Dynamics on statistical samples of potential energy surfaces

Keith D. Ball and R. Stephen Berry

Citation: The Journal of Chemical Physics **111**, 2060 (1999); doi: 10.1063/1.479474
View online: http://dx.doi.org/10.1063/1.479474
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/111/5?ver=pdfcov
Published by the AIP Publishing

Dynamics on statistical samples of potential energy surfaces

Keith D. Ball[a)]
Department of Physics and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637

R. Stephen Berry
Department of Chemistry and The James Franck Institute, The University of Chicago, Chicago, Illinois 60637

(Received 2 December 1998; accepted 5 May 1999)

Prior work [K. D. Ball and R. S. Berry, J. Chem. Phys. **109**, 8541 (1998); **109**, 8557 (1998)] has demonstrated that master equations constructed from a complete set of minima and transition states can capture the essential features of the relaxation dynamics of small systems. The current study extends this work by examining robustness of master equations based only on statistical samples of the surface topography, to make it possible to work with larger systems for which a full topographical description is either impossible or infeasible. We ask whether such ''statistical'' master equations can predict relaxation on the entire potential energy surface. Our test cases are Ar_{11} and Ar_{13}, for which we have extensive databases: 168 geometrically distinct minima and 1890 transition states for Ar_{11}, and 1478 minima and 17,357 saddles for Ar_{13} which we assume represent complete set of stationary points. From these databases we construct statistical sample sets of transition sequences, and compare relaxation predictions based on these with those obtained from the master equations representing the full potential surfaces, and with results of molecular dynamics simulations. The slowest, rate-controlling relaxation timescale converges at moderate temperatures as the number of sequences in a sample reaches ~1000, approaching convergence for as few as 100 sequences. The asymptotic value of the slowest nonzero relaxation rate is essentially identical to that from the full potential energy surface. Equilibrium properties from the statistical samples match those of the full surface. To achieve convergence within a factor of 2 of full-surface rates, the number of sequences required is approximately the same for Ar_{13} as for Ar_{11}. Precise convergence, however, appears to scale as the number of stationary points. These results reveal how the reliability and precision of kinetic predictions from statistical master equations depends on the size of the statistical database. © *1999 American Institute of Physics.* [S0021-9606(99)00629-7]

I. INTRODUCTION

The stochastic master equation is a valuable tool for revealing the kinetics on a complex potential energy surface (PES) that shape the relaxation behavior of finite systems such as clusters, molecules, and biopolymers, as well as bulk liquids, solids, and glasses. Since the master equation was first applied by Kunz *et al.* to statistical PES samples of clusters[1] and to cluster annealing,[2] it has been used to study structural relaxation processes in macromolecules, particularly the folding of lattice models of proteins[3] and heteropolymers.[4] It has also played a role in theoretical studies of biased random energy landscapes relating to the glass transition and the folding of proteins,[5–7] and has been applied recently by Becker and Karplus[8] to the relaxation of the IAN tetrapeptide, based on the PES obtained by Czerminski and Elber.[9,10]

The stochastic master equation method is a natural complement to a description by molecular dynamics (MD). First, while molecular dynamics describe precisely the mechanical behavior of all the elements of the system, a master equation is a sort of mean-field representation, in which the fast motions within the vicinity of a stationary point are replaced by their thermal averages. This means that the master equation allows us to describe phenomena on time scales far longer than we can foresee for molecular dynamics, i.e., to milliseconds and even seconds, rather than tens or perhaps hundreds of nanoseconds. The method has no practical limits on the extent of the time scale the master equation can describe. Second, the master equation directly yields a description of average behavior, obviating the need for many MD simulations required to establish reliable average kinetics. Third, the master equation transition matrix is a straightforward product of the theory used to describe the state-to-state kinetics in terms of the underlying PES. Hence the consequences of specific features of the PES or of different implications of different rate theories appear directly in the master equation solutions.

Until now, the systems studied for which a complete set of stationary points have been claimed contained ten or fewer particles, and their analysis has been handled with relative computational ease. However, as discovered for $(KCl)_5$, there were stationary points that a nominally ''exhaustive'' search method missed.[11] Even the number of known local minima for Ar_{13} has grown steadily with time. In general, one can expect that any search method is likely to

[a)]Present address: Institut für Physikalische Chemie, Technische Universität Darmstadt, Petersenstr. 20, D-64287 Darmstadt, Germany.

fail to find some of the stationary points on potential surfaces for systems of more than about ten particles.

There is no clear theoretical, or even phenomenological, indication of when a search for the stationary points on a surface has been exhaustive.[12] As the number of particles N in a system and the dimensionality $D=3N-6$ of its PES grows, the following factors exacerbate the problem of finding all stationary points: (i) The number of stationary points grows at least exponentially with N,[13,14] with all other factors being equal, so the time required to attempt a thorough search of the PES to find them must grow at least at the same rate. (ii) The computational complexity of PES search and optimization methods grows with increasing D. For example, the complexity of an MD step is $\mathcal{O}(D^2)$, while diagonalization and local optimization (quenching) algorithms typically have a minimum complexity of $\mathcal{O}(D^3)$ As an example of complexity, LJ_{38}, the generic cluster of 38 atoms bound by a Lennard-Jones interaction, is the smallest cluster whose global minimum is difficult to locate, since it is the smallest LJ_n cluster with a high divide separating a broad basin from its deepest well.[15] (iii) Furthermore, searches on a larger PES tend to take longer for the very reasons that make a larger system interesting to study. The time scale of significant conformational changes for a larger system is generally longer than for smaller molecules; for proteins, this time can range up to milliseconds, and even seconds for some glassy systems. Long time scales are a problem for dynamical algorithms because of the vast difference between the time resolution needed for a stable MD algorithm (typically $\lesssim 0.1\tau$, the typical intrawell vibration period) and the time scales of structural consequence. Roundoff error in a discrete-step algorithm inevitably accumulates; in our experience, with normal double precision simulations based on the ''velocity Verlet'' algorithm, mechanical reversibility begins to break down after about 5000 time steps, and the errors propagate rapidly from the last to higher decimal places as the duration of the simulation increases. In addition, the size of the transition matrix $\tilde{W}$ grows as the number of minima N_M in the PES sample. Given current computational capabilities, storage and diagonalization of $\tilde{W}$ for a full PES is likely to be limited to $N_M \sim 10^4$.

Necessarily, it will be large systems (compared with clusters of a few atoms) that exhibit potential for biomedical or commercial applications, at least for some time. The size of the conformational spaces of proteins gives rise to Levinthal's (pseudo)paradox,[16] which, properly put, states that proteins must have some guiding mechanism inherent in their potential surfaces because a random search of its PES would require even a small protein billions of years to fold to its native state. In essence, it is the complexity of the interaction landscapes of proteins, crystals, and, more generally, complex competing/cooperating molecular systems,[17] and the seemingly directed outcomes these systems exhibit, that drive the interest in the energy landscape paradigm.

Since one cannot in general expect to find all the stationary points on a PES, any master equation approach must rely on transition matrices constructed from only a partial knowledge of minima and transition states. Therefore, we need a sampling strategy to find a sufficient amount of the most dynamically relevant data on the PES to yield a master equation that captures the relevant, useful dynamical characteristics of the full PES. The essential problem is finding a strategy to create a robust and manageable statistical sample matrix of well-to-well transition rate coefficients on which to base the master equation. An ideal statistical sample matrix would have a spectrum of eigenvalues whose envelope exactly matches that of the full system; the sample system and the full system would have eigenvalue spectra differing only in their density, and not in any of their moments. A more limited, heuristic goal would be a matrix whose (nonzero) eigenvalues nearest zero, the smallest characteristic rates, match those of the real system. In other words, it is more important to use a master equation to reproduce the slow rates than the very fast ones. Molecular dynamics can deal with the rates up to tens, perhaps hundreds of nanoseconds, and eventually, perhaps even to microseconds with subtle algorithms that might be invented. What is needed is a method to predict or even account for rates on the millisecond time scale and beyond, from information about the PES. The goal of this paper is establishing conditions on the input to a statistically-based master equation to give a reasonable replication of the behavior of the real system that equation attempts to model.

