Theor Chem Acc (2000)
DOI 10.1007/s002140000215

Theoretical
Chemistry Accounts

Regular article

Coupling overall rotations with modal dynamics

J. Elezgaray, G. Marcou, Y. H. Sanejouand

Centre de Recherche Paul Pascal, Avenue Schweitzer, 33600 Pessac, France

Received: 3 July 2000 / Accepted: 15 September 2000 / Published online: 21 December 2000

Abstract. We consider a model for protein dynamics in which only certain collective, global motions are allowed. These directions are given by the slowest harmonics modes, as given in the reference frame of the protein. Furthermore, the latter is allowed to rotate and translate in response to interactions with other molecules. The model is obtained by projecting the (averaged) Newton equations onto this set of harmonic modes. We show that the subsequent homogenization of the time scales allows time steps one order of magnitude larger than the standard ones. This homogenization is also shown to be a necessary ingredient in order to get meaningful statistics of the trajectory.

Key words: Normal mode – Large time step – Molecular dynamics – Protein

1 Introduction

Many functional proteins are characterized by the fact that their biological function requires some internal motion of collective character. Typical examples include the binding of a substrate or, in a more general way, the transmission of conformational changes to a distant site, as in allosteric mechanisms. For instance, in the case of citrate synthase, the binding of coenzyme A induces an 18° rotation of the small domain around an axis located near residue 274, resulting in the so-called hinge-bending motion [1–3]. The latter induces the closure of the cleft between the two domains of the protein and, as a direct consequence, the closure of the substrate binding site.

An interesting question is whether these collective motions arise from the coupling with an exterior agent (ligand) or if they are already present as an intrinsic motion of the protein. This question has been settled in Ref. [4] for the case of triosephosphate isomerase where it has been shown (experimentally) that loop closure is "not ligand gated but is a natural motion of the protein". Similar results have been obtained with lysozyme [5].

From these experimental results, it seems logical to attempt a description of protein dynamics in terms of collective degrees of freedom, including only those directly related to the protein function. Such an idea underlies several molecular dynamics methods already discussed in the literature and which we now briefly mention.

In terms of the a posteriori analysis of protein trajectories generated by molecular dynamics simulations, two related methods have been extensively used. One is based on normal-mode analysis [6, 7], namely, the fluctuations of the atomic positions $\delta X_i(t)$ are described as a superposition of normal modes:

$$\delta X_i(t) = \frac{1}{\sqrt{m_i}} \sum_{n=1}^{3N_a} c_n(t)\varphi_n(i) \quad , \tag{1}$$

where N_a is the number of atoms and

$$\delta X_i(t) = X_i(t) - X_i^0$$

is the position fluctuation of the ith atom, X_i^0 is a reference conformation of the protein (the average over some trajectory or a minimum energy conformation), and $\varphi_n(i)$ is the amplitude of the nth harmonic mode at the ith atom. This description has recently been shown to give an accurate picture of X-ray diffusive scattering of protein crystals [8]. Moreover, it has been shown that, in several cases, for example, hexokinase [9], lysozyme [7], citrate synthase [10], etc., the conformational change upon ligand binding observed by crystallography is strongly correlated with (one of) the slowest harmonic modes. All these facts give some support to the idea that low-frequency normal modes contain the necessary information to accurately describe collective motions. It should also be mentioned that the slowest normal modes can be efficiently computed using extremely simplified force fields [11]. Other methods inspired from the work in Ref. [12] have also shown that the approximate computation of low-frequency harmonic modes can be done quite cheaply.

Correspondence to: J. Elezgaray

Contribution to the Symposium Proceedings of Computational Biophysics 2000

The other approach is based on the determination of the so-called essential dynamics [13]. Formally, the equality [1] is used again, but this time, the vectors $\varphi_n(i)$ are the eigenvectors of the covariance matrix $\langle \delta X_i(t) \delta X_j(t) \rangle$ ($\langle , \rangle$ means time average). In the following, these eigenvectors are noted $\Psi_n(i)$ and are termed anharmonic modes. For a stationary process, the $\Psi_n(i)$ vectors provide an optimal decomposition of the fluctuations $\delta X_i(t)$. In the present context, this means that the description provided by the anharmonic modes is, by construction, more accurate than that provided by the harmonic modes. However, it should not be concluded from this that harmonic modes are uninteresting. In fact, anharmonic modes have at least two main drawbacks: the covariance matrix needs some potentially long trajectory to be computed [14] and so far there is no approximate method to compute the $\Psi_n(i)$ vectors. In spite of this, the anharmonic decomposition has been extensively used [13, 15–18] in the a posteriori description of molecular dynamics trajectories, as well as data coming from crystallographic structures. In particular, further analysis of the correlation matrix has revealed the existence of different types of anharmonic modes, describing a rich hierarchy in the protein motion, strongly correlated to the roughness of the energy landscape [19]. The method has also been used as a basis for a phase-space exploring algorithm [20].

