A Random Loop Model for Long Polymers

Manfred Bohn* and Dieter W. Heermann
Institute of Theoretical Physics, University of Heidelberg, Philosophenweg 19, D-69120 Heidelberg, Germany
(Dated: May 10, 2007)

Models for technical polymers are too simple to describe the processes of long polymers in living organisms, as they are permanently altered by enzymes and protein complexes. We present a new polymer model based on a random walk with arbitrary harmonic long-range interactions that takes into account the dynamic behaviour of the polymer chain where attractive interactions change in a random manner. We derive a general expression for the thermal ensemble average of the squared physical distance between two arbitrary beads for a chain where beads interact via harmonic interactions. We show that the average over the disorder of loops (i.e., bead-bead interactions) is equivalent to the average over an adequate ensemble of diagonal-dominated random matrices.

PACS numbers: 05.40.Fb, 82.35.Pq, 82.35.Lr, 87.15.-v, 87.15.Aa
Keywords: Looped Polymers, Biopolymers, Random Matrices, Random Walk

Long macromolecules often can be described by quite simple polymer models. The basic idea behind polymer models is to assume the macromolecule to be composed of equal subunits (monomers) that are repeated along the chain. Mainly three basic polymer models are used: (a) the random walk (RW) model where it is assumed that the orientations between chain segments are completely uncorrelated and no volume interactions are taken into account, (b) the self-avoiding walk model (SAW) takes excluded volume into account, while (c) the globular state model furthermore includes temperature-dependent attractive interactions (cf. ref. [1, 2]). One characteristic feature of a polymer model is the end-to-end distance $\langle R^2 \rangle$. For the basic models above, the end-to-end distance scales like

$$\langle R^2 \rangle = l^2 N^{2\nu} \tag{1}$$

in the limit of large N, where l is the linker length, N the chain length and ν a constant depending on the model used.

These models describe well the situation for many technical polymers, as long as one looks at a length scale larger than the persistence length l_p, i.e., the length over which correlations in the bond angles become irrelevant. A polymer can be regarded as flexible above l_p. But there are situations, where these models can no longer correctly describe the situation. This is the case, for example, when long-range interactions can occur as in complex biological systems. A prime example for a biological polymer is the chromatin fiber situated inside the cell nucleus, where many enzymes and proteins are engaged in maintaining the living organism. These quite large molecules permanently access and alter the DNA and therefore its folding behaviour. Experiments have shown that permanent building and dissolving of loops occur mediated by so-called "transcription factories" [3, 4, 5]. The genomic code for different proteins which have to work together and therefore find each other has to be somehow spatially clustered for the time the code is being transcribed, while some time later in cell evolution another clustering of genes might be necessary to produce other proteins [6].

So it is very unlikely that in nature looped polymers stay in a static loop conformation, but that they are rather dynamic with different protein mediated interactions. This raises the question whether the models for technical polymers can be regarded as a correct and sufficient description of polymers in living systems.

In this paper we present a polymer model whose main feature is that it allows for random and dynamic loops on the basis of a random walk chain (for reasons of mathematical tractability we do not introduce self-avoidance). By *random* we mean that two arbitrary beads may interact with one another with a certain probability $\mathcal{P}$. *Dynamic* means that the interactions, i.e. the loops, may change in the course of time (see fig. 1). Thereby we advance to a more realistic description of polymers in living systems. Within this model we have to average over two ensembles. The first one is the thermal ensemble given by a certain set of interactions between the chain's beads. We are going to present a very general analytical result here. Then we have to average over the ensemble of different loop configurations, which comes into play by the dynamics of loop building and unfolding. This can only be done numerically.

Let us consider a chain with N segments. Let $\mathbf{x}_0, \ldots, \mathbf{x}_N$ be the three-dimensional equilibrium positions of the $N+1$ beads. We denote the components of the bead positions as $\mathbf{x}_i = (x_i, y_i, z_i)$. For reasons of mathematical tractability we consider the chain to be phantom, i.e., the beads are virtual and do not occupy volume in space, meaning that different beads can pass through each other. The linkers, too, are phantom. The probability density for a bead conformation $(\mathbf{x}_0, \ldots, \mathbf{x}_N)$

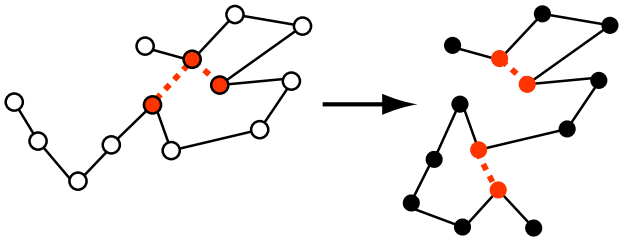

FIG. 1: Within the Random Loop Model one averages over the thermal disorder with one fixed loop configuration and over the disorder of loops. Shown are two different loop configurations of the ensemble.

