Dominant Kinetic Paths on Biomolecular Binding-Folding Energy Landscape

Jin Wang,[1,2,*,†] Kun Zhang,[1] Hongyang Lu,[1] and Erkang Wang[1,*,‡]

[1]*State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin 130021, People's Republic of China*

[2]*Department of Chemistry and Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794, USA*

(Received 20 September 2005; published 28 April 2006)

The identification of kinetic pathways is a central issue in understanding the nature of flexible binding. A new approach is proposed here to study the dynamics of this binding-folding process through the establishment of a path integral framework on the underlying energy landscape. The dominant kinetic paths of binding and folding can be determined and quantified. In this case, the corresponding kinetic paths of binding are shown to be intimately correlated with those of folding and the dynamics becomes quite cooperative. The kinetic time can be obtained through the contributions from the dominant paths and has a U-shape dependence on temperature.

DOI: 10.1103/PhysRevLett.96.168101 PACS numbers: 87.15.−v

Biomolecular recognition is an important issue in modern molecular biology [1–3]. The conventional wisdom is that structure determines the functions. In other words, the molecules have a well-defined shape before interacting with each other and keep rigid during the binding process. However, more and more experimental evidence has shown that some biomolecules cannot be stable alone without binding to the partners [4–7]. This implies that significant folding occurs as the binding proceeds. Therefore binding and folding are intimately coupled. Here flexibility rather than rigidity is crucial for binding as well as for biological function [4–7].

To understand the interplay between binding and folding dynamically, one needs first to have a good description of the corresponding degrees of freedom. One way to do that is to employ atomic detailed calculations. This way of doing it often runs into the trouble of not being able to sample enough of the configurational space. On the other hand, it is possible to use a phenomenological approach by identifying the quasireaction coordinate or order parameter mimicking the binding and folding process. The approach used here attempts to study the binding-folding phenomena common in nature with at least two order parameters, Q_b and Q_f. Here Q_b is the fraction of native binding spatial contacts and Q_f is the fraction of the native folding spatial contacts (see Fig. 1 from binding of two proteins where one protein is rigid but the interface and the other protein is flexible). This minimal representation is used to study the thermodynamics of the binding-folding process [7]. It is found that the folding and binding processes are often intimately coupled in nature. The crucial question one needs to address is how the dynamics actually occurs.

Identifying the important dynamic flow of paths that the binding complex takes to reach the native state is crucial in uncovering the fundamental kinetic mechanisms of the binding-folding process and has been a central issue in the experimental community [8]. So far, very limited efforts have been put on the actual kinetic binding intermediate process or the identification of kinetic paths connecting the initial and final states [9,10]. We will quantify the kinetic paths for the flexible binding by developing a path integral formalism. Path integral formulations have been developed successfully in studying many different areas in physics and chemistry [11–14]. The advantage of this approach is that it addresses the fundamental issues of kinetic pathways directly. The paths can be identified and quantitatively determined (See Fig. 2).

Another important question is related to how the many possible degrees of configuration could fall to the unique native state basin. The most natural and simple way of resolving this so-called Levinthal paradox [15] is that the underlying energy landscape should be funneled to guarantee both the thermodynamic stability and specificity [1,2,16,17]. This should also lead to faster kinetics [18]. Under this funneled energy landscape, in general there is a dominant flow of paths towards the native state basin. Thus the kinetics can be obtained through studying the behavior of the paths. By approximating the path integral using a dominant-path approach, we will describe with realistic parameters estimated from the current available data the

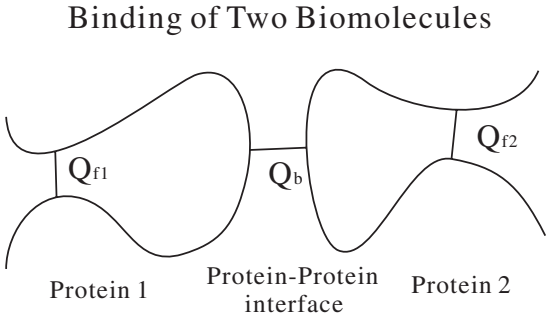

FIG. 1. Illustration of flexible binding (Q_b) coupled with folding or large conformational change (Q_{f1}, Q_{f2}).

