Dynamical and Statistical Mechanical Characterization of Temperature Coupling Algorithms

M. D'Alessandro,[†] A. Tenenbaum,[‡] and A. Amadei*,[†]

Department of Chemical Sciences and Technology, University of Rome "Tor Vergata" Via della Ricerca Scientifica 1, 00133, Rome, Italy, and Unità di Roma 1−INFM, and Physics Department, University of Rome "La Sapienza" Piazzale Aldo Moro 2, 00185, Rome, Italy

Received: October 3, 2001; In Final Form: March 11, 2002

In this article, we investigate molecular dynamics (MD) trajectories of a butane molecule, as obtained using different types of thermostats. Results show that at low temperature, where the harmonic approximation holds, the Nose'−Hoover (NH) thermostat fails to reproduce the statistical mechanical behavior, even using simulation lengths of millions of time steps, whereas the Gaussian isokinetic (IG) thermostat reproduces quite well the expected statistical mechanical values. The Berendsen's coupling (BC) provides good results for basic properties such as the average potential and kinetic energies but fails in reproducing the canonical fluctuations. Moreover, using the speed of divergence of initially nearby trajectories in phase space as a measure of the dynamical chaoticity, we found that the NH thermostat provides very slow divergence for the physical phase space degrees of freedom, concentrating most of its chaoticity in the dynamics of the thermostat virtual degree of freedom. On the contrary, the IG thermostat provides always highly diverging trajectories in phase space, characterized by a high chaoticity of each degree of freedom. Finally, the BC thermostat provides a moderate chaotic behavior for all of the degrees of freedom. Such results suggest that even assuming for both the "rigorous" algorithms (NH and IG) a full ergodic behavior the NH thermostat could require an extremely long time to achieve convergence of the time averaged properties.

1. Introduction

In classical molecular dynamics (MD) computer experiments time averages are assumed to be equivalent to ensemble averages (ergodic hypothesis). In fact, the Nekhoroshev theorem[1] suggests that this hypothesis can in general be valid in the infinite time limit; but for finite time simulations, one should be concerned with the time required for the trajectory to fill densely the phase space or at least to have a sampling sufficient to achieve that time averages of the observables of interest coincide with the ensemble averages.

In this paper, we analyze the results of computer experiments at constant temperature. The most widely used "rigorous" thermostating algorithms for MD simulations are the Gaussian isokinetic (IG) and Nose'−Hoover (NH) thermostats. The use of these algorithms, assuming ergodicity for the simulated trajectory, provides by way of time averages the correct statistical mechanical values of the system's properties. Even assuming a full ergodic behavior, the minimal time necessary to achieve a sufficient sampling of phase space still remains an open problem. The NH thermostat has been extensively studied for one-dimensional harmonic systems, where incorrect phase-space sampling was reported.[2−8] In many of these papers, more sophisticated NH-like algorithms have been proposed, which provide a correct phase space sampling at least of a system of one-dimensional harmonic oscillators. However, such more complex thermostats cannot be easily applied in general as the coupling parameters, necessary to obtain a correct sampling, are difficult to evaluate for each different case; moreover, they require a rather heavy computational load. For these reasons, the NH thermostat is still widely used for the simulation of realistic systems, assuming that the phase space sampling improves with the increase of dimension and complexity of the simulated system, although no systematic investigation has been performed in this direction.[4−6,8] On the other hand, the IG thermostat can produce a correct canonical distribution in the configurational space[9,10] but not in the momenta space, as the total kinetic energy is constrained. However, the momenta space properties can be easily obtained via basic statistical mechanics once the configurational distribution is available.[10] Note that the IG thermostat is not equivalent to the NH limit case of infinitely fast thermostat relaxation,[11] because the latter thermostat still allows kinetic energy fluctuations, whereas the IG method really constrains the total kinetic energy to a fixed value.

In this work, we investigate MD simulations of a small but yet dynamically complex system like a butane molecule, to obtain information on the reliability of MD simulations when different thermostats are used. We systematically compare three of the most used thermostating algorithms: the IG, the NH, and finally the Berendsen's coupling (BC);[12] the latter does not reproduce exactly the canonical distribution but is one of the most used thermostats because of its stability and efficiency in MD simulations. At low temperature, where the harmonic approximation holds, we could directly compare the behavior of significant quantities computed through time averages with

* To whom correspondence should be addressed. Phone: ++39-06-72594905. E-mail: andrea.amadei@uniroma2.it.
[†] University of Rome "Tor Vergata" Via della Ricerca Scientifica 1.
[‡] University of Rome "La Sapienza" Piazzale Aldo Moro 2.

10.1021/jp013689i CCC: $22.00
Published on Web 04/18/2002

the corresponding statistical mechanical values. At higher temperature, we performed an analysis of the dynamics of the system measuring the level of chaos associated with the trajectories. As a matter of fact, it has been found that there exists a threshold for the energy per degree of freedom (DOF), called strong stochasticity threshold (SST[13,14]), such that when the energy is higher a fast diffusion in phase space occurs (Anosov diffusion), whereas when the energy is lower, the dynamical diffusion is definitely slower (Arnold diffusion). To have equivalence between time and ensemble averages on the typical simulation time scale, the system must clearly be endowed with an energy per DOF higher than the SST. Nevertheless, this condition is not sufficient. It has been shown that even in systems that exhibit a chaotic behavior as a whole different DOFs may be endowed with quite different levels of chaos.[15] If such is the case, time averages of functions depending on the microscopic variables that have a slow diffusion in the phase space may turn out to be different from the ensemble averages, at least over times that are typical for a computer experiment. Here we used the rate of divergence in phase space of initially nearby trajectories, considering the whole system as well as each DOF, as a measure of the dynamical chaoticity of the system, according to Lyapunov's theory.[15−17] Fast diverging trajectories for each DOF should correspond to a chaotic sampling of phase space, with presumably a higher efficiency (fast diffusion), whereas trajectories characterized by slow divergence of even a subset of the DOFs should correspond to a partly "ordered" sampling of phase space, with presumably a lower efficiency (slow diffusion).

