Supporting Information for

# Minimum Free-Energy Paths for the Self-Organization of Polymer Brushes

Ignacio Gleria,[a] Esteban Mocskos[a,b] and Mario Tagliazucchi*,[c,d]

a. Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales, Departamento de Computación. Buenos Aires, Argentina.

b. CONICET. Centro de Simulación Computacional para Aplicaciones Tecnológicas (CSC). Buenos Aires, Argentina.

c. Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales, Departamento de Química Inorgánica, Analítica y Química-Física. Buenos Aires, Argentina. E-mail: mario@qi.fcen.uba.ar

d. CONICET-Universidad de Buenos Aires. Instituto de Química Física de los Materiales, Medio Ambiente y Energía (INQUIMAE). Buenos Aires, Argentina.

## 1. Discretization and Numerical Solving

### 1.1 Probability-Evolution Equation

The calculation of the minimum free-energy path (MFEP) using the string method requires to evolve the probability of each conformation of each chain in the system according to the following equation:

$$\frac{d\Omega^{k}\left(j,\alpha\right)}{dt}=-\frac{\partial\beta F}{\partial\Omega^{k}\left(j,\alpha\right)} \tag{S1}$$

which is eq (13) in the main text. The right hand side of eq (S1) can be obtained from eq (9) in the main text:

$$\frac{\partial\beta F}{\partial\Omega\left(j,\alpha\right)}=\frac{\partial\beta F}{\partial P\left(j,\alpha\right)}\left(\frac{\partial f\left(P\left(j,\alpha\right)\right)}{\partial P\left(j,\alpha\right)}\right)^{-1} \tag{S2}$$

where $\partial\beta F/\partial P\left(j,\alpha\right)$ is given by eq (6) in the main text:

$$\begin{aligned}\frac{\partial\beta F}{\partial P\left(j,\alpha\right)}&=\ln P\left(j,\alpha\right)+\ln\left(q(j)\right)\\&+\int n_{p}\left(\alpha,\mathbf{r},j\right)\left(-\frac{v_{p}}{v_{s}}\left(\ln\left(\rho_{s}(\mathbf{r})v_{s}\right)\right)-\int\beta\varepsilon g\left(\left|\mathbf{r}-\mathbf{r}'\right|\right)\left\langle\rho_{p}(\mathbf{r}')\right\rangle d\mathbf{r}'\right)d\mathbf{r}\end{aligned} \tag{S3}$$

The evaluation of $\partial\beta F/\partial P\left(j,\alpha\right)$ is the most expensive part of the calculation and therefore we optimized its numerical computation. We will explain next our approach for calculating eq. (S3). For simplicity, we will write the equation for a fixed *j* and, therefore, we will drop this index from our notation. In practice, the evaluation of $\partial\beta F/\partial P\left(j,\alpha\right)$ needs to be performed for all *j*, but this process can be trivially parallelized. In our code, the parallelization in *j* is distributed among processors using the MPI (*Message Passing Interface*) standard.

We first discretize our equations into a lattice. Since our problem is two-dimensional, we discretize our system in the *x* and *y* directions in square cells of side $\delta$ (where $\delta$ is a discretization parameter that we choose to be equal to 0.5 nm in the present work). The discretized system has size $D_x$ in the *x* dimension and $D_y$ in the *y* dimension. While in principle two indexes can be used to label each cell ($i_x$ and $i_y$, representing the coordinates of the cell in *x* and *y* dimensions), the matrix formulation of the problem requires to use only one index to label each cell. Thus, we define this index as $i = i_x + (i_y-1)*D_x$ (note that $1 \le i \le D_x \cdot D_y$). We now discretize eq (S3) as:

$$\frac{\partial\beta F}{\partial P\left(\alpha\right)}=\ln P\left(\alpha\right)+\ln\left(q\right)+\sum_{i}n_{p}\left(\alpha,i\right)\left(-\frac{v_{p}}{v_{s}}\left(\ln\left(\rho_{s}(i)v_{s}\right)\right)-\sum_{i'}\beta\varepsilon g\left(i,i'\right)\left\langle\rho_{p}(i')\right\rangle\right) \tag{S4}$$

where $n_p(\alpha,i)$ is the number of polymer segments that the chain in conformation $\alpha$ has in the lattice cell *i* and $g(i,i')$ is a coefficient that the gives the strength of the VdW interactions between a segment in cell *i* and other segment in cell *i*' (see next section for its calculation).

