Journal of Mathematical Chemistry 19 (1996) 331–336

A geometric framework for polymer folding

Hernán Cendra [a], Ariel Fernández [a] and Walter Reartes [b]

[a] *Departamento de Matemática of the Universidad Nacional del Sur*
and Instituto de Matemática (INMABB-CONICET), Av. Alem 1253,
8000 Bahía Blanca, Argentina
[b] *Departamento de Física, Universidad Nacional del Sur, Av. Alem 1253,*
8000 Bahía Blanca, Argentina

Received 23 June 1995; revised 8 December 1995

The problem of finding the folding pathway of a polymer, a fundamental issue in the field of molecular biophysics, is viewed from a geometrical standpoint. A riemannian metric on conformation space related to hydrodynamical features of the system and a drift related to the potential are used to derive a diffusion equation governing the time-dependent probability in conformation space. The thermodynamic equilibrium limit is found to be consistent with Boltzmann's measure.

1. Introduction

The folding process of polymer macromolecules is an extremely complex phenomenon which has been approached using several theoretical tools. In [1] the problem of finding the folding pathway followed at the level of secondary structure has been handled using techniques from the field of discrete stochastic processes. A measure in the space of all possible folding pathways should be defined in such a way that, for a given secondary structure, it is concentrated around the most probable folding pathway. Experimental evidence of the feasibility of this kind of treatment has been furnished in [2].

From a more theoretical point of view, this type of construction is related to a certain geometrization of stochastic diffusions as described for instance by Nelson in [3]. There, a riemannian metric σ representing the scale of local fluctuations and a vector field b representing the drift, are defined on a given differentiable manifold, the configuration space of the system. This defines a stochastic diffusion. Then a related measure on path space is defined which is, in a sense, equivalent to the stochastic diffusion.

On the other hand, Nelson's work [3] is mainly concerned with stochastic quantization of systems that can be described by a given classical lagrangian with stochastic least action principles, where σ represents the interaction with the background field. Although dissipative systems are excluded from this least action principle approach we will show that Nelson's basic geometric background is still appropriate for describing difussions mimicking the search in conformation space of flexible polymer chains. Such systems are of dissipative type. In a sense, this amounts to giving a continuous version of the discrete stochastic process expounded in [1], which accurately describes the folding process resolved up to secondary structure.

Our aim in the present paper consists in establishing this fact via fundamental principles of physics, namely using some basic tools from polymer dynamics [4]. We show how to write the diffusion given by Smoluchowski equation, in Nelson's geometric terms, which paves the way for finding the corresponding measure on the space of all possible folding pathways. Thus we provide theoretical evidence that the methods of polymer dynamics as described in [4] are also good to approach the problem of polymer folding, and ultimately lead to defining a measure in the space of all possible folding pathways having the property stated above. In particular, the metric σ defined on configuration space is essentially related to the viscosity of the solvent, and thus the hydrodynamic treatment expounded in [4] applies.

2. Geometrization of conformation space

Here we study a polymer which, from the point of view of the dynamics of the folding process, can be considered as a system of N beads subject to rigid constraints, living in a solvent bath at temperature T and having friction constant ζ. The only kind of constraints to be considered in the present paper corresponds exactly to the *freely rotating chain* model of [4]. Let R_i, $i = 1, 2 \ldots N$, be the three dimensional position vectors of the beads. More precisely, this means that the value of the angles formed by two consecutive vectors $R_n - R_{n-1}$ is a fixed number independent of n, and that the euclidean norms $|R_n - R_{n-1}|$ remains fixed, for each n, along the evolution of the system. A more explicit description can be given as follows. Let x_i, $i = 1, 2, \ldots N - 3$, be given real numbers. Consider the following recursive formulas:

$$R_i = R_{i-1} + A(x_{i-3}(R_{i-1} - R_{i-2}))(R_{i-2} - R_{i-3}) , \qquad i = 4, 5 \ldots N ,$$

where for each $u, v \in \mathbf{R}^3$, the vector $A(u)v$ is obtained rotating the vector v an angle $|u|$ about the vector u. For each $x = (x_1, x_2, \ldots x_{N-3})$ and given 3-dimensional vectors R_1, R_2 and R_3, let $R(x) \in \mathbf{R}^{3N}$ be the element of $\mathbf{R}^{3N}$ defined by $R(x) = (R_1, R_2, \ldots R_N)$. Let us assume for a moment that R_1, R_2 and R_3 are fixed given points of $\mathbf{R}^3$ and R_i, $i = 4, 5, \ldots N$, are given by the previous recursive formulas. The map R is X_i-periodic in each coordinate x_i, $i = 1, 2 \ldots N - 3$, where

