J. Elezgaray[1]
Y. H. Sanejouand[2]
[1] *Centre de Recherche Paul Pascal, Avenue Schweitzer, 33600 Pessac, France*

[2] *Laboratoire de Physique Quantique, UMR5626, IRSAMC-UPS, 118 route de Narbonne, 31602 Toulouse Cedex, France*

Received 9 January 1998; accepted 9 July 1998

Modeling Large-Scale Dynamics of Proteins

Abstract: *We study a dynamical model for the large-scale motions of bovine pancreatic trypsin inhibitor in vacuum. The model is obtained by projecting Newton equations onto some set of anharmonic modes. We compare the statistics of the so-obtained trajectories with those obtained by standard techniques, and conclude that our dynamical model is able to reproduce fairly well the average properties of the large-scale motions of this protein. Moreover, it allows for time steps one order of magnitude larger than the standard ones.* © 1998 John Wiley & Sons, Inc. Biopoly 46: 493–501, 1998

Keywords: *large-scale dynamics; proteins; bovine pancreatic trypsin inhibitor; Newton equations; anharmonic modes*

INTRODUCTION

It is well known[1–6] that protein motion, as viewed by simulations of molecular dynamics, is confined to some subspace of very small dimension, say about 1% of the total number of degrees of freedom of the protein. This can be checked in several ways. The one used in this paper will be based on the existence of some particular set (basis) of modes $\{\varphi_n(i),\ n = 1, \ldots, N\}$, such that the configuration of the protein at time t is

$$X_i(t) = X_i^0 + \sum_{n=1}^{N} c_n(t)\varphi_n(i) + \text{corrections} \quad (1)$$

where the index i runs over the N_a atoms of the protein, and X_i^0 is some average configuration [X_i is the position of the i^{th} atom, and correspondingly, $\varphi_n(i)$ is the amplitude of mode n at the i^{th} atom]. Typically, if $N \sim 10^{-2}\ 3N_a$, then

$$\frac{\|\text{corrections}\|^2}{\|X_i(t)\|^2} \sim 10^{-2}$$

[the notation $\|X\|$ is just $(\Sigma X_i^2)^{1/2}$]. At least two different sets of basis have been used in the literature to perform decompositions of type (1). The more natural ones are harmonic modes, obtained from the

Correspondence to: J. Elezgaray
Biopolymers, Vol. 46, 493–501 (1998)

CCC 0006-3525/98/070493-09

diagonalization of the hessian of the potential energy computed on some stationary state X_i^s. It has been shown,[7,8] for instance, that for some proteins, the transition from one particular configuration to another one (corresponding to a "conformational change" of the protein) is fairly well reproduced by the direction pointed by the less "energetic" of these modes, i.e., the one corresponding to the coefficient $c_n(t)$ that fluctuates the more (in amplitude).

A generalization of this basis is given by the so-called anharmonic modes (or quasi-harmonic modes, cf. Refs. 9–11), obtained from the diagonalization of the correlation matrix:

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T (X_i(t) - X_i^0)(X_j(t) - X_j^0)\, dt \qquad (2)$$

These modes generalize the harmonic modes to circumstances where the nonlinearities of the force field are important, and boil down to the harmonic modes for pure harmonic motion (a fact never encountered in reality). In the following, we will prefer this description to the one obtained with the harmonic modes. The main reason of this choice is the well-known property that anharmonic modes are "optimal" for the description of the trajectories from which they have been computed.[12] This optimality property means that, for any approximated decomposition of type (1) obtained with any (orthogonal) basis $\psi_n(i)$, one has

$$\sum_{n=1}^{N} \langle c_n^2 \rangle \geq \sum_{n=1}^{N} \langle d_n^2 \rangle \qquad (3)$$

if

$$X_i(t) = X_i^0 + \sum_{n=1}^{N} d_n(t)\psi_n(i) + \text{ corrections} \qquad (4)$$

(the angle brackets denote average in time) when the anharmonic modes are used to compute the coefficients $c_n(t)$. In more intuitive terms, the anharmonic decomposition captures, *on the average,* more rapidly than any other orthogonal decomposition the fluctuations around the average configuration X_i^0. The obvious drawback of this decomposition is the fact that the correlation matrix has to be computed, which in principle means that some (long enough) trajectory is available (cf. Refs. 13 and 14 for a discussion of this point).

An important fact about this type of protein motion description is that the most important modes (in the sense that the associated fluctuations $\langle c_n^2 \rangle$ are predominant) are both large scale (i.e., distributed over significant regions of the protein) and correspond to slow motions. The question we would like to address in this paper is about the possibility of obtaining evolution equations involving only those variables $c_n(t)$ corresponding to the slow and large-scale motions of the protein. A positive answer to this question would clearly mean that, for this set of equations, large time steps can be used, thus allowing significant longer periods of numerical integration. The price to be paid is that only large scales are taken into account, small scale motions (at the scale of the angstrom) being completely overlooked.

