JCTC Journal of Chemical Theory and Computation

Efficient Calculation of QM/MM Frequencies with the Mobile Block Hessian

An Ghysels,*[,†] H. Lee Woodcock III,*[,‡] Joseph D. Larkin,[§] Benjamin T. Miller,[§] Yihan Shao,[§,⊥] Jing Kong,[⊥] Dimitri Van Neck,[†] Veronique Van Speybroeck,[†] Michel Waroquier,[†] and Bernard R. Brooks[§]

Center for Molecular Modeling, Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium, Department of Chemistry, University of South Florida, 4202 E. Fowler Avenue, CHE 205, Tampa, Florida 33620-5240, United States, Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, United States, and Q-Chem Inc., 5001 Baum Blvd, Suite 690, Pittsburgh, Pennsylvania 15213, United States

Received August 23, 2010

Abstract: The calculation of the analytical second derivative matrix (Hessian) is the bottleneck for vibrational analysis in QM/MM systems when an electrostatic embedding scheme is employed. Even with a small number of QM atoms in the system, the presence of MM atoms increases the computational cost dramatically: the long-range Coulomb interactions require that additional coupled perturbed self-consistent field (CPSCF) equations need to be solved for each MM atom displacement. This paper presents an extension to the Mobile Block Hessian (MBH) formalism for QM/MM calculations with blocks in the MM region and its implementation in a parallel version of the Q-Chem/CHARMM interface. MBH reduces both the CPU time and the memory requirements compared to the standard full Hessian QM/MM analysis, without the need to use a cutoff distance for the electrostatic interactions. Special attention is given to the treatment of link atoms which are usually present when the QM/MM border cuts through a covalent bond. Computational efficiency improvements are highlighted using a reduced chorismate mutase enzyme system, consisting of 24 QM atoms and 306 MM atoms, as a test example. In addition, the drug bortezomib, used for cancer treatment of myeloma, has been studied as a test case with multiple MBH block choices and both a QM and QM/MM description. The accuracy of the calculated Hessians is quantified by imposing Eckart constraints, which allows for the assessment of numerical errors in second derivative procedures. The results show that MBH within the QM/MM description not only is a computationally attractive method but also produces accurate results.

I. Introduction

Normal mode analysis (NMA) is a well-known technique which estimates the intrinsic vibrational frequencies of chemical systems by assuming a harmonic shape for the potential energy surface. Despite its simplicity, it is still a popular and effective approach for predicting vibrational IR and Raman spectra,[1] for identifying chemical groups,[2] or for studying the large-amplitude collective motions involved in conformational changes of biomolecules.[3] This method is based on the diagonalization of the Hessian matrix, which contains the second derivatives of the potential energy with respect to the nuclear coordinates. The calculation of this $3N_{AT} \times 3N_{AT}$ Hessian, where N_{AT} is the number of atoms,

* Corresponding author e-mail: an.ghysels@ugent.be (A.G.), hlw@mail.usf.edu (H.L.W.).
† Ghent University.
‡ University of South Florida.
§ National Institutes of Health.
⊥ Q-Chem Inc.

10.1021/ct100473f
Published on Web 01/06/2011

is computationally expensive when a quantum mechanical (QM) description is used, because a set of $3N_{AT}$ coupled perturbed self-consistent field (CPSCF) equations needs to be solved.[4−11] In contrast, the computational load of the second derivative calculation is in comparison extremely cheap in the molecular mechanics (MM) description, but force fields cannot, in general, be used to investigate chemical reactions where the change in electron density (i.e., bond making/breaking, radical processes, etc.) is a purely quantum mechanical phenomenon.

Hybrid QM/MM models aim at combining the best features of the QM and MM models: the quantum descriptions necessary for chemistry and the computational advantages of force fields. The QM/MM approach partitions the system into a QM region for the chemically interesting site and an MM region for the surrounding chemical environment.[12−15] The effective cost of a QM/MM Hessian calculation depends heavily on the treatment of the electrostatics between the QM and MM region, for which two schemes have been developed.

In a subtractive scheme like the original version of ONIOM, the potential energy is the sum of the MM energy of the whole system, plus the QM energy of the QM region, minus the MM energy of the QM region as a correction for double counting.[16,17] During the QM part of the calculation, the QM atoms are unaware of the existence of the MM atoms, and thus the electron cloud in the QM region is not influenced by the MM partial charges, i.e., mechanical embedding. As a consequence, the displacement of an MM atom does not cause a change in the QM wave function, such that the corresponding derivatives are simply equal to zero, and there is no need to solve CPSCF equations for MM atom displacements.

In an additive scheme, however, as implemented in the Q-Chem/CHARMM interface[18−20] and in many other interfaces, the potential energy consists of the QM energy of the QM atoms, the MM energy of the MM atoms, and the Coulomb and van der Waals interaction energy between QM and MM atoms.[12,21,22] Such a description provides a more accurate treatment of the long-range electrostatics, which is invaluable when studying, for example, reactions and molecular configurations. This idea has also been applied to the original ONIOM scheme to account for the polarization effects from the MM region.[23] In the additive scheme, every displacement of an MM atom leads to an additional CPSCF equation.[24] Even when the number of QM atoms is low, the QM/MM interaction term in the Hamiltonian makes the Hessian determination too costly for systems with a large number of MM atoms.

Recently, the mobile block Hessian approach was developed[25−27] to calculate frequencies in a partially optimized structure. The method groups atoms into blocks which are restricted to rigid motions during the vibrational analysis. The internal geometry is fixed, but each block is still allowed to translate or rotate as a whole. Consequently, block motions replace the individual atom motions in the CPSCF equations, thus reducing the number of CPSCF equations. Until now, no implementation was available that exploits the computational advantage offered by solving CPSCF equations of reduced dimension. In this paper, we present such an implementation for the specific case of blocks chosen in the MM region.

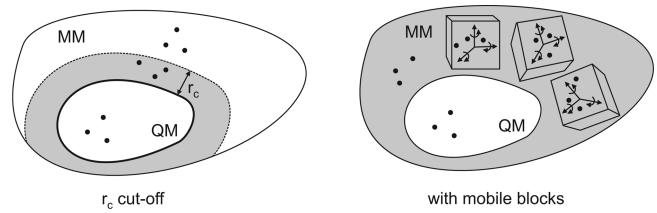

Figure 1. Cut-off technique versus mobile block Hessian approach. (a) When using a cutoff r_c to decrease the computational cost of the CPSCF, the electrostatic interaction between QM and MM atoms is neglected beyond the distance r_c. MM atoms outside the gray region do not interact with the QM atoms. (b) In the mobile block Hessian approach, all electrostatic interaction is still present. Blocks in the MM region are restricted to rigid body motions, i.e. translations and rotations of each block.

An alternative approach to economize on the number of CPSCF equations is the use of a cutoff distance r_c beyond which electrostatics between QM and MM atoms are neglected. Figure 1 illustrates the influence of a cutoff: the electrons in the QM region do not interact (Coulomb) with MM partial charges outside the gray cutoff zone. The MBH has the advantage that it does not neglect electrostatic interactions but rather restricts motions: the gray zone for MBH encompasses the whole MM region in Figure 1. Restricting the motion of distant blocks may change the overall vibrational free energy, but such effects are expected to largely cancel out when treated consistently in a thermodynamic cycle. It is the aim of this paper to show that the combination of MBH with the QM/MM description is a highly accurate and efficient approach. It therefore becomes the ideal alternative to the standard full Hessian calculation when the latter is no longer feasible. The Q-Chem/CHARMM interface now has a working parallel version for both the full QM/MM Hessian and mobile block QM/MM Hessian calculation.[28]

The following section presents the theoretical background on which the idea of MBH in a QM/MM description is based. First, the NMA equations, the frequency calculation in QM/MM, and the MBH equations are reviewed. Second, the adaptation needed for an efficient mobile block Hessian computation is outlined. Moreover, the treatment of multiple link atoms is clarified when the QM/MM border cuts through covalent bonds. It is also pointed out that MBH preserves the long-range electrostatic interactions, in contrast to the alternative approach with a cutoff distance r_c which induces an error decaying as slowly as $1/r_c^3$. The third section presents computational results of the chorismate mutase enzyme. This test case illustrates how MBH and parallellization reduce the memory requirements and CPU timings. In the fourth section, the oxidation of the bortezomib molecule is treated as a test case for the newly implemented method. This drug is used in cancer treatment since it inhibits the function of proteasomes upon binding, ultimately leading to cell death.[29,30] By imposing the Eckart constraints, the accuracy of the calculated Hessians is estimated. This

accuracy is then compared with the influence of various frequency treatments on the vibrational free energy differences: QM versus QM/MM and full Hessian versus MBH.

II. Theory and Implementation

II.A. Normal Mode Analysis. Assume that the positions of the N_{AT} atoms are described by Cartesian displacement coordinates, labeled $x = 1, ..., 3N_{AT}$, all with respect to a reference structure. A second order approximation of the potential energy surface around the reference structure is then equal to

$$V(x) \approx V(0) + G^T x + \frac{1}{2} x^T H x \tag{1}$$

where the reference energy $V(0)$ can be set to be zero. The $3N_{AT}$ dimensional gradient vector G contains the first derivatives, and the Hessian H is the $3N_{AT} \times 3N_{AT}$ matrix containing the second derivatives evaluated at the reference point. When calculating normal modes, the reference structure should be a stationary point on the potential energy surface, i.e. $G = 0$. Introducing the diagonal mass matrix M with the atomic masses on the diagonal, the normal-mode analysis (NMA) equations read

$$Hv = \omega^2 M v \tag{2}$$

Solving the NMA equations yields the eigenvalues ω^2 (frequency is $\nu = \omega/2\pi$) corresponding to the eigenvectors v.

Six frequencies should be zero because of the translational and rotational invariance of an isolated gas molecule (five for linear molecules). The corresponding normal modes represent global translations and rotations of the complete molecular system. It is possible to project out those zero frequency vectors before diagonalization, since their exact format is known. This projection amounts to imposing the Eckart constraints[26,31] and guarantees the presence of six frequencies that are identically zero even when the system is not perfectly at the stationary point on the energy surface or when the Hessian elements are inaccurate. The effect of the Eckart constraints on the frequencies is studied for the bortezomib example in section IV.

II.B. The QM/MM Full Hessian. In the additive scheme with electrostatic embedding, the system is separated into a QM and an MM region.[15] The QM region consists of N_{QM} nuclei and N_e electrons, described quantum mechanically in the Born−Oppenheimer approximation. The MM region contains N_{MM} partial charges, described classically, with $N_{QM} + N_{MM} = N_{AT}$. The Hamiltonian of the system of interacting QM and MM particles is written as

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}^{QM} + \hat{\mathcal{H}}^{QM/MM} + \mathcal{H}^{MM} \tag{3}$$

where $\hat{\mathcal{H}}^{QM}$ represents the Hamiltonian describing the QM region, i.e., the electronic kinetic energy and all electrostatic potentials generated by the electrons and QM nuclei. The QM/MM interaction Hamiltonian is

$$\hat{\mathcal{H}}^{QM/MM} = -\sum_{i=1}^{N_e}\sum_{k=1}^{N_{MM}} \frac{q_k}{|\boldsymbol{r}_i - \boldsymbol{R}_k|} + \sum_{n=1}^{N_{QM}}\sum_{k=1}^{N_{MM}} \frac{q_k Z_n}{|\boldsymbol{R}_n - \boldsymbol{R}_k|} + \sum_{n=1}^{N_{QM}}\sum_{k=1}^{N_{MM}} \left(\frac{A_{n,k}}{|\boldsymbol{R}_n - \boldsymbol{R}_k|^6} - \frac{B_{n,k}}{|\boldsymbol{R}_n - \boldsymbol{R}_k|^{12}} \right) \tag{4}$$

with q_k being the partial charge of the MM atom at position $\boldsymbol{R}_k$, Z_n the nuclear charge of the QM atom at position $\boldsymbol{R}_n$, $\boldsymbol{r}_i$ the electron positions, and $A_{n,k}$ and $B_{n,k}$ the van der Waals parameters. The QM/MM interaction Hamiltonian consists of the Coulomb interaction between MM charges and QM electrons, the Coulomb interaction between MM charges and QM nuclei, and the van der Waals interaction between MM atoms and QM atoms. The total energy of the system is thus given by

$$\begin{aligned} E_{tot} &= \langle \Phi | \hat{\mathcal{H}}^{QM} - \sum_{i=1}^{N_e}\sum_{k=1}^{N_{MM}} \frac{q_k}{|\boldsymbol{r}_i - \boldsymbol{R}_k|} | \Phi \rangle \\ &\quad + E_{nuc}^{QM/MM} + E_{vdW}^{QM/MM} + E^{MM} \\ &= E_{el}^{QM} + E_{el}^{QM/MM} + E_{nuc}^{QM/MM} + E_{vdW}^{QM/MM} + E^{MM} \end{aligned} \tag{5}$$

where $|\Phi\rangle$ is the electronic wave function for the QM atoms. The two electronic terms are calculated quantum mechanically and are referred to as the quantum part (denoted "quant"). The remaining three terms in this expression are classical (denoted "class"). This defines our decomposition of the total energy in a quantum and classical part:

$$E_{tot} = E_{quant} + E_{class} \tag{6}$$

The Cartesian Hessian H expresses the response of the total energy to $3N_{AT}$ Cartesian displacements. A general Hessian element is denoted as $H_{xy} = E_{tot}^{xy} = \partial^2 E_{tot}/\partial x \partial y$ ($x, y = 1, ..., 3N_{AT}$), with the superscripts referring to derivatives. The Hessian can be divided into submatrices as shown in Figure 2, depending on whether the x, y indices correspond to the QM−QM, MM−MM, or mixed QM−MM displacements. Not all terms in eq 5 contribute to each subblock of the Hessian: the derivatives of $E_{el,\,QM}$ and E_{MM} only contribute to the QM or MM subblock, respectively. However, the QM/MM interaction terms contribute to all subblocks of the Hessian. While the derivatives of the classical terms $E_{nuc,\,QM/MM}$ and $E_{vdW,\,QM/MM}$ are relatively easy to evaluate, the $E_{el,\,QM/MM}$ derivatives dominate the cost of the Hessian evaluation. Even if the number of QM atoms is low, the cost of the calculation still scales with N_{AT}. The main reason is that each MM atom adds three perturbations to the CPSCF equations, because the displacement of an external charge (an MM atom) leads to a change in the electronic wave function and the charge distribution. Section II.F shows how the number of perturbations can be reduced by the introduction of mobile blocks.

