Systematic Coarse-Graining of Microscale Polymer Models as Effective Elastic Chains

Elena F. Koslover*† and Andrew J. Spakowitz*†‡

†Biophysics Program, Stanford University, Stanford, California 94305, United States
‡Chemical Engineering Program, Stanford University, Stanford, California 94305, United States

ABSTRACT: One of the key goals of macromolecular modeling is to elucidate how macroscale physical properties arise from the microscale behavior of the polymer constituents. For many biological and industrial applications, a direct simulation approach is impractical due to the wide range of length scales that must be spanned by the model, necessitating physically sound and practically relevant procedures for coarse-graining polymer systems. We present a highly general systematic coarse-graining procedure that maps any detailed polymer model onto effective elastic-chain models at intermediate and large length scales. Our approach defines a continuous flow of coarse-grained models starting from the detailed microscale model, proceeding through an intermediate-scale stretchable, shearable wormlike chain, and finally resolving to a Gaussian chain at the longest lengths. We demonstrate the utility of this procedure by coarse-graining a wormlike chain with periodic rigid kinks, a model relevant for studying DNA–protein arrays and genome packing. The methodology described in this work provides a physically sound foundation for coarse-grained simulation of polymer systems, as well as analytical expressions for chain statistics at intermediate and long length scales.

1. INTRODUCTION

Polymer models are crucial to the study of a wide variety of materials ranging from DNA†‡−17 and cytoskeletal fibers5−7 in the biological context to nanotubes,8 gels,9,10 and plastics11 for technological applications. One of the key challenges in modeling polymeric materials lies in bridging the gap between the length scales of the underlying mechanics and the much larger length scales relevant to the macroscale applications of interest. For instance, polyethylene can be modeled as a freely rotating chain with angstrom-length segments,12 but this level of detail makes it impractical to simulate macroscale properties of a piece of plastic. Another example is the DNA polymer, whose molecular structure gives rise to an elastic rigidity at the nanometer length scale,1 whereas contour lengths up to several meters must be packaged, processed, and accessed by the cell.13 A meaningful approach to coarse-graining polymer models is thus essential for modeling a broad range of systems of practical relevance.

A variety of general approaches have been used for developing coarse-grained models of polymer systems. One class of models consists of united-atom formulations, where groups of atoms in a specific polymer structure are represented as coarse beads and detailed atomistic simulations are used to extract appropriate parameters for effective coarse-grained interactions.14−17 These methods can incorporate complicated nonlocal interaction potentials while reproducing the behavior of realistic polymer systems.18,19 However, the reliance on simulations makes such coarse-graining procedures computationally expensive to implement, limits them to individual specific polymer systems, and restricts the coarsening scale to mesoscopic models. An alternate approach makes use of idealized elastic models comprising either bead–spring assemblies20,21 or continuous chains22,23 to address highly coarse-grained macroscale properties of general polymeric systems, including rheology, chain size, and response to external force. In this work, we focus on the latter approach of treating polymers as elastic threads. A common procedure for modeling elastic polymers involves representing the chain as a sequence of discrete rigid links with an appropriate effective bending potential.24−26 This procedure is accurate for short discretization lengths but is unable to resolve intermediate-length behavior as the discretization is increased. An alternate approach for large-scale coarse-graining is to treat the polymer as an effective Gaussian spring. A number of works have addressed the coarse-graining of specific polymer models, such as the wormlike chain and the freely jointed chain, by mapping to bead–spring models that are valid when the contour length encompassed by each spring is quite long.1,21,27 Higher accuracy for more finely discretized models can be achieved by using alternative spring force laws, with the functional form extracted from the force–extension behavior of the polymer.28,29 The desired degree of coarse-graining arises from a balance between the relevant length scales dictating the physics of interest and the limitations of computational power required to encompass the overall size of the system. A general procedure that can bridge from finely detailed to roughly coarse-grained polymer models within a single framework remains imperative for modeling multiscale phenomena.

We present a systematic analytical procedure for generating increasingly coarse-grained models for a locally defined

Received: October 2, 2012
Revised: December 10, 2012

dx.doi.org/10.1021/ma302056v • Macromolecules XXX, XXX, XXX−XXX
polymer, while maintaining the correct statistics in the long-chain limit. The configuration statistics of any polymer chain can be fully defined by calculating all moments of the end-to-end distance \( \langle R_e^2 \rangle \) as a function of contour length \( L \). At sufficiently long lengths, any polymer can be mapped to an effective Gaussian chain where all such moments match up to an error that scales as \( 1/L \) (as a fraction of the long-chain limit). While such a mapping is used routinely to approximate a long polymer as a random walk, it fails to account for correlations in the polymer contour that become increasingly important on intermediate length scales. Our novel coarse-graining procedure maps a detailed polymer model to an effective elastic chain that matches all moments out to an error scaling as \( 1/L^2 \).

In section 2, we focus on developing a continuous polymer model that serves as a logical extension beyond the Gaussian chain, capturing behavior at intermediate length scales. We show that a stretchable, shearable wormlike chain (ssWLC) with bend-shear coupling is a universally appropriate model for coarse-graining across a range of length scales. Exact analytical expressions for low-order moments of the end-to-end distance, as well as the full structure factor, are provided for this model. We describe a procedure for extracting the model parameters when coarse-graining a detailed polymer. We then demonstrate that our procedure provides a continuous flow of increasingly coarse-grained ssWLC models from the classic Kratky–Porod inextensible wormlike chain at short length scales to the Gaussian chain at long lengths.

In section 3, we provide an example of a more complicated polymer model that can be coarse-grained using this procedure. Specifically, we show how a wormlike chain with periodic kinks can be mapped to an effective ssWLC at intermediate and long lengths. This example is motivated by its applicability to studying DNA–protein arrays. We show that, on length scales much shorter than those of a Gaussian-chain random walk, this procedure can reproduce the statistics of such kinked arrays with a simple, easy to simulate, and analytically tractable continuous polymer model.

2. SYSTEMATIC POLYMER COARSE-GRANING PROCEDURE

We describe a general approach for modeling the statistical behavior of a polymer chain based on the most localized energy contributions. Our procedure provides the lowest order correction beyond the Gaussian chain at long lengths, thereby encompassing intermediate length-scale properties. The model that emerges from this approach incorporates a local chain orientation, while simultaneously allowing stretching and shearing of the chain path relative to this orientation. Previous work by Wiggins and Nelson has shown that a discrete polymer with rigid links and an arbitrary bending potential can be renormalized to a wormlike chain at long lengths if the distribution of consecutive segment orientations is narrowly distributed around the straight angle. A recent study by Wolle et al. compared coarse-grained models of double-stranded DNA, showing how more detailed models converge to an anisotropic elastic filament at long length scales. Here we develop a more generalized result, demonstrating that any locally defined polymer chain can be mapped to a stretchable, shearable worm-like chain (ssWLC) model. When applied to a stiff chain, the ssWLC encompasses both the classic wormlike chain model at short length scales and the Gaussian chain at long lengths. The model incorporates a sliding parameter that sets the degree of coarse-graining, enabling expansion or contraction of the length scale of accuracy to suit the relevant physics of interest.

