Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit

H. Lee Woodcock,[1,a)] Wenjun Zheng,[1,b)] An Ghysels,[2] Yihan Shao,[3] Jing Kong,[3] and Bernard R. Brooks[1]
[1]*Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA*
[2]*Center for Molecular Modeling, Ghent University, Proeftuinstraat 86, B-9000 Gent, Belgium*
[3]*Q-Chem Inc., The Design Center, Suite 690, 5001 Baum Boulevard, Pittsburgh, Pennsylvania 15213, USA*

(Received 18 June 2008; accepted 13 October 2008; published online 3 December 2008)

A new vibrational subsystem analysis (VSA) method is presented for coupling global motion to a local subsystem while including the inertial effects of the environment. The premise of the VSA method is a partitioning of a system into a smaller region of interest and a usually larger part referred to as environment. This method allows the investigation of local-global coupling, a more accurate estimation of vibrational free energy contribution for parts of a large system, and the elimination of the "tip effect" in elastic network model calculations. Additionally, the VSA method can be used as a probe of specific degrees of freedom that may contribute to free energy differences. The VSA approach can be employed in many ways, but it will likely be most useful for estimating activation free energies in QM/MM reaction path calculations. Four examples are presented to demonstrate the utility of this method. © *2008 American Institute of Physics*.
[DOI: 10.1063/1.3013558]

I. INTRODUCTION

Modeling the dynamics and free energy of biomolecular systems remains a major challenge for computational biologists. Currently, minimization, molecular dynamics (MD), and harmonic analysis are the primary tools employed to examine the structure and dynamics of biomolecules. MD approaches have dominated the biosimulation community; however, the timescales required to observe global motion, which can be on the order of milliseconds to seconds, have prevented many systems from being studied.[1] Although this is a particular problem for larger systems, MD is still the most commonly employed method for three main reasons. First, harmonic analysis does not account for multiple minima although newly developed course grain methodology has mitigated this problem.[2–4] Second, there is often too much "noise" in large biomolecular systems which makes extracting the desired observables difficult. Third, although the computational cost of second derivative techniques is less computationally expensive than running long MD simulations the $O(N^3)$ memory requirements have been prohibitive until recently. For example, a 50 000 atom system would require a computer with over 150 Gb of memory.

Normal mode analysis (NMA) methods have been proposed and utilized to extend the range of biomolecules researchers are capable of examining.[5–9] This ability is predicated on the theory that harmonic dynamics are sufficient for understanding the collective motions of near-native states with low frequency normal modes dominating the qualitative dynamics.[7,10–14] This assertion has been shown to be a good approximation by numerous groups[15–23] and has led to the development of elastic network model (ENM) methods.[2–4] These methods typically replace all-atom descriptions with C_α-centered harmonic potentials that utilize a single force constant to account for all pairwise interactions within some cutoff distance (R_c). For example, the energy of an ENM representation can be defined as

$$E_{\mathrm{ENM}} = \frac{1}{2} \sum_{d_{ij}^0 < R_c} C(d_{ij} - d_{ij}^0)^2, \tag{1}$$

where d_{ij} is the distance between C_α atoms and d_{ij}^0 is the reference value from experiment (i.e., crystal structure, NMR, etc.). Given the simplicity of these models, previously impossible NMA can be performed using ENM methods. This approach has been shown to to be an effective means of extracting patterns of low frequency motion.[24]

In the current work we present the vibrational subsystem analysis (VSA) method for coupling global motion to a local subsystem. This method is a partitioning scheme that separates (and integrates out) the motion of the environment from a user defined subsystem (see Methods section) while still allowing the environmental motion to perturb the local subsystem dynamics. It was originally applied to ENM models[25] but is now generalized and extended for all-atom representations and hybrid quantum mechanical/molecular mechanical (QM/MM) potentials. Below is a brief list of possible uses.

(1) Initially described as an ENM methodology, the VSA method is well suited to describing local-global coupling in coarse-grained macromolecular systems.[26] In the current work we have extended this approach to

a)Electronic mail: hlwood@nih.gov.
b)Current address: Department of Physics, University at Buffalo, The State University of New York, Buffalo, New York 14260-1500, USA.

include a mass term in the ENM-NMA procedure which allows for better coupling of timescales since the inertial terms in the environment are handled more appropriately.

(2) Eliminating specific degrees of freedom without the detrimental effects of constraining the motion (i.e., making the system too rigid) or deleting part of the system (i.e., artificially increasing flexibility). A particularly useful illustration of this is in the "tip effect" (Ref. 27) which refers to small pieces of proteins that can protrude out of the main globular body (e.g., surface loops, unresolved density, etc.). Using the VSA method these problematic degrees of freedom can be integrated out of the environment while still being included as a perturbation to the subsystem of interest.

(3) Elimination of noise when computing harmonic vibration free energy of large biomolecular systems. The term noise is used to depict unwanted weak coupling between degrees of freedom in the environment and degrees of freedom in the subsystem that makes it difficult to calculate properties of interest. This is particularly important when performing QM/MM free energy calculations as the classical entropic effects from the environment can be included as a first order perturbation to the QM harmonic free energy. In addition, this can also be a very useful tool in describing how global motion of a protein can couple to the active site before and during biochemical reactions.