2. The Model

In this work, we used a skeletal model of a butane molecule (C_4H_{10}), where the hydrogen atoms were included into the carbons (united atoms model,[18] with four identical pseudoatoms). The mass of each pseudoatom is the mass of the whole molecule divided by 4, i.e., $m = [4(12) + 10(1)]/4 = 14.5$ amu. The potential energy consists of the sum of three different terms: the vibrational energy of the covalent bonds (stretching), the vibrational energy of the valence angles (bending), and the energy associated to the dihedral torsion. The stretching energy is represented[19] by a harmonic term $V_S = {}^1/_2\sum_{i=1}^3 k_i(b_i - b_0)^2$, where b_0 and b_i are respectively the equilibrium distance and the actual distance between atoms i and $i + 1$. The bending energy is represented[18] by a sum of quadratic terms of the cosine of the bending angle θ: $V_b = \sum_{i=1}^2 k_2(\cos\theta_i - \cos\theta_0)^2$, where θ_0 and θ_i are respectively the equilibrium angle and the actual angle between atoms i, $i + 1$, and $i + 2$. The potential energy entailed in the dihedral angle γ is represented[20] as $V_d = \sum_{i=1}^5 a_i \cos {}^i\gamma$. We have used the leapfrog algorithm for the numerical integration of the equations of motion, as it is suitable for the implementation of the thermostating techniques. The time step, as usual, is fixed as a small fraction of the shortest vibration period. In our case, we used a time step of 1 fs, which is ${}^1/_{20}$ of the stretching vibration period. With this choice, the fluctuations of the total energy, in microcanonical simulations, were less than 0.01%. The initial velocities were chosen in such a way as to obtain the required temperature and to set to zero the overall translational and angular momenta of the molecule. The used thermostats keep these constraints during the dynamical evolution. Simulations were performed in the temperature range $T = 15 \div 1100$ K using the IG,[9,10] NH,[2,11,21,22] and BC[12] thermostats, with initial equilibration runs of 0.1 ns and production runs in the range from 10 to 15 ns. Such long equilibration and production runs were used, especially when the NH thermostat is utilized,[23,24] to ensure equilibration from the initial phase space position of the MD trajectories and a good convergence of the time averages. Note that for the NH thermostat the length of the equilibration and production runs as well as the choice of the "mass" parameter τ are crucial to obtain a correct phase space sampling, in the simulation time length, as reported in the literature.[23−25] We always used time lengths of the equilibration and production runs larger than the ones used in the recent literature.[24] Concerning the "mass" parameter, we followed the indications given in a paper by Toxvaerd and Olsen,[26] where the authors identify the optimal value ($\tau = 0.01$) for an isolated butane molecule via a sensitivity analysis of different runs based on different τ values. Finally, we employed for the BC thermostat the usual relaxation time of 0.1 ps.[27]

3. Statistical and Dynamical Analysis

To compare the reliability of the two "rigorous" thermostats, we first investigated the low temperature regime. At low temperature, all classical mechanical systems, with an overall stationary point energy minimum, behave as harmonic ones, and the internal energy and heat capacity as a function of temperature can be easily obtained from statistical mechanics. One has to take into account that our trajectories, with center of mass at rest, have three constrained degrees of freedom (the center of mass coordinates). Moreover, at low temperature, where we deal with a semirigid molecule, fixing to zero the angular momentum is equivalent to constrain the three rotational degrees of freedom of the molecule. Hence, for our butane model, using basic statistical mechanics and the harmonic approximation, the molecular internal energy (considering the overall minimum as zero) and heat capacity are

$$U = \langle \mathcal{U} \rangle = \langle \mathcal{K} \rangle + \langle \mathcal{U}' \rangle = 6kT \tag{1}$$

$$C_V = \left(\frac{\partial U}{\partial T}\right)_V = \left(\frac{\partial \langle \mathcal{U} \rangle}{\partial T}\right)_V = \left(\frac{\partial \langle \mathcal{K} \rangle}{\partial T}\right)_V + \left(\frac{\partial \langle \mathcal{U}' \rangle}{\partial T}\right)_V = 6k \tag{2}$$

where k is Boltzmann's constant, $\mathcal{U}$, $\mathcal{K}$, and $\mathcal{U}'$ are the total, kinetic, and potenial energy, respectively. The total, kinetic and potential energy heat capacities, used in the previous formula, can be expressed in terms of the second central energy moments

$$\left(\frac{\partial \langle \mathcal{U} \rangle}{\partial T}\right)_V = \frac{\langle (\mathcal{U} - \langle \mathcal{U} \rangle)^2 \rangle}{kT^2} = 6k \tag{3}$$

$$\left(\frac{\partial \langle \mathcal{K} \rangle}{\partial T}\right)_V = \frac{\langle (\mathcal{K} - \langle \mathcal{K} \rangle)^2 \rangle}{kT^2} = 3k \tag{4}$$

$$\left(\frac{\partial \langle \mathcal{U}' \rangle}{\partial T}\right)_V = \frac{\langle (\mathcal{U}' - \langle \mathcal{U}' \rangle)^2 \rangle}{kT^2} = 3k \tag{5}$$

Note that the correlation $\langle (\mathcal{K} - \langle \mathcal{K} \rangle)(\mathcal{U}' - \langle \mathcal{U}' \rangle) \rangle$ in the canonical ensemble is zero even when the mass tensor is a function of the coordinates, and hence, kinetic and potential energies are not statistically independent (see appendix A). In Figure 1, we compare the average kinetic and potential energies, obtained by simulations, with the expected harmonic behavior $\langle \mathcal{K} \rangle = \langle \mathcal{U}' \rangle = 3kT$. For the average kinetic energy, we show only the values obtained by NH simulations, as in the IG simulations where the kinetic energy is by definition fixed to the correct value. In Figure 2, we compare the kinetic and potential energy heat capacity, obtained via the second central

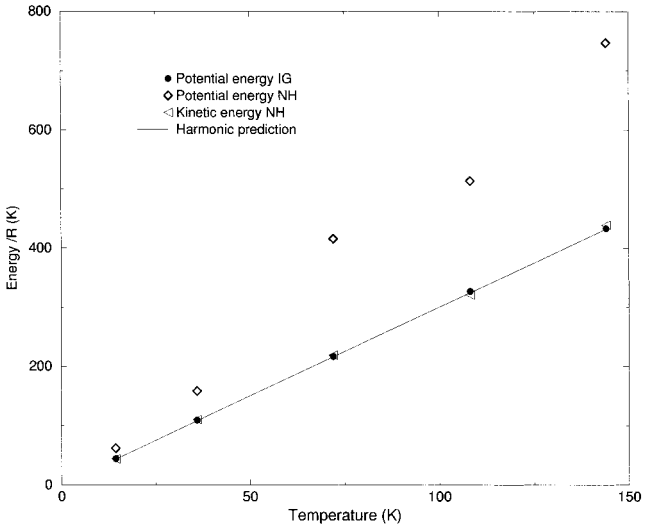

Figure 1. Average kinetic and potential energy for IG and NH simulations and theoretical harmonic behavior, as a function of temperature.

