Diffusive dynamics of protein folding studied by molecular dynamics simulations of an off-lattice model

A. Baumketner* and Y. Hiwatari
Faculty of Science, Kanazawa University, Kakuma, Kanazawa, 920-1192, Japan
(Received 12 March 2002; published 12 July 2002)

We report the results of a molecular dynamics study on the kinetic properties of a small off-lattice model of proteins. The model consists of a linear chain of monomers interacting via a number of potentials. These include hydrophobic, bond-angle, and torsion potentials. The ground-state conformation of the studied model is a β-sheet motif. Molecular dynamics simulations focused on the time evolution of the reaction coordinate measuring the similarity of a given conformation with the native state. Folding time for the studied model is calculated following the diffusive-rate formula of Bryngelson and Wolynes [J. Phys. Chem. **93**, 6902 (1989)] by using a computed separately configurational diffusion coefficient. Comparison of the folding time with the mean-first passage time obtained directly from folding simulations shows that the approximation depicting the dynamics of the reaction coordinate in protein folding as a diffusive motion on a free-energy landscape is quantitatively correct for the studied model.

DOI: 10.1103/PhysRevE.66.011905 PACS number(s): 87.15.Cc, 87.15.Aa, 82.39.−k

I. INTRODUCTION

Within the framework of the energy landscapes theory, folding of a protein can be regarded as a stochastic motion on a multidimensional statistically averaged free-energy surface [1]. The protein free-energy surface has a funnel-like structure when defined in terms of a few appropriately chosen reaction coordinates or order parameters. Although the funnel multidimensionality may prove important for larger proteins, a recent study focused on lattice protein models of the smaller proteins [2] revealed that a single reaction coordinate suffices in most cases to reasonably describe their folding dynamics. Protein motion on the energy landscape can be regarded to a first approximation as diffusive. The reaction coordinate then obeys the Brown equation of motion characterized by a single parameter, the configurational diffusion constant D [3]. Coefficient D along with the free-energy profiles are the two factors that fully determine folding time within the diffusive dynamics formalism through the known diffusive-rate formula [3]. Being just an approximation, the analytical diffusion-equation theory of the protein folding has a solid physical foundation deriving from the random energy model [4]. It was also shown to be correct in recent numerical tests [5] that employed simulations performed by the Monte Carlo method for a lattice protein model. The lattice simulations, however, suffer from at least two well-known problems that make interpretation of their results difficult. First, due to the steric constraints present in discretized spaces, closely packed structures of protein monomers may not be adequately represented in lattice models. This situation was observed by Tanaka *et al.* in molecular dynamics simulations of polyampholytes [6]. Second, there is an ambiguity in common approaches of how to map Monte Carlo moves onto physical time. As was demonstrated by Chan and Dill, kinetic sequences of events, as well as free energy profiles, simulated on a two-dimensional homopolymer lattice strongly depend on the adopted move set [7]. In view of these difficulties it is interesting to test the applicability of the physically motivated and robust idea of the diffusive dynamics by using more realistic methods and models.

In this paper we report a molecular dynamics study on the kinetics of a short off-lattice model of proteins. The basic architecture of our model was borrowed from the model introduced by Honeycutt and Thirumalai [8], which turned out to be quite successful recently [9]. Protein is modeled as a linear chain of monomers placed at the positions of C_α. Monomers can occupy any point in the phase space thus making the chain devoid of the steric constraints problem. Also, the second problem quoted above is no longer present since within the molecular dynamics method (in this paper Langevin dynamics) there is no difficulty in interpreting the simulation time steps.