It is also useful to define:

$$pot(i)=\left(-\frac{v_{p}}{v_{s}}\left(\ln\left(\rho_{s}(i)v_{s}\right)\right)-\sum_{i'}\beta\varepsilon g\left(i,i'\right)\left\langle\rho(i')\right\rangle\right) \tag{S5}$$

In order to formulate the problem in matricial form, we define the following vectors:

$$\mathbf{F_P} = \left( \frac{\partial \beta F}{\partial P(1)}, \ldots, \frac{\partial \beta F}{\partial P\left(N_{conf}\right)} \right) \tag{S6}$$

$$\mathbf{v} = \left( \ln\left(P(1) \cdot q\right), \ldots, \ln\left(P\left(N_{conf}\right) \cdot q\right) \right) \tag{S7}$$

and

$$\mathbf{pot} = \begin{pmatrix} pot(1) \\ \ldots \\ pot(D_x \cdot D_y) \end{pmatrix} \tag{S8}$$

and the matrix:

$$\mathbf{n_p} = \begin{pmatrix} n_p(1,1) & \ldots & n_p\left(1, D_x \cdot D_y\right) \\ \ldots & & \ldots \\ n_p\left(N_{conf}, 1\right) & \ldots & n_p\left(N_{conf}, D_x \cdot D_y\right) \end{pmatrix} \tag{S9}$$

We can now rewrite eq (S4) as:

$$\mathbf{F_P} = \mathbf{v} + \mathbf{n_p} \cdot \mathbf{pot} \tag{S10}$$

At this point it is important to note that $N_{\mathrm{conf}} = 2.5 \cdot 10^5$ and $D_x \cdot D_y \sim 400$. Therefore the matrix $\mathbf{n_p}$ has $\sim 10^8$ elements and the matrix multiplication in eq (S10) becomes one of the bottlenecks of the MFEP calculation. The calculation of *pot*(i) from eq (S5) is much faster than this matrix multiplication, but it requires to know $\langle \rho_p(i) \rangle$ and $\rho_s(i)$ (the density of the polymer and solvent at cell *i*, respectively). The evaluation of $\langle \rho_p(i) \rangle$ is a time consuming process by itself as it also requires a matrix multiplication ($\rho_s(i)$ can be straightforwardly determined once $\langle \rho_p(i) \rangle$ is known using the packing constraint, eq (3) in the main text).

To write down the calculation of $\langle \rho_p(i) \rangle$ in matricial form, we start with eq (2) in the main text:

$$\left\langle \rho_p(\mathbf{r}) \right\rangle = \sum_{j=1}^{N_p} \sum_{\alpha=1}^{Nconf} P(j,\alpha) n_p\left(\alpha, \mathbf{r}, j\right) \tag{S11}$$

We can rewrite this equation as

$$\left\langle \rho_p(\mathbf{r}) \right\rangle = \sum_{j=1}^{N_p} \left\langle \rho_p(\mathbf{r}, j) \right\rangle \tag{S12}$$

with

$$\left\langle \rho_p(\mathbf{r}, j) \right\rangle = \sum_{\alpha=1}^{Nconf} P(j,\alpha) n_p\left(\alpha, \mathbf{r}, j\right) \tag{S13}$$

The calculation of ⟨$\rho_p(\mathbf{r},j)$⟩ can be therefore trivially parallelized, so we will drop the index *j* in eq. (S13) for simplicity. We discretized the resulting equation, which results in:

$$\left\langle \rho_p(i) \right\rangle = \sum_{\alpha=1}^{Nconf} P(\alpha) n_p\left(\alpha, i\right) \tag{S14}$$

We now define the vectors:

$$\mathbf{P} = \begin{pmatrix} P(1) \\ \dots \\ P(N_{conf}) \end{pmatrix} \tag{S15}$$

$$\boldsymbol{\rho}_\mathbf{p} = \left( \left\langle \rho_p(1) \right\rangle, \dots, \left\langle \rho_p(D_x \cdot D_y) \right\rangle \right) \tag{S16}$$

so that eq (S14) can be rewritten in matricial form as:

$$\boldsymbol{\rho}_\mathbf{p} = \mathbf{n}_\mathbf{p}{}^T \cdot \mathbf{P} \tag{S17}$$

where the superindex *T* denotes the transverse matrix.

