PHYSICAL REVIEW E **79**, 050801(R) (2009)

All-or-none proteinlike folding transition of a flexible homopolymer chain

Mark P. Taylor,[1] Wolfgang Paul,[2] and Kurt Binder[2]
[1]*Department of Physics, Hiram College, Hiram, Ohio 44234, USA*
[2]*Institut für Physik, Johannes-Gutenberg-Universität, Staudinger Weg 7, D-55099 Mainz, Germany*
(Received 11 February 2009; published 7 May 2009)

Here we report a first-order all-or-none transition from an expanded coil to a compact crystallite for a flexible polymer chain. Wang-Landau sampling is used to construct the complete density of states for square-well chains up to length 256. Analysis within both the microcanonical and canonical ensembles shows a direct freezing transition for finite length chains with sufficiently short-range interactions. This type of transition is a distinctive feature of "one-step" protein folding and our findings demonstrate that a simple homopolymer model can exhibit protein-folding thermodynamics.

DOI: 10.1103/PhysRevE.79.050801 PACS number(s): 64.70.km, 05.10.−a, 64.60.an, 87.15.Cc

Simple model systems have long played a key role in statistical physics in elucidating complex physical phenomena (consider the Ising model as a rich example); however, this standard reductionist approach seems to be severely limited in the realm of biological systems and processes. The core of this difficulty is the highly specialized and nonequilibrium nature of the biological world. However, some biological phenomena do seem amenable to analysis via simplified or "reduced" models. The problem of protein folding is one of these [1–4]. This physical process in which a protein chain undergoes a transition from an expanded disordered state to a compact ordered state is ubiquitous across the diverse range of biological systems and can in many cases be considered an equilibrium process [5,6].

In seeking a simple framework from which to view protein folding it is common to draw an analogy with the collapse transition observed in many simple homopolymers [6–8]. However, while these two transitions are superficially similar, such a comparison does not capture the essence of protein folding, which is a symmetry-breaking transition resulting in a compact ordered rigid final state [6]. In contrast, polymer collapse is a continuous transition resulting in a compact disordered liquidlike state [7] and thus resembles the formation of a molten globule that is observed for some proteins [6]. Such proteins undergo a multistep folding process, and the final transition from molten globule to native state can be compared to the freezing transition of a collapsed homopolymer chain [9–11]. Thus a flexible homopolymer chain provides a simple model for studying multistep protein folding [12].

However, a large number of small proteins fold in a single-step process with a very distinctive all-or-none character in which an unfolded protein chain undergoes a first-order transition directly to its native (i.e., ground) state conformation [13]. Although such a first-order transition is found in a number of simple lattice models (and this feature has made such minimalist models popular in the study of protein folding) [3,4,14,15], it has previously been argued that an "all-or-none" folding transition is not possible in a flexible continuum chain model [6,14,15]. In this Rapid Communication we report such a transition for an interaction-site polymer chain of finite length with sufficiently short-range interactions. While a direct freezing transition is known to exist in colloids with short-range interactions [16], it is not obvious that such a transition should occur in a simple polymer chain. Chain connectivity greatly reduces the entropy of any "expanded phase" thereby reducing the entropic driving force needed to produce a discontinuous expanded-to-compact transition [6].

Here we consider a simple off-lattice homopolymer chain comprised of N square-well-sphere monomers connected by "universal joints" of fixed length L. The square-well (SW) potential is given by

$$u(r_{ij}) = \begin{cases} \infty & 0 < r_{ij} < \sigma \\ -\varepsilon & \sigma < r_{ij} < \lambda\sigma \\ 0 & r_{ij} > \lambda\sigma, \end{cases} \tag{1}$$

where r_{ij} is the distance between nonbonded monomers i and j and σ and $\lambda\sigma$ are the hard-core and square-well diameters, respectively. We use the well depth ε to define a dimensionless temperature $T^* = k_B T/\varepsilon$, where k_B is the Boltzmann constant, and we take the bond length to be $L=\sigma$. Since we analyze a single chain, the above potential is to be considered an effective potential that includes the effect of solvent (although here we do not consider any temperature dependence of this potential). The square-well chain has a discrete energy spectrum $E_n = -n\varepsilon$, where n is the number of square-well overlaps in the chain configuration.