If one depicts the PES as a network of pathways linking minima via transition states, the PES may be viewed as a ''web'' on which one can parametrize the nodes, or minima, of the network by some order parameter. There are then two generalized ''directions'' on the PES: $\mathcal{Q}$, that leading ''away'' from some fiduciary state (such as the global minimum, a basin bottom, or a native conformation), and the other direction $\theta_{\mathcal{Q}}$, orthogonal to $\mathcal{Q}$. On the PES, $\theta_{\mathcal{Q}}$ is in actuality a $(D-1)$-dimensional subspace. Any number of choices exist for the order parameter $\mathcal{Q}$, such as potential energy,[18] bond-order parameters,[18–20] Hamming distance,[21,22] and, as in protein studies, the fraction of occurring amino-acid pairings that match pairings in the native structure.[3,23,24] The smallest number of transitions required to reach a basin bottom can also serve as $\mathcal{Q}$.[1,25–27]

In the web picture, $\mathcal{Q}$ is measured along radial ''spokes'' of the web, while $\theta_{\mathcal{Q}}$ comprises links between spokes. In this representation, each basin bottom may be the hub of a wheel, meshed with other wheels like gears. In a topographic sense, a good PES sampling strategy should: (i) cover the widest possible range of $\mathcal{Q}$ values, (ii) capture an adequate representation of the crosslinking in the $\theta_{\mathcal{Q}}$ direction, and (iii) ensure that all of the minima and basins in the web are linked via at least one pathway.

What search strategy will produce such a web that meets these three criteria? A search employing eigenvector-following, initiated from seed minima obtained from MD, has difficulty meeting criterion (iii), since eigenvector-following is an initial value search that does not guarantee that a connection from the starting minimum to another selected minimum will be found. If the seed minima tend not to be adjacent, (i.e., one transition away from each other), then few minima will be connected across a distance of more than one transition.

Other search strategies[28–32] may be used as alternatives

to, or in conjunction with eigenvector-following, to circumvent this limitation. There are several two-point algorithms for locating saddles[33–43] which allow more direct connection of minima, eliminating the ''blind'' search for transition states, and saving much computational effort.

To obtain a partial sample of the PES that contains the features most relevant to relaxation dynamics, we propose a sequence sampling method. Starting from a typically high-energy minimum, which we call the *source minimum*, the method uses MD, Monte Carlo, or other PES exploration methods to ''walk'' along the PES, making transitions from minimum to minimum, until the stroll reaches the global minimum or some other regional basin bottom. The result is a sequence of transitions, essentially, a spoke of the basin's web. For the search to make progress toward a basin bottom, the search should be conducted at a sufficiently low temperature. Repetition of this process generates a statistical ensemble of pathways that are relevant for a given temperature.

Eigenvector-following, combined with molecular dynamics annealing[25,26] or Monte Carlo sampling techniques[44] have been used in the past to obtain ''monotonic sequences'' whose minima proceed strictly downhill in energy. While such sequences help provide a picture of the PES topography, pathway ensembles relevant to dynamical processes must necessarily include Boltzmann-weighted, nonmonotonic sequences.

There are two main advantages to using sequences as the elements of the statistical database (sequence sampling). First, by starting the sequence search from a desired state n_{init}, one can avoid collecting massive amounts of PES data that are peripheral to the process of interest. Second, relaxation on an ensemble of sequences starting from state n_{init} is approximately equivalent to relaxation starting with an initial probability distribution $P_k(t=0)=\delta_{k,n_{\text{init}}}$. An arbitrary initial state can be built from such individual delta-function probabilities; from there, relaxation can be studied on the combination of sequence networks originating from each of the initial states. Since the master equations are linear, the dynamical results can also be obtained separately on each ensemble and then combined, so long as regions important to more than one initial state are represented in the set of sequences that start with each of these initial states. In this way, we can combine results from several sequence ensembles to simulate arbitrary initial conditions.

Now we turn to the problem of how to use sequence sampling to construct a robust master equation based only on a statistical sample of linked stationary points. We do this by using as our test case a system small enough that we can construct the full master equation, yet large enough to allow us to construct a variety of statistical samples of sequences on which to base the matrices we are testing. We compare not only rates from statistically-based and full master equations; we also introduce results from molecular dynamics.

What kinds of problems are likely to arise from inadequate statistical samples? The most obvious, and most easily overcome, are simply underrepresentation of the number of sequences and inadequate representation of the distribution of accessible minima. A more subtle but obvious potential shortcoming of sequence sampling may be that it concentrates on finding spokes on the PES at the expense of finding crosslinks between them. Depending on the nature and number of sequences in a given ensemble, relaxation based on a statistical set of sequences may be strikingly different from that on the full PES. To evaluate how best to use statistical sets of sequences to capture the dynamical behavior of the full PES, we have created sets of random sequences by selecting them from a database of the full set of minima and transition states. Since we are selecting transition states from a known set of possible transitions, we do not need to use MD to produce sequences, but at each minimum simply select a transition state using a Monte Carlo selection process. We then test the convergence of their relaxation time scales to those of relaxation on the full PES as the number of sequences N_{seq} per sample increases.

We first discuss the stationary point database found for Ar_{11} and pertinent topographical features found in this database. We then describe the synthesis of sequences constructed from this database, and display the resulting relaxation time scales for sequence sets containing different numbers of sequences, compared to the time scales observed on the full PES and to the slowest time scale observed from molecular dynamics relaxation. As a brief examination into how these convergence trends might scale with the dimensionality of the potential energy surface, we also examine the convergence sequence sets constructed from a database for Ar_{13}.

II. PROPERTIES OF THE FULL PES

A. Ar_{11}

We model Ar_{11} with the pairwise Lennard-Jones interaction potential

$$V_{ij}=4\epsilon\left[\left(\frac{\sigma}{r_{ij}}\right)^{12}-\left(\frac{\sigma}{r_{ij}}\right)^{6}\right], \tag{1}$$

with $\epsilon=1.67\times10^{21}$ J and $\sigma=3.4$ Å,[45] as used previously for Ar_9.[1,2] As in prior work,[46] we initiated an eigenvector-following search seeded with 128 initial geometrically-distinct minima found from MD simulation. From these seed minima, all PES eigenvectors were followed in both directions with the initial configuration unscaled. All eigenvectors were then followed in one direction starting with the initial configuration coordinates scaled by factors of 0.92, 0.84, and 0.76. In addition, searches following the softest mode at each step were performed for both scaled and unscaled configurations. Transition state searches starting with these initial conditions were performed from all minima found as the search progressed. In total, 168 geometrically distinct minima and 1890 transition states were found. Of the latter, 1761 are nondegenerate, i.e., they are first-rank saddles. The energies and structures of the minima and 30 lowest-energy transition states are those reported by Tsai and Jordan (TJ).[47] The number of minima is slightly higher than the 152 minima reported by TJ and the 145 minima reported by Hoare.[13] Our transition-state database is a significant improvement over that of TJ, who reported 911 transition states.

In this paper, a (geometrically) distinct minimum is referred to as DMn, where n is the index number indicating the

minimum's rank in the database in order of increasing potential energy. Likewise, transition states (rank-one saddle points) are referred to as SAn. Transition states are indexed simply according to the order in which they were discovered, not in order of potential energy.