To the best of our knowledge, this question has never been addressed in the context of protein motion, except for "small" molecules (a ten amino acid peptide, a 32 atoms chain, etc.) and using harmonic modes.[15,16] Several authors[1–6] have been using the anharmonic decomposition with the purpose of both describing protein fluctuations and exploring in an efficient way the space of possible conformations, but not with the purpose of obtaining dynamical models for large scale motions. The only paper we know in the line of ours is Ref. 17, where a change of variables is made in order to describe in a hierarchical way the protein motion. We think that the approach presented here is complementary to that of Ref. 17.

In the next section, we review some of the mathematical results concerning the large scale motions of macromolecules. In the third section, we present the results obtained with our approach in the case of a small protein, bovine pancreatic trypsin inhibitor (BPTI), widely used as a model both by experimentalists and theoreticians, and the fourth section summarizes the main results we obtained.

SLOW MOTION OF MACROMOLECULES

The equations of motion of a macromolecule, in the approximation usually made in molecular dynamics, are Newton equations (written in weighted coordinates),

$$\frac{d^2X_i}{dt^2} = -\text{grad}_{X_i} U(\{X_j\})$$

with a potential energy $U(\{X_i\})$ of the form

$$U(\{X_i\}) = \frac{1}{\epsilon^2} U_{\text{fast}}(\{X_i\}) + U_{\text{slow}}(\{X_i\}) \qquad (5)$$

The first term U_{fast} includes the energy associated to the length of the bonds and the covalent angles, whereas the second term contains all the electrostatic (long-range) interactions, as well as torsional energy. The splitting of the potential energy in these two terms makes clear that, for a macromolecule, there are at least two very different scales of motion. The (small) value of $\epsilon \sim 0.1$ is such that the usefulness of a limiting equation (for $\epsilon \to 0$) seems justified.[18–20] The existence of such an equation is known at least from the work of Takens.[20] The intuition is that, as ϵ goes to zero, the molecule has to stay closer and closer to the "equilibrium" manifold defined by the constraint $U_{\text{fast}} = 0$. Thus, a possible candidate for the limiting equation is simply

$$\frac{d^2X_i}{dt^2} = -\text{grad}_{X_i} U_{\text{slow}}(\{X_j\}) - \lambda D U_{\text{fast}} \qquad (6)$$

where the last term is the Lagrange multiplier associated to the constraint $U_{\text{fast}} = 0$. However, it is known that in order to preserve time averages, correcting potentials have to be added to Eq. (6). The (probably correct) mathematical approach, as recently reviewed in Ref. 19, gives an explicit expression for this correcting potential under the form

$$U_{\text{corr}} = \sum_{i=1}^{N_a} \theta_i \omega_i,$$

the ω_i^2's being the eigenvalues of the hessian of U_{fast} and the θ_i's being constants only dependent on initial conditions. This is based on the use of a rigorous result called the "adiabatic theorem" (cf. Ref. 19), and makes no reference whatsoever to temperature or any statistical mechanics consideration. However, the resulting corrected equations

$$\frac{d^2X_i}{dt^2} = -\text{grad}_{X_i}[U_{\text{slow}}(\{X_j\}) + U_{\text{corr}}(\{X_j\})] - \lambda D U_{\text{fast}}, \qquad (7)$$

are of limited numerical interest, as the computation of U_{corr} seems, at first sight, harder than the integration of the original equation. It should be noted that from the results of Ref. 21, the correcting potential in the case of proteins has a significant effect upon the dynamics, a fact that seems to contradict the conclusions of Ref. 17.

Interestingly, a completely different approach to the computation of U_{corr} has been known since the work of Fixman.[22] The bottom line of his method is to correct time averages, possibly biased by the constraints imposed on the dynamics. The resulting correcting potential, known as "Fixman potential" is given by

$$U_{\text{Fixman}} = \delta \log\left(\sum_{i=1}^{N} \omega_i\right) \qquad (8)$$

where δ is temperature dependent and can be shown to be inaccurate in certain particular (low dimensional) cases (see Ref. 19). It should be noted that, at first sight, both U_{corr} and U_{Fixman} are difficult to compute. This leads to the conclusion that, from a computational point of view, the approaches advocated both in Refs. 19 and 22 are not immediately useful, although illuminating when viewed as first approaches to be further simplified.

