Some studies on generalized coordinate sets for polyatomic molecules

Wenjin Li^{a)} and Ao Ma^{a)}

Department of Bioengineering, The University of Illinois at Chicago, 851 South Morgan Street, Chicago, Illinois 60607, USA

(Received 6 August 2015; accepted 17 November 2015; published online 8 December 2015)

Generalized coordinates are widely used in various analyses of the trajectories of polyatomic molecules from molecular dynamics simulations, such as normal mode analysis and force distribution analysis. Here, we presented detailed discussions on the properties of some specific sets of generalized coordinates, which separate translational, rotational, and vibrational motions of a molecule from one another once the trajectories of dynamical systems are known. Efficient methods were suggested for estimating the transformation matrix between generalized and Cartesian coordinates. Some properties of the well-known BAT coordinates (bond length, angle, and torsional coordinates) were discussed as well. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4936773>]

I. INTRODUCTION

The application of generalized coordinates often provides valuable insights into various analyses of trajectories from molecular dynamics simulations of polyatomic molecules, such as normal mode analysis^{1–4} and force distribution analysis.^{5–8} From practical concerns, particular attention has been paid to generalized coordinate sets that consist of external coordinates and internal coordinates. External coordinates describe the overall translation and rotation of a system, whereas internal coordinates describe its internal or vibrational motion. It would be advantageous in applications to have external and internal coordinates separable from one another. However, this is not a trivial task due to the intrinsic coupling between the rotational motion and vibrational motion for non-rigid structures.

Conventionally, this separation is achieved under the Eckart condition,⁹ which minimizes the Coriolis term—the coupling between external rotations and internal vibrations in the molecular kinetic energy. Jellinek and Li introduced an angular velocity with respect to the center of mass for non-rigid systems, the angular velocity with which the non-rigid systems would rotate if it were a true rigid body, to eliminate the Coriolis term and thus exactly separate the rotational kinetic energy from the total kinetic energy.¹⁰ They mainly focused on systems with constant angular momentum and successfully applied their theory to rapidly rotating systems.¹¹ However, the theory is not applicable to systems with non-conserved angular momentum and the missing analysis on a complete set of coordinates further limited its applicability. A commonly adopted coordinate for describing rotational motion is Euler angles, which were originally developed for rigid body rotations.¹² For non-rigid systems, Euler angles with respect to the molecule-fixed frame (MFF) were defined recently. Although such Euler angles have been successfully used

in internal coordinate molecular dynamics simulations,^{13,14} their inclusion in a complete set of generalized coordinates introduces coupling between external and internal motions.

Here, we define Euler angles for non-rigid bodies with respect to the center of mass frame, which is a moving coordinate frame or an instantaneous reference configuration.³ To distinguish them from other definitions of Euler angles for non-rigid bodies, we name them quasi-Euler angles. In this work, we demonstrate that quasi-Euler angles in the center of mass frame and the Cartesian coordinates of the center of mass itself form the six external coordinates that are completely separable from internal coordinates for polyatomic molecules in the analysis of their dynamical trajectories. Together with any set of internal coordinates, they form a complete set of generalized coordinates. A well-known example of internal coordinates is the so called BAT coordinates that consist of bond length, angle, and torsional coordinates.^{15,16} BAT coordinates are more natural than Cartesian coordinates for describing the dynamics of polyatomic molecules, as they have already taken into account the constraints imposed by the chemical bonding connectivity. Thus, it will be of broad interest to present some properties on the generalized forces of BAT coordinates.

In Section II, we first prove the separability between the proposed external coordinates and internal coordinates. Then, we discuss the computational aspects of the transformation between generalized coordinates and Cartesian coordinates in Section III. Later, some properties of a particular set of BAT coordinates will be shown in Section IV.

II. SEPARATION OF EXTERNAL COORDINATES FROM INTERNAL COORDINATES

A. Basics

For the purpose of a self-contained presentation in this paper, we lay out some basic facts on generalized coordinates and Lagrangian dynamics.

^{a)}Authors to whom correspondence should be addressed. Electronic addresses: wenjin@uic.edu and aoma@uic.edu

1. Coordinate transformation matrices

For an unconstrained system with N atoms, let a column vector $\mathbf{r} \equiv [r_1^T r_2^T \cdots r_N^T]^T$ be the $3N$ Cartesian coordinates, with $\mathbf{r}_i \equiv [r_{ix} r_{iy} r_{iz}]^T$ the three Cartesian coordinates of atom i . For the simplification of notation, we also use $\mathbf{x} \equiv [x_1 x_2 \cdots x_{3N}]^T$ to represent the $3N$ Cartesian coordinates and $\mathbf{x} = \mathbf{r}$. The superscript $\top$ stands for the transpose of a vector or matrix and we use column vectors throughout the paper. For any set of complete and independent generalized coordinates $\xi \equiv [\xi_1 \xi_2 \cdots \xi_{3N}]^T$, the transformation between infinitesimal changes in Cartesian coordinates ($\delta\mathbf{x}$) and generalized coordinates ($\delta\xi$) is given by

$$\delta\mathbf{x} = \mathbf{A}\delta\xi, \quad (1a)$$

$$\delta\xi = \mathbf{B}\delta\mathbf{x}, \quad (1b)$$

where matrices $\mathbf{A}$ and $\mathbf{B}$ are the so-called A-matrix and B-matrix, respectively. The elements of $\mathbf{A}$ and $\mathbf{B}$ are $A_{ij} \equiv \partial x_i / \partial \xi_j$ and $B_{ij} \equiv \partial \xi_i / \partial x_j$. It immediately follows that the time derivatives of Cartesian coordinates ($\dot{\mathbf{x}}$) and generalized coordinates ($\dot{\xi}$) are related to each other as

$$\dot{\mathbf{x}} = \mathbf{A}\dot{\xi}, \quad (2a)$$

$$\dot{\xi} = \mathbf{B}\dot{\mathbf{x}}, \quad (2b)$$

Since $\mathbf{x}$ and ξ are both complete and independent coordinates, matrices $\mathbf{A}$ and $\mathbf{B}$ are of full rank and the relation below holds,

$$\mathbf{A}\mathbf{B} = \mathbf{B}\mathbf{A} = \mathbf{I}_{3N}, \quad (3)$$

where $\mathbf{I}_m$ stands for an $m \times m$ identity matrix. Assuming the first 6 coordinates in ξ are external coordinates and the rest are internal coordinates, that is $\xi \equiv [e^T q^T]^T$, with $e \equiv [e_1 e_2 \cdots e_6]^T$ the external coordinates and $q \equiv [q_1 q_2 \cdots q_{3N-6}]^T$ the internal coordinates, we can define $\mathbf{A} \equiv [\mathbf{A}_1 \mathbf{A}_2]$, where $\mathbf{A}_1$ is the $3N \times 6$ matrix that transforms changes in the six external coordinates to changes in the $3N$ Cartesian coordinates and $\mathbf{A}_2$ is the $3N \times (3N - 6)$ matrix that transforms changes in the $3N - 6$ internal coordinates to changes in the $3N$ Cartesian coordinates. Then, changes in Cartesian coordinates caused by external coordinates ($\delta\mathbf{x}_1$) and by internal coordinates ($\delta\mathbf{x}_2$) are given by

$$\delta\mathbf{x}_1 = \mathbf{A}_1\delta e, \quad (4a)$$

$$\delta\mathbf{x}_2 = \mathbf{A}_2\delta q, \quad (4b)$$

with $\delta\mathbf{x} = \delta\mathbf{x}_1 + \delta\mathbf{x}_2$. Similarly, we can define $\mathbf{B} \equiv [\mathbf{B}_1^T \mathbf{B}_2^T]^T$, where $\mathbf{B}_1$ is the $6 \times 3N$ matrix that converts changes in Cartesian coordinates to changes in external coordinates and $\mathbf{B}_2$ is the $(3N - 6) \times 3N$ matrix that converts changes in Cartesian coordinates to changes in internal coordinates,

$$\delta e = \mathbf{B}_1\delta\mathbf{x}, \quad (5a)$$

$$\delta q = \mathbf{B}_2\delta\mathbf{x}. \quad (5b)$$

From Eq. (3), we can immediately obtain the following relations:

$$\mathbf{B}_1\mathbf{A}_1 = \mathbf{I}_6, \quad (6a)$$

$$\mathbf{B}_1\mathbf{A}_2 = \mathbf{0}_{6,3N-6}, \quad (6b)$$

$$\mathbf{B}_2\mathbf{A}_1 = \mathbf{0}_{3N-6,6}, \quad (6c)$$

$$\mathbf{B}_2\mathbf{A}_2 = \mathbf{I}_{3N-6}, \quad (6d)$$

$$\mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_2 = \mathbf{I}_{3N}, \quad (6e)$$

here, $\mathbf{0}_{m,n}$ stands for an $m \times n$ zero matrix.

