A Brownian dynamics model for the chromatin fiber

L. Ehrlich, C. Münkel, G. Chirico and J. Langowski

Abstract

Motivation: We describe a Brownian dynamics model for the folding of the chromatin fiber based on the model of Woodcock et al. (Proc Natl Acad Sci USA, 90, 9021–9025, 1993). The model takes into account the elastic properties of the DNA as well as the electrostatic interaction and nucleosomal excluded-volume interaction. The solvent is described as a viscous medium, the electrostatic interactions by a screened Coulomb potential.

Results: The hydrodynamic properties and their dependence on the solvent’s ionic strength are accurately reproduced by the model for nucleosome di- and tetramers. Ionic strength-dependent changes in mobility can be attributed to partial screening of the electrostatic repulsion between different segments of linker DNA. Formation of fiber-like structures occurs on time scales of several hundred microseconds for a linear configuration of 25 nucleosomes. The model was implemented by creating user-defined data types. Use of this so-called object-oriented paradigm allowed for a high degree of component reuse in simulation, analysis and visualization contexts.

Availability: The described software is available on request.

Contact: E-mail: Lutz.Ehrlich@embl-heidelberg.de; Joerg.Langowski@dkfz-heidelberg.de

Introduction

Although detailed modeling of static chromatin structure has been carried out during the last 20 years (for reviews, see van Holde, 1989; van Holde and Zlatanova, 1995; Woodcock and Horowitz, 1995), the dynamics of the structure formation processes are still unknown. In particular, the mechanism behind the compaction from naked DNA to whole chromosomes has been described by static models, the most prominent being the concept of the ‘30nm fiber’, where the chromatin chain is arranged in a regular solenoidal array of nucleosomes.

Recent results favor a model where the 30nm fiber consists of irregular zigzag-like stretches of nucleosomes connected by linker DNA. Woodcock et al. (1993) proposed an alternative to the solenoid model, which could be shown to result in structures resembling those found in electron tomography (Horowitz et al., 1994) or in scanning force microscopy (Yang et al., 1994). In this so-called ‘3D-zig-zag’ model, the main determinants of fiber structure are two angular parameters: \( \theta \) representing the angle between linker DNA ends protruding from the core octamer, and \( \phi \) for the relative twist of two adjacent nucleosomes. Histone H1 is assumed to be responsible for maintaining a fixed entry–exit angle, \( \theta \), independent of its exact position on the nucleosome. The second angle, \( \phi \), is determined by the length of linker DNA for the specific chromatin, as every base pair leads to an additional 36° increase due to helical twist. Both angles are part of a phenomenological description of the experimental findings; in reality, the observed angles will be subject to thermal fluctuations around their equilibrium values.

How could the process of chromatin structure formation be described by a computer model? Obviously, atomic calculations are unfeasible as rearrangement processes occur on time scales of micro- to milliseconds. Instead, a simplified polymer model based on the work of Chirico and Langowski (1994) was used to describe the features of DNA on length scales of nanometers and larger. The nucleosomes are modeled as solid beads, the linker DNA protruding from these beads as stiff cylindrical segments. This model is sensible only for segments much shorter than the DNA bending persistence length \( L_p = 50 \text{nm} \). This bending persistence length determines a length scale for the directional correlation of two polymer segments. The mean of the angle between two segments of length \( s \) apart along the contour of the polymer decays as \( \exp (-s/L_p) \) (Grosberg and Khokhlov, 1994).

The linker DNA segments are elastically connected to allow for torsion, bending and stretching motion. The electrostatic repulsion of the linker DNA phosphate backbones is described by Debye–Hückel electrostatics of uniformly charged cylinders. As there is little knowledge on the ionic strength-dependent interaction of the nucleosomes, a modified Lennard–Jones potential was used to model the excluded-volume interaction.

© Oxford University Press
System and methods

To study the Brownian motion of chromatin, a system of C++ (Stroustrup, 1993) and Objective-C (Cox and Novobilski, 1991) classes was created. These classes represent user-defined data types. A particular instance of such a class is called an object. These objects contain their own variable determining their state (instance variables). In addition, they can respond to messages asking them to perform a particular subroutine (often called method or member function).

The simulations have been implemented as a hierarchy of C++ classes, using gcc2.5.8 and gcc2.7.0 compilers under NeXTSTEP3.2, Solaris2.4 and AIX4.1. Simulations were run on HP712/80, SuperSparc10 and IBM RS/6000 workstations. Visualization was done on a HP712/80 workstation. No templated code was used.

