J Math Chem
DOI 10.1007/s10910-012-9987-8

ORIGINAL PAPER

Normal mode analysis of molecular motions in curvilinear coordinates on a non-Eckart body-frame: an application to protein torsion dynamics

Janne Pesonen · Krister O. E. Henriksson · Jose Ramon López-Blanco · Pablo Chacón

Received: 17 January 2012 / Accepted: 17 February 2012

Abstract Normal mode analysis (NMA) was introduced in 1930s as a framework to understand the structure of the observed vibration-rotation spectrum of several small molecules. During the past three decades NMA has also become a popular alternative to figuring out the large-scale motion of proteins and other macromolecules. However, the "standard" NMA is based on approximations, which sometimes are unphysical. Especially problematic is the assumption that atoms move only "infinitesimally", which, of course, is an oxymoron when large amplitude motions are concerned. The "infinitesimal" approximation has the further unfortunate side effect of masking the physical importance of the coupling between vibrational and rotational degrees of freedom. Here, we present a novel formulation of the NMA, which is applied for finite motions *in non-Eckart body-frame*. Contrary to standard normal mode theory, our approach starts by assuming a harmonic potential in generalized coordinates, and tries to *avoid* the linearization of the coordinates. It also takes explicitly into account the Coriolis terms, which couple vibrations and rotations, and the terms involving Christoffel symbols, which are ignored by default in the standard NMA. We also computationally explore the effect of various terms to the solutions of the NMA equation of motions.

J. Pesonen (✉)
Department of Chemistry, University of Helsinki, P.O. Box 55 (A. I. Virtasen aukio 1), 00014 Helsinki, Finland
e-mail: janne.pesonen@helsinki.fi

K. O. E. Henriksson
Department of Physics, University of Helsinki, P.O. Box 43 (Pietari Kalmin katu 2), 00014 Helsinki, Finland

J. R. López-Blanco · P. Chacón
Department of Biological Physical Chemistry, Rocasolano Physical Chemistry Institute, CSIC, Serrano 119, 28006 Madrid, Spain

Published online: 11 March 2012

Keywords Normal modes · Curvilinear coordinates · Non-Eckart frame · Coriolis coupling · Vibration-rotation Lagrangian · Non-Euclidean metric · Finite displacements · Christoffel symbol · Equations of motion

1 Introduction

Normal mode analysis (NMA) was introduced in 1930s by Wilson and co-workers as a framework to understand the structure of the observed vibration-rotation spectrum of several small molecules [1]. Due to its relative simplicity in implementation, it achieved popularity and became the tool of the trade for researchers working on theoretical molecular spectroscopy. Anharmonic effects were included by modifying the theory somewhat (see e.g., Ref. [2]). During the past three decades NMA has also become a popular alternative to figuring out the large-scale motion of proteins and other macromolecules (see some recent reviews in Refs. [3–6], and check a recent software tool for NMA in internal coordinates in Ref. [7]). There, the aim is to reduce numbers of degrees of freedom by a judicious choice of the shape coordinates. For example, the protein large-scale motion is to a large extent determined by the torsion angles only [8] (and the number of active torsion coordinates can be far less than $3N-6$). This lead Go and co-workers [9,10] to further develope NMA as a complete mathematical framework for harmonic motions in dihedral angle space.

The basic assumption of NMA is that the potential V of the system varies quadratically with the shape coordinates, and the kinetic energy T of the system varies quadratically with the velocities (or generalized momentas, if Hamiltonian formulation is used) about a given minimum energy conformation. Perhaps as a reflection of its historical origins, NMA is usually performed in linearized internal coordinates (see, e.g., Refs. [11–18]), and not in the true geometrically defined internal coordinates. The rotational motions are almost always separated out from the internal (vibration) motions using an Eckart body-frame [19–25]. However, some of the approximations behind the "standard" NMA are unphysical. Especially problematic is the assumption that atoms move only "infinitesimally". This very idea is an oxymoron when large amplitude motions are concerned. The "infinitesimal" approximation has the further unfortunate side effect of masking the physical importance and the mathematical implications of the coupling between vibrational and rotational degrees of freedom—The Coriolis coupling is not generally zero outside the reference configuration, and the very definition of linearized shape coordinates actually depends on the choice of the body-frame.

Here, we present a novel formulation of the NMA. It is aimed specifically to describe finite motion of molecules. Contrary to standard normal mode theory, our approach starts by assuming a harmonic potential in generalized coordinates, and tries to *avoid* the linearization of the coordinates. This means that the terms involving Christoffel symbols, which are ignored by default in the standard NMA, are included to the equations of motion. We have chosen to use an *non-Eckart body-frame*, which means that the Coriolis coupling terms are also explicitly included to kinetic energy. The theory developed here is also compared to the methods presented by Wilson et. al.

[1,26], originally developed in the 1930s for small molecules, and later popularized and improved by Go et al. for the modeling of the motions of large biomolecules [10].

The main purpose of the present contribution is to understand the physical consequences of the mathematical structure of different NMA approximations rather than evaluate in detail their computational implementations. Especially, we account for the effects of *non-Euclidean metric*, which one encounters explicitly in the present curvilinear approach, and also implicitly in the "standard" NMA, when finite amplitude motions are considered (as opposed to "infinitesimal" motions). This problematics seems to be so far largely neglected in the existing litterature. We also computationally explore the effect of various terms to the solutions of the NMA equation of motions. For example, we test how well is the total energy or angular momentum preserved over the trajectories produced as the solution to normal mode equations, and how much do the trajectories differ when NMA is done in linearized instead of curvilinear coordinates.

2 Equations of motion

The Euler-Lagrange equations of motion in general curvilinear coordinates $\{q_i\}$ are [27]

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} = 0, \tag{1}$$

Now, $\dot{q}_i$ is a generalized velocity, and the Lagrangian is

$$L = T - V = \frac{1}{2}\sum_{ij} g_{ij}\dot{q}_i\dot{q}_j - V. \tag{2}$$

and the metric tensor is given by

$$g_{ij} = \sum_{\alpha}^{N} m_\alpha \frac{\partial \mathbf{x}_\alpha}{\partial q_i}\cdot\frac{\partial \mathbf{x}_\alpha}{\partial q_j}. \tag{3}$$

where $\mathbf{x}_\alpha$ is the position vector of the particle α in an inertial (laboratory) frame, and m_α is its mass. The resulting equations of motion are

$$\sum_j g_{ij}\ddot{q}_j + \sum_{jk}\Gamma_{ijk}\dot{q}_j\dot{q}_k + \frac{\partial V}{\partial q_i} = 0, \tag{4}$$

where

$$\Gamma_{ijk} = \sum_{\alpha}^{N} m_\alpha \frac{\partial \mathbf{x}_\alpha}{\partial q_i}\cdot\frac{\partial^2 \mathbf{x}_\alpha}{\partial q_j \partial q_k}. \tag{5}$$

is the Christoffel symbol of the first kind [28].

Often, the quantities in Eq. (4) are interpreted as follows: The diagonal element g_{ii} of the metric tensor represent the inertia associated with the ith degree of freedom, and the off-diagonal element g_{ij} reflects the (possible) non-orthogonality of the two degrees of freedom i and j, resulting in a contribution to the acceleration $\ddot{q}_j$. The $\Gamma_{ijj}\dot{q}_j^2$ terms represent the centrifugal effect on the ith degree of freedom by the jth generalized velocity, and the $\Gamma_{ijk}\dot{q}_j\dot{q}_k$ terms represent the Coriolis effect induced by the jth and kth generalized velocities. However, this interpretation is formal at best rather than physical, and it is not completely in accordance with the practice followed in theoretical molecular spectroscopy, where only those terms containing mixed orientational and shape coordinates merit the name Coriolis contribution (compare to Ref. [22]).

3 Harmonic potential

The potential V of a free molecule is a function of the displacements in shape coordinates

$$\Delta s_i = s_i - s_i^{(e)}, \tag{6}$$

where $s_i^{(e)}$ denotes the equilibrium value of coordinate s_i. The shape coordinates are $\{s_i\}$, by definition, invariant in all rigid motions of the molecule. The potential is required to be harmonic in the (displacement) coordinates, i.e., of the form

$$V = \frac{1}{2}\sum_{ij} f_{ij}\Delta s_i \Delta s_j. \tag{7}$$

Any potential V can be used for normal mode calculations provided it is first truncated and/or approximated as harmonic. The simplest way to do this is to evaluate

$$f_{ij} = \left.\frac{\partial^2 V}{\partial s_i \partial s_j}\right|_e \tag{8}$$

at the (local) minimum energy conformation, denoted by the subscript (e), and subsequently use the potential given by Eq. (7). Note that computation of f_{ij} scales $O(N^4)$ when the potential V has an infinite range. If the potential is restricted to some finite range, the computation scales as $O(N^3)$. This computational burden can be significantly further reduced to $O(N^2)$ by employing recursion relationships described in seminal papers of Go and coworkers [29,30].

With this potential the equations of motion become

$$\sum_j g_{ij}\ddot{q}_j + \sum_{j,k}\Gamma_{ijk}\dot{q}_j\dot{q}_k + \sum_j f_{ij}\Delta s_j = 0. \tag{9}$$

where q_i is either a rotational angle Θ_i (e.g., an Euler angle) or Δs_i. By setting $g_{ij} = m_i \delta_{ij}$ and $\Gamma_{ijk} = 0$ and replacing q_i with the Cartesian coordinates the corresponding equations for usual Cartesian dynamics are recovered. In general $\Gamma_{ijk} \neq 0$ for curvilinear coordinates, so this problem is not as easily solved as the Cartesian counterpart.

4 Normal mode calculation

4.1 Eigenequation

For the time being we ignore the Christoffel symbol, and introduce another set of coordinates $\{Q_i\}$ through a linear coordinate relation

$$q_i = \sum_j A_{ij} Q_j \tag{10}$$

The equations of motion (without the Christoffel symbol) in these new coordinates can be expressed in the matrix form as

$$\mathsf{A}^T \mathsf{g} \mathsf{A} \ddot{\mathsf{Q}} + \mathsf{A}^T \mathsf{f} \mathsf{A} \mathsf{Q} = 0, \tag{11}$$

where $\mathsf{g} = [g_{ij}]$ and $\mathsf{f} = [f_{ij}]$ are square matrices, $\mathsf{Q} = \left|Q_j\right)$ is a column vector, and the superscript T implies transpose. Because the potential energy depends only on the shape coordinates, all those entries f_{ij} of the matrix f are zero, in which either i or j (or both) index a orientational degree of freedom.