**Continuous Stretchable, Shearable Wormlike Chain.**

The path of a continuous polymer chain is defined by a space curve \( \vec{\tau}(s) \), where \( 0 \leq s \leq L \) is a coordinate marking position along the contour. The energy \( E \) of a particular configuration of the polymer can be expressed as

\[
E[\vec{\tau}(s)] = \int_0^L ds \mathcal{H}[\vec{\tau}_s, \vec{\tau}_{uw}, \vec{\tau}_{su}, \ldots]
\]

(1)

where \( \vec{\tau}_s = (\partial \vec{\tau}/\partial s), \vec{\tau}_{uw} = (\partial^2 \vec{\tau}/\partial s^2) \), and so forth. The increasing derivatives of \( \vec{\tau} \) correspond to increasing nonlocal contributions of chain deformation to the energy.

We note that this class of chains includes many commonly used models in polymer physics, including wormlike chains, freely jointed chains, and rotational isomeric state models. More general polymer models such as the twisted helical wormlike chain or the birod chain can include additional local degrees of freedom (e.g., twist density) that allow for anisotropic flexibility and intrinsic curvature. However, if we are interested specifically in the spatial distribution of the chain contour, then the formulation given by eq 1 remains fully applicable, with the Hamiltonian expressed as a function of \( \vec{\tau}(s) \) by integrating over any additional degrees of freedom.

In this manuscript, we focus specifically on “locally defined” polymer chains—ones where the correlations and interactions between different segments along the chain decay on some finite length scale. When viewed on a length scale that is much larger than the microscopic correlation length scales, all such polymers have an energy function that is effectively localized. For the general continuous chain defined by eq 1, if we zoom out far enough then all nonlocal physical effects should disappear, and to lowest order, the Hamiltonian \( \mathcal{H} \) can be approximated as \( \mathcal{H}[\vec{\tau}_s] = A |\vec{\tau}|^2 \). This approximation results in the Gaussian chain model, a single-parameter model to which all polymer models flow at sufficiently long lengths.

On shorter length scales, less localized contributions to the Hamiltonian will begin to matter, and a more general function \( \mathcal{H}[\vec{\tau}_s, \vec{\tau}_{uw}] \) is required. Here, we develop a model that expands this function out to second order, keeping only those terms permitted by symmetry considerations. For the subsequent analyses, we require the Hamiltonian to be expressed as a function of general coordinates and their first derivatives. We thus define two auxiliary variables \( \sigma(s) \) and \( \vec{u}(s) \), where \( \vec{u} \) is a unit vector and the ground-state polymer configuration must have \( \vec{\tau}_{uw}(s) = \sigma(s)\vec{u}(s) \). Essentially, the auxiliary vector \( \vec{u} \) tracks the chain orientation. We then have \( \vec{\tau}_s = \sigma \vec{u} + \sigma \vec{u} \). The first part \( \sigma \vec{u} \) is a second-order term that cannot enter linearly into the Hamiltonian due to symmetry with respect to reversal of the chain direction. As we are expanding only to second-order, this term does not contribute. Consequently, we can forego the auxiliary variable \( \sigma \) entirely and construct the Hamiltonian as a linear combination of \( \vec{\tau}_s, \vec{u}, \vec{u} \), \( \vec{u} \), and \( \vec{u} \), where we have kept only those terms of first or second order that are not disallowed by symmetry arguments. We refer to this continuous chain as the stretchable, shearable wormlike chain model (ssWLC) and for greater physical clarity express the Hamiltonian as

\[
\mathcal{H} = \frac{\varepsilon_u}{2} |\vec{u}| - \eta |\vec{\tau}|^2 + \frac{\varepsilon_u}{2} (\vec{u} \cdot \vec{u} - \gamma)^2 + \frac{\varepsilon_u}{2} |\vec{\tau}|^2
\]

(2)
where \( \mathbf{r}_1 = \mathbf{r}(s) - \mathbf{r}(\bar{u}) \). We note that the ssWLC model as presented here constitutes a modified version of previously solved elastic chain models, including the anisotropic elastic chain and the extensible shearable wormlike chain in that it includes an additional coupling term between the bend and shear degrees of freedom.

The model described by this Hamiltonian can be understood intuitively by considering the coarse-grained behavior of an arbitrarily long polymer chain, where each point along the contour has both a position \( \mathbf{r}(s) \) and an orientation \( \bar{u}(s) \). We can then look at the distribution of these quantities at position \( s + \Delta s \) relative to position \( s \), where \( \Delta s \) encompasses an intermediate length of the polymer compared to the length scale at which correlations in the chain contour die off. Such an intermediate length polymer would resist bending (Figure 1a), stretching or compression away from its ground-state structure, and resist shearing away from this orientation (Figure 1b). The polymer orientation to bend in the same direction as the bend-shear coupling term modulated by \( \epsilon_b \). We note that this model encompasses both the standard Kratky–Porod wormlike chain model in the limit \( \epsilon_b \to \infty \) and the Gaussian chain in the limit \( \epsilon_b \to 0 \) or \( \gamma = 0 \). The statistics of the ssWLC without bend-shear coupling have been previously studied, and we utilize similar approaches for calculations for our more general model.

In the subsequent sections, we develop expressions for the structure factor and low-order moments for the ssWLC with bend-shear coupling. We then show how to set the parameters of the ssWLC model in such a way as to match the moments of the end-to-end separation for a detailed polymer chain at an adjustable level of coarse-graining.

**Figure 1.** Schematic of the ssWLC model, indicating the components of the energy function corresponding to different deformations of a coarse-grained polymer chain away from its ground-state structure, including (a) bending of the chain orientation, (b) stretching or compression along the contour, (c) shearing perpendicular to the chain orientation, and (d) coupling of the bend and shear deformations.

**Distribution Function for the ssWLC.** To calculate the statistics of the ssWLC model defined by eq 2, we make use of the path integral formulation of the Green’s function for the chain distribution, defined by

\[
G(\vec{R}, s \bar{u} \bar{u}_0; L) = \int_{\vec{R}(0)=\bar{u}_0, \bar{R}(L)=\vec{R}} \mathcal{D}[\bar{u}(s)] \mathcal{D}[ar{u}(s)] \exp\left\{-\int_0^L ds \mathcal{H}[\bar{u}(s), \bar{u}(s)]\right\}
\]

where the Hamiltonian is made dimensionless by \( k_B T \). This Green’s function gives the probability density for a chain of length \( L \) starting at the origin with orientation \( \bar{u}_0 \) and ending at position \( \vec{R} \) with orientation \( \bar{u} \). It is normalized according to \[ \int G(\vec{R}, s \bar{u} \bar{u}_0; L) d\vec{R} d\bar{u} = 1 \]. Following an approach analogous to one developed for helical wormlike chains, we find the governing “Schrödinger” equation by drawing an analogy to the quantum-mechanical propagator for a coupled symmetric top and point particle with a velocity-dependent potential energy. We define the angular momentum operator \( \vec{\Lambda} \) for the chain orientation, where each component \( \Lambda_j \) corresponds to a differential rotation around the \( j \) axis. The Green’s function then obeys

\[
\left[ \frac{\partial}{\partial L} + \hat{\mathcal{H}} \right] G(\vec{R}, s \bar{u} \bar{u}_0; L) = \delta(L)\delta(\vec{R})\delta(\bar{u} - \bar{u}_0)
\]