The global minimum of Ar_{11} (DM1) is similar in conformation to the icosahedral (I_h symmetry) global minimum of Ar_{13}, except that in the former, two adjacent atoms are removed from one of the pentagonal rings encircling the I_h axis. A more symmetric reduction of the 13-atom icosahedron to an 11-particle cluster would be to remove the ''end-cap'' atoms at the poles of the I_h axis; however, this D_{5d} structure is actually DM10 (the tenth-lowest in energy).

In the monotonic sequence picture, Ar_{11} has two basins, with DM1 at the bottom of the primary basin (PB) and DM2 at the bottom of the secondary basin (SB). The existence of the stationary point at the bottom of the second basin is a consequence of its D_{3h} symmetry. In fact, it is a local minimum for this system. A sequence of rearrangements converts DM2 into DM1. The shortest, lowest-energy path between them involves two steps, through DM7:

$$\mathrm{DM2}\rightarrow \underset{V=-30.6403\epsilon}{\mathrm{SA200}} \rightarrow \mathrm{DM7}\rightarrow \underset{V=-31.3189\epsilon}{\mathrm{SA25}} \rightarrow \mathrm{DM1}.$$

Here, SA25 is the lower-energy of two DM7→DM1 transitions; the other is DM7→SA1433($V=-30.5791\epsilon$) →DM1. SA25 is the second-lowest-energy nondegenerate transition state on the Ar_{11} PES. The lowest is SA59, $V=-31.3461\epsilon$, connecting DM1 and DM9. However, since SA200 is the lowest-energy transition state leaving DM2 (albeit only the 53rd lowest in energy on the entire PES), the above pathway is also the lowest-energy pass between the two basin bottoms.

B. Ar_{13}

Using a similar eigenvector-following search strategy on the Ar_{13}, PES we found 1478 distinct minima and 17 357 transition states. This represents a modest increase in the number of minima (1328) found by TJ, and a substantial increase over their number of transition states (8453). In this database, there is only one primary basin, with each minimum accessible from the global minimum by a monotonic uphill sequence of transitions, save for one intermediate-energy minimum. This minimum might still be part of the primary basin, however, since our PES search has probably not been exhaustive. Nonetheless, for the purposes of this study, we can take the current database to be a reasonably full data set, from which we can draw significantly smaller samples. Since Ar_{13} is a small magic-number cluster, it is likely that the only significant basin on its PES is the one surrounding the global minimum.

III. SEQUENCE SAMPLING

To produce partial samples of our full PES database, we generate random sequences of minima using the following algorithm: starting from a selected source minimum, we consider all possible transition states leading from that minimum, and choose one of these at random with Boltzmann probability $P_j \propto \exp(-\Delta V_j/k_BT)$. Here ΔV_j is the barrier height of the jth transition, and T is the midpoint of the solid/liquid phase coexistence temperature regime. T is taken to be 22 and 32 K for Ar_{11} and Ar_{13}, respectively.[48] The selected transition state leads to a new minimum. This transition selection process is repeated until a basin bottom is reached, at which point that sequence terminates.

We have created 10 000 such sequences from each of two Ar_{11} source minima: DM168 (the highest-energy minimum) and DM145. We chose the latter for reasons of comparison, since it also lies high in energy on the potential surface and possesses relatively many initial transitions (12, compared to the six transitions leading from DM168). It was also chosen because it is not ''downstream'' from DM168, meaning that it cannot be accessed from DM168 by moving strictly downhill. This property helps insure that sequences emanating from DM145 will not merely be a subset of those starting from DM168. Since Ar_{11} has two basins, all Ar_{11} sequences terminate at one of the two basin bottoms, namely DM1 or DM2. In our sequence databases, relatively few sequences terminate at DM2: only 404 of those starting from DM168, and 450 of those from DM145. This relative paucity of DM2-terminating sequences may indicate that the second PES basin may play a minor role in relaxation, since most transition sequences seem to be able to find the global minimum without visiting DM2 first. The average lengths of sequences starting from the two-source minima were as follows: 6.24 transitions for DM168, and 4.52 transitions for DM145.

To treat Ar_{13}, we have likewise created 10 000 sequences, starting from DM1478 (the highest-energy minimum). These sequences have an average length of 11.73 transitions. Since the global minimum (DM1) is the only basin bottom we found, all sequences end with DM1.

For all three source minima, we constructed ensembles of N_{set} sequence sets, with each set in the ensemble containing N_{seq} sequences drawn from the pool of 10 000. We refer to this ensemble of sets as an N_{seq}-ensemble. Each sequence set serves as a PES sample for which a master equation is constructed and solved. For Ar_{11}, ensembles in the range $1 \leqslant N_{\mathrm{seq}} \leqslant 10\,000$ were constructed, while for Ar_{13}, only ensembles for $N_{\mathrm{seq}} \geqslant 5$ were constructed. For $N_{\mathrm{seq}}=1$, 2, or 5, $N_{\mathrm{set}}=2000$, while for $N_{\mathrm{seq}}=10$, 20, 50, 100, 200, and 500, $N_{\mathrm{set}}=10\,000/N_{\mathrm{seq}}$. For $N_{\mathrm{seq}}<1000$, each sequence set serves as an independent PES sample. The ensembles for which each set contains $N_{\mathrm{seq}}=1000$, 2000, 3000, or 5000 sequences have $N_{\mathrm{set}}=20$, 9, 8, or 6 sets, respectively. For these values of N_{seq}, there is some overlap of the sequences; a given sequence may be shared between one or more sets. Finally, for each source minimum the ensemble for $N_{\mathrm{seq}}=10\,000$ contained only one set, which included all sequences. Figure 1 is a schematic diagram of the random transition sequences for a typical N_{seq}-ensemble with $N_{\mathrm{seq}}=100$.

Figure 2 shows the growth in the fraction of the total set of unique minima and transition states contained in the sequence set comprised of the first N_{seq} sequences from each collection of 10 000 sequences. The figure shows that the fraction of included minima grows in a roughly logarithmic fashion. The DM145 sequence set begins to show saturation

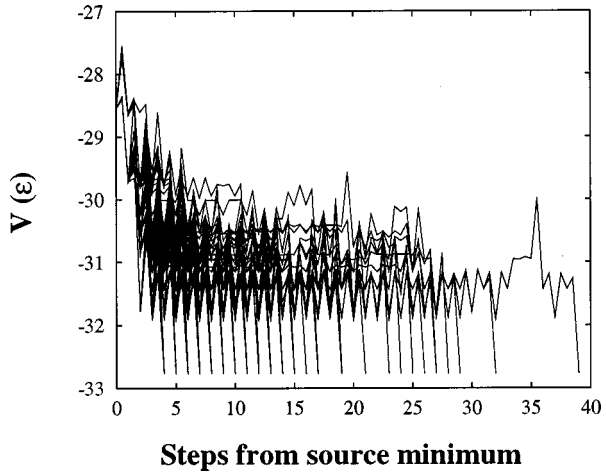

FIG. 1. Schematic representation of a typical N_{seq}-ensemble for $N_{seq}=100$ of random sequences emanating from source minimum DM168. Minima are plotted at integer step values, while transition states are plotted at half-integer values.

of the number of minima at $N_{seq}\sim 4000$, while the DM168 set does not appear to deviate from the logarithmic trend below 10^4. This latter set already contains over half the total number of Ar_{11} minima used in the present analysis. However the number of minima is almost certain to grow as more sequences are collected. The number of transition states exhibits a roughly constant logarithmic growth for $N_{seq}\gtrsim 500$ for both DM168 and DM145. In neither case does the growth appear to level off at $N_{seq}=10^4$.