Manipulation of Eq. (11) gives

$$\ddot{\mathsf{Q}} + \mathsf{A}^{-1} \mathsf{g}^{-1} \mathsf{f} \mathsf{A} \mathsf{Q} = 0. \tag{12}$$

By using the trial solution

$$Q_j(t) = a_j \sin\left(\omega_j t\right) + b_j \cos\left(\omega_j t\right) \tag{13}$$

which implies[1]

$$\dot{Q}_j(t) = \omega_j a_j \cos\left(\omega_j t\right) - \omega_j b_j \sin\left(\omega_j t\right) \tag{14}$$

and

$$\ddot{Q}_j(t) = -\omega_j^2 a_j \sin\left(\omega_j t\right) - \omega_j^2 b_j \cos\left(\omega_j t\right) = -\omega_j^2 Q_j(t) \tag{15}$$

[1] Given the initial values of the coordinates $\{q_i\}$ and the corresponding velocities $\{\dot{q}_i\}$, the appropriate constants $\left\{a_j, b_j\right\}$ can be evaluated by the procedure explained in detail in Appendix B (or vice versa, the appropriate initial values can be calculated from the given coefficients).

Eq. (12) can be written in a *normal* form as

$$\mathsf{A}^{-1}\mathsf{g}^{-1}\mathsf{f}\mathsf{A} = \Omega^2 \tag{16}$$

where the diagonal matrix Ω^2 contains the *mode frequencies* squared, ω_i^2. The first N_s diagonal elements of the matrix Ω^2 differ from zero, and the remaining three are zero. Then, one obtains the eigenvalue problem

$$(\mathsf{g}^{-1}\mathsf{f})\mathsf{A} = \mathsf{A}\Omega^2 \tag{17}$$

from which the matrix A can be solved (its columns are the eigenvectors of $\mathsf{g}^{-1}\mathsf{f}$).

It is explicitly assumed in the normal mode calculation that the elements of the metric tensor g_{ij}, and consequently, the coefficients A_{ai} are constant. Without this approximation the normal mode frequencies ω_i would depend on the amplitudes of the vibrations. Therefore it is tacitly understood that the variable metric tensor g is replaced by the constant g_0, in which the shape coordinates s_i are fixed to their reference values $s_i^{(e)}$ and the rotational angles Θ_i are fixed to their initial values $\Theta_i|_{t_0}$ at the time $t = t_0$. Hence, the eigenvalue problem reads as

$$(\mathsf{g}_0^{-1}\mathsf{f})\mathsf{A} = \mathsf{A}\Omega^2 \tag{18}$$

To emphasize that rotational angles are set to their initial value, we use the subscript 0 in g, although the initial value $s_i|_{t_0}$ of the shape coordinate s_i at $t = t_0$ need not coincide with its reference value $s_i^{(e)}$ (i.e., molecule can be deformed at the $t = t_0$).[2]

4.2 Matrix A

Because the matrix element $\left[\mathsf{g}_0^{-1}\mathsf{f}\right]_{ij}$ is zero if j indexes an orientational degree of freedom, it follows that the matrix $\mathsf{g}_0^{-1}\mathsf{f}$ has the structure

$$\mathsf{g}_0^{-1}\mathsf{f} = \left[\begin{array}{ccccc|ccc} \star & \star & \star & \cdots & \star & 0 & 0 & 0 \\ \star & \star & \star & \cdots & \vdots & 0 & 0 & 0 \\ \star & \star & \star & \cdots & \star & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & 0 & 0 & 0 \\ \star & \star & \star & \cdots & \star & 0 & 0 & 0 \end{array}\right] \tag{19}$$

where orientation angles are now indexed from $N_\mathrm{s}+1$ to $N_\mathrm{s}+3$ (N_s is the number of active shape coordinates), and the (possibly) non-zero entries of the $(N_\mathrm{s}+3)\times N_\mathrm{s}$ sub-block are denoted by a star $\star$. It follows therefore that the matrix A has the form

[2] For now on, h_0 or $h|_0$ refers to the value of a quantity h, when the rotational angles are set to their initial value, and the shape coordinates are fixed to their reference values.

$$\mathsf{A} = \left[\begin{array}{cccc|ccc} \star & \star & \dots & \star & 0 & 0 & 0 \\ \star & \star & \dots & \star & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \star & \star & \dots & \star & 0 & 0 & 0 \\ \hline \star & \star & \dots & \star & \star & \star & \star \\ \star & \star & \dots & \star & \star & \star & \star \\ \star & \star & \dots & \star & \star & \star & \star \end{array}\right] \tag{20}$$

where the first N_s columns are the eigenvectors of the non-zero frequency normal modes, and the last three columns are the eigenvectors of the zero-frequency normal modes. Hence, as expected, shape coordinates are mapped to the *non-zero* frequency normal modes

$$\Delta s_i = \sum_{j=1}^{N_\mathrm{s}} A_{\Delta s_i Q_j} Q_j \tag{21}$$

but the orientational angles are mapped to *all* normal modes

$$\Theta_i = \sum_{j=1}^{N_\mathrm{s}+3} A_{\Theta_i Q_j} Q_j \tag{22}$$

For simplicity, here an element of A is labeled by the coordinates it is associated with. This is the expected result—under any coordinate transformation, which preserves the nature of coordinates, shape coordinates are mapped to shape coordinates, but the orientational angles are generally mapped to *both* the new orientational angles *and* shape coordinates.

It must be emphasized that A is really determined by the above procedure only up to a "normalization". Here, it is chosen so that the equation

$$\mathsf{g}_0^{-1} = \mathsf{A}\mathsf{A}^T \tag{23}$$

holds true. In other words, A is now the positive "square root" of the g_0^{-1}. Also, as can be seen straightforwardly, it maps g_0 to a unit matrix 1 via the congruent transformation

$$\mathsf{A}^T \mathsf{g}_0 \mathsf{A} = 1 \tag{24}$$

and f to

$$\mathsf{A}^T \mathsf{f} \mathsf{A} = \Omega^2 \tag{25}$$

This particular choice of A does not affect the eigenvalue problem $\mathsf{A}^{-1}\mathsf{g}_0^{-1}\mathsf{f}\mathsf{A} = \Omega^2$. Nor does it change the general form of the matrix $\mathsf{A}^T\mathsf{f}\mathsf{A}$ – It has the structure

$$\mathsf{A}^T\mathsf{f}\mathsf{A} = \left[\begin{array}{cccc|ccc} \lambda_1 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_{N_\mathrm{s}} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] \tag{26}$$

where $\lambda_i > 0$.

It depends on the particular numerical algorithm whether the resulting A has this desired form. A general procedure of normalizing any A, which obeys Eq. (16), is presented in Appendix A.

4.3 On invariants

Let us now inspect the trajectories of the atoms, which result from the solution of the NMA equation of motion. The total momentum $\mathbf{P} = \sum_\alpha m_\alpha \dot{\mathbf{x}}_\alpha$ is obviously conserved, because arbitrary changes in $\{s_i\}$ and $\{\Theta_i\}$ do not change the center of mass $\mathbf{X}$. This comes as no surprise, since $\mathbf{P}$ is a constant of motion for all conservative systems.

The total energy $E = T + V$ should also be invariant for a conservative system. Now,

$$T = \frac{1}{2}\sum_\alpha m_\alpha \dot{\mathbf{x}}_\alpha^2 = \frac{1}{2}\sum_\alpha m_\alpha \sum_{ij} \frac{\partial \mathbf{x}_\alpha}{\partial q_i}\cdot\frac{\partial \mathbf{x}_\alpha}{\partial q_j}\cdot \dot{q}_i\dot{q}_j = \frac{1}{2}\sum_{ijkl} g_{ij}A_{ik}A_{jl}\dot{Q}_k\dot{Q}_l \tag{27}$$

and

$$V = \frac{1}{2}\sum_{ijkl}^{N_\mathrm{s}} f_{ij}A_{ik}A_{jl}Q_kQ_l \tag{28}$$

or, in the matrix form,

$$2E = \dot{\mathsf{Q}}^T\mathsf{A}^T\mathsf{g}\mathsf{A}\dot{\mathsf{Q}} + \mathsf{Q}^T\mathsf{A}^T\mathsf{f}\mathsf{A}\mathsf{Q} \tag{29}$$

By evaluating the above equation at the reference conformation, it would be tempting to say that the total energy equals $\sum_i^{N_\mathrm{s}} \omega_i^2\left(a_i^2 + b_i^2\right)/2$. However, this is true *only in the reference conformation*—the upper left $N_\mathrm{s} \times N_\mathrm{s}$ sub-block of $\mathsf{A}^T\mathsf{g}\mathsf{A}$ is not diagonal in an arbitrary conformation. Hence, the total energy, calculated *directly* from the trajectories of the atoms, is *not* a constant of motion (although the amount of variation depends on a particular problem, and E may stay *almost* constant). The magnitude of the non-conservation of E in the normal mode trajectories in a torsion space of a the ribosomal protein is numerically explored in Sect. 7. It can be shown by a specific numerical example (Sect. 7) that the total angular momentum

$$\mathbf{L} = \sum_{\alpha} m_\alpha \mathbf{x}_\alpha \times \dot{\mathbf{x}}_\alpha \tag{30}$$

is not a constant of motion over the trajectories produced as a solution to NMA equations. This reflects again the fact that the normal mode trajectories are not strictly speaking physical, although they can be excellent approximations.

5 The Christoffel symbol

Since the evaluation of $\partial \mathbf{x}_\alpha / \partial q_i$ scales as $O(N)$ for a fixed value of i, as does the evaluation of $\partial^2 \mathbf{x}_\alpha / \partial q_i \partial q_j$ for a fixed value of i and j, it follows that the cost of the (brute force) numerical evaluation of the Christoffel symbol Γ_{ijk} scales as $O\left(N^4\right)$, or to be more precise, $O\left(N_A^3 N\right)$, where N_A is the number of active or variable coordinates (i.e., those coordinates which are not rigidly fixed). Hence its numerical evaluation is quite expensive. We now explore the effect of including the Christoffel symbol to the original equations of motion.

