ . The reactants would be given by one subset  $\mathcal{C}_r$  and the products by another subset  $\mathcal{C}_p$ . The *Gibbs Free Energy*  $G(\mathcal{C})$  of a subset  $\mathcal{C} \subset \Psi$  is defined to be such that

$$e^{-\beta G(\mathcal{C})} = Z_{\mathcal{C}} = \int_{\mathcal{C}} \exp[-\beta H(r, p, \Omega) - \eta \Omega] d\nu.$$

Hence  $e^{-\beta G(\mathcal{C})}/Z$  is the probability the system phase point is in the set  $\mathcal{C}$ ; alternately it is the fraction of an ensemble of systems whose system phase point is in  $\mathcal{C}$ . Hence the ratio of two such probabilities, which can be equated to the ratio of the corresponding concentrations, is related in the usual way to the difference of the Gibbs free energies.  $G(\mathcal{C})$  can also be expressed in the usual way as the enthalpy minus temperature times entropy if we compute these quantities for the state conditionalized to the set  $\mathcal{C}$ :

$$\begin{aligned} U_{\mathcal{C}} &= \int_{\mathcal{C}} H(r, p, \Omega) \frac{\exp[-\beta H(r, p, \Omega) - \eta \Omega]}{Z_{\mathcal{C}}} d\nu, \\ \bar{\Omega}_{\mathcal{C}} &= \int_{\mathcal{C}} \Omega \frac{\exp[-\beta H(r, p, \Omega) - \eta \Omega]}{Z_{\mathcal{C}}} d\nu, \\ S_{\mathcal{C}} &= -k \int_{\mathcal{C}} \ln \left\{ \frac{\exp[-\beta H(r, p, \Omega) - \eta \Omega]}{Z_{\mathcal{C}}} \right\} \frac{\exp[-\beta H(r, p, \Omega) - \eta \Omega]}{Z_{\mathcal{C}}} d\nu \end{aligned}$$

The relation  $G(\mathcal{C}) = U_{\mathcal{C}} + \mathcal{P}\bar{\Omega}_{\mathcal{C}} - TS_{\mathcal{C}}$  clearly follows from these equations.

As we mentioned in section 2 we are primarily interested in the  $M$  non-hydrogen atoms of the protein. Thus we consider the reduced equilibrium state

$$\tilde{\rho}(r_1, \dots, r_M, p_1, \dots, p_M) = \int_0^\infty \int_{\mathbb{R}^{6(N-M)}} \frac{\exp[-\beta H(r, p, \Omega) - \eta \Omega]}{Z} \prod_{j=M+1}^N \left[ \frac{d^3 r_j d^3 p_j}{h^3} \right] \frac{d\Omega}{\Omega^0}.$$

This function contains the desired information about the folded conformation and its motions. We abbreviate  $\tilde{r} = (r_1, \dots, r_M)$  and  $\tilde{p} = (p_1, \dots, p_M)$ . Also abbreviate

$$d\hat{\nu} = \prod_{j=M+1}^N \left[ \frac{d^3 r_j d^3 p_j}{h^3} \right] \frac{d\Omega}{\Omega^0}.$$

We can also integrate over the variables  $p_1, \dots, p_M$  if we care only about conformations; reduced quantities of this type are usually called *conformational*. We will stay in the variables  $(\tilde{r}, \tilde{p})$ . The set  $\{(\tilde{r}, \tilde{p}) \mid \Omega > 0 \text{ and } (r, p, \Omega) \in \Psi_{E_c} \text{ for some extension of } (\tilde{r}, \tilde{p}) \text{ to } (r, p)\}$  will be called the *reduced phase space*.

If we consider subsets  $\mathcal{C}$  of the form  $\{(r, p, \Omega) \in \Psi \mid \bar{r}_j \leq r_j \leq \bar{r}_j + \Delta r_j, \bar{p}_j \leq p_j \leq \bar{p}_j + \Delta p_j, 1 \leq j \leq M\}$  and take the limit as  $\Delta r_j \rightarrow 0^+$ ,  $\Delta p_j \rightarrow 0^+$  then we are led by way of analogy with the above to the following definition of the reduced Gibbs free energy:

$$e^{-\beta \tilde{G}(\tilde{r}, \tilde{p})} = \tilde{Z}(\tilde{r}, \tilde{p}) = \int_0^\infty \int_{\mathbb{R}^{6(N-M)}} \exp[-\beta H(r, p, \Omega) - \eta \Omega] d\hat{\nu}.$$

It can be expressed in terms of reduced enthalpy and reduced entropy exactly as before:

$$\tilde{G}(\tilde{r}, \tilde{p}) = \tilde{U}(\tilde{r}, \tilde{p}) + \mathcal{P}\tilde{\Omega}(\tilde{r}, \tilde{p}) - T\tilde{S}(\tilde{r}, \tilde{p}),$$

where

$$\begin{aligned}\tilde{U}(\tilde{r}, \tilde{p}) &= \int_0^\infty \int_{\mathbb{R}^{6(N-M)}} H(r, p, \Omega) \frac{\exp[-\beta H(r, p, \Omega) - \eta\Omega]}{\tilde{Z}(\tilde{r}, \tilde{p})} d\hat{\nu}, \\ \tilde{\Omega}(\tilde{r}, \tilde{p}) &= \int_0^\infty \int_{\mathbb{R}^{6(N-M)}} \Omega \frac{\exp[-\beta H(r, p, \Omega) - \eta\Omega]}{\tilde{Z}(\tilde{r}, \tilde{p})} d\hat{\nu}, \\ \tilde{S}(\tilde{r}, \tilde{p}) &= -k \int_0^\infty \int_{\mathbb{R}^{6(N-M)}} \ln \left\{ \frac{\exp[-\beta H(r, p, \Omega) - \eta\Omega]}{\tilde{Z}(\tilde{r}, \tilde{p})} \right\} \frac{\exp[-\beta H(r, p, \Omega) - \eta\Omega]}{\tilde{Z}(\tilde{r}, \tilde{p})} d\hat{\nu}.\end{aligned}$$