Our protein model can contain either hydrophobic residues that attract each other via a Lennard-Jones potential or neutral monomers that interact via repulsive soft-core force only. A total of 16 monomers connected by fixed virtual bonds were considered; four of them are neutral and all the rest are assumed to be hydrophobic. Aside from the hydrophobic force other interactions operating among the monomers are bond-angle and torsion potentials. Ground-state conformation of the designed heteropolymer, or the native state, is a β-sheet motif. For the model we computed free-energy profiles along a reaction coordinate χ measuring the extent of similarity between a given conformation and the native state as a function of temperature. Diffusion coefficient D was computed directly from the Brown equation for the reaction coordinate by integrating the time correlation function of the force acting on variable χ at time t and the value of this variable at time zero $\chi(0)$. Availability of both energy landscapes and configurational diffusion constant enabled us to calculate folding time numerically due to the

*Permanent address: Institute for Condensed Matter Physics, 1 Svientsitsky Str., Lviv 79011, Ukraine.
Email address: andrij@icmp.lviv.ua

analytical formula of Bryngelson and Wolynes [3]. Comparison of this result with the mean-first passage time computed directly in folding simulations shows that the diffusion-equation approximation to the dynamics of the reaction coordinate produces a quantitatively correct prediction of the folding time for the studied protein model. At temperatures slightly higher than the folding temperature T_f, the diffusion-equation generated folding time τ_f that differs from that obtained in folding simulations by no more than a factor of 2. At $T<T_f$ the two folding times almost coincide.

The plan of the paper is as follows. In Sec. II we briefly describe the protein model used in the present simulations. Section III details our kinetical results obtained for the studied model and in Sec. IV we give the final conclusions.

II. OFF-LATTICE PROTEIN MODEL

To simulate the folding dynamics we consider a simple off-lattice protein model based on the model introduced by Honeycutt and Thirumalai [10]. Over the last few years this model has been shown to reproduce satisfactorily most basic aspects of folding kinetics as well as thermodynamics [9]. The protein is considered to be a linear chain of monomers placed at the positions of C_α and linked one to another by bonds of fixed length. The monomers can be either hydrophobic (H) or neutral (N) depending on whether they attract each other or repel. Among a total of 16 monomers considered four are neutral and the rest are hydrophobic. Hydrophobic monomers experience mutual attraction described by a Lennard-Jones potential,

$$V_H(r_{ij})=4e_h\left[\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \quad |i-j|\geqslant 2, \tag{1}$$

where r_{ij} is the distance between monomers i and j and the parameter σ was chosen to be equal to the bond length between neighboring monomers, 3.8 Å, for simplicity. Pairs of neutral and neutral-hydrophobic monomers are taken to interact via a repulsive soft-core potential:

$$V_{H,N}(r_{ij})=4e_h\left(\frac{\sigma}{r_{ij}}\right)^{12}, \quad |i-j|\geqslant 2. \tag{2}$$

In the simulations the strength of the hydrophobic force e_h was adopted 2 [kcal/mol], which is in good agreement with experiment [11]. Throughout the paper we will use the unit of e_h to measure energy and that of e_h/k_b to measure temperature, where k_b is the Boltzmann constant.

In addition to the hydrophobic force, all monomers are also subject to the harmonic bond-angle potential:

$$V_B(\Theta)=\frac{k_\Theta}{2}(\Theta_0-\Theta)^2, \tag{3}$$

and torsion potential:

$$V_T(\phi)=A(1-\cos\phi)+B(1-\cos 3\phi). \tag{4}$$

Here Θ is the angle formed by two consecutive virtual bonds and Θ_0 is the equilibrium value of the bond angle set to be 75°. The bond-angle parameter k_Θ was taken to be 20 $[e_h/\mathrm{Rad}^2]$. By ϕ we denote the angle between the two planes formed by three consecutive molecule's bonds. This potential has three minima favoring one *trans* conformation at $\phi=0°$ and two *gauche* conformations at $\phi=\pm 120°$. In regions on the chain where a turn is supposed to form, we put two neutral monomers and set the constants $A=0$ and $B=0.5$ $[e_h]$ for the quartets comprising these two monomers to make the energy of the *trans* and *gauche* conformations equal. Otherwise, we chose $A=B=1.5$ $[e_h]$ in order to give an energy advantage to the *trans* state over its *gauche* counterparts. The monomer sequence of the designed model is $H_3NH_3N_2H_3NH_3$. Its ground state, or native conformation, is a β-sheet motif, as shown in Fig. 1.