To study the phase behavior of this single chain we use the Wang-Landau (WL) algorithm [17] to construct the density of states $g(E_n)$, from which all thermodynamic properties can be obtained. In the WL approach one generates a sequence of chain configurations via a set of Monte Carlo (MC) moves; however, rather than accepting new configurations with the standard Metropolis criteria, one uses

$$P_{acc}(a \to b) = \min[1, w_b g(E_a)/w_a g(E_b)], \tag{2}$$

where w_a and w_b are weight factors insuring detailed balance of the particular underlying MC move. If the density-of-states function $g(E_n)$ were known, this procedure would yield a random walk through all accessible energy states of the system. In the WL method $g(E_n)$ is constructed in an iterative fashion where after each MC move the $g(E_n)$ for the current configuration is updated by a modification factor $f_m > 1$ via $g(E_n) \to f_m g(E_n)$ and a state-visitation histogram $H(E_n)$ is simultaneously updated via $H(E_n) \to H(E_n)+1$. One

periodically checks this histogram for "flatness" (i.e., all entries are within p percent of the average value), indicating an approximate equal visitation to all energy states. When flatness is achieved, the modification factor is reduced to $f_{m+1}=\sqrt{f_m}$, $H(E_n)$ is reset to zero for all states, and the $(m+1)$st level of the iteration is begun. Due to bottlenecks in configuration space the WL method occasionally leads to a highly asymmetric $H(E_n)$ which would require a prohibitively large amount of time to relax. To overcome this problem, we also periodically check for uniform growth of $H(E_n)$ (i.e., the increase in all entries is within p percent of the average increase since the last check) and use this as an alternate convergence criterion for a WL level. In the present work we assume an initial $g(E_n)=1$ for all states, take $f_0=e^1$ and $p=20$, check for flatness and uniform growth every 10^4 and 5×10^7 MC cycles, respectively, and continue the iterative process until $m\geq26$ $(f_m-1\leq10^{-8})$.

A single MC cycle consists of, on average, the following set of randomly selected move attempts: N single-bead crankshaft moves, two reptation moves, one pivot move, and N end-bridging moves [18]. For this latter move, which is found to be essential for sampling compact low-energy states for longer chains, we attempt to connect end-site 1 (N) to a nearby interior site i via deletion and reinsertion of site $i-1$ $(i+1)$. The weight factor for this bond-bridging move is $w_a=b_aJ_a$, where b_a is the number of possible bridging sites i present in state a and $J_a=1/r_{1i}$ $(1/r_{iN})$ is a Jacobian factor arising from the fixed bond-length restriction [18]. Weight factors for the other MC moves are all unity. We have verified the WL algorithm with this MC move set by direct comparison with exact $g(E_n)$ results for short $(N\leq6)$ SW chains for a range of λ [19] and by comparison with the Metropolis MC results of Zhou *et al.* [10] for $N\leq64$ SW chains with $\lambda=1.5$. For longer chains we use overlapping energy windows [20], and in all cases we carry out a preliminary WL sampling without a low-energy cutoff to estimate the ground-state energy and determine a lower E_n bound to ensure computational feasibility. For longer chains this lower bound is typically within a few percent of the ground-state energy. We have carried out a minimum of three independent WL runs for all cases and find $g(E_n)$ to be highly reproducible except in a small energy range where crystal nucleation occurs. This slight difference has minimal effect on thermodynamic properties and resulting transition temperatures agree to within $\delta T^*\approx0.003$.

In Fig. 1 we show density-of-states results, plotted as the microcanonical entropy $S(E)/k_B=\ln g(E)$, for SW chains of length $N=256$ with square-well diameters $\lambda=1.05$, 1.10, and 1.15. The approximate ground-state energies (and minimum energy used in the WL sampling) are $-E/\varepsilon=968$ (950), 975 (956), 991 (965), respectively. Note that $g(E)$ for $\lambda=1.05$ covers approximately 840 orders of magnitude. The curvature of the microcanonical entropy function can be used to directly determine the phase behavior of a finite-size system [21,22]. The signature of a discontinuous (first order) phase transition is a "convex intruder" in $S(E)$ (which produces a Maxwell-type "loop" in the microcanonical temperature $T(E)=(\partial S(E)/\partial E)^{-1}$) while a continuous (second-order) transition is signaled by the presence of an isolated inflection point in $T(E)$ (which becomes a saddle point in the thermodynamic limit). Features of both types of transitions are seen in Fig. 1 for all three well diameters. However, for $\lambda=1.05$ the continuous phase transition is located *within* the coexistence region of the first-order transition and therefore will not occur (except under metastable conditions) [16]. Thus the $N=256$, $\lambda=1.05$ SW chain undergoes a direct freezing transition without an initial collapse transition. A similar result is found for the $\lambda=1.05$ chain with $N=128$, whereas for $N=64$ two distinct transitions are found.