The corresponding growth rate in the size of the sample of stationary points for Ar_{13} is similar to that for the Ar_{11} sequence samples for $N_{seq}\leqslant 500$. However, the growth reaches its terminal logarithmic rate at a higher N_{seq} than it does in the Ar_{11} samples, causing N_{seq} to be a factor of 5–10 larger than in the Ar_{11} samples for the same fractional content. Whether this difference means that a greater number of sequences must be collected for a larger system in order to obtain master equations of the same accuracy will be addressed in the last part of the results section. Although systems of only two sizes are hardly sufficient to show a precise scaling trend in a critical value of N_{seq}, it is encouraging that the stationary point samples of Ar_{13} grow with N_{seq} at the same rate as the Ar_{11} samples.

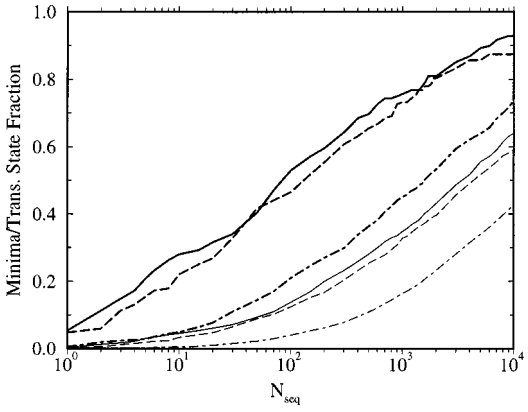

FIG. 2. The number of unique stationary points (given as the fraction of the full-PES sample) contained in the first N_{seq} random downhill sequences of Ar_{11} and Ar_{13} originating at the three source minima. Shown are the fractions of minima contained in the sequences starting from Ar_{11}, DM168 (thick solid line); Ar_{11}, DM145 (thick dashed line); Ar_{13}, DM1478 (thick dot-dashed line). The thin lines show the corresponding fractions of transition states.

IV. RESULTS AND DISCUSSION

A. Ar_{11}

We have performed our master equation calculations using the H–N model of Ref. 11, since it was found to have the best overall agreement with the results of MD relaxation from both DM145 and DM168. In the latter case, MB($\eta_P=0.1$)-based models[11] showed a marked slowdown on the longest relaxation time scale compared to MD at $T=10$ K. At higher temperatures, the results from H–N and MB($\eta_P=0.1$)-N models are both comparable to the observed MD dynamics.

For both source minima DM168 and DM145, we examine the convergence of statistics of the transition matrix eigenvalue spectrum for sequence sets with $1\leqslant N_{seq}\leqslant 10^4$ and compare these results to the properties of the eigenvalue spectrum for the full PES. Results are displayed at 10, 22, and 30 K, for which Ar_{11} is solid-like, in solid–liquid coexistence, and liquid-like, respectively.

The dynamical quantity we use as a signature in this paper is $\langle E_q(t)\rangle$, the mean quenched energy of the system at time t. We calculate the contribution of each relaxation mode to this quantity using the equation

$$\langle E_q(t)\rangle = E_q^{eq} + \sum_{k=1}^{N_M-1} A^k \exp(-\lambda_k t). \tag{2}$$

Here, λ_k is the magnitude of the eigenvalue of the kth eigenmode. Since all nonzero eigenvalues of the transient modes of the master equation are negative, we use ''eigenvalue'' to mean the eigenvalue magnitude in the rest of this paper. The contribution of each mode is given by its initial amplitude

$$A^k = c^k\left(\sum_j^{N_M} V_j P_j^k\right), \tag{3}$$

determined by the overlap c^k with the initial probability distribution and the average quenched energy contribution to that mode, taken over the potential energies V_j of all minima. These contributions are weighted by P_j^k, the occupation probability of minimum j due to mode k. The eigenmodes and eigenvalues are labeled according to their magnitude, with $k=0$ corresponding to the equilibrium solution ($\lambda_0=0$). The mean quenched energy at equilibrium E_q^{eq} is an average over all minimum energies, weighted with the respective occupation probability P_j^{eq} of each minimum j at equilibrium.

We first concentrate on the three slowest (i.e., lowest-magnitude) nonzero eigenvalues λ_1, λ_2, and λ_3, since the slowest modes determine the equilibration time of the system. The results we have found for source minima DM168 and DM145 show essentially similar convergence trends. To further test the generality of these trends, we created 5000 additional downhill sequences starting from DM143, chosen

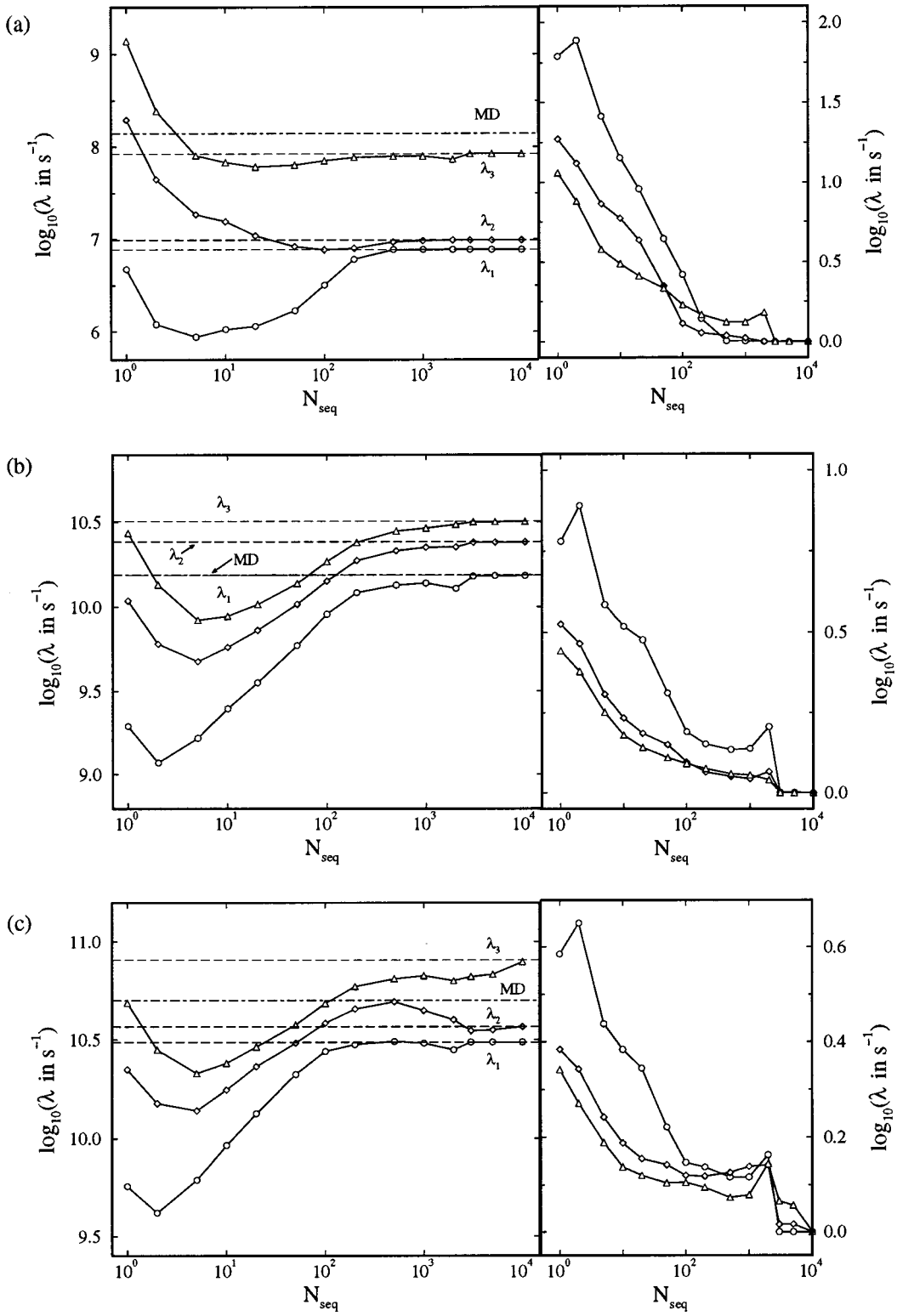

FIG. 3. Ensemble averages of the three lowest eigenvalues of the transition matrix for relaxation of Ar_{11} starting from DM168, at (a) $T=10$ K, (b) 22 K, and (c) 30 K. The eigenvalues are indicated as follows: λ_1 (circles), λ_2 (diamonds), λ_3 (triangles), where $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3$. Variances are shown in the small graphs at the right. The dashed lines are the corresponding eigenvalues of the transition matrix constructed from the full-PES database. The dot-dashed line is the slowest resolved MD time constant.

because it is not downstream from either DM145 or DM168. The convergence trends in the lowest three eigenvalues are very similar to those observed in the DM168 and DM145 data at the same temperatures. Since the results for Ar_{11} are similar for all three source minima, we will discuss only the data for DM168.