2. Kinetic energy in generalized coordinates

Now the kinetic energy of the system in Cartesian coordinates (T) can be expressed in generalized coordinates as

$$2T = \dot{\mathbf{x}}^T \mathbf{m} \dot{\mathbf{x}} = \dot{\xi}^T \mathbf{A}^T \mathbf{m} \mathbf{A} \dot{\xi} \quad (7a)$$

$$= \begin{bmatrix} \dot{e}^T & \dot{q}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_1^T \\ \mathbf{A}_2^T \end{bmatrix} \mathbf{m} \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \dot{e} \\ \dot{q} \end{bmatrix} \\ = \begin{bmatrix} \dot{e}^T & \dot{q}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_1^T \mathbf{m} \mathbf{A}_1 & \mathbf{A}_1^T \mathbf{m} \mathbf{A}_2 \\ \mathbf{A}_2^T \mathbf{m} \mathbf{A}_1 & \mathbf{A}_2^T \mathbf{m} \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \dot{e} \\ \dot{q} \end{bmatrix} \\ = \dot{e}^T \mathbf{A}_1^T \mathbf{m} \mathbf{A}_1 \dot{e} + 2\dot{e}^T \mathbf{A}_1^T \mathbf{m} \mathbf{A}_2 \dot{q} + \dot{q}^T \mathbf{A}_2^T \mathbf{m} \mathbf{A}_2 \dot{q}, \quad (7b)$$

where $\mathbf{m}$ is a $3N \times 3N$ diagonal matrix whose diagonal elements are the masses of the corresponding atoms. Assuming that $\dot{e}$ and $\dot{q}$ are independent and they separate the external part of the kinetic energy from the internal part, the Coriolis term in Eq. (7b), $\dot{e}^T \mathbf{A}_1^T \mathbf{m} \mathbf{A}_2 \dot{q}$, must vanish. Thus, we require,

$$\mathbf{A}_1^T \mathbf{m} \mathbf{A}_2 = \mathbf{0}_{6,3N-6}. \quad (8)$$

In this way, the total kinetic energy can be completely separated into an external part (T_{ext}) and an internal part (T_{int}), with

$$T = T_{\text{ext}} + T_{\text{int}}, \quad (9a)$$

$$T_{\text{ext}} = \frac{1}{2} \dot{e}^T \mathbf{A}_1^T \mathbf{m} \mathbf{A}_1 \dot{e}, \quad (9a)$$

$$T_{\text{int}} = \frac{1}{2} \dot{q}^T \mathbf{A}_2^T \mathbf{m} \mathbf{A}_2 \dot{q}. \quad (9b)$$

3. Generalized momentum and generalized force

In the Lagrangian formulation of dynamics,¹² generalized momentum $\mathbf{p}$ is defined as

$$\mathbf{p} \equiv \partial L(\xi, \dot{\xi}) / \partial \dot{\xi}, \quad (10)$$

where $L = T - V$ is the Lagrangian of the system, with V the potential energy of the system. Since we are only interested in cases where V is independent of $\dot{\xi}$, with Eq. (7a), we have

$$\mathbf{p} = \partial T(\xi, \dot{\xi}) / \partial \dot{\xi} = \frac{1}{2} \partial (\dot{\xi}^T \mathbf{A}^T \mathbf{m} \mathbf{A} \dot{\xi}) / \partial \dot{\xi} \\ = \mathbf{A}^T \mathbf{m} \mathbf{A} \dot{\xi} = \mathbf{A}^T \mathbf{m} \dot{\mathbf{x}}. \quad (11)$$

On the other hand, generalized force $\mathbf{Q}$ is given by

$$\mathbf{Q} \equiv -\partial V / \partial \xi = \mathbf{A}^T (-\partial V / \partial \mathbf{x}) \\ = \mathbf{A}^T \mathbf{F}, \quad (12)$$

where $\mathbf{F} \equiv -\partial V/\partial \mathbf{x}$ is the force vector in Cartesian coordinates.

B. External coordinates

In this section, we discuss the choice of a particular set of external coordinates that make the total kinetic energy of a system completely separated into an external and an internal part, as well as the complete separation of the kinetic energies of overall translations, rotations, and vibrations from each other.

1. The choice of external coordinates

The linear momentum $\mathbf{P}_c$ of and the angular momentum $\mathbf{L}_c$ with respect to the center of mass (COM) are conserved if the total force and the overall torque on the system are zero. We thus consider the linear momentum and angular momentum as the generalized momentum conjugate to the external coordinates. Note that

$$\mathbf{P}_c = \sum_{i=1}^N m_i \dot{\mathbf{r}}_i, \quad (13a)$$

$$\mathbf{L}_c = \sum_{i=1}^N m_i \mathbf{y}_i \times \dot{\mathbf{r}}_i, \quad (13b)$$

where $\dot{\mathbf{r}}_i$ is the velocity of atom i in Cartesian coordinates, m_i is the mass of atom i , $\mathbf{y}_i \equiv \mathbf{r}_i - \mathbf{c}$ is the coordinate of atom i in the center of mass frame (CMF). From Eq. (11), we get immediately the external part of A-matrix $\mathbf{A}_1 \equiv [\mathbf{A}_1^c \mathbf{A}_1^\theta]$, with

$$\mathbf{A}_1^c = [\mathbf{I}_3 \quad \mathbf{I}_3 \quad \cdots \quad \mathbf{I}_3]^\top, \quad (14a)$$

$$\mathbf{A}_1^\theta = [\mathbb{S}[\mathbf{y}_1] \quad \mathbb{S}[\mathbf{y}_2] \quad \cdots \quad \mathbb{S}[\mathbf{y}_N]]^\top, \quad (14b)$$

here, $\mathbf{A}_1^c$ and $\mathbf{A}_1^\theta$ are the translational and rotational part of matrix $\mathbf{A}_1$. The operator $\mathbb{S}$ transforms a vector $\mathbf{a} = [a_1 \ a_2 \ a_3]$ into a skew-symmetric matrix and is defined as

$$\mathbb{S}[\mathbf{a}] \equiv \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}. \quad (15)$$

From Eq. (12), one immediately obtains the generalized forces on the coordinates of the overall translations and rotations $\mathbf{F}_{\text{tot}} = \sum_{i=1}^N \mathbf{F}_i$ and $\boldsymbol{\tau} = \sum_{i=1}^N \mathbf{y}_i \times \mathbf{F}_i$, which are the total force and torque of the system.

Correspondingly, the external coordinates can be chosen as the coordinates conjugate to the linear momentum and angular momentum, which are the center of mass ($\mathbf{c}$) and the corresponding Euler angle ($\boldsymbol{\theta}$), respectively. For non-rigid bodies, the Euler angles are replaced by the quasi-Euler angles. From classical mechanics, we know

$$\dot{\mathbf{c}} = \frac{1}{M} \sum_{i=1}^N m_i \dot{\mathbf{r}}_i, \quad (16a)$$

$$\dot{\boldsymbol{\theta}} = \mathbf{I}^{-1} \mathbf{L}_c = \mathbf{I}^{-1} \sum_{i=1}^N m_i \mathbf{y}_i \times \dot{\mathbf{r}}_i, \quad (16b)$$

where $\dot{\mathbf{c}}$ and $\dot{\boldsymbol{\theta}}$ are the velocity of COM and the angular velocity with respect to COM, respectively, M is the total mass of the system, and $\mathbf{I} \equiv \sum_{i=1}^N m_i((\mathbf{y}_i \cdot \mathbf{y}_i)\mathbf{I}_3 - \mathbf{y}_i \mathbf{y}_i^\top)$ is the moment of inertia tensor in CMF. We assume that $\mathbf{I}$ is invertible as is the case for most polyatomic molecules. From Eqs. (16) and (2b), we immediately obtain the external part of B-matrix $\mathbf{B}_1 \equiv [\mathbf{B}_1^c \mathbf{B}_1^\theta]^\top$,

$$\mathbf{B}_1^c = \frac{1}{M} [\mathbf{m}_1 \mathbf{I}_3 \quad \mathbf{m}_2 \mathbf{I}_3 \quad \cdots \quad \mathbf{m}_N \mathbf{I}_3], \quad (17a)$$

$$\mathbf{B}_1^\theta = \mathbf{I}^{-1} [\mathbf{m}_1 \mathbb{S}[\mathbf{y}_1] \quad \mathbf{m}_2 \mathbb{S}[\mathbf{y}_2] \quad \cdots \quad \mathbf{m}_N \mathbb{S}[\mathbf{y}_N]], \quad (17b)$$

where $\mathbf{B}_1^c$ and $\mathbf{B}_1^\theta$ are the translational and rotational part of the matrix $\mathbf{B}_1$. By substituting in the explicit expressions of $\mathbf{A}_1$ and $\mathbf{B}_1$, one can see that they satisfy Eq. (6a).

