PHYSICAL REVIEW E **77**, 030401(R) (2008)

Linear self-assembly under confinement

J. R. Henderson
School of Physics and Astronomy and Centre for Self-Organising Molecular Systems, University of Leeds, Leeds LS2 9JT, United Kingdom

(Received 12 December 2007; published 4 March 2008)

An exactly solvable model is used to obtain the response to confinement of the cluster distribution of linear aggregation. A direct relevance to simulation studies of linear self-assembly in discotic solutions and in peptide tape formation is proposed. The mapping predicts, for typical simulation procedures, that a finite reservoir of solute leads to a dramatic departure from isodesmic chemical equilibria for solute-solute interaction strengths higher than only a few k_BT.

DOI: 10.1103/PhysRevE.77.030401 PACS number(s): 82.70.Uv, 05.20.−y, 61.20.Qg, 82.60.Hc

In this Rapid Communication "linear self-assembly" denotes the formation of chainlike aggregates via chemical equilibria in solvent. Examples include solutions of disk-shaped surfactants [1], the formation of peptide tapes [2], and some aspects of exotic systems such as wormlike micelles [3] and filamentous networks [4]. In the thermodynamic limit linear self-assembly reduces to isodesmic chemical equilibria and hence an exponentially decaying cluster distribution. Simulation studies are typically carried out with a finite reservoir of solute, which can dramatically affect the observed cluster distribution [5]. Also, periodic boundary conditions (PBCS) can stabilize aggregates of infinite length. This Rapid Communication proposes a mapping of the finite-size dependence of the cluster distribution of linear self-assembly to the exact cluster distribution of the one-dimensional lattice-gas (LG) model under confinement.

A wide range of issues of importance to physical chemistry are readily identified and made accessible via Widom's potential distribution theorem (PDT) applied to the LG model in the grand ensemble [6]:

$$\rho(x)/[1-\rho(x)] = z e^{-\beta v^{ext}(x)} \langle e^{-\beta \psi(x)} \rangle_c. \tag{1}$$

Here, $\rho(x)$ is the density profile at position x, β denotes the inverse temperature T in units of Boltzmann's constant k_B, and $v^{ext}(x)$ and $\psi(x)$ are energies defined by an external field and by the interparticle interactions felt by a test particle placed at position x, respectively. The subscript c on the conditional average reminds us that there must already be an empty cell fixed at position x to accommodate the test particle (in the LG model this is a probability $1-\rho(x)$ representing repulsive interactions with the test particle). The symbol z denotes the activity $e^{\beta\mu}$ with μ the chemical potential of solute and it is also convenient to introduce the quantity $a \equiv e^{\beta\epsilon}-1$ where ϵ is the attractive well-depth of the nearest-neighbor LG model. We are interested in applying the PDT to confined systems of "width" L. For example, consider evaluating the cluster distribution in PBCS; i.e., we seek c_n the probability of finding a cluster of length n with a specified monomer occupying an arbitrary position (in PBCS all cells are equivalent). We could choose any monomer, but I shall always first consider the far left member of an aggregate and then build up the aggregate by adding additional monomers to the right (conditional probability ensures that any other method of creating the aggregate must be mathematically equivalent). Note also that the cell size is the unit of length, so that the density and all cluster densities reduce to probabilities lying between 0 and 1. The probability of finding a solute in the first cell is the density ρ_L of our PBCS system of width L. However, this is not the "cluster" probability c_1, since an isolated monomer must be surrounded by solvent (which I shall equally refer to as empty cells). Instead, we must write

$$c_1 = \rho_L(1-Y_{L-1})(1-\tilde{Y}_{L-2}), \tag{2}$$

where a subscript containing L will always denote the "width" of the system to which a quantity belongs. Namely, by inserting a solute at any position in PBCS one automatically creates a "pore" of width $L-1$ defined by an external field generated by solute "walls" fixed at either end. This class of pore (or confined system) is labeled (a) in Fig. 1. Isolating the monomer means that we must then insist that the cell immediately to the right of the left-hand wall is empty of solute. This is the second factor on the right of Eq. (2); with Y denoting the density next to a filled cell in a symmetric pore. In PBCS the empty cell to the left of the monomer is also an empty cell on the far side of a pore, which is now one unit shorter and of the asymmetric form depicted in (c) of Fig. 1. Hence the final factor in Eq. (2); a tilde will always be used to denote a density next to a wall of an asymmetric pore. For true clusters ($n>1$), we follow an identical process but add $n-1$ additional solutes successively to the right of the previous solute, before capping the cluster with two solvent cells. Thus

