A note on the analysis of heterogeneous rod like chain in DNA modeling

Yan Ding, Tiejun Li

LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China.

Abstract

Two different results concerning the elastic behavior of the heterogeneous wormlike chain (WLC) [D. Bensimon et al., Europhys. Lett., 42 (1998), pp. 97] and rod like chain (RLC) [P. Nelson, Phys. Rev. Lett., 80 (1998), pp. 5810] are compared in this paper. We argue that the RLC is a more suitable model for double-stranded (ds-) DNA. As the hetero-RLC is the basic model for studying sequence-dependent ds-DNA, a rigorous path integral analysis for the effective bending persistence length is performed in the weak disorder limit. The novelty of the paper is to analyze a path integral on the Lie group $SO(3)$ with random forcing, which supplies a rigorous basis for the analysis of RLC type models.

Key words: DNA modeling, heterogeneous rod like chain, path integral, effective bending persistence length

PACS: 05.40.+j, 31.15.Kb, 87.15.By

1 Introduction

The technology for the manipulation of single molecules [6,7,16,18–20,23] opens a new way to explore the biological function of bio-macromolecules through mechanical or physical measurements. It is usually designed that the molecules are stretched by optical or magnetic tweezers, flexible microscopic cantilevers, shear or extensional flows, and so on. Especially for DNA, it is found that it shows quite nontrivial elastic properties [6,20,23]. At low external forces, DNA displays nonlinear elastic behavior because of entropy elasticity; at moderate forces, it behaves as a Hookean rod with large stretch modulus. But if the

\(^1\) E-mail: dyan, tieli@math.pku.edu.cn

Preprint submitted to Elsevier Preprint 27 December 2006
pulling force is beyond 70 pN, the molecules can suddenly be stretched into
1.7 times its native length. Besides this, DNA displays supercoiling behavior [23] for its double stranding. Various continuum models are proposed to understand this behavior of DNA in recent years [3,4,12–14,21,25] and some of them have become consensus. Among these models, wormlike chain (WLC) [12] and rod like chain (RLC) [3] play the central role.

All of the models mentioned above must be understood for homogeneous polymers in the sense that the bending rigidity coefficients and preferred bending angles are almost constant along the chain. But it is well-known that DNA is sequence dependent [17]. A famous example is that A-tract DNA will bend much more than the disordered sequences. This sequence dependence will introduce heterogeneity into the bending rigidity and bending angles phenomenologically if we consider the elastic behavior of DNA. But the realistic sequence dependent modeling is still in its infancy. Theoretically, investigating the effective behavior of a heterogeneous polymer chain is highly like a homogenization problem in mathematics [2].

There has been some models on the heteropolymers for the protein folding [10,24], RNA secondary structure analysis [5], and DNA bending and stretching [1,15]. The models presented for protein folding must be understood only for flexible polymers because there is no bending energy involved. Here we are interested in the effective behavior of heterogeneous DNA. It is worth noting that the results obtained in [1] and [15] are totally different. In [1], Bensimon et al. proposed a heterogeneous Kratky-Porod chain (the discrete counterpart of WLC) with random preferred kinks. After assuming the preferred angle satisfying a half-Gaussian distribution and taking the thermodynamic limit, they found that the effective bending persistence length \( A_{\text{eff}} \) has the following relation in the weak disorder limit

\[
\frac{A_{\text{eff}}}{A} = 1 - \sqrt{\frac{r}{2}}, \quad r \ll 1, \tag{1}
\]

where \( A \) is the bending persistence length for the homopolymer, and \( r \) is the strength of the heterogeneity. Nelson considered the heterogeneous RLC model [15], in which each point in the chain is endowed with a body-fixed frame \( \hat{E}(s) \). The bending energy is taken as the integral of the square of angular velocities along the chain with Gaussian white noise perturbations. Nelson found that the effective bending persistence length \( A_{\text{eff}} \) has the form

\[
\frac{A_{\text{eff}}}{A} \approx 1 - r, \quad r \ll 1, \tag{2}
\]

which is different from the result obtained by Bensimon et al..