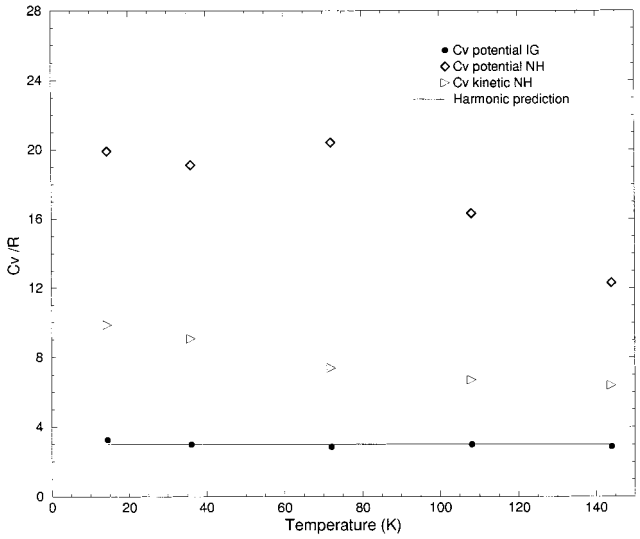

Figure 2. Kinetic and potential energy heat capacity for IG and NH simulations and theoretical harmonic behavior, as a function of temperature.

moment, with the expected harmonic behavior $(\partial\langle\mathscr{K}\rangle/\partial T) = (\partial\langle\mathscr{U}'\rangle/\partial T) = 3k$. As in the previous figure, the kinetic energy heat capacity is reported only for the NH simulations. For these latter simulations, we also compare, in Figure 3, the total energy heat capacity, obtained by the total energy second central moment and by the sum of the kinetic and potential energy second central moments, with the expected harmonic behavior given by formula 2. From these three figures, it is evident that at low temperature, where the harmonic approximation should hold, only the IG thermostat simulations show a good agreement with the expected behavior, whereas the results obtained with the NH thermostat are clearly incorrect, except for the average kinetic energy. This suggests that NH simulations, at least at low temperature, could require an extremely long time to provide well converged time averages. The exceedingly high values of the fluctuations in the NH simulation are reminiscent of a similar phenomenon found at low temperature in fcc Lennard-Jones lattices with up to 32 768 atoms.[28] In that case, the computer experiment simulated a microcanonical ensemble, and the large and long lasting fluctuations of the kinetic (or potential) energy

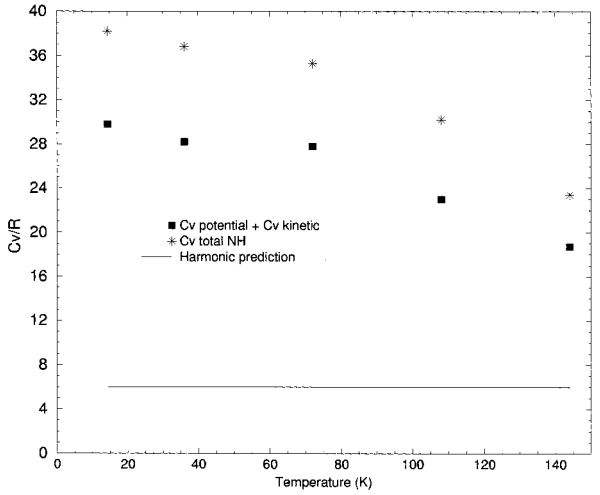

Figure 3. Total (energy) heat capacity and sum of kinetic and potential energy heat capacities for NH simulations, compared with the theoretical harmonic behavior, as a function of temperature.

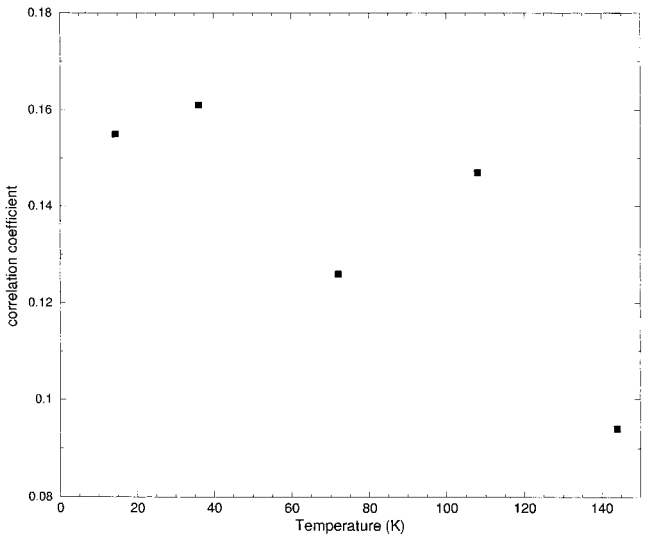

Figure 4. Correlation coefficient between average kinetic and potential energies for NH simulations, as a function of temperature.

were due to a lack of interaction among normal modes. We will resume this point later. The poor convergence of the NH simulation results is also demonstrated by the nonzero correlation coefficient

$$r = \frac{\langle(\mathscr{K} - \langle\mathscr{K}\rangle)(\mathscr{U}' - \langle\mathscr{U}'\rangle)\rangle}{[\langle(\mathscr{K} - \langle\mathscr{K}\rangle)^2\rangle\langle(\mathscr{U}' - \langle\mathscr{U}'\rangle)^2\rangle]^{1/2}}$$

between kinetic and potential energies, as shown in Figure 4. As previously mentioned, it is well-known that for a one-dimensional harmonic chain the NH thermostat provides incorrect sampling of phase space, at least within the simulation time.[2−8] The results presented here show rather clearly that such an incorrect sampling is extended to three-dimensional harmonic systems. We also studied the NH limit of infinitely fast thermostat relaxation ($\tau \rightarrow 0$). For numerical reasons in such a limit, the mass parameter must be set equal to the time step. It is worth noting that these NH-limit simulations do not reproduce the IG results and still provide incorrect properties. This is summarized by Table 1, where we use data at 140 K