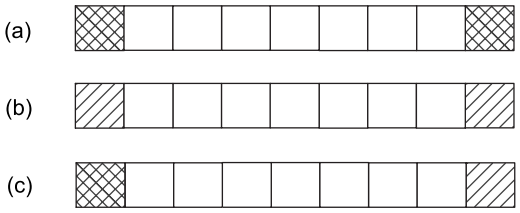

FIG. 1. Boundary conditions for confined one-dimensional LG models in the grand canonical ensemble. (a) Solute walls. (b) Solvent walls. (c) Asymmetrical boundary conditions.

$$c_n = \rho_L Y_{L-1} Y_{L-2} \cdots Y_{L-n+1}(1 - Y_{L-n})(1 - \tilde{Y}_{L-n-1}),$$
$$1 < n < L - 1; \tag{3}$$

when $n=L-1$ the final factor is dropped and finally for $n=L$ both solvent-cell factors are absent. In PBCS, c_L is a special case because it corresponds to a unique state rather than L states. In PBCS one should therefore check a calculated cluster distribution against the sum rule

$$1 = \frac{1}{\rho_L}\left\{ c_L + \sum_{n=1}^{L-1} n c_n \right\}. \tag{4}$$

Now note that the PDT can be regarded as an expression for the density $\rho_L(T,\mu)$, within the same PBCS framework, provided we also include confined systems of class (b) of Fig. 1:

$$\rho_L/(1-\rho_L) = z\{(1-\rho^w_{L-1})(1-\rho^w_{L-2}) + 2(1+a)\rho^w_{L-1}(1-\tilde{\rho}^w_{L-2}) + (1+a)^2\rho^w_{L-1}\tilde{\rho}^w_{L-2}\}, \tag{5}$$

where hereinafter I shall label a density next to a solvent wall with a superscript w.

To evaluate all of the above-defined densities and hence the cluster distribution, one can simply apply the PDT to cells adjacent to the pore walls (at fixed T,μ or equivalently, fixed a,z). From the structure of the PDT (1) it is convenient to work with quantities $\rho/(1-\rho) \equiv f$; i.e., $f_m \equiv Y_m/(1-Y_m)$ and similarly for $\tilde{f}_m$, f^w_m, $\tilde{f}^w_m$ for the various pores of width m. Starting with the limiting case $m=1$, where there are no many-body contributions to the PDT, just external field contributions from the pore walls, we have $f^w_1 = z$, $\tilde{f}_1 = \tilde{f}^w_1 = z(1+a)$, $f_1 = z(1+a)^2$. Increasing the pore "widths" one arrives straightforwardly at the following formulas:

$$f^w_m = \frac{z[1+(1+a)f^w_{m-1}]}{1+f^w_{m-1}}, \quad m > 1, \tag{6}$$

$$f_m = \frac{z(1+a)^2(1+f_{m-1})}{1+a+f_{m-1}}, \quad m > 1, \tag{7}$$

$$\tilde{f}_m = (1+a)f^w_m, \tag{8}$$

$$\tilde{f}^w_m = f_m/(1+a). \tag{9}$$

However, we have not as yet made sufficient use of conditional probability. For example, the middle term on the right of Eq. (5) has been expressed in terms of a solvent cell surrounded by a solute to the right and another solvent to the left (in PBCS), which could equally have been constructed in reverse order; i.e., $\rho^w_{L-1}(1-\tilde{\rho}^w_{L-2}) = (1-\rho^w_{L-1})\rho^w_{L-2}$. When combined with the first and last of the above relations, this results in the sum rule

$$f_m = (1+a)\left[z(1+a) - 1 + \frac{z}{f^w_m} \right], \tag{10}$$

which can also be readily proved directly from Eqs. (6) and (7) by induction and holds even at $m=1$. Thus there is only one independent pore density, which is conveniently taken to be f^w_m so that Eq. (10) replaces the need to make explicit use of the second recursion relation (7). Finally, one can use the above to confirm that at $z=1/(1+a)$ the density in PBCS is $\rho_L = 1/2$ for all values of L; which is a direct expression of the underlying Ising symmetry (corresponding to zero magnetic field).