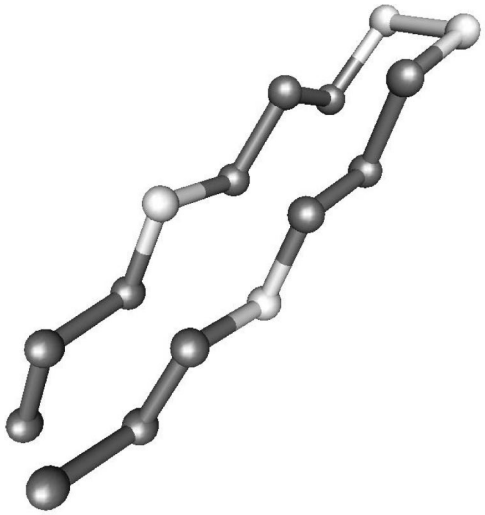

FIG. 1. Native state of the β-sheet model considered in the present study. Dark balls refer to the hydrophobic residues and the light denote the neutral residues. Picture was generated with the help of the GOPENMOL program [12].

For the model under consideration a series of Langevin dynamics simulations [13] aimed to study thermodynamical properties were performed at a few selected temperatures. The use of the Langevin dynamics method in the present work is motivated by the fact that this algorithm generates trajectory in the canonical ensemble and, more importantly, provides a means to account for the influence of the solvent degrees of freedom through the stochastic term. The latter property is most desirable in studies that aim to treat protein dynamics. The friction coefficient taken in the equations of motion was about ten times lower than the friction experienced by an alanine molecule placed in water at room temperature [14]. By employing a lower friction coefficient we were able to considerably diminish the required computational time without introducing any bias to the final results; it is known that the folding kinetics in minimal off-lattice models depends linearly on the solvent viscosity η in media with η as high as that of water [15].

During the simulations, histograms of potential energy and joint histograms of potential energy and an order parameter χ were collected. Here the structural overlapping parameter χ, defined as [16]

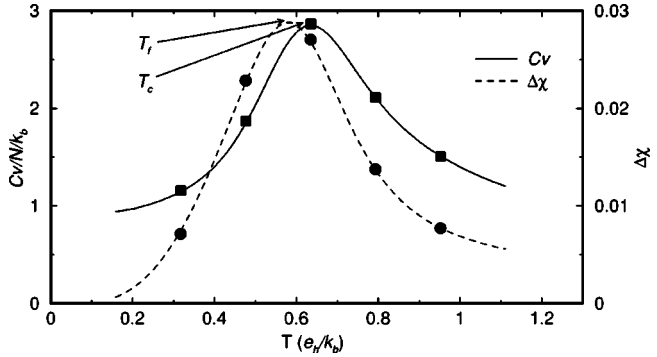

FIG. 2. Specific heat and susceptibility of the β-sheet motif studied in the present work. Symbols denote the results of the canonical simulations performed at selected temperatures. Lines show the data calculated by the multiple histogram reweighting method [18].

$$\chi = \frac{2}{(N-1)(N-2)} \sum_{i=1}^{N-2} \sum_{j=i+2}^{N} \Theta(\epsilon - |r_{ij} - r_{ij}^N|), \quad (5)$$

measures the extent of similarity between a given monomer conformation $\{\vec{r}_i\}$ and the native state $\{\vec{r}_i^N\}$. Parameter ϵ, taken to be 0.2σ in our simulations, accounts for the permissible fluctuations around the native state such that the protein is still considered folded. Θ in the above expression denotes the Heaviside step function. It is easily seen that the parameter χ is a generalization into continuous space of the fraction-of-native contacts order parameter Q, commonly used in lattice studies [3,17]. For folded states at natured conditions, χ approaches unity, while for denatured states it goes to zero. By using the multiple histogram reweighting technique [18] we computed from the accumulated in the simulations data the specific heat C_v and the susceptibility function of the order parameter $\Delta\chi = \langle\chi^2\rangle - \langle\chi\rangle^2$. In Fig. 2 we show Cv and $\Delta\chi$ as a function of temperature. The relevance of these two functions to the thermodynamics analysis rests on the means they provide for the identification of the structural transitions that occur in the present model. Specifically, the peak position of specific heat T_c indicates the well-known heteropolymeric collapse transition. This transition is accompanied by a rapid decrease of the overall size of the molecule, as measured, for example, by the radius of gyration. Another transition identified from Fig. 2 takes place at a temperature T_f given by the position of the maximum of susceptibility. This is a so-called folding transition that signifies the protein structural change from a multitude of nonspecific collapsed states at $T > T_f$ into a set of a few conformations with very high similarity to the native conformation at $T < T_f$. It was shown earlier by simulations that the minimum time required for a protein to fold is found in the vicinity of T_f [19–21].

