

Ontwikkeling van een accurate en efficiënte methode
voor normale-mode-analyse in uitgebreide moleculaire systemen:
de Mobile Block Hessian methode

Development of an Accurate and Efficient Method
for Normal Mode Analysis in Extended Molecular Systems:
the Mobile Block Hessian Method

An Ghysels

Promotor: prof. dr. M. Waroquier
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Toegepaste Natuurkunde

Vakgroep Subatomaire en Stralingsfysica
Voorzitter: prof. dr. D. Ryckbosch
Faculteit Wetenschappen
Academiejaar 2008 - 2009



ISBN 978-90-8578-280-3
NUR 928
Wettelijk depot: D/2009/10.500/38



Dit onderzoekswerk werd uitgevoerd in het Centrum voor Moleculaire
Modelling onder leiding van prof. dr. Michel Waroquier.

In de eerste plaats zou ik mijn promotor, prof. Michel Waroquier, willen bedanken. Hij overtuigde me op een of andere manier om na een spetterend Erasmusjaar in Grenoble, de poort tot de Alpen, weer terug te komen naar de Proeftuinstraat in Gent. Zo zakte ik qua onderzoeksfocus vele grootteordes naar beneden, en ruilde de seismische golven van mijn masterthesis in voor de wereld van de vibrerende molecules. Niet alleen kon ik tijdens mijn doctoraat op de steun van mijn promotor rekenen, maar ook op de vele competenties van de andere collega's, namelijk Saron Catak, Diederica Claeys, Stefaan Cottenier, Reinout Declerck, Hendrik De Cooman, Bart De Sterck, Matthias Degroote, Wim Dewitte, Karen Hemelsoet, David Lesthaeghe, Emmanuel Baribefe Naziga, Ewald Pauwels, Nooriya Sakhabutdinova, Franky Stevens, Karen Van Cauter, Diederik Van Fleteren, Marc Van Houteghem, Dimitri Van Neck, Veronique Van Speybroeck, Matthias Vandichel, Peter Vansteenkiste, Stijn Verdonck, Brecht Verstichel, and Toon Verstraelen. In het bijzonder kon prof. Veronique Van Speybroeck dankzij haar brede kennis steeds interessante toepassinggerichte pistes bedenken voor mijn eerder theoretische ideeën, en kon ik met prof. Dimitri Van Neck elk symbooltjes-gerelateerd probleem grondig uitspitten. Ik heb van hun inbreng veel geleerd. Sfeer op de werkvloer is eveneens een niet te onderschatten ingrediënt voor een geslaagd doctoraat, dus bij deze merci aan mijn dierbare bureaugenoten. I am very grateful to Bernie Brooks for inviting me to his research laboratory, where I made part of a dynamic team. The stay at NIH in Maryland certainly opened my view on the United States.

Buiten de werksfeer waren het vooral de vrienden die me aan de gang hielden. Merci voor de oppeppers, vooral aan Lydie die met haar meteorieten voor een keerpunt zorgde! Mijn klarinet en het jeugdwerk waren twee andere constanten doorheen mijn doctoraatstijd waaraan ik veel voldoening heb gehad. Tot slot petje af voor mijn familie en vooral mijn ouders voor alle ondersteuning.

An Ghysels 19 mei 2009

Leden van de examencommissie

Voorzitter

prof. dr. ir. Luc Taerwe (Universiteit Gent, IR14)

Leescommissie

prof. dr. Bernard R. Brooks (National Institutes of Health)

prof. dr. Frans Cantrijn (Universiteit Gent, WE03)

prof. dr. ir. Marie-Françoise Reyniers (Universiteit Gent, IR12)

prof. dr. Dimitri Van Neck (Universiteit Gent, WE05)

prof. dr. Michel Waroquier (Universiteit Gent, WE05)

Overige leden

prof. dr. Paul Geerlings (Vrije Universiteit Brussel)

dr. Ewald Pauwels (Universiteit Gent, WE05)

prof. dr. ir. Patrick Segers (Universiteit Gent, IR15)

dr. Stan van Gisbergen (Scientific Computing & Modelling NV)

prof. dr. ir. Veronique Van Speybroeck (Universiteit Gent, IR17)

Contents

Nederlandstalige samenvatting	xiii
--------------------------------------	-------------

English summary	xvii
------------------------	-------------

List of abbreviations	xxi
------------------------------	------------

1 Harmonic oscillator approximation	1
1.1 Introduction	1
1.2 Normal mode analysis equations	6
1.2.1 NMA in Cartesian coordinates	7
1.2.2 NMA in internal and external coordinates	9
1.3 NMA and nonstationary points	10

List of abbreviations	1
------------------------------	----------

2 The MBH method	15
2.1 PHVA	15
2.2 MBH equations	16
2.2.1 In internal coordinates	17
2.2.2 In Cartesian coordinates	18
2.3 Extensions to MBH	21
2.3.1 MBH with multiple blocks	21
2.3.2 MBH with linear blocks	21

2.3.3	General formulation of MBH	22
2.4	MBH with adjoined blocks	24
2.4.1	Restraint technique with dummy atoms	27
2.4.2	Analytical constraints	28
2.5	Procedures and implementation	32
2.5.1	Standard NMA procedure with the full Hessian	33
2.5.2	MBH in internal coordinates	35
2.5.3	MBH using group parameters as variables	35
<hr/>		
3	Validation of MBH	41
<hr/>		
3.1	Introduction	41
3.1.1	Investigated quantities	41
3.1.2	Validation tools	42
3.2	MBH versus PHVA	43
3.2.1	The difference between MBH and PHVA is a mass effect	43
3.2.2	Illustration: ethanol molecule	45
3.3	Reaction rate constants	50
3.3.1	MBH reproduces reaction rate constants	50
3.3.2	Illustration: modeling a reaction in solvent	55
3.4	Low frequency modes	58
3.4.1	MBH with adjoined blocks to describe the low frequency modes of proteins	58
3.4.2	Illustration: crambin	60
<hr/>		
4	Conclusion and perspectives	69
<hr/>		
4.1	Conclusion	69
4.2	Perspectives	70
<hr/>		
	Bibliography	75
<hr/>		
	Appendix	83
<hr/>		
	Paper I: Vibrational modes in partially optimized molecular systems	83
<hr/>		

Paper II: Cartesian formulation of the Mobile Block Hessian approach to vibrational analysis in partially optimized systems	97
<hr/>	
Paper III: Calculating reaction rates with partial Hessians: validation of the Mobile Block Hessian approach	107
<hr/>	
Paper IV: MFI Fingerprint: how pentasil-induced IR bands shift during zeolite nanogrowth	121
<hr/>	
Paper V: Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit	129
<hr/>	
Paper VI: Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach	139
<hr/>	
Paper VII: Mobile Block Hessian approach with adjoined blocks: An efficient approach for the calculation of frequencies in macromolecules	153
<hr/>	
Paper VIII: Comparative study of various normal mode analysis techniques based on partial Hessians	167
<hr/>	
List of publications	209

Nederlandstalige samenvatting

Vibratoire spectroscopie is een belangrijke techniek om de structuur van (bio)moleculen en (nano)materialen te onderzoeken. Zo is het bijvoorbeeld bijzonder geschikt om proteïnes te bestuderen in hun natuurlijke omgeving (in waterige oplossing), en kan het in vele gevallen gebruikt worden daar waar X-stralenkristallografie en nucleaire-magnetische-resonantiespectroscopie falen. Meer bepaald worden infrarood- (IR) en Ramanspectroscopie veelvuldig gebruikt om informatie in te winnen over de secundaire structuur van proteïnes. Ook in andere gebieden helpen deze technieken om de functionele groepen van materialen te identificeren, of geven ze aanleiding tot een unieke "vingerafdruk" van het materiaal door de zogenaamde skeletvibraties.

Een frequent voorkomend probleem is de preciese interpretatie van de experimenteel opgemeten spectra. Uitgebreide moleculaire systemen zijn gekenmerkt door een rijk vibrationeel spectrum, en de toekenning van de banden aan welbepaalde vibraties is erg moeilijk indien alleen experimentele data voorhanden zijn. Theoretische voorspellingen zijn hierdoor onontkenbaar een complementair hulpmiddel voor de experimenten. Elke geobserveerde band in het IR- of Ramanspectrum bestaat uit een aantal dicht bij elkaar liggende normale modes, die bekomen worden uit de normale-mode-analyse (NMA). Dit is de diagonalisatie van de volledige massagewogen moleculaire Hessiaan, de matrix die de tweedeordeafgeleiden bevat van de totale potentiële energie naar de Cartesische nucleaire coördinaten, geëvalueerd in een evenwichtspunt op het potentiële-energieoppervlak (PES).

Met NMA wordt het systeem behandeld als een set van ongekoppelde harmonische oscillatoren. De frequenties en modes bevatten informatie over de krommingen van het PES en de massadistributie in het systeem. NMA is een statische benadering die het PES lokaal aftast. De benadering is exact indien de

hogere-ordeafgeleiden, i.e. de anharmonische correcties, verwaarloosbaar zijn en is daarom een benaderende techniek die complementair is aan moleculaire dynamica of Monte Carlo-simulaties.

In uitgebreide moleculaire systemen (zoals polypeptides, proteïnes, polymeerketens, supramoleculaire assemblages, systemen in solvent, moleculen geadsorbeerd in een poreus materiaal, enz.), wordt men geconfronteerd met twee problemen. Ten eerste bevatten zulke systemen gemakkelijk enkele honderden of zelfs tienduizenden atomen, en de berekening en diagonalisatie van de volledige Hessiaan zijn computationeel een zware, indien niet onmogelijke, opgave. Ten tweede geeft een volledige berekening, indien al mogelijk, aanleiding tot een enorme hoeveelheid aan data, die nog moeilijk te interpreteren valt. Hierin ligt het doel van dit doctoraatswerk:

Dit doctoraat doelt op de berekening van accurate frequenties en modes in uitgebreide moleculaire systemen op een efficiënte manier.

Er bestaan hoofdzakelijk twee categorieën van benaderende methodes voor normale-mode-analyse: (1) de beschrijving van het PES wordt vereenvoudigd; (2) de beschrijving van het PES blijft onveranderd, maar enkel een beperkt aantal modes wordt berekend op een benaderende wijze. Dit doctoraatswerk focust op deze laatste categorie en stelt de nieuwe Mobile Block Hessian (MBH) methode voor. Het concept dat aan de basis van MBH ligt, is de opdeling van het systeem in verscheidene blokken van atomen, die als starre lichamen bewegen tijdens de vibrationele analyse met enkel translationele en rotationele vrijheidsgraden. De MBH heeft enkele varianten naargelang de blokkeuze en de manier waarop blokken aan elkaar hangen. Momenteel is de MBH geïmplementeerd in de laatste versie van CHARMM en Q-Chem, en zal het model ook beschikbaar zijn in de volgende release van ADF.

Overzicht van het doctoraatswerk

In het inleidende Hoofdstuk 1 wordt normale-mode-analyse voorgesteld als een techniek om het potentiële-energieoppervlak af te tasten binnen de harmonische oscillatorbenadering. De standaard NMA-vergelijkingen met de volledige Hessiaan worden herhaald. De problemen die gepaard gaan met frequentieberekeningen in niet-evenwichtspunten motiveren de noodzaak van een grondige theoretische studie van de normale-mode-analyse van partieel geoptimaliseerde geometrieën, zoals bij MBH het geval is.

Hoofdstuk 2 werkt de MBH-theorie uit in twee coördinatenstelsels: interne coördinaten en blokparameters. Voor de uitbreiding van MBH naar elk type van blokken (inclusief lineaire en één-atoom-blokken) en geconnecteerde blokken (gelinkt door een gemeenschappelijk “adjoining” atoom), wordt de algemene formulering in blokparameters ook uitgedrukt aan de hand van Cartesische grootheden (Cartesische Hessiaan, gradiënt). Vijf praktische implementatieschema’s voor MBH sluiten dit hoofdstuk af.

In Hoofdstuk 3 wordt de performantie van de MBH getoetst voor de reproductie van accurate frequenties en normale modes. Tijdens mijn doctoraat werd een uitgebreid aantal testvoorbeelden onderzocht om de MBH-methode te valideren. In deze samenvatting worden hieruit drie voorbeelden toegelicht. De ethanolmolecule toont aan dat MBH fysische frequenties oplevert bij een partieel geoptimaliseerde structuur, en dat MBH een verbetering vormt ten opzichte van de Partial Hessian Vibrational Analysis (PHVA) omwille van de correcte beschrijving van de massa van de blokken. De MBH is in staat om accurate reactiesnelheidsconstanten te schatten, wat geïllustreerd wordt met een aminofosfonaatreactie in solvent. Het nut van aaneengeschakelde blokken wordt aangetoond met de berekening van de laagste normale modes van crambine, een klein proteïne.

Tenslotte geeft Hoofdstuk 4 enkele besluiten omtrent de performantie van MBH. Perspectieven voor verdere verbeteringen van MBH omvatten de optimalisatie van de implementatie in frequent gebruikte simulatiepakketten, alsook verschillende gecombineerde modellen voor geavanceerde NMA.

Naast MBH vermeldt de literatuur ook andere modellen om de frequenties te berekenen in uitgebreide systemen. In het bijzonder is de Vibrational Subsystem Analysis (VSA) methode van B. R. Brooks een competitief schema. Een vergelijkende studie werd recent afgewerkt over NMA-methodes gebaseerd op Hessianen met een gereduceerde dimensie (partiële Hessianen), in samenwerking met prof. B. R. Brooks van het Laboratory of Computational Biology (NIH) in Bethesda (Maryland). PHVA blijkt hoofdzakelijk geschikt te zijn voor de reproductie van gelokaliseerde modes. MBH kan naast de gelokaliseerde modes ook globalere modes reproducen. De VSA is vooral geschikt voor de reproductie van frequenties en modes in het lagere spectrum. In partieel geoptimaliseerde structuren leiden MBH en PHVA nog steeds tot fysische frequenties. Bovendien laat MBH toe om spectra te analyseren en modes te identificeren door de blokkeuze te variëren. De vergelijkende studie werd opgenomen in de Appendix.

Nederlandstalige samenvatting

Dit doctoraatswerk leidde tot acht artikels, zes in verband met MBH – gepubliceerd, ter perse, of ingediend – en twee artikels die niet direct met de MBH gelinkt zijn. Alle publicaties zijn toegevoegd in de Appendix.

English summary

Vibrational spectroscopy is an important technique for the structural characterization of (bio)molecules and (nano)materials. For example, it is particularly suited for studying proteins in their natural environment (i.e., in aqueous solution), and can be used in many cases where other techniques such as X-ray crystallography and nuclear magnetic resonance spectroscopy cannot be employed. In particular infrared (IR) and Raman spectroscopy have been used extensively for gaining information on the secondary structure of polypeptides and proteins. Also in other fields, these techniques help to identify the functional groups in the material, or provide a unique “fingerprint” of the material, the so-called skeleton vibrations.

A frequently encountered problem in spectroscopy is the precise interpretation of the obtained experimental spectra. Many of these nanostructured systems are characterized by very complex vibrational spectra and the assignment of specific bands to particular vibrations is difficult if based solely on experimental techniques. In this field theoretical predictions form an undeniable complement to the measured spectra. Each observed band in the spectrum consists of a number of close-lying normal modes, which result from normal mode analysis (NMA). This is the diagonalization of the full mass-weighted molecular Hessian matrix, which contains the second derivatives of the total potential energy with respect to Cartesian nuclear coordinates, evaluated in an equilibrium point on the potential energy surface (PES).

By performing NMA, the system is approximated as a set of decoupled harmonic oscillators. The frequencies and modes contain information on the curvatures of the PES and the mass distribution in the system. NMA is a static approach that samples the PES exactly, if higher order derivatives, i.e. anharmonic corrections, are neglected, and is therefore an approximate analysis method complementary to molecular dynamics and Monte Carlo simulations.