Let us now turn to the main proposed application to computer simulation studies of linear self-assembly in three-dimensional canonical ensembles; namely, the effect on the cluster distribution of using a finite number of solute particles. Here, I shall make the ansatz that the activity z is controlled by the solute-solvent interactions in the three-dimensional system surrounding the aggregates and that the latter can be treated as if they were all lying along a line that ends in solvent "walls." This mapping of the one-dimensional LG model in the grand ensemble, to the three-dimensional simulation system in the canonical ensemble, is an obvious extension of the implicit solvent model. The mapping requires us to calculate the cluster distribution in a "pore" of class (b) of Fig. 1. The first thing to note is that the probability of finding a cluster of size n occupying a given site is now a function of the position of this cluster within the finite pore. That is, we have to find $c_n(x)$, where it remains convenient to continue our practice of defining the position x of the aggregate to be the cell occupied by the solute at the far left end of the aggregate. The first requirement is to calculate the density profile $\rho_L(x)$ or alternatively $f_L(x)$ as obtained directly from the PDT. In the cells neighboring one of the solvent walls, $x=1$ and L, this quantity has already been defined as f^w_L. For other values of x, the PDT in the pore evaluates to

$$f_L(x) = z\{(1-\rho^w_{L-x})(1-\rho^w_{x-1}) + (1+a)\rho^w_{L-x}(1-\rho^w_{x-1}) + (1+a)(1-\rho^w_{L-x})\rho^w_{x-1} + (1+a)^2\rho^w_{L-x}\rho^w_{x-1}\},$$
$$1 < x < L, \tag{11}$$

which is fully determined by the first recursion relation (6). Note that this formula and Eq. (5) apply to cases for which $L \geqslant 3$. The cluster distribution profiles can now be calculated from the PDT, proceeding exactly as we did for PBCS, but noting that an aggregate of length n cannot fit into the confined system unless $x \leqslant L-n+1$. The general result is

$$c_n(x) = \frac{f_L(x)}{1+f_L(x)} \frac{1}{1+\tilde{f}_{x-1}} \frac{\tilde{f}_{L-x}}{1+\tilde{f}_{L-x}} \frac{\tilde{f}_{L-x-1}}{1+\tilde{f}_{L-x-1}} \cdots \frac{\tilde{f}_{L-x-n+2}}{1+\tilde{f}_{L-x-n+2}} \frac{1}{1+\tilde{f}_{L-x-n+1}};$$
$$1 < n < L; \quad 1 < x < L-n+1. \tag{12}$$

The special cases that have to be added to this set are

$$c_1(1) = c_1(L) = \frac{f^w_L}{1+f^w_L} \frac{1}{1+\tilde{f}_{L-1}}, \tag{13}$$

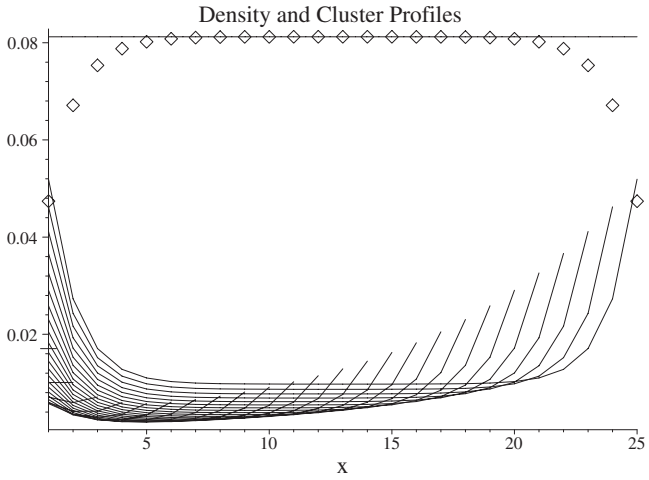

FIG. 2. Density and cluster distribution profiles for the finite-size LG model with solvent-wall boundary conditions; at solute activity $z=3/16$, solute-solute attractive well-depth $\epsilon=2.2k_BT$, and system size $L=25$. Symbols denote the density profile, reduced by a factor of 10, with the horizontal line denoting the bulk value ($L=\infty$). The curves show cluster distribution profiles $c_n(x)$ for $1 \leqslant n \leqslant L$. The larger the cluster length n, the smaller the effective value of L available to the cluster; the horizontal line centered on the ordinate denotes $c_L(1)$.