2. Exact separation of translation, rotational, and internal kinetic energies

To have the kinetic energy associated with external and internal coordinates exactly separable to each other, the Coriolis term must vanish (Eq. (8)), which we prove below:

$$\mathbf{A}_1^\top \mathbf{m} \mathbf{A}_2 = \begin{bmatrix} \mathbf{A}_1^{c\top} \\ \mathbf{A}_1^{\theta\top} \end{bmatrix} \mathbf{m} \mathbf{A}_2 = \begin{bmatrix} \mathbf{A}_1^{c\top} \mathbf{m} \\ \mathbf{A}_1^{\theta\top} \mathbf{m} \end{bmatrix} \mathbf{A}_2 \quad (18a)$$

$$= \begin{bmatrix} \mathbf{M} \mathbf{B}_1^c \\ \mathbf{I} \mathbf{B}_1^\theta \end{bmatrix} \mathbf{A}_2 = \begin{bmatrix} \mathbf{M} \mathbf{B}_1^c \mathbf{A}_2 \\ \mathbf{I} \mathbf{B}_1^\theta \mathbf{A}_2 \end{bmatrix}. \quad (18b)$$

From Eq. (6b), it is evident that $\mathbf{B}_1^c \mathbf{A}_2 = \mathbf{0}_{3,3N-6}$ and $\mathbf{B}_1^\theta \mathbf{A}_2 = \mathbf{0}_{3,3N-6}$. Therefore, Eq. (8) is satisfied, and the kinetic energy is exactly separated into external and internal parts. Note that such a relation holds for polyatomic systems under non-vanishing external forces as well. In addition, any six external coordinates, whose infinitesimal changes can be converted to changes of the six external coordinates $\mathbf{e}$ by an invertible 6×6 matrix, can separate the external kinetic energy from the internal part as well.

In addition, one may notice that $\mathbf{A}_1^{c\top} \mathbf{m} \mathbf{A}_2 = \mathbf{0}_{3,3N-6}$, $\mathbf{A}_1^{\theta\top} \mathbf{m} \mathbf{A}_2 = \mathbf{0}_{3,3N-6}$, and $\mathbf{A}_1^{c\top} \mathbf{m} \mathbf{A}_1^\theta = \mathbf{0}_{3,3}$, which means all the coupling terms among translational, rotational and internal parts of the kinetic energy vanish—therefore, they are all exactly separable from one another.

Consequently, we can further separate the external kinetic energy T_{ext} in Eq. (9a) into translational (T_{ext}^c) and rotational (T_{ext}^θ) parts,

$$T_{\text{ext}}^c = \frac{1}{2} \dot{\mathbf{c}}^\top \mathbf{A}_1^{c\top} \mathbf{m} \mathbf{A}_1^c \dot{\mathbf{c}}, \quad (19a)$$

$$T_{\text{ext}}^\theta = \frac{1}{2} \dot{\boldsymbol{\theta}}^\top \mathbf{A}_1^{\theta\top} \mathbf{m} \mathbf{A}_1^\theta \dot{\boldsymbol{\theta}}. \quad (19b)$$

After taking into account of Eq. (14), Eq. (19) can be rewritten as,

$$T_{\text{ext}}^c = \frac{1}{2} M \dot{\mathbf{c}} \cdot \dot{\mathbf{c}}, \quad (20a)$$

$$T_{\text{ext}}^\theta = \frac{1}{2} \dot{\boldsymbol{\theta}}^\top \mathbf{I} \dot{\boldsymbol{\theta}}, \quad (20b)$$

thus, T_{ext}^c and T_{ext}^θ share the same forms as the kinetic energy of the overall translation and rotation of a rigid body, and the kinetic energy for rotation in Ref. 10 (Eq. (6)) is reproduced.

III. INTERNAL COORDINATES

After achieving the complete separation between the external and the internal coordinates, we can now focus on transformation properties of internal coordinates. Although Cartesian coordinates have disadvantages in describing dynamics of complex molecules, their analytical simplicity in expressing mechanical and dynamic quantities accounts for their prevalent use. Thus, it is of great interest to discuss the computational aspects of transformations between Cartesian and generalized coordinates.

One common choice of internal coordinates is the BAT coordinates, which consists of $N - 1$ bond lengths, $N - 2$ bond angles, and $N - 3$ torsional angles. Here, we focus on the internal part of the A-matrix and B-matrix when BAT coordinates are applied.

A. B-matrix

The internal part of the B-matrix $\mathbf{B}_2$ for BAT coordinates can be found in standard textbooks of molecular spectroscopy¹⁷ and is briefly summarized in [Appendix A](#). $\mathbf{B}_2$ is a very sparse matrix. One may notice that the non-zero elements of $\mathbf{B}_2$ satisfy the following relations ($\mathbf{0}_m$ is a zero vector with m elements, see [Appendix A](#) for the explanations of other notations):

$$\begin{aligned} \frac{\partial q_{\text{bon}}}{\partial \mathbf{r}_i} + \frac{\partial q_{\text{bon}}}{\partial \mathbf{r}_j} &= \mathbf{0}_3, \\ \frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_i} + \frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_j} + \frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_k} &= \mathbf{0}_3, \\ \frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_i} + \frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j} + \frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k} + \frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_l} &= \mathbf{0}_3, \\ \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j}\right)_1 + \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k}\right)_2 &= -\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_i}, \\ \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j}\right)_2 + \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k}\right)_1 &= -\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_l}. \end{aligned} \quad (21)$$

These relations can be used for efficient computation of $\mathbf{B}_2$. As the explicit forms of $\mathbf{A}_1$ and $\mathbf{B}_2$ are given, one can check that they satisfy Eq. (6c). Now we give explicitly a complete set of generalized coordinates $\boldsymbol{\xi} \equiv [\mathbf{c}^\top \boldsymbol{\theta}^\top \mathbf{q}^\top]^\top$, which consists of three orthogonal coordinate sets, i.e., the translational, rotational, and vibrational coordinates.

B. A-matrix

In most practical contexts, B-matrix has analytical expression and can be easily evaluated, whereas this is not always the case for the A-matrix. Instead, the A-matrix might need to be calculated by inverting B-matrix and the internal part of A-matrix $\mathbf{A}_2$ is then obtained. Note that $\mathbf{A}_2$ generally does not equal to the Moore-Penrose pseudoinverse¹⁸ of $\mathbf{B}_2$ ($\mathbf{B}_2^+ = \mathbf{B}_2^\top (\mathbf{B}_2 \mathbf{B}_2^\top)^{-1}$), because $\mathbf{A}_2 \mathbf{B}_2$ is usually not a symmetric matrix. One can easily see that $\mathbf{A}_1 \mathbf{B}_1$ is a symmetric matrix only when the masses of all atoms are equal. From Eq. (6c), we have $\mathbf{A}_2 \mathbf{B}_2 = \mathbf{I}_{3N} - \mathbf{A}_1 \mathbf{B}_1$; thus, $(\mathbf{A}_2 \mathbf{B}_2)^\top = \mathbf{A}_2 \mathbf{B}_2$ holds only when all the atoms have the same mass and $\mathbf{A}_2$ equals to the pseudoinverse of $\mathbf{B}_2$ only under this condition.

For large systems, the inversion of B-matrix may suffer from numerical instability. In the following, we present a method that avoids inverting a large matrix and requires inversion of a 6×6 matrix only. We assume six “external” coordinates $\boldsymbol{\pi}$ that can be constructed as a function of all the generalized coordinates $\boldsymbol{\xi}$, that is,

$$\delta \boldsymbol{\pi} = \mathbf{E}_1 \delta \boldsymbol{\xi}, \quad (22)$$

where $\mathbf{E}_1$ is a $6 \times 3N$ matrix, with $(E_1)_{ij} \equiv \partial \pi_i / \partial \xi_j$. By construction, we require that infinitesimal changes of $\boldsymbol{\pi}$ cannot be converted to the change of $\mathbf{e}$ by an invertible 6×6 matrix and that $\boldsymbol{\pi}$ together with $\mathbf{q}$ forms a complete set of generalized coordinates. Since $\boldsymbol{\pi}$ contains internal components, we call it semi-external coordinates. We define the complete set of generalized coordinates consisting of $\boldsymbol{\pi}$ and $\mathbf{q}$ as a new set of generalized coordinates $\boldsymbol{\zeta} \equiv [\boldsymbol{\pi}^\top \mathbf{q}^\top]^\top$, where $\mathbf{q}$ is the same BAT coordinates as the internal coordinates in $\boldsymbol{\xi}$. Let matrices $\mathbf{D}$, with $D_{ij} \equiv \partial \xi_i / \partial \zeta_j$ and $\mathbf{E}$ with $E_{ij} \equiv \partial \zeta_i / \partial \xi_j$ are the transform matrix between $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$, we have

$$\delta \boldsymbol{\xi} = \mathbf{D} \delta \boldsymbol{\zeta}, \quad (23a)$$

$$\delta \boldsymbol{\zeta} = \mathbf{E} \delta \boldsymbol{\xi}, \quad (23b)$$

$$\text{with } \mathbf{D} \equiv \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{bmatrix}, \quad (23c)$$

$$\mathbf{E} \equiv \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix}. \quad (23d)$$