Here, we investigate a different question, namely how to obtain protein trajectories on the basis of the hypothesis that collective coordinates provide an accurate, although approximate, description of protein fluctuations. In other words, we do not want to analyze a trajectory generated through standard methods; our goal is to generate the protein fluctuations that can be related to the collective motions. More precisely, we will assume Eq. (1) (or its generalization, Eq. 2) and derive evolution equations involving only the collective motions, identified here with the low-frequency normal modes $\varphi_n(i)$. Then, our goal is to derive evolution equations for the amplitudes $c_n(t)$, $n = 1, \ldots N'$, in such a way that the reconstructed protein trajectory is 'similar' to the non-projected one. The parameter N' is the number of retained harmonic modes (it corresponds to a frequency cutoff) and is a free parameter of the model. It has been shown [13, 15–18] that for values $N' \sim 10^{-2} N_a$, the relative error made when truncating the sum in Eq. (1) to N' terms is of the order of 10^{-2}.

In the following we compare three types of trajectories: reference trajectories (RT), obtained from standard molecular dynamics simulations, projected RT (PRTN'), obtained from a RT trajectory by a projection onto the set $\{\varphi_n(i), n = 1, \ldots, N'\}$, and finally a projected trajectory (PTN'), obtained from the model equations discussed later. The main motivation of our approach is the intuition that, when considering only the large amplitude (slow) motions of the protein, large time steps can be used. This simple idea has already been considered by several authors. Perhaps the closest to our approach is the so called MBO(N)D method [21], in which the molecular system is substructured by collecting groups of atoms into rigid or flexible bodies. The flexibility is accounted for by an assumption generalizing Eq. (1), namely:

$$X_i(t) = T(t) + R(t) \left[X_i^0 + \frac{1}{\sqrt{m_i}} \sum_{n=1}^{N'} c_n(t) \varphi_n(i) \right] , \qquad (2)$$

where $R(t)$ [or $T(t)$] is a rotation matrix (or a translation vector), both being time-dependent. In Eq. (2), the relation between Cartesian and harmonic coordinates is assumed to hold for each of the elastic bodies of the substructuring. Here, we consider Eq. (2) applied to the whole protein. In this sense, we are considering here the limit of roughest substructuring in MBO(N)D.

It is shown later that the MBO(N)D approach requires some corrections in order to account for the protein fluctuations as measured in RT. The reason is that the MBO(N)D method was derived from methods developed in the context of dynamics simulations of large complex mechanical structures in the aerospace industry. In this case, only elastic forces are involved, and the internal dynamics of each body is characterized by rather homogenous time scales. This is not the case for proteins, where bonded and nonbonded interactions are characterized by widely different time scales and some specific treatement (time homogeneization [22, 23]) is needed. This effect is not quite noticeable when considering "small" systems, such as those considered in Ref. [24], where a particular case of the MBO(N)D method was also studied, nor in the hierarchical approach of Durup [25], in which some correcting terms cancel by construction, as explained later. To the best of our knowledge, the first place where Eq. (2) was used as a basis for a dynamical model of macromolecules is Ref. [26].

The rest of the article is organized as follows. In Sect. 2, we explain how time homogeneization [22, 23] leads to evolution equations for which the protein fluctuations and the volume explored in phase space are similar for the PTs and non PTs. We do this in the simpler case of Eq. (1) and consider the more general case in Sect. 3, where a novel algorithm for the coupling of the internal degrees of freedom and the rotational and translational degrees of freedom is given.