(4) Combining VSA with simulation approaches. For example, the use of quasiharmonic analysis to generate an effective force constant matrix from a detailed simulation can allow exploration of "hidden" couplings that ENM models cannot capture. In general, the VSA approach can be combined with numerous additional normal mode treatments; such as reduced basis methods (i.e., block normal mode method) to further reduce the number of degrees of freedom treated.[28]

(5) Inclusion of very light or mass-less particles into NMA without the need for constraints or inclusion of unwanted high frequency heat capacity. This could be particularly useful as Drude oscillator approaches become a standard classical polarization technique.

(6) One of the most popular reaction path following techniques currently in use is the intrinsic reaction coordinate (IRC) method.[29,30] The IRC method, which requires knowledge of the transition state, employs internal coordinates to map the minimum energy pathway connecting reactants and products. Although this is an effective approach, IRC mapping is typically not performed on high dimensional systems (i.e., proteins) for various reasons. One benefit of the VSA approach is that by integrating out the environment, rather than fixing or deleting it, second order methods that have previously been sparingly used (i.e., IRC) will again become available. Taking this in combination with the full QM/MM analytic second derivatives introduced in the current work QM/MM IRC calculations on biomolecular systems could easily gain mainstream use.

In Sec. II (Methods) we improve the original approach for vibrational subsystem analysis[26] and derive the formula for including environmental inertial effects in the VSA method in two different ways. Also described is the extension of the vibrational capabilities of CHARMM to utilize hybrid QM/MM potentials. In Sec. III (Results and Discussion) we give four examples of where VSA can be applied and discuss how the results correspond to full NMA; (1) VSA is applied to a simple harmonic string of beads and computed results are compared to analytically derived results. (2) VSA is applied to the torsional potential of butane with only carbon atoms defined as the subsystem. (3) VSA is applied to the cyclohexane chair to boat conversion with only carbon atoms defined as the subsystem. (4) QM (B3LYP/6-31G*) NMA, VSA, and block normal mode (BNM) methods are applied to a peptide radical rearrangement. Full NMA, VSA, and BNM results are compared and implications of the choice of subsystem is explored. Section IV (Conclusions) is an overview of the VSA method with results and future directions highlighted.

II. METHODS

A. The VSA method

To specify the vibrational subsystem methodology, the entire system is first divided into two components: (1) the subsystem, which is defined as the region of interest (i.e., part of the system that controls functionality); (2) the environment, which consists of the less important remaining portions of the molecule. The main idea is to study the coupling of these two regions and how this affects the dynamics of the subsystem. To do this the potential energy of the full system is defined as

$$2E_{\mathrm{pot}} = x^T H x = [x_s^T x_e^T] H \begin{bmatrix} x_s \\ x_e \end{bmatrix} = x_s^T H_{ss} x_s + x_s^T H_{se} x_e + x_e^T H_{es} x_s + x_e^T H_{ee} x_e, \tag{2}$$

with x_s and x_e defined as the displacements of the system and environment atoms, respectively, and the full Hessian defined as

$$H = \begin{bmatrix} H_{ss} & H_{se} \\ H_{es} & H_{ee} \end{bmatrix}, \tag{3}$$

where H_{ss}, H_{se}, and H_{ee} are the subsystem-subsystem, subsystem-environment, and environment-environment Hessians, respectively. x_e is then integrated out by setting $\partial E / \partial x_e = 0$ as a constraint, which leads to

$$x_e^0 = -H_{ee}^{-1} H_{es} x_s, \tag{4}$$

which further leads to the redefinition of the potential energy using an effective Hessian (H_{ss}^{eff})

$$2E_{\mathrm{pot}} = x_s^T \cdot H_{ss}^{eff} x_s = x_s^T (H_{ss} - H_{se} H_{ee}^{-1} H_{es}) x_s. \tag{5}$$

The fundamental approximation that is made by performing this partitioning is that the environment will be able to respond to structural changes from the subsystem by minimizing the total energy. Using this approximation and choosing an appropriate subsystem induced global changes can be

predicted. Additionally, performing a standard (NMA) employing H_{ss}^{eff} will yield the local modes that describe subsystem dynamics with a flexible environment. Following from the definition of the energy, the free energy conformational integral can also be defined and separated,

$$Z = \int e^{-(x_s^T H_{ss} x_s + x_s^T H_{se} x_s + x_e^T H_{es} x_e + x_e^T H_{ee} x_e)/2k_B T} dx_s dx_e$$
$$= \int e^{-x_{er}^T H_{ee} x_{er}/2k_B T} dx_{er} \times \int e^{-x_s^T H_{ss}^{eff} x_s/2k_B T} dx_s, \quad (6)$$

with x_{er} representing the motion of environment atoms relative to the derived positions satisfying Eq. (4) (i.e., $x_{er}=x_e-x_e^0$). This also leads to a formal splitting of the vibrational free energy via

$$F = -k_B T \log(Z)$$
$$= F_e + F_s = -k_B T \log\left(\int e^{-x_{er}^T H_{ee} x_{er}/2k_B T} dx_{er} \right)$$
$$- k_B T \log\left(\int e^{-x_s^T H_{ss}^{eff} x_s/2k_B T} dx_s \right). \quad (7)$$

Since we are primarily interested in the dynamics of the subsystem (x_s) we can ignore differences in F_e and focus entirely on changes in F_s. For example, in larger systems F_e is likely to be "noisy," but may be safe to ignore whereas F_s contains subsystem data hopefully with suitable accuracy. If either of these conditions is not satisfied then repartitioning of the subsystem and environment is required and the VSA procedure needs to be repeated. More specifically, the partition of subsystem and environment should be done several times when exploring the importance of specific degrees of freedom for a reaction coordinate of interest.