The ensemble averages for each of these eigenvalues are given in Fig. 3 for N_{seq}-ensembles of sequence sets starting

from source minimum DM168. At each temperature, the corresponding eigenvalues of the transition matrix for the full PES are given for comparison. The lowest eigenvalues at 22 and 30 K are nearly of equal magnitude. This accords with eigenvalue histograms for $(KCl)_5$ and Ar_9 that show a concentration of the eigenvalue spectrum into one peak at and above the thermodynamic coexistence region, with the lowest eigenvalues diverging only at subcoexistence temperatures.[11]

All eigenvalue plots show minima at some $N_{seq}<100$, with significantly larger dips occurring for λ_1 and above the coexistence temperature in the range $2\lesssim N_{seq}\lesssim 10$. These slower relaxation time scales occur for PES samples containing only a few sequences, because these sequences are less likely than larger samples to be interconnected at low energies, so that they have higher energy barriers to relaxation, and correspondingly slower relaxation.

In contrast, single-sequence sets act as a single basin (albeit with zero path entropy), and so relax significantly faster than the few-sequence sets. Nevertheless, their eigenvalue variance is still quite high, since they exhibit a wide array of different transition rates. The slowest relaxation mode, corresponding to the smallest eigenvalue, is uniformly slower than in the full-PES case, since any slow step along a single sequence forms a kinetic bottleneck that cannot be circumvented.

As Fig. 3 shows, limiting behavior only occurs for large values of N_{seq}, which have significant transition path entropy and connectivity. For these sets, the variances in the ensemble averages of spectral quantities become small, as the small graphs at the right of the figure show, and the slowest eigenvalue converges to some specific value $N_{seq}\rightarrow\infty$. By $N_{seq}=3000$, the eigenvalues are identical to the full-PES values, with the exception of λ_3 at $T=30\,K$. Since λ_1 is of chief interest, we mark the convergence to the full-PES dynamics as the value of N_{seq} above which λ_1 converges within a factor of 2 to the full-PES value. The critical values of N_{seq} range from roughly 130 at 10 K to 30 at 30 K. Since energy barriers are not as significant at higher temperatures, the low path entropy at small N_{seq} is not as significant as it is at lower temperatures, and hence depresses the slowest relaxation time scale by a smaller amount.

To check the validity of the master equation model, we compared the eigenvalues for sequence samples and the full PES against the slowest relaxation time constant resolved from MD. The time constants were resolved by nonlinear least-squares fitting of Eq. (2) to the simulation result for $\langle E_q(t)\rangle$, as done in previous work for $(KCl)_5$ and Ar_9.[12] Typically, 2 to 4 modes with nonzero time constants were resolved for each $\langle E_q(t)\rangle$ profile. We take the slowest of these to correspond to the slowest model eigenvalue λ_1, while generally it is not clear whether other resolved time constants correspond to λ_2 and λ_3.

At 10 K, the model value is more than a decade lower than the simulation value. This discrepancy is due mainly to the limited accuracy of the $\langle E_q(t)\rangle$ fits, since the maximum simulation length of 10^{-8} s was shorter than the time constant anticipated from the model. In contrast, the agreement at 22 K is striking, with fractional discrepancies in $\log\lambda$ of 4×10^{-4} and 8×10^{-5} for source minima DM 168 and DM 145, respectively. At 30 K, fractional discrepancies are expectedly larger, but are only at the 2% level. In absolute terms, the time constants differ by less than a factor of 2. From these results we see that our partition function models tested on systems $(KCl)_5$ and Ar_9, (Refs. 11, 46) can be extended to Ar_{11}.

The slowest mode need not dominate the relaxation behavior; there may be faster modes with larger amplitudes A^k. For the Ar_{11} source minima studied, the largest of the amplitudes of nonzero modes with time constants λ_A, $A>1$, can be from 10 to 1000 times larger than that of the λ_1 mode. (For the Ar_{13} source minimum DM1478, it can be 10^6 times as large.) In the case of Ar_{11}, the slowest predicted relaxation time is significantly closer to that observed in MD simulation than is the strongest mode. Furthermore, except for very slow relaxation, for which the value extracted from simulation is not very reliable, the MD result is generally close enough to λ_1 to assume that λ_1 is effectively the limiting relaxation rate. The relaxation spectra show that the strongest modes tend to occur in nearly-degenerate pairs with amplitudes similar in magnitude, but with opposite signs. This allows the slowest mode to dominate the relaxation energy profile $\langle E_q(t)\rangle$. Nonetheless, the convergence of this eigenvalue to its full-PES value is similar to the convergence observed for λ_1, as shown in Fig. 4.

Since the equilibrium results of relaxation are also important, we must consider not only the number and distribution of kinetic pathways, but the distribution of minima as well. We use the equilibrium value of the mean quenched energy E_q^{eq} to gauge the extent of attainment of equilibrium for each sequence set. In Fig. 5, we compare the mean values of E_q^{eq} obtained for each N_{seq}-ensemble against the corresponding value on the full PES, and compare both to the value obtained from MD. At 10 K E_q^{eq} converges to the full-PES value -32.766ϵ at $N_{seq}=10$. To three decimal places, this value is equivalent to the energy of the global minimum. At the higher temperatures, E_q^{eq} converges at successively higher values of N_{seq}.

This delayed convergence can be understood in terms of the increasing importance of the entropic term in the free energy with increasing temperature. The PES entropy at higher potential energy values is not well modeled by a PES sample constructed of only a few sequences. In particular, random sequences based on a single source minimum are likely to exclude minima at energies the same as or higher than that of the chosen source. The random sequences converge on basin bottoms, so the small entropy near the basin bottom (where the system spends virtually all of its time at equilibrium) is necessarily well-modeled, even by relatively small collections of sequences. Since the system spends more time at higher-energy PES regions at and above the coexistence temperature, larger values of N_{seq} are required to accurately model the larger entropy.