TABLE 1

	limit NH potential	IG potential	Harmonic prediction
U/R (K)	523	432	432
C_V/R	4.4	2.8	3.0

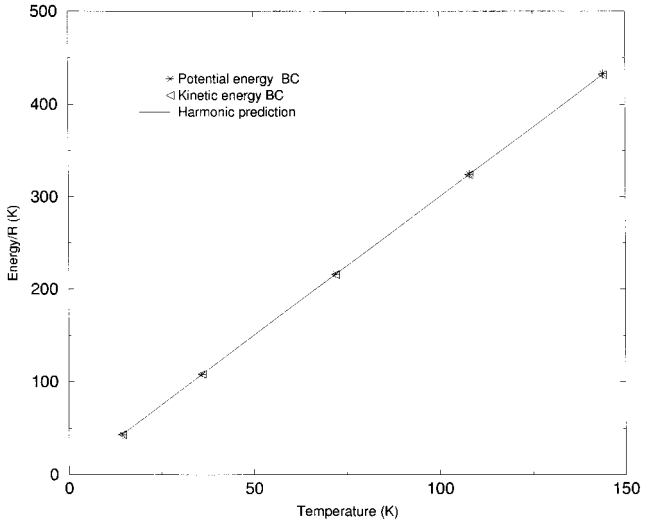

Figure 5. Average kinetic and potential energy for BC simulations and theoretical harmonic behavior, as a function of temperature.

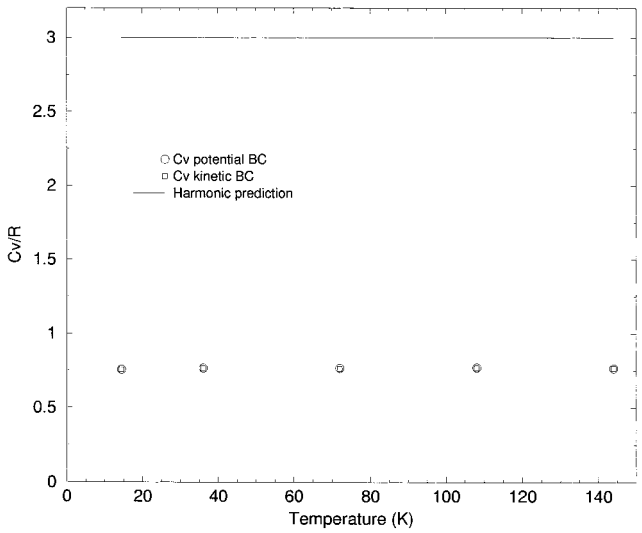

Figure 6. Kinetic and potential energy heat capacity for BC simulations and theoretical harmonic behavior, as a function of temperature.

comparing the values of average energy and heat capacity for both the limit-NH thermostat and the IG one. We also investigated the low temperature behavior of the BC thermostat, and in Figures 5 and 6, we show the average kinetic and potential energies and corresponding heat capacities as obtained by MD trajectories with the BC algorithm. This nonexact temperature coupling reproduces correctly the average potential and kinetic energies but not more complex properties based on higher order energy moments or correlations, like the heat capacity. Interestingly, the approximate fluctuation formula for the BC thermostat, recently proposed in ref 29 in order to obtain correct heat capacities values, is unable to reproduce the expected values providing still too low kinetic and potential energy heat capacities (R instead of $3R$).

At higher temperature, where the system is no more in the harmonic regime, there are no simple prescriptions for the behavior of $U(T)$ and $C_V(T)$. To ascertain the reliability of the simulations' results, we have investigated the rate of divergence

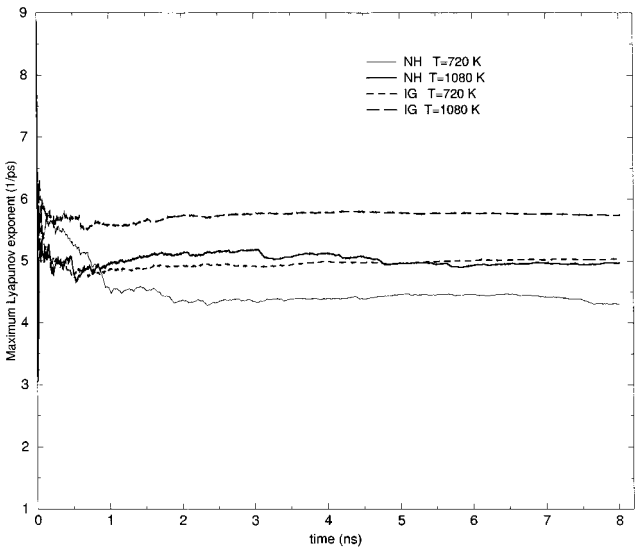

Figure 7. Maximum Lyapunov exponent for IG and NH simulations as function of time, at different temperatures.

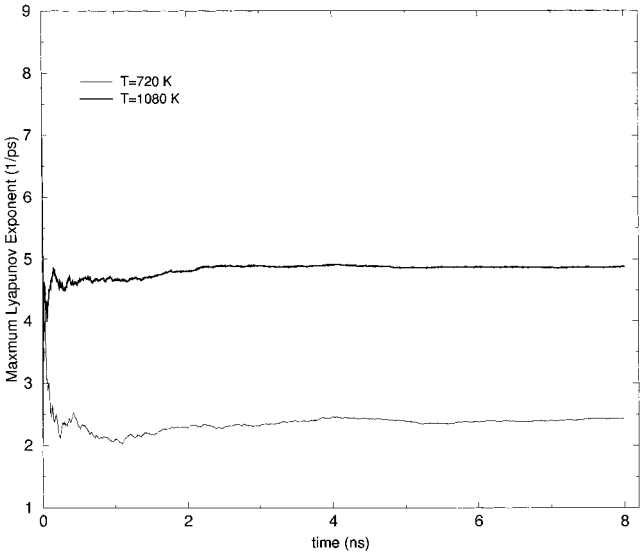

Figure 8. Maximum Lyapunov exponent for BC simulations as function of time, at different temperatures.