In extended molecular systems (like polypeptides, polymer chains, supramolecular assemblies, systems embedded in a solvent or molecules adsorbed within porous materials etc.), this procedure poses two major problems. First, the size of the relevant systems can easily reach a few hundreds or several ten thousands of atoms, and full calculations of such large systems are computationally demanding if not impossible with accurate methods. Second, even if possible, such calculations provide a large amount of data that will be increasingly difficult to interpret. Here lies the scope of this PhD work:

The aim of this PhD is the calculation of accurate frequencies and modes in extended molecular systems in an efficient manner.

Mainly two categories of approximate normal mode calculations can be identified: (1) the PES description is simplified; (2) the description of the PES is unchanged, but only a subset of the modes is calculated in an approximate way. This PhD work focuses on the latter category and presents the new Mobile Block Hessian (MBH) method and its variants. The key concept is the partitioning of the system into several blocks of atoms, which move as rigid bodies during the vibrational analysis with only rotational and translational degrees of freedom. The MBH has several variants according to the block choice and the way blocks are adjoined together. The MBH is currently implemented in the last upgrade of CHARMM and Q-Chem and the method will be available too in the next release of ADF.

Outline PhD thesis

In the introductory Chapter 1, normal mode analysis is presented as a technique to scan the potential energy surface within the harmonic oscillator approximation. The standard NMA equations with the full Hessian are revised. The problems brought up by nonstationary points motivate the necessity of a profound theoretical study of the NMA of partially optimized geometries as is the case for MBH.

Chapter 2 elaborates the MBH theory in two sets of coordinates: internal coordinates and block parameters. For the extension of the MBH to all kind of blocks (including linear, single-atom blocks) and adjoined blocks (linked by a common "adjoining" atom), the general formulation in block parameters is also linked to Cartesian quantities (Cartesian Hessian, gradient). Five practical implementation schemes for MBH conclude this chapter.

In Chapter 3, the MBH is assessed in its performance to reproduce accurate frequencies and normal modes. During my PhD, a large test set has

been investigated to validate the MBH method. In this summary, only three examples are outlined. The ethanol molecule shows how MBH yields physical frequencies for a partially optimized structure, and that MBH is an improvement with respect to the Partial Hessian Vibrational Analysis (PHVA) because of the correct mass description of the block. The MBH is capable of reproducing accurate reaction rate constants given an acceptable block choice, as is illustrated with an aminophosphonate reaction in solvent. The usefulness of adjoined blocks is demonstrated with the calculation of the lowest normal modes of crambin, a small protein.

Finally Chapter 4 gives some concluding remarks on the MBH's performance. Perspectives for the further improvements of MBH include the optimization of the implementation in frequently used program packages, as well as several combined models for advanced NMA.

Besides MBH there are other models in literature for the calculation of frequencies in extended systems. In particular, the vibrational subsystem analysis (VSA) method by B. R. Brooks is a competitive scheme. A comparative study of NMA methods based on Hessians of reduced dimension (partial Hessians) has been accomplished very recently in collaboration with prof. B. R. Brooks of the Laboratory of Computational Biology (National Institutes of Health) in Bethesda (Maryland). PHVA is found to be capable of reproducing localized modes. In addition to localized modes, the MBH can reproduce more global modes. VSA is most suited for the reproduction of the modes and frequencies in the lower spectrum. In partially optimized structures, PHVA and MBH can still yield physical frequencies. Moreover, by varying the size of the blocks, MBH can be used as an analysis tool of the spectrum. The comparative study is added in the Appendix.

This PhD work has resulted in eight papers, six related to MBH – published, in press, or submitted – and two papers not directly related to MBH. All publications are included in the Appendix.

List of abbreviations

BNM	block normal modes
CG	coarse-grained
CPHF	coupled perturbed Hartree-Fock
CPMD	Car-Parrinello molecular dynamics
ENM	elastic network model
IR	infrared
LOT	level of theory
MBH	mobile block Hessian
MC	Monte-Carlo
MD	molecular dynamics
MM	molecular mechanics
MOF	metal-organic framework
NMA	normal mode analysis
PES	potential energy surface
PCA	principal component analysis
PHVA	partial Hessian vibrational analysis
QM	quantum mechanics
QM/MM	hybrid quantum mechanics / molecular mechanics
RTB	rotation-translation blocks
SCF	self consistent field
SVD	singular value decomposition
TS	transition state
TST	transition state theory
VSA	vibrational subsystem analysis

1 Harmonic oscillator approximation

1.1 Introduction

A molecular system is in se built up from nuclei and electrons, which interact through Coulomb forces, i.e. the nuclear repulsion, the electron-electron repulsion and the electron-nucleus attraction. The potential energy is a function of the charges and positions of the nuclei and electrons. In the Born-Oppenheimer approximation, the picture is simplified by the assumption that the motion of the electrons is instantaneous compared to the motion of the nuclei. This assumption is based on the fact that the electron mass is small compared to the mass of the nuclei. The electron cloud is therefore assumed to adapt instantaneously to every displacement of the much heavier nuclei. The potential energy can then be considered as a function of the charges and the positions $\{r_{A\mu}\}$, $\mu = x, y, z$; $A = 1, \dots, N$ of the nuclei, and not the electronic positions $\{r_{i,el}\}$,

$$V = V(\{r_{A\mu}\}) = \min_{r_{i,el}} V(\{r_{A\mu} r_{i,el}\}). \quad (1.1)$$

Since the potential energy is a function of $3N$ coordinates, it represents a $3N$ -dimensional surface. Every configuration $\{r_{A\mu}\}$ corresponds with a specific potential energy V . With the knowledge of the potential energy surface (PES, also known as PEL – potential energy landscape), the equations of motion

$$m_A \ddot{r}_{A\mu} = -\frac{\partial V}{\partial r_{A\mu}}(r) \quad (1.2)$$

can be solved, given a set of initial positions and velocities. This makes the PES a very interesting object to study, as it determines all the dynamics of the

system.

The PES is not flat but typically shows several wells connected by saddle points. An example of a PES, represented as a 2D-surface, is given in Figure 1.1. Special points on the surface are the stationary points, such as the bottom of a well (minimum energy point) and a saddle point (transition state). A system in such a configuration r_0 is gradient-free: $\partial V / \partial r_{A\mu}(r_0) = 0$ for all coordinates. The reference point of a stable system is the minimum energy state. Due to temperature, the system does not remain in the reference point, but it makes small excursions around the minimum energy point. Eventually, the system can cross an energy barrier and end up in another minimum energy state. This corresponds to a chemical reaction or to a conformational change. The probability for this to happen, is proportional to the Boltzmann factor $\exp(-\beta\Delta E)$, with $\beta = 1/k_B T$, T the temperature, k_B the Boltzmann constant, and ΔE the height of the barrier. The probability is influenced by the distribution of energy levels in the well, which is in turn determined by the shape of the well, e.g., a narrow well counts less energy levels than a broad valley.

To investigate chemical reactions and conformational changes, one should scan the PES. In Monte Carlo (MC), the configurational space is sampled, while in Molecular Dynamics (MD), many trajectories on the PES are generated. Statistics based on the samples or trajectories give an estimate of the energy barrier and the shape of the well, from which the reaction kinetics or the accessible conformational changes can be deduced. Normal mode analysis (NMA) is another approach to study the characteristics of the PES locally around the minimum energy point. NMA correctly describes the PES up to second order in the coordinates. The curvatures of the PES are taken into account, which are the second derivatives of the PES with respect to the coordinates, and higher order terms are omitted. This parabolic description of the PES is referred to as the harmonic oscillator approximation or the local harmonic approximation, because a harmonic oscillator is also associated to a parabolic potential energy curve.

NMA is best illustrated with the case of a single harmonic oscillator. Consider the one-dimensional potential energy function given by the Hookean potential, which is quadratic,

$$V(r) = \frac{1}{2}k(r - a)^2 \quad (1.3)$$

with a the equilibrium length of the spring and k the spring constant. The minimum energy state corresponds with $r = a$ in the bottom of the well and a small (large) value of the curvature k corresponds with a narrow (broad) well. A particle with mass m moving in this potential has constant total energy

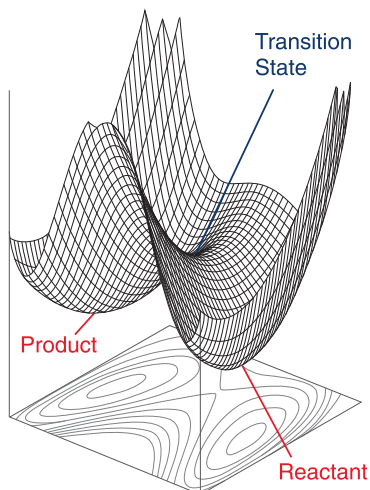


Figure 1.1: Two-dimensional representation of the potential energy surface. Reactant or product states correspond to minimum energy points; the transition state is a saddle point on the surface.

(kinetic plus potential),

$$E = \frac{1}{2}m\dot{r}^2 + \frac{1}{2}k(r - a)^2 \equiv \text{constant}, \quad (1.4)$$

and the corresponding differential equation of second order is given by

$$m\ddot{r} + k(r - a) = 0. \quad (1.5)$$

The trajectory of the particle is a harmonic oscillator: the position oscillates with a frequency $\nu = \omega/2\pi$,

$$r(t) = a + A \cos(\omega t + \phi) \quad (1.6)$$

where $\omega = \sqrt{k/m}$. The amplitude A and the phase ϕ are determined by the initial position $r(0)$ and velocity $\dot{r}(0)$. The frequency contains information both on the mass (m) and the curvature (k).

Similarly, in the $3N$ -dimensional case of an N -particle system, the harmonic approximation of the PES is a quadratic function of the coordinates. The motion of the N -particle system is composed of harmonic oscillations around the minimum of the well. More precisely, the motion is a superposition of

uncoupled harmonic oscillators, each with their own frequency. These specific uncoupled motions – the normal (mass-orthogonal) modes – are specific directions on the PES, which show characteristic main curvatures. The mass distribution is taken into account through mass-weighting of the curvatures. The frequencies and the corresponding modes are calculated by solving a generalized eigenvalue problem. Statistics based on NMA frequencies and modes are exact within the harmonic description of the PES. When the harmonic oscillator approximation is realistic, NMA is thus an ideal way to get information about the local properties of the PES.

The standard procedure to perform NMA is as follows:

1. Geometry optimization. Find an equilibrium state on the PES.
2. Calculate the second derivatives of the PES. The symmetric matrix containing all second derivatives is called the Hessian and has dimension $3N \times 3N$.
3. Solve the generalized eigenvalue problem. The eigenvalues and eigenvectors of the mass-weighted Hessian are calculated.

In extended molecular systems, like polypeptides and proteins, polymer chains, supramolecular assemblies, systems embedded in a solvent or (macro)-molecules adsorbed within porous materials etc., this procedure poses two major problems.

First, the size of the molecular systems can easily reach a few hundreds or several ten thousands of atoms, and full calculations of such large systems are computationally demanding if not impossible with accurate methods. With a quantum mechanical (QM) description, each self consistent field (SCF) cycle is expensive, and the geometry optimization represents a high computational cost. Also the calculation of QM second derivatives is very computationally demanding. In molecular mechanics (MM) simulations, more extended systems are feasible, but one is then confronted with the storage of a huge Hessian (scales as $\sim N^2$) and its diagonalization (scales as $\sim N^3$). Second, even if NMA is possible, such calculations provide a large amount of data that will be increasingly difficult to interpret. Instead of calculating every single frequency and mode, one is often interested in a specific frequency range or mode type.

The first problem can be addressed by using efficient computational methods such as linear-scaling algorithms, or by employing multi-level approaches, in which parts of the molecular system are described at an elaborated quantum

mechanical level of theory while the remaining part is treated with computationally less expensive methods such as molecular mechanics using classical force fields (QM/MM) [1–5]. In the field of biomolecules, the description of the PES is further simplified when aiming at the calculation of the lowest modes related to conformational changes. In Tirion’s approach [6], for instance, the potential energy is replaced by a very simple force field, a Hookean potential between each pair of atoms within a certain cutoff-distance. The related Elastic Network Model (ENM) only considers pairs of the C_α backbone carbons of the protein [7, 8]. It has the advantage that no energy minimization is required.

Another strategy is the partitioning of the system into blocks. In a first step the geometry of the complete system is optimized with an inexpensive method, and in a second step, the system is partially optimized at a high level of theory, while keeping the internal geometry of the blocks fixed. The computational cost of the energy minimization is thus reduced, however, the system resides in a nonequilibrium state and application of the standard NMA procedure leads to wrong frequencies, which may be spurious and even imaginary. So far, two methods have been developed for the calculation of frequencies in partially optimized structures. The Partial Hessian Vibrational Analysis (PHVA) [9–12] excludes all non-optimized atoms from the NMA by giving them an infinite mass. The Mobile Block Hessian (MBH) model [13], subject of this PhD thesis, is an improved version which takes into account the finite mass of the non-optimized atoms. The key concept is the partitioning of the system into several blocks of atoms, which move as rigid bodies during the vibrational analysis with only rotational and translational degrees of freedom. The Rotation-Translation Blocks (RTB) method developed by Durand et al. and later Tama et al. [14, 15] are closely related to MBH but can not yield physical frequencies in nonequilibrium points. The MBH has several variants according to the block choice and the way blocks are adjoined together [16].

To answer the second problem, i.e. the amount of data, the dimensionality of the NMA should be reduced, if possible decreasing the computational load simultaneously. For proteins one often chooses a model with one point mass per residue, typically located at the C_α position, which interact through a short-range parametrized harmonic force field. The Elastic Network Model belongs to this category of models. The PHVA and MBH do not alter the description of the PES but nevertheless reduce the number of normal modes, because the relative motions of atoms within a block are not considered. Another recently developed method is the Vibrational Subsystem Analysis (VSA) [17], which partitions the system into a subsystem and an environment. Only modes initiated by subsystem motions are calculated, while the environment is assumed to follow the subsystem motions in an adiabatic way.

Mainly two categories of approximate normal mode calculations can thus be identified: (1) the PES description is simplified, as e.g. in QM/MM or ENM; (2) the description of the PES is not modified, but only a subset of the modes are calculated in an approximate way, as is the case in PHVA, MBH, and VSA. This PhD work focuses on the latter category and presents the new MBH method and its variants. MBH realizes a reduced computational cost in the three steps of the NMA procedure:

1. Geometry optimization \longrightarrow Partial geometry optimization, where not all atom positions are optimized. The reference geometry is not the bottom of the well of the PES.
2. Calculate the full Hessian (second derivatives of the PES) \longrightarrow Calculate less second derivatives and/or store less second derivatives.
3. Solve the generalized eigenvalue problem \longrightarrow Diagonalize a smaller, adapted Hessian, from which less frequencies and modes evolve.

The underlying theory and the implementation are discussed in Chapter 2. Apart from the computational aspects, the main concern is to calculate those modes that are relevant for the desired application. The MBH is validated for its performance in reproducing modes and frequencies in Chapter 3. In Chapter 4 some general conclusions are drawn, leading to perspectives for future extensions of the MBH. The remainder of the present chapter revises the normal mode equations as they are commonly used in molecular modeling and points out the problems arising in case of nonequilibrium points. The latter is of great relevance for the partially optimized structure for which the MBH was designed.

1.2 Normal mode analysis equations

Solution schemes of the NMA equations are implemented in various standard ab initio molecular modeling packages. They are usually expressed in Cartesian coordinates. These rectilinear coordinates have the advantage that the metric of the space is diagonal, which facilitates the solution of the generalized eigenvalue problem. Another frequently used coordinate system is the combination of six external coordinates describing the position and orientation of the molecular system and $3N - 6$ internal coordinates describing the internal geometry of the system.