The most expensive computation steps in the calculation of the MFEP are the matrix multiplications in eqs (S10) and (S17), which involve the matrix $\mathbf{n_p}$. This matrix contains the number of segments that each polymer conformation has in each lattice site. For each chain in the system, we require to know the position of $N_{conf} \cdot L$ segments (where *L* is the chain length of the polymer), while $\mathbf{n_p}$ has a size equal to $D_x \cdot D_y \cdot N_{conf}$. The maximum fraction of non-zero elements in $\mathbf{n_p}$ is, therefore, $L/(D_x \cdot D_y)$. Using typical values of $L \sim 50$ and $D_x \sim D_y \sim 20$, we get a maximum fraction of non-zero elements of ~ 0.125 and, thus, $\mathbf{n_p}$ is a sparse matrix (the real fraction of non-zeros is actually smaller than this because some conformations have more than one segment in a given lattice cell). We store $\mathbf{n_p}$ as a compressed matrix in memory due to its large size. In previous works,[1, 2] we compressed $\mathbf{n_p}$ using a matrix of size $N_{conf} \cdot L$ where the element at the matrix position ($\alpha$,*k*) contained the position in the lattice of the segment *k* of conformation $\alpha$. In this previous works, we performed the matrix multiplications required in eqs (S10) and (S17) using loops over *L* and $N_{conf}$. Note that this procedure uses indirection to perform the matrix multiplication. In the present work, we decided to optimize the matrix multiplication step by using the subroutine *dcsrmv* from the Intel® MKL library,[3] which allows highly efficient multiplication of sparse matrices compressed in the CSR column format. The MKL implementation of our present program runs between 30% and 260% faster (depending on the parameters of the calculation and the computational architecture) than an implementation using the previous method for matrix multiplication.

### 1.2 Determination of the Poor Solvent Coefficients, *g*(*i*,*i*')

The function $g(|\mathbf{r} - \mathbf{r}'|)$ determines the distance dependence of the VdW interactions in the theory. We model this dependence with the attractive branch of the Lennard-Jones potential with a cutoff of 2.5 $\delta$, namely:

$$g\left(\left|\mathbf{r}-\mathbf{r}'\right|\right)=4\left(\frac{a}{\left|\mathbf{r}-\mathbf{r}'\right|}\right)^6 \qquad \text{for } a<\left|\mathbf{r}-\mathbf{r}'\right|<2.5\delta$$
$$g\left(\left|\mathbf{r}-\mathbf{r}'\right|\right)=0 \qquad \text{otherwise} \tag{S18}$$

where $a$ is the segment length (we used $a$ = 0.5 nm). The discretization of $g(|\mathbf{r} - \mathbf{r}'|)$ into the lattice results in the coefficients $g(i,i')$, which are required to calculate eq (S5). This coefficients are obtained by integration of $g(|\mathbf{r} - \mathbf{r}'|)$: if we denote the $x$ and $y$ coordinates of cells $i$ and $i'$ in the two dimensional lattice as ($i_x$, $i_y$) and ($i_x'$, $i_y'$), respectively, then $g(i,i')$ is given by:

$$g\left(i,i'\right)=\int_{\delta\left(i_x'-i_x-\frac{1}{2}\right)}^{\delta\left(i_x'-i_x+\frac{1}{2}\right)} dx \int_{\delta\left(i_y'-i_y-\frac{1}{2}\right)}^{\delta\left(i_y'-i_y+\frac{1}{2}\right)} dy \int_{-\infty}^{\infty} dz \; f\left(x,y,z\right) \tag{S19}$$

with

$$f(x,y,z)=\begin{cases} 4\left(\dfrac{a}{\left(x^2+y^2+z^2\right)^{1/2}}\right)^6 & \text{if } a<\left(x^2+y^2+z^2\right)^{1/2}<2.5\delta \\ 0 & \text{otherwise} \end{cases} \tag{S20}$$

## 2 Numerical Confirmation that the Barriers in the MFEP are Saddle Points of the Free-Energy Functional

In this section, we propose a method to confirm that the free-energy barriers of the MFEP found by the string method/molecular theory are actually saddle points of the free energy of the system. We have shown that taking $\partial\beta F/\partial P\left(j,\alpha\right)=0$ results in the following equation (eq (8) of the main text),

$$P\left(j,\alpha\right)=\frac{1}{q(j)}\exp\left[-\int n_p\left(\alpha,\mathbf{r},j\right)\left(v_p\beta\pi\left(\mathbf{r}\right)-\int\beta\varepsilon g\left(\left|\mathbf{r}-\mathbf{r}'\right|\right)\left\langle\rho_p(\mathbf{r}')\right\rangle d\mathbf{r}'\right)d\mathbf{r}\right] \tag{S21}$$