2 Evolution equations without overall rotations translations

In this section, we study the particular case of a single protein in vacuum, neglecting the overall translations and rotations. We start from Eq. (1) and substitute this approximation in the Newton equations. Using the orthogonality of the harmonic modes, $\sum_i \varphi_n(i) \varphi_m(i) = \delta_{n,m}$, the evolution equations for the amplitudes c_n read:

$$\begin{aligned} \frac{\mathrm{d}^2 c_n}{\mathrm{d}t^2} &= -\sum_{i=1}^{N_a} \varphi_n(i) \sqrt{m_i} \frac{\partial E}{\partial X_i} \left(X_i^0 + \frac{1}{\sqrt{m_i}} \sum_n c_n \varphi_n(i) \right) \\ &= -\frac{\partial E}{\partial c_n} \left(X_i^0 + \frac{1}{\sqrt{m_i}} \sum_m c_m \varphi_m(i) \right) , \end{aligned} \qquad (3)$$

where E stands for the energy of the system.

In order to check how well the model (Eq. 3) works, let us first consider the case of decalanine, a small peptide that forms a well defined α helix. The N and C terminal ends contain CH_3CO and $NHCH_3$ groups, respectively in order to neutralize the charges normally present at the ends of this peptide at pH 7. We use a united-atom representation for the CH and CH_3 groups, the resulting model containing 66 "atoms". All the electrostatic and van der Waals forces are taken into account, the dielectric constant is set to 1, and all the bonded interactions are as defined in the param19 force field in CHARMM [27].

The root-mean-square (rms) fluctuations of the C^{α} atoms as a function of the highest frequency included in the projection are shown in Fig. 1. Trajectories of 10 ps are obtained using either the Newton equations, then projecting the trajectory (PRT, open squares) or the projected Newton equations (Eq. 3) (PT, filled squares). The data collected from the PRT show that the approximation given by the slowest normal modes is extremely efficient: for a cutoff frequency of 200 cm^{-1} (corresponding to about 50 normal modes), the fluctuations in the PRT amount to 95% of those in the RT; however, the PT displays less than 50% of these fluctuations. Actually, as shown in Fig. 1, increasing the dimension of the projected subspace does not really help. One can conclude that the model equations (Eq. 3) are not quite satisfactory. Note that, of course, for very low temperatures, where the protein is mostly harmonic, model Eq. (3) provides the correct fluctuations.

Roughly speaking, although the approximation of Eq. (1) is reasonable, the additional approximation made in Eq. (3), namely,

$$E(X_i) \sim E\left(X_i^0 + \frac{1}{\sqrt{m_i}} \sum_{n=1}^{N'} c_n \varphi_n(i)\right) \ ,$$

is not justified, mainly because E presents strong gradients in the directions corresponding to the stiffest directions that control bond lengths and angles. As soon as small deviations from the equilibrium values of these internal coordinates occur, as is the case when using the truncated projection involved in Eq. (1), very large restoring forces take place that freeze the protein motion, thus explaining the behavior observed in Fig. 1. This intuitive argument also explains why the projected dynamics "works" for small clusters held by van der Waals forces [24] or in situations where the vectors $\varphi_n(i)$ are such that the projections of the steepest forces are mostly canceled [25].

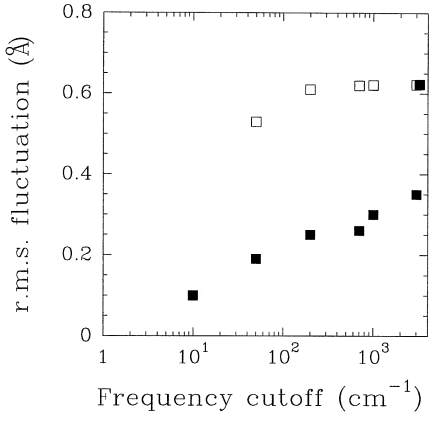

Fig. 1. Spatial fluctuations of decalanine as a function of the frequency cutoff used either to generate the projected trajectory (*PT*) (*filled squares*) using Eq. (3) or to obtain the projected reference trajectory (*PRT*) (*open squares*)

As stated in the Introduction, one way out is to "homogenize" the interactions. One possible method [23] to do this is to consider the limit in which the bonded interactions are infinitely stiff. This amounts to restraining the dynamics to the dihedral space, freeezing the bonds and angles to their equilibrium values. The resulting evolution equations can be written in several ways [23, 28]. We adopt the approach advocated in Ref. [23] and consider the evolution equations:

$$m_i \frac{d^2 X_i}{dt^2} = -\frac{\partial E_{\text{slow}}}{\partial X_i} - \lambda \frac{\partial E_{\text{fast}}}{\partial X_i} - \frac{\partial E_{\text{Fixman}}}{\partial X_i} \ . \tag{4}$$

Here, E_{slow} stands for all the energy terms minus the bond and angle contributions, E_{fast} gathers these two fast terms, and E_{Fixman} is a correcting term arising in the limiting procedure, first pointed out by Fixman [29] in the context of polymer dynamics. The factor λ is the Lagrange multiplier ensuring the constraint $E_{\text{fast}} = 0$. The mathematical procedure to obtain Eq. (4) amounts to an averaging (in time) of the fastest oscillations. The Fixman term is a consequence of the fact that the average of products of terms of zero mean is not zero. It is therefore related exclusively to the freezing of the angles [23] to their equilibrium values.