1.2.1 NMA in Cartesian coordinates

The normal mode theory is based on the solution of the Newton equations of motion with the hypothesis that the potential energy has a quadratic form. Consider a molecule with N masses m_A , $A = 1, \dots, N$. The positions are described by Cartesian coordinates $\mathbf{r}_A \equiv \{r_{A\mu}\}_{\mu=x,y,z}$ with respect to a space-fixed frame. The energy of the system reads

$$E = \frac{1}{2} \sum_{A\mu} m_A \dot{r}_{A\mu}^2 + V(\{\mathbf{r}_A\}), \quad (1.7)$$

where $\dot{r}_{A\mu}$ is a time derivative and V is the potential energy. Expanding V around a reference configuration $\{r_{A\mu}^0\}$ gives

$$V(\{\mathbf{r}_A\}) = V_0 + \sum_{A\mu} \left(\frac{\partial V}{\partial r_{A\mu}} \right)_0 \Delta_{A\mu} + \frac{1}{2} \sum_{A\mu, B\nu} \left(\frac{\partial^2 V}{\partial r_{A\mu} \partial r_{B\nu}} \right)_0 \Delta_{A\mu} \Delta_{B\nu} + \dots \quad (1.8)$$

in terms of the displacement coordinates $\Delta_{A\mu} = r_{A\mu} - r_{A\mu}^0$, where $(\cdot)_0$ denotes that the quantity is evaluated in the reference point. The constant term V_0 is the potential energy of the reference configuration. Only differences in potential energy have a physical meaning, such that V_0 can be put to zero without loss of generality. The Taylor expansion of the kinetic energy is exact in the Cartesian displacement coordinates,

$$T = \frac{1}{2} \sum_{A\mu} m_A \dot{r}_{A\mu}^2 = \frac{1}{2} \sum_{A\mu} m_A \dot{\Delta}_{A\mu}^2. \quad (1.9)$$

Collecting the displacements in one vector $x_i = r_{A\mu}$ with $i \equiv A\mu = 1, \dots, 3N$, one can expand the energy in matrix form as

$$E = \frac{1}{2} \dot{\Delta}^T M \dot{\Delta} + \Delta^T G + \frac{1}{2} \Delta^T H \Delta \quad (1.10)$$

up to second order in the displacements $\Delta_i = x_i - x_i^0$. In Eq. (1.10), the mass matrix M is a diagonal matrix containing the masses, G is the gradient vector defined by the gradients of the potential energy at the reference point, and H is the Hessian or second-derivative matrix in the reference point,

$$M_{A\mu, B\nu} = \delta_{A,B} \delta_{\mu,\nu} m_A \quad (1.11)$$

$$G_{A\mu} = \left(\frac{\partial V}{\partial r_{A\mu}} \right)_0 \quad (1.12)$$

$$H_{A\mu, B\nu} = \left(\frac{\partial^2 V}{\partial r_{A\mu} \partial r_{B\nu}} \right)_0. \quad (1.13)$$

Both M and H are symmetric matrices.

When the reference configuration is an equilibrium point, all gradients are zero: $G = 0$. The Lagrangian is given by

$$L = \frac{1}{2} \dot{\Delta}^T M \dot{\Delta} - \frac{1}{2} \Delta^T H \Delta \quad (1.14)$$

and the associated Euler-Lagrange equations are

$$m_A \ddot{\Delta}_{A\mu} + \sum_{A'\mu'} H_{A\mu, A'\mu'} \Delta_{A'\mu'} = 0, \quad (1.15)$$

or in matrix notation, $M\ddot{\Delta} + H\Delta = 0$. Apparently, the motion of a set of particles in the neighbourhood of an equilibrium point is described by a set of coupled linear differential equations of second order. The solutions of this set equations have the following time dependence of a harmonic oscillator:

$$\Delta_{A\mu}(t) = A v_{A\mu} \cos(\omega t + \phi) \quad (1.16)$$

where the scaling factor A and the phase ϕ have yet to be defined by the initial conditions, i.e. the initial positions $\Delta_{A\mu}^0$ and velocities $\dot{\Delta}_{A\mu}^0$. Substituting this form for the solution into the differential equations, we obtain

$$Hv = \lambda Mv. \quad (1.17)$$

with $\lambda = \omega^2$. This is a set of $3N$ homogeneous linear equations for the $v_{A\mu}$. It has nontrivial solutions if the determinant of the coefficients is zero:

$$\det(H - \lambda M) = 0. \quad (1.18)$$

This secular equation represents an algebraic equation of the $3N$ -th degree in λ . The eigenvalues λ determine the frequencies $\nu = \omega/2\pi$ of the motion in Eq. (1.16). For each eigenvalue λ_k , $k = 1, \dots, 3N$, the set of equations (1.17) has a solution vector $v_k = (v_{1,k}, \dots, v_{3N,k})$, which is the normal mode corresponding to the frequency ω_k : $Hv_k = \lambda_k Mv_k$. The norm of each v_k vector, that is $\sqrt{\sum_{A\mu} m_A |v_{A\mu,k}|^2}$, is arbitrary, but the vector is usually normalized.

The symmetry of the H and M matrices leads to two interesting properties. First, the eigenvalues are real. In a minimum energy state, the eigenvalues are moreover positive (≥ 0), because both M is positive definite and H is positive semidefinite. The interpretation is that any deviation from the minimum energy state leads to an increase in potential energy (or does not change, if invariance). In a transition state, one or more of the eigenvalues λ

can be negative (imaginary frequency). Second, the normal modes are mass-orthogonal, which can be written as

$$v_k^T M v_l = \delta_{k,l}, \quad (1.19)$$

assuming that the modes are normalized. The mass matrix M is the metric of the Cartesian space of $\{r_{A\mu}\}$ coordinates and defines the scalar product. It is important for the calculation of the overlap between vectors. The mass-orthonormality property illustrates that the solutions v_k represent *uncoupled* harmonic oscillators.

The generalized eigenvalue (1.17) problem is easily reduced to a standard eigenvalue problem. Introducing the mass-weighted Hessian

$$\tilde{H}_{A\mu,A'\mu'} = \frac{H_{A\mu,A'\mu'}}{\sqrt{m_A m_{A'}}} \quad (1.20)$$

or $\tilde{H} = M^{-1/2} H M^{-1/2}$, and the mass-weighted vector $\tilde{v} = M^{1/2} v$, the NMA equation becomes

$$\tilde{H} \tilde{v} = \lambda \tilde{v}. \quad (1.21)$$

which can be solved with standard eigenvalue solver routines. The non-mass-weighted normal modes are then obtained with $v = M^{-1/2} \tilde{v}$.

Once the NMA equations (1.17) are solved, the general motion of the system can be expressed as a linear combination of the motions of the $3N$ uncoupled harmonic oscillators $\{\omega_k\}$,

$$\Delta_{A\mu}(t) = \sum_k^{3N} A_k v_{A\mu,k} \cos(\omega_k t + \phi_k), \quad (1.22)$$

where A_k and ϕ_k are the amplitude and phase of the contribution by the ω_k -mode, determined by the initial positions and velocities.

1.2.2 NMA in internal and external coordinates

With any choice of $3N - 6$ internal coordinates $\{\theta_I\}$, and a body-frame whose origin lies at the center-of-mass $\mathbf{r}_{c.m.}$, the total energy can be written in a standard way [18–20] as the sum of the potential energy $W(\{\theta_I\})$ expressed in internal coordinates and kinetic energy terms arising from the center-of-mass motion, global rotation, Coriolis coupling and internal motions:

$$E = \frac{1}{2} M \dot{\mathbf{r}}_{c.m.}^2 + \frac{1}{2} \boldsymbol{\omega} \cdot \bar{\mathbf{I}} \cdot \boldsymbol{\omega} + \boldsymbol{\omega} \cdot \sum_I \mathbf{A}_I \dot{\theta}_I + \frac{1}{2} \sum_{IJ} B_{IJ} \dot{\theta}_I \dot{\theta}_J + W(\{\theta_i\}). \quad (1.23)$$

Here \mathcal{M} is the total mass and ω the angular velocity vector of the body-frame. The inertial tensor $\bar{\mathbf{I}}$, the Coriolis coupling \mathbf{A}_I between the body-frame rotation and internal velocity $\dot{\theta}_I$, and the B_{IJ} matrix, are all functions of the internal coordinates.

The potential energy $W(\{\theta_I\})$, expanded to second order about a reference geometry $\{\theta_I^0\}$, reads as

$$W(\{\theta_I\}) = V_0 + \sum_I \left(\frac{\partial W}{\partial \theta_I} \right)_0 \Delta_I + \frac{1}{2} \sum_{IJ} \left(\frac{\partial^2 W}{\partial \theta_I \partial \theta_J} \right)_0 \Delta_I \Delta_J. \quad (1.24)$$

in terms of the $3N - 6$ displacement coordinates $\Delta_I = \theta_I - \theta_I^0$. Again, the constant term V_0 can be put to zero. The set of coordinates now consists of the center-of-mass coordinates $\mathbf{r}_{c.m.}$, 3 angles to specify the orientation of the body frame (e.g. Euler angles), and the $3N - 6$ internal coordinates. The NMA equations using internal coordinates show a block structure:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & H^{(ii)} \end{pmatrix} \begin{pmatrix} v^{(c)} \\ v^{(r)} \\ v^{(i)} \end{pmatrix} = \lambda \begin{pmatrix} M^{(cc)} & 0 & 0 \\ 0 & M^{(rr)} & M^{(ri)} \\ 0 & (M^{(ri)})^T & M^{(ii)} \end{pmatrix} \begin{pmatrix} v^{(c)} \\ v^{(r)} \\ v^{(i)} \end{pmatrix} \quad (1.25)$$

where $v^{(c)}$, $v^{(r)}$, and $v^{(i)}$ have dimensions 3, 3, and $3N - 6$, respectively, and the mass matrix entries read $M_{\mu\nu}^{(cc)} = \mathcal{M}\delta_{\mu\nu}$, $M_{\mu\nu}^{(rr)} = I_{\mu\nu}(\{\theta_I^0\})$, $M_{\mu I}^{(ri)} = A_{\mu I}(\{\theta_I^0\})$, $M_{IJ}^{(ii)} = B_{IJ}(\{\theta_I^0\})$, and the Hessian elements

$$H_{I,J}^{(ii)} = \left(\frac{\partial^2 W}{\partial \theta_I \partial \theta_J} \right)_0. \quad (1.26)$$

The resulting eigenvalue problem of dimension $3N$ still has 6 zero eigenvalues corresponding to overall translation and rotation. These can be decoupled in the usual way by a congruent transformation $\tilde{v}^{(c)} = v^{(c)}$, $\tilde{v}^{(i)} = v^{(i)}$, $\tilde{v}^{(r)} = v^{(r)} + (M^{(rr)})^{-1} M^{(ri)} v^{(i)}$, yielding the final NMA equation,

$$H^{(ii)} \tilde{v}^{(i)} = \lambda [M^{(ii)} - (M^{(ri)})^T (M^{(rr)})^{-1} M^{(ri)}] \tilde{v}^{(i)}. \quad (1.27)$$

Solving this $3N - 6$ dimensional eigenvalue problem gives the $3N - 6$ internal vibrations.

1.3 NMA and nonstationary points

Normal mode analysis is developed to calculate the vibrational modes of a system in an equilibrium point on the PES. However, it happens quite

often that one is confronted with a nonequilibrium state. A first reason is that the geometry optimization is stopped once a certain threshold value is reached for the residual forces. Depending on the chosen convergence criteria, the obtained geometry will be closer to the real minimum of the well. It is also common practice to perform a partial optimization, where part of the coordinates are optimized, while the rest is kept at their input value. A common procedure is the following. First a global plausible geometry is generated with a full optimization at a low, inexpensive level of theory description, and next the active, interesting region is further optimized at a detailed, expensive level of theory, without optimizing the rest of the system. The resulting partially optimized structure is a nonequilibrium state where the forces on the atoms of the active region are zero, but the atoms in the environment still feel residual forces.

In presence of a gradient $G \neq 0$, it is in principle still possible to solve the NMA equations. Consider for instance the one-dimensional Lagrangian in terms of a generalized coordinate q ,

$$L(q, \dot{q}) = m(q)\dot{q}^2 - V(q) \quad (1.28)$$

which has a gradient term in its second order expansion in $\Delta q = q - q_0$ with q_0 the reference point,

$$L(q, \dot{q}) = m(q_0)\Delta\dot{q}^2 - V_0 - \Delta q \left(\frac{\partial V}{\partial q} \right)_0 - \frac{1}{2}\Delta q^2 \left(\frac{\partial^2 V}{\partial q^2} \right)_0 \quad (1.29)$$

or, with $m = m(q_0)$, $g = (\partial V / \partial q)_0$ and $k = (\partial^2 V / \partial q^2)_0$, and V_0 chosen to be zero,

$$L(q, \dot{q}) = m\Delta\dot{q}^2 - g\Delta q - \frac{1}{2}k\Delta q^2. \quad (1.30)$$

The solution of the corresponding Euler-Lagrange equation

$$m\Delta\ddot{q} + k\Delta q + g = 0 \quad (1.31)$$

is the sum of the general solution of the homogeneous equation and a particular solution of the inhomogeneous equation:

$$q(t) = q_0 + A \cos\left(\sqrt{\frac{k}{m}}t + \phi\right) - \frac{g}{k} \quad (1.32)$$

where the amplitude A and phase ϕ are determined by the initial position and velocity. The derivatives g and k in the reference point can be used to predict the position of the minimum energy point within the harmonic surface

assumption. The minimum would be located at $q = q_0 - g/k$ if the PES were truly parabolic. The solution given in Eq. (1.32) represents an oscillation of the mass m around the hypothetical minimum at $q = q_0 - g/k$.

In case of a multi-dimensional surface, however, the standard normal mode analysis – the diagonalization of the mass-weighted Hessian – shows some serious defects when applied to a nonequilibrium structure. Three problems arise: 1) imaginary frequency might appear, 2) three instead of six zero eigenvalues are found, and 3) the frequencies are coordinate dependent.

1. Spurious imaginary frequencies.

For a nonequilibrium point, it is not guaranteed that the curvature in every direction on the PES is positive. The Hessian is not necessarily positive semidefinite anymore, which translates itself in the occurrence of negative eigenvalues, and hence imaginary frequencies, even when not being in a transition state. The interpretation of such imaginary frequencies is not straightforward. The mode no longer corresponds to a vibration around the reference point, but rather to a non-periodic motion with a $\cosh(|\omega|t)$ -like time dependence.

Note that a transition state of a certain order n_{TS} is no nonequilibrium point, but an unstable equilibrium point. The n_{TS} imaginary frequencies that show up are related to the transition rate over the barrier.

2. Zero frequencies for invariances.

The appearance of a mode with zero frequency is the manifestation of the invariance of the PES under the motion of that specific mode. For a molecular system those invariances are the six global translations and rotations: the potential energy of the system does not change when translating or rotating the whole system. However, in a nonequilibrium point, only the three translational zeros emerge and the three rotational zeros are often absent. The rotational motions mingle with the intrinsic vibrational modes. This behavior was shown mathematically in Ref. [13]. The three eigenvalues associated with the global rotation are different from zero [21] and the absence of these Goldstone modes [22] is due to the use of the second order expansion in rectilinear Cartesian coordinates, which breaks the rotational symmetry of V . This defect is simply cured by taking coordinates that respect the symmetries of V , i.e. internal coordinates.