Methods based on the variational principle, such as Hartree-Fock or Kohn−Sham DFT, have the advantage that the $(2n + 1)$th derivative of the energy can be constructed from the nth derivative of the variational parameters (Wigner's $2n + 1$ theorem[32]). The variational

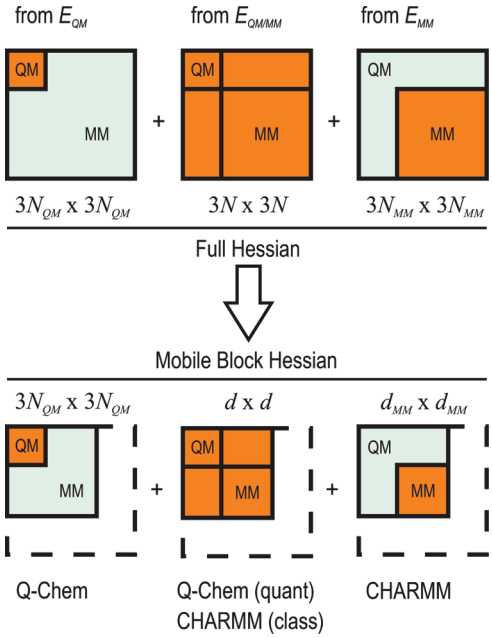

Figure 2. $\hat{\mathcal{H}}_{QM}$ and $\hat{\mathcal{H}}_{MM}$ contribute to the derivatives with respect to N_{QM} QM and N_{MM} MM atoms, respectively, whereas $\hat{\mathcal{H}}_{QM/MM}$ contributes to all Hessian elements. The mobile block Hessian H^{mb}, with blocks chosen in the MM region, is smaller in size ($d \times d$, $d = 3N_{QM} + d_{MM}$, bottom three panels) than the full Hessian H ($3N_{AT} \times 3N_{AT}$, upper three panels).

parameters are the elements of Θ that describe a unitary rotation among the molecular orbitals.[33] The parameters Θ are updated iteratively until the corresponding molecular orbitals are eigenfunctions of, e.g., the Fock operator in the case of Hartree−Fock-based methods. In accordance with Wigner's $2n + 1$ theorem, only the first derivatives of the Θ are needed for the construction of the second derivatives of the energy.

The specific discussion below applies to the Q-Chem/CHARMM interface, but in principle each QM and MM package with a suitable interface can be used for the construction of the full QM/MM Hessian. One should be aware that the implementation details may differ slightly depending on the choice of the QM and MM code. The implementation of the full QM/MM Hessian starts with the construction of the quantum contribution by the QM code. Using a compact notation,[34,35] where $\langle ... \rangle$ denotes the trace of a matrix, the Hartree−Fock energy E (corresponding to the first two terms in eq 5) calculated by the QM code reads

$$E = \langle PH_{\text{core}} \rangle + \frac{1}{2}\langle P\Pi P \rangle + \gamma = E(H_{\text{core}}, \Pi, S, \Theta) \quad (7)$$

where P ($= P(\Theta, S)$) is the density matrix, H_{core} is the core Hamiltonian matrix (including the Coulomb potential due to the MM partial charges, which corresponds to the first term in eq 4), Π represents the antisymmetrized two-electron integrals over spin orbitals, S is the overlap matrix, and γ is the nuclear repulsion energy of the QM atoms. The Fock operator is then defined as

$$F = H_{\text{core}} + P\Pi \quad (8)$$

and the standard self-consistent field (SCF) convergence criterion $E^{\Theta} = 0$ reads

$$FPS = SPF \quad (9)$$

The second derivatives with respect to atomic displacements are given by[11,33,34,36]

$$\frac{\partial^2 E}{\partial x \partial y} = E^{xy} = \langle PH^{xy}_{\text{core}} \rangle + \frac{1}{2}\langle P\Pi^{xy}P \rangle + \langle P^x H^y_{\text{core}} \rangle + \langle P^x \Pi^y P \rangle - \langle (PFP)S^{xy} \rangle - \langle (PFP)^x S^y \rangle + \gamma^{xy} \quad (10)$$

and require the density matrix response P^x, which is obtained by solving the coupled perturbated self-consistent field (CPSCF) equations for Θ^x. Since the energy is obtained by a variational method, the CPSCF equations can be derived from the identity $E^{\Theta} \equiv 0$:

$$(E^{\Theta})^x = E^{\Theta\Theta}\Theta^x + E^{\Theta H}H^x_{\text{core}} + E^{\Theta\Pi}\Pi^x + E^{\Theta S}S^x \equiv 0 \quad (11)$$

The calculation of the derivatives takes five steps:

(1) Construct $E^{\Theta H}H^x_{\text{core}} + E^{\Theta\Pi}\Pi^x + E^{\Theta S}S^x$.
(2) Solve CPSCF eq 11 for Θ^x.
(3) From Θ^x, construct P^x.
(4) From P^x, construct $F^x = H^x_{\text{core}} + \Pi^x P + \Pi P^x$.
(5) Construct E^{xy} according to eq 10.

As a result of the explicit QM/MM polarization effects, the CPSCF equations include perturbations for each MM atom, which makes step 2 and step 4 the most demanding. One should also pay attention to the memory requirements, which peak in step 2 and 4 because they scale as $6N_{MM}n_b^2$, where n_b is the number of basis functions in the basis set. To make the calculation more efficient, we have parallellized the quantum mechanical part (contribution from E^{QM}_{el} and $E^{QM/MM}_{el}$) of the full Hessian calculation. Section III discusses the timings and memory estimates in more detail using the chorismate mutase enzyme as a test system.

In the practical implementation, the next step involves passing the quantum mechanical information from Q-Chem back to CHARMM where the remaining classical terms are constructed and added to the Hessian at a relatively insignificant cost. The full QM/MM Hessian is then mass-weighted and diagonalized to obtain the frequencies and modes. The above discussion holds for Hartree−Fock calculations. The implementation for DFT is similar. The main difference is that for DFT one more term should be added to steps 1, 4, and 5 to account for the derivatives involving the exchange-correlation functional.

II.C. MBH Theory. The MBH method partitions the system into blocks of atoms.[25,26] During the geometry optimization, the position and orientation of each block are optimized, while the internal geometry of the blocks is not necessarily optimized. As a result of this partial optimization, there might be residual forces between the atoms within a block. Whereas in the subsequent vibrational analysis spurious imaginary frequencies might appear when applying the standard full Hessian NMA, the MBH is capable of reproducing physical frequencies.[25,26] The internal coordinates within multi-atom blocks are kept fixed, such that the

MBH model considers only a subset of the degrees of freedom (d). A single-atom block (free atom) is still described by its three Cartesian displacements, while a multi-atom block needs six parameters to describe the position and orientation (linear blocks are not considered here). Within our current implementation, all QM atoms are free, while MM atoms can either be free or become part of a multi-atom block. If d_b denotes the number of parameters of block b and N_{BL} is the number of blocks, the total number of parameters is given by

$$d = \sum_{b=1}^{N_{\mathrm{BL}}} d_b \tag{12}$$

with $d_b = 3$ for a single-atom block or $d_b = 6$ for a nonlinear block. A suitable choice of the block parameters is the convention introduced in ref 26. The parameters p_α, $\alpha = 1, ..., 6$, of a particular block give the position $\boldsymbol{r}'$ of each atom of the block with respect to the reference geometry $\boldsymbol{r}$ by successive rotations around the fixed z, y, and x axes of a space-fixed frame (p_6, p_5, p_4), followed by a translation (p_1, p_2, p_3):

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos p_4 & -\sin p_4 \\ 0 & \sin p_4 & \cos p_4 \end{pmatrix} \begin{pmatrix} \cos p_5 & 0 & \sin p_5 \\ 0 & 1 & 0 \\ -\sin p_5 & 0 & \cos p_5 \end{pmatrix} \times \begin{pmatrix} \cos p_6 & -\sin p_6 & 0 \\ \sin p_6 & \cos p_6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{13}$$

The position of a single-atom block only needs the parameters that describe translation. In the following, we refer to the new set of dynamical variables with indices $p = 1, ..., d$. Useful quantities are the first (Jacobian) and second derivatives of the transformation between the Cartesian displacement coordinates and the block parameters, evaluated at the reference geometry:

$$T_{xp} = \frac{\partial x}{\partial p}; \quad C^{(x)}_{pp'} = \frac{\partial^2 x}{\partial p \partial p'} \tag{14}$$

The explicit expressions for the transformation matrices T (dimension $3N_{\mathrm{AT}} \times d$) and $C^{(x)}$, $x = 1, ..., 3N_{\mathrm{AT}}$ (each of dimension $d \times d$) have been derived in refs 26 and 37.

The second derivatives evaluated at the reference point with respect to the set of d parameters yield the mobile block Hessian H^{mb}, whose elements are defined as

$$H^{\mathrm{mb}}_{pp'} = \frac{\partial^2 E_{\mathrm{tot}}}{\partial p \partial p'} = E^{pp'}_{\mathrm{tot}} \tag{15}$$

The MBH elements are now related to the full Cartesian Hessian by the following transform

$$H^{\mathrm{mb}} = T^T H T + \sum_{x=1}^{3N_{\mathrm{AT}}} G_x C^{(x)} \tag{16}$$

and similarly for the gradient

$$G^{\mathrm{mb}} = T^T G = 0 \tag{17}$$

Note that, in the case of a partially optimized structure, the Cartesian gradient G might differ from zero, but the MB gradient G^{mb} should be zero because all block parameters are supposed to be optimized.

II.D. The QM/MM Mobile Block Hessian. By combining the QM/MM description with the mobile block concept, a considerable reduction in memory and CPU time becomes possible. This section explains the modifications of the CPSCF and the changes to the QM/MM interface that are required to realize the computational profit.

The decomposition in the classical and quantum terms of the total energy (see eq 6) leads to a similar decomposition of the gradient and the Hessian:

$$G = G_{\mathrm{class}} + G_{\mathrm{quant}} \tag{18}$$

$$H = H_{\mathrm{class}} + H_{\mathrm{quant}} \tag{19}$$

where, for instance, $G_{\mathrm{class},x} = \partial E_{\mathrm{class}}/\partial x$ and similar expressions hold for G_{quant}, H_{class}, and H_{quant}. To obtain the mobile block Hessian from the standard Cartesian Hessian calculation, the same decomposition is applied to eq 16:

$$\begin{aligned} H^{\mathrm{mb}} &= H^{\mathrm{mb}}_{\mathrm{class}} + H^{\mathrm{mb}}_{\mathrm{quant}} \\ &= T^T(H_{\mathrm{class}} + H_{\mathrm{quant}})T + \sum_{x=1}^{3N_{\mathrm{AT}}} (G_{\mathrm{class},x} + G_{\mathrm{quant},x})C^{(x)} \end{aligned} \tag{20}$$

As already mentioned, the classical part of the Hessian is computationally less demanding, but the quantum part of the Hessian is expensive. Instead of constructing the full quantum Hessian H_{quant} and then projecting it to the smaller mobile block Hessian dimension (eq 16), a substantial reduction in CPU time is obtained by directly constructing the mobile block Hessian for the quantum part, $H^{\mathrm{mb}}_{\mathrm{quant}}$. The elements are similar to those in eq 10, with block displacements p as perturbations instead of Cartesian displacements x. The density matrix response P^p is obtained by solving adapted CPSCF equations, $(E^{\Theta})^p = \Sigma_x (E^{\Theta})^x T_{xp} \equiv 0$. Similar to the full Hessian calculation, the construction of the derivatives occurs in five steps. However, at the end of step 1, the terms are projected with the T transform of eq 14, such that Θ^p can be solved from the adapted CPSCF:

$$E^{\Theta\Theta}\Theta^p = -\sum_x (E^{\Theta H} H^x_{\mathrm{core}} + E^{\Theta\Pi}\Pi^x + E^{\Theta S} S^x) T_{xp} \tag{21}$$

When the p index denotes a parameter of a block in the MM region, the summation over x is reduced to the Cartesian displacements of the MM atoms within the block only, such that the right-hand side of the transform greatly simplifies to

$$E^{\Theta\Theta}\Theta^p = -\sum_x E^{\Theta H} H^x_{\mathrm{core}} T_{xp} \tag{22}$$

After the transform to p variables, the rest of the steps are all very similar with x replaced by p.

In the remainder of this section, it is discussed how the QM code interacts with the MM code; it is, for instance, essential to add and project the matrices in the correct order.