In order to decouple the equations of motion, the Christoffel symbol $\bar{\Gamma}_{ijk}$ (written in terms of the new normal coordinates Q_i) should vanish. According to tensor calculus [28] the Christoffel symbol in Q space can be written in q space as

$$\bar{\Gamma}_{ijk} = \sum_{rst} \Gamma_{rst} \frac{\partial q_r}{\partial Q_i} \frac{\partial q_s}{\partial Q_j} \frac{\partial q_t}{\partial Q_k} + \sum_{rs} g_{rs} \frac{\partial q_r}{\partial Q_i} \frac{\partial^2 q_s}{\partial Q_j \partial Q_k}. \tag{31}$$

Instead of simply setting $\bar{\Gamma}_{ijk}$ to zero, one could express q_i as a quadratic polynomial

$$q_i = \sum_j A_{ij} Q_j + \sum_{jk} B_{ijk} Q_j Q_k \tag{32}$$

in the normal coordinates, and select the constants A_{ij}, B_{ijk} so that $\bar{\Gamma}_{ijk}\big|_0 = \bar{\Gamma}_{ijk}\Big|_{\left\{\left\{s_i^{(e)}\right\}, \left\{\Theta_i|_0\right\}\right\}}$ vanishes. Now,

$$\bar{\Gamma}_{ijk}\big|_0 = \sum_{rst} \Gamma_{rst}|_0 A_{ri} A_{sj} A_{tk} + 2 \sum_{rs} g_{rs}|_0 A_{ri} B_{sjk} = 0 \tag{33}$$

Multiplying with A_{iv}^{-1} and summing over i we get

$$\begin{aligned} 0 &= \sum_{rst} \Gamma_{rst}|_0 \delta_{rv} A_{sj} A_{tk} + 2 \sum_{rs} g_{rs}|_0 \delta_{rv} B_{sjk} \\ &= \sum_{st} \Gamma_{vst}|_0 A_{sj} A_{tk} + 2 \sum_{s} g_{vs}|_0 B_{sjk} \end{aligned} \tag{34}$$

Multiplying with $\left[\mathsf{g}^{-1}\right]_{iv}\big|_0$ and summing over v we get

$$\begin{aligned} 0 &= \sum_{stv} \left[\mathsf{g}^{-1}\right]_{iv} \Gamma_{vst}\Big|_0 A_{sj} A_{tk} + 2\sum_{s} \delta_{is} B_{sjk} \\ &= \sum_{stv} \left[\mathsf{g}^{-1}\right]_{iv} \Gamma_{vst}\Big|_0 A_{sj} A_{tk} + 2B_{ijk} \end{aligned} \tag{35}$$

We then have

$$B_{ijk} = -\frac{1}{2}\sum_{stv} \left[\mathsf{g}^{-1}\right]_{iv} \Gamma_{vst}\Big|_0 A_{sj} A_{tk} \tag{36}$$

or, by introducing the Christoffel symbols of the second kind,

$$B_{ijk} = -\frac{1}{2}\sum_{st} \Gamma^{(i)}_{st}\Big|_0 A_{sj} A_{tk} \tag{37}$$

The Christoffel symbol of the second kind can be written as

$$\Gamma^{(i)}_{jk} = \sum_{\alpha=1}^{N} \left(\nabla_{\mathbf{x}_\alpha} q_i\right) \cdot \frac{\partial^2 \mathbf{x}_\alpha}{\partial q_j \partial q_k} \tag{38}$$

which offers both theoretically and computationally feasible way of evaluating $\Gamma^{(i)}_{jk}\Big|_0$.

Notice that adding the quadratic term $\sum_{jk} B_{ijk} Q_j Q_k$ does not affect the normal mode calculation—It contributes neither to the Hessian $\bar{f}_{ij}$ nor to the metric tensor $\bar{g}_{ij}$ at the reference conformation. However, the potential energy becomes up to quartic in Q_i, i.e.,

$$\begin{aligned} V \text{ (with CS)} = &\frac{1}{2}\sum_{i}^{N_\mathrm{s}} \omega_i^2 Q_i^2 + \frac{1}{2}\sum_{ijklm}^{N_\mathrm{s}} f_{ij}\left(A_{ik}B_{jlm} + A_{jk}B_{ilm}\right) Q_k Q_l Q_m \\ &+\frac{1}{2}\sum_{ijklmn}^{N_\mathrm{s}} f_{ij} B_{ikn} B_{jlm} Q_k Q_l Q_m Q_n \end{aligned} \tag{39}$$

which means that for a *given* values of $\{Q_i\}$ the value of V (with CS) is different than that of V (without CS). The similar reasoning applies also for the value of kinetic energy T (with CS)—It in general differs from T (without CS) . Furthermore, not all the coefficients $B_{\Delta s_i Q_j Q_k}$ vanish for $j = N_\mathrm{s}+1, N_\mathrm{s}+2, N_\mathrm{s}+3$. Hence inclusion of the Christoffel symbols to the normal equations of motion mixes the shape coordinates $\{\Delta s_i\}$ with the rotational coordinates $\{\Theta_1, \Theta_2, \Theta_3\}$. From the physical point of view this is unwanted, since shape coordinates should always be mapped to shape coordinates, not to rotational coordinates, although the converse need not be true.

6 Body frame and its changes

It is customary to describe motions of atoms in a molecule by introducing a concept of the *instantaneous reference configuration*. It refers to the instantenous positions $\{\mathbf{c}'_\alpha\}$ of atoms in a (hypothetical) *rigid molecule*, which can merely rotate. Because rotations preserve distances, inter-vector angles, and directed volumes, the inner products $\mathbf{c}'_\alpha \cdot \mathbf{c}'_\beta$ and the box products $\mathbf{c}'_\alpha \cdot \mathbf{c}'_\beta \times \mathbf{c}'_\gamma$ are constants, which depend only on the reference geometry of the molecule. In reality, the molecule can of course also deform. It is important to understand that the instantenous value of the set of rotational angles $\{\Theta_1, \Theta_2, \Theta_3\}$ at the time t is the same for both the actual (deforming) molecule and its reference configuration (while the shapes are of course different). We must also carefully distinguish between the initial value and the reference value of a given quantity. For example, while the reference value $s_i^{(e)}$ of the shape coordinate s_i is constant (and may, but need not, equal to $s_i|_{t_0}$), the reference value of the rotational angle is equal to the instantenous value of Θ_i, so it is variable and it in general differs from the initial value $\Theta_i|_{t_0}$.

The orientation of the instantaneous reference configuration is described by the orientation of the body frame $\{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$ (i.e., an orthonormal vector triplet attached to the molecule) with respect to some *constant* orthonormal frame $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, referred as the laboratory frame, i.e.,

$$\mathbf{u}'_i = R^\dagger \mathbf{u}_i R \tag{40}$$

The time dependent rotor [31] $R(t)$ is fully parametrized by three angles $\{\Theta_1, \Theta_2, \Theta_3\}$ (such as Euler angles). The instantaneous reference positions $\{\mathbf{c}'_\alpha\}$ are related to a set of constant vectors $\{\mathbf{c}_\alpha\}$ by the same rotation,

$$\mathbf{c}'_\alpha = R^\dagger \mathbf{c}_\alpha R \tag{41}$$

It is customary to decompose the nuclear position $\mathbf{y}_\alpha = \mathbf{x}_\alpha - \mathbf{X}$ measured from the center of mass $\mathbf{X}$ as a sum

$$\mathbf{y}_\alpha = \mathbf{c}'_\alpha + \mathbf{d}_\alpha \tag{42}$$

where $\mathbf{d}_\alpha$ is the displacement of the atom α from the instantaneous reference position (see Wilson et al. [1]). If the molecule is rigid (i.e., does not deform), then $\mathbf{d}_\alpha = 0$ and $\mathbf{y}_\alpha = \mathbf{c}'_\alpha$ at all times. The instantaneous reference vectors $\{\mathbf{c}'_\alpha\}$ and the displacement vectors $\{\mathbf{d}_\alpha\}$ are illustrated in Fig. 1 for a triatomic molecule. The bond-z body-frame (in which the instantaneous axis of rotation $\mathbf{u}'_3$ is parallel with the bond between the atoms 1 and 3) is indicated. The numerical evaluation of the instantaneous values of $\mathbf{c}'_\alpha$ and $\mathbf{d}_\alpha$ is described in Appendix D.

It should be emphasized that the decomposition of $\mathbf{y}_\alpha = \mathbf{c}'_\alpha + \mathbf{d}_\alpha$ to the displacement $\mathbf{d}_\alpha$ and the moving reference position $\mathbf{c}'_\alpha$ is not unique, but it (implicitly) depends on the choice of the body-frame through the rotor R.

In order to see the possible effect of the body-frame change $\{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\} \rightarrow \{\mathbf{u}''_1, \mathbf{u}''_2, \mathbf{u}''_3\}$ on the normal modes, it is advisable to consider it as the coordinate

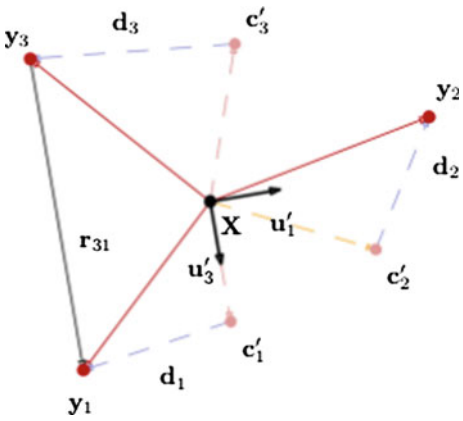

Fig. 1 A snapshot of a triatomic molecule, and the body-frame associated with it. Notice that for this particular choice of the body-frame, the axis $\mathbf{u}'_3$, the instantaneous reference position $\mathbf{c}'_1$ and the inter particle position vector $\mathbf{r}_{31}$ (also shown in the *picture*) are always parallel

transformation $\{\{s_i\}, \{\Theta'_i\}\} \rightarrow \{\{s_i\}, \{\Theta''_i\}\}$, in which the three rotational angles $\{\Theta'_i\}$ are transformed to another set of rotational angles $\{\Theta''_i\}$. The single (double) prime(s) indicates that the $\{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$ ($\{\mathbf{u}''_1, \mathbf{u}''_2, \mathbf{u}''_3\}$) is the body-frame. By using the chain rule of derivation, we can write

$$\mathsf{f}'' = \mathsf{J}_0 \mathsf{f}' \mathsf{J}_0^T \tag{43}$$

$$\mathsf{g}''_0 = \mathsf{J}_0 \mathsf{g}'_0 \mathsf{J}_0^T \tag{44}$$

where

$$\mathsf{J}_{ij} = \frac{\partial q'_j}{\partial q''_i} \tag{45}$$

is the element of the Jacobian matrix J of the coordinate transformation $\{\{s_i\}, \{\Theta'_i\}\} \rightarrow \{\{s_i\}, \{\Theta''_i\}\}$, which is induced by the change in the body-frame $\{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\} \rightarrow \{\mathbf{u}''_1, \mathbf{u}''_2, \mathbf{u}''_3\}$. J_0 is the value of J evaluated by setting shape coordinates to their reference values and the orientational coordinates to their intial values, and J_0^T signifies the transpose of J_0. The left-hand side of the normal mode equation $\mathsf{g}_0^{''-1}\mathsf{f}'' = \mathsf{A}\Omega^2\mathsf{A}^{-1}$ (where it has been re-arranged to a form, in which the unknown quantities to be solved are on the right hand side) can be written as

$$\mathsf{g}_0^{''-1}\mathsf{f}'' = \left(\mathsf{J}_0^T\right)^{-1} \mathsf{g}_0^{'-1} \mathsf{J}_0^{-1} \mathsf{J}_0 \mathsf{f}' \mathsf{J}_0^T \tag{46}$$

or

$$\mathsf{g}_0^{''-1}\mathsf{f}'' = \left(\mathsf{J}_0^T\right)^{-1} \mathsf{g}_0^{'-1} \mathsf{f}' \mathsf{J}_0^T \tag{47}$$