3. Coordinate dependence.

The coordinate dependence in case of a nonzero gradient can be proved easily with the chain rule. Suppose a set of coordinates q and a new set of coordinates $q' = q'(q)$. With J the Jacobian of the transformation, given by

$$J_{\alpha\alpha'} = \left(\frac{\partial q_\alpha}{\partial q'_{\alpha'}} \right)_0, \quad (1.33)$$

the mass matrix transforms as

$$M'_{\alpha'\beta'} = \sum_{\alpha,\beta} M_{\alpha\beta} J_{\alpha\alpha'} J_{\beta\beta'} \quad (1.34)$$

but the Hessian transforms as

$$H'_{\alpha'\beta'} = \sum_{\alpha,\beta} H_{\alpha\beta} J_{\alpha\alpha'} J_{\beta\beta'} + \sum_{\alpha} G_{\alpha} \frac{\partial^2 q_{\alpha}}{\partial q'_{\alpha'} \partial q'_{\beta'}}. \quad (1.35)$$

When $G = 0$, the same eigenvalues evolve from the NMA equations $Hv = \lambda Mv$ or $H'v' = \lambda M'v'$ and the normal modes are simply related by $v = Jv'$. Both the mass matrix and Hessian have a tensor behavior under a coordinate transformation. In presence of gradients, the second term in Eq. (1.35) breaks the tensor behavior of the Hessian. The eigenvalues are coordinate dependent: a second-order expansion of the potential energy expressed in curvilinear coordinates, which are nonlinearly related to the Cartesian coordinates, leads to different normal mode frequencies. For instance, the NMA equations in Eq. (1.25) in internal coordinates will give different eigenvalues from those of Eq. (1.17) in Cartesian coordinates. In particular, the eigenvalue equations Eq. (1.25) always generate 6 zero eigenvalues due to its construction, as the Hessian has vanishing matrix elements in the (c) and (r) subspaces, while in Eq. (1.17) the presence of 6 zero eigenvalues is not ensured.

Physical properties should not depend on the chosen coordinate system, which makes the interpretation of frequencies of a nonequilibrium point nontrivial. This phenomenon is “well known” and is related to the difference between ordinary and covariant derivatives in a non-trivial metric space [23]. Its importance and unpleasant consequences have been emphasized by several authors [24, 25]. In a completely general situation it is thus in principle not possible, due to the coordinate dependence, to define meaningful normal modes in a nonequilibrium point.

The situations of interest are not completely general but arise from physical considerations. We consider reference points on the PES that were obtained

with partial optimization, by optimizing with respect to a *subset* of coordinates, keeping the remaining coordinates fixed during the optimization. The system corresponding to the subset of coordinates that have been optimized, is actually in equilibrium, even though the complete gradient is nonzero. A normal mode analysis that is limited to the optimized coordinates will no longer give spurious imaginary frequencies and be coordinate independent. In 2002 Li and Jensen further developed the Partial Hessian Vibrational Analysis (PHVA) [11], which was initially introduced by Head and co-workers [9, 10, 26–28]. It gives an infinite mass to atoms that are not optimized. In my PhD work I developed an improved version, the Mobile Block Hessian (MBH) method, which takes into account the finite mass of the non-optimized atoms. In the following chapter, the MBH equations will be presented in terms of optimized coordinates, such that physical frequencies are guaranteed.

2 MBH for NMA of partially optimized structures

2.1 Partial Hessian Vibrational Analysis (PHVA)

In a partially optimized system, the positions of part of the atoms are optimized, while the rest of the atoms is kept immobile during the energy minimization. One assumes a certain reference structure $\{\mathbf{r}_{A_F}^0\}$ for part of the molecule consisting of N_F atoms. Keeping $\{\mathbf{r}_{A_F}^0\}$ fixed in space, the positions of the remaining $N_E = N - N_F$ atoms are optimized, resulting in an “equilibrium” configuration $\{\mathbf{r}_{A_E}^0\}$. Obviously, the full gradient is nonzero, since $(\partial V / \partial r_{A_E \mu})_0 = 0$, but $(\partial V / \partial r_{A_F \mu})_0 \neq 0$. Solutions of the NMA equations taking into account the full Hessian and mass matrix will therefore lead to unphysical results.

So far, two methods are able to surmount the defects in case of a partially optimized system. The Partial Hessian Vibrational Analysis and the new Mobile Block Hessian method are constructed in terms of optimized coordinates such that physical frequencies result.

In the first model, the Partial Hessian Vibrational Analysis, one assumes that the fixed atoms A_F do not participate in the small amplitude vibrations. Hence their displacements $\Delta_{A_F \mu}$ and velocities $\dot{\Delta}_{A_F \mu}$ are set to zero. This is consistent with a situation in which infinite masses m_{A_F} are associated to the fixed atoms.

With this assumption, the second-order expansion of the energy reduces to

$$\begin{aligned}
 E &= \frac{1}{2} \sum_{A_E\mu} m_{A_E} \dot{\Delta}_{A_E\mu}^2 + V_0 + \frac{1}{2} \sum_{A_E\mu; B_E\nu} \left(\frac{\partial^2 V}{\partial r_{A_E\mu} \partial r_{B_E\nu}} \right)_0 \Delta_{A_E\mu} \Delta_{B_E\nu} + \dots \\
 &= \frac{1}{2} \dot{\Delta}_E^T M_E \dot{\Delta}_E + V_0 + \frac{1}{2} \Delta_E^T H_E \Delta_E,
 \end{aligned} \tag{2.1}$$

where $\Delta_E, \dot{\Delta}_E, M_E, H_E$ are the displacements, velocities, mass matrix and Hessian restricted to the reduced $3N_E$ -dimensional subsystem of the non-fixed atoms with mass m_{A_E} . The constant term V_0 can be put to zero. The reduced gradient ($G_E = 0$) vanishes in this subspace, so the $3N_E$ system is in equilibrium (hence subindex E from “equilibrium”) at the reference configuration and provides coordinate-independent normal modes.

In practice, one simply needs to disregard the rows and columns related to the coordinates of the fixed atoms in the full (Cartesian) Hessian and mass matrix. The NMA equation reduces to a $3N_E \times 3N_E$ generalized eigenvalue problem:

$$H_E v_E = \lambda M_E v_E \tag{2.2}$$

The reduced mass matrix M_E is positive definite, and the presence of negative solutions λ in Eq. (2.2) gives indication that the reduced system resides in a transition state, rather than a minimum. The number of zero eigenvalues depends on the remaining symmetry of the reduced system. Usually no zeros will occur, since the fixed atoms act as an external field breaking translational and rotational invariance for the non-fixed atoms.

2.2 MBH equations

In the Mobile Block Hessian (MBH) model, the fixed atoms are allowed to participate in the small-amplitude vibrations by moving as a rigid body (a mobile block). This is a physically different situation from the previous approach, since one now takes into account the finite masses of the fixed atoms $\{m_{A_F}\}$. The MBH reduces the degrees of freedom to those that are optimized, so the problems related to nonequilibrium structures are avoided. There are no spurious imaginary frequencies, six zero frequencies are found for the translational/rotational invariances of the system, and the frequencies and modes are coordinate independent.

The NMA equations can be best constructed when using internal coordinates or using Cartesian coordinates.

2.2.1 In internal coordinates

The MBH approach is easily implemented by a suitable choice of the $3N - 6$ internal coordinates $\{\theta_I\}$, which is always possible. In the Z-matrix formalism, for instance, one can determine the first N_F atoms by $3N_F - 6$ Z-coordinates (distances, angles and dihedral angles), and the next N_E atoms can consecutively be described by $3N_E$ Z-coordinates. The $3N_F - 6$ internal coordinates describing the geometry of the fixed atoms A_F , are labeled $\{\theta_{I_F}\}$. The remaining $3N_E$ internal coordinates are labeled $\{\theta_{I_E}\}$.

The $\{\theta_{I_E}\}$ coordinates of the non-fixed atoms are optimized (hence E of “equilibrium”) while the $\{\theta_{I_F}\}$ coordinates need not be optimized. The full gradient is nonzero, since $(\partial W/\partial\theta_{I_E})_0 = 0$ but $(\partial W/\partial\theta_{I_F})_0 \neq 0$. As a result, the NMA equation (1.25) with the full Hessian will not give correct frequencies. The MBH approach is expressed in optimized coordinates only, so the MBH frequencies are coordinate independent.

Allowing the fixed block to move as a rigid body, keeping its internal geometry, can be imposed by putting the displacements Δ_{I_F} and corresponding velocities $\dot{\Delta}_{I_F}$ equal to zero in the second-order expansion of the energy. The quadratic expansion of the MBH potential energy $W(\{\theta_I\})$ becomes

$$W(\{\theta_I\}) \approx V_0 + \frac{1}{2} \sum_{I_E J_E} \left(\frac{\partial^2 W}{\partial\theta_{I_E} \partial\theta_{J_E}} \right)_0 \Delta_{I_E} \Delta_{J_E} = V_0 + \frac{1}{2} \sum_{I_E J_E} H_{I_E J_E}^{(ii)} \Delta_{I_E} \Delta_{J_E}. \quad (2.3)$$

The corresponding NMA equations are obtained by omitting the rows and columns related to the $\{\theta_{I_F}\}$ variables from the full Hessian and full mass matrix in Eq. (1.25). The resulting reduced eigenvalue problem of dimension $3N_E + 6$ still has 6 zero eigenvalues corresponding to overall translation and rotation. These can be decoupled in the usual way by a congruent transformation $\tilde{v}^{(c)} = v^{(c)}$, $\tilde{v}^{(i)} = v^{(i)}$, $\tilde{v}^{(r)} = v^{(r)} + (M^{(rr)})^{-1} M'^{(ri)} v^{(i)}$, yielding the final $3N_E$ -dimensional NMA equation,

$$H'^{(ii)} \tilde{v}^{(i)} = \lambda [M'^{(ii)} - (M'^{(ri)})^T (M^{(rr)})^{-1} M'^{(ri)}] \tilde{v}^{(i)}. \quad (2.4)$$

The primed matrices do not contain the components related to $\{\theta_{I_F}\}$. The transformed mass matrix in the right hand side of Eq. (2.4) is positive definite and the presence of negative eigenvalues λ will now unambiguously indicate that the reduced system is in a transition state.

If one has access to selected Hessian matrix elements in the chosen internal coordinates, this is obviously the most straightforward way of implementing MBH.

2.2.2 In Cartesian coordinates

The Hessian is commonly given in Cartesian coordinates as the output of an simulation program (e.g. CHARMM [29–31]), and it is of interest to express the MBH normal mode equations in terms of Cartesian quantities, i.e. the Cartesian gradient and Cartesian Hessian.

Some group-theoretical concepts

The analysis of MBH in Cartesian coordinates is greatly facilitated by introducing some (elementary) group-theoretical concepts. We first consider the relevant transformation group in 3-dimensional space that leaves the potential energy invariant. This six-parameter group consists of combinations of translations and rotations, taking coordinates \mathbf{r} to their translated/rotated positions \mathbf{r}' ,

$$r_\mu \rightarrow r'_\mu = g_\mu(\mathbf{r}, p). \quad (2.5)$$

Here p stands for the six parameters p_α , $\alpha = 1 \dots 6$. The composition of two transformations can be written as

$$\mathbf{g}(\mathbf{g}(\mathbf{r}, p), p') = \mathbf{g}(\mathbf{r}, \Phi(p, p')), \quad (2.6)$$

where $\Phi(p, p')$ are the parameters of the composition (i.e. the group multiplication law). The invariance properties of the potential energy surface are now simply expressed as

$$V(\{\mathbf{r}_A\}) = V(\{\mathbf{g}(\mathbf{r}_A, p)\}). \quad (2.7)$$

The precise parametrization of the group is arbitrary, but it is convenient to take $p_\alpha = 0$ as the identity (no translation or rotation). As an example one can take

$$\mathbf{g}(\mathbf{r}, p) = \sum_\mu p_\mu \mathbf{e}_\mu + \hat{R}_x(p_4) \hat{R}_y(p_5) \hat{R}_z(p_6) \mathbf{r} \quad (2.8)$$

where \mathbf{e}_μ , for $\mu = x, y, z$, are unit vectors along the x, y, z directions, and $\hat{R}_\mu(\phi)$ is a rotation around the μ -axis over an angle ϕ , e.g. $\mathbf{r}' = \hat{R}_z(\phi) \mathbf{r}$ has components

$$x' = x \cos \phi - y \sin \phi; \quad y' = x \sin \phi + y \cos \phi; \quad z' = z. \quad (2.9)$$

Some interesting quantities are the derivatives of the group transformation g with respect to the group parameters. The vectors $D^{(\alpha)}$ and $C^{(\alpha\alpha')}$ have

α	$D_{A\mu}^{(\alpha)}$	α	α'	$C_{A\mu}^{(\alpha\alpha')}$
1	$\delta_{\mu x}$	1	1-6	0
2	$\delta_{\mu y}$	2	1-6	0
3	$\delta_{\mu z}$	3	1-6	0
4	$\sum_{\lambda} \epsilon_{\lambda\mu x} r_{A\lambda}^0$	4	4	$\delta_{\mu x} r_{Ax}^0 - r_{A\mu}^0$
5	$\sum_{\lambda} \epsilon_{\lambda\mu y} r_{A\lambda}^0$	5	5	$\delta_{\mu y} r_{Ax}^0$
6	$\sum_{\lambda} \epsilon_{\lambda\mu z} r_{A\lambda}^0$	6	6	$\delta_{\mu z} r_{Ax}^0$
		5	5	$\delta_{\mu y} r_{Ay}^0 - r_{A\mu}^0$
		6	6	$\delta_{\mu z} r_{Ay}^0$
		6	6	$\delta_{\mu z} r_{Az}^0 - r_{A\mu}^0$

Table 2.1: Derivatives of the transformation g , for the parametrization defined in Eq. (2.8). The Levi-Civita symbol $\epsilon_{\lambda\mu\nu}$ equals 1 (-1) if $\lambda\mu\nu$ is a cyclic (anti-cyclic) permutation of xyz and zero otherwise.

components

$$D_{A\mu}^{(\alpha)} = \frac{\partial g_{\mu}}{\partial p_{\alpha}}(\mathbf{r}_A^0, 0) \quad (2.10)$$

$$C_{A\mu}^{(\alpha\alpha')} = \frac{\partial^2 g_{\mu}}{\partial p_{\alpha} \partial p_{\alpha'}}(\mathbf{r}_A^0, 0) \quad (2.11)$$

and are listed in Table 2.1 for the parametrization defined in Eq. (2.8).