This equation is valid not only for the local minima of the free energy, but also for all other stationary points (saddle points); although in practice numerical solvers fail to find the saddle points (which are unstable) and provide always the local minima of $F$ (the stable solutions). We can now replace eq. (S21) into the free-energy functional of the system, eq (5) in the main text. Combining the resulting expression with the definition of the polymer number density (eq (2) in the main text), the packing constraint (eq (3) in the main text), the probability normalization constraint (eq (4) in the main text) and the definition of $\beta\pi(\mathbf{r})$ (eq (7) in the main text) results in:[4]

$$\beta F^{stationary}=-\int\beta\pi\left(\mathbf{r}\right)d\mathbf{r}-\int\rho_s(\mathbf{r})d\mathbf{r}-\sum_{j=1}^{N_p}\ln\left(q(j)\right)$$
$$+\int\int\frac{\beta\varepsilon}{2}g(|\mathbf{r}-\mathbf{r}'|)\left\langle\rho_p(\mathbf{r})\right\rangle\left\langle\rho_p(\mathbf{r}')\right\rangle d\mathbf{r}d\mathbf{r}' \tag{S22}$$

where the superindex "*stationary*" denotes that this expression for the free energy is only valid in the stationary points of $F$, in contrast with the general expression given by eqs (1) and (5) in the main text that is always valid. Therefore, if we now plot $F$ and $F^{\text{stationary}}$ along the MFEP, we should find that $F^{\text{stationary}} = F$ only for the points that are saddle points or extrema of *F*. Figure S1 shows that this is actually what happens for the initial and final states (that are local free-energy minima) and for the barrier state of the MFEP which is, therefore, a saddle point of *F*.

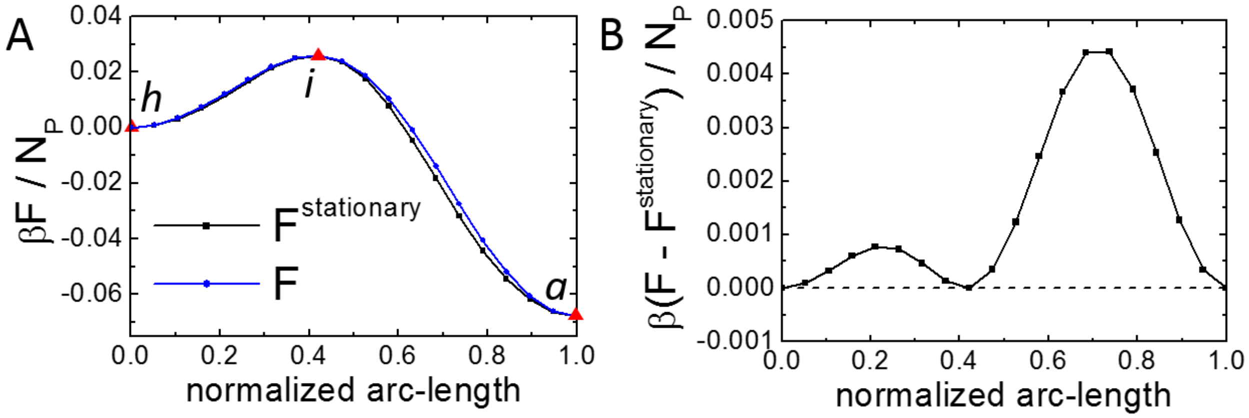

Figure S1: A. Free energy of the system, *F*, determined with eq (5) in the main text and free energy for the stationary points, $F^{\text{stationary}}$, given by eq (S22) as a function of the normalized arc-length of the string (same data as in Figure 3 of the main text). B. Difference between the free energies in A as a function of the normalized arc-length of the string.

## References

1. M. Tagliazucchi, M. O. de la Cruz and I. Szleifer, *Proc. Natl. Acad. Sci. U.S.A.*, 2010, **107**, 5300-5305.
2. M. Tagliazucchi, O. Azzaroni and I. Szleifer, *J. Am. Chem. Soc.*, 2010, **132**, 12404-12411.
3. https://software.intel.com/en-us/node/468578, (accessed 11/23/2016).
4. P. Gong and I. Szleifer, *Industrial & Engineering Chemistry Research*, 2006, **45**, 5466-5476.