Collagen Simulation Documentation

Francis G.J. Longford

March 26, 2018

1 Introduction

Many different approaches have been considered to model the behaviour of collagen. These range from the atomistic scale to the microscopic. Atomistic and coarse-grained molecular dynamics simulations have tended to focus on the mechanical properties of either single amino acid or short tropocollagen chains[1, 2, 3, 4]. Modelling of large scale collagen networks in the ECM use a much more simplified bead-and-spring approach for each fibre[5, 6, 7]. However, although these studies can report singular values of mechanical and viscoelastic properties that agree well with experimental data, a simulation to investigate the mechanism for ordering of collagen fibres has yet to be performed.

We propose that a bead-and-spring dynamical network approach to modelling the ECM could be a way forward to achieve this goal. In a similar methodology to previous studies, we suggest that a collagen fibre can be represented on a micro-metre scale by a series of atom-like beads connected by springs.

2 Modelling

2.1 Potentials

The potential energy of with respect to the length, r, of these springs is represented by a harmonic oscillator, with r_0 as the equilibrium bond length and k_b as the bond force constant.

$$V_b(r) = k_b(r - r_0)^2 \tag{1}$$

Additionally, the chains can be allowed to bend linearly by using a similar harmonic potential to model the bond angle, θ, between three beads, with θ_0 as the equilibrium bond angle and k_a as the angle force constant. The stiffening of the fibrils can therefore be modified by increasing the force constants of both potentials. It could also be assumed that since all beads are homogeneous the equilibrium bond angle would be 180, so that $\theta_0 = \pi$. For computational ease we approximate the form of this harmonic potential by a sinusoidal function

$$V_a(\theta) = k_a\left(\cos(\theta) + 1\right) \tag{2}$$

Whereas these potentials have been used several times previously to model bonding within individual collagen fibres and have clear physical analogies, the interstitial forces between fibres are less well understood. Some studies have used cross-links between fibres,(32),(35) although assigning these at the beginning of a dynamical simulation would predetermine the motion and ordering of the system if they were unable to break and form throughout. Other studies into the fibrillogenesis have suggested that lateral growth and fusion of two fibres side-by-side does not occur in the ECM, due to the twisted helical structure of fibrils.(36),(37) In which case, it would be inaccurate to explicitly assign bonds between fibres, rather than using a potential to describe these intra-fibriliar forces. Therefore, a Lennard-Jones potential could be appropriate to model the non-bonding interactions between fibres, as it has been extensively used in molecular dynamics to model dispersion

forces between non-bonded atoms.

$$V_{LJ}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3}$$

In order to model dispersion forces, this potential is applied to every pairwise distance between all beads in the simulation, although the energy parameter ϵ is significantly increase for interactions between beads at the ends of each linear chain. In this way we model the preferential localisation of cross-links between fibrils. Additionally, in order to make sure that there is an overlap in potential between bonded beads, the equilibrium bond length r_0 is set equal to $2^{1/6}\sigma$, the value of r at which there is a minima in the potential, so that $V(r_0) = -\epsilon$.

The total potential acting upon each particle in the collagen chain is therefore given by the sum- mation the of bonded and non-bonded interactions:

$$V_{TOT} = \sum^{bonds} V_b(r_{ij}) + \sum^{angles} V_a(\theta_{ijk}) + \sum^{pairwise} V_{LJ}(r_{ij}) \tag{4}$$

2.2 Forces

The force applied to each connected bead from this potential can be found by defining it as the negative of the derivative of potential with respect to bead position. Therefore, the bonded force vector $\mathbf{F}_b$ as a result of the displacement vector $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and radial distance $r_{ij} = |\mathbf{r}_{ij}|$ between beads i and j is given by:

$$\mathbf{F}_b(\mathbf{r}_i) = -\frac{\partial V_b(r_{ij})}{\partial \mathbf{r}_i} = -\frac{\partial V_b(r_{ij})}{\partial r_{ij}} \cdot \frac{\partial r_{ij}}{\partial \mathbf{r}_i} = \mathbf{r}_{ij} \frac{2k_b}{r_{ij}} (r_0 - r_{ij}) \tag{5a}$$

$$\mathbf{F}_b(\mathbf{r}_j) = -\frac{\partial V_b(r_{ij})}{\partial \mathbf{r}_j} = -\frac{\partial V_b(r_{ij})}{\partial r_{ij}} \cdot \frac{\partial r_{ij}}{\partial \mathbf{r}_j} = -\mathbf{r}_{ij} \frac{2k_b}{r_{ij}} (r_0 - r_{ij}) \tag{5b}$$