The aim of this paper is to clarify the difference between these two results, and argues that hetero-RLC is more suitable for DNA modeling. The key reason
is that DNA is an anisotropic polymer if we take into account its microscopic structure. This property will induce rotationally dependent energy, which is different from the wormlike chain. Based on this understanding, we further argue though the approach taken in [15] is very elegant and intuitive, the procedure is questionable. From the viewpoint of statistical mechanics, only the Boltzmann factor of the path integral is considered. The partition function is neglected. For the analysis of effective behavior, the stochastic terms are hidden in the partition function also. It is not so obvious that it has no contributions. We take a rigorous path integral analysis for the heterogeneous RLC in this paper. This corresponds to a random motion on the Lie group $SO(3)$. Though we find the final result is the same as that obtained by Nelson but our approach is rigorous and general for RLC type models.

The paper is organized as follows. In section 2, we state the model setup in Bensimon’s and Nelson’s papers, respectively, and the origin of the difference between them. In section 3, we give a direct calculation of the bending persistence length for homogeneous RLC, which must be known intuitively, but rarely found in literatures. Then as a simple case, we consider the heterogeneous RLC confined in a plane in section 4. We present the calculations in three dimensions in section 5. Finally we make conclusions.

2 Model setup and the origin of the difference

In [1], Bensimon et al. proposed a heterogeneous Kratky-Porod chain model with random preferred kinks (see Figure 1). Suppose the chain consists of $N$ segments of size $b$ (maximal extension $L = Nb$), the elastic energy is defined as

$$E_{\text{elastic}} = -\frac{\kappa}{b} \sum_{i=1}^{N-1} \cos(\alpha_i - \beta_i),$$

where $\kappa$ is the bending rigidity coefficients, $\alpha_i$ is the angle between the successive segments $i$ and $i + 1$ and $\beta_i$ is a preferred random orientation between these two segments. In thermal units, $\kappa = k_B T A$, where $A$ is the bending persistence length, $k_B$ is Boltzmann constant, and $T$ is the absolute temperature. Bensimon et al. found that under the condition the half Gaussian distribution of disorder $P(\beta) = \sqrt{2A_d/\pi b} \exp(-A_d\beta^2/2b)$ ($\beta > 0$) the effective bending persistence length $A_{\text{eff}}$ has the relation shown in equation (1).

The heterogeneous RLC model is set up as follows (see Figure 2). In each point $s$ on the chain it is endowed with a body-fixed frame $\hat{E}(s)$ by an orthonormal matrix $E(s)$, where the $i$-th column $E_i(s)$ is the orientation of the $i$-th axis, and particularly $\hat{E}_3(s)$ is chosen as the tangent $t(s)$ to the rod axis (see Figure 2).
Define the matrix for angular velocity

\[ \Omega = E^{-1} \cdot \dot{E} = - \sum_{i=1}^{3} \Omega_i \mathbf{T}_i, \]  

where the dot denotes \( \frac{d}{ds} \) and \( (\mathbf{T}_i)_{jk} = \varepsilon_{ijk} \). Here \( \varepsilon_{ijk} \) is the Levi-Civita’s permutation symbol. The elastic energy of heterogeneous RLC is defined as

\[ E_{\text{elastic}} = \frac{1}{2} \int_0^L ds \left[ \kappa (\Omega_1 - \zeta_1)^2 + \kappa (\Omega_2 - \zeta_2)^2 + \tilde{\kappa} (\Omega_3 - \zeta_3)^2 \right], \]

where \( \kappa = k_B T A \). \( A \) is the bending persistence length. \( \tilde{\kappa} = k_B T C \) is the twist rigidity coefficient. \( \zeta_i(s) \) are independent Gaussian white noise with

\[ \mathbb{E}\zeta_i(s) = 0, \quad \mathbb{E}\left( \zeta_i(s)\zeta_j(s') \right) = \frac{r}{A} \delta(s-s') \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \]

With the assumptions above, Nelson found that the effective bending persistence length has the form (2).

