Journal of Statistical Physics, Vol. 77, Nos. 5/6, 1994

A Measure on the Space of Polymer Folding Pathways: Preliminaries for a New Scheme of Statistical Inference

Ariel Fernández[1]

Received January 26, 1994; final May 10, 1994

Research to elucidate the cause for the robustness and expediency of biopolymer folding processes calls for a new scheme of statistical inference whereby statistical weights are directly assigned to folding pathways themselves. Such a scheme stands in contrast with traditional methods built upon a Boltzmann measure over conformation space. A rigorous result paving the way for such an approach is presented in this work, where it is proven that an appropriate measure may be defined over the space of folding pathways. Since a space endowed with a measure constitutes an ensemble, this work could be viewed as a starting point to construct a statistical mechanics of folding pathways.

KEY WORDS: Statistical mechanics; Boltzmann measure; polymer folding; Riesz–Markov representation theorem.

1. TOWARD A NEW SCHEME OF STATISTICAL INFERENCE SUITABLE FOR POLYMER FOLDING

The scarcity of theoretical approaches to explain the robustness and expediency with which a biopolymer chain finds its active folding is apparent, as recent research suggests.[1-5] For instance, statistical mechanical methods based upon the construction of a Boltzmann measure over conformation space[6] cannot account for the fact that the active structure is formed expeditiously under severe time constraints. This is especially so since such constraints force the chain to circumvent the Levinthal-like scenario.[7] That is, searching for the most stable folding by means of a random search in conformation space would take as long as the age of the

[1] Instituto de Investigaciones Bioquímicas–INIBIBB–Universidad Nacional del Sur, CONIGET, Bahía Blanca 8000, Argentina, and Frick Laboratory, Princeton University, Princeton, New Jersey 08544.

0022-4715/94/1200-1079$07.00/0

universe and thus the applicability of a thermodynamic approach to predict structures becomes questionable.

Rather, recent evidence[1,2,8–10] prompts us to attempt to introduce a *measure η on the space of folding pathways itself.* A folding pathway will be defined simply as a polymer conformation changing in time. Thus, by endowing the space of folding pathways with an appropriate measure, we would have defined an ensemble and thus paved the way for a statistical mechanics of folding pathways.

Let us now illustrate how statistical inferences could be made using the statistical mechanics of folding pathays. In the context of RNA catalysis, recent experimental evidence[8] and computer simulations[9,10] show that RNA cyclization at an internal position and RNA self-splicing are two competing processes pervasive in ribozyme function governed by two different bundles of competing folding pathways.

Thus, a theoretical approach rooted in the construction of a measure η should encompass the evaluation of integrals of the form

$$\Pr(A) = \int_A d\eta(\vartheta) \tag{1}$$

where a generic notation has been adopted in which ϑ denotes any folding pathway and $\Pr(A)$ indicates the probability of an event A which is realized by an η-measurable bunch (an open set in a suitable topology) A of folding pathways. In the context of ribozyme function, the event A might either be internal cyclization or RNA self-splicing.

In view of these considerations, the purview of this work is bound to be limited and rigorous at the same time: We shall establish that, subjecting the folding process to very general restrictions, a measure η exists over the space of folding pathways. Moreover, such a measure can be defined *constructively* based on the stochastic process whose realizations constitute the folding pathways themselves.

2. DESCRIBING THE SPACE OF FOLDING PATHWAYS

We consider a polymer chain made up of N monomeric units whose conformation is defined by $M(N)$ degrees of freedom. Each of these internal variables corresponds to a dihedral angle representing rotation around a specific bond. Such bonds might be part of the backbone chain, like those forming the sugar–phosphate backbone of RNA, or might be inherent only to the internal conformation of each residue, such as the glycosidic base–sugar bond of an RNA nucleotide.[6] Since vibrational degrees of freedom equilibrate on far shorter time scales than rotational ones, it has been rightly assumed that rotational internal variables suffice to specify a polymer conformation.[6]

Thus, we may consider in principle a conformation space X which, given the angular nature of the degrees of freedom that specify a conformation, constitutes a torus of dimension $M(N)$:

$$X = M(N)\text{-torus} \tag{2}$$

A folding pathway becomes a trajectory on X defined by a map $\vartheta: I \to X$, where I denotes a time interval. In the physically unrealistic case of an infinitely slow pathway made up of successively equilibrated states, the trajectory is determined entirely by thermodynamic or stability control. This means that the trajectory is tangent a point x to the vector field $\Phi(x) = -\text{grad}_x\, U(x)$, where $U(x)$ is the potential energy functional. This potential in turn determines the Boltzmann measure on X, the object upon which classical methods of statistical inference are based.[6]

In a more realistic context, the search in conformation space obeys a stochastic process $\xi: X \times I \to X$. This map defines a family of curves or trajectories $\xi_x: I \to X$, the realizations of the process, each one indexed by a starting or initial conformation x in X. This process must be particularly robust since only a small assortment of destination structures occur reproducibly regardless of the initial state and perturbations of the folding pathways.[4,5,9]

In accord with the introductory discussion, we shall focus on devising a proper scheme that will allow us to assign weights to folding pathways themselves. Thus, we need to introduce a proper space Θ containing all trajectories in X, define its topology $\mathfrak{T}(\Theta)$, and, finally, endow it with a measure η *induced* by the stochastic process $\xi: X \times I \to X$ which generates the trajectories.