In order to get the smoothed evolution equations for the amplitudes c_n, we take the scalar products of Eq. (4) with the normal modes φ_n. In previous work [30], we ignored the contribution coming from the Fixman potential, uniquely on the basis of results obtained with very small molecules. In those cases, it can be shown [23] that this contribution is indeed small. Concerning the $\lambda \nabla E_{\text{fast}}$ term, we have shown in Ref. [31] that, when the number of harmonic modes increases, this term becomes essential in order to maintain the compactness of the protein. Notice also that λ depends on the instantaneous conformation of the molecule in a rather involved way. Instead of computing at each time step its precise value, we will make it constant. The equations that we propose as a model for the time evolution of the amplitudes are, therefore,

$$\frac{d^2 c_n}{dt^2} = -\frac{\partial}{\partial c_n}(E_{\text{slow}} + \lambda E_{\text{fast}}) \left[X_i + \frac{1}{\sqrt{m_i}} \sum_m c_m \varphi_m(i)\right] \ . \tag{5}$$

In other terms, this model amounts to a simple rescaling of the bonded terms in E_{fast}, by the factor λ. In Ref. [30], we studied the particular case $\lambda = 0$. In fact, it is easy to see that $\lambda \ll 1$. Following the notations in Ref. [32] and assuming for the sake of simplicity that only bond length constraints are sought (k being a typical stifness constant)

$$\lambda \sim \frac{1}{k}\frac{\sum_i \frac{1}{m_i}\,\delta X_i \nabla E_{\rm soft}}{\sum_i \frac{1}{m_i}\delta X_i^2} \ll 1 \ .$$

In order to illustrate the results obtained with the model Eq. (5), we generate 200-ps equilibrated trajectories of a small protein, the carboxy terminal fragment (CTF) of the L7/L12 ribosomal protein of *E. coli*. The simulations are done in a vacuum, with a distance-dependent dielectric constant and bonded interactions as defined in the param22 force field in CHARMM. This protein is particularly interesting in the present context, because it has been shown previously [33, 34] that it displays well-defined slow, collective motions that can be quantified through the measurement of the angle between two of the three α helices present in this protein. In Fig. 2, we show the rms fluctuations of the C^α carbons, as a function of the residue number and for three different values $N' = 100,\ 150,\ 200$ of the number of retained harmonic modes. For $N' = 100,\ 150$, the trajectories PTN' were generated with a time step of 10 fs and $\lambda = 0.01$ (actually, $\lambda = 0$ would also work in these cases). For PT200, we used a time step of 5 fs, in order to ensure the stability of the trajectory. For the sake of comparison, we have also included the rms fluctuations corresponding to a trajectory with $N' = 200$ and $\lambda = 1$. It is clear from this figure that the position fluctuations of the trajectories PTN' generated by the model Eq. (5) converge much faster to the real ones than those generated by Eq. (3); however, it is also clear that some additional fluctuations are missing in the model. For instance, in Fig. 3 it is shown that the fluctuations captured by our model are lower than the ones actually taking place in the part of the phase space spanned by the first 100 harmonic modes. The situation improves for proteins interacting either with the solvent [31] or with other proteins, as shown in the next section. We address the reader to Ref. [31] for further details concerning the properties of the projected trajectories of CTF, in particular the measurement of the characteristic angle between the two helices mentioned earlier.