The QM code returns a matrix of dimension $d \times d$, with d the total number of perturbations. When assembling the Hessian in step 5, the outcome is not the MB Hessian $H_{\text{quant}}^{\text{mb}}$ as in eq 20, but only the bilinear part

$$T^T H_{\text{quant}} T = H_{\text{quant}}^{\text{mb,bil}} \tag{23}$$

is obtained by construction. The gradient correction is still lacking but will be added at the end (see further, eq 26). To add the QM code contribution correctly to the MM code contribution, the Q-Chem Hessian $H_{\text{quant}}^{\text{mb, bil}}$ is first transformed to a matrix of the standard size $3N_{\text{AT}} \times 3N_{\text{AT}}$ by a linear transform Q of dimension $d \times 3N_{\text{AT}}$:

$$H'^{\text{bil}}_{\text{quant}} = Q^T H_{\text{quant}}^{\text{mb,bil}} Q \tag{24}$$

Here, Q is the pseudoinverse of the rectangular transform matrix T, such that $QT = 1$ and

$$T^T H'^{\text{bil}}_{\text{quant}} T = T^T Q^T H_{\text{quant}}^{\text{mb,bil}} QT = H_{\text{quant}}^{\text{mb,bil}} \equiv T^T H_{\text{quant}} T \tag{25}$$

The final expression for the mobile block Hessian reads

$$H^{\text{mb}} = T^T H_{\text{class}} T + T^T Q^T H_{\text{quant}}^{\text{mb,bil}} QT + \sum_{x=1}^{3N_{AT}} (G_{\text{class},x} + G_{\text{quant},x}) C^{(x)} \tag{26}$$

where the QM code (Q-Chem) calculates $H_{\text{quant}}^{\text{mb, bil}}$ and G_{quant} and performs the Q transform, while the MM code (CHARMM) calculates H_{class} and G_{class}, assembles classical and quantum parts, and performs the T transform and the gradient correction. In practice, the $E_{\text{nuc}}^{\text{QM/MM}}$ of eq 5 is calculated by Q-Chem, but in general, this is a purely classical term which could be calculated by either the MM or QM code.

The set of linear equations to be solved in step 2 (see eq 21) now counts d equations instead of $3N_{\text{AT}}$, leading to a computational profit especially on the level of memory requirements. Moreover, the code implemented in Q-Chem also works in parallel. Using N_{P} processors, the peak memory for MBH is equal to $2d_{\text{MM}}{n_{\text{b}}}^2/N_{\text{P}}$, where $d_{\text{MM}} = 6N_{\text{BL}} + 3N_{\text{MM, free}}$ is the number of MM perturbations, compared to the peak memory $6N_{\text{MM}}{n_{\text{b}}}^2/N_{\text{P}}$ for a full QM/MM Hessian calculation. The main difference with respect to the full Hessian calculation is that the factor $3N_{\text{MM}}$ is reduced to d_{MM}, which will generally be significantly smaller. The MBH also reduces the CPU time since fewer CPSCF equations need to be constructed and solved. This means that MBH allows vibrational analysis to be performed on larger systems than is feasible with the full Hessian. In section III, concrete timings and memory estimates are discussed for the chorismate mutase test system.

II.E. Treatment of Link Atoms. It is a common practice to introduce at least one link atom when cutting a bond across the QM/MM boundaries in the electrostatical embedding scheme.[15,38] To make the MBH available for a broader range of applications, our implementation has been extended to be able to treat link atoms correctly. The introduction of a link atom generates three additional degrees of freedom, leading to an extended potential energy surface $\tilde{V}(\boldsymbol{R}, \boldsymbol{R}_{\text{LK}})$. Therefore, the full Hessian $\tilde{H}$ of a system with N_{LK} link atoms has $3N_{\text{LK}}$ extra rows/columns. Diagonalization of this extended Hessian yields $3N_{\text{LK}}$ extra frequencies, which are in essence an artifact of the QM/MM border description and are not inherent to the real physical system. Moreover, the link atoms are usually not completely optimized during the energy minimization process, and unphysical imaginary frequencies might appear as a system with constrained link atoms in a nonequilibrium state.

Hence, one has to project out the $3N_{\text{LK}}$ link atom degrees of freedom to construct a Hessian H of the dimension $3N_{\text{AT}}$. A straightforward and simple solution is to omit the rows/columns in the Hessian that correspond to the link atoms.[39] This approach coincides with the Partial Hessian Vibrational Analysis (PHVA), which can be interpreted as associating an infinite mass to the link atoms.[39−45] This procedure not only disturbs the global translational and rotational symmetry of the system, reflected by the destruction of the six zero eigenvalues of the Hessian, but also the lower frequency spectrum is affected in an unpredictable manner. Cui and Karplus propose to project out the link atom motions after making them orthogonal to the global translation/rotation vectors.[24] This method could only be applied to a system containing a single link atom which is left unconstrained during the geometry optimization. This is often not the case, because preferably constraints on the link atom's position are used to locate it between the QM host and MM host. Such constraints ensure that one of the orbitals of the QM host is effectively pointed toward the MM host, thus providing a better description of the covalent bond.

Here, we focus on a different optimization procedure, which has also been investigated in the framework of ONIOM by Dapprich et al.[17,46] A similar methodology is now extended to the framework of QM/MM where explicit QM/MM polarization effects are included. Instead of a full geometry optimization, the position of the link atom is constrained and can be written as a function of the other QM and MM atom positions: $\boldsymbol{R}_{\text{LK}} = \boldsymbol{R}_{\text{LK}}(\boldsymbol{R})$. Respecting the following notation, where x stands for the $3N_{\text{AT}}$ displacements of the QM and MM atoms and x'' for the $3N_{\text{LK}}$ link atom displacements, we can express the constraints as $x'' = x''(x)$. They reduce the dimensionality of the potential energy surface $\tilde{V}$ to a new potential energy function in $3N_{\text{AT}}$ space

$$V(x) = \tilde{V}(x, x'')|_{x''=x''(x)} \tag{27}$$

Using the chain rule, the $3N_{\text{AT}}$-dimensional gradient G and $3N_{\text{AT}} \times 3N_{\text{AT}}$ Hessian H of the new function $V(x)$ can be written as

$$G^x = \tilde{G}^x + \sum_{x''} \tilde{G}^{x''} \frac{\partial x''}{\partial x} \tag{28}$$

$$\begin{aligned} H^{xy} = \; & \tilde{H}^{xy} + \sum_{x''} \tilde{H}^{x''y} \frac{\partial x''}{\partial x} + \sum_{y''} \tilde{H}^{xy''} \frac{\partial y''}{\partial y} \\ & + \sum_{x''y''} \tilde{H}^{x''y''} \frac{\partial x''}{\partial x} \frac{\partial y''}{\partial y} + \sum_{x''} \tilde{G}^{x''} \frac{\partial^2 x''}{\partial x \partial y} \end{aligned} \tag{29}$$

with $\tilde{G}$ being the original $(3N_{\text{AT}} + 3N_{\text{LK}})$-dimensional gradient vector of $\tilde{V}(x, x'')$ and $\tilde{H}$ the $(3N_{\text{AT}} + 3N_{\text{LK}}) \times (3N_{\text{AT}} + 3N_{\text{LK}})$ matrix containing the second derivatives of $\tilde{V}(x, x'')$,

all evaluated at the reference point. The equations for the Hessian elements (eq 29) indicate that the elimination of the link atom involves both projections of the original Hessian $\tilde{H}$ (first four terms on the right-hand side) as well as a term depending on the original forces $\tilde{G}$ on the link atom (last term on the right-hand side). The main point is that the constraints imposed during the geometry optimization are also imposed during the vibrational analysis, which is the key condition for consistent and meaningful frequencies.[26] Typically, the constraints $x''(x)$ only depend on the position of the neighboring QM host (x_{QMH}) and MM host (x_{MMH}). The constraint derivatives of type $\partial x''/\partial x$ or $\partial^2 x''/\partial x \partial y$ evaluated at the reference geometry are then only nonzero if x, y corresponds to host atom displacements. Consequently, only the rows/columns in the Hessian that involve host atoms are affected by the projection, while other gradient and Hessian elements remain unchanged, i.e., $\tilde{G}^x = G^x$ and $\tilde{H}^{xy} = H^{xy}$ if x, y do not involve host atom displacements.

In our QM/MM procedure, the link atom is forced to stay colinear with the QM and MM hosts during the energy minimization, and at a fixed, scaled distance. This completely determines the positions of the link atoms. In the subsequent vibrational analysis, one can either perform a numerical or an analytical second derivative calculation. For second derivatives obtained with numerical differentiation of slightly displaced geometries, the displaced geometries are such that they respect the constraints. Since the link atom degrees of freedom are thus never sampled, the numerical Hessian yields $3N_{LK}$ zero eigenvalues. For analytical second derivatives, the Hessian $\tilde{H}$ must be reduced in size with the above projection. Specifically, the functional form of the constraints is

$$x'' = x''(x_{QMH}, x_{MMH}) = x_{QMH} + \alpha(x_{QMH} - x_{MMH}) \tag{30}$$

from which one can readily derive the relevant quantities for the projection in eqs 28−29. For instance, the scaling factor α is chosen to be 0.7261 for a link atom replacing a covalent single $C(sp^3)-C(sp^3)$ bond, it being the ratio of the equilibrium C−H and C−C distances in the CHARMM force field.[47] Note that the gradient correction in the last term of eq 29 drops out because of the linear relationship between x'' and x, i.e., $\partial^2 x''/\partial x \partial y = 0$ for x, y in $\{x_{QMH}, x_{MMH}\}$. It is clear that this projection affects only the Hessian elements of the host atoms and the link atom itself but no other Hessian elements. The projection can be performed within the QM code, since the derivatives of the classical parts vanish for these specific Hessian elements (e.g., $\tilde{H}^{x''y}_{class} = 0$).

As a last point, we mention a potential pitfall concerning the definition of the total energy of the QM/MM system with link atoms. The degree to which a link atom contributes to the total energy of the system is reason for discussion (see, e.g., refs 17, 48, and 49). As for normal-mode analysis, it is essential that both gradients and second derivatives conform to the chosen definition of the potential energy surface, that is, with the same constraints as in the energy minimization. This is readily satisfied in the Q-Chem/CHARMM interface. The user can specify the degree to which the link atom contributes to the total energy, and all derivatives are calculated accordingly. The subsequent elimination of the link atom coordinates, as in eqs 28−29, does not depend on the specific definition of the total energy of the system.

II.F. Long-Range Electrostatics. A common strategy, implemented in e.g. CP2K,[50] to reduce the computational cost of QM/MM electrostatic calculations is the introduction of a cutoff distance r_c. The Coulomb interaction between MM atoms and the QM region is neglected if the distance in between exceeds r_c. As a consequence, the number of perturbations decreases in the CPSCF equations of a QM/MM calculation, because a displacement of an MM atom which is too far away from the QM region will not affect the electronic cloud of the QM region. The full QM/MM Hessian still has its full $3N_{AT} \times 3N_{AT}$ dimension, but the reduced number of CPSCF equations facilitates its computation. In addition, the Hessian becomes more sparse, since Hessian elements $H_{A,B}$ between an MM atom A and a QM atom B are zero if they are beyond the cutoff.

Regardless of the tempting computational advantages, a cutoff strategy potentially introduces serious errors in the description of the molecular system. Electrostatics are long-range interactions with a $1/r_{AB}$ decay, leading to a $1/r_{AB}^3$ decay of the Hessian elements. A cutoff for the electrostatics leads to a shift of the potential energy and to modified vibrational frequencies with respect to the fully interacting system. It is to be anticipated that a cutoff distance r_c introduces errors on the order of $1/r_c^3$ in the Hessian elements.

The introduction of mobile blocks into the vibrational analysis has the advantage that the long-range electrostatics are not influenced. While the motion within a block is constrained, the description of the interatomic interactions is *in se* not altered. This is a major strength of the MBH approach with respect to coarse-graining methods: the MBH approach entails a correct description of long-range electrostatics and makes frequency calculations of large QM/MM systems feasible without invoking a cutoff technique.

III. Illustration of Computational Efficiency

The parallel implementation of MBH provides an efficient way to calculate vibrational frequencies of QM/MM systems. The reduced number of CPSCF equations results in computational profit on the level of both memory requirements and CPU time.

Formulas for the memory estimates of each step of sections II.B and II.D are included in Table 1 for a restricted closed-shell calculation, where n_o (n_v) denotes the number of occupied (virtual) orbitals and n_b denotes the number of basis functions. For open-shell calculations, the memory requirements double. In order to avoid load balance problems, the current implementation first solves the CPSCF equations for the QM atomic displacements followed by the MM atomic displacements.[51] Therefore, the peak memory requirement is $6N_{MM}n_b^2$ (reasonably assuming a larger number of MM atoms than QM atoms), whereas the needed disk storage still scales with $6N_{AT}n_b^2$. Parallellization over N_P processors reduces the memory by roughly a factor of N_P. The right column of Table 1 estimates the required memory per processor (besides the static work memory) for the calculation of the mobile block derivatives with N_P processors. The

Table 1. Memory Requirements Per Processor Expressed As the Number of Double Precision Floating-Point Numbers (1 number = 8 bytes) That Needs to Be Stored for Restricted Closed-Shell Hessian Calculations[a]

	full QM/MM Hessian	MB QM/MM Hessian
step 1	$3N_{MM} \times n_o n_v/N_P$	$d_{MM} \times n_o n_v/N_P$
step 2	$\ll 2 \times 3N_{MM} \times n_b^2/N_P$	$\ll 2 \times d_{MM} \times n_b^2/N_P$
step 3	$\propto n_b^2$	$\propto n_b^2$
step 4	$> 2 \times 3N_{MM} \times n_b^2/N_P$	$> 2 \times d_{MM} \times n_b^2/N_P$
step 5	$\propto n_b^2$	$\propto n_b^2$
peak	$2 \times 3N_{MM} \times n_b^2/N_P$	$2 \times d_{MM} \times n_b^2/N_P$

[a] N_P is the number of processors, n_b the number of basis functions, and n_o (n_v) the number of occupied (virtual) orbitals. The symbol $\ll$ means that the actual number is, in practice, much lower than the theoretical estimate. $\propto$ indicates the scaling, where the prefactor is independent of the system size. $>$ means that the actual number is higher than the theoretical estimate due to overhead such as the book keeping of variables. In the case of MBH, d_{MM} is the number of MM perturbations: $d_{MM} = 6N_{BL} + 3N_{MM,\,free}$, while the total number of perturbations is $d = d_{MM} + 3N_{AT}$.

peak memory is now equal to $2d_{MM}n_b^2/N_P$, where $d_{MM} = 6N_{BL} + 3N_{MM,\,free}$ is the number of MM perturbations.