Since $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ contain the same internal coordinates by design, from the definition of matrices $\mathbf{D}$ and $\mathbf{E}$, we know that $\mathbf{D}_{21}$ and $\mathbf{E}_{21}$ are $(3N - 6) \times 6$ zero matrices, $\mathbf{D}_{22}$ and $\mathbf{E}_{22}$ are $(3N - 6) \times (3N - 6)$ identity matrices. In addition, $\mathbf{D}_{11}$ and $\mathbf{E}_{11}$ are invertible 6×6 matrices, $\mathbf{D}_{12}$ and $\mathbf{E}_{12}$ are non-zero $6 \times (3N - 6)$ matrices. Thus, we have

$$\mathbf{D} \equiv \begin{bmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{0}_{3N-6,6} & \mathbf{I}_{3N-6,3N-6} \end{bmatrix}, \quad (24a)$$

$$\mathbf{E} \equiv \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{0}_{3N-6,6} & \mathbf{I}_{3N-6,3N-6} \end{bmatrix}. \quad (24b)$$

From $\mathbf{DE} = \mathbf{I}_{3N}$, we immediately have

$$\mathbf{E}_{11} = \mathbf{D}_{11}^{-1}, \quad (25a)$$

$$\mathbf{E}_{12} = -\mathbf{D}_{11}^{-1} \mathbf{D}_{12}. \quad (25b)$$

Define $\mathbf{A}^\zeta$, with $A_{ij}^\zeta \equiv \partial x_i / \partial \zeta_j$ as the matrix that transforms the infinitesimal change of $\boldsymbol{\zeta}$ to the change of Cartesian coordinates $\mathbf{x}$, we have

$$\delta \mathbf{x} = \mathbf{A}^\zeta \delta \boldsymbol{\zeta}. \quad (26)$$

In many cases, as one of them will be shown in Sec. IV and many others in the literatures,^{13–15,19,20} an analytic form of $\mathbf{A}^\zeta$ can be given. From Eqs. (26), (23), and (1), we have

$$\mathbf{D} = \mathbf{B} \mathbf{A}^\zeta, \quad (27a)$$

$$\mathbf{A} = \mathbf{A}^\zeta \mathbf{E}. \quad (27b)$$

Let $\mathbf{A}^\zeta \equiv [\mathbf{A}_1^\zeta \mathbf{A}_2^\zeta]$ and $\mathbf{D}_1 \equiv [\mathbf{D}_{11} \mathbf{D}_{12}]$. From the above equations, we have

$$\mathbf{D}_1 = \mathbf{B}_1 \mathbf{A}^\zeta, \quad (28a)$$

$$\mathbf{A}_2 = \mathbf{A}_1^\zeta \mathbf{E}_{12} + \mathbf{A}_2^\zeta. \quad (28b)$$

To summarize the above results, we have

$$\begin{aligned} A_2 &= A_1^\zeta E_{12} + A_2^\zeta, \\ \text{with } E_{12} &= -D_{11}^{-1} D_{12}, \\ D_1 &= B_1 A^\zeta, \end{aligned} \quad (29)$$

thus, we can estimate A_2 without inverting the corresponding B-matrix if the A-matrix of another generalized coordinate sets, which shares the same internal coordinates, can be estimated analytically. We note that only the inversion of a 6×6 matrix is necessary.

IV. SPECIFIC CHOICES OF INTERNAL COORDINATES

For a given molecule, there are many choices of BAT coordinates. For example, molecules often have branches, i.e., an atom forms more than two bonds; thus, the choice for bond angles and dihedral (torsional) angles is not unique. It raises the question of which BAT coordinate set one should choose. Here, we will discuss a specific subset of the BAT coordinate sets that is convenient and has wide usages. The generalized forces on this coordinate set show some interesting properties. We will discuss a molecule without cycle or a system with several molecules which can be constructed as a single tree-like molecule without cycle.

A. BKS-tree and internal coordinates

A non-cyclic system, in which the molecules contain no cyclic structures, can be represented by the so-called BKS-tree, which is a convenient representation for conformational modelling of biomolecular systems.^{19,20} In a system consisting of several molecules, all the molecules can be connected into a single tree-like molecule by virtual bonds. In a BKS-tree, atoms and bonds are represented as nodes and edges, respectively. Here, we give a brief overview for the case of a single non-cyclic molecule and the extension to cases of several molecules is straightforward (see Refs. 19 and 20 for more details of the BKS-tree).

As shown in Fig. 1, the laboratory-fixed coordinate system consists of x , y , and z -axes intersecting at the origin O , which is a virtual atom with fixed position. Y and Z are two virtual atoms on the y and z axes and their positions are also fixed.

FIG. 1. An example of a BKS-tree of a molecule with numeration of a topological tree. A chain 1-2-7-8 consisting of atoms 1, 2, 7, and 8 is highlighted in the dashed red circle. See the main text for other details.

We choose an atom of the molecule that forms only one bond with other atoms (a leaf of the molecule if considering the molecule as a tree) as the base of the tree (or the molecule), then all the other atoms can be indexed. The numeration of the tree begins from the base and always increases along any chain coming from the base^{19,20} and an example is shown in Fig. 1. A virtual bond between the origin O and the base connects the molecule with the axes and the positions of all the atoms in the molecule can be uniquely determined. The position of each atom is uniquely determined by a bond length, a bond angle, and a dihedral angle if the positions of all the atoms prior to it (the atoms with lower indices) are fixed. The position of the base (atom 1) is determined by a bond length between atoms 1 and O , an angle (formed by atoms 1, O , and Z), and a dihedral angle (formed by atoms 1, O , Z , and Y). Here, we assume that there is one virtual bond between O and Z , and another one between Z and Y . After the position of the base is determined, the positions of other nodes can be determined in a similar way: the virtual node Y is now the base of the constructed BKS-tree and the BKS-tree includes the tree-like structure and its numeration.

For each node in the BKS-tree, there is a unique chain of nodes that connects the target node to the base. As shown in Fig. 2(a), for node i , we label the first three nodes in the chain back to the base as j , k , and l , whose positions are already determined. Then the three coordinates that determine the position of node i are the bond length q_i^{bon} , bond angle q_i^{ang} , and dihedral angle q_i^{dih} . Here, q_i^{bon} is the length of the bond between nodes i and j , q_i^{ang} is the angle formed by nodes i , j , and k (the angle between the bond formed by the first two nodes $i-j$ and the bond formed by the last two nodes $j-k$), q_i^{dih} is the dihedral formed by nodes i , j , k , and l (the torsional angle between the plane formed by the first three nodes $i-j-k$ and the plane formed by the last three nodes $j-k-l$). Three unit vectors corresponding to the three coordinates are attributed to atom i but located at atom j . They determine the infinitesimal displacements of the structure due to these coordinates. The

FIG. 2. (a) General representation of a non-cyclic BKS-tree with the chain from node i back to the base highlighted. Black circle and solid lines explicitly represent the nodes and edges of the tree, respectively. The thick dashed-line circle on a node, the dashed line and circle connected with a node represent the connections to the node other than those explicitly shown. The first three nodes in the chain from node i to the base are labelled as j , k , and l , the rest of the chain is represented by the dashed lines and circles inside the big circle. Three internal coordinates that are attributed to node i are the bond length q_i^{bon} , bond angle q_i^{ang} , and dihedral angle q_i^{dih} . Here, q_i^{bon} is the length of the bond between nodes i and j , q_i^{ang} is the angle of nodes i , j , and k , q_i^{dih} is the dihedral of nodes i , j , k , and l . (b) General representation of a non-cyclic BKS-tree to show the definition of an improper dihedral angle when a node j has more than one branches. Node l is selected as the node in the first branch instead of node l' . It defines an improper dihedral.

unit vector of q_i^{bon} (e_i^{bon}) is pointing from node j to node i , the unit vector of q_i^{dih} (e_i^{dih}) is pointing from node k to node j , and the unit vector of q_i^{ang} (e_i^{ang}) is the cross product of e_i^{bon} and e_i^{dih} (see Fig. 3).

For a node with more than one branch, the dihedral angle for the node on the first branch is assigned in the way mentioned above (the first branch is the branch with indices consecutive to the node where the branch starts), whereas dihedral angles for nodes in other branches are defined differently in the BKS-tree. As shown in Fig. 2(b), the node l is taken as the node in the first branch instead of the node l' in the chain back to the base and the three coordinates that determine the position of node i are again the bond length between node i and node j , the bond angle formed by nodes i , j , and k , and the dihedral formed by nodes i , j , k , and l . Note that the selected dihedral angle is now an improper dihedral angle. At each branch, there is only one proper dihedral and all the others are improper dihedrals.