Molecular dynamics reveals a somewhat higher value for E_q^{eq}, indicating that the harmonic model relatively overestimates the low-energy contributions to partition functions, causing the downward shift in energy of the model values. The agreement between theory and simulation is reasonable,

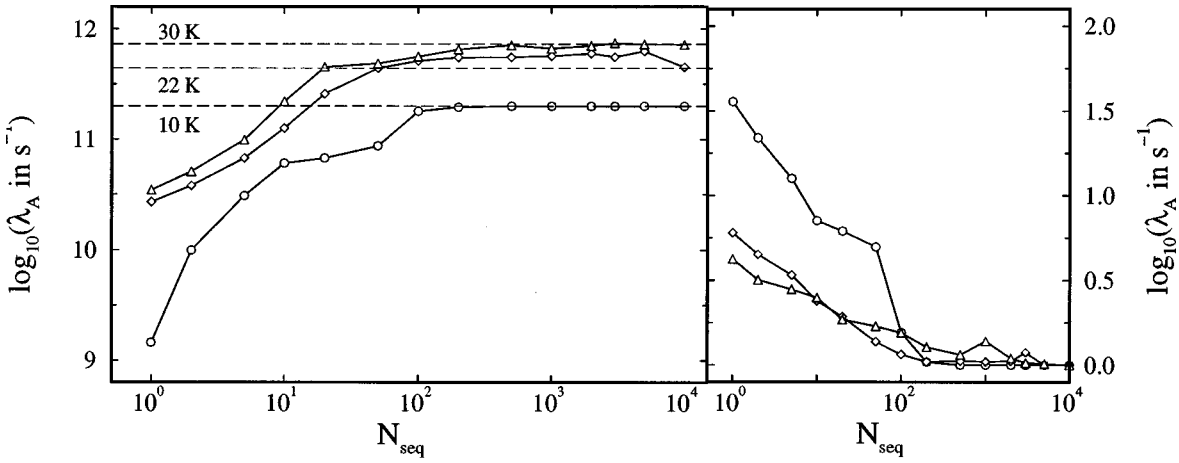

FIG. 4. Ensemble averages of the eigenvalue λ_A of the largest-amplitude relaxation mode for Ar_{11} starting from DM168, at $T=10$ K (circles), 22 K (diamonds), and 30 K (triangles). Standard deviations are represented in the graph on the right of the figure. The dashed lines are the corresponding eigenvalues of the transition matrix constructed from the full-PES database and the given initial condition.

since the two values of E_q^{eq} are much closer to each other than they are to the energy of the next-highest minimum ($V=31.9152\epsilon$).

B. Ar_{13}

The Ar_{11} results indicate that relaxation dynamics can be modeled relatively well with about 100 random sequences starting from a given source minimum. From Fig. 2, this translates to PES samples containing roughly 40% of the minima on the entire PES, and 20% of the transition states. Even though the mean sequence length tends to be longer on a larger PES, the fraction of stationary points contained in N_{seq} sequences for the larger system is likely to be much smaller than for an equal number of sequences on a smaller PES. This is desirable, since for larger systems to be tractable, the time to obtain and process the data, and the space required to store the data, must grow more slowly than the PES dimensionality. The question we address here is how the minimum necessary N_{seq} scales as the PES grows. If this number stays constant, or only increases relatively slowly, the random-sequence PES sampling approach will still be useful for large systems.

In Fig. 6, we show the time constants for the slowest three relaxation modes for the N_{seq}-ensembles from Ar_{13} source minimum DM1478, together with full-PES values for these same modes. Although λ_2 and λ_3 do not seem to converge uniformly for fewer than 10^4 sequences, λ_1 converges precisely to the full-PES value at each temperature at and above $N_{seq}=5000$ (compared to 500–3000 for Ar_{11}, depending upon temperature). For Ar_{13}, λ_1 approaches the full-PES value to within a factor of 2 at roughly the same value at low temperatures, and at high temperatures is within this factor even when $N_{seq}=5$. The main feature of the λ_1 evolution is an overshoot of the full-PES value. This overshoot increases with temperature, growing to 1.5 times the full-PES value. In addition, the overshoot occurs for lower N_{seq}, and extends over a wider N_{seq} range with increasing temperature.

The overshoot may be due to missing ''inertia,'' in the sense that sufficiently many paths exist at intermediate N_{seq} values to ensure rapid relaxation, but the occupation prob-

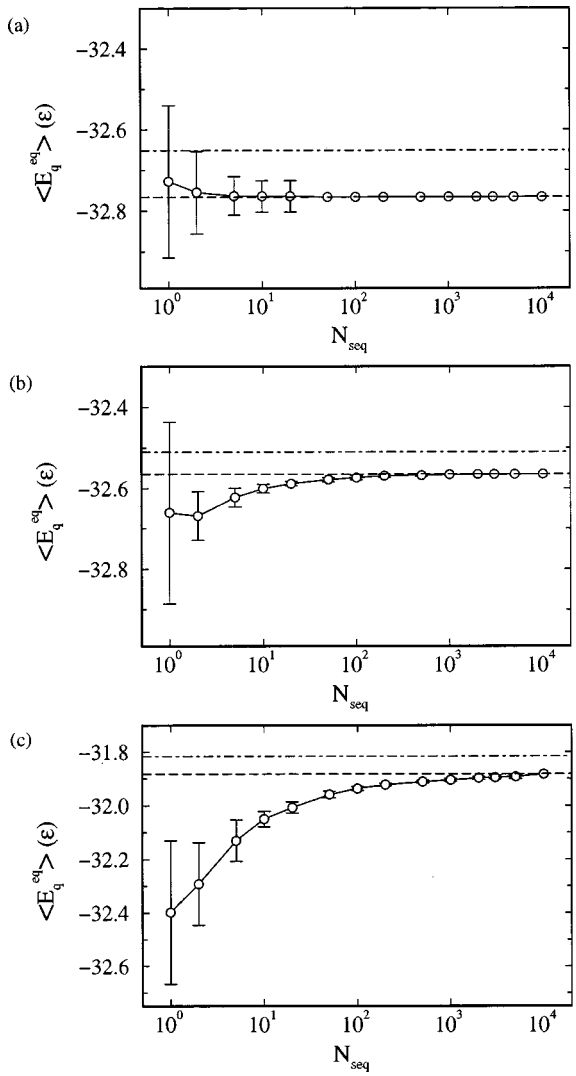

FIG. 5. Mean-quenched energy at equilibrium E_q^{eq} for sequences sets of Ar_{11} constructed from sequences starting at DM168. Values of E_q^{eq} are given at (a) 10 K, (b) 22 K, and (c) 30 K. The solid line with circles (and error bars) indicates the ensemble average (and standard deviation) of E_q^{eq}. The dashed line is the full-PES value of E_q^{eq}. The dot-dashed line is the molecular dynamics result.