where

\[
\hat{\mathcal{H}} = -\frac{1}{2l_p}\Lambda^2 + \frac{\eta^2}{2\epsilon_b} (\Lambda \cdot \bar{u})^2 - \frac{1}{2\epsilon_b} [(1 - \bar{u} \cdot \bar{u}) \cdot \nabla \bar{u}]^2
\]

\[
- \frac{1}{2\epsilon_b} (\bar{u} \cdot \nabla \bar{u})^2 + \gamma \bar{u} \cdot \nabla \bar{u} - \frac{\eta}{\epsilon_b} (\bar{u} \times \nabla \bar{u}) \cdot \vec{\Lambda}
\]

and \( l_p = \epsilon_b/(1 + \eta^2 \epsilon_b/\epsilon_b) \). We proceed to perform a Fourier transform from position \( \vec{R} \) to wavevector \( \vec{k} \), a Laplace transform from \( L \) to Laplace variable \( p \), and a spherical harmonic transform in the orientations \( \bar{u} \) and \( \bar{u}_0 \). Namely, we expand the Green’s function as

\[
\hat{G}(\vec{k}, s \bar{u} \bar{u}_0; \bar{p}) = \sum_{j,l} g_{sl,j}(\vec{k}, \bar{p}) Y^j_0(\bar{u}_0) Y^j_0(\bar{u})
\]

and project eq 4 onto the spherical harmonics. Equation 4 then becomes

\[
p g_{sl,j} + \sum_{n=0}^\infty \sum_{n=0}^\infty g_{sl,n} A_{sl} = \delta_{sl,j}
\]

where the matrix \( A \) is defined by the action of the operator \( \hat{H} \) on the spherical harmonics:

\[
A_{ll'} = \int Y^j_0(\bar{u}) \hat{H} Y^j_0(\bar{u}) d\bar{u}
\]

To simplify subsequent calculations, reorient the universe frame of reference such that \( \vec{k} = \vec{k}_z \). The elements of \( A \) are then given by

\[
\text{doi.org/10.1021/ma302056v | Macromolecules XXX, XXX, XXX--XXX}
\]
\[ A_{l,i} = \frac{(l+1)}{2p} \frac{\eta^2}{\epsilon} - \frac{\eta^2}{2\epsilon} + \frac{k^2}{2\epsilon} (1 - \epsilon) + \frac{k^2}{2\epsilon} \epsilon \]

\[ A_{l,i+1} = -i\epsilon a_i + \frac{i\epsilon}{\epsilon} a_{i+1} \]

\[ A_{l,i-1} = -i\epsilon a_i - \frac{i\epsilon}{\epsilon} (1 + 1) a_i \]

\[ A_{l,i+2} = -\frac{k^2}{2\epsilon} b_{i+2} + \frac{k^2}{2\epsilon} b_{i+2} \]

\[ A_{l,i-2} = -\frac{k^2}{2\epsilon} b_i + \frac{k^2}{2\epsilon} b_i \]

where \( [a_i = ((l-j)l-j)/(4l-1)]^{1/2}) \), \( b_i = a_i^2 - 1 \), and \( \epsilon = (a_i^2 + (d_{i-1})^2 \).

Finally, the coefficients for the spherical harmonic expansion of the Green's function can be found as

\[ g = (pI + A)^{-1} \]

We note that eq 4–9 closely parallel those developed for the helical wormlike chain in the classic treatment by Yamakawa,\(^{23}\) although the form of our Hamiltonian, and thus the 5-diagonal structure of our matrix \( A \), differ to incorporate the stretching, shearing, and coupling terms of the ssWLC model.

If we are interested only in the end distribution of the polymer, then integrating eq 5 over the orientation \( \vec{u} \) and averaging over \( \vec{u} \), leaves \( \overline{G(kp)} = g_{00}(k) \). Since there is no coupling between the indices of the coefficients, we set \( j = 0 \) in all subsequent calculations and drop the \( j \) index entirely. The structure factor of the chain is defined as

\[ S(k) = \frac{1}{L} \int_0^L ds_1 \int_0^L ds_2 \langle \exp[i\vec{k} \cdot (\vec{r}(s_1) - \vec{r}(s_2))] \rangle \]

\[ = \frac{2}{L} \mathcal{L}^{-1} \left( \frac{\overline{G(k; p)}}{p^2} \right) \]

where the angled brackets indicate an average with respect to the chain statistics, and \( \mathcal{L}^{-1} \) denotes the inverse Laplace transform from \( p \) to the chain length \( L \). The inverse Laplace transform can be performed by summing over the poles of \( g_{00} \). Namely,

\[ \mathcal{L}^{-1} \overline{g_{00}}/p^2 = A^{-1}L + (A^{-1})^2 + \sum \frac{\lambda_i A_0 A_{0i}}{\lambda_i^2} \Pi_{\lambda_i} (A_{0i} - \lambda_i)^{-1} \]

where \( \lambda_i \) are eigenvalues of \(-A\), and \( \Pi_{\lambda_i} \) denotes the determinant of the matrix with the first row and column removed. [i.e., the \((0,0)\) cofactor of the matrix]. While \( A \) is an infinite matrix, it must be truncated at a finite value of \( l \) for practical calculations. Truncation at level \( 2l \) yields exactly correct values of the moments up to \( \langle R_z^2 \rangle \) and approximately correct values of higher moments at large \( L \). In all calculations below, we truncate at \( l = 10 \).

The full distribution of the end-to-end distances for the ssWLC can be obtained by numerically inverting the Fourier–Laplace transform. That is, we find the Green’s function for a ssWLC with free ends as

\[ G(\vec{R}; L) = \frac{1}{(2\pi)^3} \int dk \exp(-ik \cdot \vec{R}) G(k; L) \]

\[ = \frac{2}{(2\pi)^3} \int_0^\infty dk \frac{k \sin(kR)}{R} \mathcal{L}^{-1}[g_{00}(k, p)] \]

where the Laplace inversion is performed by summing over poles, analogously to eq 11.

**Calculating low-order moments.** To find moments of the end-to-end distribution, we use

\[ \langle R_z^n \rangle = \left( -i \right)^n \frac{\partial^n g_{00}}{\partial k^n} \]

We make use of the fact that \( A_{ij}(k = 0) \) is a diagonal matrix, \((\partial A_{ij}/\partial k)_{k=0} \) is nonzero only when \( l' = l \pm 1 \), and \((\partial^2 A_{ij}/\partial k^2)_{k=0} \) is nonzero only when \( l' = l \) or \( l' = l \pm 2 \). Consequently, the derivatives of the matrix inverse can be expanded with the help of Stone-fence diagrams that have been employed for calculating moments of plain wormlike chains,\(^{25}\) expanded to include the \( k^2 \) terms arising from the shear and stretch components (see Figure 2). For instance, the lowest nonzero moment is given by the expression

\[ \langle R_z^2 \rangle = -2(A_{00} + p)^{-1} \frac{\partial A_{01}}{\partial k} (A_{11} + p)^{-1} \frac{\partial A_{10}}{\partial k} (A_{00} + p)^{-1} \]

\[ + (A_{00} + p)^{-1} \frac{\partial^2 A_{00}}{\partial k^2} (A_{00} + p)^{-1} \]

\[ + (A_{00} + p)^{-1} \frac{\partial A_{10}}{\partial k} (A_{11} + p)^{-1} \frac{\partial A_{10}}{\partial k} (A_{00} + p)^{-1} \]

\[ + (A_{00} + p)^{-1} \frac{\partial^2 A_{00}}{\partial k^2} (A_{00} + p)^{-1} \]

\[ \] corresponding to the stone-fence diagrams in Figure 2a. The moment \( \langle R_z^n \rangle \) is given by putting together all diagrams of order \( 2n \) that start and end at the \( l = 0 \) level. The coefficient associated with each diagram is given by \((-1)^{(n-\text{even})(2n)}/2m_n\), where \( m_n \) is the number of second-order steps. It is clear from this diagrammatic notation that only terms up to the \( l = n \) level contribute to \( \langle R_z^n \rangle \).