Performing a similar derivation as equation (5) results in an expression for the angular force $\mathbf{F}_a$ acting upon between beads i, j and k as a result of the angle θ_{ijk} that bisects vectors $\mathbf{r}_{ij}$ and $\mathbf{r}_{jk}$.

$$\mathbf{F}_a(\mathbf{r}_i) = -\frac{\partial V_a(\theta_{ijk})}{\partial \mathbf{r}_i} = -\frac{\partial V_a(\theta_{ijk})}{\partial \theta_{ijk}} \cdot \frac{\partial \theta_{ijk}}{\partial \mathbf{r}_i} = \frac{1}{r_{ij}} \left(\frac{\mathbf{r}_{jk}}{r_{jk}} + \frac{\mathbf{r}_{ij}}{r_{ij}} \sin(\theta_{ijk}) \right) \tag{6a}$$

$$\mathbf{F}_a(\mathbf{r}_j) = -\frac{\partial V_a(\theta_{ijk})}{\partial \mathbf{r}_j} = -\frac{\partial V_a(\theta_{ijk})}{\partial \theta_{ijk}} \cdot \frac{\partial \theta_{ijk}}{\partial \mathbf{r}_j} = -\frac{(\mathbf{r}_{ij} + \mathbf{r}_{jk})}{r_{ij} \cdot r_{jk}} - \sin(\theta_{ijk}) \left(\frac{\mathbf{r}_{ij}}{r_{ij}^2} - \frac{\mathbf{r}_{jk}}{r_{jk}^2} \right) \tag{6b}$$

$$\mathbf{F}_a(\mathbf{r}_k) = -\frac{\partial V_a(\theta_{ijk})}{\partial \mathbf{r}_k} = -\frac{\partial V_a(\theta_{ijk})}{\partial \theta_{ijk}} \cdot \frac{\partial \theta_{ijk}}{\partial \mathbf{r}_k} = \frac{1}{r_{jk}} \left(\frac{\mathbf{r}_{ij}}{r_{ij}} + \frac{\mathbf{r}_{jk}}{r_{jk}} \sin(\theta_{ijk}) \right) \tag{6c}$$

The non-bonded Lennard-Jones type forces between each bead are subsequently derived in the same fashion.

$$\mathbf{F}_{LJ}(\mathbf{r}_i) = -\frac{\partial V_b(r_{ij})}{\partial \mathbf{r}_i} = -\frac{\partial V_b(r_{ij})}{\partial r_{ij}} \cdot \frac{\partial r_{ij}}{\partial \mathbf{r}_i} = -\mathbf{r}_{ij} \frac{24\epsilon}{r_{ij}^2} \left[2 \left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{7a}$$

$$\mathbf{F}_{LJ}(\mathbf{r}_j) = -\frac{\partial V_b(r_{ij})}{\partial \mathbf{r}_j} = -\frac{\partial V_b(r_{ij})}{\partial r_{ij}} \cdot \frac{\partial r_{ij}}{\partial \mathbf{r}_j} = \mathbf{r}_{ij} \frac{24\epsilon}{r_{ij}^2} \left[2 \left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{7b}$$

This results on the net force acting upon a single bead i defined by:

$$\mathbf{F}_i = -\frac{\partial V_b(r_{ij})}{\partial \mathbf{r}_i} - \frac{\partial V_a(\theta_{ijk})}{\partial \mathbf{r}_i} - \frac{\partial V_{LJ}(r_{ij})}{\partial \mathbf{r}_i} \tag{8}$$

In such a way, a molecular dynamical simulation, with particle trajectories being determined by pairwise force interactions and integrated using the verlocity-Verlet algorthim could be used to simulate a simplified system of collagen fibrils in the ECM.