PROJECTION ONTO BASES OF NONHARMONIC MODES

Let us consider the decomposition

$$X_i(t) = X_i^0 + \sum_n c_n^<(t)\varphi_n^<(i) + \sum_n c_n^>(t)\varphi_n^>(i) \qquad (9)$$

This expression is *exact*, in the sense that the sets of vectors $\langle \varphi_n^<(i) \rangle$ and $\langle \varphi_n^>(i) \rangle$ form together a basis for the space of possible configurations of the protein. In the following, it will be assumed that the set $\langle \varphi_n^<(i) \rangle$ spans the "large-scale" motions of the macromolecule, whereas the set $\langle \varphi_n^>(i) \rangle$ describes the complementary motions. This classification can be done according to the value $\langle c_n^2 \rangle$, so that the vector $\langle \varphi_n(i) \rangle$ will be "small scale" if $\langle c_n^2 \rangle <$ *threshold*. This threshold is the first parameter of the present model.

The exact (in the sense of molecular dynamics) averaged evolution equations, projected into the set of large-scale vectors, are simply

$$\frac{d^2c_n^<}{dt^2} = -\sum_i \varphi_n^<(i)\text{grad}_{X_i} U_{\text{corr+slow}}(\{X_j\}) = -\text{grad}_{c_n^<} U_{\text{corr+slow}}(\{c_m^<\}, \{c_m^>\}) \qquad (10)$$

where

$$\text{grad}U_{\text{corr+slow}} \equiv \text{grad}(U_{\text{slow}}(\{X_j\}) + U_{\text{corr}}(\{X_j\})) + \lambda DU_{\text{fast}}$$

takes into account both the correcting potential and the Lagrange multiplier associated to the constraint $U_{fast} = 0$. Obviously, Eqs. (10) are not closed, because the small-scale coordinates appear also in the right-hand side of Eqs. (10), and we need to include the small-scale equations in order to get a solvable system of equations if no further modeling is made.

The simplest solution of the closure problem, where Eqs. (10) are replaced by

$$\frac{d^2c_n^<}{dt^2} = -\text{grad}_{c_n^<}U_{\text{corr+slow}}(\{c_m^<\}, \{c_m^> = 0\}) \quad (11)$$

should be expected to be *wrong*, unless some particular care is taken in the choice of the basis vectors. In fact, when writing Eq. (11), we are assuming that (h1) imposing $\langle c_m^> = 0 \rangle$ is equivalent to the constraint $U_{\text{fast}} = 0$. The (h1) can be rigorously imposed when working only with dihedral coordinates and fixing covalent distances and angles to their equilibrium values. This approach is being developed by some other groups.[23,24] It seems to be somewhat cumbersome and will not be pursued here. On the other hand, imposing this condition in an *approximate* way seems easier. For instance, using the set of hierarchical coordinates introduce in Ref. 17 probably amounts to this in an indirect way. Here, we will assume that (h1) is approximately true. We do not have any rigorous argument to support this besides the numerical results presented below. Our intuition is based on the fact that, for a large enough set of large-scale vectors, the difference $X_i(t) - \sum_n c_n^<(t)\varphi_n^<(i)$ will be small on the average, and that the constraint $U_{\text{fast}} = 0$ is approximately satisfied by a typical protein configuration. Thus, we also expect that $\text{grad}_{c_n^<}U_{\text{fast}}[\sum_n c_m^<(t)\varphi_m^<(i)]$ will be small. In the following we will also make the assumption that (h2) the projection of the correcting potential onto the set of large-scale vectors is small. This second assumption (h2) is harder to check directly, unless the value of the correcting potential is effectively computed. Here, we will assume that (h2) is true, and check the consequences of this. This leads us to consider the model equations:

$$\frac{d^2c_n^<}{dt^2} = -\text{grad}_{c_n^<}U_{\text{slow}}(\{c_m^<\}, \{c_m^> = 0\}) \quad (12)$$

Admittedly, this model is somewhat counterintuitive, since we are neglecting all the interactions involving bond lengths and covalent angles. The next section will discuss the consequences of Eq. (12).

NUMERICAL RESULTS FOR BPTI

In this section, we describe several numerical simulations performed on BPTI, using the model Eqs. (12). All the parameters of the force field are those of the program Amber 4.1.[25] The electrostatic energy has been computed with a distance-dependent dielectric function (IDIEL = 0 Amber option) and with no cutoff. First, we need an approximation of the nonharmonic vectors $\varphi_n(i)$. These have been obtained from a "standard" trajectory of 400 ps, where temperature is fixed to the value $T = 300$ K using Berendsen's algorithm.[26] The fluctuations $\langle c_n^2 \rangle$ decrease so fast to zero that the projections onto the first 50 modes capture, on the average, around 90% of the total fluctuations. In the following, all the simulations of the model Eqs. (12) will be done with this subset of large-scale vectors (simulations performed with 100 modes do not exhibit significant differences).