When inverting the Laplace transform in chain length \((p \rightarrow L)\), each diagram contributes a term scaling as \( L^{n-2} \) where \( t \) is the number of times that the diagram touches the \( l = 0 \) level. Thus, for all moments of \( R_z \) only the diagrams confined to the lowest levels contribute in the long length limit. We expand the moments at long length as
\[ \langle R_z^{2n} \rangle = \sum_{k=0}^{n} C_{2n}^{(k)} L^k + E(L) \]  

(15)

where \( E(L) \) includes all terms that decay exponentially with length. The two highest order terms in this expansion are then given by

\[ \langle R_z^2 \rangle = aL + b + \ldots \]

\[ \langle R_z^4 \rangle = 3a^2L^2 + (12ab + c)L + \ldots \]

\[ \langle R_z^{2n} \rangle = \frac{(2n)!}{2^n n!} a^n L^n + \left[ \frac{(2n)!}{2^{n-2}4!(n-2)!} a^{n-2}c + \frac{(2n)!n}{2^{n}(n-1)!} a^{n-1}b \right] L^{n-1} + \ldots \]

where \( a \) and \( b \) are obtained from Laplace inversion of the diagrams in Figure 2a, and \( c \) from the inversion of the last five diagrams in Figure 2b.

We note that the development of the moment calculations from eq 9 up to this point can be generalized to any polymer model with an appropriate definition of the matrix \( A \). In the case where \( A \) contains higher-order terms in \( k \), the stone-fence diagrams must be modified to include higher-order steps, as will be seen in section 3 of this manuscript. The expression for the long-length limit of the moment expansion given in eq 16 is fully general. Consequently, if two polymer models match in the three expansion coefficients \( C_2^{(0)}, C_2^{(1)}, C_4^{(1)} \), then all subsequent moments will also match up to the two highest order terms in \( L \). This matching procedure is analogous to the standard mapping of a polymer to the Gaussian chain by finding its Kuhn length, \( C_2^{(0)} \). The Gaussian mapping returns moments that deviate from those of the original chain with an error that scales as \( 1/L \), yielding errors that scale as \( 1/L^2 \).

For the continuous ssWLC model, we can plug the expressions for matrix \( A \) defined in eq 8, into the diagrams illustrated in Figure 2, and invert the Laplace transform to get

\[ a = \frac{2}{3} \left( \gamma + \frac{2\eta}{\epsilon} \right) l_p + \frac{2}{3\epsilon \ell} + \frac{1}{3\epsilon \ell} \]

\[ b = -\frac{2}{3} \left( \gamma + \frac{2\eta}{\epsilon} \right) l_p \]

\[ c = \frac{32}{45} \left( \gamma + \frac{2\eta}{\epsilon} \right) \left( \gamma - \frac{\eta}{\epsilon} \right) \left( \gamma + \frac{3\eta}{\epsilon} \right) + 4\gamma + \frac{2\eta}{\epsilon} \left( \frac{3}{5\ell} + \frac{2}{5\ell} l_p + \frac{8}{45} \frac{1}{\epsilon \ell} - \frac{1}{\epsilon \ell} \right) l_p + \frac{16\gamma}{45} \left( \frac{1}{\ell} - \frac{1}{\epsilon \ell} \right) \left( \gamma - \frac{\eta}{\epsilon} \right) + \left( \gamma + \frac{2\eta}{\epsilon} \right) \left( \gamma + \frac{3\eta}{\epsilon} \right) \]

(17)

Finally, we note that the persistence length for the ssWLC can easily be found using

\[ \left\langle \tilde{u}_o \tilde{u}_o \right\rangle = \frac{1}{l_p \alpha} \left( l_p^2 \left( 3\alpha + 1 \right) + 4\left( 1 - \alpha \right) l_p^2 C_2^{(1)} \right) + l_p^2 \left[ 144C_2^{(0)} + 12\alpha^2 C_2^{(1)} (15\alpha^2 - 6\alpha - 5) + 64\alpha^2 \left( C_2^{(1)} \right)^2 + 10(1 - 2\alpha - 3\alpha^2) C_4^{(1)}/2 \right] \]

(18)

where the final step is inversion of the Laplace transform. Thus, the parameter \( l_p = e_\eta/(1 + \eta^2 e_\eta/\epsilon_\ell) \) gives the persistence length for the chain orientation.

Coarse-Graining with the ssWLC. To map an arbitrary polymer model to an equivalent ssWLC that will exhibit matching statistics at long lengths, we need to know the low-order moments of the polymer model. Specifically, we require the coefficients \( C_2^{(0)}, C_2^{(1)} \) corresponding to the linear and constant terms of \( \langle R_z^2 \rangle \) and the coefficient \( C_4^{(1)} \) for the linear term of \( \langle R_z^4 \rangle \). For a continuous chain, these terms can be found using the procedure described in the previous section. For a discrete chain, an analogous procedure is described in section 3 of this manuscript. Furthermore, in the absence of an analytically tractable form, these quantities can be found by running direct simulations of the detailed polymer in question.

We want to find appropriate parameters for the ssWLC such that these low-order moment terms match those of the detailed polymer. Our model has 5 parameters \( (\epsilon_\eta, e_\eta, e_\ell, \gamma, \eta) \) with which to match the three terms \( (C_2^{(0)}, C_2^{(1)}, C_4^{(1)}) \). We remove one of the excess degrees of freedom by defining a dimensionless parameter \( \alpha = \eta^2 e_\eta/\epsilon_\ell \) and setting this equal to a constant. This allows us to decouple the persistence length (which measures correlations in the chain orientation) from the shear modulus. As a result, the chain orientation distribution now depends entirely on \( e_\eta \) (bending modulus) while the stretch and shear modulus, as well as \( \gamma \) influence only the position of the chain ends. The \( \alpha \) parameter can be set to any positive constant, and could, in principle, be varied to minimize the error in any additional statistical average term of the polymer model (e.g., \( C_4^{(0)} \) or \( C_4^{(1)} \)). However, there is no clear choice of a single additional statistic that would improve the overall accuracy of the coarse-grained model. For all results presented in this manuscript, we found a constant value of \( \alpha \) to be a reasonable choice.