Group parameters as dynamical variables

The movement of the block as a rigid body can be described in a natural way by treating the six group parameters in Eq. (2.5) as dynamical variables, such that the instantaneous position of each atom F in the block is given by the result of a common translation/rotation of the reference position $\{\mathbf{r}_F^0\}$ in the block,

$$\mathbf{r}_F(t) = \mathbf{g}(\mathbf{r}_F^0, p(t)). \quad (2.12)$$

The velocities of the N_F atoms in the block are

$$\dot{r}_{F\mu} = \sum_{\alpha} \frac{\partial g_{\mu}}{\partial p_{\alpha}}(\mathbf{r}_F^0, p) \dot{p}_{\alpha}. \quad (2.13)$$

In order to express the MBH normal mode equations in terms of the Cartesian Hessian we choose as variables the 6 parameters p_α , combined with the $3N_E$ Cartesian coordinates $\{\mathbf{r}_E\}$ of the remaining atoms. The potential energy expressed in the new coordinates becomes

$$\tilde{V}(p, \{\mathbf{r}_E\}) = V(\{\mathbf{g}(\mathbf{r}_F^0, p)\}, \{\mathbf{r}_E\}). \quad (2.14)$$

Obviously, the partially optimized MBH reference structure $\{\mathbf{r}_A^0\}$ corresponds to $p = 0$ and $\{\mathbf{r}_E^0\}$ in the present coordinates. At this reference point the first derivatives of \tilde{V} indeed vanish,

$$\tilde{G}_\alpha = \left(\frac{\partial \tilde{V}}{\partial p_\alpha} \right)_0 = \sum_{F\mu} G_{F\mu} D_{F\mu}^{(\alpha)} = 0 \quad (2.15)$$

$$\tilde{G}_{E\mu} = \left(\frac{\partial \tilde{V}}{\partial r_{E\mu}} \right)_0 = G_{E\mu} = 0. \quad (2.16)$$

The normal mode equations in the group coordinates are of dimension $3N_E + 6$ and read in standard form,

$$\tilde{H}v = \lambda \tilde{M}v, \quad (2.17)$$

where we used the obvious notation

$$\tilde{H}_{\alpha,\alpha'} = \left(\frac{\partial^2 W}{\partial p_\alpha \partial p_{\alpha'}} \right)_0, \quad (2.18)$$

$$\tilde{H}_{\alpha,E\mu} = \left(\frac{\partial^2 W}{\partial p_\alpha \partial r_{E\mu}} \right)_0, \quad (2.19)$$

$$\tilde{H}_{E\mu,E'\mu'} = \left(\frac{\partial^2 W}{\partial r_{E\mu} \partial r_{E'\mu'}} \right)_0. \quad (2.20)$$

The \tilde{H} elements can be expressed in terms of Cartesian quantities. A direct evaluation starting from Eq. (2.14) yields

$$\tilde{H}_{\alpha,\alpha'} = \sum_{F\mu,F'\mu'} H_{F\mu,F'\mu'} D_{F\mu}^{(\alpha)} D_{F'\mu'}^{(\alpha')} + \sum_{F\mu} G_{F\mu} C_{F\mu}^{(\alpha\alpha')} \quad (2.21)$$

$$\tilde{H}_{\alpha,E\mu} = \sum_{F'\mu'} H_{E\mu,F'\mu'} D_{F'\mu'}^{(\alpha)} \quad (2.22)$$

$$\tilde{H}_{E\mu,E'\mu'} = H_{E\mu,E'\mu'}. \quad (2.23)$$

The corresponding mass matrix \tilde{M} is to be derived from the kinetic energy. Expanding around the total reference structure $\{\mathbf{r}_A^0\}$, the kinetic energy [see

Eq. (2.13)] is:

$$T = \frac{1}{2} \sum_{F\mu} m_F \left(\sum_{\alpha} D_{F\mu}^{(\alpha)} \dot{p}_{\alpha} \right)^2 + \frac{1}{2} \sum_E m_E \dot{\mathbf{r}}_E^2. \quad (2.24)$$

This determines the mass matrix components,

$$\tilde{M}_{\alpha,\alpha'} = \sum_{F\mu} m_F D_{F\mu}^{(\alpha)} D_{F\mu}^{(\alpha')}, \quad (2.25)$$

$$\tilde{M}_{\alpha,E\mu} = 0, \quad (2.26)$$

$$\tilde{M}_{E\mu,E'\mu'} = \delta_{E,E'} \delta_{\mu,\mu'} m_E. \quad (2.27)$$

Note that the mass matrix is block-diagonal in the block structure induced by the present choice of the 6 coordinates p_{α} and $3N_E$ coordinates $r_{E\mu}$.

2.3 Extensions to MBH

2.3.1 MBH with multiple blocks

The MBH has been extended to the case of multiple mobile blocks, where the position and orientation of each block are optimized with respect to each other [13]. An example of such a partitioning of a molecular system is shown in Figure 2.1. Six block parameters, three translational and three rotational, are assigned to each block. The MBH is closely related to the Rotation-Translation Blocks (RTB) [15] and Block Normal Modes (BNM) [32, 33] method but incorporates gradient corrections to account for the forces on the fixed atoms, whereas RTB/BNM can only treat fully optimized systems.

2.3.2 MBH with linear blocks

In case of a linear block, only five (three translational and two rotational) parameters are required. The MBH equations have been adapted to accommodate linear blocks [16].

Linear blocks in the partitioning can be used to fix bond lengths. It is the NMA analogon to the frequently used SHAKE algorithm in molecular dynamics [34], where typically C-H bonds are fixed to allow for a longer time step. In NMA it is interesting to eliminate the high frequency stretches when one is only interested in the lower spectrum, thereby reducing the dimensionality of the NMA.

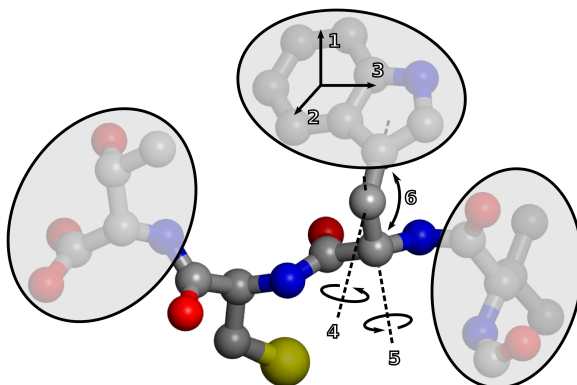


Figure 2.1: Schematic representation of the basic idea behind the MBH method. The shaded blocks symbolize the parts of the molecule of which the internal geometry is kept fixed during the partial geometry optimization. In the MBH approach they are described as rigid bodies with six degrees of freedom (translations and rotations).

2.3.3 General formulation of MBH

An all-block formulation has been developed to write down the MBH equations in a consistent way [16]. It considers three types of blocks: nonlinear blocks, linear blocks, and single atoms. The block types have six, five, and three degrees of freedom respectively. The total number of resulting frequencies is then $d = \sum_b d_b$ with d_b the number of degrees of freedom of block b . Six frequencies are zero corresponding with global translations/rotations, and $d - 6$ frequencies correspond to vibrational modes.

Group parameters as dynamical variables

The MBH equations (2.4) describe motions of a N_E single atoms and one nonlinear block. Here we introduce the formalism for a system containing an arbitrary number of blocks. We assume that *all* the atoms in the molecule have been partitioned into blocks, with block dimension equal to 6, 5 or 3. We start again from some general group-theory concepts. The rigid-body motion of a nonlinear block b can be conveniently described by six parameters $p_{b\alpha}$ ($\alpha = 1, \dots, 6$) of a translation/rotation group. The instantaneous positions $\mathbf{r}_A(t)$ of the atoms A in the block are simply generated by applying a common

transformation with parameters p_b to the reference positions r_A^0 ,

$$\mathbf{r}_A(t) = \mathbf{g}(\mathbf{r}_A^0, p_b(t)), \quad \forall A \in b. \quad (2.28)$$

The parametrization of the transformation group is arbitrary, but we will stick to the convention introduced in (2.8), of successively rotating along the fixed z -, y - and x -axes of a space-fixed frame, followed by a translation,

$$\mathbf{g}(\mathbf{r}, p) = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos p_4 & -\sin p_4 \\ 0 & \sin p_4 & \cos p_4 \end{pmatrix} \begin{pmatrix} \cos p_5 & 0 & \sin p_5 \\ 0 & 1 & 0 \\ -\sin p_5 & 0 & \cos p_5 \end{pmatrix} \begin{pmatrix} \cos p_6 & -\sin p_6 & 0 \\ \sin p_6 & \cos p_6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (2.29)$$

The group variables p_b of the various blocks are used as the dynamical variables.

Three types of blocks are distinguished, depending on the number of its degrees of freedom d_b . The transformation $\mathbf{g}(\mathbf{r}_A^0, p_b)$ in Eq. (2.28) then depends on the number d_b .

1. *For a normal block* (containing at least three noncollinear atoms), $d_b = 6$ and all parameters p_1, \dots, p_6 are needed.
2. *For a single-atom block*, $d_b = 3$ and only the translational parameters p_1, p_2, p_3 are retained in the transformation (2.28).
3. *For a linear block* (containing two or more collinear atoms), $d_b = 5$. The translational parameters p_1, \dots, p_3 and two of the rotational parameters are now retained in the block transformation (2.28). For such a linear block, we must exclude rotations about a particular axis (e.g. the μ -axis with $\mu = x, y, \text{ or } z$) of the space-fixed frame. A suitable choice of the μ -axis is e.g. the axis with largest projection $|r_{A\mu}^0 - r_{B\mu}^0|$ for two atoms in the linear block.

For a particular block, the allowed transformations on the coordinates of its atoms thus depends on a number d_b of block variables, determined by the type of block. Note that the reduced set of p -variables in case of a linear block or free atom can be complemented with zeros for the missing variables in the transformation $\mathbf{g}(\mathbf{r}_A^0, p_b)$ in Eq. (2.28). To avoid confusion, we will denote by \hat{p}_b this complemented set of six parameters.

Derivation of the normal mode equations

The potential energy surface (PES) expressed in the set of $\{p_{b\alpha}\}$ variables becomes

$$\tilde{V}(\{p_b\}) = V(\{\mathbf{g}(\mathbf{r}_A^0, \hat{p}_{b(A)})\}), \quad (2.30)$$

where $V(\{\mathbf{r}_A\})$ is the PES in Cartesian coordinates, and $b(A)$ is the block to which atom A belongs. The equilibrium position of the blocks is taken as the reference point, i.e. the gradients in the $p_{b\alpha}$ variables

$$\tilde{G}_{b\alpha} = \left(\frac{\partial \tilde{V}}{\partial p_{b\alpha}} \right)_0 = \sum_{A \in b, \mu} G_{A\mu} D_{A\mu}^{(\alpha)} = 0, \quad (2.31)$$

all vanish. The NMA equations are given by

$$\tilde{H}v = \lambda \tilde{M}v \quad (2.32)$$

A direct evaluation starting from Eq. (2.30) yields the Hessian in the $p_{b\alpha}$ variables,

$$\begin{aligned} \tilde{H}_{b\alpha, b'\alpha'} &= \left(\frac{\partial^2 \tilde{V}}{\partial p_{b\alpha} \partial p_{b'\alpha'}} \right)_0 \\ &= \sum_{A \in b, \mu} \sum_{A' \in b', \mu'} H_{A\mu, A'\mu'} D_{A\mu}^{(\alpha)} D_{A'\mu'}^{(\alpha')} + \delta_{b, b'} \sum_{A \in b, \mu} G_{A\mu} C_{A\mu}^{(\alpha\alpha')} \end{aligned} \quad (2.33)$$

and the mass matrix elements are given by

$$\tilde{M}_{b\alpha, b'\alpha'} = \delta_{b, b'} \sum_{A \in b, \mu} m_A D_{A\mu}^{(\alpha)} D_{A\mu}^{(\alpha')}. \quad (2.34)$$

The form is identical to the Hessian and mass matrix derived in (2.17), except that now the indices α pertaining to a certain block b need not correspond to all six parameters of the full translation/rotation group, but can be restricted to five or three in order to accomodate linear blocks or single atoms.

2.4 MBH with adjoined blocks

The MBH method has also been extended to the case in which the blocks may be linked to each other, by one or two common atoms (adjoining atoms) [16], as schematically shown in Figure 2.2. We distinguish between three different partitioning schemes. In the original implementation of the MBH method,

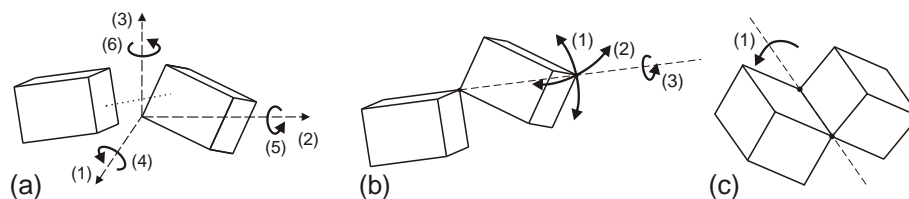


Figure 2.2: Degrees of freedom in the relative motion of (nonlinear) adjoined blocks. (a) Non-adjoined blocks have six degrees of freedom. The dotted line visualizes the bonding/nonbonding interaction between the blocks. (b) Blocks adjoined by one adjoining atom have three degrees of freedom. (c) Blocks adjoined by two adjoining atoms (hinge type connection) have one degree of freedom.

atoms could only be part of one MBH block (Fig. 2.2a). In the current MBH approach atoms can be part of two or more rigid blocks. Such a common atom, i.e. an adjoining atom, plays the role of a joint between the blocks. Two basic linkages can be realized with adjoining atoms: blocks connected by one adjoining atom (Fig. 2.2b) and blocks connected by two adjoining atoms (Fig. 2.2c). Unlinked blocks have six degrees of freedom in their relative motion, but blocks connected by an adjoining atom, have only three (the three translational degrees of freedom are lost). The remaining three degrees of freedom correspond to rotations about an axis through the adjoining atom. If blocks are connected by two adjoining atoms to a previous block, the linkage is a hinge type connection. Only one degree of freedom remains, corresponding to a rotation about the axis through the two adjoining atoms.

This extension is important for instance in very large (bio)systems, where one would like to have the possibility to calculate specific normal modes at decreased computational cost, since the number of degrees of freedom is further reduced by the introduction of adjoining atoms. The linkage between blocks is a fundamental difference with the RTB method. Moreover the MBH method is valid for partially optimized systems, even in the present adjoined version, extending the possible applications.

The various ingredients entering a possible partition of a molecule into blocks are (1) nonlinear blocks, (2) linear blocks, and (3) single atoms. These ingredients can now be adjoined together in several ways, as specified in Figure 2.3. For each topology a molecular example is given. A topology might even contain a ring structure (Fig. 2.3f) or a branched chain of blocks (Fig. 2.3g).

The extension of the MBH method to the case of adjoined blocks has been implemented in an intuitive way using penalty functions and in a more mathematically involved way using analytical constraints. Both implementations

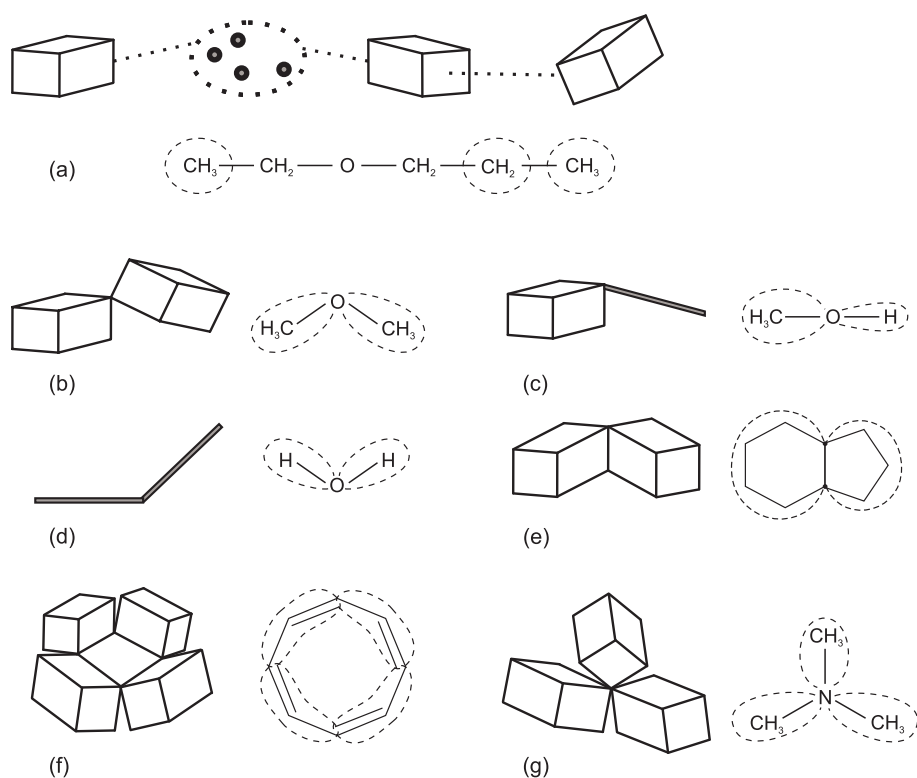


Figure 2.3: With the ingredients (nonlinear blocks, linear blocks, single atoms) various topologies can be constructed. In each case a molecular example is given. (a) An example of non-adjoined blocks. Nonlinear blocks have six, linear blocks have five degrees of freedom. (b)-(d) Some examples of adjoined blocks with one adjoining atom. (e) A hinge-type connection with blocks having two adjoining atoms gives rise to a hinge bending motion with one rotational degree of freedom. (f) A ring structure consisting of singly adjoined blocks. (g) Example of three blocks with one adjoining atom, shared by the three blocks.

will be shortly presented below.