III. RESULTS AND DISCUSSION

Following Bryngelson and Wolynes [3] we assume that the dynamics of the structural overlap function χ is governed by the Brown equation,

$$\dot{\chi}(t) = \beta D F[\chi(t)] + \delta R(t), \quad (6)$$

where $\beta = 1/k_b T$, D is the configurational diffusion coefficient, and $\delta R(t)$ is the stochastic force exerted on the degree of freedom χ. Here we assume a further simplification to the original diffusion equation of Ref. [3] by neglecting the dependence of D on the reaction coordinate χ. It was noted earlier that this approximation is rather crude [5], especially at high values of χ. Nevertheless, we use it in the present work partially in order to test how it works for the present model and partially because there is no straightforward way of evaluating $D(\chi)$ directly from simulations. The regular force that appears in Eq. (6) is related to the free-energy profile $U(\chi)$ along the reaction coordinate as

$$F[\chi] = -\frac{\partial U(\chi)}{\partial \chi}. \quad (7)$$

The free-energy funnel is easily available in simulations from the distribution function $P(\chi)$ of the order parameter χ as $U(\chi) = -k_b T \log[P(\chi)]$. Stochastic force δR in Eq. (6) is modeled as a Gaussian noise with a zero mean and variance $\langle \delta R^2 \rangle = 2D$. δR taken at a moment t is neither correlated with its value at any previous time nor is it correlated with the reaction coordinate χ at any time, including time t. By multiplying Eq. (6) by $\chi(0)$ and taking statistical average we can derive an equation for the time autocorrelation function $\Phi(t) = \langle \chi(t)\chi(0) \rangle$. Integration of this equation yields an expression for the diffusion constant,

$$\beta D = -\frac{\Phi(+\infty) - \Phi(0)}{\int_0^\infty \Psi(t)dt} = -\frac{\Delta\chi}{\int_0^\infty \Psi(t)dt}, \quad (8)$$

where we defined the time correlation function of the regular force at time t and the reaction coordinate at time zero as $\Psi(t) = \langle F[\chi(t)]\chi(0) \rangle$. Equation (8) provides a means to compute configurational diffusion from numerical simulations. Once the distribution function of the order parameter $P(\chi)$ at a given temperature is known, the time correlation function $\Psi(t)$ can be easily evaluated. It is worth noting here that since the only input quantity to Eq. (8) is the trajectory of $\chi(t)$, the above approach of computing D is equally applicable to both off-lattice minimal models of proteins as well as to fully microscopic macromolecular models that take into account every atom of the system. In Fig. 3 we display the time correlation function $\Psi(t)$ calculated for the present protein model at a number of different temperatures. All the curves in this figure have negative initial values, which means that the force acting on the reaction coordinate and the reaction coordinate itself are anticorrelated. This is understandable since the generalized force $F(\chi)$ tends to restore equilibrium whenever the variable χ deviates from its mean value $\langle\chi\rangle$ given by the condition $F(\langle\chi\rangle) = 0$. Hence, $F[\chi]$ and χ must have opposite signs. Figure 3 shows that functions $\Psi(t)$ calculated at all temperatures above and below T_f have long-decaying tails. These slowly decaying tails are specific signatures of arrested dynamics that commonly