Evidently, the two products $\mathsf{g}_0^{'-1}\mathsf{f}'$ and $\mathsf{g}_0^{''-1}\mathsf{f}''$ are similar, so their eigenvalues and eigenvectors do *not* depend on the choice of the body-frame. Starting with a given potential $V(s_1, s_2, \ldots)$,*both* the equations $\mathsf{g}_0^{'-1}\mathsf{f}' = \mathsf{A}\Omega^2\mathsf{A}^{-1}$ and $\mathsf{g}_0^{''-1}\mathsf{f}'' = \mathsf{A}\Omega^2\mathsf{A}^{-1}$

Table 1 The first five non-zero eigenfrequencies ω_i for protein 1AB3, when the orientational angles are (a) included, and (b) excluded

Mode	Eigenfrequency (fs^{-1})	
	(a)	(b)
1	5.8507×10^{-4}	5.3266×10^{-4}
2	7.7745×10^{-4}	2.3495×10^{-3}
3	9.9830×10^{-4}	3.9234×10^{-3}
4	1.2310×10^{-3}	1.8527×10^{-2}
5	1.4476×10^{-3}	2.5289×10^{-2}
…	…	…

(or rather $\mathsf{g}_0'^{-1}\mathsf{f} = \mathsf{A}\Omega^2\mathsf{A}^{-1}$ and $\mathsf{g}_0''^{-1}\mathsf{f} = \mathsf{A}\Omega^2\mathsf{A}^{-1}$, as the Hessians are identical) will result the same normal frequencies ω_i and (relative) amplitudes.

It is emphasized that the above similarity relation holds generally only when the rotational degrees of freedom are included in the calculation. Although the Hessian involves only the vibrational degrees of freedom (i.e., the entries involving rotational degrees of freedom are zero), at least some elements in the metric tensor, which couple rotational and vibrational degrees of freedom differ from zero, even in the reference conformation, unless one uses an Eckart body-frame (See Sect. 6.1). Then the resulting set of eigenmodes contains three eigenfrequencies which are zero. In practice, the resulting non-zero modes may differ significantly from the modes obtained when one takes only the vibrational degrees of freedom into account. This is demonstrated in Sect. 7, Table 1. It should be emphasized that the normal modes derived in the absence of orientational coordinates are not physically meaningful unless an Eckart frame is used.

6.1 Eckart condition

If the body-frame is determined from the *Eckart condition* [19–24]

$$\sum_{\alpha=1}^{N} m_\alpha \mathbf{c}'_\alpha \times \mathbf{y}_\alpha = \sum_{\alpha=1}^{N} m_\alpha \mathbf{c}'_\alpha \times \mathbf{d}_\alpha = 0 \tag{48}$$

then all the Coriolis terms

$$g_{s_i\Theta_j} = \sum_{\alpha}^{N} m_\alpha \frac{\partial \mathbf{x}_\alpha}{\partial s_i} \cdot \frac{\partial \mathbf{x}_\alpha}{\partial \Theta_j} \tag{49}$$

vanish *in the reference conformation* (but not necessarily in other conformations). While explicit analytical formulas of Eckart axes are not known for large molecules (apart from planar molecules that is, see Ref. [32]), the rotation matrix associated with an Eckart frame happens to be the matrix that minimizes the root mean square

deviation (RMSD) of the resulting conformation from the reference conformation by rotating the molecule rigidly, and it can be found straightforwardly [33].

7 Test case

The theory developed above was tested on the ribosomal protein S15 from Thermus Thermophilus, cataloged with PDB code 1AB3 at the RCSB Protein Databank. This protein consists of a single chain containing 88 residues. The backbone torsion angles ϕ, ψ and side-chain torsion angle χ (collectively denoted from now on as $\{\phi_i\}$ for simplicity of notation) in each residue—when applicable—were used as degrees of freedom, together with the three orientational angles $\{\Theta_1, \Theta_2, \Theta_3\}$ (defined in terms of a *bond-z body-frame*, which is defined by the coordinates of three atoms). This resulted in a total of 177 degrees of freedom. All calculations were carried out using the previously developed methodology [34,35] as implemented in the computer code TOD.

The atomic interaction potential is a harmonic Cartesian potential:

$$V = \frac{1}{2} \sum_{\alpha\beta} F_{\alpha\beta} \lambda_{\alpha\beta}^2 \tag{50}$$

where

$$\lambda_{\alpha\beta} = \left|\mathbf{r}_{\alpha\beta}\right| - r_{\alpha\beta}^{(e)} = \left|\mathbf{x}_\beta - \mathbf{x}_\alpha\right| - r_{\alpha\beta}^{(e)} \tag{51}$$

For simplicity, put $F_{ij} = 1.0 \times 10^{-3}$ eV/Å^2. The Hessian is now

$$f_{ij} = \left.\frac{\partial^2 V}{\partial\phi_i \partial\phi_j}\right|_e = \sum_{\alpha\beta} F_{\alpha\beta} \left.\frac{\partial}{\partial\phi_i}\left(\lambda_{\alpha\beta}\frac{\partial\lambda_{\alpha\beta}}{\partial\phi_j}\right)\right|_e = \sum_{\alpha\beta} F_{\alpha\beta} \left.\frac{\partial\lambda_{\alpha\beta}}{\partial\phi_i}\frac{\partial\lambda_{\alpha\beta}}{\partial\phi_j}\right|_e \tag{52}$$

where

$$\frac{\partial\lambda_{\alpha\beta}}{\partial\phi_i} = \hat{\mathbf{r}}_{\alpha\beta} \cdot \frac{\partial\mathbf{r}_{\alpha\beta}}{\partial\phi_i} \tag{53}$$

Eigenmodes were calculated for two cases. In case (a) the overall orientational degrees of freedom were included, and in the unphysical case (b) they were not. A comparison of the lowest modes, shown in Table 1, indicates the modes are not identical in these cases. All modes calculated in case (a) are shown in Fig. 2.

The dynamics was also compared in the presence and absence of the quadratic term $\Gamma_{ijk}\dot{q}_j\dot{q}_k$ in the equation of motion. The distribution of all the constants A_{ij} and B_{ijk} is shown in Figs. 3 and 4, respectively. From the figures it is clear that most elements are clustered around zero, without being exactly zero or even "infinitesimally" close to zero. Since B_{ijk} gives a quadratic contribution to the degrees of freedom q_i, this is sufficient to change the dynamics when the Christoffel symbol is included. Note that when only a_1 differs from zero, then $Q_1(t) = a_1 \sin(\omega_1 t)$ by Eq. (13). In that case,

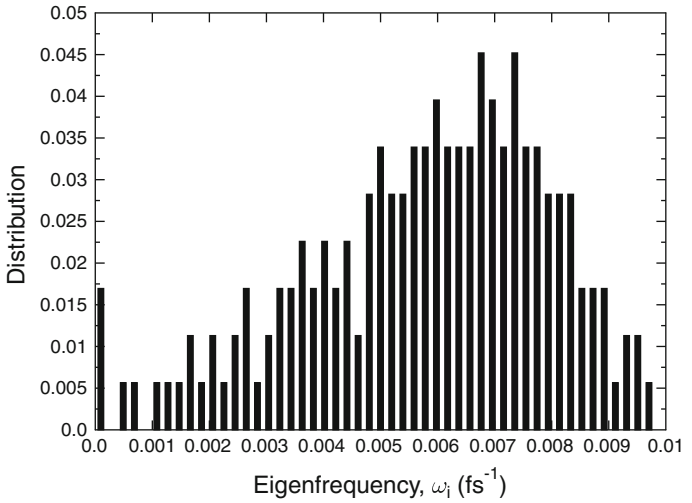

Fig. 2 *Plot* of the distribution of the eigenfrequencies ω_i

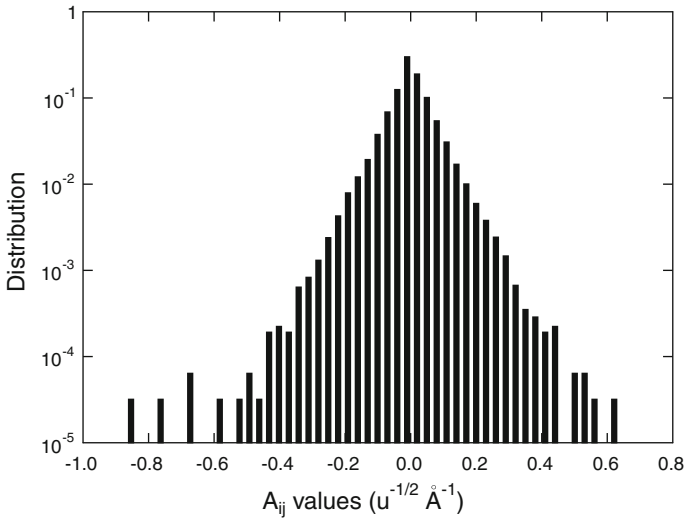

Fig. 3 *Plot* of the distribution of the elements A_{ij} (u stands for the atomic mass unit, or Dalton, $1.660538782(83) \times 10^{-27}$ kg)

when using the Christoffel symbol $q_i = \sum_j A_{ij} Q_j + \sum_{j,k} B_{ijk} Q_j Q_k \sim O(a_1^2)$ by Eq. (32), but when leaving it out $q_i = \sum_j A_{ij} Q_j \sim O(a_1)$. The difference in behavior of the displacement is therefore sensitive to the prefactor of the single excited mode. If it is much smaller than one, the effect of Christoffel symbols becomes neglible.