A number of classes were used from the libg++ (version 2.0) and LAPACK++ (version 1.0/3) class libraries. Whereas libg++ provides basic and advanced data types such as lists, strings and random number generators, LAPACK++ is an object-oriented version of the LAPACK library of linear algebra subroutines. Together with an adaptation of these well-known Fortran subroutines, which let the user solve systems of linear equations, linear least squares or eigenvalue problems, they provide dedicated data types such as vectors and matrices.

Algorithm

For a realistic description of chromatin dynamics in a viscous solvent, two aspects of system-solvent interaction have to be considered. First, the system is subjected to microscopic collisions with solvent molecules, resulting in a random (Langevin) force in a macroscopic description of the Brownian motion (Chandrasekhar, 1943). Second, the viscosity of the solvent will couple the displacements of different parts of the simulated molecule (hydrodynamic coupling). These observations led to the development of a Brownian dynamics algorithm with detailed hydrodynamic interactions by Ermak and McCammon (1978). In this work, the stiff cylindrical segments representing linker DNA are used to compute the elastic and electrostatic force contributions.

Geometrical parameters

Segment geometry. Chromatin is modeled as an array of cylindrical segments and solid spheres attached to them. The cylinders represent linker DNA between nucleosomes; they are connected at joint positions \( r_i \). In modeling terms, each of these segments can either be of type 'DNASegment' (linker DNA without a chromatosome) or 'NucleosomeSegment' (linker DNA with a chromatosome). These segment types are different with respect to their hydrodynamic and excluded-volume interactions (see below); for the calculations involved, they are replaced by solid spheres of different radii. A nucleosome consisting of two such segments is shown in Figure 1. Here, segment \( i \) is of type 'DNASegment', whereas segment \( (i + 1) \) is of type 'NucleosomeSegment'. In the following, the term 'segment' will generically refer to both types, unless otherwise stated.

Fig. 1. Sketch of the chromatin model. At the joint of the segments \( i \) and \( (i + 1) \), the local coordinate system \( S_{i+1} \) and a nucleosomal sphere are depicted. The nucleosomal sphere's center is positioned at a distance of \( \sigma_{\text{Nuc}} \) from the segments' joint at the position \( r_i \). \( b_i \) and \( b_{i+1} \) denote the current segment lengths. The hydrodynamic spheres representing the linker DNA are positioned at \( r_i \) and \( r_{i+2} \); their radii are given by \( \sigma_i \) and \( \sigma_{i+2} \), respectively. These beads are the additional beads for the 'end segments' as described in the text.
For hydrodynamic purposes, segments of type `DNASegment` are represented by spherical `beads` with a given radius \( \sigma_i \), at the joint between segment \((i - 1)\) and \(i\). The first and last segments in an open configuration have additional beads attached to their very ends, respectively (cf. the beads at the start of segment \(i\) and at the end of segment \((i + 1)\) in Figure 1).

To determine the bead radius \( \sigma_i \) for a given chromatin configuration, the diffusion coefficient for a solid cylinder is computed first. The cylinder represents a double strand of DNA with a radius of 1.2 nm, whereas its length is chosen equal to the contour length of the linker DNA simulated. The computation of the diffusion coefficient follows the formalism described by Garcia de la Torre and Bloomfield (1981). The resulting diffusion coefficient is compared to that of a rigid assembly of \(N\) solid beads of radius \(\sigma_o\), \(N\) being the number of segments simulated for the particular configuration. These beads are each spaced one equilibrium segment length apart. The diffusion coefficient for such an assembly is given in Garcia de la Torre and Bloomfield (1981). Finding the bead radius for which the diffusion coefficients become equal is then formulated as finding the root of a function \(F(\sigma) = 0\). This root is then determined by an iterative procedure.

A segment of type `NucleosomeSegment` is represented by a larger bead of radius \(\sigma_{Nuc}\) whose surface coincides with the connection of two cylinders representing linker DNA protruding from the chromatosome. This larger bead is centered at the bisector of the angle formed between these linker DNA ends (see Figure 1).

A complete chromatin configuration is thus built by catenating nucleosomes depicted in Figure 1, connecting the linker DNA ends. The nucleosomes are the source of an additional excluded-volume force which is computed from the core positions \( r_{Nuc}^i \) (see below). The nucleosomal bead radius \(\sigma_{Nuc}\) is chosen in such a way that the resulting bead’s volume equals that of an entire core octamer plus the DNA wound around it. The volume of this DNA is approximated by the sum of the volumes of the hydrodynamic beads which would be necessary to represent it as linker DNA. The resulting nucleosomal bead radius \(\sigma_{Nuc}\) is then used in computing the hydrodynamic interaction tensor (see below).