To more realistically describe the system we also introduce mass into the subsystem partitioning. The addition of mass allows better timescale coupling as inertial terms in the environment are treated more consistently,

$$HV = \lambda MV, \quad (8)$$

where Eq. (8) is the secular equation and H the full Hessian, V the eigenvectors/normal modes, λ the eigenvalues/frequencies squared, and M the kinetic energy matrix with mass elements along the diagonal. Eq. (8) can be written as

$$H_{ss}V_s + H_{se}V_e = \lambda M_s V_s,$$
$$H_{es}V_s + H_{ee}V_e = \lambda M_e V_e, \quad (9)$$

where subindex s or e refers to system or environment, respectively. Elimination of V_e from Eq. (9) gives

$$H_{se}(\lambda M_e - H_{ee})^{-1} H_{es} V_s = (\lambda M_s - H_{ss}) V_s, \quad (10)$$

with $(\lambda M_e - H_{ee})^{-1} \approx -H_{ee}^{-1} - \lambda H_{ee}^{-1} M_e H_{ee}^{-1}$ in the limit of small λ thus leading to an expansion of a first order generalized eigenvalue problem,

$$(H_{ss} - H_{se}H_{ee}^{-1}H_{es})V_s \approx \lambda(M_s + H_{se}H_{ee}^{-1}M_e H_{ee}^{-1}H_{es})V_s. \quad (11)$$

The left side of Eq. (11) has the same effective Hessian as Eq. (5) while the right hand side of Eq. (11) has an effective kinetic energy matrix that includes the inertial effects of the motion of the environment.

Another equivalent derivation uses Lagrange multipliers. If $\mathcal{L}$ denotes the Lagrangian of the system, each of the constraints $f_k(x)=0$ can be taken into account by adding a term $\lambda_k f_k(x)$ to $\mathcal{L}$. Hence an additional set of variables $\{\lambda_k\}$, the Lagrange multipliers, are introduced and the Lagrangian of the constrained system becomes

$$\mathcal{L} = T - V + \sum_k \lambda_k f_k(x). \quad (12)$$

In the case of VSA, one keeps all atoms in the environment force free, and the Lagrangian reads

$$\mathcal{L} = T - V - \sum_e^{3n_e} \lambda_e \frac{\partial V}{\partial x_e}, \quad (13)$$

with n_e defined as the number of atoms in the environment. In order to get the harmonic oscillator approximation, the Lagrangian is expanded up to second order in the displacements of x_s, x_e, and λ_e.

$$\mathcal{L} = \tfrac{1}{2}\dot{x}_e^T M_e \dot{x}_e + \tfrac{1}{2}\dot{x}_s^T M_s \dot{x}_s - \tfrac{1}{2}x_e^T H_{ee} x_e - \tfrac{1}{2}x_s^T H_{ss} x_s - x_e^T H_{ee} x_s$$
$$- \lambda_e^T H_{ee} x_e - \lambda_e^T H_{es} x_s. \quad (14)$$

In the unconstrained case, the Euler–Lagrange equations lead to Newton equations, which are first order. Here the VSA constraints generate $3n_e$ extra equations for the Lagrange multipliers,

$$M_s \ddot{x}_s + H_{ss} x_s + H_{se} x_e + H_{se} \lambda_e = 0, \quad (15)$$

$$M_e \ddot{x}_e + H_{es} x_s + H_{ee} x_e + H_{ee} \lambda_e = 0, \quad (16)$$

$$H_{es} x_s + H_{ee} x_e = 0. \quad (17)$$

Eliminating x_e and λ_e from Eqs. (15)–(17), the VSA eigenvalues λ and vectors $V' = (V_s, V_e, V_\lambda)$ are easily found from the following eigenvalue problem for the subsystem:

$$(H_{ss} - H_{se}H_{ee}^{-1}H_{es})V_s = \lambda(M_s + H_{se}H_{ee}^{-1}M_e H_{ee}^{-1}H_{es})V_s \quad (18)$$

and consequently

$$V_e = -H_{ee}^{-1}H_{es}V_s, \quad (19)$$

$$V_\lambda = -\lambda H_{ee}^{-1} M_e H_{ee}^{-1} H_{es} V_s. \quad (20)$$

These equations are formally identical to those derived with perturbation theory above.

B. Hybrid QM/MM analytic second derivatives

Expanding on the functionality of subsystem partitioning, we implement hybrid QM/MM analytic second derivatives into the Q-CHEM software package and interface this with VIBRan module of CHARMM.[5–7,31,32] Both restricted and unrestricted wave-functions are supported at the HF and

TABLE I. Overview of the integrals that can be neglected (denoted with an ×) when considering QM, QM/MM, and MM/MM blocks of the Hessian matrix.

	QM-QM	QM-MM/MM-QM	MM-MM
PH^{xy}	√	√	√
$P\|\|^{xy}P$	√	×	×
$(PFP)S^{xy}$	√	×	×
P^yH^x	√	√	√
$P^y\|\|^xP$	√	√	×
$(PFP)^yS^x$	√	√	×
γ^{xy}	√	√	√

DFT levels of theory, whereas previous work in this area focused only on RHF and RDFT treatments.[33] We briefly review standard SCF derivative theory and illustrate how this differs from applying hybrid QM/MM methodology.