Following the above procedure, a unique generalized coordinate set is defined once a BKS-tree is constructed. This set contains N bond lengths, N bond angles, and N dihedral angles. The positions of node i and all the nodes attached to node i depend on q_i^{bon} and q_i^{ang} . The positions of all the nodes attached to node j are dependent on q_i^{dih} if q_i^{dih} is a proper dihedral, whereas the positions of all the nodes attached to node i depend on q_i^{dih} if q_i^{dih} is an improper dihedral.

If the position $\mathbf{r}_\alpha$ of a node α depends on the three coordinates that are attributed to node i , its partial derivative to these coordinates are given as below:^{19,21}

$$\frac{\partial \mathbf{r}_\alpha}{\partial q_i^{\text{bon}}} = \mathbf{e}_i^{\text{bon}}, \quad (30a)$$

$$\frac{\partial \mathbf{r}_\alpha}{\partial q_i^{\text{ang}}} = \mathbf{e}_i^{\text{ang}} \times (\mathbf{r}_\alpha - \mathbf{r}_j), \quad (30b)$$

$$\frac{\partial \mathbf{r}_\alpha}{\partial q_i^{\text{dih}}} = \mathbf{e}_i^{\text{dih}} \times (\mathbf{r}_\alpha - \mathbf{r}_j), \quad (30c)$$

FIG. 3. Unit vectors attributed to node i . Three unit vectors are located at node j . The unit vector of q_i^{bon} (e_i^{bon}) points from node j to node i , the unit vector of q_i^{dih} (e_i^{dih}) points from node k to node j , the unit vector of q_i^{ang} (e_i^{ang}) is the cross product of e_i^{bon} and e_i^{dih} .

where $\mathbf{r}_j$ is the position of node j ; e_i^{bon} , e_i^{ang} , and e_i^{dih} are the unit vectors defined in Fig. 3. Thus, the A-matrix of such a BKS-tree coordinate set can be estimated analytically.

As an example for the BKS-tree in Fig. 1, the six semi-external coordinates are the bond between atoms 1 and O , the two bond angles (one by atoms 1, O , and Z ; the other by atoms 2, 1, and O), and the three dihedral angles (the first by atoms 1, O , Z , and Y ; the second by atoms 2, 1, O , and Z ; the third by atoms 3, 2, 1, and O). They are the coordinates that formed with at least one virtual atom. Apparently, these coordinates can change even when there is no overall translational and rotational motions of the molecule around the center of mass. On the other hand, their changes can be converted to overall translation and rotation of the molecule (but not necessarily the rotation around the center of mass) and have no effects on the internal coordinates. In short, they are semi-external coordinates. The rest $3N - 6$ coordinates are the internal coordinates which are uniquely determined once the BKS-tree is constructed.

B. Remove forces on external coordinates

Before we move on to discuss various properties of the generalized forces on the internal coordinates, we will first show that the forces acting solely on the external coordinates can be removed from the system without side effects. Therefore, all the conclusions we draw on systems with no force acting on their external coordinates will be equally valid for systems with net forces acting on their external coordinates.

Let $\mathbf{Q}_e$ and $\mathbf{Q}_q$ be the generalized forces acting on the external and internal coordinates, respectively. From Eq. (12), we have

$$\mathbf{Q}_e = \mathbf{A}_1^\top \mathbf{F}, \quad (31a)$$

$$\mathbf{Q}_q = \mathbf{A}_2^\top \mathbf{F}. \quad (31b)$$

Define $\mathbf{F}_{\text{ext}}$ as the force on each atom that originates from the generalized force acting solely on the external coordinates ($\mathbf{Q}_e$) and the force from $\mathbf{Q}_q$ as $\mathbf{F}_{\text{int}}$, that is,

$$\mathbf{F} = \mathbf{F}_{\text{ext}} + \mathbf{F}_{\text{int}}, \quad (32a)$$

$$\mathbf{F}_{\text{ext}} = \mathbf{B}_1^\top \mathbf{Q}_e, \quad (32b)$$

$$\mathbf{F}_{\text{int}} = \mathbf{B}_2^\top \mathbf{Q}_q. \quad (32c)$$

Then, we have

$$\mathbf{Q}_q = \mathbf{A}_2^\top \mathbf{F} \quad (33a)$$

$$= \mathbf{A}_2^\top \mathbf{F}_{\text{ext}} + \mathbf{A}_2^\top \mathbf{F}_{\text{int}} \quad (33b)$$

$$= \mathbf{A}_2^\top \mathbf{B}_1^\top \mathbf{Q}_e + \mathbf{A}_2^\top \mathbf{F}_{\text{int}} \quad (33c)$$

$$= \mathbf{A}_2^\top \mathbf{F}_{\text{int}}. \quad (33d)$$

Here, we note that $\mathbf{B}_1 \mathbf{A}_2 = \mathbf{0}_{6,3N-6}$; thus, we can subtract $\mathbf{F}_{\text{ext}}$ from the force on each atom, so the force on each atom is $\mathbf{F}_{\text{int}}$ only and the generalized forces on internal coordinates remain untouched. In this way, we can transform a system with non-vanishing forces on external coordinates into a system with zero such forces if we are only concerned with the generalized forces on internal coordinates. After removing the forces on external coordinates, the generalized forces on the internal

coordinates can be given by $\mathbf{Q}_q = \mathbf{B}_2^{\dagger} \mathbf{F}_{\text{int}}$ or $\mathbf{Q}_q = \mathbf{A}_2^{\dagger} \mathbf{F}_{\text{int}}$ (see Appendices B and C). Thus, we can replace $\mathbf{A}_2$ with $\mathbf{A}_2^{\dagger}$ in computing the generalized force for a system without generalized forces on external coordinates. This provides a computationally more efficient way as $\mathbf{A}_2^{\dagger}$ can be obtained analytically if proper semi-external coordinates are chosen.

Therefore, without loss of generality, we will discuss systems with no forces on their external coordinates, that is,

$$\mathbf{Q}_e = \mathbf{A}_1^{\dagger} \mathbf{F} = \mathbf{0}_6 \quad (34a)$$

$$\text{or } \mathbf{F}_{\text{tot}} = \sum_{i=1}^N \mathbf{F}_i = \mathbf{0}_3 \quad \text{and} \quad \boldsymbol{\tau} = \sum_{i=1}^N \mathbf{y}_i \times \mathbf{F}_i = \mathbf{0}_3. \quad (34b)$$

We note that the condition that the forces on external coordinates are zero does exclude the situation that there exist non-vanishing forces on individual atoms from interactions between the molecule and the environment.

C. Some invariant properties of generalized forces

1. Generalized force on a bond is independent of the choice of BKS-tree

In a non-cyclic BKS-tree, the bond between node i and node j separates all the nodes into two non-overlapping sets of nodes, S_i and S_j (see Fig. 4(a)). If the bond q_{bon} between nodes i and j is included in the set of internal coordinates, there are two cases: (a) the base is chosen from set S_j and nodes in set S_i are attached to node j ; (b) the base is chosen from set S_i and nodes in set S_j are attached to node i .

In case (a), the positions of nodes in set S_i depend on the bond q_{bon} , whereas the positions of nodes in set S_j do not. Therefore, the generalized force on bond q_{bon} is given by

$$\begin{aligned} Q_{\text{bon}}^a &= \sum_{\alpha=1}^N \mathbf{F}_{\alpha} \cdot \frac{\partial \mathbf{r}_{\alpha}}{\partial q_{\text{bon}}} = \sum_{\alpha \in S_i} \mathbf{F}_{\alpha} \cdot \frac{\partial \mathbf{r}_{\alpha}}{\partial q_{\text{bon}}} \\ &= \sum_{\alpha \in S_i} \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ji}, \end{aligned} \quad (35)$$

where $\mathbf{e}_{ji} = (\mathbf{r}_i - \mathbf{r}_j) / \|\mathbf{r}_i - \mathbf{r}_j\|$ and $\sum_{\alpha \in S_i}$ means a summation over all the nodes in set S_i . Similarly, in case (b), the

generalized force on bond q_{bon} is given by

$$Q_{\text{bon}}^b = \sum_{\alpha \in S_j} \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ij}. \quad (36)$$

Then,

$$\begin{aligned} Q_{\text{bon}}^a - Q_{\text{bon}}^b &= \sum_{\alpha \in S_i} \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ji} - \sum_{\alpha \in S_j} \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ij} \\ &= \sum_{\alpha \in S_i} \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ji} + \sum_{\alpha \in S_j} \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ji} \\ &= \sum_{\alpha=1}^N \mathbf{F}_{\alpha} \cdot \mathbf{e}_{ji} = 0. \end{aligned} \quad (37)$$

Eq. (37) used the condition that the total force on the system is zero, i.e., $\sum_{\alpha=1}^N \mathbf{F}_{\alpha} = \mathbf{0}_3$. Therefore, we have $Q_{\text{bon}}^a = Q_{\text{bon}}^b$ and the generalized force on the same bond is always the same and independent of the choice of BKS-tree.