The nucleosomal bead position \(\vec{r}_i^{Nuc}\) is recomputed at each time step of the simulation by:

\[
\vec{r}_i^{Nuc} = \vec{r}_i + \sigma_{Nuc} \left[ \frac{\vec{u}_{i-1} - \vec{u}_i}{|\vec{u}_{i-1} - \vec{r}_i|} \right]
\]

where \(\vec{u}_i = \vec{r}_{i+1} - \vec{r}_i\) and \(\vec{u}_i = \vec{u}_i/|\vec{u}_i|\). Implicitly, this means that nucleosomal DNA does two full turns around the histone core without having any superhelical pitch upon leaving the core. This means that the linker DNA enters and leaves and nucleosome bead at the same point.

**Local coordinate systems.** The linker DNA’s elastic degrees of freedom can be described by attaching a local coordinate system \(S_i\) to each segment \(i\). The origin of the coordinate \(S_i\) system is placed at the joint between segment \(i\) and \((i - 1)\). Each coordinate system’s \(z\)-axis is given by \(\vec{u}_i\), a unit vector along the segment. A second unit vector perpendicular to \(\vec{u}_i\), \(\vec{f}_i\), along with a third, \(\vec{v}_i = \vec{u}_i \times \vec{f}_i\), constitutes \(S_i\). Two adjacent coordinate systems \(S_i\) and \(S_{i+1}\) can then be related via a rotation parameterized by three Euler angles \((\xi_{1i}, \xi_{2i}, \xi_{3i})\) needed to rotate \(S_i\) into \(S_{i+1}\). These Euler angles are used to describe the bending and torsion of linker DNA. The first angle, \(\xi_{1i}\), denotes a rotation around the unit vector \(\vec{u}_i\). The second Euler angle, \(\xi_{2i}\), rotates \(\vec{u}_i\) into \(\vec{u}_{i+1}\) around the new \(\vec{f}_i\), whereas the last rotation, \(\xi_{3i}\), is executed around the axis \(\vec{u}_{i+1}\). From this it can be seen that \(\xi_{2i}\) describes the angle between segment \((i)\) and \((i + 1)\), and \(\Phi_i = \xi_{1i} + \xi_{3i}\) is the relative torsion between them. At the start of a simulation, the position and relative orientation of all local coordinate systems are set to initial values by specifying the origin coordinates \(\vec{r}_i\) and the relative twists \(\Phi_i\) between coordinate systems \(S_i\) and \(S_{i+1}\) for all \(i\). The specification is completed by setting the vector \(\vec{f}_0\) for the first segment to an arbitrary unit vector [usually \(\vec{f}_0 = (0,0,1)\)].

**Bending angles.** To describe the bending of the linker DNA, a new angle variable \(\nu_i\) is introduced. It is given as the angle between \(\vec{u}_{i+1}\) and a unit vector \(\vec{B}_i\). This vector represents the elastic equilibrium orientation of segment \((i + 1)\) with respect to segment \((i)\). It is uniquely defined in the coordinate system \(S_i\) by two angles \((\theta_i, \phi_i)\). In the case of free DNA, these angles are set to \((\theta_i, \phi_i) = (0, 0)\). In the case of nucleosomes, they are used to model the protruding linker DNA. A non-zero value for \(\theta_i\) denotes a permanent bend between segments \((i)\) and \((i + 1)\) as the elastic equilibrium. The relative twist between nucleosomes is described by the choice of \(\phi_i\). As the addition of a single base pair of linker DNA gives rise to an additional twist of 36° between nucleosomes, the angles \(\phi_i\) can only be set to values of the set \([0°, 36°, ..., 324°]\).

The angles \((\theta_i, \phi_i)\) are independent of the Euler angles \((\xi_{1i}, \xi_{2i}, \xi_{3i})\) connecting the local coordinate systems. The input parameters \((\theta_i, \phi_i)\) are kept constant during simulations, whereas \((\xi_{1i}, \xi_{2i}, \xi_{3i})\) and \(\nu_i\) vary according to the dynamics of the configuration studied. These quantities can be derived from the segments joint coordinates and the relative twists of the coordinate systems \(S_i\).

**Brownian dynamics.**

The discretized Langevin equation for the evolution of the local coordinate system origins \(\vec{r}_i\) is (cf. Chirico and Langowski, 1994):

\[
\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \sum_{j=1}^{N} \mathbf{D}_j \vec{r}_j \frac{\Delta t}{k_B T} + \vec{R}_i(t)
\]
Here, $k_B T$ represents Boltzmann’s constant multiplied by the absolute temperature, $\delta t$ is the time step, $\vec{F}_j$ is the total force acting on bead $j$ and $\vec{D}_j$ is the hydrodynamic tensor coupling the displacements of beads $i$ and $j$. For equal bead radii, an approximate expression has been given by Rotne and Prager (1969).