These two different results seem subtle at the first glance. But a careful checking shows the symmetry is the central issue. In Bensimon’s heterogeneous
Kratky-Porod chain model, it is supposed that the bending of the chain has the axial symmetry. This implies that not only there is no torsion energy, but also the rod is uniform in any direction (see the left figure for the uniform rod in Fig. 3). It makes the heterogeneity $\beta$ satisfies the half Gaussian distribution. But in Nelson’s heterogeneous RLC model, the bending of the chain in different directions will cost different energy. This can be explained by taking into account the anisotropy of the DNA’s microscopic structure. It is well-known that DNA is an unsymmetric double helix with the major grooves and minor grooves [17]. The molecules in the backbone will naturally prefer to be bent to the major groove side because there are less molecules in this side (see the right figure in Fig. 3). This intuition is reasonable, but not always correct. As a counterexample, though the AdMLP TATA-box bends flexibly towards the major groove, the A-tract variant A29 bends towards the minor groove [22]. In any case, the axial symmetry does not exist. This symmetry-breaking makes the heterogeneity in Nelson’s model takes the form of Gaussian white noise, which is more reasonable for DNA modeling.

![Uniform Rod vs Anisotropic Rod](image)

Fig. 3. Comparison between uniform rods (the WLC), and the anisotropic rods (the RLC). For the case of uniform rods, it is reasonable to assume the axial symmetry. For the case of RLC, the chain will be more easily bent to the major groove side intuitively since there are less molecules in this side.

We will follow Nelson’s heterogeneous RLC model in this paper. In Nelson’s analysis, only the heterogeneity in the elastic energy $E_{\text{elastic}}$ is considered. But it must also appear in the partition function $Z$, which is neglected by Nelson. We will give a rigorous path integral analysis for this heterogeneous model in the continued sections. This analysis will be useful for further work in RLC.

### 3 Path integral analysis of homogeneous RLC

To analyze the effective bending persistence length of heterogeneous RLC, we will give an analysis for the homo-polymer at first to show the essential ingredients and the basic result.
Note that the angular velocity can be represented as $\Omega_i = \frac{1}{2} Tr(\Omega \cdot T_i)$. We then define the Lagrangian of the RLC

$$\mathcal{L}(\dot{E}, E) = \frac{1}{2} \sum_{i=1}^{3} \kappa_i \Omega_i^2,$$  \hspace{1cm} (7)$$

where $\kappa_1 = \kappa_2 = \kappa$, $\kappa_3 = \tilde{\kappa}$. Define the propagator or Green’s function $[9,11,26]

$$G(E, s | E', s') = \int_{E(s)=E'}^{E(s)=E} \mathcal{D}E(t) \exp \left( - \frac{1}{k_B T} \int_s^{s'} L(\dot{E}, E) dt \right) \delta(E(s) - E),$$

where $s \geq s'$. The path integral analysis of this system is on the Lie group space $SO(3)$. To perform the analysis, we introduce the local coordinates representation for ortho-normal frames — the Euler angles $\phi(s), \theta(s)$ and $\psi(s)$. The matrix $E(s)$ can be decomposed to

$$E(s) = e^{-\phi(s)T_3} e^{-\theta(s)T_1} e^{-\psi(s)T_3}. \hspace{1cm} (8)$$

The components of the angular velocity have the form

$$\Omega_1 = \sin \psi \sin \theta \dot{\phi} + \cos \psi \dot{\theta},$$
$$\Omega_2 = \cos \psi \sin \theta \dot{\phi} - \sin \psi \dot{\theta},$$
$$\Omega_3 = \dot{\psi} + \cos \theta \dot{\phi}.$$ 

Define $\theta(s) = (\phi(s), \theta(s), \psi(s))$, then the Lagrangian becomes a quadratic form as

$$\frac{1}{k_B T} \mathcal{L}(\dot{\theta}, \theta) = \frac{A}{2} (\dot{\phi}^2 \sin^2 \theta + \dot{\theta}^2) + \frac{C}{2} (\dot{\psi}^2 + \dot{\phi}^2 \cos^2 \theta + 2 \dot{\phi} \dot{\psi} \cos \theta)$$

$$= \frac{1}{2} g_{ij} \dot{\theta}^i \dot{\theta}^j, \hspace{1cm} (9)$$

where the matrix

$$g_{ij} = \begin{pmatrix} A \sin^2 \theta + C \cos^2 \theta & 0 & C \cos \theta \\ 0 & A & 0 \\ C \cos \theta & 0 & C \end{pmatrix}. \hspace{1cm} (10)$$