Using Pople's compact notation,[34] where $\langle\ \rangle$ denotes the trace of a matrix, the SCF energy can be defined as

$$E = \langle PH\rangle + \tfrac{1}{2}\langle P||P\rangle + \gamma, \tag{21}$$

where P is the density matrix, H is the core Hamiltonian, and $||$ are the antisymmetrized two-electron integrals over spin orbitals μ, ν, λ, σ, and γ is the nuclear repulsion energy. The Fock operator is then defined as

$$F = H + P|| \tag{22}$$

and the standard SCF can be written as

$$FPS = SPF, \tag{23}$$

with S defined as the overlap matrix. Now, taking the second derivative with respect to atomic perturbations the following expression is derived:

$$\frac{\partial^2 E}{\partial x \partial y} = \langle PH^{xy}\rangle + \frac{1}{2}\langle P||^{xy}P\rangle + \langle P^yH^x\rangle + \langle P^y||^xP\rangle - \langle PFPS^{xy}\rangle - \langle (PFP)^yS^x\rangle + \gamma^{xy}. \tag{24}$$

At this point the coupled perturbed equations (CPHF for Hartree-Fock and CPKS for DFT) need to be solved. However, in contrast to standard QM theory the Hessian matrix has four blocks: QM-QM, QM-MM, MM-QM, and MM-MM with the QM-MM and MM-QM block being equivalent. At first glance solving the CP equations would appear to be impossible for very large (thousands of atoms) systems, but given the nature of classical point charge terms involving overlap matrix derivatives (S^x, S^{xy}) and two-electron integral derivatives $(||^x, ||^{xy})$can be dropped (Table I). This simplification, combined with improved techniques for solving CP equations[34] lead to the case where problems that were previously impossible to examine are now feasible.

C. Classical, quantum, and hybrid QM/MM transition state searching

In addition to implementing QM/MM analytic second derivatives, we also added eigenvector following routines to the Newton–Raphson minimization routines in CHARMM. This is done in a flexible manner that allows efficient minimization both in and out of a harmonic well. If the system can be approximated harmonically then standard Newton–Raphson minimization occurs making use of the energy (E), gradient of the energy (∇E) and Hessian of the energy ($\nabla^2 E$). However, if the system cannot be approximated as a set of harmonic oscillators then a line search is performed along the specified eigenvector(s). In addition to the three pieces of information already gathered (vide supra) two additional gradient calculations are performed along the positive and negative directions of eigenvector(s). These five pieces of information are then fit to a third order polynomial ($f(x) = ax^3 + bx^2 + cx + d$) and solved via a linear least squares fit procedure.

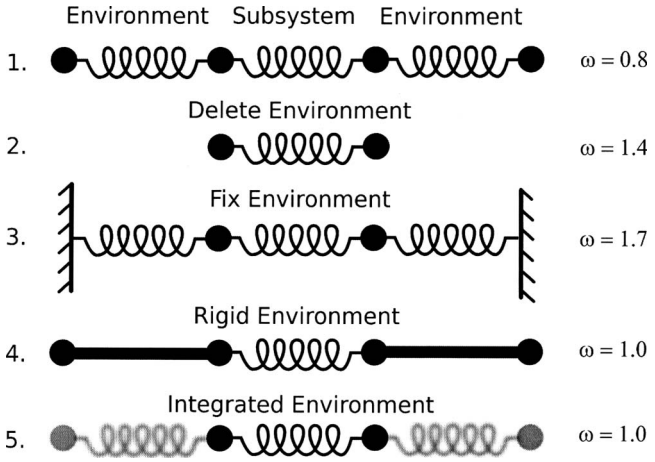

FIG. 1. Illustration and results of harmonic beads on strings. The subsystem is defined as the middle two beads with various treatments applied to the "environment." (1) Full treatment, (2) environment deleted, (3) environment fixed, (4) rigid environment, and (5) environment integrated out (i.e., VSA).

Using classical QM and QM/MM potentials we have adapted the current minimization procedure to change the direction of eigenvector(s) searching. Typically, minimization occurs along all eigenvectors, however, in the new routines saddle points can be searched for by specifying the order of the stationary point that is desired. For example, searching for a transition state (first order saddle point) occurs by determining if the eigenvalue is below a user defined threshold. If this criteria is met then all eigenvectors except that which corresponds to the lowest eigenvalue will be minimized and the specified eigenvector will be maximized. Since eigenvalues are used to determine the mode to follow it is useful to initially perturb the structure in the direction that the desired saddle point is located. This can be done easily with coordinate fixing or restraining potentials in CHARMM or other software.

III. RESULTS AND DISCUSSION

In the following section the results of four test cases are reported: (1) harmonic beads on springs, (2) classical torsional barrier of butane, (3) classical chair to boat inversion of cyclohexane, and (4) quantum mechanical eclipsed-antitorsional barrier of retinol.

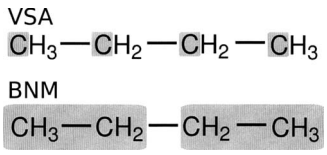

FIG. 2. Illustration of the schemes used to examine the butane anti to eclipsed rotation. The top scheme was employed with the VSA method with the bottom scheme being used with the BNM method.

A. Harmonic beads on springs

The first, and simplest, test system consists of four beads with the mass of each set to 1.0 amu. The beads are connected by springs of force constant $k=1.0$ kcal/mol/Å (Fig. 1). In this example, the subsystem (middle two beads and the spring connecting them) and the environment (the outer two beads and the springs connecting them to the subsystem) are initially defined. Five cases are constructed and examined with frequencies being computed and compared: (1) full explicit system of four beads and three springs; (2) subsystem with environment deleted; (3) full explicit system with environmental beads being fixed (i.e., their masses being set to infinity); (4) rigid environment approximation (i.e., MBH);[35–37] (5) integration of environmental degrees of freedom (i.e., VSA). For each system a Hessian matrix was constructed, mass weighted, and diagonalized to yield the eigenvalues (i.e., frequencies ω).