The initial values of the coordinates and their velocities were determined by setting the a_i and b_i values (and not by specifying intial values for $\{s_i\}$ and $\{\dot{s}_i\}$). By the equipartition theorem every degree of freedom, both kinetic and potential, contributes

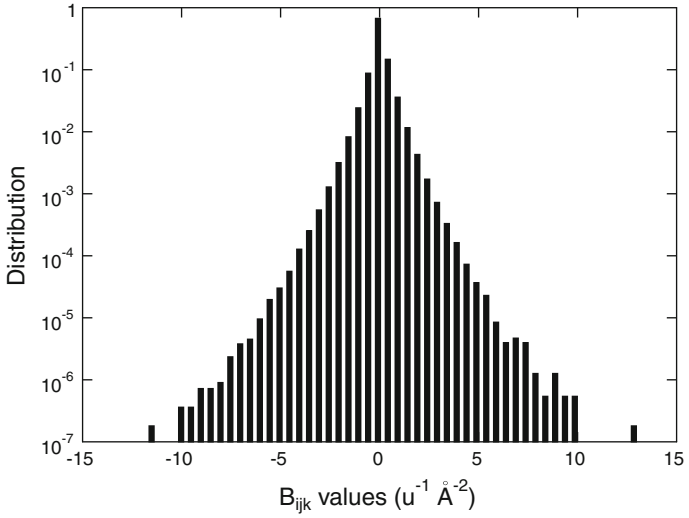

Fig. 4 *Plot* of the distribution of the elements B_{ijk}

a term $k_B T/2$ to the total energy E, which is given by Eq. (29). The time average over one period is independent of the mode frequencies,

$$\langle E \rangle = \sum_i \frac{\omega_i^2}{2}(a_i^2 + b_i^2) = N_s k_B T \tag{54}$$

For the simplicity of comparison, the coefficients $b_i = 0$ are set to zero for all i. This implies that the molecule is initially in the undeformed state (i.e., $\Delta s_i = 0$ for all i). The initial rates of change were then determined from the a_i solely. If only the single mode i is excited, this results in $a_i = \sqrt{2K_B T}/\omega_i$. The corresponding distribution of the intial torsional velocities is presented in Fig. 5 for the temperature of $T = 300\,\mathrm{K}$.

In order to quantify the impact of the Christoffel symbol one may directly compare the resulting different motion of the protein. One possible measure for the difference between the two cases—with and without the Christoffel symbol—is the root-mean-square deviation

$$\mathrm{RMSD} = \sqrt{\frac{1}{N}\sum_{\alpha=1}^{N}\left(\mathbf{x}_\alpha - \mathbf{x}'_\alpha\right)^2} \tag{55}$$

Here N is the number of atomic positions in the protein, $\mathbf{x}_\alpha$ is the position of atom α when the motion is generated using the Christoffel symbol, and $\mathbf{x}'_\alpha$ is the position of atom α when the motion is generated without it. The calculated RMSD is shown in Fig. 6 during a full period, for an initially undeformed molecule at a $T = 300\,\mathrm{K}$, using single modes (a) $i = 1$, (b) $i = 2$, and (c) $i = 3$.

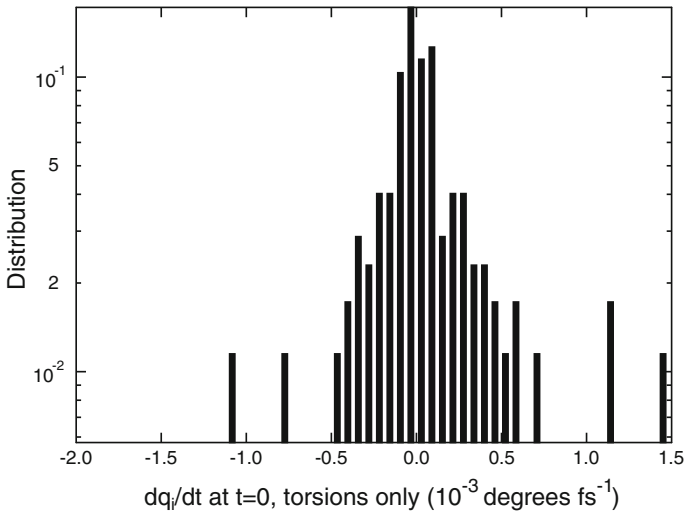

Fig. 5 *Plot* of the distribution of the initial velocities of the torsions, when only the lowest mode is excited, and molecule is initially undeformed and at a $T = 300\,\mathrm{K}$

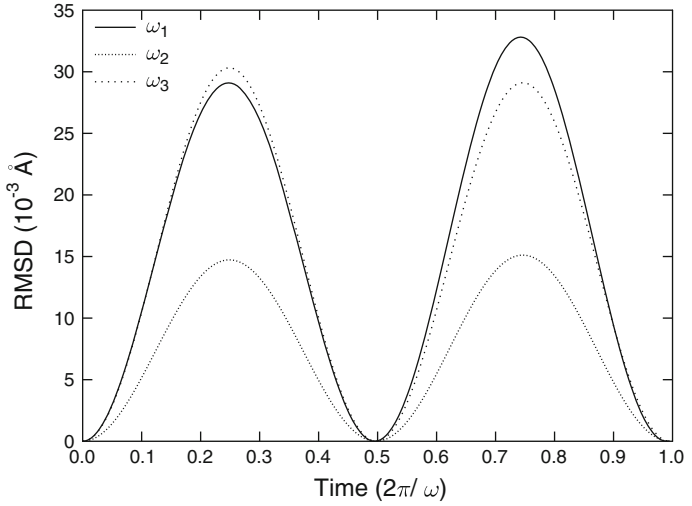

Fig. 6 RMSD during one period of oscillation, using normal mode frequencies of ω_1, ω_2, and ω_3. The molecule is initially undeformed and at a $T = 300\,\mathrm{K}$

The Torsion Angle RMSD is defined as

$$\mathrm{TARMSD} = \sqrt{\frac{1}{N_\phi}\sum_{i=1}^{N_\phi}\left(\phi_i - \phi_i'\right)^2} \tag{56}$$

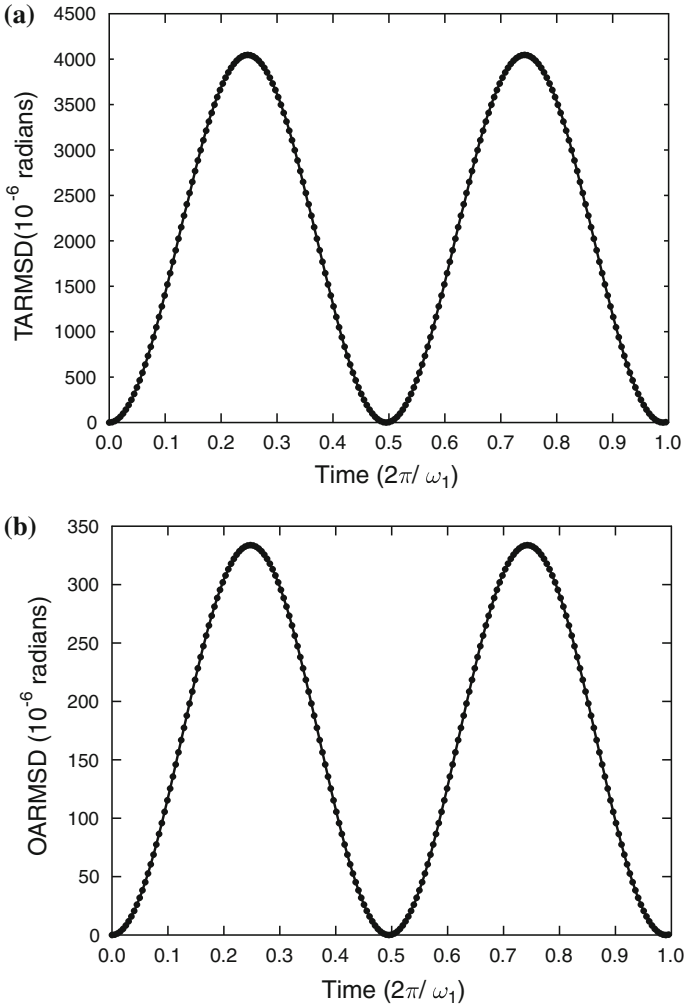

Fig. 7 **a** Torsion Angle RMSD and **b** Orientational Angle RMSD during one period of oscillation, using a normal mode frequency of ω_1. The molecule is initially undeformed and at a $T = 300\,\mathrm{K}$

Here N_ϕ is the number of torsion angles, ϕ_i is the torsion angle i when the motion is generated using the Christoffel symbol, and ϕ_i' is the torsion angle i when the motion is generated without it. The Orientational Angle RMSD is defined in a similar fashion, but it only concerns the three orientational angles $\{\Theta_1, \Theta_2, \Theta_3\}$. The calculated TARMSD and OARMSD values are shown in Fig. 7 during a full period. Values of $a_1 = \sqrt{2K_BT}/\omega_1, b_1 = 0$ and $a_i = b_i = 0$ for $i \neq 1$ were used with $T = 300\,\mathrm{K}$. Again the difference in the dynamical behavior is neglible.