$$(\theta^1, \theta^2, \theta^3) = (\phi, \theta, \psi)$, and the summation convention is implied. Correspondingly the measure in the integrand will be replaced by the Haar measure

$$dE = \frac{1}{8\pi^2} \sin \theta d\phi d\theta d\psi := Jd\theta, \hspace{1cm} (11)$$
where \( J = \frac{1}{8\pi} \sin \theta, d\theta = d\phi d\theta d\psi, \) and \( \theta \in [0, \pi), \phi, \psi \in [0, 2\pi). \)

The quadratic form (9) reminds us the path integral in the curved space with metric \((g_{ij}).\) Its determinant \( g = A^2 C \sin^2 \theta, \) which agrees with the Haar measure up to a constant. This suggests that we may view the path integral as a Brownian motion on the new manifold. From the stochastic differential geometry or path integral theory in [8,11], we obtain the equation for \( G(\theta, s|\theta', 0) \)

\[
\partial_s G = \frac{1}{2} \Delta_\theta G, \tag{12}
\]

where \( \Delta_\theta \) is the Laplace-Beltrami operator with respect to \( \theta \)

\[
\Delta_\theta = \frac{1}{\sqrt{g}} \partial_i (\sqrt{g} g^{ij} \partial_j).
\]

Define the transition probability for RLC

\[
p(\theta, s|\theta', 0) = \frac{1}{A} G(\theta, s|\theta', 0), \tag{13}
\]

where

\[
A = \int G(\theta, s|\theta', 0) J d\theta
\]

is the normalization constant. We have the equation for \( p \)

\[
\partial_s p = \frac{1}{A} \partial_s G - \frac{1}{A^2} G \cdot \partial_s A = \frac{1}{2} \Delta_\theta p \tag{14}
\]

with initial condition \( p(\theta', s|\theta, 0) = \delta(\theta - \theta'), \) where the delta function \( \delta(\theta - \theta') \) is understood in the sense that

\[
\int f(\theta) \delta(\theta - \theta') J d\theta = f(\theta')
\]

for any continuous function \( f. \)

The bending persistence length of a polymer is characterized by the correlation function of the unit tangent vectors along the chain. We have

\[
E_3(0) \cdot E_3(s) = \sin \theta(0) \sin \theta(s) \cos(\phi(0) - \phi(s)) + \cos \theta(0) \cos \theta(s).
\]

Define the correlation function \( C(s) = \langle E_3(0) \cdot E_3(s) \rangle, \) where \( \langle \cdot \rangle \) means the thermal average with respect to all possible paths. Then

\[
C(s) = \int \int E_3' \cdot E_3 f_{eq}(E') p(E, s|E', 0) dE dE' = \int \int E_3' \cdot E_3 f_{eq}(\theta') p(\theta, s|\theta', 0) J d\theta d\theta'. \tag{15}
\]
where \( f_{eq}(\theta') = 1 \) is the equilibrium density. We have
\[
\frac{d}{ds} C(s) = \int \int E_3' \cdot E_3 f_{eq}(\theta') \partial_s p(\theta, s|\theta', 0) J J' d\theta d\theta' = -\frac{1}{A} C(s),
\]
where the last equality is obtained by integration by parts twice. From the initial condition \( C(0) = 1 \), we obtain the correlation function
\[
C(s) = \exp \left(-\frac{s}{A}\right).
\]
And it is obvious that \( A \) is exactly the bending persistence length of homogeneous RLC.

4 Path integral analysis of heterogeneous RLC: 2D case

To present the essence of the path integral analysis for heterogeneous RLC, we first consider a simplified two dimensional case in the absence of some technical calculations.