2. Generalized force on bond angles

As shown in Fig. 4(b), the bond angle consists of nodes i , j , and k . Set S_i includes node i and all the nodes attached to node i if node i is attached to node j . Set S_k includes node k and all the nodes attached to node k if node k is attached to node j . Set S_j includes all the nodes other than node j and nodes in sets S_i and S_k .

If the bond angle q_{ang} formed by nodes i , j , and k is included in the set of internal coordinates, there are two cases: (a) the base is chosen from set S_i and nodes in set S_k are attached to node j ; (b) the base is chosen from set S_k and nodes in set S_i are attached to node j .

In the case of (a), the generalized force on the angle is given by

$$Q_{\text{ang}}^a = \sum_{\alpha \in S_k} \mathbf{F}_{\alpha} \cdot (\mathbf{e}_{jk} \times \mathbf{e}_{ij}) \times (\mathbf{r}_{\alpha} - \mathbf{r}_j) \quad (38)$$

Similarly, in the case of (b), the generalized force on the angle is given by

$$Q_{\text{ang}}^b = \sum_{\alpha \in S_i} \mathbf{F}_{\alpha} \cdot (\mathbf{e}_{ji} \times \mathbf{e}_{kj}) \times (\mathbf{r}_{\alpha} - \mathbf{r}_j). \quad (39)$$

Therefore,

$$\begin{aligned} Q_{\text{ang}}^a - Q_{\text{ang}}^b &= \sum_{\alpha \in S_k \cup S_i} \mathbf{F}_{\alpha} \cdot (\mathbf{e}_{jk} \times \mathbf{e}_{ij}) \times (\mathbf{r}_{\alpha} - \mathbf{r}_j) \\ &= (\mathbf{e}_{jk} \times \mathbf{e}_{ij}) \cdot \sum_{\alpha \in S_k \cup S_i} (\mathbf{r}_{\alpha} - \mathbf{r}_j) \times \mathbf{F}_{\alpha}. \end{aligned} \quad (40)$$

If node j have only two edges, then the summation in Eq. (40) vanishes because

$$\begin{aligned} \sum_{\alpha \in S_k \cup S_i} (\mathbf{r}_{\alpha} - \mathbf{r}_j) \times \mathbf{F}_{\alpha} &= \sum_{\alpha=1}^N (\mathbf{r}_{\alpha} - \mathbf{r}_j) \times \mathbf{F}_{\alpha} \\ &= \sum_{\alpha=1}^N \mathbf{r}_{\alpha} \times \mathbf{F}_{\alpha} - \mathbf{r}_j \times \sum_{\alpha=1}^N \mathbf{F}_{\alpha} = 0. \end{aligned} \quad (41)$$

Here, we used the conditions that the total torque $\sum_{\alpha=1}^N (\mathbf{r}_{\alpha} \times \mathbf{F}_{\alpha})$ and the total force $\sum_{\alpha=1}^N \mathbf{F}_{\alpha}$ of the system are zero. Thus, Q_{ang}^a

FIG. 4. (a) General representation of a non-cyclic BKS-tree with a bond between nodes i and j . Set S_i includes node i and all the nodes attached to node i , if node i is attached to node j . Set S_j includes node j and all the nodes attached to node j if node j is attached to node i . (b) General representation of a non-cyclic BKS-tree with a bond angle consisting of nodes i , j , and k . Set S_i includes node i and all the nodes attached to node i if node i is attached to node j . Set S_k includes node k and all the nodes attached to node k , if node k is attached to node j . Set S_j includes all the nodes other than node j and nodes in sets S_i and S_k .

equals Q_{ang}^b and is independent of the choice of BKS-tree. If node j has more than two edges, Q_{ang}^a is not necessarily equal to Q_{ang}^b , and the generalized forces on the same bond angle for two coordinate sets can be different.

3. Generalized force on a proper dihedral is independent of the choice of BKS-tree

As shown in Fig. 5(a), a proper dihedral q_{dih} consists of nodes i, j, k , and l . S_i includes node i and nodes attached to it, if node i is attached to node j ; S_l includes node l and nodes attached to it, if node l is attached to node k ; S_j includes nodes attached to node j except those in S_i , if node j is attached to node k ; S_k includes nodes attached to node k except those in S_l , if node k is attached to node j . There are only two cases that the dihedral q_{dih} consisting of nodes i, j, k , and l will be included in the set of internal coordinates. (a) The base is chosen from set S_l and node i is chosen as the node in the first branch; (b) the base is chosen from set S_i and node l is chosen as the node in the first branch.

In the case of (a), all the nodes attached to node j are dependent on the dihedral q_{dih} . We define this set of nodes as set S_a , which is the union of sets S_i and S_j . The generalized force on the dihedral q_{dih} is then given by

$$Q_{\text{dih}}^a = \sum_{\alpha=1}^N \mathbf{F}_\alpha \cdot \frac{\partial \mathbf{r}_\alpha}{\partial q_{\text{dih}}} = \sum_{\alpha \in S_a} \mathbf{F}_\alpha \cdot \frac{\partial \mathbf{r}_\alpha}{\partial q_{\text{dih}}} = \sum_{\alpha \in S_a} \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_j) \quad (42a)$$

$$= \sum_{\alpha \in S_a} \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_k). \quad (42b)$$

Similarly, in the case of (b), all the nodes attached to node k are dependent on dihedral q_{dih} . We define this set of nodes as set S_b , which is the union of sets S_l and S_k . The generalized

force on the dihedral q_{dih} is then given by

$$Q_{\text{dih}}^b = \sum_{\alpha \in S_b} \mathbf{F}_\alpha \cdot \mathbf{e}_{jk} \times (\mathbf{r}_\alpha - \mathbf{r}_k). \quad (43)$$

Then,

$$\begin{aligned} Q_{\text{dih}}^a - Q_{\text{dih}}^b &= \sum_{\alpha \in S_a} \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_k) - \sum_{\alpha \in S_b} \mathbf{F}_\alpha \cdot \mathbf{e}_{jk} \times (\mathbf{r}_\alpha - \mathbf{r}_k) \\ &= \sum_{\alpha \in S_a \cup S_b} \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_k) \\ &= \sum_{\alpha=1}^N \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_k) \\ &= \sum_{\alpha=1}^N \mathbf{e}_{kj} \cdot (\mathbf{r}_\alpha - \mathbf{r}_k) \times \mathbf{F}_\alpha \\ &= \mathbf{e}_{kj} \cdot \sum_{\alpha=1}^N (\mathbf{r}_\alpha - \mathbf{r}_k) \times \mathbf{F}_\alpha = 0. \end{aligned} \quad (44)$$

Here, we used the conditions that the total torque $\sum_{\alpha=1}^N (\mathbf{r}_\alpha \times \mathbf{F}_\alpha)$ and the total force $\sum_{\alpha=1}^N \mathbf{F}_\alpha$ of the system are zero. Therefore, the generalized force on a proper dihedral is independent of the choice of BKS-tree if the base is set as one of the leaves of the molecule.

4. Generalized forces on the same improper dihedral in different coordinate sets are generally not equal

As shown in Fig. 5(b), the improper dihedral consists of nodes i, j, k , and l . Set S_i includes node i and all the nodes attached to it, if node i is attached to node j . Set S_l includes node l and all the nodes attached to it, if node l is attached to node j . Set S_j includes nodes attached to node j but excluding those in sets S_i and S_l , if node j is attached to node k . Set S_k includes node k and all the nodes attached to it, if node k is attached to node j . There are two cases that the improper dihedral q_{imp} consisting of nodes i, j, k , and l will be included in the set of internal coordinates: (a) the base is from set S_k and node i is the node in the first branch; (b) the base is from set S_k and node l is the node in the first branch.

In the case of (a), the generalized force on the improper dihedral is given by

$$Q_{\text{imp}}^a = \sum_{\alpha \in S_l} \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_j). \quad (45)$$

Similarly, in the case of (b), the generalized force on the improper dihedral is given by

$$Q_{\text{imp}}^b = \sum_{\alpha \in S_i} \mathbf{F}_\alpha \cdot \mathbf{e}_{kj} \times (\mathbf{r}_\alpha - \mathbf{r}_j). \quad (46)$$

Apparently, Q_{imp}^a is not necessarily equal to Q_{imp}^b .

V. CONCLUDING REMARKS

We discussed a set of generalized coordinates, in which external coordinates and internal coordinates are separated

FIG. 5. (a) General representation of a non-cyclic BKS-tree with a proper dihedral by nodes i, j, k , and l . S_i includes node i and nodes attached to it, if node i is attached to node j ; S_l includes node l and nodes attached to it, if node l is attached to node k ; S_j includes nodes attached to node j except those in S_i , if node j is attached to node k ; S_k includes nodes attached to node k except those in S_l , if node k is attached to node j . (b) General representation of a non-cyclic BKS-tree with an improper dihedral consisting of nodes i, j, k , and l . Set S_i includes node i and all the nodes attached to it, if node i is attached to node j . Set S_l includes node l and all the nodes attached to it, if node l is attached to node j . Set S_j includes nodes attached to node j but excluding those in sets S_i and S_l , if node j is attached to node k . Set S_k includes node k and all the nodes attached to it, if node k is attached to node j .

exactly in the sense that there is no coupling between their kinetic energies. The full A-matrix and B-matrix were provided and an efficient computational method was proposed. The B-matrix can be evaluated analytically, whereas the A-matrix can be computed without inverting the corresponding B-matrix—only the inversion of a 6×6 matrix is required instead. These explicit information of the A-matrix and B-matrix of the proposed generalized coordinate set are expected to be very useful in various studies, especially in the analysis of results from molecular dynamics simulations.