$$c_1(1 < x < L) = \frac{f_L(x)}{1+f_L(x)} \frac{1}{1+\tilde{f}_{x-1}} \frac{1}{1+\tilde{f}_{L-x}}, \tag{14}$$

$$c_{1<n<L}(1) = c_{1<n<L}(L+1-n)$$

$$= \frac{f_L^w}{1+f_L^w} \frac{\tilde{f}_{L-1}}{1+\tilde{f}_{L-1}} \cdots \frac{\tilde{f}_{L-n+1}}{1+\tilde{f}_{L-n+1}} \frac{1}{1+\tilde{f}_{L-n}}, \tag{15}$$

$$c_L(1) = \frac{f_L^w}{1+f_L^w} \frac{\tilde{f}_{L-1}}{1+\tilde{f}_{L-1}} \cdots \frac{\tilde{f}_1}{1+\tilde{f}_1}. \tag{16}$$

Note that the symmetry of the confined system is a good first check on the derivation of these relations. The quantity of immediate physical relevance is the average probability of finding an aggregate of size n anywhere in the confined system:

$$\bar{c}_n \equiv \frac{1}{L} \sum_{x=1}^{L-n+1} c_n(x), \tag{17}$$

which can be compared directly with the cluster distribution in the thermodynamic limit. A stringent test of the above formulas is to check them against the average density of the confined system:

$$\sum_{n=1}^{L} n\bar{c}_n = \frac{1}{L} \sum_{x=1}^{L} \rho_L(x). \tag{18}$$

In the finite size systems with solvent walls, the density and cluster distribution profiles reveal two, apparently related, relationships with well-known colloidal physics. Namely, solute avoids contact with the walls leaving an interfacial solvent-rich region, while in contrast the cluster distribution profiles are peaked at the walls (see Fig. 2). At first sight this may appear contradictory, but one should recall that a cluster at say the left boundary cannot move any further to the left, so that its contribution to the density profile at the wall is reduced by a factor of n over the contribution the same cluster would give to the bulk density in the thermodynamic limit. The avoidance of solute from the boundary walls is directly related to the physics of drying (wetting by

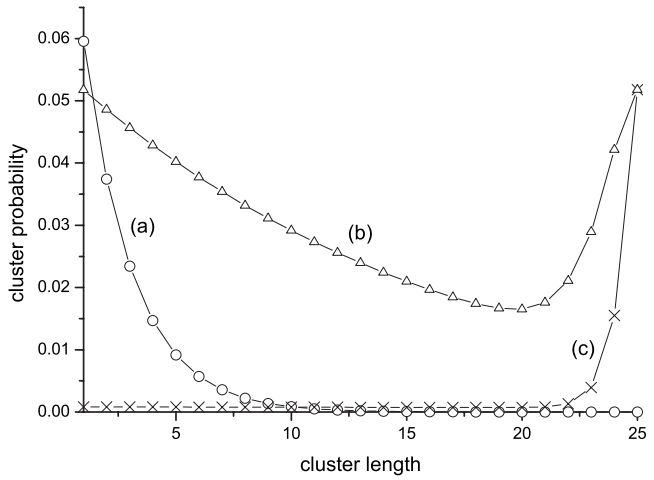

FIG. 3. Cluster probability distributions for the finite-size LG model with solvent-wall boundary conditions. In all cases the solute activity is fixed at $z=3/16$ and the system size is $L=25$. (a) $\bar{c}_n$ versus n at solute-solute attractive well depth $\epsilon=1.573k_BT$. (b) $10\bar{c}_n$ at $\epsilon=2.573k_BT$. (c) $2\bar{c}_n$ at $\epsilon=3.573k_BT$.