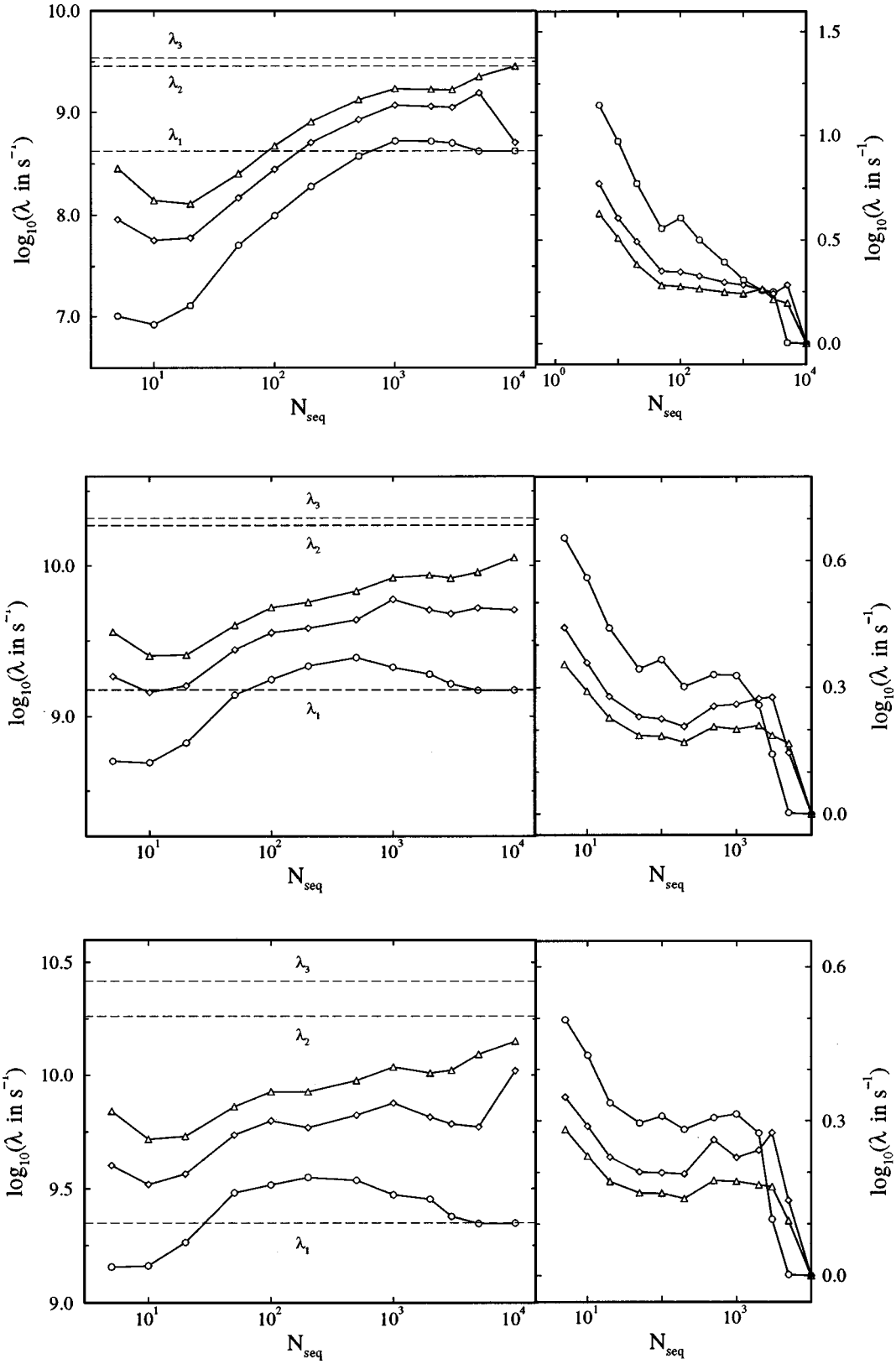

FIG. 6. Ensemble averages of the three lowest eigenvalues of the transition matrix for relaxation of Ar_{13} starting from DM1478, at (a) $T=20$ K, (b) 32 K, and (c) 40 K. The eigenvalues are indicated as follows: λ_1 (circles), λ_2 (diamonds), λ_3 (triangles), where $\lambda_1 \leqslant \lambda_2 \leqslant \lambda_3$. Variances are shown in the small graphs at the right. The dashed lines are the corresponding eigenvalues of the transition matrix constructed from the full-PES database.

ability is concentrated into fewer minima. Therefore, relaxation involves fewer activated processes than would be significant on the full PES, allowing relaxation to occur more quickly. In any case, the overshoot stays within a reasonable range. In general, whether our sequence sample must contain more samples depends upon the accuracy we require: if a factor of two is permissible, we need the same number of sequences as for Ar_{11}. For accuracies of 10% or so, we may

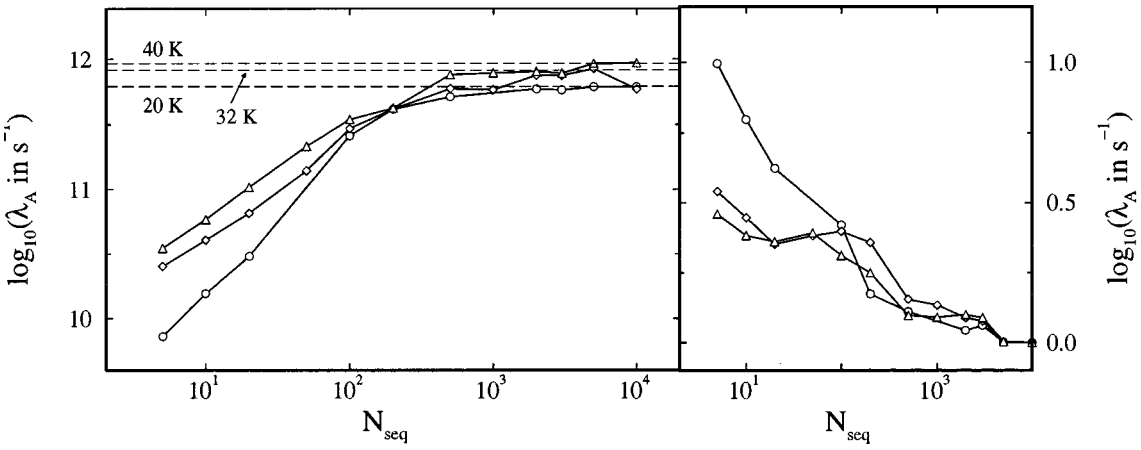

FIG. 7. Ensemble averages of the eigenvalue λ_A of the largest-amplitude relaxation mode for Ar_{13} starting from DM1478, at $T=20$ K (circles), 32 K (diamonds), and 40 K (triangles). Standard deviations are represented in the graph on the right of the figure. The dashed lines are the corresponding eigenvalues of the transition matrix constructed from the full-PES database and the given initial condition.

need to collect ten times as many sequences for Ar_{13} as for Ar_{11}. Given the limited accuracy of the rate theory used in the model relaxation, a factor of 2 may suffice. In any case, for larger systems one should observe the convergence of relaxation rates with N_{seq} in order to be certain of the bounds of the overshoot region.

For an additional point of comparison to Ar_{11}, Fig. 7 indicates that the time constant λ_A of the largest-amplitude mode converges to the same degree of accuracy for Ar_{13} with roughly five times as many sequences as needed for Ar_{11}, but otherwise converges smoothly and without overshoot. Below liquid-like temperatures, the convergence of E_q^{eq} shown in Fig. 8 converges about as rapidly for Ar_{13} as it does for Ar_{11}, but more slowly at higher temperature. Comparing the 30 K results for Ar_{11} to the 40 K results for Ar_{13} reveals that E_q^{eq} convergence to within 1% of the full-PES value occurs when $N_{seq} \approx 8-10$, while the same level of convergence occurs for Ar_{13} only when $N_{seq} \geqslant 200$. Given the higher PES entropy on the Ar_{13} surface, slower convergence is to be expected. However, for a system to relax to a specific PES region (e.g., a basin bottom), the temperature will necessarily need to be low. Therefore, convergence of E_q^{eq} is sufficiently rapid at temperatures where the equilibrium is important for relaxation studies.

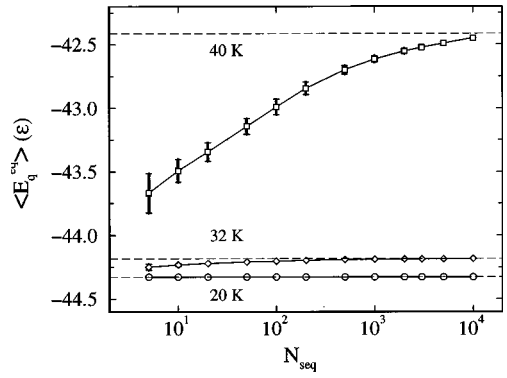

FIG. 8. Mean-quenched energy at equilibrium E_q^{eq} on Ar_{13} sequence sets constructed from sequences starting at DM1478. Values of E_q^{eq} are given at 20 K (circles), 32 K (diamonds), and 40 K (triangles). Error bars represent symmetric standard deviations of E_q^{eq}. Dashed line is the full-PES value of E_q^{eq}.