2.3 Fibril Dynamics

As mentioned previously, MD simulations are propagated through time by the calculation of net intra- and inter-molecular forces acting upon each particle. These forces are defined as proportional to the negative of the derivative of the potential energy function for each interaction. Therefore the net force per particle is a vector summation of all the individual covalent and non-covalent force contributions from its surrounding neighbours. The net force in each atom is then used to propagate the system through time using the atomic mass m from Newton's equations of motion.

$$\mathbf{F}_i(t) = m_i \mathbf{a}_i(t) \tag{9}$$

This leads to a set of differential equations for the position $\mathbf{r}_i(t)$ and velocity $\mathbf{r}_i(t)$ of each bead at any given moment in time t.

2.4 Temperature

In a dynamical system particles obey the fluctuation-dissipation theorem, whereby deviations of properties away from the ensemble average are counter-acted statistically by an equivalent reverse process. Thermal fluctuations alter particle velocities, which impact upon the particle collision rate and so vary the amount of kinetic energy dissipated. Therefore the average temperature of a system can be represented in terms of kinetic energy and consequently average particle velocities. At equilibrium, classical systems adhere to the Maxwell-Boltzmann distribution of energy states. This means that the total kinetic and potential energies will fluctuate around their ensemble averages with a probability given by a normal distribution. The total kinetic energy K_{TOT} is given in terms of individual particle velocities by

$$K_{TOT} = \frac{1}{2} \sum_{i}^{N} m_i |\mathbf{v}_i|^2 \tag{10}$$

The instantaneous temperature T of a system containing N particles is related to K_{TOT} via the Boltzmann constant k_B and number of degrees of freedom of each particle N_f.

$$k_B T = \frac{1}{2}\frac{K_{TOT}}{N_f} = \frac{1}{2N_f}\sum_{i}^{N} m_i |\mathbf{v}_i|^2 \tag{11}$$

Therefore we can constrain the velocities of each bead in the system by setting a fixed parameter $k_B T$, to which this relation must hold.

2.5 SHG Image Recreation and Alignment Analysis

In order to produce imitation images to compare with SHG results, we convolute a 2D discrete bead density distribution with a Gaussian function. The resulting image intensity map $I(x, y)$ is designed to mimic experimental images of collagen intensity produced by SHG analysis of prostate biopsies[8].

Given a set of bead positions $\mathbf{r}$ for N beads, a histogram $H(x, y)$ is created by binning the frequency of each position across a discrete grid. The number of grid points in each dimension is given by selecting an appropriate level of resolution. We chose to set the length of each fibre as 1 μm, so that for a fibre n beads long the radius of one bead equivalent to $\sigma = (2n)^{-1}$ μm. A typical SHG slide possesses a resolution of ..., leading to the conversion factor...

The value of this discrete density distribution at each pixel $H(x, y)$ is then convoluted by integrating a Gaussian function with a variance of σ^2 over a circular area with a maximum radius of r_c to yield the intensity map $I(x, y)$.

$$I(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \oint_{x'^2+y'^2 \leq r_c^2} H(x'-x, y'-y) \exp\left(-\frac{x'^2+y'^2}{2\sigma^2}\right) dx'dy' \tag{12}$$

For simulations in 3D, we also include a discrete distribution along the z axis, and weight

$H(x, y, z)$ by the radial distance viewed when from the plane $z = 0$. In both cases we are left with an intensity function containing two variables $I(x, y)$ only.

$$I(x,y) = \frac{1}{\sqrt{2\pi}\sigma} \oint\limits_{x'^2+y'^2+z'^2 \leq r_c^2} H(x'-x, y'-y, z') \exp\left(-\frac{x'^2+y'^2+z'^2}{2\sigma^2}\right) dx'dy'dz' \quad (13)$$