2.4.1 Restraint technique with dummy atoms

A partitioning of the system into adjoined blocks can be realized using a restraint technique and dummy atoms. This has the advantage that all MBH formulas in Ref. [35] need not be changed. Only the Hessian must be complemented to impose the restraints. The procedure works for an arbitrary number of adjoining atoms connecting an arbitrary number of blocks, so both linkages with one adjoining atom or two adjoining atoms can be treated.

Suppose an adjoining atom A belongs to s_A (the share number) blocks. We can formally replace the adjoining atom by s_A dummy atom copies A_1, \dots, A_{s_A} at the same reference position. The mass of the dummy atoms is the original mass m_A divided by s_A . Each of the dummy atoms is assigned to one of the blocks. In this way a strict partition (no adjoining atoms anymore) of the atoms over the blocks is restored, and the former MBH computational routines can be used.

We must still impose the fact that the dummy copies describe in reality the same adjoining atom A , i.e. during the motion the coordinates $\mathbf{r}_{A_1} = \dots = \mathbf{r}_{A_{s_A}}$ should coincide. This can be done by adding to the potential energy harmonic terms of the form

$$V_A = \frac{1}{2} \kappa \sum_{i=2}^{s_A} |\mathbf{r}_{A_1} - \mathbf{r}_{A_i}|^2, \quad (2.35)$$

for each adjoining atom. Choosing the spring constant κ sufficiently large will ensure that during the motion the positions of the dummy copies (nearly) coincide; otherwise a large energy penalty is generated. At the level of the normal modes, the introduction of the dummy copies will lead to more vibrations than in the true system. However, the addition of the V_A terms to the potential energy will result in a clear energy separation between the physical modes where the dummy atoms coincide, and the unphysical modes involving stretches between dummy copies of the same atom. These unphysical modes are all at much higher energies, and can be easily deleted from the frequency list.

The value of κ to be used depends on the following considerations. On the one hand, κ must be in the decoupling regime, so that the frequency spectrum of the physical modes does not depend on its value. On the other hand, an exaggeratedly large value for κ can lead to numerical instability as the condition number of the Hessian matrix deteriorates. In practice, a value of about 10^7 Hartree/(Bohr length)² was found to provide excellent results.

2.4.2 Analytical constraints

An alternative implementation imposes the constraints in an analytical way. After the partitioning of the molecular system into blocks (non-adjoined - adjoined - free atoms) each atom A is characterized by the share number s_A (i.e. the number of blocks to which A belongs), and the corresponding block indices $b(A)_1, \dots, b(A)_{s_A}$. The positions of the atoms are still governed by the transformations (2.28). However, the transformation parameters $\{p_b\}$ of the blocks are no longer independent variables when adjoining atoms are present. An adjoining atom belongs to more than one block, and the fact that the position of the adjoining atom must be the same when transformed according to any of these blocks, imposes constraints among the parameters of the blocks to which the adjoining atom belongs. For an adjoining atom L which is shared by s_L blocks $b(L)_1, b(L)_2, \dots, b(L)_{s_L}$, these constraints can be neatly expressed as

$$\mathbf{g}(\mathbf{r}_L^0, \hat{p}_{b(L)_1}) = \mathbf{g}(\mathbf{r}_L^0, \hat{p}_{b(L)_2}) = \dots = \mathbf{g}(\mathbf{r}_L^0, \hat{p}_{b(L)_{s_L}}). \quad (2.36)$$

When all adjoining atoms are considered, the set of constraints of the form in Eq. (2.36) therefore impose relationships between the group parameters $\{p_b\}$, identifying the allowed motions in the presence of adjoining atoms.

These relationships may also be viewed as defining an allowed hypersurface $p_b \equiv p_b(y)$ in $\{p_b\}$ space, which can be parametrized in terms of k independent variables $y_\pi, \pi = 1, \dots, k$. These y variables will be taken as the dynamical variables in the normal mode analysis. The number k of independent degrees of freedom is as yet unknown and depends on the number of adjoining atoms and the topological structure of the links between the various blocks. Constructing a complete functional dependence $p_b(y)$ can be hideously complicated. For the purpose of a vibrational analysis, however, we only need small excursions from the reference geometry. As we shall see, this simplifies things considerably.

Construction of the Hessian

The potential energy surface expressed in terms of the y coordinates becomes

$$W(y) = \tilde{V}(\{p_b(y)\}) = V(\{\mathbf{g}(\mathbf{r}_A^0, \hat{p}_{b(A)_1}(y))\}). \quad (2.37)$$

Because of the constraints in Eq. (2.36), it is immaterial which block $b(A)_i$ is chosen to describe the motion of an adjoining atom in the Cartesian PES on the right of Eq. (2.37). For simplicity we always take the first block $b(A)_1$.

We assume that the reference geometry has been optimized in the presence of the adjoining atom constraints, i.e. it is gradient-free for all the independent variables y_π ,

$$0 = G'_\pi = \left(\frac{\partial W}{\partial y_\pi} \right)_0 = \sum_{b\alpha} \tilde{G}_{b\alpha} \left(\frac{\partial p_{b\alpha}}{\partial y_\pi} \right)_0. \quad (2.38)$$

The Hessian at the reference geometry is easily calculated using Eq. (2.37),

$$\begin{aligned} H'_{\pi,\tau} &= \left(\frac{\partial^2 W}{\partial y_\pi \partial y_\tau} \right)_0 \\ &= \sum_{b\alpha} \sum_{b'\alpha'} \tilde{H}_{b\alpha,b'\alpha'} \left(\frac{\partial p_{b\alpha}}{\partial y_\pi} \right)_0 \left(\frac{\partial p_{b'\alpha'}}{\partial y_\tau} \right)_0 + \sum_{b\alpha} \tilde{G}_{b\alpha} \left(\frac{\partial^2 p_{b\alpha}}{\partial y_\pi \partial y_\tau} \right)_0. \end{aligned} \quad (2.39)$$

The gradient and Hessian elements that appear in Eqs. (2.38-2.39) can be expressed in terms of the Cartesian ones using Eqs. (2.31) (not necessarily zero in this case) and (2.33). It is important however to realize that Eqs. (2.31) and (2.33) are only valid for a strict partition of the atoms, where each atom belongs to a unique block. In the present situation such a partition is done automatically by selecting the first block for each atom.

Solution of the unknown coefficients $\left(\frac{\partial p_{b\alpha}}{\partial y_\pi} \right)_0$ and $\left(\frac{\partial^2 p_{b\alpha}}{\partial y_\pi \partial y_\tau} \right)_0$ in Eqs. (2.38-2.39) is surprisingly easy. Using the y variables, the constraint equations in Eq. (2.36) can be re-expressed as identities. In a succinct notation, we have, for all Cartesian components μ and all adjoining atoms L , a sequence of $s_L - 1$ identities,

$$\prod_{i=1}^{s_L} \left[= g_\mu(\mathbf{r}_L^0, \hat{p}_{b(L)_i}(y)) \right]. \quad (2.40)$$

The first and second derivatives of Eq. (2.40) with respect to the y variables can be worked out using the chain rule,

$$\prod_{i=1}^{s_L} \left[= \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial p_{b(L)_i\alpha}}{\partial y_\pi} \right)_0 \right], \quad (2.41)$$

$$\prod_{i=1}^{s_L} \left[= \sum_{\alpha\alpha'} C_{L\mu}^{(\alpha\alpha')} \left(\frac{\partial p_{b(L)_i\alpha}}{\partial y_\pi} \right)_0 \left(\frac{\partial p_{b(L)_i\alpha'}}{\partial y_\tau} \right)_0 + \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial^2 p_{b(L)_i\alpha}}{\partial y_\pi \partial y_\tau} \right)_0 \right], \quad (2.42)$$

in terms of the familiar MBH coefficients $D_{L\mu}^{(\alpha)}$ and $C_{L\mu}^{(\alpha\alpha')}$ tabulated in Table 2.1.

For any particular $\pi (= 1, \dots, k)$, Eq. (2.41) expresses the fact that the first-order derivatives of the transformation parameters, $\left(\frac{\partial p_{b\alpha}}{\partial y_\pi}\right)_0 = x_{b\alpha}^{(\pi)}$, must be a solution $x^{(\pi)}$ to the homogeneous linear system of $3 \sum_L (s_L - 1)$ equations

$$Kx^{(\pi)} = 0 \quad (2.43)$$

where the matrix K has entries (with L specifying the adjoining atom; $\mu = 1, \dots, 3$; and $i = 1, \dots, s_L - 1$):

$$K_{\mu Li, b\alpha} = \begin{cases} D_{L\mu}^{(\alpha)} & \text{if } b = b(L)_i \\ -D_{L\mu}^{(\alpha)} & \text{if } b = b(L)_{i+1} \\ 0 & \text{otherwise.} \end{cases} \quad (2.44)$$

The linear system (2.43) therefore immediately determines the number of independent y variables, as the dimension of the null space of the matrix K .

For a particular combination $\pi\tau$, Eq. (2.42) then fixes the corresponding second-order derivatives of the transformation parameters, $\left(\frac{\partial^2 p_{b\alpha}}{\partial y_\pi \partial y_\tau}\right)_0 = x_{b\alpha}^{(\pi\tau)}$, as a solution $x^{(\pi\tau)}$ of the set of inhomogeneous equations

$$Kx^{(\pi\tau)} = y^{(\pi\tau)} \quad (2.45)$$

where the vector on the right has components

$$y_{\mu Li}^{(\pi\tau)} = \sum_{\alpha\alpha'} C_{L\mu}^{(\alpha\alpha')} [x_{b(L)_{i+1}\alpha}^{(\pi)} x_{b(L)_{i+1}\alpha'}^{(\tau)} - x_{b(L)_i\alpha}^{(\pi)} x_{b(L)_i\alpha'}^{(\tau)}]. \quad (2.46)$$

We conclude that the Hessian expressed in the y variables is completely determined by solving the linear systems in Eq. (2.43) and Eq. (2.45). Numerically, this can be done efficiently by performing a singular value decomposition (SVD) of the matrix K . When determining the null space of K , a singular value is considered being zero if it is smaller than a small threshold value to account for numerical inaccuracies. In practice, this never leads to ambiguities.

Construction of the mass matrix and the normal mode equations

With the Hessian given by Eq. (2.39), the further derivation of the normal modes proceeds in a standard fashion. The instantaneous velocities of the atoms in the molecule are given by the defining property

$$\dot{r}_{A\mu} = \sum_{\alpha\pi} \frac{\partial g_\mu}{\partial p_\alpha} (r_A^0, \hat{p}_{b(A)_1}(y)) \left(\frac{\partial p_{b(A)_1\alpha}}{\partial y_\pi}(y) \right) \dot{y}_\pi. \quad (2.47)$$

For a contribution of an adjoining atom to the kinetic energy one can pick arbitrarily a block $b(A)$ to which A belongs; any other choice is equivalent on account of Eq. (2.41). The kinetic energy in the y variables, expanded to second-order, therefore reads

$$T = \frac{1}{2} \sum_{A\mu} m_A \left(\sum_{\alpha\pi} D_{A\mu}^{(\alpha)} x_{b(A)1\alpha}^{(\pi)} \dot{y}_\pi \right)^2 = \frac{1}{2} \sum_{\pi\tau} M'_{\pi\tau} \dot{y}_\pi \dot{y}_\tau. \quad (2.48)$$

The last identity in Eq. (2.48) defines the mass matrix M' corresponding to the use of y as the dynamical variables; its entries are given by

$$M'_{\pi\tau} = \sum_{A\mu} m_A \sum_{\alpha\alpha'} x_{b(A)1\alpha}^{(\pi)} x_{b(A)1\alpha'}^{(\tau)} D_{A\mu}^{(\alpha)} D_{A\mu}^{(\alpha')}. \quad (2.49)$$

The final normal mode equations are represented by the generalized eigenvalue problem of dimension k ,

$$H'v = \lambda M'v, \quad (2.50)$$

with $v = \omega/2\pi$ the vibrational frequency ($\lambda = \omega^2$) and v the corresponding normal mode eigenvector.

Comparison with restraints technique with dummy atoms

By adding a strong harmonic coupling between dummy copies of the adjoining atoms, the system behaves as if the blocks were linked. The restraints technique has the advantage that the non-adjoined version of the MBH can be used without additional implementations. The dummy atom technique was shown to work well, but has a significant disadvantage in the sense that the introduction of the dummy atoms leads to an extension of the dynamical system. Thereby the dimension of the Hessian temporarily increases in size, and for very large molecular systems this can be quite bothersome and time-consuming. Moreover, the strong coupling between the dummies can cause numerical instabilities in the diagonalization of the Hessian for very large systems.

The exact treatment of the MBH with analytical constraints is able to overcome the shortcomings of the dummy atom concept, because it avoids the extra degrees of freedom inherent to dummy atoms. The main idea is to introduce block parameters for each block and then, in a second step, to impose the link constraints on the set of block parameters. Some implementation efforts are required, but a considerable reduction in computer time can be realized.

2.5 Procedures and implementation

This section translates the theoretical results from the previous sections into procedures which can be implemented. An overview of the different MBH routines summarizes the several possible implementation schemes. The implementation into large-scale modeling packages is important for any new method in molecular modeling, as the break-through of a method depends not only on its accuracy but also on its availability in the users' favorite simulation programs, and on its computational performance.

We here present five general schemes:

- Standard NMA
- MBH in internal coordinates
- MBH in block parameters, no adjoined blocks
- MBH in block parameters, using dummy atoms
- MBH in block parameters, using analytical constraints

Each scheme was implemented and tested in Python [36]. E.g. for the standard NMA scheme, the frequencies were compared with the standard output of existing packages (Gaussian03 [37], CP2K [38], ADF [39], CHARMM [29], Q-Chem [40]).

Partial geometry optimization

MBH is able to yield physical frequencies in partially optimized systems. The examples of Figures 2.4 and 2.5 will be used to illustrate the procedure to perform such a partial geometry optimization.

In the example of Figure 2.4 two blocks are defined, the atoms 2–5 and the atoms 7–9. In the *Z*-formalism, the set of internal coordinates consists of distances d_{ij} , angles a_{ijk} , and dihedral angles d_{ijkl} . The internal coordinates in boldface in Figure 2.4 are kept fixed during the optimization (this option is, e.g., provided in the modeling package Gaussian03 [37]). Note that one can as well keep the atoms of block 1 at their absolute positions, while internal coordinates should be fixed for the other blocks. This is equivalent, as we are interested in the partially optimized *internal* geometry and not the absolute space frame coordinates. In Figure 2.4, for instance, the positions \mathbf{r}_2 – \mathbf{r}_5 and the internal coordinates r_{78} , r_{89} , a_{789} of the second block can be held constant.

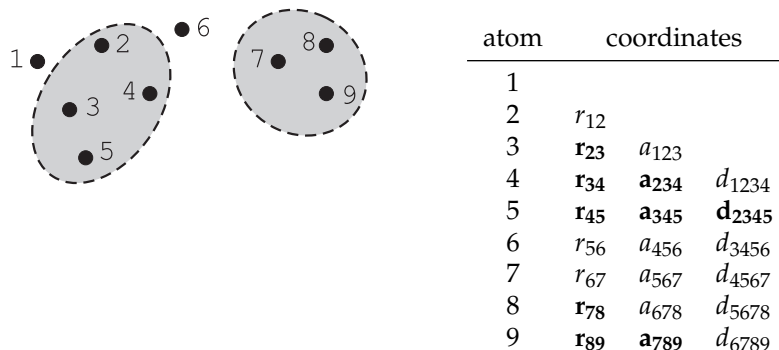


Figure 2.4: System of 9 atoms with 2 MBH blocks. In the Z-matrix construction, the atoms of each block (2–5 and 7–9) should be numbered consecutively. The internal coordinates are distances d_{ij} , angles a_{ijk} , and dihedral angles d_{ijkl} . The Z-coordinates in boldface are kept fixed during the vibrational analysis.