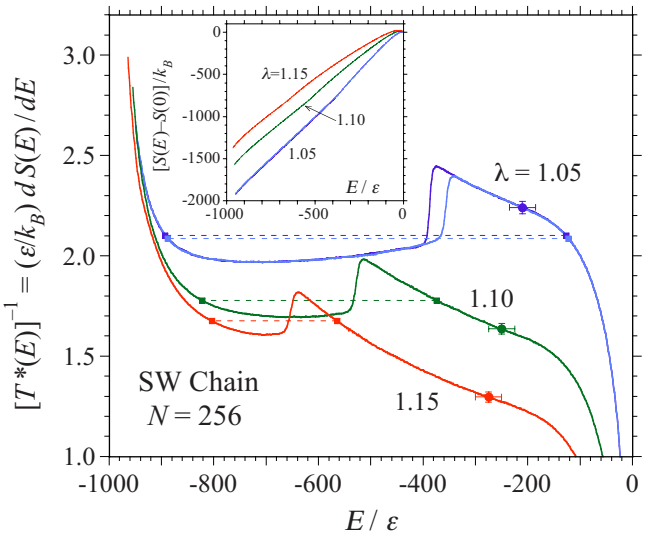

FIG. 1. (Color online) Derivative of the microcanonical entropy $dS(E)/dE$ vs energy E for a SW chain of length $N=256$ and SW diameter $\lambda\sigma$ as indicated. The Maxwell-type "loops" indicate discontinuous phase transitions and the dashed tie-lines, obtained via an equal areas construction, give the transition temperatures and energies of the coexisting states. The filled circles locate the positions of isolated inflection points in $T(E)$ which signal the presence of a continuous phase transition. For $\lambda=1.05$ we show results from the two WL runs with the largest disparity. Here the transition temperatures differ by $\Delta T^*=0.003$. Inset: Microcanonical entropy $S(E)$ vs energy E.

To further characterize this direct freezing transition we show in Fig. 2 the average chain size, given by the mean-square radius of gyration $\langle R_g^2\rangle_E=\frac{1}{N^2}\sum_{i<j}^{N}\langle r_{ij}^2\rangle_E$, versus energy E. For the short-range interaction potential $(\lambda=1.05)$, the coexistence bounds of the first-order freezing transition are seen to connect a high-energy state populated by expanded chain conformations with a low-energy state of very compact chains (which visual analysis shows to be crystallites with hexagonal order). In contrast, for the longer-range interaction potentials the coexisting high- and low-energy states at freezing are both comprised of compact chain conformations (with liquidlike and crystallite structure, respectively).

To make a more direct link with experimental results we also analyze these single chain phase transitions within the canonical ensemble. In this case the energy probability distribution, at canonical temperature T, is given by

$$P(E,T)=g(E)e^{-E/k_BT}\Big/\sum_E g(E)e^{-E/k_BT}, \qquad (3)$$

and canonical ensemble results are obtained by averages over this probability function. Thus, for example, in the Fig. 2