Initial conditions for the simulation of Eq. (12) are obtained by projection of a configuration of BPTI taken from the standard trajectory used to compute the anharmonic modes. Initially, the time step has to be less than 1 fs in order to thermalize the configuration of the protein in the large-scale subspace. This step is necessary because the projection affects the small-scale structure of the protein, and particularly the mere geometry of the residues. The thermalization is again done with Berendsen's algorithm[26] by imposing that the temperature is T_{slow}, the second parameter of the model. The production period of the dynamics is done at constant energy. In principle, one would say that $T_{\text{slow}} = N_{\text{slow}}T/3N_{\text{atoms}} \sim 6$ K, i.e., the kinetic energy per degree of freedom has to be preserved in the model trajectory. This is equivalent to saying that the equipartition theorem applies to the motion of the molecule, something that can only be approximately true (mainly because the force field is not linear). In the following, we use this value of T_{slow} as a rough estimate of what the temperature should be, and subsequently tune this parameter in order to recover the correct value of the fluctuations. It turns out that $T_{\text{slow}} \sim 20$ K gives reasonable results, as shown below.

The time step allowed by the model Eqs. (12) is obviously much larger than 1 fs. One way to obtain the maximum time step is to measure how the ratio between the total energy fluctuations and the kinetic energy fluctuations varies in simulations at constant energy. Empirically, it is found that the maximum time step corresponds roughly to a value of this ratio

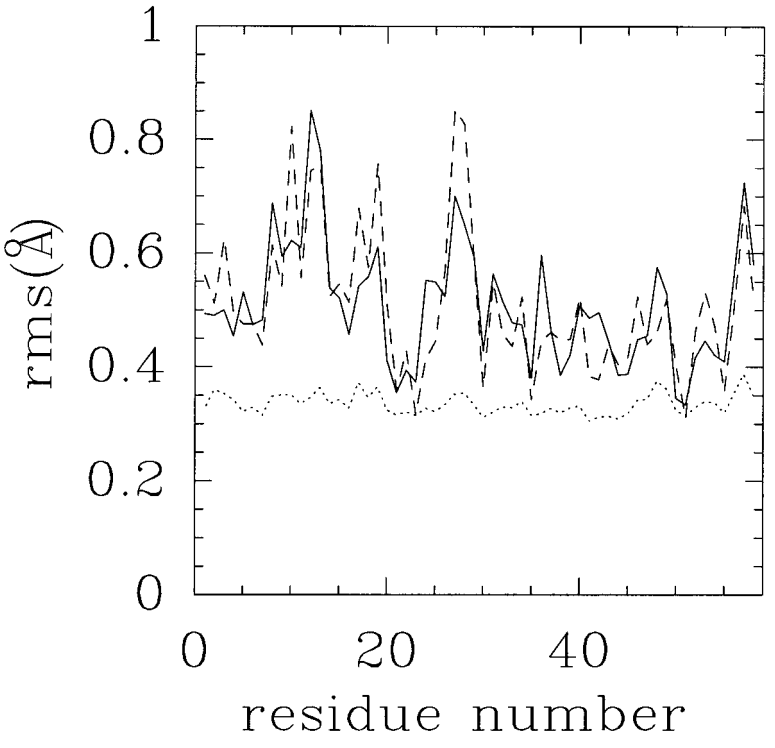

FIGURE 1 The rms of the position of the C^{α} atoms as a function of the residue number for the standard trajectory (continuous line) and for the trajectory projected onto 50 vectors (dashed line). The dotted line has been obtained by projecting onto 200 vectors the trajectory obtained with standard Newton equations.

of about 10%. This criterion yields a maximum value for the time step of 10 fs. Actually, it is found that the trajectory with a time step of 20 fs remains stable (i.e., the fluctuations in the total energy do not exceed 10% of the kinetic energy fluctuations) for about 100 ps, after which the protein suddenly "explodes." It is likely that algorithms more sophisticated than the Verlet algorithm, with, for instance, a control of the energy at each time step, would allow gains in central processing unit time of the order of 20, but this has not been tried yet.