The example of Figure 2.5 shows several cases for the share number, which introduce other types of constraints in the geometry optimization. In Figure 2.5c the share number of atom 3 is three. During the optimization, the displacement of each block for the geometry update is limited to rotations about an axis through the adjoining atom. In Figure 2.5b, the relative motion of the blocks is restricted to a rotation about the common axis.

2.5.1 Standard NMA procedure with the full Hessian

We recall the standard procedure to perform NMA with the full Hessian. In practice, two situations are possible: either internal coordinates are used, either Cartesian quantities are used.

Scheme “standard NMA”

1. **Full geometry optimization.**
Find a minimum energy state on the PES.
2. **Calculate the second derivatives of the PES.**
 - If internal coordinates $\{\theta_I\}$ are used, construct $H^{(ii)}$ and the corresponding mass matrix of Eq. (1.25).

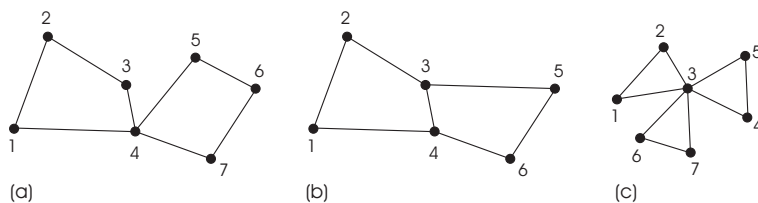


Figure 2.5: Example of adjoining atoms. (a) Two blocks $[1,2,3,4],[4,5,6,7]$ having one adjoining atom (4) in common. Keeping the position of $[1,2,3,4]$ fixed, the relative motion of $[4,5,6,7]$ is limited to rotations about any axis through the adjoining atom. (b) Two blocks $[1,2,3,4],[3,4,5,6]$ having two adjoining atoms (3,4) in common (=hinge). Keeping the position of $[1,2,3,4]$ fixed, the relative motion of $[3,4,5,6]$ is limited to hinge-bending motions about the axis through the two adjoining atoms. (c) Three blocks $[1,2,3],[3,4,5],[3,6,7]$ having one adjoining atom (3) in common. The share number of atom 3 is three.

- If Cartesian coordinates $\{r_{A\mu}\}$ are used, construct the Cartesian Hessian H and mass matrix M of Eq. (1.13).

3. Calculate eigenvalues and eigenvectors.

Diagonalize the appropriate mass-weighted Hessian.

- If internal coordinates $\{\theta_I\}$ are used, solve the NMA Eqs. (1.27).
- If Cartesian coordinates $\{r_{A\mu}\}$ are used, solve the NMA equations (1.17).

In extended systems, the calculation, storage and diagonalization of the Hessian might become problematic. With an efficient implementation, the MBH reduces the computational cost of the second derivative calculation, the storage and the diagonalization, since less degrees of freedom are considered. An additional step for MBH is the definition of the blocks. Instead of a complete energy minimization a partial geometry optimization is performed. Depending on the features of the simulation program, all derivatives or a subset of the derivatives should be calculated, which are used to construct the MBH matrices. In the last step, a smaller matrix than the full Hessian is to be diagonalized. In case of adjoined blocks, the link constraints require some additional manipulations.

2.5.2 MBH in internal coordinates

The MBH is easiest implemented if user-defined internal coordinates are available. The MBH formulated in internal coordinates can handle any topology, even with linear blocks, a branched chain of blocks or a ring structure of blocks.

Scheme "MBH in internal coordinates"

- 1. Block selection and constraints selection.**
Choose blocks and number the atoms of the blocks consecutively.
- 2. Partial geometry optimization.**
Minimize the energy while the internal geometry of each block is kept fixed.
- 3. Calculation of second derivatives.**
The second derivatives at the partially optimized (reference) structure should be calculated at the same level of theory as used in the partial optimization procedure. Construct $H^{(ii)}$ and $M^{(ii)}$ of Eq. (2.4) by omitting the rows and columns of the Hessian corresponding to fixed internal coordinates.
Note that it is of computational profit to calculate, if possible, only the Hessian elements corresponding to the $\{\theta_{I_E}\}$ coordinates, and not those corresponding to the fixed $\{\theta_{I_F}\}$ coordinates.
- 4. Calculate eigenvalues and eigenmodes.**
The frequencies and normal modes follow from solving the Eqs. (2.4).

2.5.3 MBH using group parameters as variables

Via Cartesian coordinates or directly

When the derivatives with respect to internal coordinates are not readily available, it is natural to use the group parameters $\{p_{b\alpha}\}$ of Eq. (2.5) to express the MBH normal mode equations, as those parameters represent translations and rotations of the blocks. If it is possible to calculate second derivatives with respect to the group parameters directly (e.g. using finite difference approximation), one may choose the set $\{p_{b\alpha}\}$ as coordinates and calculate the MBH matrices in group parameters (\bar{H}, \bar{M}) directly without calculating the full Hessian. However, this option is seldom readily available. Therefore

the procedure is here based on the Cartesian quantities (Hessian, gradient, masses), as provided by most molecular modeling packages.

We first consider the instructive MBH scheme in case no adjoined blocks are present (the share number s_A is 1 for all atoms of the system), before tackling cases where adjoined blocks are involved.

Scheme “MBH without adjoined blocks”

1. **Block selection.**

Choose the blocks.

2. **Partial geometry optimization.**

Minimize the energy while the internal geometry of each block is kept fixed.

3. **Calculation of second derivatives.**

The second derivatives at the partially optimized (reference) structure should be calculated at the same level of theory as used in the partial optimization procedure.

Starting from the Cartesian Hessian and gradient, first calculate the coefficients $D_{A\mu}^{(\alpha)}$ and $C_{A\mu}^{(\alpha\alpha')}$ using Table 2.1. Construct \tilde{H} using Eq. (2.21-2.22-2.23). Construct \tilde{M} using Eq. (2.25-2.26-2.27).

4. **Calculate eigenvalues and eigenmodes.**

The frequencies and normal modes follow from solving the Eqs. (2.32).

For a linear block ($d_b = 5$) one should decide upon an axis ($\mu = x, y, z$) of the space-fixed frame and omit rotations around this axis, following the discussion in Section 2.3.3.

Six more frequencies/modes will be found than in the procedure using internal coordinates. Those have zero frequency and correspond with the global translational and rotational modes of the system. One can obtain a reduction in the number of required Cartesian Hessian matrix elements, by exploiting the six invariances of the Hessian. Making use of the six predictable zero frequency modes (global translations and rotations), it is possible to avoid the calculation of six rows/columns of the Hessian corresponding to the parameters of a certain block B . The reduction is considerable if the number of atoms in block B is large. This property was proved in Ref. [35].

Next we consider the case where adjoined blocks are present. As shown in Section 2.4, there are two possibilities to impose the link constraints between blocks: using a restraints technique with dummy atoms or using analytical constraints. The implementation schemes will be summarized consecutively. Again, we start from the Hessian and gradient in Cartesian coordinates.

Scheme “MBH with dummy atoms”

- 1. Block selection and constraints selection.**
Identify the blocks b in the molecular system, and their respective type and number d_b of transformation parameters. Assign the atoms to the various blocks, by listing for each atom A the blocks to which it belongs and the share number s_A .
- 2. Partial geometry optimization.**
Minimize the energy while the internal coordinates of each block are kept fixed.
- 3. Design dummy atoms.**
For each adjoining atom ($s_A > 1$), extend positions, mass matrix, gradient vector, Hessian matrix in order to include the $s_A - 1$ additional dummy atoms in the molecular system, as described in Sec.2.4.1. Add the artificial dummy atom coupling terms to the appropriate matrix elements of the Hessian.
- 4. Calculation of second derivatives.**
This part remains unchanged, see step 3 of the “MBH without adjoined blocks” scheme.
- 5. Calculate eigenvalues and eigenmodes.**
This part remains unchanged, see step 4 of the “MBH without adjoined blocks” scheme.

Scheme “MBH with analytical link constraints”

- 1. Block selection and constraints selection.**
Identify the blocks b in the molecular system, and their respective type and number d_b of transformation parameters. Assign the atoms to the various blocks, by listing for each atom A the blocks to which it belongs and the share number s_A .

2. Partial geometry optimization.

Minimize the energy while the internal coordinates of each block are kept fixed.

3. Determination of the actual degrees of freedom.

With $s = \sum_A 3(s_A - 1)$ and $d = \sum_b d_b$, construct the $s \times d$ matrix K according to Eq. (2.44). Perform a singular value decomposition, $K = VSW^T$, where W is an orthogonal $d \times d$ matrix, S is a diagonal $s \times d$ matrix containing the singular values, and V is an orthogonal $s \times s$ matrix. The number k of zero singular values equals the number of actual degrees of freedom, i.e. the number of y_π variables defined in Sec.2.4.2. In practice, a singular value is considered zero below a small threshold value, to account for numerical inaccuracy. The k columns of W corresponding to the zero singular values can be taken as the orthonormal basis vectors $x^{(\pi)}$ in Eq. (2.43), spanning the null space of K ; they are collected in a $d \times k$ matrix X .

4. Construction of the $(d \times 1)$ gradient \tilde{G} , and of the $d \times d$ Hessian \tilde{H} and mass matrix \tilde{M} .

These quantities correspond to using the d block parameters $p_{b\alpha}$ as dynamical variables, and can be obtained in terms of the Cartesian $(3N_a \times 1)$ gradient G and the $3N_a \times 3N_a$ Hessian H and mass matrix M . They read $\tilde{G} = U^T G$, $\tilde{M} = U^T M U$, and $\tilde{H} = U^T H U + \tilde{R}$, where the $3N_a \times d$ transformation matrix U is defined as $U_{A\mu, b\alpha} = D_{A\mu}^{(\alpha)} \delta_{b, b(A)_1}$. In nonequilibrium, \tilde{H} receives a gradient correction \tilde{R} given by $\tilde{R}_{b\alpha, b'\alpha'} = \delta_{b, b'} \sum_{A \in b, \mu} G_{A\mu} C_{A\mu}^{(\alpha\alpha')} \delta_{b, b(A)_1}$. The coefficients $C^{(\alpha\alpha')}$ and $D^{(\alpha)}$ are tabulated in Table 2.1.

5. Construction of the $k \times k$ Hessian H' and mass matrix M' .

These quantities correspond to using the genuine independent variables y_π , $\pi = 1, \dots, k$, as dynamical variables. According to Eqs. (2.39, 2.49) they read $M' = X^T \tilde{M} X$, and $H' = X^T \tilde{H} X + R'$. In nonequilibrium, H' receives a gradient correction R' given by $R'_{\pi, \tau} = \sum_{b\alpha} \tilde{G}_{b\alpha} x_{b\alpha}^{(\pi\tau)}$. The $d \times 1$ vectors $x^{(\pi\tau)}$ are obtained as a solution to the inhomogeneous linear system in Eq. (2.45). Since the singular value decomposition of the matrix K has already been constructed in step 3, this requires little additional effort.

6. Calculate eigenvalues and eigenmodes.

Calculate the k eigenvalues and eigenvectors of $M'^{-1/2} H' M'^{-1/2}$. These yield the desired frequencies and normal modes of the linked block system.

Both procedures for adjoined blocks were implemented in Python [36] and their equivalence was tested numerically [41].

3 Validation of the MBH

3.1 Introduction

3.1.1 Investigated quantities

As was shown in Chapter 2, the MBH method yields coordinate independent frequencies, even when the molecular system is partially optimized, due to the gradient corrections. The MBH should also be validated for its performance in reproducing the physical quantities generated by NMA. In general, one can be interested in (1) the atomic motion involved in the vibrational modes themselves, e.g. in order to explore configurational space and study conformational changes; (2) the values of the frequencies, e.g. to analyze fingerprint regions at characteristic frequencies in vibrational spectroscopy; (3) ratios of partition functions leading to thermodynamical properties, e.g. the free energy difference determining the reaction rate constant.

During my PhD, the MBH and other partial Hessian methods have been tested on a series of test cases ranging from 9 to 8252 atoms, as tabulated in Table 3.1. Moreover, a test set of reactions has been investigated for the reproduction of free energies in Ref. [42] (not included in the table). In this chapter, the main features of the MBH are illustrated with three examples. In Section 3.2 the MBH is compared to the PHVA in its ability to reproduce localized modes in a partially optimized system, with ethanol as a small test example. In Section 3.3, the MBH is validated for the reproduction of accurate reaction rate constants and applied to a reaction in explicit solvent. In Section 3.4 an MBH partitioning with adjoined blocks is applied to reproduce the lowest frequency modes of

system	atoms	degrees of freedom	Reference
ethanol	9	27	[13]
<i>n</i> -octyl-ether	51	153	[13]
dimethylether	9	27	[16]
isobutane	14	52	[16]
<i>n</i> -alkanes	11–32	33–96	[16]
alanine-20-peptide	203	603	[16]
crambin	648	1944	[16]
caspase-1	8252	24756	[41]
amines	13–31	39–93	[43]
quinine	48	144	[43]
bis-cinchona	106	318	[43]
LAO binding protein	3649	10947	[43]

Table 3.1: Investigated test cases with number of atoms (N), number of degrees of freedom ($3N$), and reference to the corresponding paper.

the protein crambin. A more complete, extended validation can be found in the papers included in the Appendix.

3.1.2 Validation tools

The MBH frequencies/modes are compared to those obtained with a full Hessian calculation, which serve as the benchmark values. We used several tools to evaluate the performance of the MBH numerically.

Square overlap

Assume that all modes have been normalized, such that $|\langle v|v\rangle|^2 = 1$. Note that $v^T M v = 1$ is the expression if non-mass-weighted modes are used, so here the modes $|v\rangle$ are implicitly assumed to be mass-weighted. The square overlap P_{ij} between an MBH mode $|v_i^{MBH}\rangle$ and a full Hessian mode $|v_j^{bench}\rangle$ is defined as

$$P_{ij} = |\langle v_i^{MBH} | v_j^{bench} \rangle|^2 \quad (3.1)$$

The square overlap is a number lying in a range 0 – 100% and quantifies how much two modes resemble each other.

Cumulative square overlap

The cumulative square overlap P_j for a full Hessian mode $|v_j^{bench}\rangle$ is defined as

$$P_j = \sum_i |\langle v_i^{MBH} | v_j^{bench} \rangle|^2 \quad (3.2)$$

The cumulative square overlap is also a number lying in the range 0 – 100%; it does not necessarily become 100% because the MBH modes do not form a complete basis. The cumulative square overlap expresses to what extent the full Hessian mode lies in the space spanned by the MBH modes.

Overlap plot

Frequencies can be compared by plotting the MBH frequencies as a function of the full Hessian frequencies. By attributing a color to each data point, the overlap information can be added to the plot, such that the plot combines a frequency and mode comparison. An example is given in Figure 3.6 for crambin. It is a visual tool to evaluate the performance of the MBH frequencies and modes simultaneously.

Tama factor

Following the procedure introduced by Tama [15], the first n MBH frequencies can be plotted against the first n full Hessian frequencies. In the low frequency range, a straight line can be fitted to the resulting plot:

$$v_{MBH} = d \cdot v_{bench}. \quad (3.3)$$

The slope d is a measure for the frequency over- or underestimation. Note that this procedure is not applicable to all frequency ranges, because the number of MBH frequencies is *a priori* lower than the number of benchmark frequencies.