It can be interfered from these graphs that although the detailed dynamics with and without the Christoffel symbol differs, such differences depends crucially on the amplitude. For illustrative modes of Fig. 6, maximal divergences found at maximal

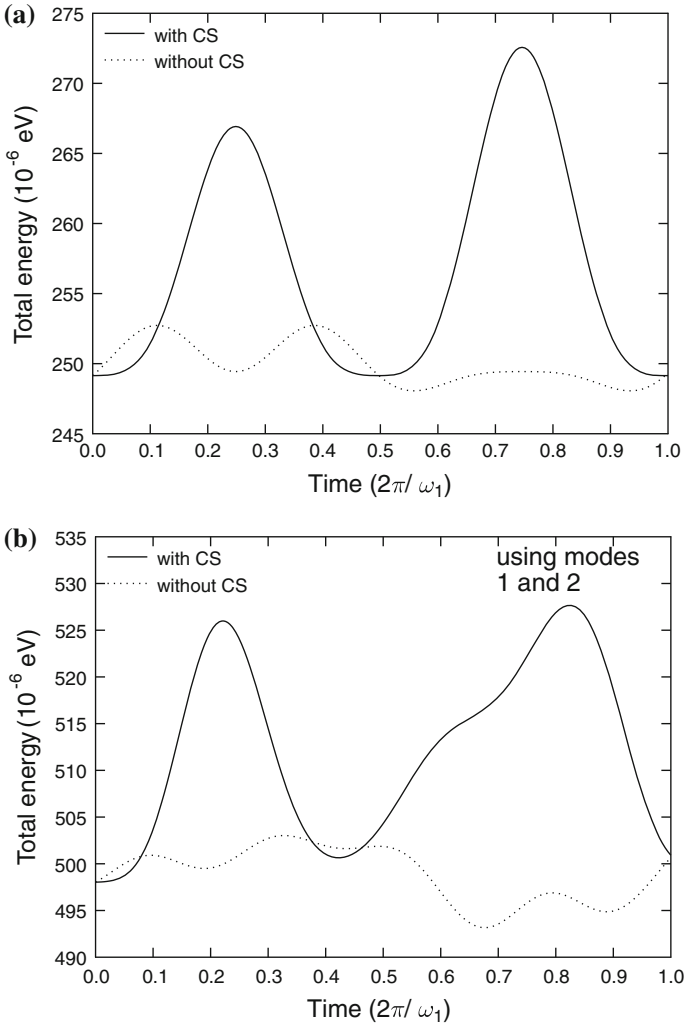

Fig. 8 Total energy $E = T + V$, calculated directly from the trajectories of the atoms, when **a** the lowest **b** the two lowest vibrational normal modes are excited. The molecule is initially undeformed and at a $T = 300\,\mathrm{K}$

sinus amplitudes ($\pi/2$ and $3\pi/2$) were around 30×10^{-3} Å, so the difference in this case is rather small. Similar observations can be made by looking to torsional and rotational space (Fig. 7), where the difference of the detailed dynamics with and without the Christoffel symbol is small.

The total energy E and the magnitude $|\mathbf{l}|$ of the total internal angular momentum $\mathbf{l} = \sum_\alpha m_\alpha \mathbf{y}_\alpha \times \dot{\mathbf{y}}_\alpha$ (where $\mathbf{y}_\alpha = \mathbf{x}_\alpha - \mathbf{X}$) are plotted in Figs. 8 and 9, respectively. Both quantities are calculated directly from the trajectories of the atoms and two cases are depicted: (a) the lowest and (b) the two lowest vibrational normal modes are excited (the molecule is again in $T = 300\,\mathrm{K}$ and initially at a undeformed state, i.e., $b_i = 0$

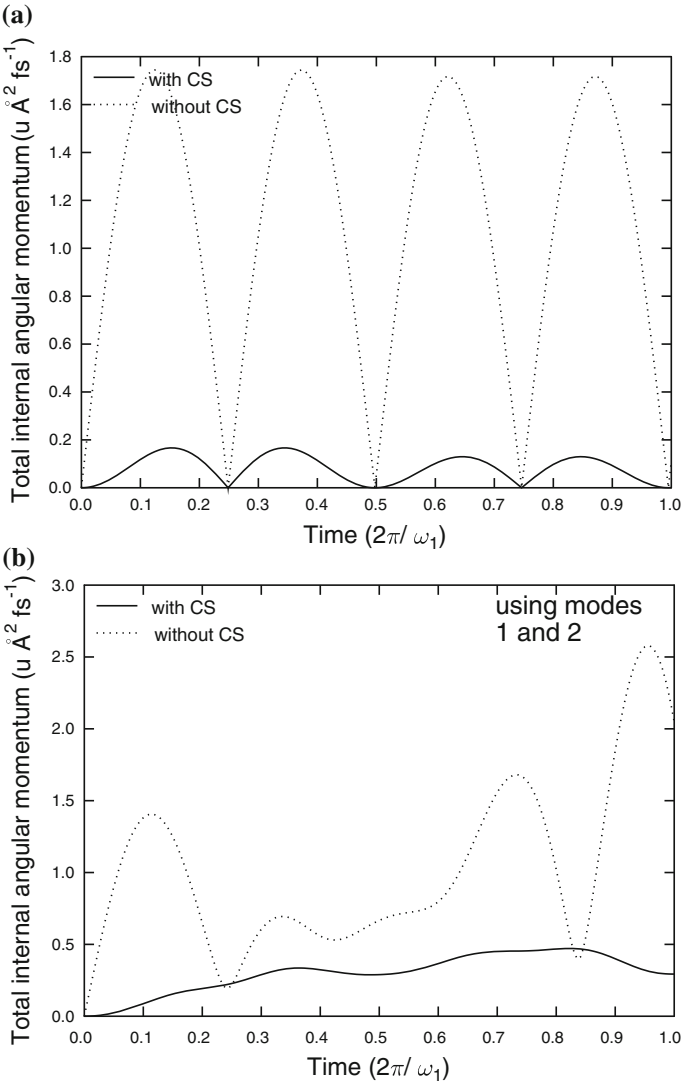

Fig. 9 Magnitude $|\mathbf{l}|$ of the total internal angular momentum $\mathbf{l} = \sum_\alpha m_\alpha \mathbf{y}_\alpha \times \dot{\mathbf{y}}_\alpha$, calculated directly from the trajectories of the atoms, when **a** the lowest **b** the two lowest vibrational normal modes are excited. The molecule is initially undeformed and at a $T = 300\,\mathrm{K}$

for all i). As can be clearly seen, both quantities vary during the period of the normal motion, but the scale of variation is small. Omitting the Christoffel symbols in the original equation of motion results in less variation in of E (which is almost constant over the trajectories obtained as a solution to the NMA calculation) than including them. On the other hand, keeping the Christoffel symbols in the original equation of motion results in less variation in $|\mathbf{l}|$ than omitting them.

Finally, we compare the trajectories produced by the curvilinear normal mode calculation to those produced as a result of the linearized version pioneered by Wilson

et al. in the 1930s, and later popularized by Go et al. for the modeling of the motions of large biomolecules. The eigenvalues (frequencies) and eigenvectors in the linear and curvilinear NMA calculation are identical, but the trajectories produced by these two approaches differ. This is due to the fact that the shape coordinates $\{\bar{s}_i\}$ in linear NMA are *not* equal to their geometrically defined counterparts $\{s_i\}$ in the curvilinear NMA. One measure of the deviation of the trajectories resulted as the solution to curvilinear NMA and linear NMA is the root-mean-square deviation of the corresponding displacements

$$\mathrm{DRMSD} = \sqrt{\frac{1}{N}\sum_{\alpha=1}^{N}\left(\mathbf{d}_\alpha - \bar{\mathbf{d}}_\alpha\right)^2} \tag{57}$$

along, say one period of the lowest vibrational normal mode. DRMSD is obviously equal to the root-mean-square deviation of the atomic positions, as $\mathbf{c}'_\alpha$ is the same in the both cases. In the above equation, $\mathbf{d}_\alpha$ is the displacement resulted by the curvilinear NMA calculation, and

$$\bar{\mathbf{d}}_\alpha = \sum_{ij=1}^{N_\mathrm{s}} \boldsymbol{\xi}_{s_i}^{(\alpha)} A_{\Delta s_i Q_j} Q_j \tag{58}$$

is the displacement resulted from the corresponding linear NMA when the Christoffel symbols are omitted from the equations of motion (and $\boldsymbol{\xi}_{s_i}^{(\alpha)} = \partial \mathbf{x}_\alpha / \partial s_i|_e$ is the reference value of the tangent vector associated with the particle α and curvilinear shape coordinate s_i at the time t). See Appendix D for instructions how to evaluate the reference values of tangents and displacements in practice. As can be seen from Fig. 10, the trajectories resulted from linear NMA calculation for our particular test system deviate slightly from the trajectories obtained as a result of the corresponding curvilinear NMA calculation. The choice of using linear NMA instead of curvilinear one alters the resulting trajectories by a factor 1.3 more than the exclusion of the Christoffel symbols from the curvilinear calculation. A short review of the Lagrangian linear NMA, in which the relevant key features are clarified, and some new proofs are derived is given in Appendix C.

8 Conclusion

We have presented a more general alternative to the commonly used Eckart-frame NMA. Here, the orientational degrees of freedom must be included to obtain valid eigenfrequencies from Eq. (18), as shown in Table 1. Although the detailed dynamics with and without the Christoffel symbol is initially the same, differences appeared when moving along the modal curvilinear coordinates. We have also shown that the trajectories obtained as solution to NMA calculation (both in the absence and presence of the Christoffel symbol in the equations of motion), do not keep the total energy or the total angular momentum constant.

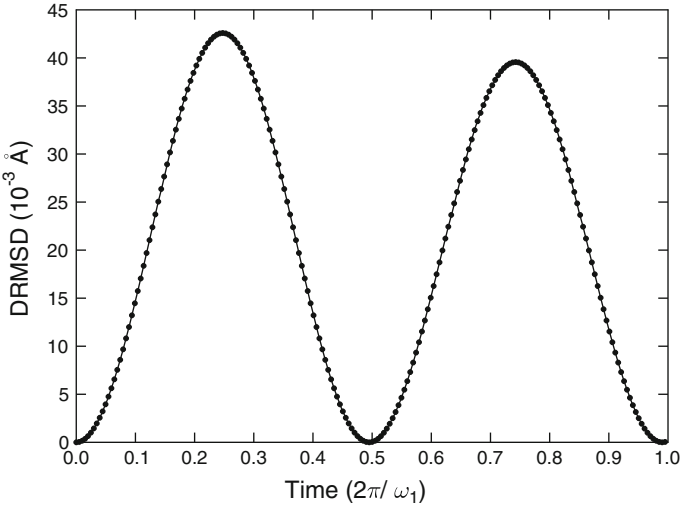

Fig. 10 The root-mean-square deviation of the nuclear displacements resulting from linear v.s. curvilinear normal mode calculation (without the Christoffel symbol). Only the lowest mode is excited. The molecule is initially undeformed and at a $T = 300\,\mathrm{K}$

The theory is also compared to the linearized version popularized by Wilson et al. in the 1930s, and later popularized and improved by Go et al. for the modeling of the motions of large biomolecules. Although the eigenvalues (frequencies) and eigenvectors produced by the linear and curvilinear NMA calculation are identical, the trajectories produced by these two approaches differ. This is due to the fact that the shape coordinates in linear NMA are *not* equal to their geometrically defined counterparts in the curvilinear NMA.

Acknowledgments K. O. E. H. is funded from the Research Grant *Motions in Macromolecular Function: New Approaches to Visualize and Simulate Protein Flexibility*, awarded 2008 by the *Human Frontier of Science Program* (*HFSP*). This study was also supported by the *Ministerio de Educacin y Ciencia of Spain* BFU2009-09552 (PC).