Results for case 1 are taken as the benchmark and determined to be $\omega=0.8$ cm^{-1}. Case 2, where the environment is deleted results in a near doubling of the frequency yielding $\omega=1.4$ cm^{-1}. Fixing the environment (case 3) enhanced the motion even further and resulted in a frequency of 1.7 cm^{-1}. Approximating the environment as rigid rods (case 4) produced $\omega=1.0$ cm^{-1} which is the same result obtained for the VSA (case 5) treatment where the environment was integrated out. It should be noted that the results ($\omega=1.0$ cm^{-1}) for cases 4 and 5 do not have to be the same. However, given the simplicity of the system and because the masses and springs are symmetric the results work out to be the same. For more complicated systems this will not be the case. It should also be noted that the kinetic energy matrix was included [Eq. (18)] in all results and is clearly needed to account for the inertia of the environment. For example, the frequency of the VSA result would be 2.0 cm^{-1} if inertial effects are ignored.

TABLE II. Vibrational subsystem analysis and BNM results for butane in the global minimum (i.e., anti conformation). All frequencies are listed in cm^{-1}.

VSA Freq.	Full NMA Freq. (% overlap)	BNM freq.	Full NMA Freq. (% overlap)
122	122 (100)	128	122 (100)
294	290 (100)	308	290 (100)
397	392 (100)	445	392 (98)
993	866 (77)	511	219 (82)
	1107 (56)		743 (51)
1047	1012 (80)	711	256 (72)
	985 (52)		829 (49)
			1058 (35)
			1161 (36)
1146	1036 (83)	991	866 (72)
	1351 (45)		1036 (34)
			1107 (58)

TABLE III. Vibrational subsystem analysis and BNM results for butane at the highest transition state (i.e., eclipsed conformation). All frequencies are listed in cm^{-1}.

VSA Freq.	Full NMA Freq. (% overlap)	BNM Freq.	Full NMA Freq. (% overlap)
−165	−175 (99)	−154	−175 (98)
314	310 (100)	320	310 (99)
516	501 (99)	540	304 (82)
			684 (50)
870	814 (92)	569	501 (97)
1093	968 (64)	703	286 (69)
	1074 (65)		1160 (40)
1182	1018 (42)	946	814 (81)
	1110 (70)		1111 (54)

B. Torsional barrier of butane

All atom and VSA NMA calculations are carried out to examine the eclipsed-antibarrier of butane. Block normal mode (BNM) calculations are also performed and compared to VSA results. The different subsystems employed for VSA and BNM calculations are illustrated in Fig. 2 and results are presented in Tables II and III. It should be noted that both the VSA and BNM calculations are set up to produce twelve normal modes which is approximately a 70% reduction in system size. Additionally, the newly developed saddle point optimization routines are used to locate the eclipsed transition state.

We also compute activation free energies (ΔG_{298}) for all atom, VSA and BNM approaches. The VSA and BNM results $\Delta G_{298}=5.5$ and 5.5 kcal/mol, respectively, are in very good agreement with the all atom result, $\Delta G_{298}=5.6$ kcal/mol. Although, the ΔG_{298} results agree well between VSA and BNM it is clear that the VSA method did a better job of reproducing high frequency motion in both the anti and eclipsed rotomers of butane. For example, the eclipsed rotomer where hydrogen bumping is known to play a much larger role in determining conformational free energies, the maximum frequency computed with VSA is 1182 cm^{-1} compared to 946 cm^{-1} with BNM (Table III). This difference can be attributed to the inclusion of the environmental hydrogen motion in the VSA modes, whereas in the BNM procedure the internal motion within blocks is constrained.

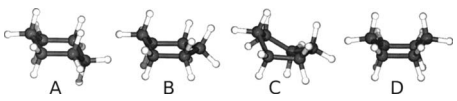

FIG. 3. Illustration of the four extreme points located along the cyclohexane interconversion from the chair to the boat form. The reaction proceeds from left to right with the chair (a) being the global minimum, (b) the global transition state (i.e., reclined chair), (c) a local twist-boat minimum, and (d) the boat form (a local transition state).

TABLE IV. Relative free energies for the boat to chair inversion of cyclohexane. Results are listed for both the standard NMA and VSA. All values are listed in kcal/mol and are computed using the global minimum as the reference. See Fig. 3 for definition of states.

Conformer	Full NMA	VSA
TS [chair (A) - transition state (B)]	9.8	10.1
Twist-boat [chair (A) - twist boat (C)]	6.5	6.7
boat [chair (A) - boat (D)]	7.1	7.3

To further examine mode mapping in the VSA and BNM approaches, mode overlap matrices are computed. This is done by taking the dot product of eigenvectors generated by the VSA and BNM methods with eigenvectors of full NMA. This analysis allows determination of which modes from the full NMA the VSA/BNM results correspond to. Results of this analysis are listed in parentheses in Tables II and III.

C. Chair to boat inversion of cyclohexane

The chair to boat interconversion pathway of cyclohexane is examined using both the all atom and VSA models (Fig. 3). The subsystem employed includes only the ring carbons of cyclohexane which is a 67% reduction in system size. Additionally, we compute the ΔG_{298} between the four extreme points located along the pathway: (1) the global minimum chair conformer, (2) the global maximum/transition state, (3) a local minimum twist-boat conformer, and (4) the boat conformer which is a local transition state. All transition states are located using the newly implemented saddle point finder. The free energies are defined relative to the global minimum (i.e., chair conformer) and are listed in Table IV.