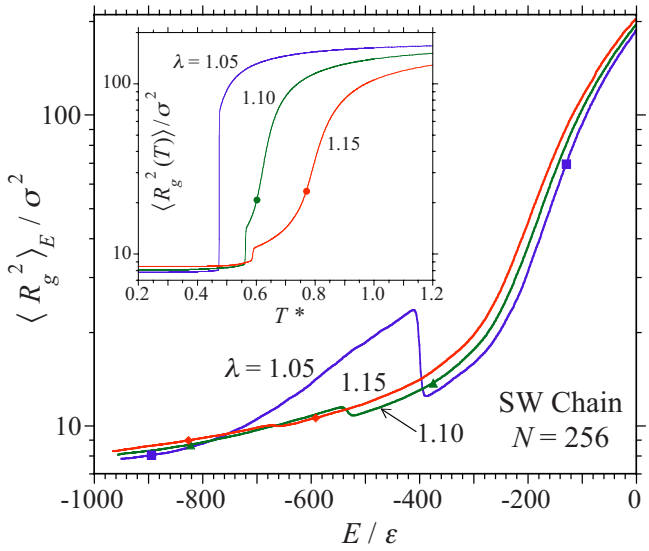

FIG. 2. (Color online) Microcanonical mean-square radius of gyration $\langle R_g^2\rangle_E$ vs energy E for a SW chain of length $N=256$ and SW diameter $\lambda\sigma$ as indicated. The symbols mark the coexisting energies for the discontinuous (freezing) transition. The steps in these functions correspond to the formation of partially crystalline transition state structures. Inset: canonical mean-square radius of gyration $\langle R_g^2(T)\rangle$ vs temperature T^*. Symbols locate the continuous (coil-globule) transition and vertical sections locate the freezing transition.

inset we show the canonical mean-square radius of gyration $\langle R_g^2(T)\rangle=\Sigma_E\langle R_g^2\rangle_E P(E,T)$ versus temperature T. With decreasing temperature these curves all show a smooth reduction in chain dimensions interrupted by a fairly abrupt decrease in chain size at the freezing transition. As is evident from these curves, chain-size itself does not provide a clear experimental signature for the presence or absence of a coil-globule transition.

To locate phase transitions in the canonical ensemble one typically examines the heat capacity $C(T)=d\langle E\rangle/dT$, where

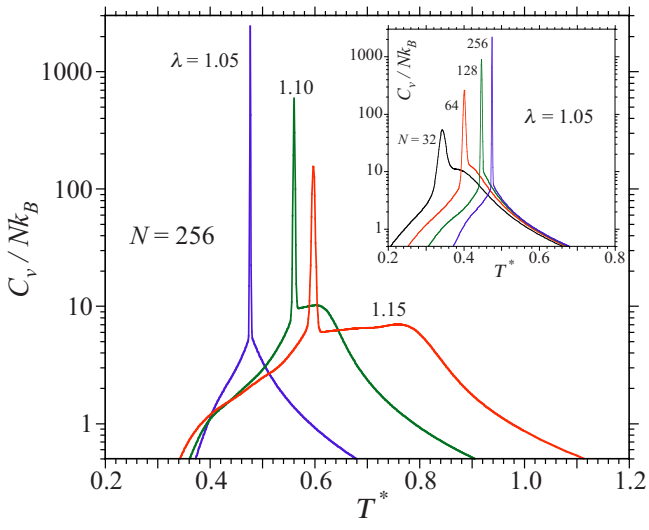

FIG. 3. (Color online) Specific heat per monomer C/Nk_B vs temperature T^* for a SW chain of length $N=256$ and SW diameter $\lambda\sigma$ as indicated and, in the inset, for $\lambda=1.05$ and N as indicated.

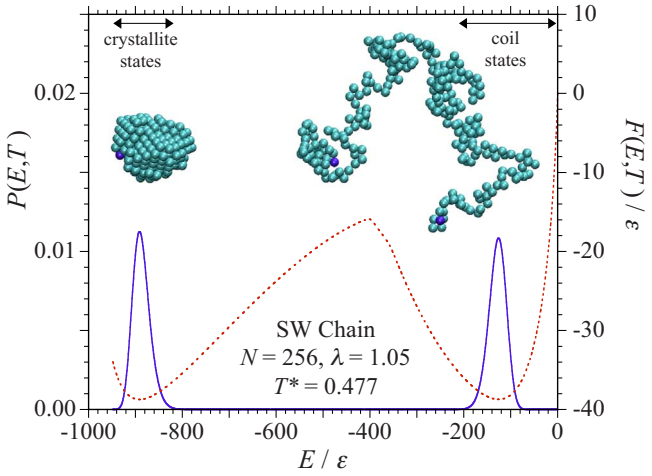

FIG. 4. (Color online) Potential-energy probability distribution $P(E,T)$ (solid line, left scale) and free energy $F(E,T)=E-TS(E)$ (dashed line, right scale) vs energy E for a SW chain with $N=256$ and $\lambda=1.05$ at freezing. The bimodal distribution is equally weighted at the transition. The simulation snapshots show representative chain conformations for the most probable energies $-E/\varepsilon=891$ and 126.