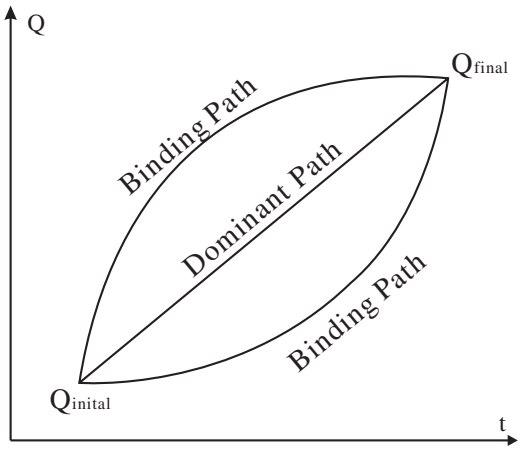

FIG. 2. Possible kinetic binding paths from initial to final configuration (Q_i, Q_f) and time.

dynamics and the degrees of cooperativity in the binding-folding process. The kinetic time scale can be obtained by summing over the appropriate weighted contributions from the dominant paths.

To proceed, let us briefly review the formulation of the thermodynamic energy landscape of the binding-folding process [2,7]. We first start with the energy function of a polypeptide chain with an interface. $H^f = \sum \epsilon_{ij}^f \sigma_{ij}^f$ and $H^b = \sum \epsilon_{ij}^b \sigma_{ij}^b$ and total energy is $H = H^f + H^b$, where H^f and H^b are the energy functions of polypeptide chain energy (folding) and interface contacts (binding). The ϵ_{ij}'s are the contact-energy strengths, while σ_{ij} is the contact variable equal to 1 when there is a spatial contact and zero when there is no spatial contact within a cutoff distance.

By employing the microcanonical ensemble and thermodynamic relationships, we can easily obtain the free energy of the system: $F(Q_f, Q_b) = N\delta E_f Q_f + CN\delta E_b Q_b - NS_0(Q_f, Q_b)T - N\frac{\Delta E_f^2(1-Q_f)(1+\gamma_f Q_f)}{2kBT} - CN\frac{\Delta E_b^2(1-Q_b)(1+\gamma_b Q_b)}{2kBT}$, where N is the number of the amino-acid residues, δE_f is the energy gap or bias towards the native folded state, δE_b is the energy gap or bias towards the native binding state, and ΔE_f is the roughness or spread of the folding energy while ΔE_b is the roughness or spread of the binding energy. γ_f and γ_b are the inhomogeneity coefficients for folding and binding. C is the scaling constant measuring the relative strengths of binding versus folding. S_0 is the entropy of the configurations $S_0 = \ln \Omega$.

The entropy function can be fitted with a simple function by noticing that the entropy of the completely native folding and binding state $S(1, 1)$ is zero and the entropy of completely unfolded and unbinding state is $S(0, 0)$; the entropy of native folded but completely unbinding state is $S(1, 0)$ and the entropy for completely unfolded and native binding state is $S(0, 1)$. These quantities can all be estimated. So the entropy function has a form through the interpolation of these know entropies: $S_0(Q_f, Q_b) = (1 - Q_f)(1 - Q_b)S(0, 0) + Q_f(1 - Q_b)S(1, 0) + Q_b(1 - Q_f)S(0, 1)$.

Under the free-energy profiles, the equation of motion for native contact vector $\mathbf{Q} = (Q_f, Q_b)$ formation can be formulated as $d\mathbf{Q}/dt = -\frac{\partial \beta \mathbf{F}(\mathbf{Q})}{\partial \mathbf{Q}} + \eta$. Because of the long time scale, the binding-folding motion are overdamped. Therefore the second derivatives of $\mathbf{Q}$ with respect to time t may be ignored. Here $-\partial \beta \mathbf{F}(\mathbf{Q})/\partial \mathbf{Q}$ is the gradient force that the motion of $\mathbf{Q}$ vector would follow and η is the noise term assumed to be Gaussian and uncorrelated in time (white) for simplicity. The correlation of the noise is given by $\langle \eta(\mathbf{Q}, t)\eta(\mathbf{Q}, 0)\rangle = 2\mathbf{D}(\mathbf{Q})\delta(t)$. The $\mathbf{D}(\mathbf{Q})$ is the $\mathbf{Q}$-dependent diffusion coefficient tensor (or matrix). The binding-folding process has many degrees of freedom; therefore when looking at the motion along the reduced two-dimensional order parameter or reaction coordinate $\mathbf{Q}$, there is an effective noise or friction force from the rest of the other dimensions.