The quasi-Euler angles given in Eq. (16b) are defined with respect to an instantaneous reference configuration. Although these coordinates are different from conventional ones, they could be very useful for the purposes of analysis. Recently, curvilinear coordinates on a non-Eckart body-frame and delocalized internal coordinates, which are defined with respect to an instantaneous reference configuration as well, were shown to be useful in normal mode analysis³ and mechanochemical force analysis,⁸ respectively. In the current work, the inclusion of quasi-Euler angle as rotational coordinates in a complete set of coordinates provides an exact decoupling between external and internal motions. In particular, we explicitly provided the complete A-matrix and B-matrix for such a coordinate set. However, A_1^θ and B_1^θ in Eqs. (14b) and (17b) respectively, are functions of Cartesian coordinates, not the generalized coordinates themselves. The complete A- and B-matrices therefore should not be used to simulate the evolution of polyatomic molecules. Instead, they are designed to analyse the properties of the molecules given the trajectories of the molecules in Cartesian coordinates are known. All the above-mentioned analyses are based on classical mechanics, although the trajectories can be provided by either classical or quantum mechanical simulations.

In the following, we discuss a few examples where the results from the current paper could be useful. One potential application is to use the insights from Eq. (4) for structural alignment of configurations from molecular dynamics simulations. With A and B matrices at hand, the changes in Cartesian coordinates caused by external coordinates ($\delta\mathbf{x1}$) and by internal coordinates ($\delta\mathbf{x2}$) can be evaluated by Eq. (4). Then, for two configurations connected by a dynamic trajectory, optimal structural superposition of these configurations can be achieved by simply removing $\delta\mathbf{x1}$. This approach provides an alternative to the widely adopted method of minimizing the mass-weighted root-mean-square-deviation between the two target configurations.²² The optimal superposition of configurations is usually the preconditioning step in the widely applied normal mode analyses¹ and principal component analyses.²³ Another example of potential application is to the analysis of vibrational energy relaxation.²⁴ The separation between vibrational and translational-rotational energies given in Eq. (9) could be used to remove the energies associated with the overall translation and rotation of polyatomic molecules, which are simply ignored as a well-accepted approximation in the current practice of the field. Furthermore, the application to study protein structural changes induced by shear flows might be promising.^{25–27}

In addition, some of the invariant properties of the BAT coordinates derived from a BKS-tree were presented. Although we demonstrate such invariant properties for a system with no forces on the external coordinates, the conclusions hold for systems with non-vanishing external forces as we can safely remove external forces and the generalized forces on internal coordinates remain the same. These invariant properties on the generalized forces of the BAT coordinates highlighted the importance of applying BAT coordinates in the analyses of the system which are related to the generalized force. Examples could be force distribution analysis^{5–8} and studies in mechanochemistry.^{28–30}

ACKNOWLEDGMENTS

The work is supported by the NIH Grant No. R01 GM086536 awarded to A.M.

APPENDIX A: THE INTERNAL PART OF B-MATRIX

The elements of B-matrix can be calculated as follows.¹⁷

If $q_{\text{bon}} = r_{ji}$ is the length of a bond between atoms i and j with $r_{ji} = \|\mathbf{r}_i - \mathbf{r}_j\|$, $\mathbf{e}_{ji} = (\mathbf{r}_i - \mathbf{r}_j)/\|\mathbf{r}_i - \mathbf{r}_j\|$ is the unit vector pointing from atom j to atom i . Then,

$$\frac{\partial q_{\text{bon}}}{\partial \mathbf{r}_i} = \mathbf{e}_{ji}, \quad (\text{A1a})$$

$$\frac{\partial q_{\text{bon}}}{\partial \mathbf{r}_j} = \mathbf{e}_{ij} = -\mathbf{e}_{ji}, \quad (\text{A1b})$$

$$\frac{\partial q_{\text{bon}}}{\partial \mathbf{r}_k} = \mathbf{0}_3 \quad \text{if } k \neq i \text{ or } j. \quad (\text{A1c})$$

If $q_{\text{ang}} = \theta$ is the bond angle formed by atoms i , j , and k with

$$\cos(\theta) = \mathbf{e}_{ji} \cdot \mathbf{e}_{jk}, \quad (\text{A2})$$

Then,

$$\frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_i} = \frac{\cos(\theta)\mathbf{e}_{ji} - \mathbf{e}_{jk}}{r_{ji} \sin(\theta)}, \quad (\text{A3a})$$

$$\frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_j} = \frac{(r_{ji} - r_{jk} \cos(\theta))\mathbf{e}_{ji} + (r_{jk} - r_{ji} \cos(\theta))\mathbf{e}_{jk}}{r_{ji}r_{jk} \sin(\theta)}, \quad (\text{A3b})$$

$$\frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_k} = \frac{\cos(\theta)\mathbf{e}_{jk} - \mathbf{e}_{ji}}{r_{jk} \sin(\theta)}, \quad (\text{A3c})$$

$$\frac{\partial q_{\text{ang}}}{\partial \mathbf{r}_l} = \mathbf{0}_3 \quad \text{if } l \neq i, j, \text{ or } k. \quad (\text{A3d})$$

If $q_{\text{dih}} = \phi$ is the dihedral angle (or torsion angle) formed by atoms i , j , k , and l with

$$\cos(\phi) = \frac{(\mathbf{e}_{ij} \times \mathbf{e}_{jk}) \cdot (\mathbf{e}_{jk} \times \mathbf{e}_{kl})}{\sin(\theta_2) \sin(\theta_3)}, \quad (\text{A4})$$

θ_2 is the angle formed by atoms i , j , and k , while θ_3 is the angle formed by atoms j , k , and l , and

$$\cos(\theta_2) = \mathbf{e}_{ji} \cdot \mathbf{e}_{jk}, \quad (\text{A5a})$$

$$\cos(\theta_3) = \mathbf{e}_{kj} \cdot \mathbf{e}_{kl}. \quad (\text{A5b})$$

Then,

$$\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_i} = -\frac{\mathbf{e}_{ij} \times \mathbf{e}_{jk}}{r_{ji} \sin^2(\theta_2)}, \quad (\text{A6a})$$

$$\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j} = \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j}\right)_1 + \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j}\right)_2, \quad (\text{A6b})$$

$$\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k} = \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k}\right)_1 + \left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k}\right)_2, \quad (\text{A6c})$$

$$\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_l} = -\frac{\mathbf{e}_{lk} \times \mathbf{e}_{kj}}{r_{lk} \sin^2(\theta_3)}, \quad (\text{A6d})$$

$$\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_m} = \mathbf{0}_3 \quad \text{if } m \neq i, j, k \text{ or } l, \quad (\text{A6e})$$

with

$$\left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j}\right)_1 = \frac{(r_{jk} - r_{ji} \cos(\theta_2))(\mathbf{e}_{ij} \times \mathbf{e}_{jk})}{r_{ji} r_{jk} \sin^2(\theta_2)}, \quad (\text{A6f})$$

$$\left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_j}\right)_2 = \frac{\cos(\theta_3)(\mathbf{e}_{lk} \times \mathbf{e}_{kj})}{r_{jk} \sin^2(\theta_3)}, \quad (\text{A6g})$$

$$\left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k}\right)_1 = \frac{(r_{jk} - r_{kl} \cos(\theta_3))(\mathbf{e}_{kl} \times \mathbf{e}_{jk})}{r_{jk} r_{kl} \sin^2(\theta_3)}, \quad (\text{A6h})$$

$$\left(\frac{\partial q_{\text{dih}}}{\partial \mathbf{r}_k}\right)_2 = \frac{\cos(\theta_2)(\mathbf{e}_{ij} \times \mathbf{e}_{jk})}{r_{jk} \sin^2(\theta_2)}. \quad (\text{A6i})$$

APPENDIX B: PSEUDOINVERSE OF $\mathbf{B}_2$ MATRIX GIVES EXACT GENERALIZED FORCES ON INTERNAL COORDINATES WHEN FORCES ON EXTERNAL COORDINATES ARE ZERO

Given a system with no generalized forces on external coordinates and the force $\mathbf{F}$ on each atom, we prove that

$$\mathbf{Q}_q^\top = \mathbf{F}^\top \mathbf{A}_2 = \mathbf{F}^\top \mathbf{B}_2^+ \quad (\text{B1})$$

as follows:

$$\mathbf{F}^\top \mathbf{A}_2 = \mathbf{F}^\top \mathbf{B}_2^+ \quad (\text{B2a})$$

$$\iff \mathbf{F}^\top \mathbf{A}_2 = \mathbf{F}^\top \mathbf{B}_2^T (\mathbf{B}_2 \mathbf{B}_2^T)^{-1} \quad (\text{B2b})$$

$$\iff \mathbf{F}^\top \mathbf{A}_2 \mathbf{B}_2 \mathbf{B}_2^T = \mathbf{F}^\top \mathbf{B}_2^T \quad (\text{B2c})$$

$$\iff \mathbf{F}^\top (\mathbf{I}_{3N} - \mathbf{A}_1 \mathbf{B}_1) \mathbf{B}_2^T = \mathbf{F}^\top \mathbf{B}_2^T \quad (\text{B2d})$$

$$\iff \mathbf{F}^\top \mathbf{B}_2^T - \mathbf{F}^\top \mathbf{A}_1 \mathbf{B}_1 \mathbf{B}_2^T = \mathbf{F}^\top \mathbf{B}_2^T \quad (\text{B2e})$$

$$\iff \mathbf{F}^\top \mathbf{B}_2^T = \mathbf{F}^\top \mathbf{B}_2^T, \quad (\text{B2f})$$

here, Eqs. (6e) and (34a) are used.