of nearby trajectories by means of the maximum Lyapunov exponent,[30,31] defined as

$$\lambda_1 = \lim_{t\to\infty} \frac{1}{t} \ln \frac{|\mathbf{w}(t)}{\mathbf{w}(0)} \tag{6}$$

where $\mathbf{w}$ is the distance vector in phase space of two trajectories starting from close points ($\mathbf{w}(0)$ tending to zero). Note that in such a limit the vector $\mathbf{w}$ defines the "tangent space" of the phase space of the system. A positive value of λ_1 means that the trajectories will diverge exponentially in the long-time scale. In Figure 7, we show the maximum Lyapunov's exponent λ_1 vs time, for both the IG and NH simulations at different temperatures. From the figure, it is clear that in the IG simulations the convergence is faster; moreover, we obtain larger values of λ_1. As previously mentioned, the larger the value of λ_1, the faster the divergence of the trajectories in phase space; we can hence conclude that the IG simulations provide a more chaotic sampling of phase space. Similarly, in Figure 8, we show the maximum Lyapunov's exponent for the BC trajectories. For this last thermostat, we observe a convergence rate similar to

the IG case, but the values are lower than the IG ones. To investigate the chaoticity of single DOFs in the system, it is convenient to use generalized coordinates which are, in general, different from the Cartesian ones. We used the essential dynamics (ED)[32−37] method to separate DOFs with possibly a different level of chaos. In this approach, which is based on the principal component analysis of the atomic positional fluctuations, it is possible to separate the DOFs of the system into two different sets: an "essential" one, which is responsible for the larger portion of the positional fluctuations, and the remaining one, containing most of the DOFs, characterized by nearly constrained motions. The definition of the essential DOFs is obtained from the diagonalization of the covariance matrix of the atomic displacements, with the essential DOFs being identified by the eingenvectors associated to the larger eigenvalues. The eigenvalues are the average squared fluctuations (variances) in the configurational space, occurring along the corresponding eigenvectors, and hence their sum provides a measure for the amount of sampled configurational phase space. This is a different but complementary information with respect to that obtained from the maximum Lyapunov exponent, and we have found that they are in agreement, as the total squared fluctuation of the simulations with the IG thermostat is always higher than that with the NH and BC thermostats (data not shown). Because we are interested in the chaoticity of single DOFs, i.e., in the rate of divergence of single DOFs on initially nearby trajectories, we can quantify this property using the coherence angles CAs defined by[15−17]

$$\cos^2 \alpha_i = \lim_{t\to\infty} \frac{1}{t} \int_0^t \frac{|\mathbf{w}(t')|^2}{|\mathbf{w}(t')|^2} \, dt' \tag{7}$$

where $\mathbf{w}_i$ is the vector obtained by projecting $\mathbf{w}$ on the ith DOF plane, defined by a coordinate and its conjugated momentum (the momentum of an eigenvector DOF can be easily obtained projecting the Cartesian momenta onto the eigenvector); α_i is the ith CA corresponding to the asymptotic value (infinite time limit) of the angle between the ith DOF plane and the direction corresponding to the maximum Lyapunov's exponent (see appendix B). A DOF associated with a small CA will diverge with a high rate, being aligned close to the maximum Lyapunov's exponent direction (maximum expansion direction); it will hence be characterized by a highly chaotic dynamics. On the contrary, DOFs associated with high CA values will be characterized by a slow divergence and, hence, by a less chaotic dynamics. As $\sum_{i=1}^{N} \cos^2 \alpha_i = 1$, it is possible to define an em average angle of the DOFs α as

$$\alpha = \arccos\sqrt{\frac{1}{N}\sum_{i=1}^{N}\cos^2 \alpha_i} = \arccos\sqrt{\frac{1}{N}} \tag{8}$$

Deviations of α_i from the average angle give a measure of the different level of chaos of the DOFs. In the case of IG and BC simulations, the total number of DOFs is 12, and hence the average angle is 73°; for NH simulations, there is one more DOF, and the average angle becomes 74°. Because the CAs are defined by asymptotic time averages, they become meaningful only when a good convergence is reached. Such a convergence was obtained for the IG, NH, and BC simulations on a time scale between 10 and 15 ns. In Figure 9, we report the CAs for two high temperature IG and NH simulations. As expected, in all simulations, the angles of the DOFs 10, 11, and 12 are exactly equal to 90°, being therefore orthogonal to the maximum expansion direction. These DOFs, corresponding to the zero eigenvalues of the covariance matrix, are in fact associated with the three exact holonomic constraints that fix the center of mass of the molecule. For the other DOFs, the situation is quite different. In the case of the IG thermostat, the physically relevant DOFs, from 1 to 9, are collected in a rather narrow strip around the average angle, and this means that these DOFs behave chaotically and in a similar way (the structure of this spectrum of CAs will be analyzed in a forthcoming paper[38]). On the contrary, from the simulations with the NH thermostat, the DOFs from 1 to 9 exhibit angles close to 90°, and only the angle of the 13th DOF, corresponding to the added thermostat variable, is quite low (16° at $T = 720$ K and 20° at $T = 1080$ K). These last results mean that the direction associated in the tangent space with the added variable is close to the maximum expansion direction, wich is almost orthogonal to the other, physical DOFs. The 13th DOF entails hence most of the chaoticity of the system, whereas the motions in the physical subspace are characterized by a very low chaos. This kind of pattern is probably the reason underlying the large, noncanonical undulations of the instantaneous kinetic energy found some years ago in a two-dimensional Lennard-Jones fluid simulated with a NH thermostat.[23] In that case, as in the case of the fcc lattice mentioned before,[28] a poor energy transfer among DOFs restricted the diffusion of the simulated trajectory in the phase space, producing results not matching the statistical mechanical expectations. In the simulations of the butane molecule we performed using the NH thermostat, only the thermostat variable seems to be endowed with a motion chaotic enough to allow an efficient sampling of its phase subspace. In Figure 10, we also compare at the same temperatures as in Figure 9 the IG CAs with the ones obtained by the NH thermostat in the limit of infinitely fast relaxation. It is worth noting that once again the limit-NH thermostat is not equivalent to the IG one and provides virtually the same CAs pattern of the NH simulations using the optimal "mass" parameter. In Figure 11, we present the same analysis for two BC and microcanonical (MC) high-temperature simulations. From this last figure, it is clear that BC and MC simulations provide a set of CAs similar to the IG ones.