We can now formulate the dynamics for the probability of starting from initial configuration Q_{initial} at $t = 0$ and end at the final configuration of Q_{final} at time t, with the Onsager-Machlup functional [13] as $P(\mathbf{Q}_{\text{final}}, t, \mathbf{Q}_{\text{initial}}, 0) = \int D\mathbf{Q} \exp\left[-\int dt \left(\frac{1}{4}\frac{\{(d\mathbf{Q}/dt)+[\mathbf{D}(\mathbf{Q})\partial\beta F(\mathbf{Q})/\partial\mathbf{Q}]\}^2}{\mathbf{D}(\mathbf{Q})} - \frac{1}{2}\frac{\partial \mathbf{D}(\mathbf{Q})[\beta F(\mathbf{Q})/\partial \mathbf{Q}]}{\partial \mathbf{Q}}\right)\right] = \int D\mathbf{Q} \exp[-\int L(\mathbf{Q}(t))dt]$.

The integral over $D\mathbf{Q}$ represents the sum over all possible paths connecting $\mathbf{Q}_{\text{initial}}$ at time $t = 0$ to $\mathbf{Q}_{\text{final}}$ at time t. The exponential factor gives the weight of each path. So the probability of binding-folding dynamics from nonnative configurations $\mathbf{Q}_{\text{initial}}$ to native configuration $\mathbf{Q}_{\text{final}}$ is equal to the sum of all possible paths with different weights. The $L(\mathbf{Q}(t))$ is the Lagrangian or the weight for each path (Fig. 2).

Notice that not all the paths give the same contribution. We can approximate the path integrals with a set of dominant paths. Since each path is exponentially weighted, the other subleading path contributions are often small and can be ignored. One can easily use this observation to find the paths with the optimal weights. The dominant paths should satisfy the Euler-Lagrangian equation (see Fig. 2), $\frac{d}{dt}\frac{\partial L}{\partial \dot{\mathbf{Q}}} - \frac{\partial L}{\partial \mathbf{Q}} = 0$, and the resulting equation becomes $\ddot{\mathbf{Q}} - \frac{1}{2}\frac{\partial \mathbf{D}(\mathbf{Q})/\partial \mathbf{Q}}{\mathbf{D}(\mathbf{Q})}\dot{\mathbf{Q}}^2 - 2\mathbf{D}(\mathbf{Q})\frac{\partial V(\mathbf{Q})}{\partial \mathbf{Q}} = 0$, where $V(\mathbf{Q}) = \frac{\partial \beta F(\mathbf{Q})}{\partial \mathbf{Q}}\frac{\mathbf{D}(\mathbf{Q})}{4}\frac{\partial \beta F(\mathbf{Q})}{\partial \mathbf{Q}} - \frac{\mathbf{D}(\mathbf{Q})}{2}\frac{\partial^2 \beta F(\mathbf{Q})}{\partial \mathbf{Q}^2} - \frac{1}{2}\frac{\partial \mathbf{D}(\mathbf{Q})}{\partial \mathbf{Q}}\frac{\partial \beta F(\mathbf{Q})}{\partial \mathbf{Q}}$. The equation of motion of Q has the acceleration term $\ddot{Q}$, the frictional (positive and negative) term $\frac{1}{2}\frac{\partial \mathbf{D}(\mathbf{Q})/\partial \mathbf{Q}}{\mathbf{D}(\mathbf{Q})}\dot{\mathbf{Q}}^2$, and the force term $2\mathbf{D}(\mathbf{Q})\frac{\partial V(\mathbf{Q})}{\partial \mathbf{Q}}$. Define $-\partial U(\mathbf{Q})/\partial \mathbf{Q} = 2\mathbf{D}(\mathbf{Q})\partial V(\mathbf{Q})/\partial \mathbf{Q}$. Then the problem becomes one of a two-dimensional particle moving in a potential well U with friction.