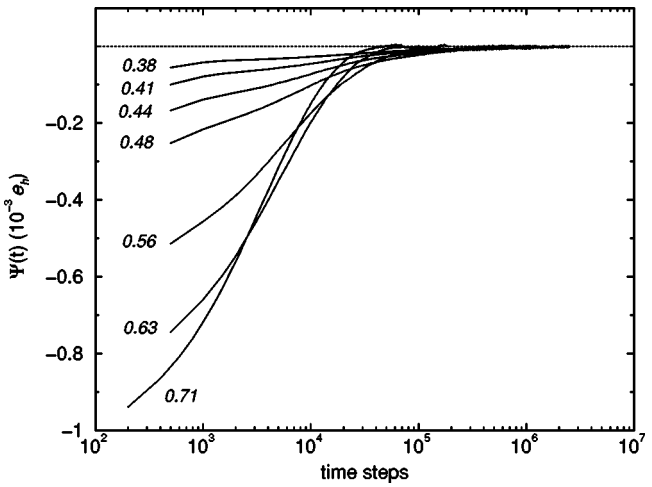

FIG. 3. Time correlation function $\Psi(t)$ of the regular force acting on variable χ with this variable itself, computed as a function of temperature. Number next to each curve denotes the corresponding temperature.

arise in disordered materials, particularly in supercooled liquids and glasses [22]. They are an interesting subject for a theoretical study especially in the context of the protein folding problem. More detailed study of the relaxation functions $\Psi(t)$ and $\Phi(t)$ will be presented elsewhere [23] and here it is enough to note that the relaxation rate of $\Psi(t)$ decreases gradually with temperature. That means the absolute value of the integral from the denominator of Eq. (8) grows with T. Using the time correlation functions shown in Fig. 3 and following Eq. (8) we calculated the configurational diffusion constant D. It is shown in Fig. 4 as a function of inverse temperature β. The diffusion coefficient is a monotonically decreasing function of β, thereby reflecting the fact that it is more difficult for the molecule to change conformations at a lower temperature, where the dynamic mechanism of trapping in local minima starts to be prevalent. At sufficiently low temperatures $T<T_f$, configurational dynamics clearly obeys the Arrhenius law, as can be seen from Fig. 4. At $T>T_f$, the dependence of D on temperature becomes non-

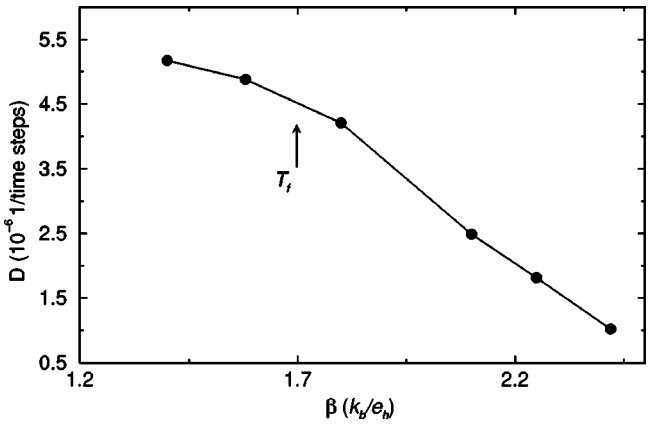

FIG. 4. Configurational diffusion coefficient D computed for the studied protein model as a function of the inverse temperature.

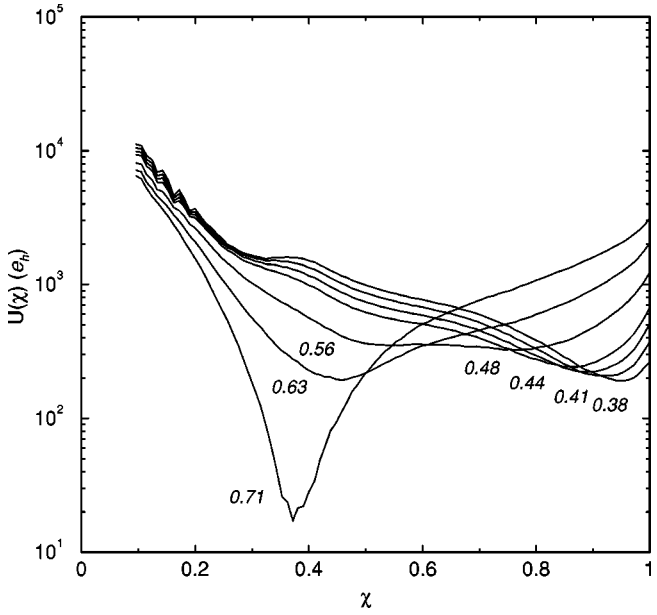

FIG. 5. Free-energy profiles computed for the studied protein model at varying temperatures. The data were obtained by applying the histogram reweighting routine [18].