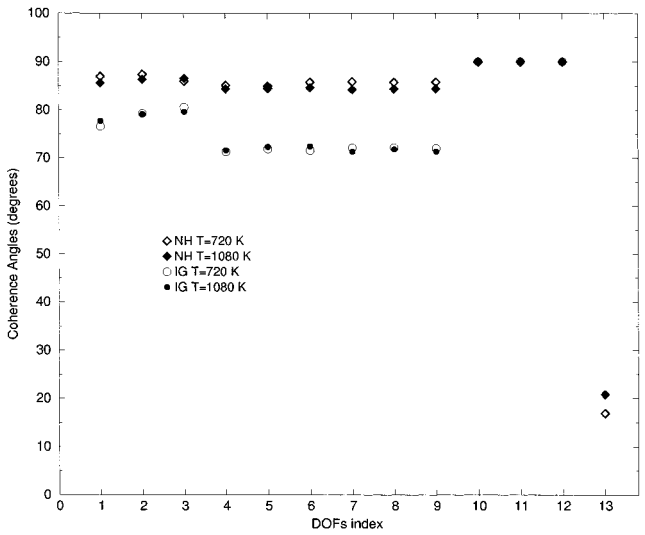

Figure 9. Coherence angles spectrum for IG and NH simulations at different temperatures.

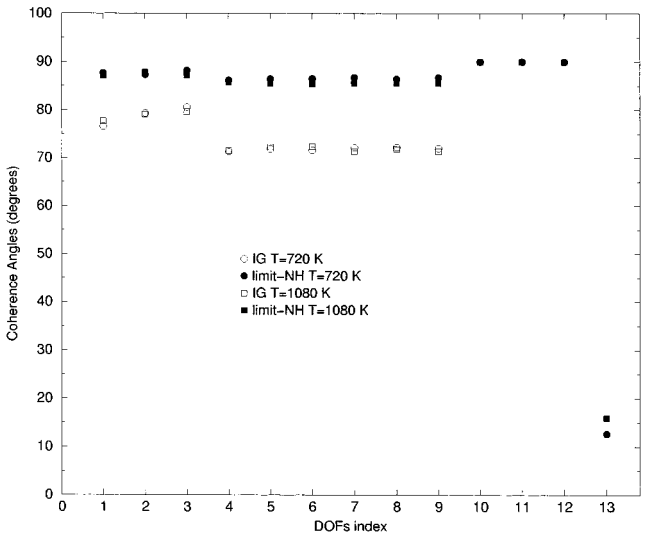

Figure 10. Coherence angles spectrum for IG and limit NH ($\tau \rightarrow 0$) simulations at different temperatures.

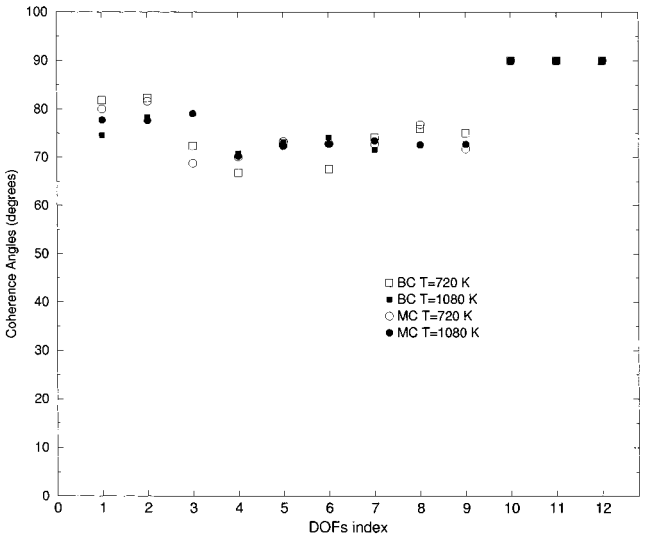

Figure 11. Coherence angles spectrum for BC and MC simulations at different temperatures.

4. Conclusions

Molecular dynamics simulations in ensembles other than the microcanonical one involve the use of specific thermostating algorithms that should reproduce the correct statistical mechanics of the simulated system. Even assuming that the ergodic hypothesis holds, this is possible only if in the time of the simulation the time averaged properties converge, i.e., the simulated trajectory fills densely all of the physically allowed phase space. The correct choice of an efficient thermostating algorithm is essential. In this work, we compared the MD results obtained using the IsoGaussian, IG, and the Nose' Hoover, NH thermostats, which should provide in principle the correct canonical statistical mechanics. We also compared the previous coupling methods with the BC thermostat, which is not exact but is widely used.

At low temperature, where the harmonic approximation should hold, the direct comparison of some significant physical properties, such as the internal energy and heat capacity, with the expected theoretical values clearly shows that the NH thermostat is unable to reproduce the expected behavior, whereas the IG thermostat does. Interestingly, the BC thermostat does provide correct average energy values, but it cannot reproduce properly the energy fluctuations. At higher temperature, where there are no simple theoretical prescriptions, we used some indirect indicators (maximum Lyapunov exponent λ_1 and coherence angles) to obtain information on the efficiency of MD sampling. Our results indicate that the IG algorithm produces trajectories characterized by a highly chaotic behavior, suggesting a fast sampling of phase space. On the contrary, the NH thermostated trajectories seem to be endowed with a partly ordered dynamics in the physical phase space, with most of the chaoticity concentrated in the behavior of the added nonphysical variable (thermostat variable). The BC simulations, similarly to the MC simulations, are somewhat between the NH and IG ones, showing a moderate chaoticity of the dynamics. From these results, we can conclude that, at least for our small system and for the time lengths investigated, only the IG simulations can reproduce accurately the physical behavior, providing the required sampling and the expected statistics. The phase space sampling of the NH thermostated trajectories probably requires exceedingly long times to provide correct time averages, being characterized by a poorly chaotic behavior. Finally, the BC simulations, although unable to provide the correct fluctuations and correlations, seem to converge and be reliable as far as more basic properties are concerned.

Acknowledgment. We thank Prof. H. Posh for his advice in the choice of the model. We acknowledge Professors A. Di Nola and H. J. C. Berendsen for stimulating discussions. We also gratefully acknowledge the computational help of R. Spezia.