The second extra degree of freedom is used to adjust the length scale of accuracy for the effective model. We recall that the development of the ssWLC as a general model for locally defined polymer chains rested on the assumption that chain correlations decay on a sufficiently small length scale compared to the overall chain length. Consequently, increasing the persistence length \( l_p \) also raises the length scale above which the model is accurate. Thus, the parameter \( l_p \) functions as a slider to specify the requisite level of coarse-graining. When simulating the ssWLC, the chain should be discretized into segments much shorter than the persistence length, so increasing \( l_p \) allows for coarser discretization of the chain contour and fewer degrees of freedom for the simulation. We will discuss a systematic procedure for selecting effective parameters for a discretized ssWLC in a subsequent manuscript.

For fixed values of \( l_p \) and \( \alpha \), we solve for the remaining parameters to match \( C_2^{(1)}, C_2^{(0)} \) and \( C_4^{(1)} \), yielding the solutions

\[ e_\ell^{-1} = \frac{1}{l_p^2 (12\alpha - 4)} \left( 6C_2^{(0)} (3\alpha + 1) + 4(1 - \alpha) l_p^2 C_2^{(1)} \right) + l_p [144C_2^{(0)}]^2 + 12l_p^2 C_2^{(0)} C_4^{(1)} (15\alpha^2 - 6\alpha - 5) + 64l_p^2 \alpha^2 (C_2^{(1)})^2 + 10(1 - 2\alpha - 3\alpha^2) C_4^{(1)}/2 \]
\[ \varepsilon_{\parallel}^{-1} = -2\varepsilon_{\perp}^{-1} + 3 \left( C_2^{(0)} + \frac{C_2^{(1)}}{L_p} \right) \]  
\[ \gamma = \frac{4ae_1 L_p - 6C_2^{(0)}}{4L_p^2(1 + \alpha)} - \frac{ae_2}{L_p(1 + \alpha)} \]

The above three equations constitute a simple, systematic procedure for mapping a polymer model of interest onto an effective ssWLC model, with \( L_p \) setting the length scale of accuracy. As this length scale becomes very large (\( L_p \rightarrow \infty \)), we reach the limiting behavior \( \gamma \rightarrow 0, \varepsilon_{\perp}^{-1} \rightarrow C_1^{(1)}, \varepsilon_{\parallel}^{-1} \rightarrow C_2^{(1)} \). In this limit, the local trajectory of the chain position \( \bar{r} \) becomes decoupled from the orientation vector \( \bar{u} \). Path integration over all values of \( \bar{u}(s) \) results in the position statistics dictated by the parameters \( (\gamma, \varepsilon_{\perp}, \varepsilon_{\parallel}) \) analogous to those of the Gaussian chain. We note here that the orientation vector \( \bar{u}(s) \) is not identical to the tangent vector of the original chain being coarse-grained. Instead the \( \bar{u} \) provides an indication of the anisotropy inherent in the chain distribution under the ssWLC model. By forcing a constant term in the lowest-order moment \( (R^2) \) to match that of the original chain, we maintain an initial rigidity in the chain at shorter lengths, in contrast to a purely Gaussian chain which is flexible on all length scales. As the coarse-graining scale increases, these orientation vectors are rigidified (increasing \( L_p \)) in order to maintain this initial correlation. The length scale of accuracy increases concomitantly, so that the ssWLC provides an accurate description of the detailed polymer statistics only for chain lengths much larger than \( L_p \).

As an example of this systematic zooming-out procedure, we coarse-grain a plain, inextensible wormlike chain \(^{23} \) by mapping to a ssWLC at increasingly longer length scales. The low-order moments for a plain WLC with persistence length \( L_p^{(0)} = 1 \) can be found from eq 16–17 by setting \( \gamma = 1, \varepsilon_{\perp}^{-1} = \varepsilon_{\parallel}^{-1} = \eta = 0 \). In Figure 3, we plot the parameters of the effective ssWLC as a function of the coarse-graining scale \( L_p \). In the limit \( L_p \rightarrow L_p^{(0)} \), the plain WLC chain limit is recovered. We note that the shear modulus decreases toward the Gaussian chain value faster than the stretch modulus, implying that the shearing degree of freedom predominates in the short to intermediate length-scale behavior of the WLC. This result is in agreement with analytic calculations, which show spreading over the end-to-end vector away from a fixed initial orientation at much shorter lengths than compression along the orientation vector (see Figure 5 in ref 37).

The coarse-graining procedure described here defines a continuous flow from the plain WLC to the Gaussian chain. To demonstrate this progression more explicitly, we calculate the structure factor \( S(k) \) — the Fourier transform of the density correlations for segments along the chain. At small values of the wavevector \( k \) this quantity describes long length scale behavior, while larger values of \( k \) correspond to shorter length scales. Thus, the structure factor addresses polymer statistics across the full range of length scales.

Figure 4 compares the structure factor of the effective ssWLC to those of the plain WLC and the Gaussian chain, illustrating the transition between the two limits as the coarse-graining parameter \( L_p \) is varied. At small values of \( k \) the models fall on a universal curve, as all chains map to the Gaussian at sufficiently long length-scales. Larger values of \( k \) show a divergence between the behavior of the WLC and the Gaussian, due to the different local structures of these chains. The length scale of accuracy for the ssWLC model can be approximated by considering the maximal value of \( k \) at which the structure factor closely matches that of the original WLC. As shown in Figure 4, this \( k \) value decreases with increasing \( L_p \). Thus, mapping to a ssWLC with a larger value of \( L_p \) is accurate only at longer length scales, validating the use of this parameter as a slider to determine the degree of coarse-graining.

### 3. Coarse-graining a WLC with Kinks

We employ our coarse-graining procedure to explore the statistics of a wormlike chain with twist and periodic fixed kinks. This polymer is of particular interest for studying DNA-protein arrays. Above length scales of about ten basepairs, DNA is known to behave as a wormlike chain.\(^3\) Many DNA-binding proteins locally bend or kink the substrate DNA upon binding, and the elastic polymer properties of DNA have been shown to impact the cooperative binding of such proteins.\(^3\) The formation of long arrays consisting of many proteins binding and kinking the DNA is particularly relevant to studying chromatin condensation, where the first step in the packaging hierarchy consists of wrapping DNA around histone proteins to

---

**Figure 3.** Parameters from mapping a plain inextensible wormlike chain with persistence length \( L_p^{(0)} = 1 \) to an effective ssWLC, plotted as a function of the coarse-graining scale \( L_p \). The effective bending modulus is given by \( \varepsilon_{\parallel} = 2L_p \) and the bend-shear coupling by \( (\eta = (\varepsilon_{\perp}/\varepsilon_{\parallel})^{1/2}) \).

**Figure 4.** Structure factor for increasingly coarse-grained polymer models, ranging from the plain inextensible wormlike chain (top solid line), to the Gaussian chain (bottom dashed line). Intermediate curves correspond to effective ssWLC models with \( L_p = 1.2, 2, 4, 10 \), from top to bottom. Inset: value of the wavevector \( k \) at which the structure factor matches that of the plain wormlike chain to within an error of 10%.
form periodically spaced nucleosomes. Previous work has shown that the local geometry and linker spacing of the kinks induced by nucleosomes play an important role in determining the structure and elastic properties of nucleosome arrays and compact chromatin fibers.