In 2D case, the frames can be chosen as
\[
E = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix},
\]
where \( \theta(s) \) is the angle between the tangent \( t(s) = E_1 \) and the fixed horizontal axis and \( \theta(s) \in [0, 2\pi) \). Then the Lagrangian is simplified as
\[
\frac{1}{k_B T} \mathcal{L}(\dot{\theta}, \theta) = \frac{A}{2} (\dot{\theta} - \zeta)^2,
\]
where the heterogeneity is introduced as the Gaussian white noise
\[
\mathbb{E} \zeta(s) = 0, \quad \mathbb{E} (\zeta(s)\zeta(t)) = \frac{r}{A} \delta(s-t).
\]
And the measure in the integrand is \( J d\theta = \frac{1}{2\pi} d\theta \). From the path integral theory [11,8], the Hamiltonian of the system is
\[
\mathcal{H} = \frac{1}{2} (\dot{\theta} + A\zeta) \frac{1}{A} (\dot{\theta} + A\zeta).
\]
And the Green’s function satisfies
\[ \partial_s G = \hat{\mathcal{H}} G, \] (21)
where \(\hat{\mathcal{H}}\) is the quantization of Hamiltonian \(\mathcal{H}\) by replacing \(\dot{\theta}\) with \(-\partial_{\theta}\)
\[
\hat{\mathcal{H}} = \frac{1}{2}(\partial_{\theta} - A\zeta)\frac{1}{A}(\partial_{\theta} - A\zeta).
\] (22)
Correspondingly, the transition probability density defined in (13) satisfies a stochastic partial differential equation
\[
\partial_s p = \left( \frac{1}{2A} \partial_{\theta\theta} - \zeta \partial_{\theta} \right) p, \quad p|_{s=0} = \delta(\theta - \theta').
\] (23)
In 2D case we have \(t(0) \cdot t(s) = \cos(\theta(0) - \theta(s))\), and the correlation function
\[
\mathcal{C}(s) = \int \int \cos(\theta' - \theta) f_{eq}(\theta') p(\theta, s|\theta', 0) J J' d\theta d\theta'.
\]
From (23) we can obtain the Taylor expansion of \(p(\theta, s|\theta', 0)\) around \(s = 0\) to second order in integral form by recursively using the equation
\[
p(\theta, s|\theta', 0) = \delta(\theta - \theta') + \int_0^s \left( \frac{1}{2A} \partial_{\theta\theta} - \zeta(s') \partial_{\theta} \right) p(\theta, s'|\theta', 0) ds'.
\] (24)
Taking expectation with respect to the white noise and using (19), we obtain
\[
E\mathcal{C}(s) = 1 - \frac{1 + r}{2A}s + o(s).
\] (25)
Thus the effective bending persistence length \(A_{eff} \approx \frac{2A}{1 + r}\) in 2D and we have
\[
\frac{A_{eff}}{2A} \approx \frac{1}{1 + r} \approx 1 - r, \quad r \ll 1,
\] (26)
where the denominator is \(2A\) because the bending persistence length is \(2A\) in two dimensional case.

5 Path integral analysis of heterogeneous RLC: 3D case

In 3D heterogeneous case, the Lagrangian has the form
\[
\frac{1}{k_B T} \mathcal{L}(\dot{\theta}, \theta) = \frac{1}{2} g_{ij} \dot{\theta}^i \dot{\theta}^j + a_i \dot{\theta}^i
\]
after dropping some constant terms, where the metric \((g_{ij})\) is the same as (10) and \((a_i)\) has components

\[
a_i = \begin{pmatrix}
-A \sin \theta (\zeta_1 \sin \psi + \zeta_2 \cos \psi) - C \cos \theta \zeta_3 \\
-A \zeta_1 \cos \psi + A \zeta_2 \sin \psi \\
-C \zeta_3
\end{pmatrix}.
\]

We obtain the equation for Green’s function

\[
\partial_s G = \hat{\mathcal{H}}_\theta G,
\]

where

\[
\hat{\mathcal{H}}_\theta = \frac{1}{2} \frac{1}{\sqrt{g}} (\partial_j + a_j) \left( \sqrt{g} g^{ij} (\partial_i + a_i) \right).
\]

(27)

From equation (13), the transition probability density satisfies

\[
\partial_s p = (\hat{\mathcal{H}}_\theta - \overline{\hat{\mathcal{H}}}_\theta) p, \quad p|_{s=0} = \delta(\theta - \theta'),
\]

(28)

where

\[
\overline{\hat{\mathcal{H}}}_\theta = \int \hat{\mathcal{H}}_\theta p J d\theta.
\]