The random displacements $\vec{R}'_i(t)$ have to fulfill:

$$\langle \vec{R}'_i(t) \rangle = 0 \quad \text{and} \quad \langle \vec{R}'_i(t) \otimes \vec{R}'_i(t) \rangle = 2D_y \delta t$$ \hfill (3)

The generation of these random displacements with a given covariance matrix $2D \delta t$ involves a Cholesky decomposition and is described in Ermak and McCammon (1978).

The relative torsion between two segments, $\Phi_i$, is the sum of the first and last Euler angles $\xi_{1i}$ and $\xi_{1i}$ needed to rotate coordinate system $S_i$ into $S_{i+1}$. Its time evolution is governed by:

$$\Phi_i(t + \delta t) = \Phi_i(t) + D_{r1} \frac{\delta t}{k_B T} + \delta(t)$$ \hfill (4)

where $D_r$ is the rotational diffusion coefficient, $T_r$ is the torque acting on segment $i$ and $f_i(t)$ is a random torsion, subject to the conditions ($\delta_i$ denotes the Kronecker delta):

$$\langle f_i(t) \rangle = 0 \quad \text{and} \quad \langle f_i(t) f_j(t) \rangle = 2D_r \delta_{ij} \delta t$$ \hfill (5)

Equations (2) and (4) are sufficient to compute the bead positions and orientations for each segmental coordinate system (cf. Chirico and Langowski, 1994).

The potentials for the forces as well as the hydrodynamic tensor $D_y$ are described in the following section.

**Interactions**

**Stretching.** Stretching of the linker DNA along the contour of the double strand is described by a harmonic potential:

$$U^{(0)}(b_i) = \alpha^{(0)} \frac{k_B T}{2} (b_i - b_0)^2$$ \hfill (6)

with $b_i$ the length of the cylindrical segment ($b_0$ denoting its equilibrium value). $\alpha^{(0)}$ is a parameter determining the strength of the potential. It is computed to be $\alpha^{(0)} = (bb_0)^{-2}$; a typical value is $\alpha^{(0)} = 238 \text{nm}^2$ for a linker segment length of 8.1 nm (24 bp) and $\delta = 0.008$. The particular choice of $\delta$ was made in accordance with simulations of supercoiled DNA (Chirico and Langowski, 1994). Recent experimental results (Smith et al., 1996) suggest a value of the same order of magnitude, namely $\delta = 0.028$.

**Bending.** Bending of adjacent linker DNA segments is described by:

$$U^{(b)}(\theta_i) = \alpha^{(b)} \frac{k_B T}{2} \theta_i^2$$ \hfill (7)

The parameter $\alpha^{(b)}$ is related to the bending persistence length $L_p \approx 50 \text{nm}$ of free DNA via $\alpha^{(b)} = \frac{L_p}{2} \sigma_{ij}$ (Grosberg and Khoklov, 1994). In our model, we choose the flexibility of the permanent bends used for the entry/exit angle of linker DNA protruding from the histone octamer to have the same value. A typical value for $\alpha^{(b)}$ is 5.98.

**Torsion.** The torsion interaction energy is modeled by a harmonic potential:

$$U^{(\Phi)}(\Phi_i) = \alpha^{(\Phi)} \frac{k_B T}{2} (\Phi_i)^2$$ \hfill (8)

The potential parameter is determined by $\alpha^{(\Phi)} = C/k_B T b_0$, with $C$ being the torsional rigidity of DNA. In the literature, values in the range of $C = (1.4 - 3.0) \times 10^{-28} \text{ J m}$ are given (Allison et al., 1989); recent studies converge on a value of $C = 2.6 \times 10^{-28} \text{ J m}$ (Fujimoto and Schurr, 1990; Schurr et al., 1992). In the simulations, $C = 2.0 \times 10^{-28} \text{ J m}$ was used, resulting in $\alpha^{(\Phi)} = 5.99$ for $b_0 = 8.1 \text{ nm}$ and $T = 25^\circ \text{C}$.