When $\mathbf{D}(\mathbf{Q})$ is a constant, the friction term is zero. For simplicity, we assume the diffusion coefficient tensor matrix is diagonal with only two elements (D_{ff} and D_{bb}) present, while the nondiagonal elements to account for the

kinetic coupling between Q_b and Q_f are zero ($D_{fb} = D_{bf} = 0$).

We can also write out explicitly the equations of motion in the scalar form as $\ddot{Q}_f - 2D_{ff}\frac{\partial V}{\partial Q_f} = 0$ and $\ddot{Q}_b - 2D_{bb}\frac{\partial V}{\partial Q_b} = 0$. Notice that the frictional term becomes zero under the current assumption of a Q-independent diffusion coefficient.

By solving these two equations with initial points of $Q_f = Q_b = 0$ and end points at $Q_f = Q_b = 1$, we can obtain the dominant-path contribution to the weight of the paths. By substituting the dominant-path solution back into the path integral formulation, we can obtain the expression for the time of the kinetic process from non-native states to native state.

We have used the results of the bioinformatics survey of database with 500 binding complexes (the two sets of parameters) to infer the kinetic mechanism of binding folding [7]. Our choices of the parameters are as follows: Set I parameters are for the average proteins often more stable, and set II parameters are for less stable and more floppy proteins. The energy gap of folding and binding are given as $\delta E_f = -11.8$ kJ/mol, $\delta E_b = -9.3$ kJ/mol (set I) and $\delta E_f = -10.3$ kJ/mol, $\delta E_b = -12.9$ kJ/mol (set II) The other related parameters for inhomogeneity coefficients γ, unfolding and unbinding entropy $S(0, 0)$, binding entropy alone $S(1, 0)$, folding entropy alone $S(0, 1)$, the roughness or variance of the energy, and diffusion coefficients D are the same for both set I and set II: $\gamma_f = 1.0$, $\gamma_b = 0$, $S(0, 0) = 3 + \frac{9.7}{200}k_B$, $S(1, 0) = \frac{9.7}{200}k_B$, $S(0, 1)/S(0, 0) = 0.75$, $\Delta E_f = 3.4$ kJ/mol, $\Delta E_b^2 = 3.4$ kJ/mol, $D_{ff} = D_{bb} = 1/\mathrm{s}$, $C = 0.2$ [7].

The dominant kinetic paths are shown in Fig. 3 for different free-energy landscapes with parameter sets I and II. The underlying landscapes are all downhill and funneled towards the native state.

We can see clearly that for the parameter set I for the more-stable proteins, the kinetic process proceeds with a significant fraction of folding initially and then proceeds with the completion of the binding process. The path is less diagonal. The folding and binding are not very strongly coupled. On the other hand, with the parameter set II for the more floppy proteins, we see that significant binding occurs first and then proceeds with folding and binding together towards the native state. The path is more diagonal. So in this case, the folding and binding process are more cooperatively coupled together. Parameter sets I and II give different underlying landscapes which determine different degrees of coupling between folding and binding. The recent experiments on single molecule flexible binding [19] show clearly that there are distinct conformational states of bound and loosely bound states corresponding to native binding and partially binding states with less well-defined structures. This implies strong coupling between binding and folding during the binding complex formation. Our theory and kinetic path picture qualitatively provides an explanation for that.