Appendix A

Consider a molecular system in the classical limit, where the kinetic energy is given by

$$\mathcal{K} = \boldsymbol{\pi}^T \frac{\tilde{M}^{-1}}{2} \boldsymbol{\pi} \tag{9}$$

where $\boldsymbol{\pi}$ are the conjugated momenta of the generalized coordinates $\boldsymbol{\xi}$ and $\tilde{M}$ is the mass tensor of the system, a symmetric matrix, which is in general not diagonal and depends on the coordinates. The covariance between kinetic and potential energy can be obtained from

$$\begin{aligned}\langle \Delta\mathcal{K}\Delta\mathcal{U}' \rangle &= \frac{\int e^{-\beta\mathcal{U}} \Delta\mathcal{U}' \, d\xi \int e^{-\beta\mathcal{K}} \Delta\mathcal{K} \, d\pi}{\int e^{-\beta\mathcal{U}'} \, d\xi \int e^{-\beta\mathcal{K}} \, d\pi} \\ &= \frac{\int e^{-\beta\mathcal{U}'} (\det \tilde{M})^{1/2} \Delta\mathcal{U}' (\langle K \rangle_\xi - \langle K \rangle) \, d\pi}{\int e^{-\beta\mathcal{U}'} (\det \tilde{M})^{1/2} \, d\xi}\end{aligned} \tag{10}$$

where

$$\Delta\mathcal{K} = \mathcal{K} - \langle \mathcal{K} \rangle \tag{11}$$

$$\Delta\mathcal{U}' = \mathcal{U}' - \langle \mathcal{U}' \rangle \tag{12}$$

$$\langle K \rangle_\xi = \frac{\int e^{-\beta\mathcal{K}} \mathcal{K} \, d\pi}{\int e^{-\beta\mathcal{K}} \, d\pi} \tag{13}$$

and the integrals in the last equation are defined at fixed $\boldsymbol{\xi}$. Using in eq 13 the $\boldsymbol{\pi}$ sets that diagonalize the mass tensor at

each configuration, we obtain that the corresponding average kinetic energy, evaluated at fixed coordinates, is

$$\langle K\rangle_{\xi} = \frac{1}{2}\sum_{i=1}^{\mathcal{N}} \frac{\langle \pi_i^2\rangle_{\xi}}{M_i} = \left(\frac{kT}{2}\right)\mathcal{N} \tag{14}$$

$$\langle \pi_i^2\rangle_{\xi} = kTM_i \tag{15}$$

where M_i is the ith diagonal element of the diagonalized mass tensor, $\mathcal{N}$ is the total number of degrees of freedom, and each M_i is a function of the coordinates $\boldsymbol{\xi}$. Note that because we used an orthogonal transformation of the momenta to diagonalize the mass tensor, and hence its inverse, the corresponding Jacobian in eq 13 is 1. From the last equations, we readily obtain

$$\langle K\rangle = \langle K\rangle_{\xi} = \left(\frac{kT}{2}\right)\mathcal{N} \tag{16}$$

which must be valid for any possible configuration, and hence $\langle \Delta\mathcal{K}\Delta\mathcal{U}'\rangle = 0$

Appendix B

In this appendix, we show how we can relate the dynamics in the tangent space and the expansion directions characterized by the Lyapunov exponents with the coherence angles. We will use a simplified mathematical approach without addressing the more complex details. A mathematically rigorous derivation of the relevant properties of the tangent space is given in refs 30 and 31. A detailed description of the CAs can be found in refs 15−17.

Consider the Hamiltonian form of the equations of motion defined by a first order set of ordinary differential equations

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, t) \tag{17}$$

where $\boldsymbol{x}(t, \boldsymbol{x}_0)$ is the phase space point describing the trajectory which is fully defined as a function of time and initial position $\boldsymbol{x}_0$. Varying the previous equation with respect to the initial position $\boldsymbol{x}_0$ and considering that this derivation commutes with the time derivative, we obtain

$$\dot{\boldsymbol{w}} = \tilde{A}\boldsymbol{w}$$

where

$$w_i = \sum_j \left(\frac{\partial x_{\mathrm{i}}}{\partial x_{0,j}}\right)\Delta x_{0,j}$$

and

$$A_{i,j} = \left(\frac{\partial f_i}{\partial x_j}\right)$$

hence

$$\boldsymbol{w}(t) = \tilde{T}_{t-\tau}\boldsymbol{w}(t-\tau) \tag{18}$$

$$\tilde{T}_{t-\tau} = \tilde{I} + \tilde{A}_{t-\tau}\tau \tag{19}$$

with τ a differential time interval. Dividing the whole time t into a set of L differential intervals we obtain, for $t \geq \tau$:

$$\boldsymbol{w}(t) = \Pi_{l=1}^{L}\, \tilde{T}_{(l-1)\tau}\boldsymbol{w}_0 \tag{20}$$

where $\boldsymbol{w}_0 = \boldsymbol{w}(0)$. Expressing the matrix product in terms of exponential matrixes, we have

$$\Pi_{l=1}^{L}\tilde{T}_{(l-1)\tau} = \Pi_{l=1}^{L}\mathrm{e}^{\tilde{E}_{(l-1)\tau}} = \mathrm{e}^{\tilde{Z}\mathrm{L}} \tag{21}$$

where

$$\tilde{Z} = \frac{1}{L}\sum_{l=1}^{L}\tilde{E}_{(l-1)\tau} \tag{22}$$

hence

$$\boldsymbol{w}(t) = \mathrm{e}^{\tilde{Z}t/\tau}\boldsymbol{w}_0 \tag{23}$$

Let us assume that in the limit $L \rightarrow \infty$ $\tilde{Z}$ converges to a stable value, i.e., the time average becomes stationary. Given the transformation $\tilde{B}$ that diagonalizes $\tilde{Z}$, it follows from the Taylor expansion of the exponential matrix that it diagonalizes also $\mathrm{e}^{\tilde{Z}}$; we can therefore write for the long time scale:

$$\boldsymbol{w}(t) = (\tilde{B}e^{\tilde{\Theta}}\tilde{B}^{-1})^{t/\tau}\,\boldsymbol{w}_0 = \tilde{B}e^{\tilde{\Theta}t/\tau}\tilde{B}^{-1}\boldsymbol{w}_0 \tag{24}$$

with

$$\tilde{\Theta} = \tilde{B}^{-1}\tilde{Z}\tilde{B} \tag{25}$$

and

$$(\mathrm{e}^{\tilde{\Theta}t/\tau})_{i,i} = \mathrm{e}^{\Theta_{i,i}t/\tau} \tag{26}$$

where the time independent matrix $\tilde{\Theta}$ is diagonal and hence $\mathrm{e}\tilde{\Theta}$ is also diagonal; $\Theta_{i,i}$ is in general a complex number such that $\Theta_{i,i}/\tau = \lambda_i + (-1)^{1/2}\gamma_i/\tau$, where λ_i corresponds to the ith Lyapunov exponent and the ith column of $\tilde{B}$ provides the corresponding eigenvector in the original basis set. Note that the convergence of $\tilde{Z}$ is the key point in this derivation, and we assume it without addressing all of the necessary mathematical requirements.