can be regarded as a Boltzmann distribution

$$G(\mathbf{x}_0, \dots \mathbf{x}_N) = C \, \exp\left(-\frac{U}{k_B T}\right) \quad , \tag{2}$$

where C is a normalization constant and $U = U(\mathbf{x}_0, \dots \mathbf{x}_N)$ is the total potential energy of the chain. The potential for a Gaussian chain where adjacent beads along the chain interact with each other via harmonic potential, has the following form,

$$U_{\text{Gaussian}} = \frac{\kappa}{2} \sum_{j=1}^{N} \parallel \mathbf{x}_j - \mathbf{x}_{j-1} \parallel^2 \ .$$

For now, we keep U arbitrary. We want to obtain an expression for the distance probability density $P(\parallel \mathbf{x}_I - \mathbf{x}_J \parallel)$ of arbitrary chosen beads $\mathbf{x}_I$ and $\mathbf{x}_J$. This can be done by first calculating the marginal density $P(\mathbf{x}_I, \mathbf{x}_J)$

$$P(\mathbf{x}_I, \mathbf{x}_J) = \int \cdots \int \prod_{\substack{i=0 \\ i \neq I,J}}^{N} d\mathbf{x}_i \, C \, \exp\left(-U(\mathbf{x}_0, \dots, \mathbf{x}_N)/k_B T\right) \ . \tag{3}$$

To get the joint probability density $P_3(\mathbf{x}_I - \mathbf{x}_J) = P_3(x_I - x_J, y_I - y_J, z_I - z_J) := P_3(u, v, w)$ where $u = x_I - x_J$, $v = y_I - y_J$ and $w = z_I - z_J$ we have to integrate over x_I, y_I and z_I keeping $x_J = x_I + u$, $y_J = y_I + v$, $z_J = z_I + w$

$$P_3(u, v, w) = \iiint dx_I \, dy_I \, dz_I P(\mathbf{x}_I, \mathbf{x}_J)|_{\substack{x_J = x_I + u \\ y_J = y_I + v \\ z_J = z_I + w}} \tag{4}$$

The distance probability density of the distance between bead I and bead J, $P(r_{IJ})$ with $r_{IJ} = \parallel \mathbf{x}_I - \mathbf{x}_J \parallel$, can then be obtained by integrating over the sphere

$$P(r_{IJ})dr = \int d\phi \int d\theta \sin\theta \, r_{IJ}^2 \times P_3(r_{IJ}\cos\phi\sin\theta, r_{IJ}\sin\phi\sin\theta, r_{IJ}\cos\theta) \ . \tag{5}$$

From the distance probability distribution we obtain the mean square physical distance between bead I and bead J:

$$\langle r_{IJ}^2 \rangle_{\text{thermal}} = \int dr \, r^2 \, P(r) \tag{6}$$

where the average is performed over the thermal ensemble.

We stay very general, assuming that each bead may interact via a harmonic potential with any other, resulting in the potential

$$U = \frac{1}{2} \sum_{i<j}^{N} \kappa_{ij} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2 = \frac{1}{4} \sum_{\substack{i,j=0 \\ j \neq i}}^{N} \kappa_{ij} \parallel \mathbf{x}_i - \mathbf{x}_j \parallel^2 \ . \tag{7}$$

This potential has already been proposed in ref. [7], but not been elaborated further. It allows the building of loops of various sizes and constitutes the basis of our Random Loop Model.

We now eliminate the degrees of freedom stemming from the translational invariance of the problem by setting $\mathbf{x}_0 = \mathbf{0}$ (the absolute position of the chain in space is irrelevant for distances between beads).

Due to the Gaussian character of the probability density G, the spatial dimensions factorize,

$$G(\mathbf{x}_1, \dots, \mathbf{x}_N) = G_1(x_1, \dots, x_N) \cdot G_1(y_1, \dots, y_N) \cdot G_1(z_1, \dots, z_N)$$

and we can concentrate on the one-dimensional density function G_1. By an easy calculation omitted here, we can rewrite the one-dimensional probability G_1 density as follows

$$G_1(x_1, \dots, x_N) = C_1 \exp(-\frac{1}{2}\mathbf{X}^T K \mathbf{X}) \tag{8}$$

where $\mathbf{X} = (x_1, \dots, x_N)^T$ and

$$K = \begin{pmatrix} \sum_{\substack{j=0 \\ j \neq 1}}^{N} \kappa_{1j} & -\kappa_{12} & -\kappa_{13} & \dots & -\kappa_{1N} \\ -\kappa_{21} & \sum_{\substack{j=0 \\ j \neq 2}}^{N} \kappa_{2j} & -\kappa_{23} & \dots & -\kappa_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\kappa_{N1} & -\kappa_{N2} & \dots & \dots & \sum_{\substack{j=0 \\ j \neq N}}^{N} \kappa_{Nj} \end{pmatrix} \tag{9}$$