**Electrostatic interaction of linker DNA.** The electrostatic repulsion of the linker DNA backbones is partially shielded by the monovalent cations in the solvent, resulting in a Debye–Hückel electrostatic interaction on a length scale given by the Debye length $1/\kappa$, where $\kappa^2 = 8\pi^2 k_B T D / e^2$ (Tanford, 1961). Here, $I$ is the ionic strength, $e$ the proton charge and $k_B$ the Boltzmann constant. $T$ denotes the absolute temperature and $D$ is the dielectric constant of water at that given temperature. For physiological salt conditions, $1/\kappa \approx 1 \text{ nm}$. The electrostatic energy due to the interaction of segment $i$ and $j$ is:

$$U^{(e)}(\Phi_i) = \alpha^{(e)} \frac{k_B T}{2} \int_0^{\Phi_i} \int_0^{\Phi_i} e^{-\kappa \mid \Phi_i-\Phi_j \mid} d\Phi_i d\Phi_j$$ \hfill (9)

$\lambda_i$ and $\lambda_j$ being integration parameters. The vector $\vec{d}_{ij} = \vec{d}_{ij} (\lambda_i, \lambda_j)$ parameterizes the connecting line between the two segments $i$ and $j$. Its explicit form is given by:

$$\vec{d}_{ij}(\lambda_i, \lambda_j) = \lambda_j \left( \begin{array}{c} \sqrt{1 - \gamma_{ij}}^2 \gamma_{ij}^2 + \frac{\sigma_{ij}}{b_0 \gamma_{ij}} \\ 0 \\ -\gamma_{ij} \end{array} \right)$$ \hfill (10)

The parameters $(\gamma_{ij}, \sigma_{ij}, \gamma_{ij}, \sigma_{ij})$ describe the relative position of two segments $i$ and $j$ in space. To compute them from the bead positions, the ‘shifted midpoints’ of the segments $i$ and $j$, $L_i^{+} = 50 \text{nm}$ of free DNA via $\alpha^{(b)} = \frac{L_p}{2} \sigma_{ij}$ (Grosberg and Khoklov, 1994). In our model, we choose the flexibility of the permanent bends used for the entry/exit angle of linker DNA protruding from the histone octamer to have the same value. A typical value for $\alpha^{(b)}$ is 5.98.
Brownian dynamics model for chromatin fiber

\[ \vec{r}_t^* \text{ and } \vec{r}_t^\prime \text{ are needed. With } \vec{r}_t^{(m)} = (\vec{r}_{t+1} + \vec{r}_t)/2 \text{ as the midpoint of segment } i, \text{ they are given by:} \]

\[ \vec{r}_t = \begin{cases} \vec{r}_t + \frac{1}{2} \vec{u}_t \quad \text{if } |\vec{r}_t^{(m)} - \vec{r}_t| \leq |\vec{r}_t^{(m)} - \vec{r}_{t+1}| \\ \vec{r}_{t+1} - \frac{1}{2} \vec{u}_t \quad \text{else} \end{cases} \quad (11) \]

The four parameters can then be extracted from the bead positions by using:

\[ \varepsilon_j = |\vec{v}_j|, \quad \tilde{\varepsilon}_j = \frac{1}{b_0}(\vec{r}_t^* - \vec{r}_t^\prime) \quad (12) \]

\[ \gamma_{ij} = \frac{\vec{u}_i \times \vec{v}_j}{|\vec{v}_j|} \quad (13) \]

\[ \gamma_{ji} = \frac{\vec{u}_j \times \vec{v}_i}{|\vec{v}_i|} \quad (14) \]

\[ \sigma_{ij} = \frac{[\vec{u}_i \times \vec{v}_j][\vec{u}_j \times \vec{v}_i]}{|[\vec{u}_i \times \vec{v}_j][\vec{u}_j \times \vec{v}_i]|} \quad (15) \]

To determine the electrostatic force acting on two segments, these four parameters are computed. An approximate value for the interaction integral is then interpolated from the table of pre-computed energy values. The electrostatic force is obtained by numerically computing the gradient with respect to the dynamical variables (i.e., the bead positions). The pre-computed energy tables are four-dimensional arrays, with the indices corresponding to distinct intervals of the four parameters \( (\varepsilon_j, \gamma_{ij}, \gamma_{ji}, \sigma_{ij}) \). The number of intervals per parameter can be specified in the simulation input file, whereas the maximum and minimum value for each parameter are set before computing the energy table. For a given set of indices, the electrostatic interaction energy is then determined by two-dimensional Romberg integration (Press et al., 1995).

As the charged DNA backbones are surrounded by a layer of cations, an effective (dimensionless) charge density \( \xi \) was computed according to the mechanism proposed by Stigter (1977). The parameter \( \alpha^{(c)} \) is then determined by \( \alpha^{(c)} = \xi^2 D_k T / e^2 \). A salt concentration of 100 mM NaCl results in a value of \( \alpha^{(c)} = 26.1/nm \) at \( T = 25°C \).