Arrhenius. This behavior is in good agreement with the previous simulation study of diffusive protein dynamics performed for lattice models [5].

The configurational diffusion coefficient describes the ruggedness of the energy landscapes or, in other words, how freely the chain can jump among local minima and make transitions among strata of conformational subspaces that have, common degree of similarity to the native state. Another factor that influences the dynamics of the reaction coordinate is the energy landscape itself, or more specifically, its gradient. The free-energy profile determines the systematic tendency of χ to drift towards the global minimum. Combined with the information on D it predicts folding time for a diffusive motion on the rugged energy surface according to a formula derived by Bryngelson and Wolynes [3]

$$\tau_f = \frac{1}{D}\int_{\chi_{unf}}^{\chi_{fol}} dx \int_0^x dy e^{\beta[U(x)-U(y)]}, \tag{9}$$

where χ_{unf} is the structural overlap in the unfolded, and χ_{fol} in the folded ensembles. In our simulations we took a value of $\chi=0.3$ to be characteristic of unfolded states and that of $\chi=0.9$ to represent fully folded molecules. The free-energy profiles computed for the present protein model by using the histogram reweighting technique [18] are shown in Fig. 5. For temperatures above $T_f=0.59$, the free-energy curve has one minimum that corresponds with the unfolded state. Folding in this case is uphill and thus requires a considerable time. At temperatures around the folding temperature, the free-energy surface has two minima at low and high values of χ with each minimum corresponding to the folded and unfolded states, respectively. This type of the free-energy surface indicates that the folding transition in the present model is a first-order like. The free-energy barrier that the

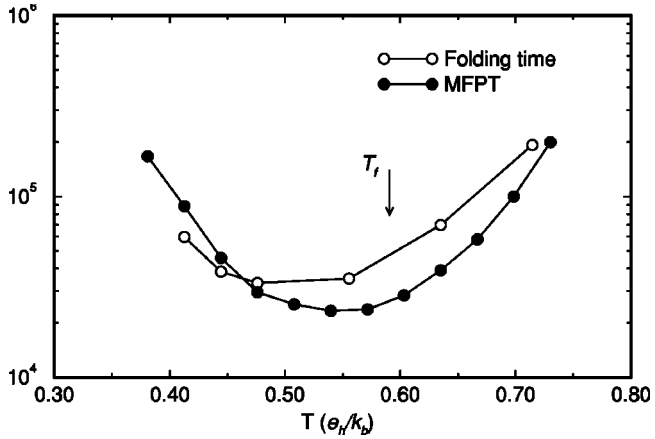

FIG. 6. Folding time (in time steps) calculated due to formula (9) and the mean-first passage time computed for the studied model in folding experiments.

protein has to overcome on its route to the native state is small, however, up to k_bT in magnitude at the transition temperature T_f. It is unlikely that this small barrier may serve as a time-limiting step in the folding of the present model. More plausibly, folding in our case is controlled exclusively by the diffusive process at $T \sim T_f$. At low temperatures the folding reaction is entirely downhill as can be seen in Fig. 5 and the folding time is again determined by D alone.