From now on, we assume the κ_{ij}'s to be given in units of $k_B T$. Up to now we did not make any assumptions concerning the spring constants (i.e. basically the matrix entries) κ_{ij}. In the following we only assume that K is a *symmetric* and *regular* matrix. So G_1 in eq. (8) turns out to be a multivariate normal distribution with mean $\mu = 0$ and covariance matrix $\Sigma = K^{-1}$. Evaluation of the integrals (3)-(5) by standard methods yields

$$P(\parallel \mathbf{x}_I - \mathbf{x}_J \parallel) \equiv P(r_{IJ}) = \tilde{C} \, r_{IJ}^2 \, \exp\left[-\frac{1}{2}\frac{1}{\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}} r_{IJ}^2\right] \ .$$

Here

$$\Sigma = K^{-1} = (\sigma_{ij})_{i,j}$$

and $\tilde{C}$ is the normalization constant. We can rewrite this using

$$\Gamma = \frac{1}{2}\frac{1}{\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ}}$$

$$P(r_{IJ}) = \tilde{C}\; r_{IJ}^2\; \exp\left[-\Gamma r_{IJ}^2\right]$$

and with the correct normalization we obtain

$$P(r_{IJ}) = \frac{4}{\sqrt{\pi}}\,\Gamma^{\frac{3}{2}}\; r_{IJ}^2\; \exp\left[-\Gamma r_{IJ}^2\right] \tag{10}$$

and finally

$$\left\langle r_{IJ}^2 \right\rangle_{\text{thermal}} = \frac{3}{2\Gamma} = 3(\sigma_{JJ} + \sigma_{II} - 2\sigma_{IJ})\;. \tag{11}$$

The bracket delimiters here denote the average over the thermal ensemble of $N+1$ beads interacting via a given, *static* (i.e. time-independent), harmonic potential.

The average over the thermal ensemble as given by eq. (11) only depends on the matrix K, which contains all the information about the chain's interactions. Our model is supposed to have random and dynamic loops. Building random loops means that we have to fill the matrix K in a random manner. The average over the disorder of loops is then equivalent to an average over the ensemble of random matrices.

First of all we have to specify the subset of all regular and symmetric real matrices K that is of interest. Of course, the matrices must satisfy the form of eq. (9), i.e., the principal diagonal entries are the negative sum of all other entries in the specific row. Furthermore, as we want to maintain a Gaussian chain, the secondary diagonals have to be filled with $-\kappa$ where κ is a parameter defining the interaction strength.

Building random loops means that we have to have some of the $\kappa_{ij} = \kappa_{ji} \neq 0$ ($|i-j| > 1$) with a certain probability $\mathcal{P}$. How this is done depends on the physical situation one wants to model. For a chain with $N+1$ beads (represented by a $N \times N$-matrix) we have

$$\mathcal{N} = \sum_{n=1}^{N-2} n = \frac{(N-2)(N-1)}{2} \tag{12}$$

independent entries that can be filled in a random fashion.

Once we have chosen the set of random matrices appropriate for the situation we can perform an averaging over this ensemble, which is in fact an average over the disorder of loops.

In this paper, we present results for one specific type of random interactions, i.e. random loops: We assume the polymer to form loops in a random manner, but allow only loops of a certain range of sizes ℓ determined through the condition $1 < l_1 \leq \ell \leq l_2 \leq N$. This can be achieved by setting $\kappa_{ij} = \kappa_{ji} = \kappa$ with probability $\mathcal{P}$ and $\kappa_{ij} = \kappa_{ji} = 0$ with probability $1-\mathcal{P}$ if $l_1 \leq i-j \leq l_2$, and $\kappa_{ij} = \kappa_{ji} = 0$ otherwise. Mathematically this yields a random matrix K with a band structure.

To obtain the mean square displacement between two beads we have to average over two ensembles. The first one is the thermal ensemble given by eq. (11), which is an average over one static interaction configuration. Then we have to average over the disorder of different loop configurations, equivalent to averaging over the specific ensemble of random matrices. In the following, the bracket delimiters without subscript denote these two averages

$$\left\langle R^2 \right\rangle = \left\langle \left\langle R^2 \right\rangle_{\text{thermal}} \right\rangle_{\text{loops}}\;.$$

Averaging over the matrix ensemble cannot be done analytically, so we have to approximate by averaging numerically over a smaller set of representative matrices.