The biological relevance of modeling locally perturbed DNA structures has motivated a number of computational and theoretical studies of the statistics of kinked DNA. Analytical results have been developed for a “kinkable WLC” comprising an elastic polymer with spontaneously formed, randomly inserted, fully flexible kinks. The mean square end-to-end distance for a chain with rigid kinks held at a fixed dihedral angle without twist fluctuations has also been calculated. More recently, a general form for the end distribution of elastic chains with kinks of limited flexibility was derived by Zhou and Chirikjian, though no explicit expressions for a twisted chain are separated by linker chains of length $l$.

Figure 5. (a) Schematic of the kinked WLC model. $\Omega$ is the chain orientation at each point along the contour. $\Omega_0$ is the rotation matrix defining the kink. $G_k$ is the propagator across a single chain unit. Kinks are separated by linker chains of length $L_n$. (b) Ground-state structure of the kinked WLC forms a helix. (c) Snapshot of a kinked WLC from simulations incorporating thermal fluctuations away from the ground-state helix (Figure 5c).

Low-Order Moments for the Kinked WLC. To find an effective ssWLC matching to the kinked WLC at long length scales, we need to calculate the terms $C_3^{(1)}$, $C_3^{(0)}$, $C_4^{(0)}$ for the low-order moments of the detailed polymer model. To calculate these quantities, we begin by finding the local propagator $[G_k(\vec{R}, \Omega_0, \Omega_0)]$, corresponding to the distribution function for position and orientation at the end of one repeating unit $(\vec{R}, \Omega_0)$, relative to the position and orientation at the start of the unit $(\vec{R}_0 = 0, \Omega_0)$. As before, we carry out a Fourier transform of the position $(\vec{R} \rightarrow \vec{k})$ and expand the propagator in terms of Wigner-functions $D_l^m$, a generalization of the spherical harmonics for the full three-angle rigid-body rotation. The transformed propagator can then be expressed as

$$\hat{G}_k = \sum_{l_0, j_0, m_0} B_{l_0 j_0}^{m_0}(\vec{k}) D_{l_0}^{m_0}(\Omega_0) D_{j_0}^{m_0}(\Omega)$$

The elements of matrix $B(\vec{k})$ in the above expansion have been derived in previous work by a convolution in position and rotation space of the propagator for the elastic linker polymer and that of the rigid kink. For completeness and consistency of notation, we include the expressions for $B$ and its derivatives in the Appendix. This matrix has a structure analogous to the matrix $A$ in eq 8, but where each element indexed by $l_0$ is replaced with a $(2l + 1) \times (2l + 1)$ matrix corresponding to the $j$ indices. Thus, $B(k = 0)$ is block diagonal, $(\partial B/\partial k)_{l_0,0}$ is block-tridiagonal, and so forth.

The partition function $G_N$ for the entire kinked WLC, consisting of $N$ repeating units, is found from the convolution in both position and orientation space of $N$ copies of the local propagator. These convolutions turn into multiplications under the Fourier and Wigner-function transforms, giving

$$\hat{G}_N = \sum_{l_0, j_0, m_0} [B_{l_0 j_0}^{m_0}(\vec{k}) D_{l_0}^{m_0}(\Omega_0) D_{j_0}^{m_0}(\Omega)]$$

Calculation of the low order moments, $\langle R_N^2 \rangle$, for the kinked WLC thus reduces to finding derivatives of $B^N$ with respect to $k$. This can be accomplished using the same stone-fence diagrams described in the preceding section, with the addition of triple and quadruple steps. For instance, $\langle R_k^2 \rangle$ is given by

$$\langle R_k^2 \rangle = -\frac{\partial^2 [B^N]_{00}}{\partial k^2}$$

where all the matrices are evaluated at $k = 0$. Here the submatrix $B_{l_0}$ is nonzero only for $l = l'$, corresponding to points in the stone-fence diagrams. $B_{l_0}$ is nonzero for $l = l \pm 1$.

dx.doi.org/10.1021/ma302056v | Macromolecules XXXX, XXX, XXX--XXX
corresponding to single steps in the diagrams, and $B_\ell$ is nonzero for $\ell = \ell' \pm 2$ corresponding to double steps. The two terms in eq 24 thus correspond to the two diagrams in Figure 2a.

Performing the eigendecomposition of $B_\ell$ yields

$$[B_\ell]^n = \sum_{j=1}^{2l+1} \lambda_j^{(j)n} U_j^{(j)}$$  \hspace{1cm} (25)

where $\lambda_j^{(j)}$ are the eigenvalues and $U_j^{(j)}$ is the dyadic product of the $j^{th}$ right and left eigenvectors.

The discrete convolutions in eq 24 can then be carried out analytically to extract the linear and constant terms of the lowest moment as

$$L_\epsilon C_2^{(1)} = -B_{00}^{(0)} - \sum_{j=1}^{3} \frac{2}{1 - \lambda_j^{(0)}} B_{0j}^{(0)} U_j^{(0)} B_{10}^{(0)}$$ \hspace{1cm} (26)

$$C_2^{(0)} = \sum_{j=1}^{3} \frac{2}{(1 - \lambda_j^{(0)})^2} B_{0j}^{(0)} U_j^{(0)} B_{10}^{(0)}$$ \hspace{1cm} (27)

The analogous calculations for the next highest moment, $\langle R^4 \rangle$, can be found in the Appendix. $C_2^{(1)}$ corresponds to 1/3 of the Kuhn length that can be used to approximate the kinked WLC as a Gaussian chain at very long lengths. The additional components $C_2^{(0)}$, $C_4^{(1)}$ allow us to find an effective ssWLC model as described in section 2.

**RESULTS**

We perform the coarse-graining procedure described in the previous sections for the kinked WLC with kinks $\Omega_h$ defined by Euler angles $(\pi/4, \pi/4, \pi/4)$ separated by varying lengths $l_p$ of a semiflexible polymer with all physical parameters set to match those of DNA. That is, the linker polymer chains have bend persistence length $l_p^{(0)} = 50$ nm, twist persistence length $l_t = 100$ nm, and natural twist $\tau = 2\pi/(10.5 \text{ bp} \times 0.34 \text{ nm}/ \text{ bp})$. Because of the nonzero twist $\tau$, changing the linker length $L_u$ alters the relative alignment of the kinks and results in very different ground-state structures for the polymer.

Figure 6 illustrates the parameters of the effective ssWLC as a function of the linker length, for a coarse-graining scale of $l_p = 70$ nm. As expected, kinked chains that are naturally more extended have larger values of $\chi$, the relaxed extension of the effective shearable chain. The parameter $\gamma$ is always less than the ground-state helix height $h$ per contour length because the coarse-grain model includes contraction of the overall polymer size due to thermal fluctuations.