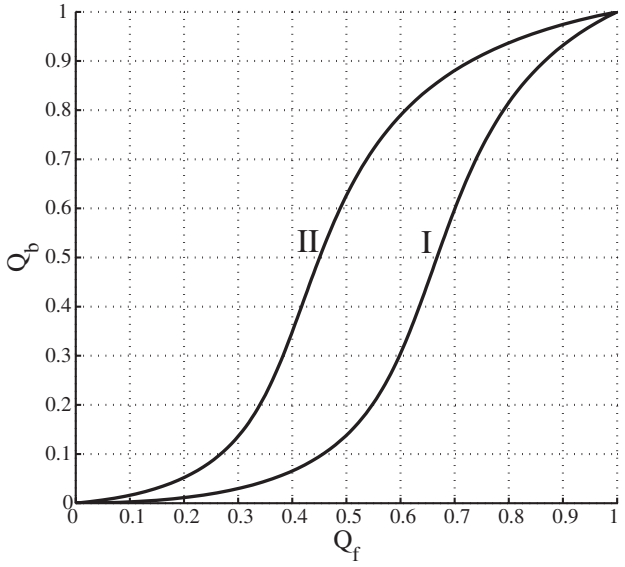

FIG. 3. Free-energy profile and dominant kinetic paths with respect to Q_f and Q_b for more-stable proteins (parameter set I), and free-energy profile and dominant kinetic paths with respect to Q_f and Q_b for more floppy proteins (parameter set II).

The kinetics of binding show complex temperature dependence in the folding and binding experiments [8,20]. Very often, the U-shape dependence of the kinetic time on temperature is found (the Chevron rollover) [8,20,21]. By varying temperature, the underlying energy landscape structures can be probed [8]. The kinetic time for binding is plotted in Fig. 4 versus temperature. The kinetic time is shown to have a U shape, similar to the experimental Chevron phenomena for folding [21]. At high temperatures, the native state is unstable so the kinetic time increases with temperature. At low temperatures, local trapping becomes possible, so the kinetic time decreases as temperature decreases. This explains why the time has a U-like shape.

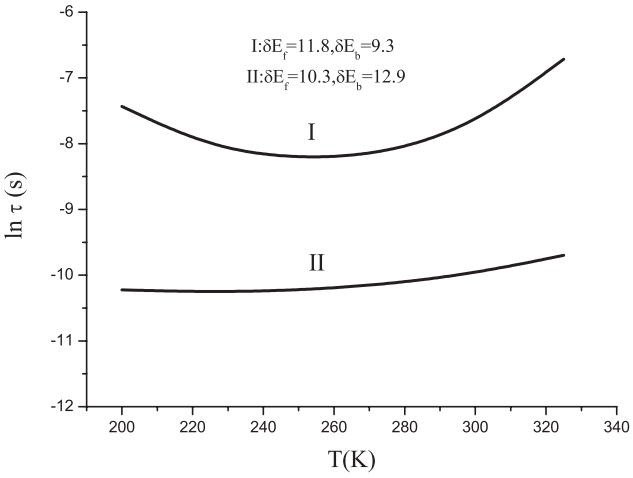

FIG. 4. The kinetic time of binding folding versus temperature.

Our formalism applies to folding too when we freeze the binding degree of freedom [22]. The folding alone will also have the Chevron behavior [21]. Usually it is sharper than binding. This is because the binding tends to facilitate folding through the cooperative interactions. The effect of cooperative interactions tends to make the rollover shallower than folding alone [21,23]. From another perspective, folding represents large conformational change. It can speed up binding compared with the more rigid or stable one (Fig. 4). The difference between folding and binding is mainly the connectivity since the underlying dominant driving force is quite similar (hydrophobic interactions). Folding can be seen as self-binding. One expects qualitatively similar although quantitatively different kinetic behavior versus temperature.

It is worth pointing out that binding experiments with wide temperature ranges are hard to perform. More kinetic experiments are needed and important for exploring the mechanisms of binding-folding coupling. Furthermore, we took the average parameters for the quantification of landscape from the bioinformatics survey [7]. The individual protein binding complex can be quite different and shows different shapes of underlying free-energy landscapes. As a result, the rollover in Fig. 4 can be different. The rollover may become sharp for some protein binding complexes while broad for others. Experimentalists need to choose the system with care to see this kinetic rollover behavior. We are working towards an atomic detailed model now to study temperature dependence of kinetics for specific proteins instead of the qualitative averaged one here.

In Fig. 4, we can also see that the kinetic process is faster for flexible binding (parameter set II). The more-stable folding implies that the binding process starts first with significant folding and then proceeds with binding. So it is essentially a rigid-binding process. The more flexible binding implies that significant binding starts first and induces the folding. The binding and folding are thus intimately coupled together. So as we have shown here the flexible binding (binding folding coupled together) has a kinetic advantage (faster) over rigid binding (folding first and then binding). Binding with large conformational changes helps to reach the kinetic specificity rather than the rigid one. This is due to the larger capture radius for the flexible binding. It is analogous to fly casting in fishing [7]. This suggests a new set of experiments to test the mechanism of flexible binding.