Let $\mathfrak{T}(X)$ be the topology on X induced by the metric topology $\mathfrak{T}(\mathfrak{R}^{M(N)})$ of $\mathfrak{R}^{M(N)}$ ($\mathfrak{R}$ are the real numbers), the space in which X is embedded. That is,

$$\mathfrak{T}(X) = \{A \cap X;\ A \in \mathfrak{T}(\mathfrak{R}^{M(N)})\} \tag{3}$$

Let us define now a product topological space of copies or replicas of X which contains in principle all continuous and discontinuous folding pathways with associated time span $|I|$:

$$\Omega = \prod_{t \in I} X_t; \qquad X \equiv X_t \qquad \text{for all } t \tag{4}$$

Thus, $\Omega \supset \Theta$, where $\Theta = C(I \to X)$ is the space of continuous maps of the interval I on X. This space Θ is endowed with the topology $\mathfrak{T}(\Theta)$ *inherited* from the product topology $\prod_{t \in I} \mathfrak{T}(X_t)$ of Ω. Moreover, Θ is *naturally* endowed with a measure μ induced by the product Boltzmann measure $\mathrm{Pr}_{\mathrm{B}} = \prod_{t \in I} \mu_{\mathrm{B},t}$ ($\mu_{\mathrm{B},t}$ is the Boltzmann measure on the replica X_t of conformation space) defined on $\mathfrak{P}(\prod_{t \in I} \mathfrak{T}(X_t))$, the minimal sigma algebra of sets generated by the product topology.

For every $x \in X$, let $\xi_x \in \Theta$ be a specific realization of the stochastic process ξ: $X \times I \to X$. This realization represents a specific folding pathway with associated time span $|I|$ starting with conformation x at $t = 0$. The collection of such realizations constitutes a subset $\xi(X) = \{\xi_x \in \Theta, x \in X\}$ of Θ which is comprised of all the folding pathways that are determined by the generating rules that define the stochastic process ξ.[4,9]

It is not the aim of this work to actually specialize the map to any specific folding process.[4,5,9] It suffices to indicate that in the specific case where folding operates under time constraints and kinetic control governs the folding pathways, a realization ξ_x may be defined and simulated computationally by means of the following general Markov process:

For each time $t \in I$, we define a map $t \to J(x, t) = \{j\colon 1 \leqslant j \leqslant n(x, t)\}$, where $J(x, t)$ = collection of elementary events representing conformational changes which are feasible at time t given that the initial conformation x has been chosen at time $t = 0$, and $n(x, t)$ = number of possible elementary events at time t. Associated to each event there is a unimolecular rate constant $k_j(x, t)$ = rate constant for the jth event in $J(x, t)$,[4] which may take place at time t for a process that starts with conformation x. The mean time for an elementary refolding event is the reciprocal of its unimolecular rate constant. Thus, the only elementary, events allowed are elementary refolding events that satisfy $k_j(x, t)^{-1} \leqslant |I|$.

At this point we may define the Markov process by introducing a random variable $r \in [0, \sum_{j=1}^{n(x,t)} k_j(x, t)]$ whose probability of taking any particular value of the interval is the reciprocal of the length of the interval. Let r^* be a realization of r, that is, a chosen value of r, such that if

$$\begin{gathered} \sum_{j=0}^{j^*-1} k_j(x, t) < r^* \leqslant \sum_{j=0}^{j^*} k_j(x, t) \\ k_0(x, t) = 0 \qquad \text{for any } x,\ t \end{gathered} \tag{5}$$

then the event $j^* = j^*(x, t)$ is chosen at time t for the folding process that starts with conformation x. That is, we have partitioned the interval by adding progressively one rate constant at a time. Obviously, the largest rate constant will produce the largest subinterval in the partition, and thus the choice of r will have the highest probability of falling in that particular subinterval. Thus the map $t \to j^*(x, t)$ for fixed initial condition x constitutes a realization of the Markov process which unambiguously determines the trajectory ξ_x.

3. THE EXISTENCE OF A MEASURE ON THE SPACE OF FOLDING PATHWAYS

At this point we shall formulate and prove the following result:

Theorem. The stochastic process ξ indexed by a starting conformation $x \in X$ induces a measure η on Θ which satisfies the relation

$$\eta A = \int_A \chi_{\xi(X)}(\vartheta)\, d\mu(\vartheta) \tag{6}$$

where μ = measure in Θ induced by the product Boltzmann measure, as defined in Section 2; $\chi_{\xi(X)}(\vartheta) = 1$ if there exists $x \in X$ such that $\vartheta = \xi_x$, and $\chi_{\xi(X)}(\vartheta) = 0$ otherwise.