In order to measure the alignment of the collagen fibril network we adopt the same methodology as Garcia *et al.*[8], who employed the FibrilTool plugin[9] of the ImageJ software package[10]. This tool calculates the tangent to the fibril structure unit vector $\mathbf{t}(x, y)$:

$$\mathbf{t}(x,y) = (\mathbf{t}_x, \mathbf{t}_y) = \frac{(-\partial I/\partial y,\ \partial I/\partial x)}{\sqrt{(\partial I/\partial x)^2 + (\partial I/\partial y)^2}} \quad (14)$$

Which is transformed into components of the 2x2 nematic tensor $\mathbf{n}(x, y)$ for each pixel.

$$\mathbf{n}(x,y) = \begin{pmatrix} \mathbf{n}_{xx} & \mathbf{n}_{xy} \\ \mathbf{n}_{yx} & \mathbf{n}_{yy} \end{pmatrix} = \begin{pmatrix} \mathbf{t}_x^2 & \mathbf{t}_x\mathbf{t}_y \\ \mathbf{t}_x\mathbf{t}_y & \mathbf{t}_y^2 \end{pmatrix} \quad (15)$$

The nematic tensor for a selected area then becomes the average local tensor $\langle\mathbf{n}\rangle$ of the constituent pixels. The eigenvalues $n_1 > n_2$ and corresponding eigenvectors $\mathbf{e}_1, \mathbf{e}_2$ of $\langle\mathbf{n}\rangle$ are used to measure the anisotropy of collagen fibrils in the sampled image area $\xi = n_1 - n_2$.

Considering the definition of our Gaussian convoluted intensity map $I(x, y)$ is another Gaussian convoluted intensity map $I(x, y)$, we choose to implement the FibrilTool methodology ourselves, rather than exporting images for analysis with ImageJ. Consequently, we

define the derivatives in equation (14) by the following:

$$\frac{\partial I(x,y)}{\partial x'} = -\frac{1}{\sqrt{2\pi}\sigma^3} \oint\limits_{x'^2+y'^2 \leq r_c^2} x' H(x'-x, y'-y) \exp\left(-\frac{x'^2+y'^2}{2\sigma^2}\right) dx'dy' \tag{16a}$$

$$\frac{\partial I(x,y)}{\partial y'} = -\frac{1}{\sqrt{2\pi}\sigma^3} \oint\limits_{x'^2+y'^2 \leq r_c^2} y' H(x'-x, y'-y) \exp\left(-\frac{x'^2+y'^2}{2\sigma^2}\right) dx'dy' \tag{16b}$$

Which can then be easily transformed into $\mathbf{t}(x,y)$ and $\mathbf{n}(x,y)$ via equations (14) and (15) respectively.

3 Computational Routines

The simulation has been written in 2D and 3D.

3.1 Calculation of Angles

In order to efficiently calculate the force and energy components of each multi-body term, we neglect to calculate each angle θ_{ijk} between connected beads i, j and k explicitly, and instead use sinusoidal functions (equations (2), (6)). We then calculate the terms $\sin(\theta_{ijk})$ and $\cos(\theta_{ijk})$ from the cross and dot products respectively of the displacement vectors $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ and $\mathbf{r}_{jk} = \mathbf{r}_j - \mathbf{r}_k$.

$$\sin(\theta_{ijk}) = \frac{\mathbf{r}_{ij} \otimes \mathbf{r}_{jk}}{|\mathbf{r}_{ij}||\mathbf{r}_{jk}|} \qquad \cos(\theta_{ijk}) = \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{|\mathbf{r}_{ij}||\mathbf{r}_{jk}|} \tag{17}$$

The cross-product for two dimensional vectors is undefined, and so when running a simulation in 2D, we used the following trigonometric definitions.

$$\sin(\theta_{ijk}) = \frac{\det(\mathbf{r}_{ij}, \mathbf{r}_{jk})}{|\mathbf{r}_{ij}||\mathbf{r}_{jk}|} \qquad \cos(\theta_{ijk}) = \frac{\mathbf{r}_{ij} \cdot \mathbf{r}_{jk}}{|\mathbf{r}_{ij}||\mathbf{r}_{jk}|} \tag{18}$$

Where the determinant of two 2D vectors is defined by.

$$\det(\mathbf{a}, \mathbf{b}) = \mathbf{a}_y \mathbf{b}_x - \mathbf{a}_x \mathbf{b}_y \tag{19}$$

We use numerical python (NumPy) to calculate each sinusoidal term, due to its speed when handling vector arrays, which is close to that of compiled C.