It is worth pointing out that the binding involves two molecules and the reaction is bimolecular. Kinetics is in general concentration dependent. Our study here should be considered as the saturation limit where the concentration becomes constant. Another effect of nonadditive cooperative interactions from solvents can also be incorporated into the current formalism. These are topics for future study.

J. W. would like to thank Professor Peter Wolynes, Professor Jose Onuchic, Professor Andy McCammon, and Professor George Stell for helpful discussions. K. Z., H. L., and E. W. are supported by the National Science Foundation of China (20575963). J. W. is supported by the National Science Foundation (USA), Petroleum Research Fund.

*Corresponding author.
†Electronic address: jin.wang.1@stonybrook.edu
‡Electronic address: ekwang@ciac.jl.cn

[1] P. A. Rejto and G. M. Verkhivker, Proc. Natl. Acad. Sci. U.S.A. **93**, 8945 (1996); J. Janin, Proteins: Struct., Funct., Genet. **25**, 438 (1996); C. J. Tsai *et al.*, Protein Science **8**, 1181 (1999).
[2] J. Wang and G. M. Verkhivker, Phys. Rev. Lett. **90**, 188101 (2003).
[3] J. A. McCammon, Curr. Opin. Struct. Biol. **8**, 245 (1998).
[4] D. E. Koshland, Jr., Proc. Natl. Acad. Sci. U.S.A. **44**, 98 (1958).
[5] H. J. Dyson and P. E. Wright, Curr. Opin. Struct. Biol. **12**, 54 (2002).
[6] A. K. Dunker *et al.*, Biochemistry **41**, 6573 (2002).
[7] B. A. Shoemaker, J. J. Portman, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **97**, 8868 (2000); G. A. Papoian and P. G. Wolynes, Biopolymers **68**, 333 (2003).
[8] H. Frauenfelder, F. Parak, and R. D. Young, Annu. Rev. Biophys. Biophys. Chem. **17**, 451 (1988); H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, Science **254**, 1598 (1991).
[9] M. Berkowitz and J. A. McCammon, J. Chem. Phys. **75**, 957 (1981).
[10] R. Olender and R. Elber, J. Chem. Phys. **105**, 9299 (1996); R. Elber *et al.*, J. Phys. Chem. B **103**, 899 (1999).
[11] N. Wiener, *Generalized Harmonic Analysis and Tauberian Theorems* (MIT Press, Boston, 1964).
[12] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[13] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953); P. Hanggi, Z. Phys. B **75**, 275 (1989).
[14] K. L. C. Hunt and J. Ross, J. Chem. Phys. **75**, 976 (1981).
[15] C. Levinthal, *Proceedings in Mossbauer Spectroscopy in Biological Systems* (University of Illinois Press, Urbana, 1969), p. 22.
[16] R. A. Goldstein, Z. A. Luthey-Schulten, and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **89**, 4918 (1992).
[17] C. J. Tsai, D. Xu, and R. Nussinov, Folding Des. **3**, R71 (1998).
[18] C. L. Lee, G. Stell, and J. Wang, J. Chem. Phys. **118**, 959 (2003); Y. Zhou, C. Zhang, G. Stell, and J. Wang, J. Am. Chem. Soc. **125**, 6300 (2003).
[19] X. Tan *et al.*, J. Phys. Chem. B **108**, 737 (2004).
[20] L. S. Itzhaki *et al.*, J. Mol. Biol. **254**, 260 (1995); B. Kuhlman *et al.*, J. Mol. Biol. **284**, 1661 (1998); H. Nguyen *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 3948 (2003); H. Yang *et al.*, Science **302**, 262 (2003).
[21] H. Kaya and H. S. Chan, Phys. Rev. Lett. **85**, 4823 (2000); J. Mol. Biol. **315**, 899 (2002).
[22] J. Wang, K. Zhang, H. Lu, and E. K. Wang, Biophys. J. **89**, 1612 (2005).
[23] J. Wang, C. Lee, and G. Stell, Chem. Phys. **316**, 53 (2005).