$\langle E\rangle=\Sigma_E EP(E,T)$. In Fig. 3 we show $C(T)$ for $N=256$ SW chains with a range of λ and for SW chains with $\lambda=1.05$ for a range of N. The sharp low-temperature spike, seen in all of these curves, locates the chain freezing transition, while the high-temperature peak/shoulder, found for larger λ and smaller N, locates the polymer collapse (i.e., coil-globule) transition [9,10]. This latter feature is absent for $N=128$ and 256 with $\lambda=1.05$, consistent with our microcanonical analysis. Similarly, transition temperatures obtained from the locations of the maxima in $C(T)$ all agree with the microcanonical results. Note, however, that it is only through the microcanonical analysis that we can definitively confirm the absence of a coil-globule transition since in the canonical analysis this latter transition may be present but masked by the nearby $C(T)$ freezing peak.

A distinguishing feature of a first-order phase transition in the canonical ensemble is a bimodal $P(E,T)$ distribution near the transition. As seen in Fig. 4, the $N=256$, $\lambda=1.05$ chain exhibits such a bimodal distribution at the freezing transition ($T^*=0.477\pm0.002$). The simulation snapshots in Fig. 4 illustrate the coexistence of an expanded coil state with a compact crystallite of hexagonal structure. A free-energy barrier separates these coexisting states and the peak of this barrier, shown in Fig. 4, coincides with the step in $\langle R_g^2\rangle_E$ seen in Fig. 2. Both of these features are associated with the formation of a "folding nucleus" (i.e., a small crystallite attached to one or more unfolded chain segments).

A direct first-order transition from an ensemble of expanded coil states to a compact ordered state is a distinguishing feature of single-step protein folding [6]. To make some contact with the protein folding problem we compare our results for the direct freezing of the $N=128$, $\lambda=1.05$ SW chain with typical experimental results for the folding of a small (~100 amino acid) protein. We assume an equilibrium folding temperature of 333 K (60 °C) which sets the SW depth to $\varepsilon=k_BT/T^*\approx1.5$ kcal/mol (where $T^*=0.446$).

This is roughly equivalent (same integrated well area) to a Lennard-Jones potential with a well depth of 0.17 kcal/mol, similar to values used in atomistic modeling of proteins [23]. The equilibrium energy (enthalpy) difference between denatured and native states for a small protein is ~ 1 kcal/mol/residue [6] compared with $\Delta E/N \approx 300\varepsilon/128 = 3.5$ kcal/mol/site for the $N=128$ SW chain. The thermal folding of a small protein typically spans $\sim 10-20$ K [6,15] compared with $\Delta T^* \approx 0.01$ or ~ 8 K for the SW chain (see Fig. 3) and $C(T)$ for this model satisfies the experimental calorimetric two-state criterion of $\Delta H_{vH} \approx \Delta H_{cal}$ [15]. A typical equilibrium free-energy barrier for a single-step folder is in the range 5–15 kcal/mol [6,8] comparable with the SW chain value of $\Delta F_{barrier} \approx 8\varepsilon \approx 12$ kcal/mol. An Arrhenius analysis of this barrier height for the SW chain (i.e., a plot of $\Delta F_{barrier}/k_B T$ vs $1/T$) shows linear behavior for both folding and unfolding resulting in a Chevron plot characteristic of proteins [6,15]. The "stability" or room-temperature free-energy difference between native and denatured states for a small protein is typically 2–18 kcal/mol [1,6,13] compared to $24\varepsilon = 36$ kcal/mol for the SW chain. This limited set of comparisons shows that our homopolymer model has proteinlike thermodynamics. Of course a SW chain is far from being a realistic model for a protein. However, the fact that we find close correspondence between the freezing of a flexible homopolymer chain and the folding of a small protein suggests that a chain with sufficiently short-range interactions may serve as a good starting point for the development of simple continuum models (perhaps analogous to the widely studied HP lattice models [14]) for single-step protein folding. The short-range potential provides a model with a very low-entropy ground state (see Fig. 1) analogous to the low-entropy native state of a protein.