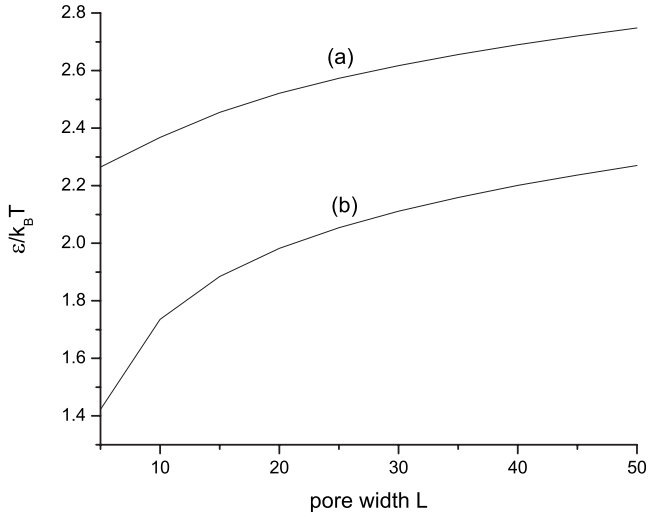

FIG. 4. Cluster probability "transition" lines at a solute activity of $z=3/16$. (a) The solute-solute well depth for crossover cluster distributions $\bar{c}_L=\bar{c}_1$ as a function of system size, from one-dimensional LG systems confined by solvent walls [class Fig. 1(b)]. Figure 3(b) corresponds to the value at $L=25$. Above the transition line the cluster distribution is dominated by the largest aggregates [see Fig. 3(c)] while below the line the distribution of aggregates tends rapidly to the isodesmic form [see Fig. 3(a)]. (b) The corresponding transition line under periodic boundary conditions; $c_L=c_1$.

vapor at the interface between a hard-wall and saturated liquid) because the implicit solvent model with solvent walls maps onto a single-component model with hard-wall boundaries. The fact that clusters prefer to sit at a hard wall is known in colloidal physics as the depletion effect. In particular, note that a cluster benefits from being next to say the left solvent wall because this removes the need to accommodate a solvent cell on the left of the cluster that is otherwise needed to separate it from the remaining solute. These drying and depletion effects decay smoothly from the boundaries, with apparently the same decay length, typically over about five solute diameters (LG cells). This is because there are no phase transitions in the one-dimensional Ising/LG model and so complete drying is unreachable. Figure 3 displays three representative cluster distributions in systems of "width" 25 bounded by solvent walls. Each system has the same solute chemical potential, chosen to map onto a typical liquid state of a simulation model. All that is varied is the nearest-neighbor attractive well depth or, equivalently, the temperature. In case (a), high temperature, the cluster distribution in the confined system is very close to that in the thermodynamic limit $L \to \infty$ and so displays the expected exponential character. In case (b) the temperature is low enough so that in an unconfined system there would be significant contributions from aggregate lengths out to about 100. Here, the confined system has responded by favoring both small clusters and large clusters, over intermediate length aggregates. The last example, (c), is at an even lower temperature where the cluster distribution would prefer to decay exponentially out to around aggregates of length 1000, but cannot because it is forced into a pore of width 25. Here, the system responds by favoring the largest possible aggregates over all others, the opposite to the behavior in the thermodynamic limit where the cluster distribution is isodesmic.

An interesting aspect of Fig. 3 is that case (b) corresponds to a quite moderate solute-solute attractive well depth ϵ and the other two distributions belong to systems that are identical to (b) apart from ϵ being greater or lower by exactly k_BT, respectively. This rapid "transition" between two contrasting cluster distributions (a) and (c) has arisen because linear self-assembly is extremely sensitive to temperature in the region of $k_BT \simeq \epsilon/2$. In this physically relevant regime, doubling the attraction strength increases the aggregation number in the thermodynamic limit from small to huge. The curve labeled (a) in Fig. 4 plots this "transition" as a function of confinement, with Fig. 3(b) belonging to $L=25$, and illustrates that this dramatic physics will always be relevant. Figure 4(b) shows a similar transition line for finite-size systems with periodic boundary conditions. In both discotic and peptide solutions, especially the latter where typically ϵ receives contributions from around 10 hydrogen bonds [2], the simulation will often correspond to states well above these transition lines. In which case, the one-dimensional LG model predicts that instead of observing the equilibrated exponential cluster distribution appropriate to the thermodynamic limit, one would instead observe a state of almost total polymerization [5].

[1] J. R. Henderson, J. Chem. Phys. **113**, 5965 (2000), and references therein.

[2] C. Whitehouse, J. Fang, A. Aggeli, M. Bell, R. Brydson, C. W. G. Fishwick, J. R. Henderson, C. M. Knobler, R. W. Owens, N. H. Thomson, D. A. Smith, and N. Boden, Angew. Chem., Int. Ed. **44**, 1965 (2005), and references therein.

[3] M. E. Cates, Macromolecules **20**, 2289 (1987); M. Kroger and R. Makhloufi, Phys. Rev. E **53**, 2531 (1996).

[4] M. Kroger, O. Peleg, Y. Ding, and Y. Rabin, Soft Matter **4**, 18 (2008).

[5] See for example, R. G. Edwards, J. R. Henderson, and R. L. Pinning, Mol. Phys. **86**, 567 (1995).

[6] B. Widom, J. Chem. Phys. **39**, 2808 (1963).