$$X_i = \frac{2\pi}{|R_{i+2} - R_{i+1}|} , \qquad i = 1, 2, \ldots N - 3 .$$

Therefore, R induces an embedding $R \colon T \to \mathbf{R}^{3N}$, where T represents a $(N - 3)$-dimensional torus. On the other hand, for fixed R_1, the system formed by the three points R_1, R_2 and R_3 is a rigid body, thus its position can be parametrized in a one-to-one and onto manner by a point in $SO(3)$. Finally, since R_1 may range freely over $\mathbf{R}^3$, we obtain a natural extension of the previous embedding

$$R: \ \mathbf{R}^3 \times SO(3) \times T \to \mathbf{R}^{3N}.$$

Let $Q = \mathbf{R}^3 \times SO(3) \times T$ for simplicity. Thus Q is a $(N+3)$-dimensional manifold, which is the configuration space of the system and we have an embedding R of Q into $\mathbf{R}^{3N}$. Any motion of the polymer is a curve or *path* in Q, in particular, so is the folding pathway (see [1] for a more detailed description of the process of polymer folding). Forces acting on each bead are of several types: forces of mutual atraction and repulsion of short rank among beads, random forces due to thermal fluctuations, forces of viscosity type and the forces of the constraints. However, the tremendous complexity of this system, especially when the number N is high, makes it almost useless to try to find exact solutions. The statistical point of view turns out to be clearly more appropriate, and we shall explain how to deal with it. According to [4], 3.8, let g be the riemannian metric on Q inherited from the euclidean metric $\cdot$ (dot) in $\mathbf{R}^{3N}$, through the embedding R. Thus, in local coordinates q^i, $i = 1, 2, \ldots n$, where n is the dimension of Q, we have the following:

$$g_{ij} = \frac{\partial R}{\partial q^i} \cdot \frac{\partial R}{\partial q^j}.$$

There is a well known natural volume element associated to the riemannian metric g, which in local coordinates can be expressed as follows:

$$\sqrt{g}\, dq^1 dq^2 \ldots dq^{N+3}.$$

Let $\phi/\sqrt{g}$ be the probability density with respect to the previous volume, that is, for any given measurable subset D of Q, the probability of finding the polymer in the set D is proportional to the integral of $\phi/\sqrt{g}$ over D with respect to the volume. Then, according to [4], (3.162), (with a slight change of notation), the equation governing the time evolution of ϕ is the following:

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial q^i}\left[h^{ij}\left(k_B T \frac{\partial \phi}{\partial q^j} + \frac{\partial U}{\partial q^j}\phi\right) - V^i \phi\right]. \tag{1}$$

Here h^{ij} is (the inverse matrix of) a metric on Q that depends on the friction constant ζ and the constraints, k_B is the Boltzmann constant, $U + k_B T \log\sqrt{g}$ is a potential on Q that describes the attraction and repulsion forces among the beads (this can be calculated from experimental data in particular cases, [1]) and V is a vector field on Q that takes into account the macroscopic flow, if any, of the solvent. Next, we must switch to a more appropriate notation. Set

$$k_B T h^{ij} = \sigma^{ij},$$

$$W = \frac{U}{k_B T},$$

$$Z_i = -\sigma_{ij} V^j,$$

thus Z is a 1-form on Q. Then eq. (1) becomes

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial q^j}\left[\sigma^{ij}\frac{\partial \phi}{\partial q^i} + \sigma^{ij}\left(\frac{\partial W}{\partial q^i} + Z_i\right)\phi\right]. \tag{2}$$

3. Folding dynamics as viewed in the new geometric context

We shall write Smoluchowski equation in a more intrinsic (not depending on the choice of local coordinates) way. The expression of the codifferential δ of a given 1-form α in local coordinates is the following [5]:

$$\delta\alpha = -\frac{1}{\sqrt{\sigma}}\left(\frac{\partial}{\partial q^i}\left(\sqrt{\sigma}\sigma^{ij}a_j\right)\right), \tag{3}$$

where σ_{ij} is a given metric on Q and $\sqrt{\sigma}$ is the square root of the determinant of the matrix σ_{ij}. Recall that the *Laplace–De Rham* operator is by definition, $\delta d + d\delta$. Following [3] we shall call Δ the *Dohrn–Guerra Laplacian* which, for functions and 1-forms coincide with *minus* the Laplace–De Rham operator. For a given function ρ on Q, we know that $\delta\rho = 0$, then, in local coordinates, we have the following:

$$\Delta\rho = \frac{1}{\sqrt{\sigma}}\frac{\partial}{\partial q^i}\left(\sqrt{\sigma}\sigma^{ij}\frac{\partial \rho}{\partial q^j}\right). \tag{4}$$

Using the previous formulas, it follows by a straightforward calculation that Smoluchowski equation (2) can be written intrinsically as follows:

$$\frac{\partial \rho}{\partial t} = \Delta\rho + \delta(\rho b),$$

where b is a 1-form on Q, given in local coordinates by the following:

$$b_i = -\frac{\partial}{\partial q^i}\left(W + \log\sqrt{\sigma}\right) - Z_i$$

and

$$\rho = \frac{\phi}{\sqrt{\sigma}}.$$

Finally, set

$$\sigma_{ij} = 2\tau_{ij}$$

and

$$b_i = 2c_i .$$

Then the previous formula becomes

$$\frac{\partial \rho}{\partial t} = \frac{1}{2}\Delta\rho + \delta(\rho c) . \tag{5}$$

We should remark that in (5), δ and Δ are the codifferential and the Laplacian with respect to the new metric τ_{ij}.