Figure 1 represents the rms fluctuations of the position of the C^{α} atoms as computed from two trajectories in vacuum of BPTI, one using the standard Verlet algorithm and no projections (time step 1 fs), the other one using the same algorithm, with a projection onto the first 50 anharmonic modes (time step 10 fs). For comparison, we have also included the rms obtained when projecting the *nonaveraged* Newton equations, to emphasize the fact that averaging is a necessary ingredient if large-scale dynamics is sought. The agreement is fair, taking into account the approximations made. This could be interpreted by saying that these backbone fluctuations result from a collective motion well described by the projections made. Hayward and colleagues[27] reached the same conclusion, and even showed that the main fluctuation pattern of the backbone can be well described by just the projection onto the first anharmonic mode. It should be noted that in their paper, the fluctuations of the C^{α} atoms are about a factor two larger than ours. This is due to the fact that, starting from the configuration found in the Protein Data Bank (let's call it 1), BPTI stays for 100–200 ps around some configuration, then moving to a neighbor state (configuration 2). The factor 2 is found when considering portions of the trajectory where this transition is present. Our simulations show that for at least 1.2 ns, the protein stays in this last configuration, which we used as a starting point to our simulations of model Eqs. (12). It should be emphasized that starting from configuration 1, the projected equations do not yield the transition 1–2 (at least this has not been observed in our simulations). This is probably one of the (possible) weaknesses of the present approach, and will be studied more thoroughly in further works, in the case of proteins for which conformational changes are well documented and biologically relevant.

We next compare the energy characteristics of the two trajectories. The average and rms fluctuations

Table I Comparison of the Different Energy Terms (in kcal/mol) Computed with Standard Newton Equations (Time Step 1 fs) or with the Projected Equations (Time Step 10 fts)

Type of Energy	Average (Standard)	RMS (Standard)	Average (Projected)	RMS (Projected)
Dihedral	206	11	123	4
1-4 non bonded	134	4	114	8
1-4 electrostatic	681	8	673	6
Van der Waals	−261	13	−283	9
Electrostatic	−1761	41	−1763	13
H-bond	−22	4	−20	5

of the different energy terms, as computed by the AMBER 4.1 force field are listed in Table I. The long-range energy terms give comparable mean values, whereas the torsional contribution is wrong by a factor of almost 2, something that should be expected on a model of the large-scale dynamics. In order to further investigate this discrepancy, we have computed the histograms of the values taken by the dihedral angles of the backbone with a rms fluctuation greater than 30°. It turns out that those are the same for the projected and nonprojected trajectories, at least in the asymptotic regime (configuration) reached by the protein after 200 ps. The histograms, shown in Figure 2, also show a *qualitative* resemblance. However, when looking to the frequency of transitions between the two values taken by the first of these dihedral angles, it appears that transitions are much less frequent (at least a factor 10) in the projected trajectory (see, for instance, Figure 3) than in the nonprojected trajectory. The reason for this discrepancy is clear: projecting into large-scale modes selects those torsional motions that are *collective,* that is, imply large-scale motions. This is not the case for all the transitions shown in Figure 3, as those changes in the dihedral angle are correlated to simultaneous changes in neighboring angles, the total boiling down to local deformations. This point can be further checked by projecting the standard trajectory into the same subset of anharmonic eigenvalues, and computing the torsion angles in the projected structure. This operation actually filters all these local deformations.

In Figure 4 we compare the regions of phase space projected onto the two first eigenvectors explored by the projected and nonprojected trajectories. The agreement between the two should be expected from the closeness of the rms fluctuations of the position of the C^{α} atoms, and is in agreement with Hayward et al.'s conclusion that most of these fluctuations are related to the variations of the c_n coefficients for small values of n.

Finally, in Figure 5, we compare some correlation functions of the two first projection coefficients c_n. In Figure 5(a) [respectively 5(b)], the normalized correlation function

$$\frac{\langle (c_n(t) - \bar{c}_n)(c_n(t-\tau) - \bar{c}_n)\rangle}{\langle (c_n(t) - \bar{c}_n)^2\rangle}$$

is shown for $n = 1, 2$. Clearly, the decay of this function is comparable in both cases up to 1 ps, whereas the projected dynamics seems to more quickly lose the memory of its state (the correlation function decays quicker) for time delays greater than 1 ps. In Figures 5(c,d), we compare the normalized correlation functions of the time derivatives $\dot{c}_n$, $n = 1, 2$. This time, the overall decay seems comparable in both cases, although the decay (for time delays less than 1 ps) of the correlation function of $\dot{c}_2$ is slower in the projected dynamics. The mean squared displacement $\mathrm{msd}_n(t) = \langle [c_n(t) - c_n(0)]^2 \rangle$ [represented in Figs. 5(e,f)] quantifies the way each of the coordinates explores the phase space. A power law behavior $\mathrm{msd}_n(t) \sim t^{2H}$ as $t \to \infty$ of this quantity is typical of random walks ($H = 0.5$ for a pure brownian motion). Figures 5(e,f) show that the short time behavior ($t < 50$ ps) is quite different for the standard and projected trajectories, whereas it reaches comparable values for time differences of the order of 100 ps. It should be mentioned that the asymptotic value of the plateau reached by the msd is slightly higher for the standard trajectory. This is probably an indication that the projected trajectory explores in a less efficient way the phase space of the protein than the standard trajectory. The eventual power law character of these plots[35] is difficult to decide from our statistics (the maximum trajectory length is 1 ns). In this sense, it is interesting to compare the behavior observed by Garcia et al. on the crambin (Fig. 7 of Ref. 35), where the fluctuations on time scales on the