3.2 Velocity-Verlet Integration

A Verlet integration routine is used to propagate the acceleration $\mathbf{a}_i(t)$ and velocity $\mathbf{v}_i(t)$ vectors through a time step Δt in order to solve the positions $\mathbf{r}_i(t + \Delta t)$ and velocities $\mathbf{v}_i(t + \Delta t)$. It is assumed that the forces remain constant during a finite difference in time Δt, which is reversible and therefore the change in dynamical variables can be approximated as a Taylor series expansion.

$$\begin{aligned} \mathbf{r}_i(t + \Delta t) &= \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 + ... \\ \mathbf{r}_i(t - \Delta t) &= \mathbf{r}_i(t) - \mathbf{v}_i(t)\Delta t - \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 + ... \\ \therefore \quad \mathbf{r}_i(t + \Delta t) &= 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \mathbf{a}_i(t)\Delta t^2 \end{aligned} \tag{20}$$

If we wish to explicitly solve $\mathbf{v}_i(t)$, this normally requires both $\mathbf{r}_i(t - \Delta t)$ and $\mathbf{r}_i(t + \Delta t)$ to be already known. However, the Velocity Verlet algorithm[11] uses a half-step velocity

calculation, allowing for $\mathbf{v}_i(t+\Delta t)$ to be estimated at the same time as $\mathbf{r}_i(t+\Delta t)$.

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 \tag{21a}$$

$$\mathbf{a}_i(t+\Delta t) = \frac{\mathbf{F}_i(t+\Delta t)}{m_i} \tag{21b}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{1}{2}\left[\mathbf{a}_i(t) + \mathbf{a}_i(t+\Delta t)\right]\Delta t \tag{21c}$$

Note that the forces are recalculated halfway through the algorithm based on the updated nuclear positions. An appropriate length of time step Δt is critical to maintain the conservation of energy and ensure the integration method is time-reversible. A larger time step will increase the speed of simulation and computational efficiency, though at the cost of numerical accuracy. In order to avoid bead overlaps, leaping to physically unrealistic interactions and excessive potential energies Δt is required to be shorter than the fastest motions in the system.

3.3 Langevin Dynamics

The Langevin equation is a stochastic differential equation, describing a slower-moving system coupled to one containing fast degrees of freedom that are implicitly taken into account by a random force. This seems an appropriate description of motion for large collagen proteins containing 1000s of atoms immersed in the ECM blood-solution, mainly comprised of water. In order to replicate these dynamics we employ a Langevin thermostat, which includes an explicit term for the drag, or friction force as a function of bead velocity, as well as a stochastic force $\mathbf{R}(t)$, representing thermal "noise". The drag force felt by particle i is defined as proportional to its velocity and a friction coefficient ξ.

$$F_i(t) = -\xi \mathbf{v}_i(t) \tag{22}$$

The friction coefficient is related to the collision frequency γ_i of a particle via $\gamma_i = \xi/m_i$, so that γ_i^{-1} is a measure of the velocity relaxation time. Relaxation times are usually expressed, since they can be parametrised using fluid viscosities via Stoke's law. Therefore (9) can be amended with these terms included.

$$F_i(t) = m_i \mathbf{a}_i(t) - \gamma_i m_i v_i(t) + \mathbf{R}_i(t) \tag{23}$$

We include this constraint in the velocity-Verlet algorithm via the second-order integration method proposed by Vanden-Eijnden and Ciccotti[12], which is accurate to order Δt^2. In which case the stochastic force at each time step $\mathbf{R}_i(t)$ is split into two terms, $\eta_i(t)$ and $\xi_i(t)$, both taken from a Gaussian distribution with a zero mean and variance of $2k_B T\gamma_i/m_i$. Initially we calculate $\mathbf{C}_i(t)$, an approximant of the integral of $\mathbf{v}_i$ between each time-step.

$$\mathbf{C}_i(t) = \frac{1}{2}\left[\mathbf{a}_i(t) - \gamma_i \mathbf{v}_i(t)\right]\Delta t + \left(\frac{1}{2}\eta_i(t) + \frac{1}{2\sqrt{3}}\xi_i(t)\right)\Delta t^{3/2} \tag{24}$$

This term is then incorporated into equations (21) via the following routine.