More extended chains have higher values of the stretch modulus $\epsilon_{\parallel}$ and lower values of the shear modulus $\epsilon_{\perp}$. Although the semiflexible polymer connecting the kinks is itself inextensible, a compact helical spring created by the kinks (such as the 49 bp structure in Figure 6c) can be easily stretched by exploiting the bend and twist flexibilities of the polymer. By contrast, more extended chains (such as the 45 bp structure in Figure 6c) can be easily sheared by displacing the ends perpendicular to the helix via bending of the linker polymers. We note that for extremely compact kinked chains there is no solution for the effective ssWLC with the selected value of the coarse-graining parameter $l_p$. While solutions exist at higher coarse-graining levels, they require a negative value of $\gamma$ (or, equivalently, a negative value of $\eta$). Such points are left

**Figure 6.** Mapping of kinked wormlike chains with different linker lengths onto an effective ssWLC model. The coarse-grained ssWLC has fixed persistence length $l_p = 70$ nm and bending modulus $\epsilon_{\parallel} = 140$ nm. (a) Dashed line gives the height along the helix axis ($h$) per unit contour length of the ground-state helix. Solid line plots the natural stretch ($\gamma$) for the effective coarse-grained model. (b) Inverse shear ($\epsilon_{\perp}^{-1}$) and stretch ($\epsilon_{\parallel}^{-1}$) moduli for the effective shearable chain as a function of linker length. (c) Ground-state structures for three different kinked chains corresponding to the marked points in parts a and b.

**Figure 7.** Distribution of the end-to-end distance from simulations of the kinked WLC, compared to effective coarse-grained models. Blue circles: distributions for a kinked WLC with 15 kinks, separated by contour lengths of $L_u = 40,45,49$ bp. Solid green lines: analytical results for distribution of the effective ssWLC, with $l_p = 70$ nm and parameters $\gamma$, $\epsilon_{\parallel}$, $\epsilon_{\perp}$, $\eta$ as labeled. Dashed red lines: Gaussian chain distributions, with effective Kuhn length $\kappa$. 
out of Figure 6. We do not attempt to address these extreme cases, since for such compact structures the statistics of the polymer will be dominated by steric interactions between different coils of the chain, an effect that is not included in our model.

The accuracy of the coarse-grained model is illustrated in Figure 7, which shows the distribution of the end-to-end distances obtained by direct simulation of the kinked WLC, compared to the analytical results for the ssWLC with parameters given in Figure 6. Simulations were carried out by discretizing the chains into segments much shorter than both the persistence length and the twist $1/\tau$ and sampling from the multivariate normal distribution to obtain bend and twist perturbations for each segment in the kinked WLC. The simulated chains consisted of 15 kinks with linker lengths 40, 45, and 49 bp (where 1 bp =0.34 nm). The ground-state structures of the kinked chains are shown in Figure 6c. The total chain lengths simulated ($4l^{(0)}_p, 4S^{(0)}_p, S^{(0)}_p$ for the three different linker lengths) are in the intermediate regime where Gaussian statistics do not accurately reproduce the behavior of the detailed kinked chain. We see that in this regime the effective ssWLC model provides a greatly improved approximation to the chain statistics.

The greatest deviation from the kinked chain occurs for the most extended kinked array, with 45bp linkers. Here, the finite extensibility of the linker chain, which is not reproduced by the ssWLC model, has the greatest effect on the end-to-end distribution, as evidenced by the extended tail of the ssWLC distribution at large $R$. A finite length constraint could, in principle be implemented in discretized simulations based on the ssWLC model. Nonetheless, even in this special case the shearable chain provides a significantly better model of the kinked chain statistics than does the Gaussian.

These results demonstrate that the coarse-graining procedure described here accurately reproduces the end-to-end statistics of the kinked WLC model at lengths much shorter than those that allow the polymer to be treated as a Gaussian chain. We note that although the kinked DNA model is based on a twisted polymer, with the twist degrees of freedom significantly impacting the overall structure of the kinked chain, it is nonetheless possible to map this detailed polymer onto our ssWLC model, which lacks twist. This example highlights the generality of the mapping procedure described in this manuscript. All locally defined polymer chains can be approximated as a ssWLC on intermediate to long length scales.

4. CONCLUSIONS

In this work, we present a fully general analytical method for coarse-graining polymer models by mapping to an effective shearable wormlike chain. Our approach allows for systematic zooming in on a polymer model by decreasing the coarse-graining parameter $l_p$. For large values of $l_p$, this procedure is equivalent to approximating the polymer as a Gaussian chain, whereas smaller values more accurately model the elastic properties of the polymer at intermediate length scales, including its resistance to stretching, bending, and shearing away from the structure of lowest free energy. For a continuous semiflexible chain, this coarse-graining approach defines a continuous flow from the original chain model to the Gaussian chain. For more complicated polymer models, the smallest coarse-graining scale at which an effective mapping can be found is limited to those values of $l_p$ where the effective $\epsilon_\perp, \epsilon_\parallel$ are positive real numbers.

By mapping a locally complicated polymer model onto an effective ssWLC, we gain both the ability to perform more coarse-grained simulations and the ability to analytically calculate certain aspects of the chain statistics. Effective models with larger values of $l_p$ can be discretized into correspondingly larger segments, allowing for more efficient simulations. In a subsequent publication,\(^{36}\) we address the topic of how best to map a polymer onto a discrete shearable wormlike chain for arbitrarily large discretization. The systematic reduction of simulation degrees of freedom is key to in silico studies of large-scale polymer systems such as genomic DNA, cytoskeletal filament networks, or entangled polymer meshes.

The structure factor and the end-to-end distance distribution for the ssWLC can be calculated exactly using the methods described in section 2. The joint distribution of chain positions and orientations is also analytically tractable via an analogous procedure. Thus, aspects of the chain statistics such as the looping probability with or without fixed end orientations can be approximated for the polymer chain on intermediate to long length scales.

As an example of our coarse-graining procedure we map a kinked WLC, consisting of an inextensible, twisted, semiflexible polymer with periodic kinks onto an effective shearable wormlike chain. We show that this mapping yields a significant improvement over the Gaussian chain in reproducing chain statistics at intermediate lengths. This example can be applied to studying the role of DNA-bending proteins in modulating the large-scale structure of genomic DNA. Future work will include examining the looping behavior of DNA–protein arrays, including the nucleosome arrays that form the lowest level of hierarchical chromatin packing, across a large span of length scales.

Several recent studies have raised concerns regarding the applicability of elastic chain models in studying realistic DNA structures at short lengths.\(^{39–51}\) However, we demonstrate in this work the importance of the elastic chain as a general, unifying model for describing long-length polymer behavior. Local effects, such as the binding of kinking proteins, can be subsumed into the effective parameters of the coarse-grained ssWLC model. We therefore emphasize that the influx of new measurements regarding the behavior of short DNA segments should not lead to complete abandonment of elastic chain models. Instead, information on local polymer structure can be incorporated into the framework of an elastic model for describing chain statistics at long lengths. While there are situations where physical constraints are such that small length scale effects must be acknowledged, coarse-grained elastic chains such as the ssWLC remain critical to bridging from locally detailed models to large-scale behavior.