(29)

Some further calculations show that

\[
\hat{\mathcal{H}}_\theta - \overline{\hat{\mathcal{H}}}_\theta = \hat{\mathcal{H}}^1_\theta + \hat{\mathcal{H}}^2_\theta.
\]

(30)

where

\[
\hat{\mathcal{H}}^1_\theta = \frac{1}{2} \Delta_\theta, \quad \hat{\mathcal{H}}^2_\theta = g^{ij} a_i \partial_j.
\]

(31)

Now we suppose the heterogeneity is the Gaussian white noise as before except that \(\zeta_3\) is replaced with \(\zeta_3 + \omega\), where \(\omega\) represents the intrinsic twist of DNA. It is found that this does not affect the result.

We proceed as those in section 3 and section 4 to obtain the effective bending persistence length. After taking Taylor expansion of \(p(\theta, s|\theta', 0)\) to second order in integral form, we have

\[
\mathbb{E}C(s) = \mathbb{E} \left( \int \int E' \cdot E_{3, f_{eq}(\theta')} p(\theta, s|\theta', 0) J J' d\theta d\theta' \right)
\]

\[
= \mathbb{E} \left( \int \int E' \cdot E_{3, f_{eq}(\theta')} \left[ \delta + \int_0^s \hat{\mathcal{H}}^1_\theta \delta(\theta - \theta') ds' + \int_0^s \hat{\mathcal{H}}^2_\theta \delta(\theta - \theta') ds' \right. \right.
\]

\[
+ \left. \int_0^s \int_0^s \hat{\mathcal{H}}^2_\theta \delta(\theta - \theta') ds'' ds' \right] J J' d\theta d\theta' \right) + o(s),
\]

(32)
where $\hat{\mathcal{H}}_\theta^1$ and $\hat{\mathcal{H}}_\theta^2$ depend on $s, s'$ and $s''$ through the heterogeneity $\zeta_i(s)$. Direct calculation shows

$$E\left( \int \int E_3' \cdot E_3 f_{eq}(\theta') \delta JJ'd\theta d\theta' \right) = 1,$$

$$E\left( \int \int E_3' \cdot E_3 f_{eq}(\theta') \left( \int_0^s \hat{\mathcal{H}}_\theta^1 \delta ds' \right) JJ'd\theta d\theta' \right) = -\frac{1}{A}s,$$

$$E\left( \int \int E_3' \cdot E_3 f_{eq}(\theta') \left( \int_0^s \int_0^{s'} \hat{\mathcal{H}}_\theta^2 \delta ds' \int_0^{s''} \delta ds'' \right) JJ'd\theta d\theta' \right) = 0,$$

$$E\left( \int \int E_3' \cdot E_3 f_{eq}(\theta') \left( \int_0^s \hat{\mathcal{H}}_\theta^2 \int_0^{s'} \hat{\mathcal{H}}_\theta^2 \delta ds' ds'' \right) JJ'd\theta d\theta' \right) = -\frac{r}{A}s. \quad (33)$$

Here we used the following relation in (33)

$$\int_0^s \int_0^{s'} \delta(s' - s'') ds'' ds' = \frac{s^2}{2}.$$

Substituting the results above into (32), we obtain

$$E \mathcal{C}(s) = 1 - \frac{1 + r}{A}s + o(s). \quad (34)$$

So the effective bending persistence length $A_{eff} \approx \frac{A}{1+r}$ in 3D and we have

$$\frac{A_{eff}}{A} \approx 1 - r \quad (r \ll 1). \quad (35)$$

This result coincides with the formula obtained by Nelson [15] in the weak disorder limit.

6 Conclusion

In this paper, we clarify the difference between two results concerning the effective bending persistence length for heteropolymers in [1] and [15]. Based on the heterogeneous RLC model, we present a rigorous analysis. This corresponds to a path integral on the Lie group $SO(3)$ with random forcing. As the current heterogeneous RLC model is too idealized, the future work will be investigating the effective behavior in the base pair level.

Acknowledgement

T. Li acknowledge the support by National Science Foundation of China grants 10401004 and 20490222, and the National Basic Research Program under the grant 2005CB321704.
References