The average errors in the three ΔG_{298} examined is 0.25 kcal/mol with the majority of that being contributed by the global max transition state (0.3 kcal/mol). Using this as a guide, it is expected that hydrogen motion should be most critical at the global transition state. Further analysis and results listed in Table V confirm this hypothesis. For example, examining the internal coordinate derivatives shows that an important mode associated with H–C–C angle bending and H–C–C–C torsional motion does not overlap significantly with the VSA modes at the global TS. In contrast, this mode does map into VSA modes at the chair, twist-boat and boat conformers. This mode encompasses H1–C1 motion which is important because the global TS is a "reclined chair" with all but one ring carbon (C4) being approximately coplanar (Fig. 3). Therefore, the C1 carbon is associated with the major motion of a normal chair becoming a reclined chair. To further demonstrate this, we perform additional VSA calculations where the hydrogens connected to C1 were included in the subsystem and found that the average error in ΔG_{298} dropped from 0.25 to 0.16 kcal/mol, which is in

TABLE V. Cyclohexane boat to chair conversion results comparing full NMA to VSA. All frequencies are listed in cm^{-1} and degenerate modes are designated *a* and *b*.

Chair		Transition state		Twist-boat		Boat	
VSA Freq.	Full NMA Freq. (% overlap)	VSA Freq.	Full NMA Freq. (% overlap)	VSA Freq.	Full NMA Freq. (% overlap)	VSA Freq.	Full NMA Freq. (% overlap)
241_a	236_a (100) 236_b (0)	−219	−234 (100)	93	93 (100)	−88	−88 (100)
241_b	236_b (100) 236_a (0)	24	24 (100)	253	243 (100)	222	215 (100)
402	377 (100)	327	313 (99)	297	286 (100)	314	302 (100)
464_a	454_a (99)	478	462 (98)	481	478 (100)	481	478 (100)
464_b	454_b (99)	480	473 (99)	483	461 (97)	488	467 (97)
675	565 (86) 905 (51)	733	644 (79) 893 (46)	688	598 (87) 888 (44)	689	596 (86) 888 (49)
862	800 (91) 1079 (38)	822	785 (89) 839 (31)	854	797 (85) 932 (51)	850	804 (87) 944 (44)
973_a	880 (68) 938_a (65) 938_b (0)	951	900 (90)	962	906 (86) 888 (41)	954	899 (94)
973_b	880 (68) 938_a (0) 938_b (65)	972	883 (79) 1035 (44)	977	874 (78) 1046 (49)	981	869 (80) 1043 (50)
1144_a	777 (35) 1076_a (85) 1374_a (0)	1171	1072 (76) 1342 (39)	1164	1089 (72) 1372 (46)	1172	1093 (73) 1365 (51)
1144_b	777 (35) 1076_b (0) 1374_b (85)	1196	1100 (75) 1389 (44)	1190	1069 (76) 1376 (53)	1188	1070 (76) 1382 (56)
1200	1079 (76) 1184 (65)	1245	1141 (68) 1342 (34) 1352 (41)	1211	1114 (75) 1337 (65)	1214	1118 (74) 1335 (64)

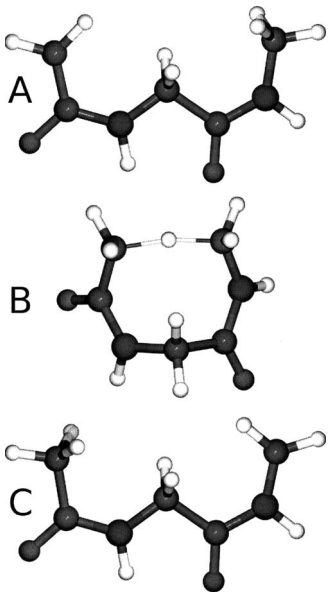

FIG. 4. Illustration of the reactant (a), transition state (b), and product (c) of the hydrogen transfer reaction in the glycine-glycine dipeptide radical.

agreement with the error attached to the other states where the H–C–C and H–C–C–C mode is mapped correctly.

The VSA modes capture the important motions of ring interconversion. Specifically, in the global minimum chair state degenerate modes exist in the full NMA. These modes are correctly reproduced in the VSA results. Modes associated with the following frequencies were determined to be degenerate: 241, 464, 973, and 1144 cm^{-1} (Table V). In an effort to clarify the degeneracy of these modes, we project out (via rotation) the degenerate mode and determine how the two modes overlapped with corresponding VSA modes. This is done by diagonalizing the 2×2 overlap matrix between the two degenerate full Hessian modes and the two degenerate VSA modes. This yields the linear combinations of the full Hessian modes that best overlap with the VSA modes.

D. Quantum mechanical peptide radical rearrangement

The harmonic limit free energy for intramolecular hydrogen atom transfer in a model dipeptide radical (Gly-Gly) is examined using full NMA, VSA, and BNM at the B3LYP/6-31G(d) level of theory (Fig. 4). The definitions of the various subsystems and blocking schemes are illustrated in Fig. 5. The Gly-Gly dipeptide radical has been recently studied as a benchmark for hydrogen transfer reactions in peptides.[38] In the current study, the ΔE, ΔG_{298}, and forward and reverse $\Delta G^{\ddagger}$ are reported. Moran *et al.* found the ΔE and barriers at the G3(MP2)-RAD level of theory to be −5.5, 18.4, and 23.9 kcal/mol, respectively. This is in relatively good agreement with the B3LYP/6-31G(d) results presented in Table VI although B3LYP over stabilizes the reactant by 0.7 kcal/mol [Fig. 4(a)]. The free energies are defined relative to the global minimum.