In the infinite time limit, we hence have, from eq 24

$$w_i = \mathrm{e}^{\Theta_{1,1}t/\tau}w_1'(0)B_{i,1} \tag{27}$$

where $\boldsymbol{w}'(0) = \tilde{B}^{-1}\boldsymbol{w}_0$ and hence $w_1'(0)$ is the component of $\boldsymbol{w}_0$ along the eigenvector corresponding to the largest Lyapunov exponent, obtained expressing the vector of the tangent space in the eigenvectors basis set. From the last equation, it is easy to derive eq 6 and the angle between the maximum expansion direction and a given subspace:

$$\cos^2\alpha = \frac{\sum_i {b_{i,1}}^2}{\sum_i {B_{i,1}}^2} \tag{28}$$

where $b_{i,1}$ is the component along the ith vector of the original basis set of the projection on a given subspace of the eigenvector corresponding to λ_1. The value of $\cos^2\alpha_i$ as computed through eq 7 can be formally expressed through the last equation when the subspace is defined by a coordinate and its conjugated momentum.

References and Notes

(1) Nekhoroshev, N. N. *Russ. Math. Surv.* **1977**, *32*, 1.
(2) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(3) Hoover, W. G.; Holian, B. L. *Phys. Lett. A* **1996**, *211*, 253.
(4) Winkler, R. G. *Phys. Rev. A* **1992**, *45*, 2250.
(5) Martina, G. J.; Klein, M. L. *J. Chem. Phys.* **1992**, *97*, 2635.
(6) Bulgac, A.; Kusnezov, D. *Phys. Rev. A* **1990**, *42*, 5045.
(7) Posch, H.; Hoover, W. G. *Phys. Rev. A* **1989**, *39*, 2175.

(8) Posch, H.; Hoover, W. G.; Vesely, F. *Phys. Rev. A* **1986**, *33*, 4253.
(9) Evans, D. J.; Morris, G. P. *Phys. Lett.* **1983**, *98A*, 433.
(10) Amadei, A.; et al. *J. Chem. Phys.* **2000**, *112*, 9.
(11) Evans, D. J.; Holian, B. L. *J. Chem. Phys.* **1985**, *83*, 4069.
(12) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(13) Pettini, M.; Landolfi, M. *Phys. Rev. A* **1990**, *41*, 768.
(14) Pettini, M.; Cerruti-Sola, M. *Phys. Rev. A* **1991**, *44*, 975.
(15) D'Alessandro, M.; Tenenbaum, A. *Phys. Rev. E* **1995**, *52*, R2141.
(16) D'Alessandro, M.; D'Aquino, A.; Tenenbaum, A. *Physica A* **1997**, *240*, 115.
(17) D'Alessandro, M.; D'Aquino, A.; Tenenbaum, A. *Phys. Rev. E* **2000**, *62*, 4809.
(18) Toxvaerd, S. *J. Chem. Phys.* **1990**, *93*, 4290.
(19) $b_0 = 0.39$ Å. Posch H. private communication.
(20) Ryckaert, J. P.; McDonald, I. R.; Klein, M. L. *Mol. Phys.* **1989**, *67*, 957.
(21) Nosè, S. *J. Chem. Phys.* **1984**, *81*, 511.
(22) Evans, D. J. *J. Chem. Phys.* **1983**, *72*, 3297.
(23) Holian, B. L.; Voter, A. F.; Ravelo, R. *Phys. Rev. E* **1995**, *52*, 2338.
(24) Cho, K.; Joannopoulos, J. D. *Phys. Rev. A* **1992**, *45*, 7089.
(25) Di Tolla, F. D.; Ronchetti, M. *Phys. Rev. E* **1993**, *48*, 1725.
(26) Toxvaerd, S.; Olsen, O. H. *Phys. Scr., T* **1990**, *33*, 98.
(27) van Gusteren, W. F.; Berendsen, H. J. C.; Biomos, B. V. University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands.
(28) Simonazzi, R.; Tenenbaum, A. *Phys. Rev. E* **1996**, *54*, 964.
(29) Morishita, T. *J. Chem. Phys.* **2000**, *113*, 2976.
(30) Osedelec, V. I. *Trans. Moscow Math. Soc.* **1968**, *19*, 197.
(31) Benettin, G.; Galgani, L.; Strelcyn, J. M. *Phys. Rev. A* **1976**, *14*, 2338.
(32) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. *Proteins: Struct., Funct. Genet.* **1993**, *17*, 412.
(33) van Alten, D. M. F.; Amadei, A.; Linssen, A. B. M.; Eijsink, V. G. H.; Vriend, G.; Berendsen, H. J. C. *Proteins: Struct., Funct. Genet.* **1995**, *22*, 45.
(34) van Alten, D. M. F.; Findlay, J. B. C.; Amadei, A.; Berendsen, H. J. C. *Protein Eng.* **1996**, *8*, 1129.
(35) Amadei, A.; Linssen, A. B. M.; de Groot, B. L.; van Alten, D. M. F.; Berendsen, H. J. C. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615.
(36) de Groot, B. L.; Amadei, A.; van Alten, D. M. F.; Berendsen, H. J. C. *J. Biomol. Struct. Dyn.* **1996**, *13*, 741.
(37) de Groot, B. L.; Amadei, A.; Scheek, R. M.; van Nuland, N. A. J.; Berendsen, H. J. C. *Proteins: Struct., Funct. Genet.* **1996**, *26*, 314.
(38) D'Alessandro, M.; Tenenbaum, A.; Amadei, A. Manuscript in preparation.