To illustrate the efficiency, a small part (330 atoms) of the chorismate mutase enzyme (1COM in the Protein Data Bank) is taken as a test example.[52] At the B3LYP/6-31+G* level of theory, the number of occupied orbitals is $n_o = 59$, and the number of virtual orbitals is $n_v = 261$, totalling $n_b = 320$ basis functions. The full QM calculation would need roughly 1.5 TB of memory, which is not feasible at present. Now consider a QM/MM calculation where the system is divided into a 24 QM atoms and 306 MM atoms. The same level of theory, B3LYP/6-31+G*, is used for QM atoms while the MM atoms are described by the PARAM27 force field of CHARMM.[53,54] This QM/MM Hessian calculation only requires about 2.5 GB, a drastic reduction with respect to the full QM calculation.

In the next step, the system is divided into mobile blocks. By dividing the 306 MM atoms into 20 blocks, with one residue per block, a mere 0.5 GB is sufficient for the MBH calculation on one processor. This illustrates that MBH allows for vibrational analysis on larger systems than is feasible with the full Hessian. For example, the largest system in which the full Hessian calculation fits on a single processor with a typical 8 GB of memory consists of roughly 24 QM atoms and 600 MM partial charges. The largest system for the approximate MBH calculation has 24 QM atoms and 4500 MM atoms, assuming an average of 15 atoms per block.

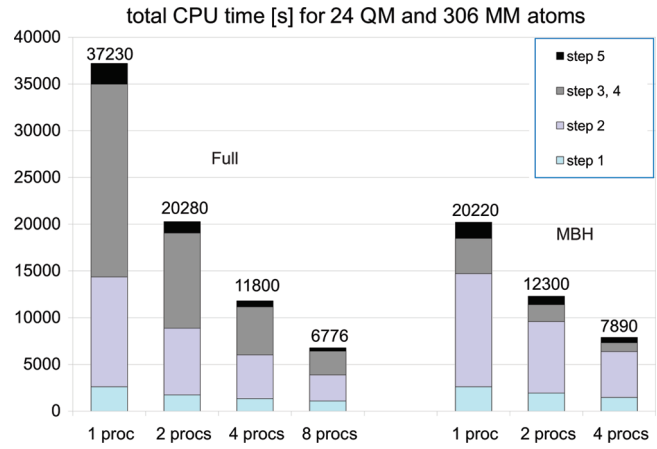

Figure 3. Chorismate mutase test system: CPU times for a QM/MM system with 24 QM atoms and 306 MM atoms. Parallellization speeds up the QM/MM calculatation: the full Hessian is calculated on 1, 2, 4, and 8 processors (left); the mobile block Hessian on 1, 2, and 4 processors (right).

Thus, MBH increases the size of the QM/MM systems that can be addressed by 7.5 times. Even larger system sizes are feasible when grouping multiple residues per block. In addition, parallellization of the code further reduces the required memory per processor.

Figure 3 shows the CPU times of the frequency calculation for the chorismate mutase test system, described with the same QM/MM description of 24 QM atoms and 306 MM atoms. The timings of the full Hessian calculation are compared to those of the MBH, where again the 306 MM atoms are divided into 20 blocks. The MBH CPU time on one processor (20 220 s) reduces to 54% of the full Hessian CPU time on one processor (37 230 s). The CPU times of steps 3 and 4 are affected the most, while the CPU time of step 2 remains unaltered or might even increase. The latter is due to the use of an iterative subspace algorithm for solving the CPSCF equations. The number of trial vectors in the CPSCF solution subspace is largely unaffected by blocking the MM atoms. In our particular example, the number of basis vectors is actually slightly increased because of the mixing of different atomic displacements in the blocks. This leads to a subsequent small increase in the execution time for the CPSCF step, which is almost strictly proportional to the number of basis vectors. For the full Hessian calculation, a speedup of a factor of 5.5 is realized when using eight processors instead of one processor. For the MBH calculation, the speedup is 2.6 when using four processors instead of one processor. This shows that MBH indeed reduces the computational efficiency, and the parallel implementation even more. Of course, the effective speedup depends on the block choice, where larger blocks lead to more impressive speedups.

IV. Application: The Bortezomib Drug

IV.A. Oxidative Deboronation of Bortezomib. Boronic acids (R-B(OH)$_2$) play an important role in a variety of medical applications due to the ability of boron to mimic the tetrahedral transition state of an sp^3 hybridized carbon. A particularly interesting example is the drug bortezomib (originally codenamed PS-341, marketed as Velcade),[29,30] which is

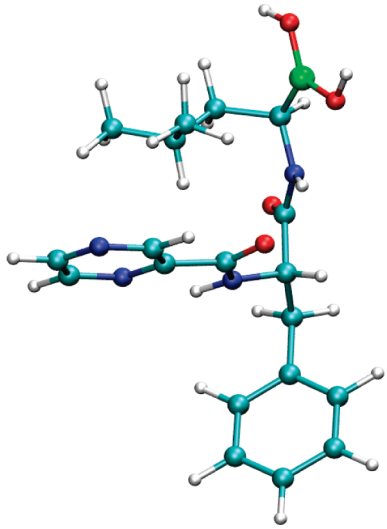

Figure 4. The drug bortezomib: 3D model of the reactant (REA).

used to treat multiple myeloma and mantle cell lymphoma, two types of hematologic cancer. Bortezomib is a boronic acid analog of a Phe−Leu dipeptide coupled to a 2-carboxyl-pyrazine group (Figure 4) that binds to the catalytic site of the 26S proteasome with high specificity.[55,56] The proteasome, a large multicatalyic protease complex, regulates protein expression and degradation of ubiquitinylated proteins, cleaning abnormal or misfolded proteins from the cell. Inhibition of this cellular pathway by bortezomib ultimately results in apoptosis due to an accumulation of damaged or misfolded proteins in the cell through a number of possible mechanisms.[57]

The chemical activity of bortezomib is largely due to the boronic acid moiety, which appears to bind with the active site N-terminal threonine residue of the proteasome.[58−60] In a recent article, Larkin et al. reported the results of a computational study of the model system boroglycine ($H_2N-CH_2-B(OH)_2$), using H_2O_2 and H_2O as reactive oxygen species.[61] The oxidative deboronation, which is suggested as the principal pathway for the metabolism of bortezomib, is found to be exothermic and endothermic for the reactions with H_2O_2 and H_2O, respectively. With the computational improvements in Q-Chem/CHARMM, we can now systematically study the full bortezomib molecule (53 atoms) instead of the smaller model system (11 atoms).

Figure 5 illustrates the oxidative deboronation reaction, where an oxygen cleaves the boron acid and takes the position of the boron. The vibrational free energy difference ΔG_{vib} of the reaction is studied since this quantity is calculated with the vibrational frequencies. Two reactive oxygen reagents are considered: H_2O_2 and H_3C-OH. Methanol is chosen as the second reagent because it better describes—compared to water in ref 61— the alcohol group of the threonine residue of the 26S proteasome on which the bortezomib molecule is believed to bind. The respective products are an alcohol and an ether:

$$R-B-(OH)_2 + H_2O_2 \rightarrow R-OH + B-(OH)_3$$

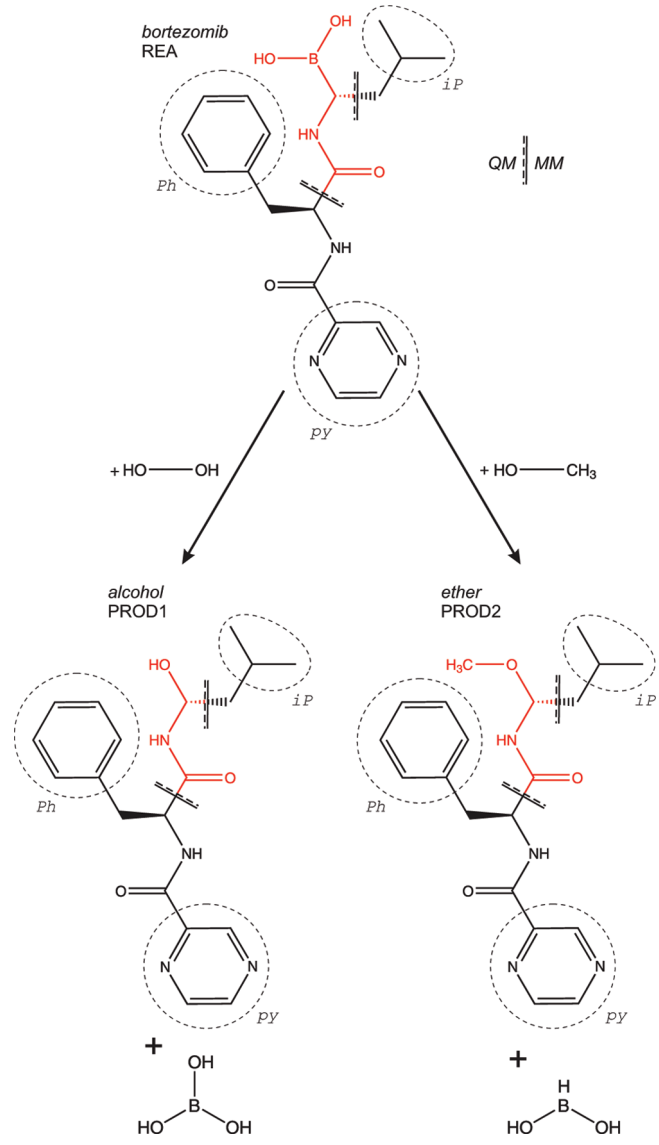

Figure 5. Bortezomib—Structure of the drug bortezomib with indication of the QM and MM regions. QM atoms are colored red. Link atoms are placed at the QM/MM border. Oxidative deboronation of bortezomib (REA) with the oxygen reagens H_2O_2 or CH_3-OH results in an alcohol (PROD1) or ether (PROD2) product, respectively. The mobile blocks Ph, py, and iP are indicated with a dashed line.

$$R-B-(OH)_2 + H_3C-OH \rightarrow R-O-CH_3 + H-B-(OH)_2$$

where R designates the Phe−Leu moieties coupled to the pyrazine. The bortezomib reactant will be referred to as REA and the alcohol and ether product as PROD1 and PROD2, respectively.

With this application, we aim at presenting a workable model for the computation of vibrational frequencies in a QM/MM approach with the inclusion of mobile blocks in the MM region. We restrict ourselves to the thermochemistry of the reactants versus the products, as the reaction path leads along several intermediate transition states.[61] A profound assessment can only be done if we dispose of a high-level benchmark study involving accurate optimized geometries for the bortezomib structure. Therefore, section IV.B is devoted to the creation of the QM and QM/MM geometries.

Table 2. Bortezomib−RMSD in Å between QM Geometries Optimized at Different Levels of Theory[a]

	QM geometries	B3LYP	PBE	B3LYP-D	PBE-D
REA	B3LYP	−			
	PBE	0.10	−		
	B3LYP-D	0.57	0.50	−	
	PBE-D	0.68	0.60	0.12	−
	RI-MP2/cc-pvtz	0.58	0.51	0.09	0.12
PROD1	B3LYP	−			
	PBE	0.64	−		
	B3LYP-D	0.75	0.34	−	
	PBE-D	0.82	0.42	0.11	−
	RI-MP2/cc-pvtz	0.78	0.45	0.14	0.09
PROD2	B3LYP	−			
	PBE	0.07	−		
	B3LYP-D	0.91	0.86	−	
	PBE-D	0.97	0.93	0.09	−
	RI-MP2/cc-pvtz	0.94	0.89	0.16	0.14

[a] The basis set is 6-311++G(d,p) except for RI-MP2. REA refers to the reactant bortezomib, while PROD1 and PROD2 refer to the alcohol and ether products formed after oxidative deboronation, as indicated in Figure 5.

Section IV.C focuses on the calculation of the QM and QM/MM frequencies, either derived from full Hessians or from MBH. Section IV.D compares the QM/MM difference in the vibrational Gibbs free energy with the benchmark full QM value. Moreover, we discuss the influence of the introduction of mobile blocks in the MM region.

IV.B. Creating QM and QM/MM Geometries. The benchmark geometries are generated by performing a full QM calculation with Q-Chem. The input geometry is taken from the Protein Data Bank (2F16 in the Protein Data Bank) where the proteasome 26S and water are removed to select the relevant conformation. The geometry of each structure is first optimized in order to calculate frequencies subsequently using analytical second derivatives. Since it is not required that the gradient within a given block be zero in order to apply MBH, the geometry optimization could also be done with rigid blocks, where the rigidity is imposed via the SHAPES facility in CHARMM. To increase the accuracy, the convergence criteria and numerical accuracy options are set slightly tighter than is typical for a plain geometry optimization: the tolerance for the gradient is reduced to 0.00015 hartree/Bohr, and the tighter (75 302) atomic integration grid is chosen with 75 radial points and 302 Lebedev angular points for each atom. The B3LYP[62,63] and PBE (=PBE1PBE)[64−66] levels of theory are used for this example with the 6-311++G(d,p) basis set, as well as the B3LYP-D and PBE-D functionals where an empirical correction term is added to account for dispersion effects.[67] In addition, the structure is optimized at the more expensive RI-MP2/cc-pVTZ[68,69] level to validate the accuracy of the B3LYP(-D) and PBE(-D) geometries. A frequency calculation is however currently not practical at the RI-MP2 level of theory because its second derivatives implementation is based on the numerical differentiation of the analytical gradient. Table 2 lists the mass-weighted root-mean-square distances (RMSD) between the QM geometries optimized at different levels of theory, which are calculated on the basis of the non-hydrogen atoms after aligning the structures. Structures with an RMSD below 0.25 Å are considered to have close-lying geometries; when the RMSD is higher, the structures are considered less similar. The RMSD between B3LYP and PBE structures is low, except in the isolated case of the PROD1 product. Similarly, the B3LYP-D and PBE-D geometries lie close to each other, but they differ from the geometries without dispersion. In all cases, the B3LYP-D and PBE-D geometries lie closer to the RI-MP2 geometries than B3LYP and PBE, and preference should be given to DFT methods including dispersion. Visualization of the structures shows that the distance between the Leu and Phe moieties decreases under the influence of the dispersion forces. Apparently, the dispersion in this rather large molecular system plays a more important role on the geometry than the level of theory.