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \mathbf{C}_i(t) \tag{25a}$$

$$\mathbf{a}_i(t+\Delta t) = \frac{\mathbf{F}_i(t+\Delta t)}{m_i} \tag{25b}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \left\{\frac{1}{2}\left[\mathbf{a}_i(t+\Delta t) + \mathbf{a}_i(t)\right] - \gamma_i \mathbf{v}_i(t)\right\}\Delta t + \eta(t)\sqrt{\Delta t} - \gamma_i \mathbf{C}(t) \tag{25c}$$

References

[1] M. J. Buehler, "Nature designs tough collagen: Explaining the nanostructure of collagen fibrils," *Proceedings of the National Academy of Sciences*, vol. 103, no. 33, pp. 12285–12290, 2006.

[2] A. Gautieri, A. Russo, S. Vesentini, A. Redaelli, and M. J. Buehler, "Coarse-grained

model of collagen molecules using an extended martini force field," *Journal of Chemical Theory and Computation*, vol. 6, no. 4, pp. 1210–1218, 2010.

[3] A. Gautieri, S. Vesentini, A. Redaelli, and M. J. Buehler, "Hierarchical structure and nanomechanics of collagen microfibrils from the atomistic scale up," *Nano Letters*, vol. 11, no. 2, pp. 757–766, 2011. PMID: 21207932.

[4] I. Streeter and N. H. de Leeuw, "A molecular dynamics study of the interprotein interactions in collagen fibrils," *Soft Matter*, vol. 7, pp. 3373–3382, 2011.

[5] A. M. Stein, D. A. Vader, D. A. Weitz, and L. M. Sander, "The micromechanics of three-dimensional collagen-i gels," *Complexity*, vol. 16, no. 4, pp. 22–28, 2011.

[6] S. Nachtrab, S. C. Kapfer, C. H. Arns, M. Madadi, K. Mecke, and G. E. Schrder-Turk, "Morphology and linear-elastic moduli of random network solids," *Advanced Materials*, vol. 23, no. 22-23, pp. 2633–2637, 2011.

[7] L. Zhang, S. Lake, V. Lai, C. Picu, V. Barocas, and M. Shephard, "A coupled fiber-matrix model demonstrates highly inhomogeneous microstructural interactions in soft tissues under tensile load," *J Biomech Eng.*, vol. 135, pp. 011008–011008–9, 2012.

[8] A. M. Garcia, F. L. Magalhes, J. S. Soares, E. Paulino-Jr, M. F. de Lima, M. Mamede, and A. M. de Paula, "Second harmonic generation imaging of the collagen architecture in prostate cancer tissue," *Biomed. Phys. Eng. Express*, vol. 4, p. 025026, 2018.

[9] A. Boudaoud, A. Burian, D. Borowska-Wykret, M. Uyttewaal, R. Wrzalik, D. Kwiatkowska, and O. Hamant, "Fibriltool, an imagej plug-in to quantify fibrillar structures in raw microscopy images," *Nat. Protocols*, vol. 9, p. 457463, 2014.

[10] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri, "Nihimage to imagej: 25 years of image analysis," *Nat. Methods*, vol. 9, p. 671675, 2012.

[11] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters," *The Journal of Chemical Physics*, vol. 76, no. 1, pp. 637–649, 1982.

[12] E. Vanden-Eijnden and G. Ciccotti, "Second-order integrators for langevin equations with holonomic constraints," *Chemical Physics Letters*, vol. 429, no. 1, pp. 310 – 316, 2006.