Hydrodynamic interaction. The hydrodynamic interaction tensor can be approximated for polymeric systems with beads of two different radii (Garcia and Bloomfield, 1977). Its explicit form is:

\[ \mathbf{D}_{ij} = \delta_{ij} \frac{k_B T}{6 \pi \eta r_i} (1 - (\delta_{ij} - 1)) \frac{k_B T}{8 \pi \eta r_j} \]

\[ \times \left[ \left( 1 + \frac{\vec{r}_{ij} \otimes \vec{r}_{ij}}{r_{ij}^2} \right) + \frac{2}{r_{ij}^2} \left( \frac{1}{3} I - \frac{\vec{r}_{ij} \otimes \vec{r}_{ij}}{r_{ij}^2} \right) \right] \quad (16) \]

where \( \vec{r}_{ij} = \vec{r}_i - \vec{r}_j, r_{ij} = |\vec{r}_{ij}| \text{ and } I \text{ is the } 3 \times 3 \text{ unit matrix. For a segment representing a nucleosome, the nucleosomal bead's position } \vec{r}_i^* \text{ is used instead of } \vec{r}_i. \)

The nucleosome beads are interacting with DNA or nucleosomal beads through a modified Lennard–Jones repulsion. The modulus of the force is limited to \( 10^4 k_B T / nm \) and is subject to a cut-off at bead distances greater than 1–2 nm. No such term is used for the interaction of two DNA beads as their repulsion is established by electrostatic interaction. The force acting on a nucleosome bead \( i \) due to interaction with a DNA bead \( j \) is given by:

\[ \vec{F}_{ij}^{(c)} = \frac{\vec{r}_{ij}}{|\vec{r}_{ij}|} \sigma_{ij} k_B T \left[ \left( \frac{\sigma}{r_{ij}} \right)^{14} - \left( \frac{\sigma}{r_{ij}} \right)^{7} \right] \quad (17) \]

where \( \vec{F}_{ij} = -\vec{F}_{ji} \) and \( r_{ij} \) is the distance between the bead surfaces. The parameter \( \sigma \) is the cut-off range for the excluded volume interaction. Its value is set to \( \sigma = 1–2 \text{ nm} \). The second parameter \( e \) determines the relative strength of the repulsive force; it is set to \( e = 0.002/nm^2 \). These values were chosen to minimize bead overlap during the simulations.

When two nucleosomal beads interact, the nucleosome position \( \vec{r}_i^N \) is used instead of the bead position \( \vec{r}_i. \)

Implementation

Typically, a study of a particular chromatin system can be divided into three separate fields: simulation, analysis and visualization of the trajectories. Instead of having different monolithic program packages for each of these three fields, one major design goal was to facilitate reuse of classes between them. The informal description of the classes is grouped according to the three aspects of usage.

Simulation

For the core simulation classes, the design goals were:

(i) Portability across different platforms. Therefore, no architecture-dependent code (i.e., graphics routines, numerical routines) should be used in these classes.

(ii) Ease of use. The classes should have an interface which makes the set-up of new simulations easy for non-specialists.

(iii) Facilitation of parallelization. The simulation of larger systems (i.e., in the range of \( n = 100-1000 \) segments) in a message-passing implementation requires that as much data as possible is maintained inside the objects to be transferred between processors. Besides, numerical algorithms should be able to exploit parallel architectures.
As a result of these design considerations, software components were organized as follows:

The central class is \texttt{Configuration}, which represents the concept of a physical configuration of DNA or nucleosome segments. These are stored in a list of objects of class \texttt{DNA\_Segment} or \texttt{Nucleosome\_Segment}, respectively. These segment objects contain all data structures needed to compute their local (i.e. elastic) interactions, whereas the \texttt{Configuration} object is used for the long-range interactions (i.e. electrostatic and hydrodynamic). The hydrodynamic interaction tensor is an instance variable of class \texttt{LASP\_DM\_Double}, the \texttt{LAPACK++} class representing positive semidefinite matrices. The Cholesky factorization needed is part of the object's instance methods, using the BLAS libraries for linear algebra computations. Since assembly-level optimized code is often available, the most efficient implementation of the numerical routine can be chosen for each platform.

The physical parameters are contained in a \texttt{Parameters} object, which is attached to a \texttt{Configuration} object. At each moment of (simulation) time, a configuration can be represented by a \texttt{Snapshot} object, which is the means of communication for \texttt{Configuration} and \texttt{Trajectory} objects. The latter are used for permanent storage of snapshots and communication for \texttt{Configuration} and \texttt{Trajectory} objects.

The physical parameters of a simulation (e.g. salt concentration, time step) are changed via the according parameters from file names needed in the simulation. If the user wants to simulate an open configuration, then the flag \texttt{Open\_chain} has to be given. The parameter \texttt{Accuracy\_limit} is used to determine the maximum number of intervals in a one-dimensional Romberg integration used in the electrostatics computations.