By numerical integration of Eq. (9) we calculated the folding time for the present model of β protein. It is shown in Fig. 6 along with the mean-first passage time τ_{MFP}. The mean-first passage time was calculated as follows. From a simulation at temperature $T=1$ an ensemble of 500 monomer configurations was generated over sufficiently long time intervals to avoid statistical correlation. At this high temperature the structural overlap χ is about 0.3, i.e., sufficiently small. The chain then populates mostly expanded conformations and we can assume that the set of these conformations reproduces the unfolded ensemble of the protein. From each of these initial states a folding simulation at the target temperature was initiated. The simulation was halted as soon as χ reached its maximum value 1, and the time of the first passage was recorded. Final value for τ_{MFP} was averaged over all 500 independent trajectories. To check the accuracy of the final result the array of the initial configurations was doubled to reach 1000 and the whole simulation process repeated. The resulting τ_{MFP} did not differ visually from the one shown in Fig. 6. It can be seen from Fig. 6 that the analytical formula (9) produces an overall very good agreement with the result of the direct simulation. The curve of $\tau_f(T)$ reproduces quite well the main trends of the mean-first passage time, especially slowing down at low and high temperatures. At high temperatures $T > T_f$, the values of both τ_f and τ_{MFP} differ by no more than two times. At lower temperatures the agreement is even better, where the two folding times almost coincide. The conclusion that we are naturally led to draw here is that the diffusive dynamics on the free-energy funnels picture devised by Bryngelson and Wolynes for proteins is quantitatively correct when applied to the present protein model. This is a rather surprising result, given all the serious approximations made to derive Eq. (5), especially the negligence of the diffusion coefficient dependence on χ. Nevertheless, Fig. 6 clearly shows that the dynamics of the reaction coordinate can be satisfactorily described by the Brown equation (6).

IV. CONCLUSIONS

In this paper we applied the molecular dynamics method to study kinetical properties of a small off-lattice protein model. We focused on the dynamics of the reaction coordinate χ that measures similarity between a given conformation and the native state. By using direct evaluation of the configurational diffusion constant D we have shown that the dynamics of χ can be satisfactorily described by the diffusive Brown equation, as suggested by Bryngelson and Wolynes [3]. Folding time calculated from the diffusive-rate formula and the mean-first passage time calculated in simulations directly agree to within a factor of 2 over a wide range of temperatures. Particularly good agreement is observed at lower temperatures, below the folding transition.

ACKNOWLEDGMENT

One of us (A.B.) benefited from discussions with P. G. Wolynes when this work was carried out.

[1] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[2] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, Proteins: Struct., Funct., Genet. **32**, 136 (1998).
[3] J. D. Bryngelson and P. G. Wolynes, J. Phys. Chem. **93**, 6902 (1989).
[4] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U.S.A. **84**, 7524 (1987).
[5] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
[6] M. Tanaka, A. Y. Grosberg, V. S. Pande, and T. Tanaka, Phys. Rev. E **56**, 5798 (1997).
[7] H. S. Chan and K. A. Dill, J. Chem. Phys. **99**, 2116 (1993).
[8] J. D. Honeycutt and D. Thirumalai, Proc. Natl. Acad. Sci. U.S.A. **87**, 3526 (1990).
[9] D. Thirumalai and D. K. Klimov, Curr. Opin. Struct. Biol. **9**, 197 (1999).
[10] J. D. Honeycutt and D. Thirumalai, Biopolymers **32**, 695 (1992).
[11] T. E. Creighton, *Protein Folding* (Freeman, New York, 1992).
[12] http://laaksonen.csc.fi/gopenmol/gopenmol.html
[13] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1986).
[14] *CRC Handbook of Chemistry and Physics*, edited by R. C.

Weast (CRC Press, Boca Raton, FL, 1975).
[15] D. Klimov and D. Thirumalai, Phys. Rev. Lett. **79**, 317 (1997).
[16] T. Veitshans, D. Klimov, and D. Thirumalai, Fold Des **2**, 1 (1996).
[17] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci, Proc. Natl. Acad. Sci. U.S.A. **92**, 3626 (1995).
[18] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[19] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **103**, 4732 (1995).
[20] N. D. Socci and J. N. Onuchic, J. Chem. Phys. **101**, 1519 (1994).
[21] A. Baumketner and Y. Hiwatari (unpublished).
[22] W. Götze and L. Sjögren, Rep. Prog. Phys. **55**, 241 (1992).
[23] A. Baumketner and Y. Hiwatari (unpublished).