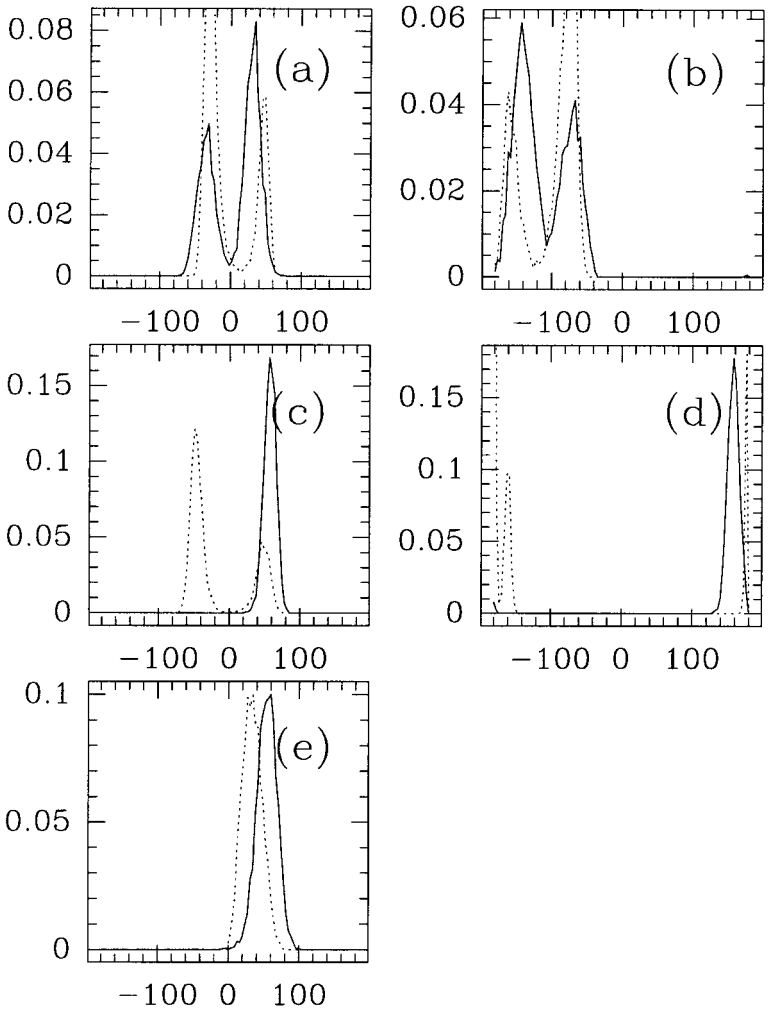

FIGURE 2 Histogram of the values taken, respectively, by the dihedrals: (a) ψ of Cys 5, (b) ϕ of Leu 6, (c) ϕ of Thr 11, (d) ϕ of Lys 26, and (e) ϕ of Ala 27. Continuous line: standard trajectory (400 ps); dashed line: projected trajectory (2 ns).

order of nanoseconds are clearly exhibited. To see whether this difference is due to the presence of water molecules on the simulation in Ref. 35 or not will require additional investigation.

CONCLUSIONS

In this paper, we have presented preliminary results on the application of a computing scheme for the simulation of the large-scale dynamics of a biomolecule. The method is based on appropriate projections of the equations of motion, averaged in time, onto some set of anharmonic modes. The results presented show that the method is stable for time steps one order of magnitude larger than those used in the standard approach, and that it yields correct results for quantities related to large-scale motions, such as electrostatic energy or torsional motions of the backbone. The main disadvantage of the method is the need to compute the anharmonic modes. According to Ref. 4, fairly short trajectories already provide a good approximation of these vectors, but further investigations are likely to be required along this direction, particularly because, as noted in Ref. 14, for large proteins (say above 300 residues), the characteristic time of the slowest motions becomes slower. The use of alternative representations such as the hierarchical set of coordinates introduced in Ref. 17 or the approximate large-scale harmonic modes of Ref. 28 offer probably a natural way to extend the considerations of the present work. Also, further tests are needed to assert the possibility of the so-generated trajectories to effectively describe experimentally observed confor-