Fig. 2. A sample input file. This file is a collection of the external parameters needed for a simulation. These parameters are given in the format (keyword: value). Strings appearing after the hash sign are considered comments, with the exception of the two keywords \texttt{DATA} and \texttt{FILES} separating external parameters from file names needed in the simulation. If the user wants to simulate an open configuration, then the flag \texttt{Open\_chain} has to be given. The parameter \texttt{Accuracy\_limit} is used to determine the maximum number of intervals in a one-dimensional Romberg integration used in the electrostatics computations.

\begin{verbatim}
# PARAMETERS
Number_of_segments: 50
DNA_length._nm: 405.
Linking_number_difference: 0
Ionic_strength._M: 0.005
Kuhn_length._nm: 100.0
Temperature._grad_C: 20.
Streching_potential_constant: 0.008
Time_step._ps: 5
Steps_between_HI-calculations: 50
Number_of_tilt_angle_entries: 10
Number_of_twist_angle_entries: 18
Number_of_distance_entries: 15
Accuracy_limit: 6

# FILES
Conformational_file: sv40.cnf
Read_energy_table_from : sv40.1=5mM

By guest on June 3, 2013 http://bioinformatics.oxfordjournals.org/Downloaded from
the same time. By using the ConfigurationIterator class, the segments comprising of configuration can be accessed and local data (e.g. bead positions) can be extracted. As this analysis often consists of computing regressions, interpolation, least-squares fits or autocorrelations of the spatial coordinates, a hierarchy of analysis classes was developed to handle these standard operations. The design for these classes was taken from a collection of design patterns (a design pattern being a standard design solution of a particular problem class) compiled by Gamma et al. (1994). Here, the so-called Visitor pattern was used. Acting on objects of class DataSet [which contain sorted two-dimensional points (x, y) extracted from the coordinates], the various algorithms have been coded as objects. By using both the analysis objects and the I/O facilities offered by the simulation classes, a typical analysis program can be developed very rapidly. Besides, these programs tend to be no longer than one or two pages of source code. The complete source for measurement of the diffusion coefficient using an autocorrelation method, for example, contains less than 100 lines of code. As another example, computation of the radius of gyration (a quantity important in monitoring salt-induced compaction of chromatin) takes one page of source code. Typically, a piece of analysis code consists of looping over all snapshots in a trajectory. With t as a Trajectory object and c a Configuration object, this is achieved by a statement like t.useFile (fn), where fn denotes the name of the trajectory file. Single configurations are then accessed by a statement like c.readSnapshot (t.nextSnapshot()). After attaching a ConfigurationIterator it to c, the position of the hydrodynamic sphere for the first segment can then be accessed through (it.start())->getBead1().

Users who want to analyze trajectories according to their needs have to write a new analysis program. In our experience, even scientists without training in computer science could easily adapt analysis programs to suit their particular analysis method. To analyze new trajectories (representing different chromatin configurations), existing analysis programs can be used with new input trajectory files.

Visualization

Visualization of the simulation results should give the spectator a three-dimensional impression of the Brownian motion of chromatin. TIFF frames representing snapshots of the trajectories were generated with the NeXT 3DKit. This class library serves as an Objective-C (Cox and Novobilski, 1991) system of classes encapsulating the RenderMan API (Upstill, 1989). By mixing the Objective-C visualization classes with the C++ simulation classes, all the I/O and iteration facilities could be reused. The resulting hybrid language is called 'Objective-C++'. It is supported by the gcc2.5.8 port done by NeXT as part of NeXTSTEP3.2. Three-dimensional scenes are described by the RenderMan RIB files (similar to the Postscript page description language). These graphical descriptions of simulated trajectories can be rendered on a variety of architectures by using publicly available rendering tools [e.g. the Blue Moon Rendering Tools (http://www.seas.gwu.edu/student/gritz/bmrt.html)]. If a user wants to visualize different or multiple trajectories of the same chromatin configuration, the appropriate trajectory files have to be provided. For different chromatin configurations, the code has to be adapted and recompiled.

Instead of inspecting the trajectories by using rendering tools, their descriptions can be processed to produce a series of TIFF files. These were compressed into MPEG II movies to make visualization of results easily available on other platforms. Several of these files are accessible on the WWW via http://www.dkfz-heidelberg.de/Macromol/ehrlich/chromatin.html.

Discussion

In this section, preliminary results and limitations of the computer model will be discussed. A more detailed discussion of the results and their biological implications will be the subject of a forthcoming publication.