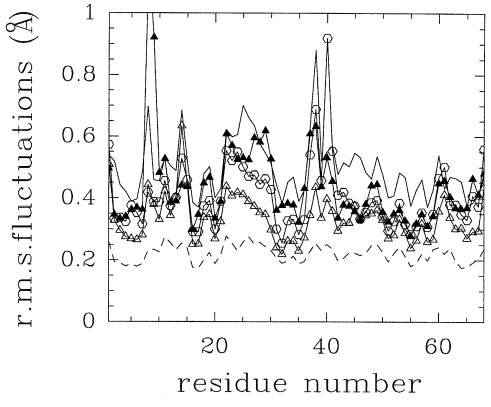

Fig. 2. The root mean square (*rms*) of the position of the C^α atoms as a function of the residue number for the reference trajectory (*RT*) (*continuous line*), PT100 (*open triangles*), PT150 (*open hexagons*), PT200 (*filled triangles*), and PRT200 (*dashed line*)

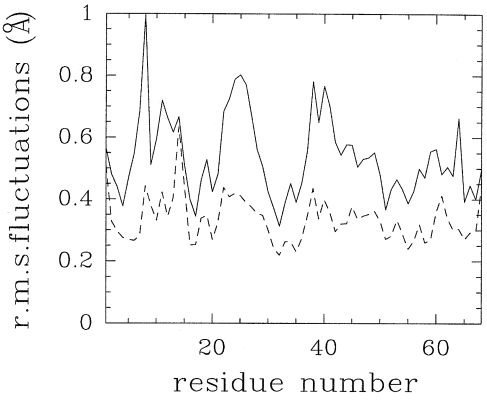

Fig. 3. The rms of the position of the C^α atoms as a function of the residue number for the RT projected onto 100 modes (PRT100) (*continuous line*) and for PT100 (*dashed line*)

3 Including overall rotations and translations

We now turn to the more general case of Eq. (2) and derive evolution equations for both c_n, R, and T. The equations for the overall translations are quite straightforward and are the same as in the case of a rigid body, namely:

$$\sum_i m_i \ddot{T} = \sum_i F_i \ . \tag{6}$$

The equations for c_n and R are derived in the Appendix. Here, the main point is that there are fewer unknowns ($\dot{\Omega}$ and $\ddot{c}_n$) than the number of equations ($3N$) to be satisfied; therefore, the solution is not uniquely defined. We define the problem in the least-squares sense and show in the Appendix that it can be solved analytically.

The general structure of the evolution equations for the overall rotations, $R(t)$, and mode amplitudes, $c_n(t)$, (Eqs. A3, A4 in the Appendix) is of the form.

$$\ddot{a} = f(a, \dot{a}) \ . \tag{7}$$

Because of the presence of $\dot{a}$ in the right-hand side of this equation, the Verlet algorithm cannot be applied and the Lobatto algorithm is needed instead:

$$\dot{a}_{n+1/2} = \dot{a}_n + \frac{\mathrm{d}t}{2} f(a_n, \dot{a}_{n+1/2}),$$

$$a_{n+1} = a_n + \mathrm{d}t\ \dot{a}_{n+1/2},$$

$$\dot{a}_{n+1} = \dot{a}_{n+1/2} + \frac{\mathrm{d}t}{2} f(a_{n+1}, \dot{a}_{n+1/2}) \ .$$

In the present context, the variable $a(t)$ stands for the set of unknowns $R(t)$, $\{c_n(t), n = 1, \ldots N'\}$. We have found it useful to use the quaternion representation of the rotation matrix,

$$R = \begin{pmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2) \\ 2(q_1q_2+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_0q_1) \\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{pmatrix}, \tag{8}$$

with the constraint that $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The angular velocity, Ω, and its time derivative can also be

expressed as combinations of the time derivatives $\dot{q}_i$ and $\ddot{q}_i$ as explained in Ref. [35]. The resulting equations are then exactly of the form of Eq. (7). The first of the three steps of the Lobatto algorithm is an implicit equation, easily solved by iteration. It should be noted, however, that this iteration must be done in the following way (l is here the iteration index, the time index has not been added in the sake of clarity)

$$\mathbf{I}\dot{\Omega}_{l+1} = -\sum_i \sqrt{m_i}(Z_i \wedge x_i) + \sum_i (\sqrt{m_i} \sum_n \ddot{c}_{n,l} R\varphi_n(i) \wedge x_i),$$

$$\ddot{c}_{n,l+1} = \sum_i (Z_i \cdot R\varphi_n(i)) - \dot{\Omega}_{l+1} \cdot \tilde{\varphi}_n \ ,$$

with initial conditions $\ddot{c}_{n,l=0} = 0$, otherwise the matrix that has to be inverted becomes quickly singular as the number of harmonic modes increases.