In precise terms, *the μ-measurable function $\chi_{\xi(X)}$ is the Radon–Nikodym derivative of η with respect to μ.*

Proof. The space X is compact when endowed with topology $\mathfrak{T}(X)$; thus, by Tikhonov's theorem, Ω is compact with the product topology, and Θ is also compact when endowed with the topology inherited from the product topology. Since Θ is also Hausdorff, we shall apply the Riesz–Markov representation theorem.[11] Consider the space of continuous functionals $C(\Theta)$; then, given a functional F in the dual space $C(\Theta)^*$, there exists a measure η on Θ such that

$$F(h) = \int_\Theta h(\vartheta)\, d\eta(\vartheta) \qquad \text{for } any\ h \text{ in } C(\Theta) \tag{7}$$

Since there are no restrictions on F, we take

$$F(h) = \int_X h(\xi_x)\, d\mu_{\mathrm{B}}(x) \tag{8}$$

Thus, we have shown that η is induced by the stochastic process ξ.

The measure η may be constructed as follows: Let $A \in \mathfrak{T}(\Theta)$; then we define its measure as

$$\eta A = \operatorname{Sup}\{F(h), 0 \leqslant h \leqslant 1, h \in C(\Theta), A \supset \operatorname{support}(h)\} \tag{9}$$

This real functional defined on open sets may be canonically extended to a *regular* measure over $\mathfrak{P}(\prod_{t \in I} \mathfrak{T}(X_t) \cap \Theta)$.[11]

Consider now the set $D(A)$ of functionals $f(\vartheta)$ of the form

$$f(\vartheta) = \left\{\int_I \chi_{\pi_t(A)}(\pi_t\vartheta)\, f(t) \exp[-\beta U(\pi_t\vartheta)]\, dt\right\} \times \left\{|I| \int_X \exp[-\beta U(x)]\, \delta x\right\}^{-1} \tag{10}$$

where $\pi_t: \Omega \to X_t$ is the canonical projection; $\beta = 1/k_B T$ (T is the temperature, k_B the Boltzmann constant); $0 \leqslant f(t) \leqslant 1$ is *any* continuous real function; $\chi_{\pi_t(A)}$ is the characteristic function of the projection of A on the replica X_t; and δx is the differential volume in conformation space X.

The set $D(A)$ is *dense* in $G(A) = \{0 \leqslant h \leqslant 1, h \in C(\Theta), A \supset \text{support}(h)\}$. Therefore we have

$$\eta A = \text{Sup}\{F(h), h \in D(A)\} \tag{11}$$

This equation enables us to compute the measure of A, thus verifying Eq. (6):

$$\begin{aligned} \eta A &= \left\{\int_X \int_I \chi_{\pi_t(A)}(\pi_t \xi_x) \exp[-\beta U(\pi_t \xi_x)]\, dt\, \delta x\right\} \\ &\quad \times \left\{|I| \int_X \exp[-\beta U(x)]\, \delta x\right\}^{-1} \\ &= \int_A \chi_{\xi(X)}(\vartheta)\, d\mu(\vartheta) \end{aligned} \tag{12}$$

To avoid confusion, the reader should be reminded that μ is induced by the product Boltzmann measure $\text{Pr}_B = \prod_{t \in I} \mu_{B,t}$ ($\mu_{B,t}$ is the Boltzmann measure on the replica X_t). The result given in Eq. (12) results by replacing the set $G(A)$ by the set $D(A)$ in the definition given by Eq. (9), a valid procedure since any functional in $G(A)$ is the limit of a sequence of functionals in $D(A)$.

This completes the proof of the theorem. QED

ACKNOWLEDGMENT

This work was financially supported by the Camille and Henry Dreyfus Foundation of New York City and by Fundación Antorchas (Argentina).

REFERENCES

1. R. Jaenicke, *Angew. Chem. Int. Ed. Engl.* **23**:295 (1984).
2. T. E. Creighton, *Bioessays* **8**:57 (1988).
3. T. E. Creighton, *Proc. Natl. Acad. Sci. USA* **85**:5082 (1988).
4. A. Fernández, *Phys. Rev. Lett.* **64**:2328 (1990); *Phys. Rev. A Rapid Commun.* **45**:R8348 (1992).

5. E. I. Shakhnovich, G. Farztdinov, A. M. Gutin, and M. Karplus, *Phys. Rev. Lett.* **67**:1665 (1991).
6. C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, New York, 1980).
7. C. Levinthal, *J. Chim. Phys.* **65**:44 (1968).
8. S. Partono and A. Lewin, *Mol. Cell. Biol.* **8**:2562 (1988).
9. A. Fernández, *J. Theor. Biol.* **157**:487 (1992).
10. A. Fernández, A. Lewin, and H. Rabitz, *J. Theor. Biol.* **164**:121 (1993).
11. E. Nelson, *Ann. Math.* **69**:630 (1959).