To calibrate parameters of the model, such as the angles θ and φ and the nucleosomal bead radius, dinucleosomes were simulated at various salt concentrations of the solvent. The diffusion coefficients extracted from the trajectories were compared with existing diffusion measurements for chicken erythrocyte dinucleosomes by Yao et al. (1990).

The segment length was set to 12.71 nm (~38 bp) to give the 76 bp of linker DNA for this particular chromatin [chicken erythrocyte chromatin has been reported to contain (216 ± 5)bp DNA (van Holde, 1989), of which 146bp are wound around the histone octamer]. This linker DNA content is divided into two segments, as shown in Figure 1. A comparison of the measured and simulated diffusion coefficients can be found in Figure 3. Good quantitative agreement between simulation and experiment was found for the parameters θ = 140°, φ = 0° and σ_Nuc = 5.95 nm. The choice of θ = 140° is in accordance with results from cryoelectron microscopy by Bednar et al. (1995). In their work, it was found that the average angle between linker DNA protruding from the core at 20 mM NACl was ~40° for chicken erythrocyte trinucleosomes. In our model, this observed bending angle corresponds to (180° − θ).

Second nucleosome tetramers were simulated by catenating four nucleosomes depicted in Figure 1. The segment length was chosen to be 15.98 nm (47 bp) to compare simulations to sedimentation measurements of sea urchin sperm chromatin done by Thomas et al. (1986). Diffusion coefficients were extracted from the sedimentation data using the Svedberg equation with a specific volume of \( V = 0.604 \text{ cm}^3/\text{g} \). Bead radii and angular parameters were the same as in the dinucleosome simulations. The diffusion
Fig. 3. Comparison of simulated and measured dinucleosome diffusion coefficients at various salt conditions. Experimental data were taken from Yao et al. (1990); the bars represent the spectrum of measured values (errors were reported to be 3% for each single measurement).

coefficients in the 20–60 mM NaCl range were in good agreement with the observed sedimentation data.

For di- and tetramers, the polymer model used thus gave a realistic description of chromatin on length scales of several nanometers and reproduces the correct increase in mobility upon an increase in the salt concentration. Therefore, the dominant interaction for chromatin condensation can be attributed to the partial screening of linker DNA electrostatic repulsion, as has been hypothesized earlier by Clark and Kimura (1990).

After the two series of simulations, a nucleosome 25mer was simulated for 400 μs, starting from an extended conformation. Here, all segments initially formed a straight line, while the permanent bending parameter θ was set to 140° for all nucleosomes. The second angular parameter, ϕ, was set to a random value from the set [0°, 36°, ..., 324°] for each nucleosome. The segment length was chosen to be 12.71 nm (~38 bp) to compare with data given by Woodcock (1994) for chromatin with a linker content of 76 bp. After 200 μs, the configuration had condensed to a fiber-like structure with an approximate fiber diameter of 45 nm, as compared to 42 nm measured by Woodcock (1994) for this particular chromatin. As opposed to the solenoid model for the chromatin fiber, this structure showed no helical order. The compaction ratio observed for the simulated fiber was less than that for the fibers observed in electron microscopy (1.2 nucleosomes per 10 nm of fiber as opposed to observed 4–6 nucleosomes per 10 nm). We consider this to be an indication that the inter-nucleosomal interactions are not yet modeled correctly.

As all simulated chromatin configurations are built by catenating a number of nucleosomes depicted in Figure 1, we implicitly assume that there is still overhanging linker DNA for open configuration (i.e. no exonuclease digestion has occurred).

Despite the surprising good agreement of hydrodynamic quantities, several interesting topics cannot yet be studied with the current model:

(i) As the positioning of the nucleosomes on the linker DNA is static in the course of the simulations (at least with the current version of the code), nucleosome sliding cannot be modeled. This motion might become interesting in rearrangement processes needed to start transcription.

(ii) Only monovalent cations are included in the solvent description.

(iii) As the DNA wound around the octamer is not explicitly modeled, its dynamics on the octamer surface as well as its unwinding upon H1 depletion cannot be studied.

In summary, our polymer model for chromatin leads to a realistic description of its hydrodynamic and structural properties. With the advent of better and faster hydrodynamic approximations, modeling of other diffusion-controlled processes (such as the onset of transcription or chromatin looping) becomes possible. Such models can be useful in elucidating how the structural properties of chromatin fibers affect processes like transcription and chromatin rearrangement in chromosomes. Besides, an adaptation of the code to entirely new scenarios for interacting chain molecules should be possible without great effort. Source code for the described classes is available from the authors.
References


Received on September 17, 1996: revised on January 1, 1997: accepted on January 10, 1997