In order to check the model Eqs. (A3) and (A4), we generate molecular dynamics trajectories of the melittin tretamer in vacuum. Melittin is the major component of the venom of the honey bee that is responsible for lysis of the cell membrane [36, 37]. This amphiphilic polipeptide contains 26 residues and adopts an α-helical structure in very different environments, from detergent micelles [38] and nonpolar solvents [39] to concentrated aqueous salt solutions [40]. However, the tretamer association is only found in aqueous salt solutions. Each α helix is actually separated into two segments by a proline residue at position 14. Segment 1–13 is hydrophobic, whereas the segment from Ala^{15} to Gln^{26} is amphiphilic. This arrangement is at the origin of the tetramer formation through hydrophobic contacts.

The initial configuration of the tretamer is such as given in the Protein Databank (code 1MLT). The simulations are done in vacuum, with a distance-dependent dielectric constant, $\epsilon = 80$, and the CHARMM-22 force field for bonded interactions. The equilibrated RT is 500-ps long. In generating the PT, we encountered similar difficulties as in Ref. [30], namely, the starting configuration, when projected, generates a great number of high-energy steric contacts that have to be removed through a sequence of extremely small time steps. Of course, increasing the number of modes included in the description fixes this problem partially, however this imposes severe conditions on the time step that could be used subsequently. In all the following, we used 44 harmonic modes for each of the four helices, resulting in an overall counting of 200 degrees of freedom. The time step used is 10 fs (the resulting speed-up is around 8), the average rms of the total energy being less than 0.1 kcal/mol for the 500-ps trajectory. Time steps of 20 fs could also be used, resulting in a speed-up of 17 and a total energy rms less than 0.2 kcal/mol. The position fluctuations of the C^α carbons along the backbone of the four α helices are represented in Fig. 4, with the removal for each helix of the overall translation–rotation motions. The overall agreement seems quite satisfactory: the PTs capture the pronounced mobility of the peptide ends (residues 1 and 26 for each helix), but are, on average, somewhat more rigid around the Pro^{14} residue. We have also represented the observed fluctuations in PT44 with no rescaling ($\lambda = 1$) of the bonded interactions: it is clear that the rescaling considerably improves the similarity

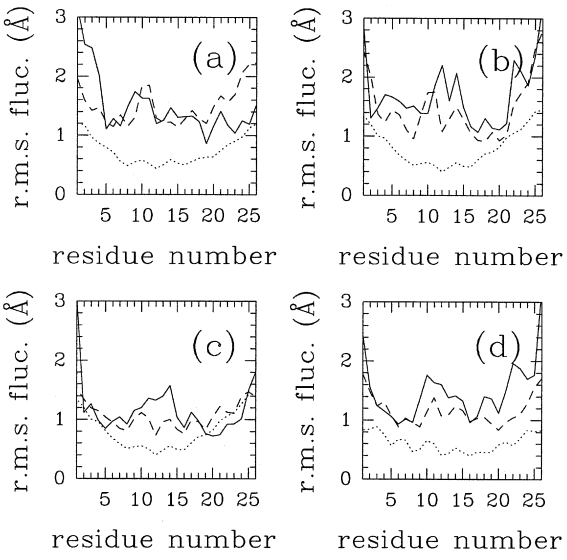

Fig. 4a–d. The rms position fluctuations of the melittin tetramer C^α carbons. **a–d** correspond to helices 1–4: RT (*continuous line*), PT44 (*dashed line*), PRT55 (*short-dashed line*)

between the PTs and non-PTs. Further insight is gained by measuring the angle between segments 1–13 and 15–26. This is done using the command HELIX of CHARMM and removing the four terminal residues from each end. As can be seen in Fig. 5, the projection onto 44 modes preserves well the average behavior of this angle, measured on the first helix; however, the fluctuations are less pronounced than in the RT.

4 Conclusions

In this article we have discussed a simple algorithm to integrate the evolution equations of projected versions of protein structures with time steps of about 10 fs. Projecting provides a smoothed force field in which some rescaling of the bonded interactions has to be taken into account. This idea is quite similar to that used in the simulation of large polyelectrolytes with blocking techniques [41]. We have shown that such rescaling considerably improves the convergence of (the characteristics of) the PTs to the non-PTs. The same idea has been extended to the more general case where several flexible molecules interact. In this case, the coupling with possible overall rotations of each molecules has to be taken into account.