For the QM/MM calculation, the REA, PROD1, and PROD2 molecules are divided into a QM and MM region, whereas the small H_2O_2 and CH_3-OH reactants and the $B-(OH)_3$ and $H-B-(OH)_2$ products are still described completely on the QM level. The MM region consists of 42 atoms: the phenyl group (Phe), the iso-butyl group (Leu), the pyrazine (ring with nitrogens), and some neighboring atoms, as shown in Figure 5. The reactive site is made part of the QM region, colored red in Figure 5. In addition, the amide bond is chosen to belong entirely to the QM region, since preliminary tests in which the QM/MM border crosses the amide bond turned out to break the partially delocalized nature of the amide bond with even a nonplanar geometry in some cases. This brings the number of QM atoms to 11, 8, and 11 for the REA, PROD1, and PROD2 molecules, respectively. Each structure has two link atoms where the QM/MM border cuts through covalent C−C bonds, as indicated in Figure 5. The combination of the LONEPAIR and SHAKE commands[70] of CHARMM keeps each link atom colinear with its MM and QM host at a relative distance of 0.7261 (see eq 30) during the geometry optimization.

In this QM/MM calculation, the same QM functionals are used as in the full QM case (the B3LYP, PBE, B3LYP-D, or PBE-D level with the 6-311++G(d,p) basis set), while the MM force field is based on the PARAM27 parameter set[53,54] of CHARMM. The root-mean-square distances (RMSD) between the QM/MM geometries optimized at these four levels of theory are summarized in Table 3. As in the QM calculations, the B3LYP and PBE structures are similar, and the B3LYP-D and PBE-D structures also lie close to each other. Contrary to the full QM calculations, the comparison of B3LYP and PBE with B3LYP-D and PBE-D shows that the inclusion of dispersion interactions only has a minor influence on the geometry. The reason is that the dispersion contribution is limited to the small subset of QM atoms; hence it barely affects the relative orientation between the Leu and Phe moieties. This behavior is confirmed by comparing the QM/MM structure with its respective QM structure, of which the RMSD is also included in Table 3. Indeed, QM/MM structures calculated in the absence of dispersion lie closer to their QM counterpart than structures including dispersion. For instance, the RMSD between the

Table 3. Bortezomib−RMSD in Å between QM/MM Geometries Optimized at Different Levels of Theory[a]

	QM/MM geometries	B3LYP	PBE	B3LYP-D	PBE-D
REA	B3LYP	–			
	PBE	0.05	–		
	B3LYP-D	0.11	0.07	–	
	PBE-D	0.15	0.11	0.05	–
	correspondingQM	0.18	0.20	0.43	0.50
PROD1	B3LYP	–			
	PBE	0.14	–		
	B3LYP-D	0.16	0.03	–	
	PBE-D	0.23	0.09	0.07	–
	correspondingQM	0.18	0.60	0.63	0.62
PROD2	B3LYP	–			
	PBE	0.16	–		
	B3LYP-D	0.18	0.04	–	
	PBE-D	0.12	0.05	0.07	–
	correspondingQM	0.24	0.16	0.87	0.88

[a] In addition, each QM/MM geometry is compared with its respective QM geometry calculated at the same level of theory. REA refers to the reactant bortezomib, while PROD1 and PROD2 refer to the alcohol and ether products formed after oxidative deboronation, as indicated in Figure 5.

QM/MM and QM structure is 0.18 Å with the B3LYP functional, while it is 0.43 Å with the B3LYP-D functional.

IV.C. Frequency Calculations. Hessians and gradients are calculated analytically at the same level of theory as the geometry optimization. The Hessian is diagonalized after mass-weighting to obtain the frequencies and vibrational modes. In this paper, we used simultaneously the program TAMkin[71] to derive frequencies, modes, and thermodynamic properties. With its batch processing features, it provides a handy interface to extract the relevant molecular information from the large number of Q-Chem/CHARMM output files, to compute the frequencies, to write mode trajectory files for visualization of the vibrational eigenmodes, and to derive the Gibbs free energy differences.

Frequencies are derived from three different types of Hessians:

A. QM Full Hessian. The frequency run (and geometry optimization) is performed with a QM description. The full $3N_{AT} \times 3N_{AT}$ Hessian is calculated with Q-Chem.

B. QM/MM Full Hessian. The frequency run (and geometry optimization) is performed with a QM/MM description. The full $3N_{AT} \times 3N_{AT}$ Hessian is calculated with Q-Chem/CHARMM. CHARMM can derive MBH frequencies from this full Hessian.

C. QM/MM Mobile Block Hessian, Reduced CPSCF. The frequency run (and geometry optimization) is performed with a QM/MM description. The mobile block Hessian of reduced dimension $d \times d$ is calculated by Q-Chem/CHARMM with the reduced number of CPSCF equations and diagonalized in CHARMM, directly leading to the MBH frequencies.

In cases A and B, a Hessian of full size $3N_{AT} \times 3N_{AT}$ is constructed. Its diagonalization is performed by Q-Chem (case A), by CHARMM (case B), or by TAMkin (case A or B) and will be referred to as the full Hessian vibrational analysis (FHVA) in the remainder of the discussion. From these full Hessians, one can also derive MBH frequencies through projection and a gradient correction, as explained in ref 26 (cases A and B) and implemented in CHARMM[19] and in TAMkin.[71] But the direct method to attain MBH frequencies is case C, where frequencies evolve directly from the diagonalization of the mobile block Hessian itself, without prior construction of the full Hessian. This is the new implementation which is the subject of this paper and which is now available via Q-Chem/CHARMM.

The calculations are performed on four processors with the parallel version of Q-Chem and Q-Chem/CHARMM. The QM Hessian computation in case A is the most time-consuming one (~3 days). The QM/MM Hessian calculations in cases B and C are considerably faster, taking approximately 0.5% of the time for a full QM calculation (~30 min). The speedup realized by the reduced CPSCF implementation versus a full CPSCF QM/MM implementation is moderate since the number of MM atoms and blocks is rather small in the system under study compared to the chorismate mutase test system of section III.

In the current implementation, the mobile blocks should be part of the MM region and should not contain any MM host atom of the link atoms. Figure 5 proposes three plausible blocks, of which the internal motions are suspected not to matter when estimating the vibrational free energy difference: the phenyl group (Ph), the iso-propyl group (iP), and the pyrazine (py). Frequencies are calculated with one block or with multiple blocks simultaneously. For instance, the method MBH_{Ph} indicates that the vibrational analysis is performed assuming that the atoms of block Ph vibrate coherently as a whole. In the method labeled $MBH_{Ph, iP, py}$, the system contains three blocks (Ph, iP, py), while the remaining atoms can vibrate individually. The introduction of blocks reduces the number of frequencies; for instance, 87 frequencies remain for the bortezomib reactant with three blocks, which is 55% of the original 159 frequencies.

The accuracy of the vibrational frequencies is largely influenced by the geometry convergence criteria and the quality of the Hessian elements, which mainly depends on the numerical integration grid for calculating the electron integrals in the CPSCF equations. A good assessment of the accuracy is the value of the lowest six eigenvalues of the Cartesian Hessian. In principle, those should be zero at a minimum or maximum energy point because of the invariance of the potential energy surface of a gas phase molecule under global translations and rotations.[26,72,73] Global rotations with a nonzero frequency are caused by a poor geometry convergence, such that the Eckart conditions are not fulfilled at the reference point and the global rotations may mix up with the internal vibrations. Global translations with a nonzero frequency signal inaccurate Hessian elements. The influence of the inaccuracies is validated by first projecting out the global translations/rotations from the full Hessian before diagonalization, thus creating six "hard" zero eigenfrequencies and removing any translational/rotational contribution in the low lying eigenmodes. This approach is equivalent to imposing Eckart constraints on the vibrational motions[1] and will therefore be denoted with the superscript "Eck". The Eckart projection is performed by TAMkin or by the RAISE option in CHARMM. Each standard method, FHVA or MBH, thus has a corresponding method $FHVA^{Eck}$

or MBH^{Eck} where the six lowest frequencies are hard zeros. The Eckart projection mainly affects the low frequency spectrum. The difference between the standard method and these Eckart methods is a measure of the Hessian's accuracy.

IV.D. Discussion. A first point of interest is the influence of MBH on the individual frequencies and modes. Such detailed studies have been performed recently in, e.g., refs 25, 27, 37, 74, and 75. It was found that the block choice determines which local modes and/or global modes are well described by MBH. For the bortezomib system under study, all atoms of the reactive site are considered completely mobile free atoms. The spectator groups are fairly rigid during the reaction, and their internal geometry can be kept fixed. With this plausible block choice, a reasonable similarity is to be expected between MBH modes and frequencies and the benchmark full Hessian results. This can be verified by calculating the overlap between the modes, a number lying between 0 and 100%, defined as

$$O_{ij} = |\langle \nu_i^{MBH} | \nu_j \rangle|^2 \tag{31}$$

where $|\nu_i^{MBH}\rangle$ is the ith mass-weighted MBH mode with frequency ν_i^{MBH} and $|\nu_j\rangle$ is the jth mass-weighted FHVA mode with frequency ν_j. The overlap data of REA are plotted in Figure 6, where the QM/MM MBH is derived from the full QM/MM Hessian by projection (case B of ssection IV.C). A dark dot indicates a high overlap between the modes, and a dot located close to the diagonal of the plot indicates that the corresponding frequencies are almost equal. The plot shows good agreement between the FHVA modes of REA with the MBH^{Ph} and $MBH^{Ph,\ iP,\ py}$ modes, and an excellent agreement between FHVA and MBH for frequencies below 250 cm^{-1}.

A second point of interest is the thermodynamic quantities derived from the frequencies. The frequencies serve directly as input quantities for the vibrational free energy G_{vib}, which makes it an interesting parameter for studying the influence of the MBH model. The MBH however reduces the number of frequencies; hence a better quantity is the *difference* in vibrational free energy ΔG_{vib} between the products and the reactants, calculated as

$$\Delta G_{vib} = G_{vib}(PROD1) + G_{vib}(B(OH)_3) - G_{vib}(REA) - G_{vib}(H_2O_2) \tag{32}$$

$$\Delta G_{vib} = G_{vib}(PROD2) + G_{vib}(HB(OH)_2) - G_{vib}(REA) - G_{vib}(H_3COH) \tag{33}$$

In the harmonic oscillator approximation, the vibrational free energy G_{vib} is derived from the vibrational partition function Q_{vib} by the well-known relation $G_{vib} = -k_B T \ln Q_{vib}$, where k_B is the Boltzmann constant and T the temperature. The vibrational partition function is built from the individual contributions of the harmonic frequencies.[76−79] The entropic part of the vibrational free energy, $-T\Delta S_{vib}$, is also reported, which is derived from the relation $S_{vib} = k_B(\partial \ln Q_{vib})/(\partial T)$. When the harmonic oscillators are treated classically (high temperature limit) instead of quantum mechanically, the free energy difference is purely entropic.

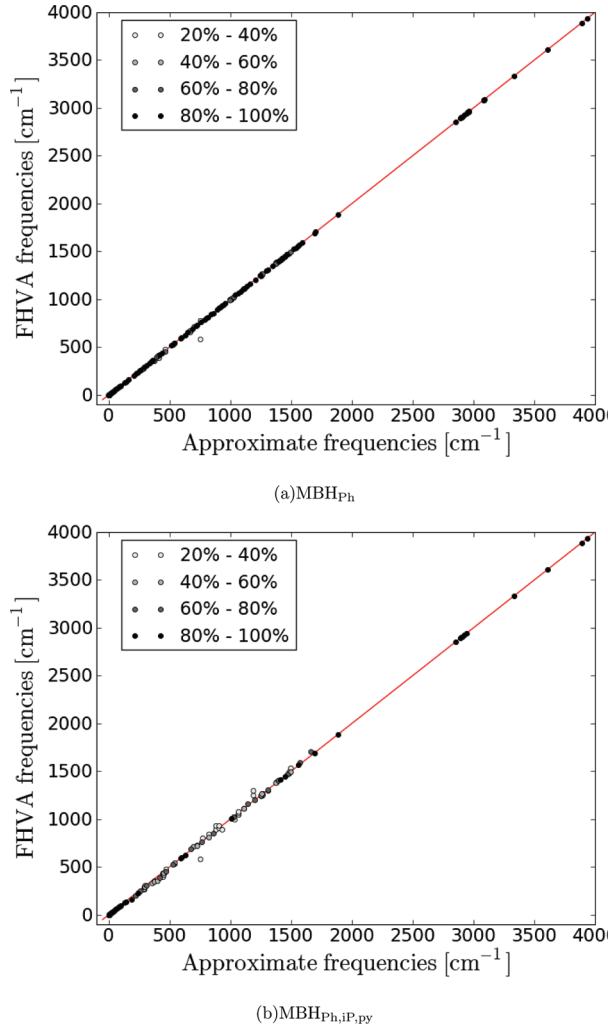

Figure 6. Bortezomib—Evaluation of the MBH frequencies/modes of REA on the basis of the QM/MM Hessian (case B). The overlap $O_{ij} = |\langle \nu_i^{MBH} | \nu_j \rangle|^2$ between the mass-weighted MBH modes $|\nu_i^{MBH}\rangle$ and FHVA modes $|\nu_j\rangle$ is plotted as a function of the respective frequencies. A dark colored dot indicates a high overlap between the modes; overlap values below 20% are not shown.