We emphasize that the method presented here is highly general and can be applied to modeling any locally defined polymer on a sufficiently large length scale. However, nonlocal interactions cannot be incorporated directly into the model. Steric exclusion effects, for instance, can be included in the local but not the global sense. That is, steric interactions between nearby segments along the chain contour can be incorporated into the effective elastic parameters of the model. Depending on the structure of the underlying self-avoiding polymer, there can then be an intermediate length scale where the ssWLC is applicable but the inherent rigidity of the chain makes distal
stereic effects negligible. In some cases, such as the highly compact helical structures seen for the kinked WLC, stereic interactions between distal segments will dominate the structure and the ssWLC model described here cannot be used. Similarly, other nonlocal effects such as topological entanglements, global constraints on twist and writhe, and hydrodynamic interactions are not currently included in the model. However, these global effects could be incorporated into large-scale simulations using the effective ssWLC as a starting point for the polymer model. In cases where the chain twist around its axis as well as the overall contour path is of interest, the ssWLC model can be expanded to include the twist degree of freedom and the concomitant coupling between twist and position. The optimal procedure for handling these aspects of coarse-grained systems is an area deserving future study.

The methodology described in this work is designed to serve as a springboard for the development of coarse-grained polymer models, an indispensable task for researchers interested in studying macromolecular properties on macroscopic length scales.

### APPENDIX

#### Unit Propagator for Kinked WLC

Here we derive the propagator \( G_i(k, \Omega \mid \Omega_0; p) \) for the position and orientation at the end of one unit of the kinked WLC relative to the position and orientation at the start of the unit. A unit of the kinked chain consists of a wormlike chain with twist \( \gamma_{\Omega} \) of length \( L_0 \) followed by a rigid kink defined by the rotation matrix \( \Omega_k \). If we define \( \Omega^\tau \) as the orientation at the end of the linker chain prior to the kink, then the Fourier-Laplace-transformed propagator across the linker only is given by an expansion in Wigner functions,

\[
G_{ll}(k, \Omega \mid \Omega_0; p) = \sum_{l_p, j_p, m_p} g_l^l(k) D_{ll}^{m_p}(\Omega_0) D_{ll}^{m_p}(\Omega)
\]

where the coefficients \( g_l^l \) can be expressed in matrix form as \( g^l = (A^l + pI)^{-1} \)

\[
A^l_{l, l+1} = \frac{l(l+1)}{2p} + \left( \frac{1}{2l} - \frac{1}{2l_0^0} \right) l^2 + i\tau
\]

\[
A^l_{l, l-1} = -ika_l
\]

To get the propagator for the full unit of the chain, we note that the orientation following the kink is given by \( \Omega = \Omega^{\tau} \Omega_k \) and then make use of the Wigner function expansion for a composition of rotations

\[
D_{ll}^{m}(\Omega) = \frac{8\pi^2}{2l+1} \sum_n D_{ll}^{m}(\Omega_0) D_{ll}^{m}(\Omega_k)
\]

We can define a kink matrix \( M \) that is block-diagonal in the \( l \) indices

\[
M_{ll} = \frac{8\pi^2}{2l+1} D_{ll}^{m}(\Omega_k)
\]

and treat the linker propagator matrix \( g \) as diagonal in the \( j \) indices and block-tridiagonal in the \( l \) indices. The final expression for the unit propagator is then given by eq 22 with the matrix of coefficients,

\[
B = g M
\]

The derivatives of \( g \) with respect to \( k \) can be found using stone-fence diagrams with only single-order steps allowed. The diagrams for the first 4 derivatives are shown in Figure 8. The inverse Laplace transform \( p \to L_0 \) can be easily carried out for each diagram by summing over the poles. We note that since \( A \) is a symmetric matrix for this system, all \( g \) derivatives are symmetric as well. The derivatives of \( B \) are then found by right-multiplying the derivatives of \( g \) by \( M \).

#### Low Order Moments for Kinked WLC with Many Kinks

The first two nonzero moments for the end-to-end distance of a kinked WLC with \( N \) kinks can be expressed as,

\[
\langle R^2_k \rangle = C_1^{(1)}(L_0 N) + C_1^{(0)} + \ldots
\]

\[
\langle R^2_k \rangle = 3(C_2^{(1)})^2(L_0 N)^2 + C_4^{(1)}(L_0 N) + C_4^{(0)} + \ldots
\]

where \( R \) refers to terms that decay exponentially with \( \eta \). The terms \( C_2^{(0)}, C_4^{(1)} \) are found from eq 24–27. We perform analogous calculations to find the term \( C_4^{(1)} \) using the stone-fence diagrams in Figure 2b, with additional diagrams incorporating third and fourth derivatives of the unit propagator matrix \( B \). Discrete convolutions are carried out over the powers \( \eta \), with the total of the powers equal to the

\[
\begin{align*}
\text{Figure 8. Stone-fence diagrams for calculating } k \text{ derivatives of the} \\
g_{ll} \text{ in the Wigner-function decomposition of the} \\
\text{propagator for the twisted wormlike chain. Each point on the diagram} \\
\text{represents a factor of } (1/(p + A_j)), \text{ and each step represents a factor of} \\
(\partial A_{ll+1}/\partial k), \text{ evaluated at } k = 0.
\end{align*}
\]

\[
\begin{align*}
\text{Figure 9. Additional stone-fence diagrams for calculating } \langle R^2 \rangle \\
\text{for the} \\
kinked \text{ WLC. Together with the diagrams in Figure 2b these represent} \\
\text{all the terms in eq S9.}
\end{align*}
\]
number of kinks $N$ minus the number of steps in the diagram. Each step in the diagram corresponds to a term of $\mathcal{O}(B_T^m / \delta u^m)$, with the order of the diagram equal to the order of the step. The coefficient for each diagram is $4! / (2^m 3^{m-1} m!)$, where $m$ is the number of $m$-th-order steps. For example, the first diagram in Figure 2b contributes the following term:

$$
24 \sum_{n_1} \sum_{n_2} \sum_{n_3} \sum_{n_4} \frac{[B_{00} B_{11} B_{10} B_{00} B_{01} B_{10} B_{11} B_{10} B_{11} B_{10}]}{B_{00} N^{n_4-n_2-n_3-n_4-n_4}}
$$

(S8)

Using the eigendecomposition of $B$ given by eq 25, together with the normalization condition $B_{00} = 1$, we perform the discrete convolutions directly to give

$$
L_n c_4^{(1)} = 24 \sum_{i,j=1}^{3} \left[ \frac{5 - 3(\lambda_1^{(i)} + \lambda_2^{(i)}) + \lambda_1^{(i)} \lambda_2^{(i)}}{2(\lambda_1^{(i)} - 1)^2(\lambda_2^{(i)} - 1)^2} \right] B_{00} U_{10}^{(i)} B_{11} U_{10}^{(j)} + \sum_{i=1}^{3} \left[ \frac{\lambda_2^{(i)} - 3}{2(\lambda_2^{(i)} - 1)^2} \right] B_{00} U_{10}^{(i)} B_{10}^{(i)} - 3B_{00}^{(i)} B_{00}^{(j)} B_{11} U_{10}^{(j)} B_{10}^{(i)}
$$

(S9)

This provides the final term necessary for mapping the kinked WLC to a coarse-grained ssWLC model.

**AUTHOR INFORMATION**

**Corresponding Author**

*E-mail: lenafabr@stanford.edu (E.F.K.); ajspakow@stanford.edu (A.J.S.)*

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

We thank Jay D. Schieber and Ekaterina Pilyugina for helpful discussions over the course of this work. Funding was provided by the Fannie and John Hertz foundation and the National Science Foundation CAREER Award program.

**REFERENCES**