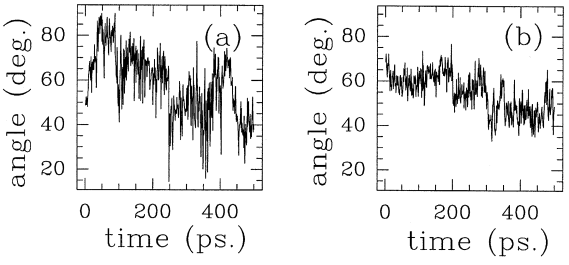

Fig. 5. Angle between the 1–13 and 15–26 segments of helix 1: **a** RT, **b** PT44

Although the results are quite satisfactory, in particular regarding the energy conservation properties, there is clearly room for improvement, in particular concerning the flexibility of the projected configurations. We are currently exploring the addition of a fluctuating forcing and a frictionlike term, coupled by the fluctuation–dissipation theorem. The presence of both terms is quite clear from theoretical considerations, such as those developed in Ref. [42]. Such an extended model boils down to generalized Langevin equations in the space of the projected components. The model considered here neglects completely such terms. It should also be noted that these additional terms can, in principle [43], be computed from short nonprojected simulations.

Acknowledgements. The authors acknowledge computer resources from Pôle M3PEC (U. Bordeaux I) and from IDRIS (under contract no. 990871).

Appendix

Here, we derive the evolution equations for the amplitudes, $c_n(t)$, rotation, $R(t)$, and translation, $T(t)$. We use throughout the notation $\dot{a} \equiv \mathrm{d}a/\mathrm{d}t$. We start from Eq. (2) and derive twice with respect to time, obtaining

$$\begin{aligned}\Gamma_i - \Gamma_\mathrm{g} = {} & \dot{\Omega} \wedge x_i + \Omega \wedge (\Omega \wedge x_i) \\ & + 2\Omega \wedge R\left(\frac{1}{\sqrt{m_i}} \sum_n d_n \varphi_n(i)\right) \\ & + R \frac{1}{\sqrt{m_i}} \sum_n \dot{d}_n \varphi_n(i) \ , \end{aligned} \tag{A1}$$

where

$$\dot{c_n} \equiv d_n, \quad x_i \equiv X_i - X_\mathrm{g}, \quad \Gamma_i = \ddot{x}_i, \quad \Gamma_\mathrm{g} = \ddot{X}_\mathrm{g} \ ,$$

X_g being the position of the center of mass and Ω the angular velocity. Equation (A1) is easily obtained by repeatedly applying the relation

$$\dot{A} = \dot{A}_\mathrm{moving} + \Omega \wedge A$$

relating the time derivatives of a time-depedent vector, A, in a fixed reference ($\dot{A}$) and a moving reference, $\dot{A}_\mathrm{moving}$. We rewrite Eq. (1) as

$$0 = \ddot{\Sigma}_i - Z_i + \dot{\Omega} \wedge \sqrt{m_i} x_i \ , \tag{A2}$$

with

$$\begin{aligned} Z_i = {} & \sqrt{m_i}(\Gamma_i - \Gamma_\mathrm{g}) - \sqrt{m_i}\Omega \wedge (\Omega \wedge x_i) \\ & - 2\Omega \wedge R \sum_n d_n \varphi_n(i) \end{aligned}$$

and

$$\ddot{\Sigma}_i = R \sum_n \ddot{c}_n \varphi_n(i) \ .$$

A simple counting shows that the system Eq. (A2) is ill-posed, as it has fewer unknowns ($N' + 3$) than equations ($3N_\mathrm{a}$). A simple way out is to define the unknowns $\ddot{c}_n$ and $\dot{\Omega}$ as given by the least-squares solution of Eq. (A1), i.e., they minimize the quantity

$$\frac{1}{2} \sum_i (-Z_i + \ddot{\Sigma}_i + \dot{\Omega} \wedge \sqrt{m_i} x_i)^2 \ .$$

It is then straightforward to obtain

$$\mathbf{I}\dot{\Omega} = -\sum_i \sqrt{m_i}(Z_i \wedge x_i) + \sum_i \left(\sqrt{m_i} \sum_n \ddot{c}_n R \varphi_n(i) \wedge x_i \right) \ , \tag{A3}$$

$$\ddot{c}_n = \sum_i [Z_i \cdot R\varphi_n(i)] - \dot{\Omega} \cdot \tilde{\varphi}_n \ , \tag{A4}$$

where

$$\tilde{\varphi}_n \equiv \sum_i \sqrt{m_i} x_i \wedge R\varphi_n(i) \ ,$$

and $\mathbf{I}$ the instantaneous inertia matrix, i.e.,

$$\mathbf{I}_{11} = \sum_i x_i^2 x_i^2 + x_i^3 x_i^3, \quad \mathbf{I}_{12} = -\sum_i x_i^1 x_i^2, \ldots \ , \tag{A5}$$