Equation (5) is precisely the *Fokker–Planck* equation corresponding to a smooth markovian diffusion, and leads to the definition of a measure on path space that concentrates precisely on the most probable folding pathway, as described in [3]. This should lead, at least for naturally selected molecules, to the same folding pathway found computationally in [1,2].

These facts constitute our main result. As a theoretical application that takes advantage of the geometric framework, let us show how we can easily derive Boltzmann's distribution law in the thermodynamic limit.

Let us assume an equilibrium state of the system. In particular the macroscopic flow is zero, i.e. $Z = 0$. Besides, since $\rho = \rho(q)$ is an equilibrium solution of (5), the following equality holds:

$$0 = \delta d\rho + \delta(\rho dP) , \tag{6}$$

where $P = W + \log\sqrt{\sigma}$. Now we introduce a function r on Q defined as follows:

$$\rho = r\exp(-P) . \tag{7}$$

A straightforward calculation shows that (6) is equivalent to the following:

$$\delta(dr\exp(-P)) = 0 ,$$

which immediately implies that

$$\int \delta(dr\exp(-P))r = 0 ,$$

where the integral is taken with respect to the volume element canonically defined by the metric σ on the manifold Q. Since the differential operator δ is the adjoint of the differential operator d, with respect to the metric $\langle , \rangle$ induced by σ on forms, we have the following:

$$\int \langle \exp(-P)dr, dr\rangle = 0 .$$

Since $\exp(-P) > 0$ and $\langle , \rangle$ is positive definite, we can conclude that $dr = 0$, that is, r is a constant.

4. Concluding remarks

By placing the folding process of flexible polymer chains in the proper hydrodynamical context, we obtain an appropriate riemannian geometry on conformation space. This leads naturally to a Smoluchowski-type equation for the time dependent distribution of conformation probabilities. We verify the consistency and coherence of our treatement by studying the thermodynamic limit and finding the equilibrium distribution, which corresponds to a Boltzmann-type distribution when the volume canonically associated to the metric is adopted.

This research should be placed within a paradigm in which the statistical weights are placed not on conformations but on the folding pathways themselves [6]. Thus within a biologically-relevant in vitro context, the functional folding becomes the destination structure of the folding pathway which carries the highest statistical weight [6]. This destination structure must emerge within biologically-relevant timescales, typically incommensurate with the thermodynamic limit of infinite times [1,2,6]. Such an approach takes into account the time constraints associated to the problem and thus differs essentially from current paradigms in which the active conformation is identified with the global free energy minimum [7].

Considering folding pathways has led to a time-dependent distribution of folding products which has been determined experimentally in specific folding contexts [8,9]. Furthermore, our kinetically-controlled simulations of RNA folding pathways in vitro have been recently probed experimentally [10,11]. However, there is a need to provide the proper theoretical underpinnings to these findings, furnishing the underlying generic equation to be satisfied by the time-dependent distribution of folding products. This general issue has been addressed in this work.

References

[1] A. Fernández, Phys. Rev. Lett. 64 (1990) 2328.

[2] A. Fernández, J. Phys. A: Math. Gen. 27 (1994) 6039.

[3] E. Nelson, *Quantum Fluctuations* (Princeton University Press, Princeton, NJ, 1985).

[4] M. Doi and S.F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1986).

[5] R. Abraham, J.E. Marsden and T. Ratiu, *Manifolds, Tensor Analysis and Applications* (Addison-Wesley, 1983).

[6] A. Fernández, Ann. Phys. 4 (1995) 600.

[7] J. Novotny, Modeling protein folding and structure, *Protein Folding*, eds. L.M. Gierasch and J. King (Am. Assoc. Adv. Sci., Washington, DC, 1990) pp. 249–251.

[8] T.E. Creighton, Understanding protein folding pathways and mechanisms, *Protein Folding*, eds. L.M. Gierasch and J. King (Am. Assoc. Adv. Sci., Washington, DC, 1990) pp. 157–170.

[9] K. Chou, G. Némethy, M. Pottle and H.A. Scheraga, Biochem. 24 (1985) 7948.

[10] P.P. Zarrinkar and J.R. Williamson, Science 265 (1994) 918.

[11] A. Fernández, G. Appignanesi and H. Cendra, Chem. Phys. Lett. 242 (1995) 460.