The alternative would be to derive the free energy differences from a molecular dynamics (MD) simulation, which performs a more realistic sampling of the reaction coordinate. In this paper, only the harmonic limit is considered, which amounts to sampling a local harmonic approximation of the potential energy surface. Even in cases where the harmonic limit is insufficient to accurately describe the reaction process, one can still learn from the harmonic limit result by comparing it to explicit MD sampling. Since MD basically solves Newton's classical equations of motion and not the time-dependent Schrodinger equation, MD results should be compared with the results derived from the classical oscillators instead of the quantum oscillators partition function. For completeness, we have therefore calculated some classical oscillator results besides the quantum oscillator results.

Tables 4−7 list the vibrational free energies and the vibrational entropic contributions, calculated with quantum/classical oscillators, and without/with Eckart projection. The

Table 4. Deboronation of Bortezomib with the Oxygen Reagents H_2O_2[a]

		B3LYP				PBE			
Hessian	NMA method	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$
case A	FHVA	2.40	2.47	1.65	1.88	2.50	2.24	1.57	1.77
QM	MBH_{Ph}	2.40	2.47	1.65	1.88	2.48	2.23	1.57	1.77
full	MBH_{iP}	2.33	2.40	1.68	1.88	2.30	2.01	1.41	1.59
	MBH_{py}	2.40	2.47	1.66	1.88	2.50	2.23	1.56	1.77
	$MBH_{Ph,\,iP}$	2.33	2.40	1.68	1.88	2.28	2.00	1.41	1.58
	$MBH_{Ph,\,py}$	2.40	2.47	1.65	1.88	2.48	2.22	1.56	1.76
	$MBH_{iP,\,py}$	2.33	2.40	1.68	1.89	2.29	2.00	1.40	1.58
	$MBH_{Ph,\,iP,\,py}$	2.33	2.40	1.68	1.88	2.27	1.99	1.40	1.57
case B	FHVA	2.77	2.81	2.16	2.33	2.93	2.96	2.28	2.47
QM/MM	MBH_{Ph}	2.77	2.81	2.16	2.33	2.94	2.97	2.28	2.47
full	MBH_{iP}	2.76	2.80	2.16	2.33	2.95	2.98	2.32	2.50
	MBH_{py}	2.78	2.82	2.16	2.34	2.92	2.95	2.27	2.46
	$MBH_{Ph,\,iP}$	2.76	2.80	2.16	2.33	2.96	2.99	2.32	2.51
	$MBH_{Ph,\,py}$	2.78	2.82	2.16	2.34	2.93	2.96	2.28	2.47
	$MBH_{iP,\,py}$	2.76	2.81	2.17	2.34	2.94	2.97	2.31	2.49
	$MBH_{Ph,\,iP,\,py}$	2.76	2.81	2.17	2.34	2.95	2.98	2.32	2.50
case C	MBH_{Ph}	2.81	2.87	2.21	2.38	2.97	3.03	2.34	2.53
QM/MM	MBH_{iP}	2.82	2.87	2.23	2.40	2.99	3.03	2.36	2.55
mobile block	MBH_{py}	2.79	2.84	2.18	2.35	2.93	2.97	2.28	2.47
(several)	$MBH_{Ph,\,iP}$	2.79	2.83	2.19	2.36	2.97	3.00	2.34	2.52
	$MBH_{Ph,\,py}$	2.81	2.84	2.18	2.36	2.94	2.97	2.29	2.48
	$MBH_{iP,\,py}$	2.80	2.84	2.20	2.37	2.98	3.01	2.34	2.53
	$MBH_{Ph,\,iP,\,py}$	2.80	2.83	2.19	2.36	2.99	3.02	2.35	2.54

[a] The vibrational free energy difference of eq 32 and its entropic part are calculated (in kcal/mol) with several NMA models. The superscript "cl" indicates the use of classical instead of quantum oscillators. The superscript "Eck" indicates that Eckart conditions are applied. MBH and MBH^{Eck} frequencies are derived from the QM full Hessian (case A) and the QM/MM full Hessian (full CPSCF, case B) with CHARMM or TAMkin or are obtained from the direct QM/MM mobile block Hessian (reduced CPSCF, a different Hessian for each block choice, case C). The basis set is 6-311++G(d,p).

Table 5. Deboronation of Bortezomib with the Oxygen Reagens H_2O_2 Continued[a]

		B3LYP-D				PBE-D			
Hessian	NMA method	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$
case A	FHVA	1.78	1.87	1.12	1.33	2.67	2.36	1.50	1.74
QM	MBH_{Ph}	1.86	1.94	1.13	1.36	2.69	2.39	1.52	1.76
full	MBH_{iP}	1.99	2.08	1.24	1.46	2.77	2.73	1.85	2.08
	MBH_{py}	1.80	1.89	1.13	1.34	2.62	2.31	1.49	1.72
	$MBH_{Ph,\,iP}$	2.07	2.15	1.25	1.49	2.80	2.75	1.87	2.10
	$MBH_{Ph,\,py}$	1.88	1.96	1.13	1.37	2.64	2.33	1.51	1.74
	$MBH_{iP,\,py}$	2.01	2.10	1.25	1.48	2.72	2.67	1.84	2.06
	$MBH_{Ph,\,iP,\,py}$	2.09	2.17	1.26	1.50	2.75	2.70	1.86	2.08
case B	FHVA	2.75	2.76	1.99	2.20	2.28	2.26	1.49	1.71
QM/MM	MBH_{Ph}	2.69	2.70	1.93	2.14	2.28	2.26	1.50	1.71
full	MBH_{iP}	2.68	2.69	1.94	2.15	2.35	2.36	1.61	1.82
	MBH_{py}	2.74	2.75	1.98	2.19	2.28	2.26	1.49	1.71
	$MBH_{Ph,\,iP}$	2.62	2.63	1.88	2.08	2.35	2.37	1.62	1.83
	$MBH_{Ph,\,py}$	2.69	2.69	1.92	2.13	2.29	2.26	1.50	1.71
	$MBH_{iP,\,py}$	2.67	2.68	1.93	2.14	2.35	2.36	1.61	1.82
	$MBH_{Ph,\,iP,\,py}$	2.61	2.62	1.87	2.07	2.35	2.36	1.61	1.82
case C	MBH_{Ph}	2.73	2.75	1.97	2.18	2.29	2.28	1.51	1.72
QM/MM	MBH_{iP}	2.71	2.72	1.96	2.17	2.41	2.41	1.65	1.86
mobile block	MBH_{py}	2.76	2.77	1.99	2.21	2.32	2.28	1.52	1.73
(several)	$MBH_{Ph,\,iP}$	2.65	2.65	1.90	2.11	2.38	2.39	1.64	1.85
	$MBH_{Ph,\,py}$	2.70	2.71	1.94	2.15	2.30	2.25	1.49	1.70
	$MBH_{iP,\,py}$	2.69	2.70	1.95	2.16	2.36	2.36	1.61	1.81
	$MBH_{Ph,\,iP,\,py}$	2.65	2.66	1.90	2.11	2.38	2.39	1.64	1.85

[a] See caption of Table 4.

three parts in each table correspond to the three different cases of Hessians as explained above. Variations $\delta\Delta G_{vib}$ in a particular column in the table are due to the description level (QM versus QM/MM) and the NMA models (FHVA versus MBH). Table 8 summarizes these variations $\delta\Delta G$ as well as deviations caused by the functional and the application of the Eckart conditions. The deviations will now be discussed in detail on the basis of the results obtained from quantum oscillators. Classical oscillator results, indicated with a superscript "cl" in Tables 4−7, lead to conclusions

Table 6. Deboronation of Bortezomib with the Oxygen Reagens CH_3OH[a]

		B3LYP				PBE			
Hessian	NMA method	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$
case A	FHVA	−0.13	−0.03	0.62	0.83	−1.11	−1.15	−0.41	−0.20
QM	MBH_{Ph}	−0.12	−0.03	0.62	0.83	−1.14	−1.19	−0.43	−0.23
full	MBH_{iP}	−0.23	−0.13	0.63	0.81	−1.33	−1.34	−0.44	−0.28
	MBH_{py}	−0.14	−0.04	0.62	0.83	−1.12	−1.16	−0.41	−0.20
	$MBH_{Ph,\,iP}$	−0.22	−0.12	0.63	0.81	−1.36	−1.37	−0.46	−0.31
	$MBH_{Ph,\,py}$	−0.13	−0.04	0.62	0.84	−1.15	−1.19	−0.44	−0.23
	$MBH_{iP,\,py}$	−0.23	−0.14	0.63	0.81	−1.33	−1.35	−0.44	−0.28
	$MBH_{Ph,\,iP,\,py}$	−0.23	−0.13	0.63	0.82	−1.36	−1.38	−0.47	−0.31
case B	FHVA	0.38	0.43	1.26	1.42	0.16	0.20	1.12	1.26
QM/MM	MBH_{Ph}	0.33	0.38	1.21	1.36	0.16	0.19	1.11	1.25
full	MBH_{iP}	0.43	0.47	1.33	1.47	0.18	0.21	1.14	1.28
	MBH_{py}	0.39	0.43	1.27	1.42	0.17	0.21	1.13	1.26
	$MBH_{Ph,\,iP}$	0.38	0.43	1.28	1.42	0.17	0.20	1.14	1.27
	$MBH_{Ph,\,py}$	0.34	0.38	1.21	1.37	0.17	0.20	1.12	1.26
	$MBH_{iP,\,py}$	0.43	0.47	1.33	1.47	0.18	0.22	1.15	1.29
	$MBH_{Ph,\,iP,\,py}$	0.38	0.43	1.28	1.43	0.18	0.21	1.14	1.28
case C	MBH_{Ph}	0.37	0.44	1.27	1.42	0.20	0.27	1.18	1.32
QM/MM	MBH_{iP}	0.48	0.53	1.37	1.52	0.21	0.25	1.18	1.32
mobile block	MBH_{py}	0.40	0.44	1.27	1.42	0.18	0.21	1.13	1.27
(several)	$MBH_{Ph,\,iP}$	0.41	0.45	1.29	1.44	0.18	0.21	1.14	1.28
	$MBH_{Ph,\,py}$	0.35	0.39	1.23	1.38	0.18	0.22	1.14	1.28
	$MBH_{iP,\,py}$	0.46	0.50	1.35	1.50	0.22	0.26	1.19	1.32
	$MBH_{Ph,\,iP,\,py}$	0.43	0.47	1.32	1.47	0.22	0.25	1.18	1.32

[a] See caption of Table 4.

Table 7. Deboronation of Bortezomib with the Oxygen Reagens CH_3OH Continued[a]

		B3LYP-D				PBE-D			
Hessian	NMA method	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$	ΔG_{vib}	ΔG_{vib}^{Eck}	$-T\Delta S_{vib}^{Eck}$	$\Delta G_{vib}^{Eck,\,cl}$
case A	FHVA	0.63	0.65	1.30	1.51	−0.59	−0.56	0.36	0.50
QM	MBH_{Ph}	0.58	0.59	1.29	1.48	−0.59	−0.56	0.36	0.50
full	MBH_{iP}	0.75	0.77	1.42	1.64	−0.60	−0.57	0.36	0.50
	MBH_{py}	0.61	0.63	1.30	1.51	−0.62	−0.59	0.35	0.49
	$MBH_{Ph,\,iP}$	0.69	0.71	1.40	1.61	−0.61	−0.58	0.36	0.50
	$MBH_{Ph,\,py}$	0.55	0.57	1.29	1.47	−0.62	−0.59	0.36	0.49
	$MBH_{iP,\,py}$	0.72	0.74	1.42	1.63	−0.63	−0.60	0.35	0.49
	$MBH_{Ph,\,iP,\,py}$	0.66	0.68	1.40	1.60	−0.63	−0.60	0.35	0.48
case B	FHVA	0.59	0.59	1.27	1.47	0.37	0.38	1.17	1.35
QM/MM	MBH_{Ph}	0.65	0.66	1.33	1.53	0.37	0.38	1.17	1.35
full	MBH_{iP}	0.64	0.65	1.34	1.53	0.37	0.38	1.19	1.36
	MBH_{py}	0.59	0.60	1.29	1.48	0.37	0.38	1.18	1.35
	$MBH_{Ph,\,iP}$	0.70	0.71	1.40	1.59	0.37	0.38	1.18	1.36
	$MBH_{Ph,\,py}$	0.66	0.67	1.34	1.54	0.37	0.38	1.17	1.35
	$MBH_{iP,\,py}$	0.65	0.66	1.35	1.54	0.37	0.39	1.19	1.36
	$MBH_{Ph,\,iP,\,py}$	0.71	0.72	1.41	1.60	0.37	0.39	1.19	1.36
case C	MBH_{Ph}	0.68	0.70	1.37	1.57	0.42	0.44	1.22	1.41
QM/MM	MBH_{iP}	0.68	0.68	1.37	1.56	0.43	0.44	1.23	1.41
mobile block	MBH_{py}	0.61	0.62	1.30	1.49	0.39	0.40	1.18	1.36
(several)	$MBH_{Ph,\,iP}$	0.73	0.74	1.42	1.61	0.39	0.40	1.20	1.37
	$MBH_{Ph,\,py}$	0.67	0.68	1.36	1.55	0.39	0.41	1.20	1.38
	$MBH_{iP,\,py}$	0.67	0.68	1.37	1.56	0.40	0.42	1.22	1.39
	$MBH_{Ph,\,iP,\,py}$	0.75	0.76	1.45	1.64	0.41	0.42	1.22	1.39

[a] See caption of Table 4.

similar to those of the quantum oscillator results and are not discussed separatedly.