with

$$x_i = (x_i^1, x_i^2, x_i^3) \ .$$

References

1. Remington S, Weigand G, Huber R (1982) J Mol Biol 158: 111–152
2. Hubert R, Bennett W (1983) Biopolymers 22: 261–279
3. Wiegand G, Remington S (1986) Annu Rev Biophys Chem 15: 97–117
4. Williams JC, McDermott AE (1995) Biochemistry 34: 8309–8319
5. Mchaourab HS, Oh KJ, Fang CJ, Hubbell WL (1997) Biochemistry 36: 307–316
6. Go N, Noguti T, Nishikawa T (1983) Proc Natl Acad Sci USA 80: 3696–3700
7. Brooks B, Karplus M (1985) Proc Natl Acad Sci USA 82: 4995–4999
8. Faure P, Micu A, Perahia D, Doucet J, Smith J, Benoit J (1994) Nat Struct Biol 1: 124–128
9. Bennet W, Steitz T (1980) J Mol Biol 140: 210–230
10. Marques O, Sanejouand Y-H (1995) Proteins 33: 557
11. Tirion M (1996) Phys Rev Lett 77: 1905–1908
12. Durand P, Trinquier G, Sanejouand Y-H (1994) Biopolymers 34: 759–771
13. Amadei A, et al (1993) Proteins 17: 412–425
14. Balsera MA, Wriggers W, Oono Y, Shulten K (1996) J Phys Chem 100: 2576–2572
15. Horiuchi T, Go N (1990) Proteins 10: 106–116
16. Kitao A, et al (1991) J Chem Phys 158: 447–472
17. van Aalten DMF, et al (1995) Proteins 22: 45–54
18. Garcia AE (1992) Phys Rev Lett 68: 2696–2699
19. Kitao A, Hayward S, Go N (1998) Proteins 33: 496–517
20. Amadei A, Linssen ABM, de Groot BL, van Aalten DMF, Berendsen HJC (1996) J Biol Struct Dyn 13: 615
21. Chun HM, Padilla CE, Chin DN, Watanabe M, Karlov VI, Alper HE, Soosar K, Blair KB, Becker OM, Caves LS, Nagle R, Haney DN, Farmer DL (2000) J Comput Chem 21: 159
22. Reich S (1995) Physica D 89: 28–42

23. Schütte C, Bornemann FA (1997) Physica D 102: 57–77
24. Askar A, Space B, Rabitz H (1995) J Phys Chem 99: 7330–7338
25. Durup J (1991) J Phys Chem 95: 1817
26. Buchner M, Ladanyi BM (1991) Mol Phys 73: 1127
27. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4: 187–217
28. Mazur A, Abagyan R (1989) J Biomol Struct Dyn 6: 815
29. Fixman M (1978) J Chem Phys 69: 1527; (b) Fixman M (1978) J Chem Phys 69: 1538
30. Elezgaray J, Sanejouand YH (1998) Biopolymers 46: 493–501
31. Elezgaray J, Sanejouand YH (2000) J Comput Chem 21: 1274–1282
32. Frenkel D, Smit B (1996) Understanding molecular simulation. Academic, London p 239
33. Sanejouand Y-H, Tapia O (1995) J Phys Chem 99: 5698
34. Aqvist J, Leijonmarck M, Tapia O (1989) Eur Biophys J 16: 327
35. Rapaport DC (1995) The art of molecular dynamics simulation. Cambridge University Press, Cambridge
36. Habermann E (1972) Science 177: 314–322
37. Bernèche S, Nina M, Roux B (1998) Biophys J 75: 1603–1618
38. Inagaki F, Shimada I, Kawaguchi K, Hirano M, Terasawa I, Ikura T, Go N (1989) Biochemistry 28: 5985–5991
39. Bazzo R, Tappin MJ, Pastore A, Harvey TS, Carver JA, Campbell ID (1988) Eur J Biochem 173: 139–146
40. Terwillinger TC, Eisenberg D (1982) J Biol Chem 257: 6010–6015
41. Peterson C, Sommelins O, Söderberg B (1996) Phys Rev Lett 76: 1079
42. Grabert H, Hanggi P, Talkner P (1980) J Stat Phys 22: 537
43. Romiszowski P, Yaris R (1991) J Chem Phys 94: 6751