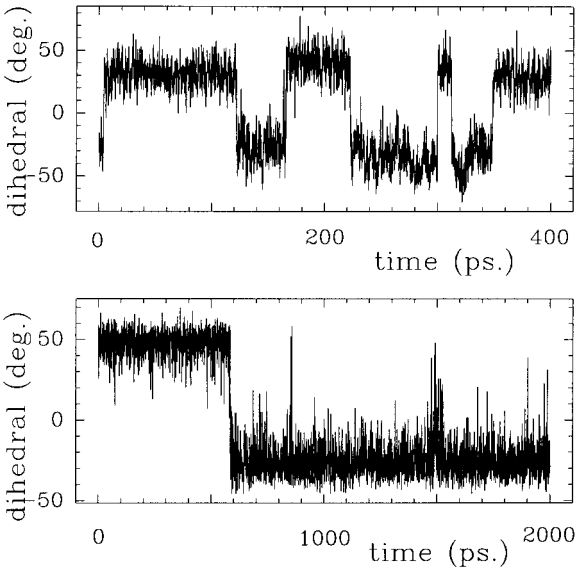

FIGURE 3 Time series of the dihedral (ψ of Cys 5) for the standard trajectory (a) and the projected trajectory (b). Note that the length of the (b) trajectory is five times longer than that of (a).

mational changes, such as the hinge bending motion of cytrate synthase.[29] This is a challenging task, since those motions are expected to involve time scales well beyond the nanosecond and will probably require the interfacing with numerical integration schemes more elaborated than the Verlet scheme used here.

Note that the scheme introduced in this work and tested on a small protein in vacuo can be extended to the study of large biomolecules in solution in a rather straightforward way. First, because within the last few years, it has become possible to compute a few modes of proteins as large as the aspartate transcarbamylase (2757 amino acids[30]). Second, because it is already possible to solve the equations of solvent motions with large time steps, either by using water models with "heavy" hydrogens, that is, with 10 amu

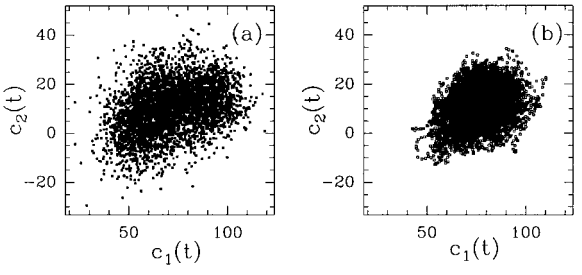

FIGURE 4 Projections onto the subspace spanned by the first two eigenvectors of the (a) projected trajectory and (b) the standard trajectory.

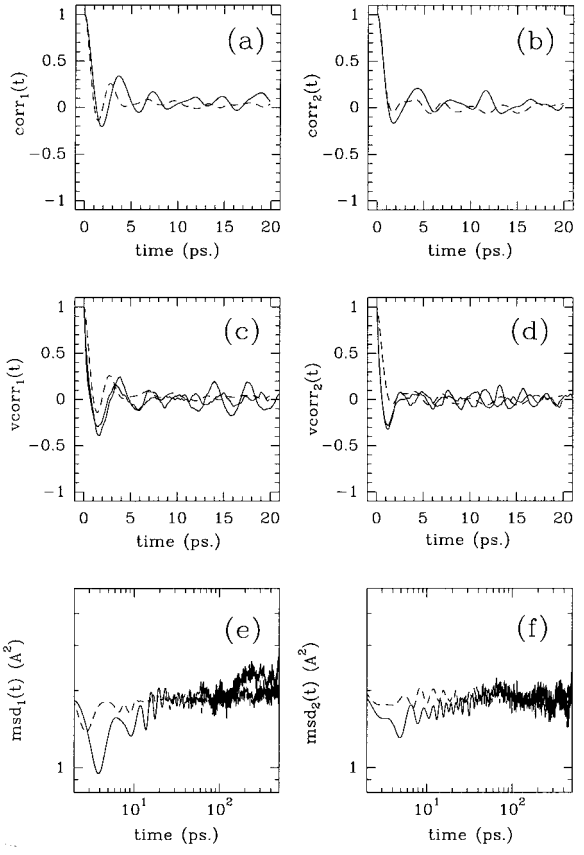

FIGURE 5 Correlation functions for the standard (solid line) and projected trajectories (dashed line): (a,b) The normalized correlation functions for, respectively, $c_1(t)$ and $c_2(t)$. (c,d) The normalized correlation functions of the time derivatives $\dot{c}_1(t)$ and $\dot{c}_2(t)$. (e,f) The msd function for, respectively, $c_1(t)$ and $c_2(t)$.