The harmonic limit free energy of the full system results in a ΔG_{298} of −5.1 kcal/mol and a $\Delta G^{\ddagger}_{\mathrm{fwd}}$ of 20.4 kcal/mol. We next test the VSA method using a small subsystem

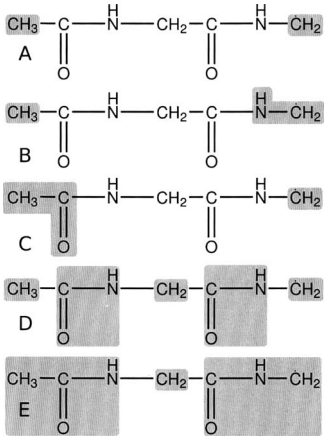

FIG. 5. Definition of subsystem and blocking schemes used to examine the hydrogen transfer reaction in the glycine-glycine dipeptide radical. Schemes (a), (b), and (c) were applied to the vibrational subsystem analysis method while schemes (d) and (e) were employed with the BNM method. See Table VI for results.

[Fig. 5(a)] consisting of only the terminal CH_3 and radical CH_2 groups. This yields good agreement with the results from the full NMA having a 0.7 kcal/mol RMS error (RMS error is determined using both the free energy and free energy of activation) while reducing the system size by over 60% (i.e., reducing the number of modes from 54 down to 21). We also examine two blocking schemes and employ the BNM method. The least restrained of these [Figure 5(d)] consists of five blocks and keeps the transfer hydrogen completely free (total of 33 modes, a nearly 40% increase from the VSA method). The results of this BNM scheme again agree nicely, however, the RMS error increases slightly to 1.0 kcal/mol (nearly a 20% increase over VSA results).

Moran *et al.* found that energetically carbon centered radicals are "more stabilized by electron donation from an adjacent N-atom lone pair than delocalization into an adjacent carbonyl group." Using the VSA method and schemes 5(b) and 5(c) we explore the entropic contributions to this stabilization. Although both schemes 5(b) and 5(c) agree well with full NMA free energy results, there is a significant shift in relative state stabilization. For example, examining results from schemes 5(b) and 5(c) shows a reverse in stabilization of the product state [Fig. 4(c)] relative to the reactant

TABLE VI. QM Full normal mode and VSA free energy results for hydrogen transfer in the glycine-glycine dipeptide radical. $\Delta G/\Delta E$, $\Delta G^{\ddagger}_{\mathrm{fwd}}/\Delta E^{\ddagger}_{\mathrm{fwd}}$ and $\Delta G^{\ddagger}_{\mathrm{rev}}/\Delta E^{\ddagger}_{\mathrm{rev}}$ refer to the (free) energy of reaction, forward, and reverse (free) energy of activation. For comparison standard QM energy of reaction and barriers are listed. All calculations were performed at the B3LYP/6-31G(d) level of theory in Q-CHEM with results reported in kcal/mol.

System	ΔE	$\Delta E^{\ddagger}_{\mathrm{fwd}}$	$\Delta E^{\ddagger}_{\mathrm{rev}}$
Full	−4.8	19.0	23.8
System	ΔG_{298}	$\Delta G^{\ddagger}_{\mathrm{fwd}}$	$\Delta G^{\ddagger}_{\mathrm{rev}}$
Full	−5.1	20.4	25.5
VSA I [Fig. 5(a)]	−5.2	19.3	24.5
VSA II [Fig. 5(b)]	−5.6	19.2	24.9
VSA III [Fig. 5(c)]	−4.8	19.7	24.4
BNM I [Fig. 5(d)]	−5.4	18.9	24.3
BNM II [Fig. 5(e)]	−4.9	18.1	22.9

[Fig. 4(a)]. More specifically, when the N–H group is included in the subsystem the product state is destabilized, however, the reactant state is more destabilized. This leads to a lowering for the forward and reverse barriers while ΔG_{298} is increased. In contrast, when the C=O group is included in the subsystem the product state is destabilized more so than the reactant state. Again, this leads to both the forward and reverse barriers being lowered, but with the ΔG_{298} also being decreased.

The results obtained for this model system, employing VSA coupled with quantum mechanics, are not only able to extend the understanding of radical biochemistry but also serve as an example of the powerful and accurate analysis that can be undertaken using these new techniques. For example, comparing ΔG_{298} for scheme 5(b) (−5.6 kcal/mol) and scheme 5(c) (−4.8 kcal/mol) indicates that inclusion of the N–H group in the subsystem has a stabilizing effect on the reaction whereas inclusion of the C=O group actually exerts a destabilization. This tends to confirm the assertion of Moran *et al.* that the N–H group is more beneficial to the stabilization of the Gly-Gly radical.

IV. CONCLUSIONS

In the current work we presented the detailed derivation, implementation, and testing of the VSA methodology and the coupling of this with quantum mechanical and newly implemented hybrid QM/MM analytic second derivative techniques. Four illustrative examples were presented, ranging from simply solved analytical models (the harmonic beads) to complicated isomerizations where localized hydrogen motion was analyzed and determined to be critical to complicated radical rearrangements that employed QM methods coupled with full normal mode analysis, BNM analysis, and VSA.