Thus the reduced enthalpy is for the whole system, not just the protein, and it is averaged over the other degrees of freedom, where the protein degrees of freedom  $(\tilde{r}, \tilde{p})$  are held fixed. The reduced entropy is for a conditionalized state describing the non-protein degrees of freedom.

Also note that  $\tilde{\rho}(\tilde{r}, \tilde{p}) = e^{-\beta\tilde{G}(\tilde{r}, \tilde{p})}/Z$ , and hence

$$Z = \int_{\mathbb{R}^{6M}} e^{-\beta\tilde{G}(\tilde{r}, \tilde{p})} \prod_{j=1}^M \left[ \frac{d^3 r_j d^3 p_j}{h^3} \right].$$

It is for this reason that the static protein folding problem is usually stated in terms of minimizing  $\tilde{G}(\tilde{r}, \tilde{p})$ , although exactly what Gibbs free energy one is supposed to minimize is usually not made clear.

**Static Problem.** *Characterize the regions of reduced phase space where  $\tilde{\rho}(\tilde{r}, \tilde{p})$  is maximum, or where  $\tilde{G}(\tilde{r}, \tilde{p})$  is minimum.*

Despite our identification of the minimizers of  $\tilde{G}(\tilde{r}, \tilde{p})$  with the folded phase points, we cannot infer from any of our discussion that on a trajectory the quantity  $\tilde{G}(\tilde{r}, \tilde{p})$  has a tendency to decrease. This cannot be true in general, because folded proteins are known to spontaneously unfold.

For each choice of the thermodynamic variables  $N$ ,  $T$ , and  $\mathcal{P}$ , we obtain an equilibrium distribution  $\rho$ . The dependence of this distribution on these variables relates to *in vitro* folding experiments. For example there is experimental evidence that when  $N$  is very large, the distribution experiences a “first-order phase transition” as  $T$  is increased. Essentially at higher than usual temperatures the unfolded state becomes more stable than the folded state. This is called *temperature induced denaturation*. The great bulk of experimental information relates most directly to the static description.

In section 2 we conjectured that there is a function  $\rho_\infty(r, p)$  which represents the large-time limit of the protein-folding process in the micro-canonical ensemble. We can integrate out the non-protein degrees of freedom to obtain  $\tilde{\rho}_\infty(\tilde{r}, \tilde{p})$ . This can be compared to  $\tilde{\rho}(\tilde{r}, \tilde{p})$ . In the latter case the protein molecule is being adequately and realistically thermalized. In the former this is more and more true in the thermodynamic limit.

**Correspondence Problem.** *How do the functions  $\tilde{\rho}(\tilde{r}, \tilde{p})$  and  $\tilde{\rho}_\infty(\tilde{r}, \tilde{p})$  compare to one another?*

**Acknowledgments.** The author would like to thank his many colleagues for stimulating and encouraging conversations. Especially Ralph Howard (mathematics), Pavel Mazur (physics), Rick Creswick (physics), Benjamin Gimarc (chemistry), Mark Berg (chemistry), John Dawson (chemistry), Lukasz Lebioda (chemistry), Stephen Kistler (chemistry), Catherine Murphy (chemistry), Mihaly Czako (biology), and Bob Best (genetics).

#### REFERENCES

- B. R. Balian, *From Microphysics to Macrophysics, Volume 1*, Springer-Verlag, New York, 1991.
- BKP. C. L. Brooks, M. Karplus, B. M. Pettitt, *Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics*, Advances in Chemical Physics **71** (1988).
- C. R. E. Christoffersen, *Basic principles and techniques of molecular quantum mechanics*, Springer Advanced Texts in Chemistry, Springer-Verlag, New York, 1989.
- Cr. T. E. Creighton, *Proteins, Structures and Molecular Principles*, W.H. Freeman and Co., New York, 1983.
- Cr2. T. E. Creighton (ed.), *Protein Folding*, W.H. Freeman and Co., New York, 1992.
- M. P.G. Mezey, *Potential Energy Hypersurfaces*, Elsevier, New York, 1987.
- McQ. D.A. McQuarrie, *Statistical Mechanics*, Harper & Row, New York, 1973.
- MvH. C.K. Matthews, K.E. van Holde, *Biochemistry*, Benjamin/Cummings, New York, 1996.
- Neu. A. Neumaier, *Molecular modeling of proteins and mathematical prediction of protein structure*, SIAM Rev. **39** (1997), 407–460.
- NRC. National Research Council, *Mathematical Challenges from Theoretical/Computational Chemistry*, National Academy Press, Washington D.C., 1995.
- P. R.K. Pathria, *Statistical Mechanics*, Pergamon Press, Oxford, 1972.