Finally, the results presented here have more general implications for the phase behavior of polymer solutions. In particular, we confirm the recent prediction based on a lattice polymer model of the disappearance of a collapsed globule phase for a chain with sufficiently short-ranged interactions [20,24]. A finite-size scaling analysis of the present model shows that the $(T$-$\lambda)$-phase diagram for a long SW chain exhibits a tricritical point near $\lambda = 1.15$ where the continuous coil-globule and the first-order globule-crystal transitions merge into a single first-order coil-crystal transition. This behavior is analogous to the disappearance of the liquid phase for a SW fluid with $\lambda \leq 1.25$ [25].

We thank David Landau, Jutta Luettmer-Strathmann, and Peter Virnau for helpful discussions. Financial support through the Deutsche Forschungsgemeinschaft (Grant No. SFB 625/A3) and the National Science Foundation (Grant No. DMR-0804370), and a sabbatical leave from Hiram College are also gratefully acknowledged.

[1] K. A. Dill *et al.*, Annu. Rev. Biophys. **37**, 289 (2008).
[2] Y. Chen *et al.*, Arch. Biochem. Biophys. **469**, 4 (2008).
[3] A. Kolinski and J. Skolnick, Polymer **45**, 511 (2004).
[4] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).
[5] C. B. Anfinsen, Science **181**, 223 (1973).
[6] A. V. Finkelstein and O. B. Ptitsyn, *Protein Physics* (Academic Press, London, 2002).
[7] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* (AIP Press, New York, 1994).
[8] S. J. Hagen and W. A. Eaton, J. Mol. Biol. **297**, 781 (2000).
[9] Y. Zhou, C. K. Hall, and M. Karplus, Phys. Rev. Lett. **77**, 2822 (1996).
[10] Y. Zhou *et al.*, J. Chem. Phys. **107**, 10691 (1997).
[11] M. P. Taylor, J. Chem. Phys. **114**, 6472 (2001).
[12] Y. Zhou and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **94**, 14429 (1997); A. Voegler Smith and C. K. Hall, J. Chem. Phys. **113**, 9331 (2000); H. Jang, C. K. Hall, and Y. Zhou, Biophys. J. **82**, 646 (2002).
[13] S. E. Jackson, Folding Des. **3**, R81 (1998); K. W. Plaxco *et al.*, Biochemistry **39**, 11177 (2000).
[14] K. A. Dill *et al.*, Protein Sci. **4**, 561 (1995).
[15] H. S. Chan, Proteins **40**, 543 (2000); H. Kaya and H. S. Chan, *ibid.* **40**, 637 (2000); **52**, 510 (2003).
[16] A. P. Gast, C. K. Hall, and W. B. Russel, J. Colloid Interface Sci. **96**, 251 (1983); M. H. J. Hagen and D. Frenkel, J. Chem. Phys. **101**, 4093 (1994).
[17] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001); Phys. Rev. E **64**, 056101 (2001).
[18] F. A. Escobedo and J. J. de Pablo, J. Chem. Phys. **102**, 2636 (1995).
[19] M. P. Taylor, J. Chem. Phys. **118**, 883 (2003).
[20] F. Rampf, K. Binder, and W. Paul J. Polym. Sci., Part B: Polym. Phys. **44**, 2542 (2006).
[21] D. H. E. Gross, *Microcanonical Thermodynamics* (World Scientific, Singapore, 2001).
[22] C. Junghans, M. Bachmann, and W. Janke, Phys. Rev. Lett. **97**, 218103 (2006); J. Hernandez-Rojas and J. M. Gomez Llorente, *ibid.* **100**, 258104 (2008).
[23] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[24] F. Rampf, W. Paul, and K. Binder, Europhys. Lett. **70**, 628 (2005); W. Paul, T. Strauch, F. Rampf, and K. Binder, Phys. Rev. E **75**, 060801(R) (2007).
[25] D. L. Pagan and J. D. Gunton, J. Chem. Phys. **122**, 184515 (2005); H. Liu, S. Garde, and S. Kumar, *ibid.* **123**, 174505 (2005).