As an example, we show the results for the mean square displacement $\left\langle R^2 \right\rangle$ in relation to the contour length for the special case of allowing all loop sizes between 2 and N. We have three adjustable parameters within this model: a) the chain length N, b) the probability $\mathcal{P}$ and c) the interaction strength κ. The basic traits of this model can be seen in fig. 2. At short contour lengths the mean square displacement grows like a random walk, but soon a leveling-off can be observed due to the attractive long-range interactions. Thus, this model yields completely different traits than a simple random walk or self-avoiding walk model. Note that the probabilities $\mathcal{P}$ are chosen very small, because one only wants to have a few loops. The average number of loops per configuration is $\mathcal{N}\mathcal{P}$. For $\mathcal{P} = 4 \cdot 10^{-5}$ and $N = 1000$ we have about 20 loops.

Starting from a biological point of view we have developed a new polymer model that is closer to biological systems than "static" models (with regard to the structure of interactions) for technical polymers. It has been shown by biologists that polymers in living systems are permanently altered by enzymes and protein complexes. This of course requires the development of new models that are able to predict the average properties of such a polymer. Our model is based on a random-walk-like chain, but it allows rather arbitrary harmonic interactions between any of the chain's beads. The dynamics of a biological system comes in when we average over the disorder of loops. We have shown that a configuration of interactions can be described by a random matrix ensemble, eq. (9), with a specific form where the principal diagonal contains the sum over all other entries in the row and the secondary diagonal is filled by a spring constant κ. All other entries can be chosen in a random fashion. The mean square displacement $\left\langle R^2 \right\rangle$ then depends on the

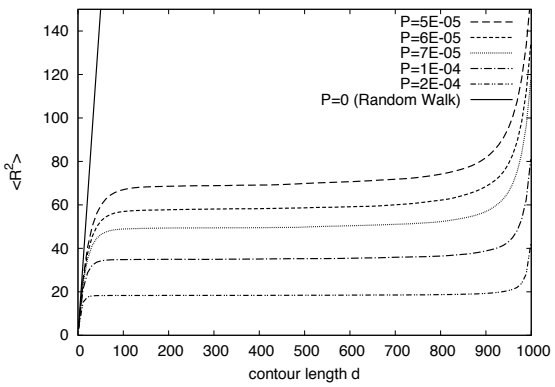

FIG. 2: Mean square physical distances in relation to genomic separation. The chain length is $N = 1000$, each bead interacts with any other with probability $\mathcal{P}$. The average denotes an average over the thermal ensemble (analytical) and the disorder of loops (numerical).

average properties of the ensemble of random matrices. This ensemble is to be defined according to the situation. We proposed a model where loop sizes of a certain range $[l_1, l_2]$ are allowed, equivalent to a band structure of the matrices K.

The model leaves much room for further investigations. There is a wide variety of random matrix ensembles one can choose, resulting in different behaviour of the mean square displacement. There is also theoretical interest in studying the diagonal-dominated random matrix ensembles resulting from our model. It was shown recently that these matrices belong to a new universality class [8].

M. Bohn gratefully acknowledges funding from the Landesgraduiertenförderung. We are also grateful for the fruitful discussions with Prof. R. van Driel, Prof. H. Horner and M. Hellmann.

* Electronic address: bohn@tphys.uni-heidelberg.de

[1] A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, 1994).

[2] P.-G. de Gennes, *Scaling concepts in polymer physics* (Ithaca, N.Y., Cornell University Press, 1979).

[3] M. J. Lercher, A. O. Urrutia, and L. D. Hurst, Nat Genet **31**, 180 (2002), ISSN 1061-4036, URL `http://dx.doi.org/10.1038/ng887`.

[4] P. R. Cook, Nat Genet **32**, 347 (2002), URL `http://dx.doi.org/10.1038/ng1102-347`.

[5] D. A. Jackson, S. J. McCready, and P. R. Cook, J Cell Sci Suppl **1**, 59 (1984).

[6] C. Lanctôt, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer, Nat Rev Genet **8**, 104 (2007), URL `http://dx.doi.org/10.1038/nrg2041`.

[7] R. Sachs, G. Engh, B. Trask, H. Yokota, and J. Hearst, PNAS **92**, 2710 (1995), http://www.pnas.org/cgi/reprint/92/7/2710.pdf, URL `http://www.pnas.org/cgi/content/abstract/92/7/2710`.

[8] D. W. Heermann and M. Bohn, arXiv:0705.1241v1 [cond-mat.stat-mech] (2007), URL `http://arxiv.org/abs/0705.1241v1`.