A Normalization of A

Each of the first $j = 1, 2, \ldots, N_\mathrm{s}$ columns (i.e., the eigenvectors of $\mathsf{g}_0^{-1}\mathsf{f}$ associated with the non-zero eigenvalues) of A can be multiplied by some non-zero real without affecting the eigenvalue equation—The resulting new A is still a solution to Eq. (17). Now, for any choice of A, we have

$$\mathsf{A}^T\mathsf{g}_0\mathsf{A} = \left[\begin{array}{cccc|ccc} \mu_1 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & \mu_2 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \mu_{N_\mathrm{s}} & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & \star & \star & \star \end{array}\right] \tag{59}$$

where $\mu_i > 0$, and the $\star$ indicates again a possibly non-zero entry. In order to properly normalize A, we first divide its first $j = 1, 2, \ldots, N_\mathrm{s}$ columns by $\sqrt{\mu_j}$, i.e., set

$$A_{ij} \to \frac{A_{ij}}{\sqrt{\mu_j}} \tag{60}$$

for $i = 1, 2, \ldots, N_\mathrm{s} + 3$ and $j = 1, 2, \ldots, N_\mathrm{s}$. This guarantees that

$$\mathsf{A}^T \mathsf{g}_0 \mathsf{A} = \left[\begin{array}{cccc|ccc} 1 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & \star & \star & \star \\ 0 & 0 & 0 & 0 & \star & \star & \star \end{array}\right] \tag{61}$$

The lower right 3×3 sub-block of A is then substituted by

$$A_{ij} \to \Upsilon_{i-N_\mathrm{s}, j-N_\mathrm{s}} \tag{62}$$

where $i, j = N_\mathrm{s} + 1, N_\mathrm{s} + 2, N_\mathrm{s} + 3$, and the 3×3 matrix Υ is determined from

$$\Upsilon^T \mathsf{g}_0^{(\mathrm{rot})} \Upsilon = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{63}$$

i.e., the jth column of Υ is the jth eigenvector of the symmetric matrix

$$\mathsf{g}_0^{(\mathrm{rot})} = \left.\begin{bmatrix} g_{\Theta_1\Theta_1} & g_{\Theta_1\Theta_2} & g_{\Theta_1\Theta_3} \\ g_{\Theta_1\Theta_2} & g_{\Theta_2\Theta_2} & g_{\Theta_2\Theta_3} \\ g_{\Theta_1\Theta_3} & g_{\Theta_2\Theta_3} & g_{\Theta_3\Theta_3} \end{bmatrix}\right|_{\left\{\left\{s_i^{(e)}\right\}, \left\{\Theta_i|_{t_0}\right\}\right\}} \tag{64}$$

divided by the square root of the respective eigenvalue.

B Initial conditions

Let $q_i(t_0)$ and $\dot{q}_i(t_0)$ be the initial conditions at time $t = t_0$. The constants a_j, b_j can be solved from the initial conditions

$$q_i(t_0) = \left(\sum_r b_r A_{ir}\right) + \sum_{j,k} B_{ijk} \left(\sum_s b_s A_{js}\right)\left(\sum_u b_u A_{ku}\right). \tag{65}$$

Multiplying with A^{-1}_{wi} and summing over i gives

$$b_w = \sum_i A^{-1}_{wi} q_i(t_0) - \sum_{i,j,k} A^{-1}_{wi} B_{ijk} \left(\sum_s b_s A_{js} \right) \left(\sum_u b_u A_{ku} \right). \tag{66}$$

Iterative solution gives b_r for all r. A similar treatment for the initial velocities reveals that the coefficients a_r can be obtained by iterative solution of

$$\begin{aligned} \omega_w a_w = & \sum_i A^{-1}_{wi} \dot{q}_i(t_0) - \sum_{i,j,k} A^{-1}_{wi} B_{ijk} \left(\sum_s \omega_s a_s A_{js} \right) \left(\sum_u b_u A_{ku} \right) \\ & - \sum_{i,j,k} A^{-1}_{wi} B_{ijk} \left(\sum_s b_s A_{js} \right) \left(\sum_u \omega_u a_u A_{ku} \right). \end{aligned} \tag{67}$$

using the known coefficients b_u.

C Linear NMA

Here we shortly review the older approach to normal modes, first developed by Wilson and co-workers [1,26] and later popularized and improved by Go et al. for large biomolecules [10]. We clearly point out what are the actual approximations in that approach, and how the omission of the rotational degrees of freedom affects the solutions of the normal mode equations. So far, this problematics has been almost completely neglected in the existing literature. The treatment is throughly based on Lagrangian formulation.

C.1 Linearized shape coordinates

It is not the curvilinear shape coordinates, which were used in the classical work of Wilson and co-workers [1], but instead their *linearized counterparts*. The linearized counterpart $\bar{s}_i$ of the curvilinear shape coordinate s_i is *defined* by [36]

$$\Delta \bar{s}_i \left(\mathbf{d}_1, \mathbf{d}_2, \ldots \right) = \sum_{\alpha=1}^{N} \mathbf{d}_\alpha \cdot \boldsymbol{\xi}_\alpha^{(s_i)} \tag{68}$$

where $\Delta \bar{s}_i = \bar{s}_i - \bar{s}_i^{(e)} = \bar{s}_i - s_i^{(e)}$, and

$$\boldsymbol{\xi}_\alpha^{(s_i)} = \nabla_{\mathbf{x}_\alpha} s_i \big|_e \tag{69}$$

is the value of the vector derivative $\nabla_{\mathbf{x}_\alpha} s_i$ evaluated at the instantaneous moving reference conformation $\mathbf{y}_1 = \mathbf{c}'_1, \mathbf{y}_2 = \mathbf{c}'_2, \ldots$. It must be emphasized that although $\bar{s}_i^{(e)} = s_i^{(e)}$, in general $\bar{s}_i$ is a different coordinate than s_i, and its value need not

be equal to that of s_i in any other conformation besides in the reference conformation. Also, the types of motion that $\bar{s}_i$ and s_i describe are generally different—For example, a change in linearized bond-angle induces also a change in bond lengths (this is often accounted by saying that the linearized bond-angle possesses stretching characteristics), where as changing the corresponding curvilinear bond-angle does not alter the lengths of the bonds spanning the angle. Also, as surprising as it is, the explicit relations $\bar{s}_i = f_i(s_1, s_2, \ldots)$ depend implicitly on the choice of the body-frame $\{\mathbf{u}'_1, \mathbf{u}'_2, \mathbf{u}'_3\}$—Changing the body-frame also generally changes these relations (see Ref. [36]; Incidentally, this is also what Eckart states, although in a slightly different form, in the first page of his article).

C.2 Displacements

The nuclear displacement $\mathbf{d}_\alpha$ is related linearly (around the moving reference position $\mathbf{c}'_\alpha$) to the changes in the linearized shape coordinates. Mathematically,

$$\mathbf{d}_\alpha = \sum_{i=1}^{3N-6} \boldsymbol{\xi}_{s_i}^{(\alpha)} \Delta \bar{s}_i \tag{70}$$

where

$$\boldsymbol{\xi}_{s_i}^{(\alpha)} = \left.\frac{\partial \mathbf{d}_\alpha}{\partial s_i}\right|_e = \left.\frac{\partial \mathbf{y}_\alpha}{\partial s_i}\right|_e = \left.\frac{\partial \mathbf{x}_\alpha}{\partial s_i}\right|_e \tag{71}$$

is the *instantaneous reference value* of the tangent vector associated with the particle α and coordinate s_i at the time t [36]. Notice that the second last equality follows from the fact that $\mathbf{c}'_\alpha$ does not depend on the shape coordinates, and the last equality follows from the fact that the center of mass $\mathbf{X}$ does not depend on the shape coordinates. Notice also that the reference value of a tangent vector is *not* generally equal to its initial value (this rather obvious fact is often blurred in the older presentations, which are restricted to "infinitesimal" rotations and vibrations, and which seem to take it granted that the molecule is initially at the undeformed state).

C.3 Hessian

In order to obtain insight to the effects of changing the body-frame, it is best to start with the*curvilinear* shape coordinates, instead of their linearized counterparts. For one thing, any potential V is always a function of the curvilinear shape coordinates, and it does not depend on rotational coordinates. Hence, the Hessian f_{ij} in curvilinear shape coordinates $\{s_i\}$ is related by its counterpart $\bar{f}_{ij}$ in linearized shape coordinates $\{\bar{s}_i\}$ by

$$f_{ij} = \left.\frac{\partial^2 V}{\partial s_i \partial s_j}\right|_e = \left.\sum_{kl}\left(\frac{\partial^2 V}{\partial \bar{s}_k \partial \bar{s}_l}\frac{\partial \bar{s}_k}{\partial s_i}\frac{\partial \bar{s}_l}{\partial s_j} + \frac{\partial V}{\partial \bar{s}_k}\frac{\partial^2 \bar{s}_k}{\partial s_i \partial s_j}\right)\right|_e$$
$$= \left.\sum_{kl}\frac{\partial^2 V}{\partial \bar{s}_k \partial \bar{s}_l}\frac{\partial \bar{s}_k}{\partial s_i}\frac{\partial \bar{s}_l}{\partial s_j}\right|_e = \sum_{kl}\bar{f}_{kl}\frac{\partial \bar{s}_k}{\partial s_i}\frac{\partial \bar{s}_l}{\partial s_j} \tag{72}$$

Because it is true for *any* choice of the body-frame that

$$\left.\frac{\partial \bar{s}_i}{\partial s_j}\right|_e = \delta_{ij} \tag{73}$$

(this is the consequence of the definition in Eq. (68), See Ref. [36], page 044319–8), it follows that

$$f_{ij} = \bar{f}_{ij} \tag{74}$$

i.e., the Hessian in a set of curvilinear shape coordinates is identical with that in their linearized counterparts.

C.4 Metric tensor

Because $\boldsymbol{\xi}_{s_i}^{(\alpha)} \cdot \boldsymbol{\xi}_{s_j}^{(\alpha)}$ are constant, it also follows that the "vibrational" elements of the covariant metric tensor,

$$g_{\bar{s}_i \bar{s}_j} = \sum_{\alpha}^{N} m_\alpha \boldsymbol{\xi}_{s_i}^{(\alpha)} \cdot \boldsymbol{\xi}_{s_j}^{(\alpha)} \tag{75}$$

are constants, when the linearized shape coordinates are utilized. However, the Coriolis elements

$$g_{\bar{s}_i \Theta_j} = \sum_{\alpha}^{N} m_\alpha \boldsymbol{\xi}_{s_i}^{(\alpha)} \cdot \frac{\partial \mathbf{x}_\alpha}{\partial \Theta_j} \tag{76}$$

generally are not (they may depend on the shape coordinates $\{\bar{s}_1, \bar{s}_2, \ldots\}$). As a consequence, the use of Eckart axes eliminate Coriolis terms only in the reference conformation, and not in an arbitrary conformation.