First, consider the FHVA results of the reaction of bortezomib with H_2O_2 to form the alcohol product PROD1 (Tables 4 and 5). The vibrational contribution to the reaction free energy is unfavorable, since the ΔG_{vib} values are positive. The QM ΔG_{vib} [FHVA] results range from 1.78 to 2.40 kcal/mol, and the QM/MM ΔG_{vib} [FHVA] results range from 2.28 to 2.93 kcal/mol. Both the internal energy and the entropic part are positive for all four functionals. The results of the reaction with methanol to form the ether product PROD2 are more complex (Tables 6 and 7). The QM ΔG_{vib} [FHVA] results range from −1.11 to 0.63 kcal/mol, where only the B3LYP-D functional has a positive value, whereas the QM/MM ΔG_{vib} [FHVA] results range from 0.16 to 0.59 kcal/mol, all being positive. The vibrational

Table 8. Bortezomib—Average Deviations of the Absolute Value of the Vibrational Free Energy (δG_{vib}) and Average Deviations of the Vibrational Free Energy Difference ($\delta \Delta G_{vib}$) Caused by Several Calculation Parameters: The QM versus QM/MM Description, The Choice of Functional (B3LYP, PBE, B3LYP-D, PBE-D), The NMA Model (FHVA, FHVA^{Eck}, MBH, MBH^{Eck}) and the Mobile Block Hessian Implementation (case B versus case C)

source of deviation	δG_{vib} [kcal/mol]	
FHVA vs FHVA^{Eck}	0.18 (QM)	0.02 (QM/MM)
source of deviation	**$\delta \Delta G_{vib}$ [kcal/mol]**	
QM vs QM/MM (FHVA)	0.62	
functional (FHVA)	0.53 (QM)	0.20 (QM/MM)
FHVA vs FHVA^{Eck}	0.11 (QM)	0.02 (QM/MM)
MBH vs MBH^{Eck}	0.10 (QM)	0.03 (QM/MM)
MBH vs FHVA	0.09 (QM)	0.04 (QM/MM)
MBH^{Eck} vs FHVA^{Eck}	0.11 (QM)	0.04 (QM/MM)
case B vs case C (MBH)	–	0.03 (QM/MM)

effect on the reaction kinetics in the QM description is therefore unclear for this reaction when comparing the four QM functionals. The internal energy difference of the quantum oscillators is always negative, but the entropic part depends heavily on the choice of the potential. The QM/MM description is more consistent: the vibrational free energy contribution is systematically unfavorable for the reaction ($\Delta G_{vib} > 0$), with the internal energy difference being negative and the entropic part positive.

As mentioned in the previous subsection, the effect of imposing the Eckart constraints on the absolute free energy values is a good measure for the accuracy of the Hessian. The differences between the values G_{vib} and G_{vib}^{Eck} are taken up in Table S1 of the Supporting Information and the average deviations in Table 8. It is found that QM Hessians are very sensitive to the Eckart constraints with an average shift in G_{vib} of 0.18 kcal/mol, while the sensitivity of QM/MM Hessians is noticeably better with an average shift of 0.02 kcal/mol. Note that these average deviations are based on the rather large REA, PROD1, and PROD2 molecules, since those have floppy modes with low-lying frequencies, which are absent in the smaller species. The different effect on QM and QM/MM Hessians can be explained, on one hand, by the geometry convergence of a QM system being delicate—since it depends on QM gradients which are sensitive to numerical integration errors themselves as well. On the other hand, the analytical QM second derivatives are particularly sensitive to the numerical integration accuracy of two electron integrals and convergence criteria of iterative loops (e.g., SCF loop, CPSCF loop). With 50 or more QM atoms in the QM description and only 11 or less atoms in the QM region of the QM/MM description, the QM/MM calculations are thus more accurate. These average deviations should be considered as errors inherent to the calculated data, originating from the present computational settings, just like experimental data having a limited accuracy imposed by the experimental setup.

The effect of the Eckart conditions on the reaction free energy can be seen in Tables 4−7 by comparing ΔG_{vib} with ΔG_{vib}^{Eck}. For QM Hessians, the deviation between FHVA and FHVA^{Eck} is on average about 0.11 kcal/mol and, for QM/MM Hessians, 0.02 kcal/mol. Not surprisingly, a deviation between FHVA and FHVA^{Eck} brings along a comparable deviation between MBH and MBH^{Eck}. The rather small deviations illustrate that the errors on the absolute vibrational free energies occasianally cancel out when taking the difference, but this behavior is not guaranteed. The expected accuracy of the free energy difference therefore must be on the same order as the accuracy of the absolute free energy values themselves, which is indeed the case.

Next, the influence of the use of MBH frequencies on the free energy difference is discussed, by comparing ΔG_{vib} [FHVA] with ΔG_{vib} [MBH]. The MBH approximation produces an average error of 0.09 and 0.04 kcal/mol when derived from QM and QM/MM Hessians, respectively, where the average is taken over all levels of theory and all block choices. The *absolute* free energies are drastically reduced by the MBH approach by over 145 kcal/mol when three blocks are used (data not shown); however, this significant shift is consistent between reactants and products such that it mostly cancels out when considering the free energy *differences* in eqs 32 and 33. This means that MBH alters the FHVA results relatively little. A comparison of ΔG_{vib}^{Eck} [FHVA] and ΔG_{vib}^{Eck} [MBH] shows similar errors of MBH^{Eck} with respect to FHVA^{Eck}. This significant cancellation of errors is in agreement with an earlier study on the reproduction of reaction rate constants with MBH by Ghysels et al.,[74] where errors canceled out in the difference $G(ts) - G(rea)$ between the transition state and the reactants. In the present study, the cancellation implies that the internal motions of the proposed blocks Ph, iP, and py are not crucial for the reactive behavior of the chemically active boron center, as expected. Indeed, the use of MBH has as much effect as have the Eckart constraints, which is a measure of the best accuracy that can be obtained with the given data (i.e., the Hessians).

When applied to transition state geometries, our approach can be used to estimate tunneling corrections, kinetic isotope effects, and local free energy estimates in the harmonic limit, without the cost of explicit conformational sampling. Indeed, for the estimates to be meaningful, only the frequency and the character of the lowest modes need be accurate. For instance, the kinetic isotope effect is closely related to the ratio of partition functions, which are governed by the low frequency modes. The studies in refs 25, 27, 37, 71, 74, and 75 have shown that MBH errors in partition functions, in free energy differences, and in reaction rates indeed cancel out when comparing two conformations if the same block choice is applied. Also, the application in the present paper supports the idea of the cancellation of errors. This makes MBH a promising method when it is applied on transition states, in particular for examining enzymatic reactions employing the QM/MM description.

The direct calculation of the QM/MM mobile block Hessian (case C, bottom part of tables) should yield the exact same frequencies as those based on the projection of the QM/MM full Hessian (case B, middle part of tables). The difference of 0.03 kcal/mol on average is indeed minimal and should be explained by numerical inaccuracies such as the finite convergence criteria in the CPSCF routine in Q-Chem and the numerical integration. In conclusion, the

new MBH implementation in the parallel Q-Chem/CHARMM interface is capable of reproducing the reference (FHVA) ΔG with satisfying accuracy.

Moreover, close inspection of the MBH values corresponding to one, two, or three blocks confirms the product rule as established in ref 74. In this paper, the correction in reaction rate due to the introduction of multiple blocks was found to be the product of the corrections due to the presence of each block individually. Similarly, the small deviations in free energy $\Delta\Delta G$ due to the introduction of single blocks approximately add up to the deviation in free energy due to multiple blocks, e.g., $\Delta\Delta G(\mathrm{MBH_{Ph}}) + \Delta\Delta G(\mathrm{MBH_{py}}) + \Delta\Delta G(\mathrm{MBH_{iP}}) \approx \Delta\Delta G(\mathrm{MBH_{Ph,\,py,\,iP}})$.

The variations in ΔG_{vib} induced by the choice of NMA model, more specifically the MBH, are negated by different errors. Indeed, the choice of a QM versus a QM/MM treatment shifts the FHVA values by 0.62 kcal/mol on average. The choice of the functional, on the other hand, is responsible for an average shift of 0.53 and 0.20 kcal/mol in the QM and QM/MM cases, respectively. These values lie higher than the typical errors encountered by MBH (0.10 for QM, 0.04 for QM/MM) such that MBH can be considered a minor source of deviations of the vibrational free energy difference, on the same order as the inherent accuracy of the data (0.10 for QM, 0.03 for QM/MM). Table 8 displays the hierarchy of the errors. The introduction of mobile blocks in the vibrational analysis has a significantly lower effect on ΔG compared to other computation parameters such as the functional and the QM versus QM/MM treatment. Therefore, the MBH approach not only is favorable because of the reduction in computational cost and memory requirements but is also capable of reproducing the full Hessian results with satisfactory to excellent accuracy.

V. Conclusion

The computation of vibrational frequencies from the analytical second derivatives matrix is a bottleneck in the hybrid QM/MM description due to the long-range Coulomb interactions when the electrostatic embedding scheme is employed. This is even the case for a small number of QM atoms, since additional CPSCF equations need to be solved for each MM atom displacement. Instead of using a cutoff technique that neglects interactions beyond a certain cutoff distance r_c, we introduce mobile blocks in the MM region. These blocks can translate/rotate as a whole, but their internal degrees of freedom are frozen in the vibrational analysis. Consequently, fewer CPSCF equations need to be solved, leading to a reduction in both computation time and memory requirements. In the chorismate mutase example, MBH decreases the CPU time to 54% of the full Hessian calculation, and the memory is reduced by a factor of roughly 5, when using mobile blocks of 15 atoms each. The computational profit further increases with increasing block size.

In this paper, the MBH formalism is established in view of the QM/MM interface between Q-Chem and CHARMM. A parallel version of MBH in the Q-Chem/CHARMM interface is now implemented in the latest version. Moreover, special attention is paid to the treatment of link atoms. The presence of link atoms creates artificial degrees of freedom which should be projected out of the Hessian in accordance with the definition of the total energy and the constraints imposed during the geometry optimization. Our suggestion is to impose the link atom to be located at a fixed scaled distance collinearly with the QM host atom and MM host atom. Formulas for the corresponding projection have been developed, and this projection is now available in the Q-Chem/CHARMM interface.

As an illustrative example, the vibrational free energy of bortezomib and the products after oxidative deboronation with the reagents H_2O_2 and methanol have been studied extensively, with four different levels of theory and a series of MBH block choices. Our results for this particular test system show an inherent error of 0.10 kcal/mol for the QM and 0.03 kcal/mol for the QM/MM vibrational free energy differences, which is quantified by imposing Eckart constraints. The introduction of mobile blocks introduces an error at a similar order of magnitude: 0.10 kcal/mol for QM and 0.04 kcal/mol for QM/MM vibrational free energy differences. Therefore, the considered block choices are reasonable approximations, especially given that much larger deviations are caused by the choice of functional (0.53 to 0.20 kcal/mol) or by the QM versus QM/MM description (0.62 kcal/mol). MBH is thus not only a computationally attractive method but also an adequate approximate approach for the calculation of thermodynamic quantities such as vibrational free energy differences.

Acknowledgment. This work is supported by the Fund for Scientific Research—Flanders (FWO), the Research Board of Ghent University (BOF), and BELSPO in the frame of IAP 6/27. This work is also supported by the Intramural Research Program of the National Heart, Lung and Blood Institute, National Institutes of Health (NIH). Funding was also received from the European Research Council under FP7 with ERC grant agreement number 240483. H.L.W. would like to acknowledge NIH (1K22HL088341-01A1) and the University of South Florida (start-up) for funding. Y.S. and J.K. would like to thank NIH for a Small Business Innovative Research grant (GM073408). Computational resources and services used in this work were provided by the Lobos cluster of the National Institutes of Health.

Supporting Information Available: The effect of imposing the Eckart constraints is a good measure for the accuracy of the Hessian; Table S1 contains the deviation of the vibrational free energy $\delta G_{\mathrm{vib}} = G_{\mathrm{vib}} - G_{\mathrm{vib}}^{\mathrm{Eck}}$ for the molecules REA, PROD1, and PROD2 calculated with the B3LYP, PBE, B3LYP-D, and PBE-D functionals and 6-311++G(d,p) basis set. It shows that QM Hessians are very sensitive to the Eckart constraints, while the sensitivity of the QM/MM Hessians is noticeably better.

This information is available free of charge via the Internet at http://pubs.acs.org/.

References

(1) Wilson, E. B.; Cross, P. C.; Decius, J. C. *Molecular Vibrations*; Dover Publications: New York, 1980.

(2) Fessenden, R. J.; Fessenden, J. S. *Organic chemistry*, 4th ed.; Brooks/Cole Publishing Company: Belmont, CA, 1990; pp 323−339.

(3) Cui, Q.; Bahar, I. *Normal Mode Analysis: Theory and applications to biological and chemical systems*; Chapman & Hall/CRC, Taylor & Francis Group: Boca Raton, FL, 2006; Mathematical and Computational Biology Series.

(4) Pulay, P. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules. I. Theory. *Mol. Phys.* **1969**, *17*, 197.

(5) Pople, J. A.; Krisnhan, R.; Schlegel, H. B.; Binkley, J. S. Derivative studies in Hartree-Fock and Moller-Plesset theories. *Int. J. Quant. Chem.* **1979**, *Symp. 13*, 225–41.

(6) Saxe, P.; Yamaguchi, Y.; Schaefer, H. G. Analytic 2nd derivatives in restricted Hartree-fock theory - a method for hig-spin open-shell molecular wave-functions. *J. Chem. Phys.* **1982**, *77*, 5647–5954.

(7) Osamura, Y.; Yamaguchi, Y.; Saxe, P.; Vincent, M. A.; Gaw, J. F.; Schaefer, H. F. Unified theoretical treatment of analytic first and second energy derivatives in open-shell Hartree-Fock theory. *J. Chem. Phys.* **1982**, *72*, 131–139.

(8) Osamura, Y.; Yamaguchi, Y.; Saxe, P.; Fox, D. J.; Vincent, M. A.; Schaefer, H. F. Analytic 2nd derivative techniques for self-consistent-field wave-functions - a new approach to the solution of the coupled perturbed Hartree-Fock equations. *J. Mol. Struct.* **1983**, *103*, 183–196.