We reiterate that the VSA method will be useful in a variety of situations: examination of local-global motion, performing accurate NMA while eliminating unwanted degrees of freedom, eliminating excess noise from large NMA (i.e., QM/MM), employing dynamic simulations coupled with VSA via quasi harmonic analysis, and integration of light particles during NMA (e.g., application to polarizable models).

Using the test cases designed and employed we were able to illustrate that the VSA method performs at least as well as the BNM and in many cases can outperform it while using substantially smaller subsystems (i.e., with fewer modes included). However, we feel that the primary strength of the VSA method is its use as an analysis tool. As stated previously, the VSA can be employed to study complicated local-global couplings in proteins using course grained models, but we have also shown that the VSA can be a powerful tool when analyzing complicated electronic systems. For example, using the VSA method we were able to demonstrate that carbon centered radicals in peptides are not only governed by electronic stabilization but also will be influenced by entropic factors related to motion of electronically important substituents. The VSA as an analysis tool has limitless possibilities.

ACKNOWLEDGMENTS

We thank Damian Moran from Macquarie University for helpful insight and fruitful discussions. This research was supported, in part, by the Intramural Research Program of the NIH, NHLBI. Additionally, A.G. was supported by the Fund for Scientific Research—Flanders. Y.S. and J.K. wish to acknowledge the financial support from the National Institutes of Health through SBIR Grant No. GM065617. Use of the LoBoS (www.lobos.nih.gov) supercomputer is also acknowledged and appreciated.

[1] R. Elber, Curr. Opin. Struct. Biol. **15**, 151 (2005).
[2] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996).
[3] T. Haliloglu, I. Bahar, and B. Erman, Phys. Rev. Lett. **79**, 3090 (1997).
[4] I. Bahar, A. R. Atilgan, and B. Erman, Folding Des. **2**, 173 (1997).
[5] B. R. Brooks, D. Janezic, and M. Karplus, J. Comput. Chem. **16**, 1522 (1995).
[6] D. Janezic and B. R. Brooks, J. Comput. Chem. **16**, 1543 (1995).
[7] D. Janezic, R. M. Venable, and B. R. Brooks, J. Comput. Chem. **16**, 1554 (1995).
[8] N. Go, T. Noguti, and T. Nishikawa, Proc. Natl. Acad. Sci. U.S.A. **80**, 3696 (1983).
[9] E. C. Dykeman and O. F. Sankey, Phys. Rev. Lett. **100**, 028101 (2008).
[10] D. A. Case, Curr. Opin. Struct. Biol. **4**, 285 (1994).
[11] A. Kitao and N. Go, Curr. Opin. Struct. Biol. **9**, 164 (1999).
[12] J. P. Ma, Structure (London) **13**, 373 (2005).
[13] A. Amadei, A. B. Linssen, and H. J. Berendsen, Proteins **17**, 412 (1993).
[14] B. R. Brooks and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **80**, 6571 (1983).
[15] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, Biophys. J. **80**, 505 (2001).
[16] B. Isin, P. Doruker, and I. Bahar, Biophys. J. **82**, 569 (2002).
[17] O. Keskin, S. R. Durell, I. Bahar, R. L. Jernigan, and D. G. Covell, Biophys. J. **83**, 663 (2002).
[18] S. Kundu and R. L. Jernigan, Biophys. J. **86**, 3846 (2004).
[19] F. Tama and Y. H. Sanejouand, Protein Eng. **14**, 1 (2001).
[20] C. Xu, D. Tobi, and I. Bahar, J. Mol. Biol. **333**, 153 (2003).
[21] W. J. Zheng and S. Doniach, Proc. Natl. Acad. Sci. U.S.A. **100**, 13253 (2003).
[22] W. J. Zheng and B. R. Brooks, J. Mol. Biol. **346**, 745 (2005).
[23] M. Delarue and Y. H. Sanejouand, J. Mol. Biol. **320**, 1011 (2002).
[24] W. G. Krebs, V. Alexandrov, C. A. Wilson, N. Echols, H. Y. Yu, and M. Gerstein, Proteins **48**, 682 (2002).
[25] W. J. Zheng and B. R. Brooks, Biophys. J. **88**, 3109 (2005).
[26] W. Z. Zheng and B. R. Brooks, Biophys. J. **89**, 167 (2005).
[27] M. Y. Lu, B. Poon, and J. P. Ma, J. Chem. Theory Comput. **2**, 464 (2006).
[28] G. H. Li and Q. Cui, Biophys. J. **86**, 743 (2004).
[29] K. Fukui, J. Phys. Chem. **74**, 4161 (1970).
[30] C. Gonzalez and H. B. Schlegel, J. Chem. Phys. **90**, 2154 (1989).
[31] Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. Distasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Daschel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C.-P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, D. C. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, Phys. Chem. Chem. Phys. **8**, 3172 (2006).
[32] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaninathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[33] Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).

[34] M. Frisch, M. Head-Gordon, and J. Pople, Chem. Phys. **141**, 189 (1990).
[35] A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, J. Chem. Phys. **126**, 224102 (2007).
[36] A. Ghysels, D. Van Neck, and M. Waroquier, J. Chem. Phys. **127**, 164108 (2007).
[37] A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier, J. Chem. Theory Comput. **4**, 614 (2008).
[38] D. Moran, R. Jacob, G. P. F. Wood, M. L. Coote, M. J. Davies, R. A. J. O'Hair, C. J. Easton, and L. Radom, Helv. Chim. Acta **89**, 2254 (2006).