C.5 Christoffel symbol

While the Christoffel symbols

$$\Gamma_{\bar{s}_i \bar{s}_j \bar{s}_k} = \sum_{\alpha}^{N} m_\alpha \left.\frac{\partial \mathbf{y}_\alpha}{\partial s_i}\right|_e \cdot \left[\frac{\partial}{\partial \bar{s}_j}\left(\left.\frac{\partial \mathbf{y}_\alpha}{\partial s_k}\right|_e\right)\right] \tag{77}$$

containing only the vibrational degrees of freedom vanish identically, those

$$\Gamma_{\bar{s}_i \Theta_j \bar{s}_k} = \sum_{\alpha}^{N} m_\alpha \left. \frac{\partial \mathbf{y}_\alpha}{\partial s_i} \right|_e \cdot \left[\frac{\partial}{\partial \Theta_j} \left(\left. \frac{\partial \mathbf{y}_\alpha}{\partial s_k} \right|_e \right) \right] \tag{78}$$

which connect two vibrational, and one rotational degrees of freedom do not necessarily vanish.

C.6 Normal mode calculation

The eigenvalue problem is again that of Eq. (17). As we have seen, the Hessian is the same as in the corresponding curvilinear case. The *reference value* of the metric tensor g associated with the three orientational angles $\{\Theta_i\}$ and the active linearized shape coordinates is also the same as that in the curvilinear case. Hence, the eigenvalues (frequencies) and eigenvectors produced by the linear and curvilinear NMA calculation are identical.

Also, as in the curvilinear case, the total momentum $\mathbf{P}$ is conserved, but the total energy E and the total angular momentum $\mathbf{L}$ are not (now the variable Coriolis elements cause this non-constancy).

D Numerical evaluation of trajectories and displacements

In order to evaluate trajectories, we must be able to numerically evaluate the nuclear position $\mathbf{x}_\alpha(t_n)$ at the time $t = t_n$. As a first step, the values of $\{\Theta_1(t_n), \Theta_2(t_n), \Theta_3(t_n)\}$ and $\{s_i(t_n)\}$ are calculated from the solution of the curvilinear NMA calculation.

D.1 Displacements at the time $t = t_n$, curvilinear NMA

The atomic trajectories $\{\mathbf{x}_\alpha(t_n)\}$ at the time $t = t_n$ are calculated through the recursive method previously developed in Refs. [34] (algorithm 2 in that reference). The internal position is given by $\mathbf{y}_\alpha = \mathbf{x}_\alpha - \mathbf{X}$. The value of the reference positions $\left\{\mathbf{c}'_\alpha(t_n)\right\}$ at the time $t = t_n$ can be calculated using the same procedure by setting $s_i = s_i^{(e)}$ and $\{\Theta_1, \Theta_2, \Theta_3\} = \{\Theta_1(t_n), \Theta_2(t_n), \Theta_3(t_n)\}$, i.e.,

$$\mathbf{c}'_\alpha(t_n) = \mathbf{y}_\alpha\left(\left\{s_i^{(e)}\right\}, \Theta_1(t_n), \Theta_2(t_n), \Theta_3(t_n)\right) \tag{79}$$

Once the nuclear positions $\{\mathbf{y}_\alpha\}_n$ and their reference positions $\left\{\mathbf{c}'_\alpha\right\}_n$ have been calculated at the time $t = t_n$, the displacements are given by

$$\mathbf{d}_\alpha(t_n) = \mathbf{y}_\alpha(t_n) - \mathbf{c}'_\alpha(t_n) \tag{80}$$

D.2 Displacements at the time $t = t_n$, linear NMA

In the linear NMA the displacements are given by Eq. (58). In order to utilize that equation, the reference value of the tangents at the time $t = t_n$ need to be evaluated. They are obtained easily by setting $s_i = s_i^{(e)}$ and $\{\Theta_1, \Theta_2, \Theta_3\} = \{\Theta_1(t_n), \Theta_2(t_n), \Theta_3(t_n)\}$ to Eq. (9) in Ref. [35], i.e.,

$$\boldsymbol{\xi}_{s_i}^{(\alpha)}(t_n) = \frac{\partial \mathbf{x}_\alpha}{\partial s_i}\left(\left\{s_i^{(e)}\right\}, \Theta_1(t_n), \Theta_2(t_n), \Theta_3(t_n)\right) \tag{81}$$

References

1. E.B. Wilson, J.C. Decius, P.C. Cross, *Molecular Vibrations* (Dover, New York, 1980)
2. G. Sørensen, A new approach to the hamiltonian of nonrigid molecules, in *Large Amplitude Motion in Molecules II. Topics in Current Chemistry*, vol. 82, ed. by F. L. Boschke (Springer, Berlin, 1979), pp. 97–175
3. Q. Cui, I. Bahar (eds.), *Normal Mode Analysis—Theory and Practice to Biological and Chemical Systems* (Chapman & Hall, CRC, London, 2006)
4. L. Skjaerven, S.M. Hollup, N. Reuter, Normal mode analysis for proteins. J. Mol. Struct. THEOCHEM. **898**(1–3), 42–48 (2009)
5. E.C. Dykeman, O.F. Sankey, Normal mode analysis and applications in biological physics. J. Phys. Condens. Matter. **22**(42), 423202 (2010)
6. I. Bahar et al., Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins. Chem. Rev. **110**(3), 1463–1497 (2010)
7. J.R. Lopez-Blanco, J.I. Garzon, P. Chacon, iMod: multipurpose normal mode analysis in internal coordinates. Bioinformatics **27**(20), 2843–2850 (2011)
8. A. Kitao, S. Hayward, N. Go, Comparison of normal mode analyses on a small globular protein in dihedral angle space and cartesian coordinate space. Biophys. Chem. **52**(2), 107–114 (1994)
9. T. Noguti, N. Go, Dynamics of native globular proteins in terms of dihedral angles. J. Phys. Soc. Jpn. **52**, 3283–3288 (1983)
10. N. Go, T. Noguti, T. Nishikawa, Dynamics of a small globular protein in terms of low-frequency vibrational modes. PNAS **80**, 3696–3700 (1983)
11. M. Levitt, C. Sander, P.S. Stern, Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. **181**, 423–447 (1985)
12. M. Tirion, Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys. Rev. Lett. **77**, 1905–1908 (1996)
13. F. Tama, Y.-H. Sanejouand, Conformational change of proteins arising from normal mode calculations. Protein Eng. **14**(1), 1–6 (2001)
14. F. Tama, W. Wriggers, C.L. Brooks III, Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J. Mol. Biol. **321**, 297–305 (2002)
15. J.A. Kovacs, P. Chacon, R. Abagyan, Predictions of protein flexibility: first-order measures. Proteins **56**, 661–668 (2004)
16. H.W.T. van Vlijmen, M. Karplus, Normal mode calculations of icosahedral viruses with full dihedral flexibility by use of molecular symmetry. J. Mol. Biol. **350**, 528–542 (2005)
17. M. Rueda, P. Chacon, M. Orozco, Thorough validation of protein normal mode analysis: a comparative study with essential dynamcis. Structure **15**, 565–575 (2007)
18. S. Fuchigami, S. Omori, M. Ikeguchi, A. Kidera, Normal mode analysis of protein dynamics in a non-eckart frame. J. Chem. Phys. **132**(10), 104109 (2010)
19. C. Eckart, Some studies concerning rotating axes and polyatomic molecules. Phys. Rev. **47**, 552–558 (1935)
20. R.J. Malhiot, S.M. Ferigle, Eckart conditions in Wilson's treatment of molecular vibrations. J. Chem. Phys. **22**(4), 717–719 (1954)

21. J.D. Louck, H.W. Galbraith, Eckart vectors, Eckart frames, and polyatomic molecules. Rev. Mod. Phys. **48**(1), 69 (1976)
22. R.G. Littlejohn, M. Reinsch, Gauge fields in the separation of rotations and internal motions in the n-body problem. Rev. Mod. Phys. **69**(1), 213–276 (1997)
23. R.G. Littlejohn, K. Mitchell, Gauge theory of small vibrations in polyatomic molecules, in *Geometry, Mechanics, and Dynamics*, ed. by P. Newton, P. Holmes, A. Weinstein (Springer, New York, 2002), pp. 407–428
24. A.V. Meremianin, Body frames in the separation of collective angles in quantum n-body problems. J. Chem. Phys. **120**(17), 7861–7876 (2004)
25. T.M. Yanao, K. Takatsuka, Kinematic effects associated with molecular frames in structural isomerization dynamics of clusters. J. Chem. Phys. **120**(19), 8924–8936 (2004)
26. S. Califano, *Vibrational States* (Wiley, London, 1976)
27. H. Goldstein, C. Poole, J. Safko, *Classical Mechanics*, 3rd edn. (Addison Wesley, San Francisco, 2002)
28. D.C. Kay, *Tensor Calculus (Schaum's outline series)* (McGraw-Hill, New York, 1988)
29. T. Noguti, N. Go, A method of rapid calculation of a second derivative matrix of conformational energy for large molecules. J. Phys. Soc. Jpn. **52**(10), 3685–3690 (1983)
30. W. Braun, S. Yoshioki, N. Go, Formulation of static and dynamic conformational energy analysis of biopolymer systems consisting of two or more molecules. J. Phys. Soc. Jpn. **53**(9), 3269 (1984)
31. D. Hestenes, *New Foundations for Classical Mechanics*, 2nd edn. (Kluwer Academic Publishers, Dordrecht, 1999)
32. H. Wei, Eckart frames for planar molecules. J. Chem. Phys. **118**, 7202 (2003)
33. K.N. Kudin, A.Y. Dymarsky, Eckart axis conditions and the minimization of the root-mean-square deviation: Two closely related problems. J. Chem. Phys. **122**, 224105 (2005)
34. J. Pesonen, K.O.E. Henriksson, Polymer conformations in internal (polyspherical) coordinates. J. Comput. Chem. **31**(9), 1873–1881 (2010)
35. K.O.E. Henriksson, J. Pesonen, Polymer dynamics in torsion space. J. Comput. Chem. **31**(9), 1882–1888 (2010)
36. J. Pesonen, Kinetic energy operators in linearized internal coordinates. J. Chem. Phys. **128**(4), 044319 (2008)