APPENDIX C: THE INTERNAL PART OF A-MATRIX FOR GENERALIZED COORDINATES WITH SEMI-EXTERNAL COORDINATES ($\mathbf{A}_2^\zeta$) GIVES THE EXACT GENERALIZED FORCES ON INTERNAL COORDINATES WHEN FORCES ON EXTERNAL COORDINATES ARE ZERO

From Eq. (27b), we have

$$\mathbf{A}_1 = \mathbf{A}_1^\zeta \mathbf{E}_{11} \quad \text{and} \quad \mathbf{A}_2 = \mathbf{A}_1^\zeta \mathbf{E}_{12} + \mathbf{A}_2^\zeta \quad (\text{C1a})$$

$$\Rightarrow \mathbf{A}_1^\zeta = \mathbf{A}_1 \mathbf{E}_{11}^{-1} \quad \text{and} \quad \mathbf{A}_2^\zeta = \mathbf{A}_2 - \mathbf{A}_1 \mathbf{E}_{11}^{-1} \mathbf{E}_{12} \quad (\text{C1b})$$

and $\mathbf{F}^\top \mathbf{A}_2^\zeta$ gives the exact generalized forces on internal coordinates, since

$$\mathbf{F}^\top \mathbf{A}_2^\zeta = \mathbf{F}^\top (\mathbf{A}_2 - \mathbf{A}_1 \mathbf{E}_{11}^{-1} \mathbf{E}_{12}) \quad (\text{C2a})$$

$$= \mathbf{F}^\top \mathbf{A}_2 - \mathbf{F}^\top \mathbf{A}_1 \mathbf{E}_{11}^{-1} \mathbf{E}_{12} \quad (\text{C2b})$$

$$= \mathbf{F}^\top \mathbf{A}_2 - \mathbf{0}_6^\top \mathbf{E}_{11}^{-1} \mathbf{E}_{12} \quad (\text{C2c})$$

$$= \mathbf{F}^\top \mathbf{A}_2 = \mathbf{Q}_q^\top, \quad (\text{C2d})$$

here, Eq. (34a) is used.

¹G. Mathias and M. D. Baer, "Generalized normal coordinates for the vibrational analysis of molecular dynamics simulations," *J. Chem. Theory Comput.* **7**, 2028–2039 (2011).

²P. Y. Ayala and H. B. Schlegel, "Identification and treatment of internal rotation in normal mode vibrational analysis," *J. Chem. Phys.* **108**, 2314–2325 (1998).

³J. Pesonen, K. O. Henriksson, J. R. López-Blanco, and P. Chacón, "Normal mode analysis of molecular motions in curvilinear coordinates on a non-Eckart body-frame: An application to protein torsion dynamics," *J. Math. Chem.* **50**, 1521–1549 (2012).

⁴K. Kamiya, Y. Sugawara, and H. Umeyama, "Algorithm for normal mode analysis with general internal coordinates," *J. Comput. Chem.* **24**, 826–841 (2003).

⁵W. Stacklies, M. C. Vega, M. Wilmanns, and F. Gräter, "Mechanical network in titin immunoglobulin from force distribution analysis," *PLoS Comput. Biol.* **5**, e1000306 (2009).

⁶W. Li, S. A. Edwards, L. Lu, T. Kubar, S. P. Patil, H. Grubmüller, G. Groenhof, and F. Gräter, "Force distribution analysis of mechanochemically reactive dimethylcyclobutene," *ChemPhysChem* **14**, 2687–2697 (2013).

⁷T. Stauch and A. Dreuw, "A quantitative quantum-chemical analysis tool for the distribution of mechanical force in molecules," *J. Chem. Phys.* **140**, 134107 (2014).

⁸T. Stauch and A. Dreuw, "On the use of different coordinate systems in mechanochemical force analyses," *J. Chem. Phys.* **143**, 074118 (2015).

⁹C. Eckart, "Some studies concerning rotating axes and polyatomic molecules," *Phys. Rev.* **47**, 552 (1935).

¹⁰J. Jellinek and D. Li, "Separation of the energy of overall rotation in any N-body system," *Phys. Rev. Lett.* **62**, 241 (1989).

¹¹J. Jellinek and D. Li, "Vibrations of rapidly rotating N-body systems," *Chem. Phys. Lett.* **169**, 380–386 (1990).

¹²H. Goldstein, *Classical Mechanics* (Pearson Education India, 1965).

¹³S.-H. Lee, K. Palmo, and S. Krimm, "A new formalism for molecular dynamics in internal coordinates," *Chem. Phys.* **265**, 63–85 (2001).

¹⁴S.-H. Lee, K. Palmo, and S. Krimm, "A new method to calculate B-matrix elements for external rotations with application to the B-matrix internal coordinate molecular dynamics formalism," *Chem. Phys. Lett.* **342**, 643–651 (2001).

¹⁵N. Gö and H. A. Scheraga, "On the use of classical statistical mechanics in the treatment of polymer chain conformation," *Macromolecules* **9**, 535–542 (1976).

¹⁶J. H. Frederick and C. Woywod, "General formulation of the vibrational kinetic energy operator in internal bond-angle coordinates," *J. Chem. Phys.* **111**, 7255–7271 (1999).

¹⁷E. B. Wilson, J. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw Hill, 1955).

¹⁸R. Penrose, "A generalized inverse for matrices," *Math. Proc. Cambridge Philos. Soc.* **51**, 406–413 (1955).

¹⁹A. K. Mazur and R. A. Abagyan, "New methodology for computer-aided modelling of biomolecular structure and dynamics I. Non-cyclic structures," *J. Biomol. Struct. Dyn.* **6**, 815–832 (1989).

²⁰A. Mazur, V. Dorofeev, and R. Abagyan, "Derivation and testing of explicit equations of motion for polymers described by internal coordinates," *J. Comput. Phys.* **92**, 261–272 (1991).

²¹H. B. Thompson, "Calculation of Cartesian coordinates and their derivatives from internal molecular coordinates," *J. Chem. Phys.* **47**, 3407 (1967).

²²W. Kabsch, "A solution for the best rotation to relate two sets of vectors," *Acta Crystallogr., Sect. A: Cryst. Phys., Diff., Theor. Gen. Crystallogr.* **32**, 922–923 (1976).

²³F. Sittel, A. Jain, and G. Stock, "Principal component analysis of molecular dynamics: On the use of Cartesian vs. internal coordinates," *J. Chem. Phys.* **141**, 014111 (2014).

- ²⁴P. H. Nguyen and G. Stock, "Nonequilibrium molecular-dynamics study of the vibrational energy relaxation of peptides in water," *J. Chem. Phys.* **119**, 11350–11358 (2003).
- ²⁵I. B. Bekard, P. Asimakis, J. Bertolini, and D. E. Dunstan, "The effects of shear flow on protein structure and function," *Biopolymers* **95**, 733–745 (2011).
- ²⁶P. Szymczak and M. Cieplak, "Stretching of proteins in a uniform flow," *J. Chem. Phys.* **125**, 164903 (2006).
- ²⁷P. Szymczak and M. Cieplak, "Proteins in a shear flow," *J. Chem. Phys.* **127**, 155106 (2007).
- ²⁸J. Ribas-Arino, M. Shiga, and D. Marx, "Understanding covalent mechanochemistry," *Angew. Chem., Int. Ed.* **48**, 4190–4193 (2009).
- ²⁹J. Liang and J. M. Fernández, "Mechanochemistry: One bond at a time," *ACS Nano* **3**, 1628–1645 (2009).
- ³⁰W. Li and F. Gräter, "Atomistic evidence of how force dynamically regulates thiol/disulfide exchange," *J. Am. Chem. Soc.* **132**, 16790–16795 (2010).