masses,[31] or by using one of the multiple time step algorithms recently developed, like those proposed by Berne and his co-workers (see, for instance, Ref. 32). What is yet not clear is whether the first hydration shell of the protein, which is highly organized (see, for instance, Ref. 33), will have to be treated like the rest of the solvent, or like a part of the protein. A preliminary study of the standard trajectory of a protein in a water solution, needed to compute the anharmonic modes, will help to clarify this point. A recent work on the low-frequency normal modes of a protein, namely, Ras-p21, embedded in a water shell,[34] suggests that the water molecules of the first hydration shell of the protein could be well described as a part of the protein, from the point of view of the large-scale dynamics of the system. Note that only a moderate number of water molecules seems to be critically needed (350 in the case of myoglobin), as shown in a study of series of molecular dynamics simulations performed at various hydration levels.[36]

The authors acknowledge computer resources from Pôle MNI (U. Bordeaux I), and from IDRIS (under contract no. 980871). One of us (YHS) thanks Martin Karplus for fruitful discussions during work on his Ph.D. thesis, and D. Perahia who initiated and directed his work at that time.[15]

REFERENCES

1. Horiuchi, T. & Go, N. (1990) *Proteins* **10,** 106–116.
2. Kitao, A., et al. (1991) *J. Chem. Phys.* **158,** 447–472.
3. Garcia, A. E. (1992) *Phys. Rev. Lett.* **68,** 2696–2699.
4. Amadei, A., et al. (1993) *Proteins* **17,** 412–425.
5. van Aalten, D. M. F., et al. (1995) *Proteins* **22,** 45–54.
6. Hayward, S. & Go, N. (1995) *Ann. Rev. Phys. Chem.* **46,** 223–250.
7. Brooks, B. & Karplus, M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4995–4999.
8. Marques, O. & Sanejouand, Y.-H. (1995) *Proteins* **33,** 557–560.
9. Karplus, M. & Kushick, J. N. (1981) *Macromolecules* **14,** 325–332.
10. Perahia, D., et al. (1990) *Biopolymers* **29,** 645–677.
11. Ichiye, T. & Karplus, M. (1991) *Proteins* **11,** 205–217.
12. Lumley, J. L. (1971) *Stochastic Tools in Turbulence,* Academic Press, New York.
13. Teeter, M. M. & Case, D. A. (1990) *J. Phys. Chem.* **94,** 8091–8097.
14. Balsera, M. A., Wriggers, W., Oono, Y. & Schulten, K. (1996) *J. Phys. Chem.* **100,** 2567–2572.
15. Sanejouand, Y.-H. (1990) Ph.D. thesis, Orsay, France.
16. Askar, A., Space, C. & Rabitz, H. (1995) *J. Phys. Chem.* **99,** 7330–7338.
17. Durup, J. (1991) *J. Phys. Chem.* **95,** 1817–1829.
18. Reich, S. (1995) *Phys. D* **89,** 28–42.
19. Schütte, C. & Bornemann, F. A. (1997) *Phys. D* **102,** 57–77.
20. Takens, F. (1980) in *Global Theory of Dynamical Systems, Evanston 1979,* Nitecki, Z. & Robinson, C., Eds., Springer-Verlag, Berlin.
21. van Gunsteren, W. F. & Karplus, M. (1982) *Macromolecules* **15,** 1528–1544.
22. Fixman, M. (1978) *J. Chem. Phys.* **69,** 1527, 1538.
23. Noguti, T. & Go, N. (1985) *Biopolymers* **249,** 527–546.
24. Mazur, A. K., Dorofeev, V. E. & Abagyan, R. A. (1991) *J. Comp. Phys.* **92,** 261–272.
25. Perlman, D. A., et al. (1995) AMBER 4.1, University of California, San Francisco.
26. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984) *J. Chem. Phys.* **81,** 3684–3690.
27. Hayward, S., Kitao, A. & Go, N. (1994) *Protein Sci.* **3,** 936–943.
28. Tirion, M. M. (1996) *Phys. Rev. Lett.* **77,** 1905–1908.
29. Wiegand, G. & Remington, S. J. (1987) *Ann. Rev. Biophys. Biophys. Chem.* **15,** 97.
30. Thomas, A., Field, M. J., Mouawad, L. & Perahia, D. (1996) *J. Mol. Biol.* **257,** 1070–1087.
31. Pomes, R. & McCammon, J. A. (1990) *Chem. Phys. Lett.* **166,** 425–428.
32. Tuckerman, M. E. & Berne, B. J. (1991) *J. Chem. Phys.* **95,** 8362–8364.
33. Alary, F., Durup, J. & Sanejouand, Y. H. (1993) *J. Phys. Chem.* **97,** 13864–13876.
34. Ma, J. & Karplus, M. *J. Mol. Biol.*, to appear.
35. García, A. E., Blumenfeld, R., Hummer, G. & Krumhansl, J. A. (1997) *Phys. D* **107,** 255–239.
36. Steinbach, P. J. & Brooks, B. R. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9135–9139.