V. CONCLUSIONS

We have proposed a novel scheme for acquiring PES samples consisting of sets of randomly-generated sequences of minima from which a master equation can be constructed. We determine how extensive such a sequence set must be to mimic the dynamical behavior on the full PES, and find that the slowest dynamical time scales approach those of full-PES values at $N_{seq} \sim 100$ or below, with essentially complete convergence occurring for $N_{seq} > 1000$. With increasing temperature, the onset of convergence tends to occur with fewer sequences in the PES sample.

At phase coexistence temperatures and below, the average occupation probabilities at equilibrium, as measured by the mean-quenched energy, converge to the full-PES results at the same or smaller N_{seq} values as those needed for relaxation time scale convergence. Although the convergence of this equilibrium property requires higher N_{seq} values for liquid-like temperatures, relaxation at lower temperatures is more interesting, since only at these temperatures will the system equilibrate to specific, isolated final states of low entropy.

Since we have only examined two systems of different sizes, we have not been able to determine an explicitly quantitative scaling rule for the number of sequences or fraction of stationary points that a PES sample must contain in order to yield master equation results mirroring those of the full PES. We did however find that the Ar_{13} PES, with roughly an order-of-magnitude more minima than the Ar_{11} PES, required up to ten times as many sequences to get precise convergence to full-PES values. As Fig. 2 indicates, Ar_{13} and Ar_{11} require databases with about the same fraction of the total number of stationary points in order to yield precise convergence. If this level of precision is needed, then the master equation approach may have to be abandoned for methods requiring less extensive knowledge of the stationary

points of the PES. However if convergence to within a factor of two will suffice, then the master equation method may be a promising tool for studying relaxation. In general, this less stringent convergence criterion should suffice, since the transition rates calculated by transition-state theory can themselves be no more accurate than a factor of two, and sometimes are only within an order-of-magnitude of the observed rates.

We have also found two qualitative differences between the time scale convergences of Ar_{11} and Ar_{13}: namely, that Ar_{13} time scales are closer to the full-PES values even for PES samples comprised of very few sequences, and that, as Fig. 6 shows, the time scales for Ar_{13} tend to overshoot the full-PES value for an intermediate number of sequences. These observations suggest that in practice one may need sample a given large PES with a variety of sample sizes, and monitor the convergence of the results in a fashion similar to (if less thorough than) our work in this paper. Nonetheless, our results indicate that the sequence-sampling method yields the appropriate statistical ensemble of PES pathways with a only a moderate amount of data acquisition, and is likely to yield insight into dynamics for still larger systems.

ACKNOWLEDGMENT

This research was supported by a Grant from the National Science Foundation.

[1] R. S. Berry and R. Breitengraser-Kunz, Phys. Rev. Lett. **74**, 3951 (1995).
[2] R. E. Kunz, P. Blaudeck, K. H. Hoffmann, and R. S. Berry, J. Chem. Phys. **108**, 2576 (1998).
[3] P. E. Leopold, M. Montal, and J. N. Onuchic, Proc. Natl. Acad. Sci. USA **89**, 8721 (1992).
[4] M. Cieplak, M. Henkel, J. Karbowski, and J. R. Banavar, Phys. Rev. Lett. **80**, 3654 (1998).
[5] V. Gates, E. Kangaroo, M. Roachcock, and W. C. Gall, Physica D **15**, 289 (1985).
[6] J. Wang, J. Onuchic, and P. Wolynes, Phys. Rev. Lett. **76**, 4861 (1996).
[7] A. Fernández and G. Appignanesi, Phys. Rev. Lett. **78**, 2668 (1997).
[8] O. M. Becker and M. Karplus, J. Chem. Phys. **106**, 1495 (1997).
[9] R. Czerminski and R. Elber, Proc. Natl. Acad. Sci. USA **86**, 6963 (1989).
[10] R. Czerminski and R. Elber, J. Chem. Phys. **92**, 5580 (1990).
[11] K. D. Ball and R. S. Berry, J. Chem. Phys. **109**, 8557 (1998).
[12] K. D. Ball, Ph.D. thesis, The University of Chicago, 1998.
[13] M. R. Hoare, Adv. Chem. Phys. **40**, 49 (1979).
[14] F. H. Stillinger and T. A. Weber, Phys. Rev. A **28**, 2408 (1983).
[15] J. P. K. Doye and D. J. Wales, J. Chem. Phys. **110**, 6896 (1999).
[16] C. Levinthal, in *Mossbauer Spectroscopy in Biological Systems*, edited by P. DeBrunner, J. Tsibris, and E. Munck (University of Illinois, Urbana, IL, 1969), pp. 22–24, proceedings of a meeting held at Allerton House, Monticello, IL.
[17] *Molecular Evolution on Rugged Landscapes: Proteins, RNA and the Immune System*, edited by A. S. Perelson and S. A. Kauffman (Addison Wesley Longman, Reading, MA, 1991).
[18] R. M. Lynden-Bell and D. J. Wales, J. Chem. Phys. **101**, 1460 (1994).
[19] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, Phys. Rev. B **28**, 784 (1983).
[20] R. M. Lynden-Bell, J. S. van Duijneveldt, and D. Frenkel, Mol. Phys. **80**, 801 (1993).
[21] R. Hamming, *Coding and Information Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1980).
[22] M. Matsumoto (unpublished).
[23] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).
[24] J. N. Onuchic, P. G. Wolynes, Z. Luthey-Schulten, and N. D. Socci, Proc. Natl. Acad. Sci. USA **92**, 3626 (1995).
[25] K. D. Ball *et al.*, Science **271**, 963 (1996).
[26] R. S. Berry, N. Elmaci, J. P. Rose, and B. Vekhter, Proc. Natl. Acad. Sci. USA **94**, 9520 (1997).
[27] A. Heuer, Phys. Rev. Lett. **78**, 4051 (1997).
[28] D. D. Frantz, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **97**, 5713 (1992).
[29] H. Grubmüller, Phys. Rev. E **52**, 2893 (1995).
[30] D. L. Freeman and J. D. Doll, Annu. Rev. Phys. Chem. **47**, 43 (1996).
[31] M. S. Head, J. A. Given, and M. K. Gilson, J. Phys. Chem. A **101**, 1609 (1997).
[32] K.-K. Han, Phys. Rev. E **54**, 6906 (1996).
[33] R. Elber and M. Karplus, Chem. Phys. Lett. **139**, 375 (1987).
[34] S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).
[35] I. V. Ionova and E. A. Carter, J. Chem. Phys. **98**, 6377 (1993).
[36] A. Matro, D. L. Freeman, and J. D. Doll, J. Chem. Phys. **101**, 10458 (1994).
[37] L. R. Pratt, J. Chem. Phys. **85**, 5045 (1986).
[38] R. Elber, in *Recent Developments in Theoretical Studies of Proteins*, edited by R. Elber (World Scientific, Singapore, 1996).
[39] D. M. Deaven, N. Tit, J. R. Morris, and K. M. Ho, Chem. Phys. Lett. **256**, 195 (1996).
[40] S. Huo and J. E. Straub, J. Chem. Phys. **107**, 5000 (1997).
[41] J. R. Gunn, J. Chem. Phys. **106**, 4270 (1997).
[42] A. Ulitsky and D. Shalloway, J. Chem. Phys. **106**, 10099 (1997).
[43] J. M. Goodman and W. C. Still, J. Comput. Chem. **12**, 1110 (1991).
[44] J. P. K. Doye and D. J. Wales, Z. Phys. D **40**, 194 (1997).
[45] N. Bernardes, Phys. Rev. **112**, 1533 (1958).
[46] K. D. Ball and R. S. Berry, J. Chem. Phys. **109**, 8541 (1998).
[47] C. J. Tsai and K. D. Jordan, J. Phys. Chem. **97**, 11227 (1993).
[48] T. L. Beck, J. Jellinek, and R. S. Berry, J. Chem. Phys. **87**, 545 (1987).