(9) Yamaguchi, Y.; Frisch, M. J.; Gaw, J.; Schaefer, H. F.; Binkley, J. S. Analytic evaluation and basis set dependence of intensities of infrared spectra. *J. Chem. Phys.* **1986**, *84*, 2262.

(10) Frisch, M. J.; Yamaguchi, Y.; Gaw, J.; Schaefer, H. F.; Binkley, J. S. Analytic Raman intensities from molecular electronic wave-functions. *J. Chem. Phys.* **1986**, *84*, 531.

(11) Frisch, M.; Head-Gordon, M.; Pople, J. Direct analytic SCF 2nd derivatives and electric-field properties. *Chem. Phys.* **1990**, *141*, 189–196.

(12) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

(13) Singh, U. C.; Kollman, P. A. A Combined Abinitio Quantum-Mechanical and Molecular Me-chanical Method for Carrying out Simulations on Complex Molecular-Systems - Applications to the Ch3cl + Cl- Exchange-Reaction and Gas-Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718–730.

(14) Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum-Mechanical and Molecular Mechanical Potential for Molecular-Dynamics Simulations. *J. Comput. Chem.* **1990**, *11*, 700–733.

(15) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(16) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A multilayered integrated MO+MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt(P(t-Bu)(3))(2)+H-2 oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

(17) Dapprich, S.; Komaromi, I.; Byun, K.; Morokuma, K.; Frisch, M. J. A new ONIOM imple-mentation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *THEOCHEM* **1999**, *461*, 1–21.

(18) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in methods and algorithms in a modern quantum chemistry program package. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172–3191.

(19) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

(20) Woodcock, H. L.; Hodoscek, M.; Gilbert, A. T. B.; Gill, P. M. W.; Schaefer, H. F.; Brooks, B. R. Interfacing Q-Chem and CHARMM to perform QM/MM reaction path calculations. *J. Comput. Chem.* **2007**, *28*, 1485–1502.

(21) Gao, J.; Truhlar, D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.

(22) Senn, H. M.; Thiel, W. QM/MM methods for biological systems. In *Atomistic Approaches in Modern Biology: from Quantum Chemistry to Molecular Simulations*; Springer-Verlag: Berlin, 2007; Vol. 268, pp 173−290.

(23) Vreven, T.; Byun, K. S.; Komromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. Combining Quantum Mechanics Methods with Molecular Mechanics Methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815.

(24) Cui, Q.; Karplus, M. Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations. *J. Chem. Phys.* **2000**, *112*, 1133–1149.

(25) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. Vibrational modes in partially optimized molecular systems. *J. Chem. Phys.* **2007**, *126*, 224102.

(26) Ghysels, A.; Van Neck, D.; Waroquier, M. Cartesian formulation of the Mobile Block Hessian approach to vibrational analysis in partially optimized systems. *J. Chem. Phys.* **2007**, *127*, 164108.

(27) Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Catak, S.; Brooks, B. R.; Van Neck, D.; Waroquier, M. Comparative study of various normal mode analysis techniques based on partial Hessians. *J. Comput. Chem.* **2010**, *31*, 994–1007.

(28) Currently described functionality is included in CHARMM version 36a3 and later and the current development version of Q-Chem (scheduled to be released as part of version 4.0).

(29) Adams, J.; Behnke, M.; Chen, S.; Cruickshank, A. A.; Dick, L. R.; Grenier, L.; Klunder, J. M.; Ma, Y. T.; Plamondon, L.; Stein, R. L. Potent and selective inhibitors of the proteasome: dipeptidyl boronic acids. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 333–338.

(30) Adams, J.; Kaufmann, M. Development of the proteasome inhibitor velcade (bortezomib). *Cancer Invest.* **2004**, *22*, 304–311.

(31) Goldstone, J.; Salam, A.; Weinberg, S. Broken symmetries. *Phys. Rev.* **1962**, *127*, 965–970.

(32) Angyan, J. G. Wigner's (2n+1) rule for nonlinear Schrodinger equations. *J. Math. Chem.* **2009**, *46*, 1–14.

(33) Head-Gordon, M.; Pople, J. A. Optimization of wave function and geometry in the finite basis hartree-fock method. *J. Phys. Chem.* **1988**, *92*, 3063–3069.

(34) Ochsenfeld, C.; Head-Gordon, M. A reformulation of the coupled perturbed self-consistent field equations entirely within a local atomic orbital density matrix-based scheme. *Chem. Phys. Lett.* **1997**, *270*, 399–405.

(35) Woodcock, H. L.; Zheng, W. J.; Ghysels, A.; Shao, Y. H.; Kong, J.; Brooks, B. R. Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. *J. Chem. Phys.* **2008**, *129*, 214109.

(36) Liang, W.; Zhao, Y.; Head-Gordon, M. An efficient approach for self-consistent-field energy and energy second derivatives in the atomic-orbital basis. *J. Chem. Phys.* **2005**, *123*, 194106.

(37) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Brooks, B. R.; Waroquier, M. Normal modes for large molecules with arbitrary link constraints in the Mobile Block Hessian approach. *J. Chem. Phys.* **2009**, *130*, 084107.

(38) Das, D.; Eurenius, K. P.; Billings, E. M.; Sherwood, P.; Chatfield, D. C.; Hodoscek, M.; Brooks, B. R. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J. Chem. Phys.* **2002**, *117*, 10534–10547.

(39) Li, H.; Jensen, J. H. Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.* **2002**, *107*, 211–219.

(40) Jin, S. Q.; Head, J. D. Theoretical Investigation of Molecular Water-Adsorption on the Al(111) Surface. *Surf. Sci.* **1994**, *318*, 204–216.

(41) Calvin, M. D.; Head, J. D.; Jin, S. Q. Theoretically modelling the water bilayer on the Al(111) surface using cluster calculations. *Surf. Sci.* **1996**, *345*, 161–172.

(42) Head, J. D. Computation of vibrational frequencies for adsorbates on surfaces. *Int. J. Quantum Chem.* **1997**, *65*, 827–838.

(43) Head, J. D.; Shi, Y. Characterization of Fermi resonances in adsorbate vibrational spectra using cluster calculations: Methoxy adsorption on Al(111) and Cu(111). *Int. J. Quantum Chem.* **1999**, *75*, 815–820.

(44) Head, J. D. A vibrational analysis with Fermi resonances for methoxy adsorption on Cu(111) using ab initio cluster calculations. *Int. J. Quantum Chem.* **2000**, *77*, 350–357.

(45) Besley, N. A.; Metcalf, K. A. Computation of the amide I band of polypeptides and proteins using a partial Hessian approach. *J. Chem. Phys.* **2007**, *126*, 035101.

(46) Derat, E.; Bouquanta, J.; Humbel, S. On the link atom distance in the ONIOM scheme. An harmonic approximation analysis. *THEOCHEM* **2003**, *632*, 61–69.

(47) MacKerel, A., Jr.; Brooks, C., III; Nilsson, L.; Roux, B.; Won, Y.; Karplus, M. In *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; Schleyer, v. R. et al., Eds.; John Wiley & Sons: Chichester, U. K., 1998; Vol. 1, pp 271−277.

(48) Bakowies, D.; Thiel, W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(49) Swart, M. AddRemove: A new link model for use in QM/MM studies. *Int. J. Quantum Chem.* **2003**, *91*, 177–183.

(50) CP2K Developers Home Page. http://cp2k.berlios.de (accessed October 15, 2010).

(51) Korambath, P. P.; Kong, J.; Furlani, T. R.; Head-Gordon, M. Parallelization of analytical Hartree-Fock and density functional theory Hessian calculations. Part I: parallelization of coupled-perturbed Hartree-Fock equations. *Mol. Phys.* **2002**, *100*, 1755–1761.

(52) Woodcock, H. L.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F.; Brooks, B. R. Exploring the quantum mechanical/molecular mechanical replica path method: a pathway optimization of the chorismate to prephenate Claisen rearrangement catalyzed by chorismate mutase. *Theor. Chem. Acc.* **2003**, *109*, 140–148.

(53) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R., Jr.; Evanseck, J. D.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J. Phys. Chem.* **1998**, *102*, 3586–3616.

(54) MacKerell, A. D., Jr; Feig, M.; Brooks, C. L., III. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(55) Fahy, B. N.; Schlieman, M. G.; Virudachalam, S.; Bold, R. J. Schedule-dependent molecular effects of the proteasome inhibitor bortezomib and gemcitabine in pancreatic cancer. *J. Surg. Res.* **2003**, *113*, 88–95.

(56) Nawrocki, S. T.; Carew, J. S.; Pino, M. S.; Highshaw, R. A.; Andtbacka, R. H. I.; Dun-ner, K.; Pal, A.; Bornmann, W. G.; Chiao, P. J.; Huang, P.; Xiong, H.; Abbruzzese, J. L.; McConkey, D. J. Aggresome disruption: a novel strategy to enhance bortezomib-induced apoptosis in pancreatic cancer cells. *Cancer Res.* **2006**, *66*, 3773–3781.

(57) Nawrocki, S. T.; Carew, J. S.; Pino, M. S.; Highshaw, R. A.; Dunner, K.; Huang, P.; Abbruzzese, J. L.; McConkey, D. J. Bortezomib borate anion in a hydrogen-bonded host lattice sensitizes pancreatic cancer cells to endoplasmic reticulum stress-mediated apoptosis. *Cancer Res.* **2005**, *65*, 11658–11666.

(58) McCormack, T.; Baumeister, W.; Grenier, L.; Moomaw, C.; Plamondon, L.; Pramanik, B.; Slaughter, C.; Soucy, F.; Stein, R. L.; Zuhl, G.; Dick, L. R. Active site-directed inhibitors of phodococcus 20 S proteasome: kinetics and mechanism. *J. Biol. Chem.* **1997**, *272*, 26103–26109.

(59) Pekol, T.; Daniels, J. S.; Labutti, J.; Parsons, I.; Nix, D.; Baronas, E.; Hsieh, F.; Gan, L.-S.; Miwa, G. Human metabolism of the proteasome inhibitor bortezomib: identication of circulating metabolites. *Drug Metab. Dispos.* **2005**, *33*, 771–777.

(60) Labutti, J.; Pearsons, I.; Huang, R.; Miwa, G.; Gan, L.-S.; Daniels, J. S. Oxidative deboronation of the peptide boronic acid proteosome inhibitor bortezomib: contributions from reactive oxygen species in this novel cytochrome P450 reaction. *Chem. Res. Toxicol.* **2006**, *19*, 539–546.

(61) Larkin, J. D.; Markham, G. D.; Milkevitch, M.; Brooks, B. R.; Bock, C. W. Computational Investigation of the Oxidative Deboronation of Boroglycine, H_2N-CH_2-$B(OH)_2$, Using H_2O and H_2O_2. *J. Phys. Chem. A* **2009**, *113*, 11028–11034.

(62) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic be-havior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(63) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B* **1988**, *37*, 785–789.

(64) Ernzerhof, M.; Perdew, J. P.; Burke, K. Coupling-constant dependence of atomization energies. *Int. J. Quantum Chem.* **1997**, *64*, 285.

(65) Ernzerhof, M.; Scuseria, G. E. Assessment of the Perdew-Burke-Ernzerhof exchange-correlation functional. *J. Chem. Phys.* **1999**, *110*, 5029–5036.

(66) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable pa-rameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–69.

(67) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dis-persion correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(68) Feyereisen, M.; Fitzegerald, G.; Komornicki, A. Use of approximate integrals in ab initio theory. An application in MP2 energy calculations. *Chem. Phys. Lett.* **1993**, *208*, 359–363.

(69) Woon, D. E.; Dunning, T. H. J. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975–2988.

(70) Vangunsteren, W. F.; Berendsen, H. J. C. Algorithms for Macromolecular Dynamics and Constraint Dynamics. *Mol. Phys.* **1977**, *34*, 1311–1327.

(71) Ghysels, A.; Verstraelen, T.; Hemelsoet, K.; Van Speybroeck, V.; Waroquier, M. TAMkin: a versatile package for vibrational analysis and chemical kinetics. *J. Chem. Inf. Model.* **2010**, 1736–1750.

(72) Grochowski, P. Rotational symmetry of the molecular potential energy in the Cartesian coordinates. *Theor. Chem. Acc.* **2008**, *121*, 257–266.

(73) Brandhorst, K.; Grunenberg, J. Efficient computation of compliance matrices in redundant internal coordinates from Cartesian Hessians for nonstationary points. *J. Chem. Phys.* **2010**, *132*, 184101.

(74) Ghysels, A.; Van Speybroeck, V.; Verstraelen, T.; Van Neck, D.; Waroquier, M. Calculating reaction rates with partial Hessians: Validation of the mobile block Hessian approach. *J. Chem. Theory Comput.* **2008**, *4*, 614–625.

(75) Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Van Neck, D.; Brooks, B. R.; Waroquier, M. Mobile Block Hessian approach with adjoined blocks: an efficient approach for the calculation of frequencies in macromolecules. *J. Chem. Theory Comput.* **2009**, *5*, 12031215.

(76) Mc Quarrie, D. A.; Simon, J. D. *Physical Chemistry - a molecular approach*; University Science Books: Sausalito, CA, 1997; pp 1075−1079.

(77) Brooks, B. R.; Janezic, D.; Karplus, M. Harmonic-Analysis of Large Systems. 1. Methodology. *J. Comput. Chem.* **1995**, *16*, 1522–1542.

(78) Janezic, D.; Brooks, B. R. Harmonic-Analysis of Large Systems. 2. Comparison of Different Protein Models. *J. Comput. Chem.* **1995**, *16*, 1543–1553.

(79) Janezic, D.; Venable, R. M.; Brooks, B. R. Harmonic-Analysis of Large Systems. 3. Comparison with Molecular-Dynamics. *J. Comput. Chem.* **1995**, *16*, 1554–1566.

CT100473F