3.2 MBH and PHVA for partially optimized structures

3.2.1 The difference between MBH and PHVA is a mass effect

The PHVA assigns an infinite mass to the non-optimized coordinates, whereas the MBH takes into account the finite mass of the block. This can be depicted

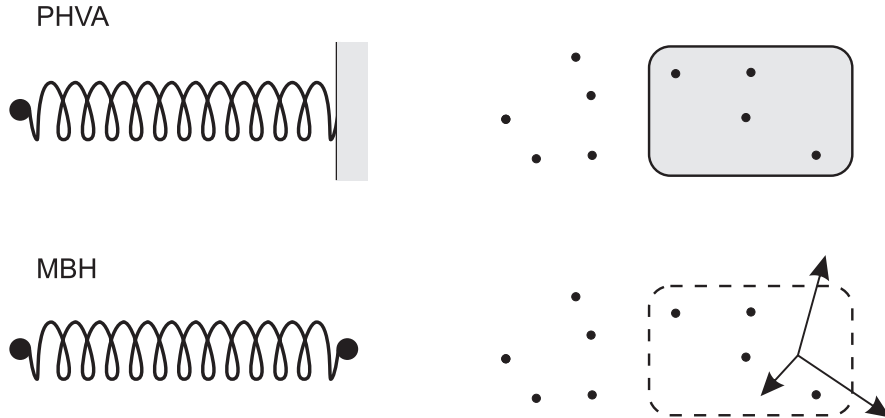


Figure 3.1: The difference between the PHVA and the MBH is a mass effect. The PHVA is best illustrated with a mass m connected by a spring to a wall; the fixed atoms form an immobile block. In the MBH the mass m is connected to a finite mass m' ; the fixed atoms form a mobile block that can translate and rotate.

by the schematic drawing in Figure 3.1. In the PHVA it is as if a mass m is connected by a spring to an immobile wall with infinite mass, while in the MBH the mass m is connected by a spring to some finite mass m' . Given the same spring constant k , the obtained frequency will nevertheless be different. In the PHVA situation, the frequency is $\sqrt{k/m}$, while in the MBH situation it becomes $\sqrt{k/\mu}$, with $\mu = mm'/(m + m')$ the reduced mass.

Because of the infinite mass in the PHVA model, the system is not invariant under global translations and rotations, so the six zero eigenvalues are lacking. In the MBH approach six zero eigenvalues still result from the NMA, which illustrates that MBH does not violate the global invariances.

The MBH and PHVA are both able to produce physical frequencies and modes, and reproduce localized modes accurately. The problems encountered in the standard NMA on a partially optimized structure are avoided by the introduction of the blocks. However, the PHVA shows spurious low frequencies because of the immobile block assumption for the fixed atoms. This is illustrated with a simple system, the ethanol molecule containing a well localized O-H stretch. It has the advantage that the exact frequencies are readily available and can be compared with those predicted by PHVA and MBH in partially optimized geometries.

3.2.2 Illustration: ethanol molecule

Benchmark spectrum

The entire molecule has been optimized at a so-called “high” level (B3LYP/6-31+g(d)), and this optimized geometry will be used further on as the reference. Benchmark frequencies of all present normal modes are obtained by diagonalizing the mass-weighted full 27×27 Hessian in Eq. (1.17), and are tabulated in the first column of Table 3.2.

An exact treatment should generate six eigenvalues exactly equal to zero corresponding to the global translation and rotation. In practice the values differ slightly from zero (varying between -1 and 8 cm^{-1}). Translational frequencies are sensitive to numerical errors in the construction of the Hessian. Rotational frequencies are in addition affected by the small residual forces due to the finite convergence criteria. The effect of the almost zero frequencies on the other 21 frequencies is here negligible: projecting out the overall translation and rotation gives six eigenmodes exactly equal to zero and does not affect the vibrational frequencies. Some of the modes have a clear interpretation: the two lowest frequencies (244 cm^{-1} , 295 cm^{-1}) correspond with internal rotations of the methyl top and the hydroxyl group, while the highest (3756 cm^{-1}) is associated with the highly localized O-H stretching mode.

PHVA and MBH spectrum

We applied the PHVA and MBH methods to two different cases: first to the fully optimized structure (Table 3.2) and then to several partially optimized structures (Table 3.3). The block choices are given in the heading of the tables. The shaded box indicates which atoms belong to the block.

The PHVA frequencies are calculated with Eq. (2.2). For the MBH frequencies, the “MBH without adjoining atoms” scheme was used of Section 2.5.3 since there are no adjoining atoms. The ordering of the calculated frequencies in the tables is determined by the maximum of the square of the overlap $|\langle v^{bench} | v \rangle|^2$ between the benchmark mode with frequency ν_{bench} and the calculated mode with frequency ν . For modes without a pronounced maximum overlap, this is of course rather arbitrary.

PHVA and MBH applied to the equilibrium structure

Frequencies are calculated for the fully optimized geometry while the part of the Hessian chosen to be included in the vibrational analysis is varied. In this

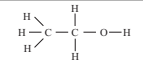
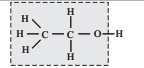
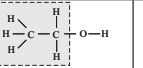
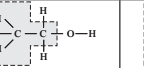
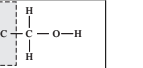
									
Full	Projected	PHVA	MBH	PHVA	MBH	PHVA	MBH	PHVA	MBH
-1	0	-	-	-	-	72	-	59	-
-1	0	-	-	-	-	242	-	94	-
0	0	-	-	-	-	678	-	211	-
1	0	-	-	-	-	-	-	355	-
2	0	-	-	-	-	-	-	-	-
8	0	-	-	-	-	-	-	-	-
internal rotation:									
244		-	-	-	-	-	249	-	244
295		275	289	299	289	279	297	292	295
C-C-O stretch:									
417		-	-	-	491	-	458	-	419
mixed modes:									
825		-	-	272	807	823	992	1024	831
902		-	-	-	-	-	1011	-	906
1037		-	-	-	-	-	-	748	1040
1104		-	-	-	1031	-	-	1058	1111
1185		-	-	-	-	-	-	-	1195
1270		1164	1208	1224	1245	1210	1218	1226	1277
1305		-	-	715	-	1254	1269	1262	1308
1419		-	-	-	-	-	-	-	-
1461		-	-	-	-	1317	1378	1445	1452
1505		-	-	-	-	-	-	-	-
1521		-	-	-	-	-	-	-	-
C-H stretch:									
1544		-	-	-	-	1501	1534	1541	1541
2995		-	-	245	-	2928	2977	2995	2995
3021		-	-	-	-	2877	3015	3023	3023
3049		-	-	-	-	-	-	-	-
3116		-	-	-	-	-	-	-	-
3124		-	-	-	-	-	-	-	-
O-H stretch:									
3756		3645	3708	3755	3756	3755	3756	3756	3756

Table 3.2: Normal mode frequencies (in cm^{-1}) of ethanol derived from the benchmark geometry, which corresponds with the geometry optimization obtained at B3LYP/6-31+g(d). The rigid body is composed of the atoms in the shaded region. In the left column translational and rotational frequencies from the full Hessian calculation are plotted before as well as after projection. Vibrational frequencies are not affected by this projection. The PHVA and MBH frequencies were ordered according to the maximum overlap with the benchmark modes.

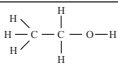
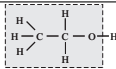
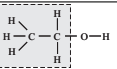
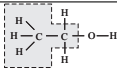
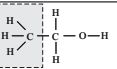
												
bench	Full PHVA MBH	Full PHVA MBH	Full PHVA MBH	Full PHVA MBH	Full PHVA MBH							
-1	-258	-	-264	-	-254	71	-	-251	60	-		
-1	-130	-	-132	-	-80	243	-	-90	95	-		
0	-62	-	-75	-	-61	678	-	-66	212	-		
1	-1	-	-2	-	-1	-	-	-1	355	-		
2	-1	-	0	-	1	-	-	1	-	-		
8	1	-	1	-	1	-	-	1	-	-		
244	91	-	42	245	-	124	-	248	124	-	246	
295	294	281	295	284	296	285	289	280	298	289	294	297
417	403	-	-	396	-	490	397	-	458	397	-	419
825	760	-	-	758	272	813	793	824	992	793	746	839
902	873	-	-	877	-	-	880	-	-	883	-	909
1037	1021	-	-	1025	-	1032	1025	-	1013	1029	1025	1041
1104	1070	-	-	1096	-	-	1097	-	-	1098	1058	1116
1185	1147	-	-	1156	-	-	1177	-	-	1177	-	1201
1270	1257	1161	1205	1261	1224	1244	1266	1210	1218	1267	1225	1280
1305	1267	-	-	1277	715	-	1298	1252	1268	1300	1262	1313
1419	1384	-	-	1385	-	-	1388	-	-	1390	-	-
1461	1422	-	-	1429	-	-	1452	1315	1376	1454	-	-
1505	1489	-	-	1488	-	-	1488	-	-	1488	1444	1452
1521	1505	-	-	1506	-	-	1507	-	-	1508	-	-
1544	1527	-	-	1527	-	-	1544	1502	1535	1544	1542	1542
2995	3140	-	-	3137	-	-	2995	2877	2978	2994	2995	2995
3021	3162	-	-	3162	-	-	3023	2929	3016	3022	3023	3023
3049	3177	-	-	3173	-	-	3162	-	-	3162	-	-
3116	3242	-	-	3243	-	-	3243	-	-	3242	-	-
3124	3253	-	-	3252	-	-	3251	-	-	3251	-	-
3756	3751	3640	3703	3755	3755	3755	3756	3756	3756	3756	3756	3756

Table 3.3: Normal mode frequencies (in cm^{-1}) of ethanol derived on basis of partially optimized geometries at the B3LYP/6-31+g(d) level of theory. The rigid body is composed of atoms in the shaded region and its geometry is originating from a geometry optimization of the whole molecule at the low HF/STO-3g level. Benchmark frequencies are given in the left column for comparison. The PHVA and MBH frequencies were ordered according to the maximum overlap with the benchmark modes.

case there are no remaining residual forces. The results are given in Table 3.2 for various rigid body sizes. The shaded box indicates the part of the molecule that is not included in the calculation of the Hessian. For instance, the second column of the table reports the three frequencies corresponding with the modes generated by a rigid body and one single atom that can vibrate. The only non-fixed atom in the vibrational analysis is the hydrogen atom of the hydroxyl group.

Only three PHVA and MBH modes are calculated in the case where only the H-atom of the hydroxyl group is taken into account (second column of Table 3.2). Visualization of the three modes revealed that they all qualitatively correspond with a normal mode of the benchmark frequency spectrum, but that their values are underestimated (benchmark values are 295 cm^{-1} , 1270 cm^{-1} and 3756 cm^{-1}). In the PHVA method the underestimation is even more pronounced due to the use of infinite masses. This is a general conclusion: *the PHVA fails in accurately reproducing the benchmark values especially in the small and medium frequency region, while the quantitative agreement of the MBH predictions of localized modes - taking place in the non-fixed region - with the benchmark values is manifestly present.* The agreement is even very close in the last case where the rigid body consists of a fixed methyl group. The O-H stretch mode is always present in the frequency spectrum, only in the case of one relaxed atom (the hydrogen) the obtained frequency is slightly underestimated.

Looking at lower frequencies, the MBH approach gives consistently better results than the PHVA, because the reduced mass effect is taken into account. As the total mass of the fixed block decreases, it is obvious that the PHVA induces spurious low frequency modes of the order of $60 - 100\text{ cm}^{-1}$, corresponding with translations/rotations of the non-fixed atoms in the field of their environment (the fixed atoms). Summarizing, when working with molecules in the gas phase, the use of the MBH model is highly recommended as soon as the total mass of the relaxed atoms becomes of the same order as the total mass of fixed block.

PHVA and MBH applied to partially optimized structures

We introduce partially optimized geometries, and verify whether the relevant calculated frequencies in the active site can be reproduced by both the PHVA and MBH models in an accurate way with respect to the benchmark frequencies. Initially the ethanol molecule has been optimized at the lower HF/STO-3g level. Then the system was partially optimized at the higher B3LYP/6-31+g(d) level while keeping the atoms of the rigid block fixed at their initial HF/STO-3g

positions. Results are collected in Table 3.3. The atoms belonging to the shaded box were not optimized at the high level but were kept fixed at their low level geometries.

For each case the frequencies resulting from a full Hessian (calculated at B3LYP/6-31+g(d)) diagonalization, i.e. from a standard normal mode analysis as well as the normal modes resulting from the PHVA and the MBH method are tabulated. The standard frequency analysis gives always a number of spurious imaginary frequencies, as could be expected since residual forces on the non-optimized atoms disturb the evaluation of the frequencies. In addition, the positive frequencies deviate substantially from the benchmark values given in Table 3.2. This means that the normal frequencies generated by standard procedures in program packages such as Gaussian03 [37] in molecules whose atomic positions have not been optimized at the same level of theory, are far from being accurate.

By applying the PHVA or the MBH model the unphysical imaginary frequencies disappear, but the resulting frequencies differ significantly between the two methods. The MBH results converge rapidly to the benchmark values, highlighting the efficiency of the proposed MBH model. A striking resemblance is even observed in the last column where the fixed body is restricted to the ending methyl group. The low frequency spectrum of the PHVA method however deviates largely from the benchmark values. This is entirely due to the reduced mass effect, inducing spurious unphysical modes.

Finally we compare the MBH results of the optimized and partially optimized geometry. While the full Hessian frequencies are very sensitive to the exact molecular geometry, it is remarkable that PHVA and MBH frequencies from the partially optimized geometry are very close to those obtained with the benchmark geometry with the same block size. This indicates that the PHVA and MBH models are less sensitive to the exact internal geometry of the fixed block.

Conclusions

We summarize the main conclusions of the ethanol example. The MBH and PHVA are both able to get rid of the spurious imaginary frequencies, which are inherent to partially optimized structures. The MBH yields six frequencies corresponding with global translations and rotations, the PHVA does not. Both the MBH and PHVA reproduce the highly localized modes well, but the PHVA yields spurious modes in the low frequency range, because of the infinite mass assigned to the fixed atoms.

3.3 Reaction rate constants

3.3.1 MBH reproduces reaction rate constants

Frequencies are an essential ingredient in predicting reaction rate coefficients due to their input in the vibrational partition functions. The question arises whether the MBH method is able to describe the chemical reaction kinetics in an accurate way. Chemical kinetics in static approaches are still widely based on transition state theory (TST) [44–47]. Key parameters here are the reaction energy barrier between the reactants and activated complex (the transition state) and the vibrational frequencies. In the harmonic oscillator approximation, the molecular partition function is factorized in a translational, rotational and vibrational contribution, where the latter is completely determined by the eigenfrequencies. The accurate (QM) calculation of frequencies is thus essential.

Frequencies are usually obtained by NMA. This is the main bottleneck in *ab initio* predictions of chemical kinetics in large molecular systems, since frequency calculations are computationally very demanding. An accurate *ab initio* description of chemical kinetics of reactions in gas phase is nowadays perfectly practicable for moderate-sized molecules, but once the molecular environment comes into play (e.g. solvent), one has to adapt the level of theory to make the computation feasible [48]. This puts a heavy burden on the accuracy of the numerical results. If a molecular mechanics (MM) force field is used instead of a quantum mechanics (QM) or hybrid (QM/MM) [1–4] description, the frequency calculation becomes less problematic, though even at the full MM level other issues, such as the storage and manipulation of the huge Hessian matrices associated with very large systems, can become prohibitive in real applications. Anyway, chemical reactions inherently involve bond breaking and charge transfer, so it is essential to provide a QM description for (at least) the reactive region and a full MM description is usually no option.

In addition, there are computational limitations in the geometry optimization of extended systems at a high level of theory (LOT). Very often one goes over to a partial optimization: the interesting region containing the active site is optimized at a high LOT, while the environment is kept fixed at a low LOT geometry. This approach permits to get an *ab initio* description of the chemically active site in large molecular systems, but at the same time, it creates several new problems. One of them is the extraction of accurate frequencies for the relevant vibrational modes. All partially optimized systems are nonequilibrium structures, and instead of the standard full Hessian NMA, a partial Hessian method (PHVA, MBH) should be used.

In principle, the expression of the rate constant k includes all normal vibrational modes in reactants and activated complex. It is inherent to both MBH and PHVA approaches that the number of frequencies is always smaller than in a standard frequency calculation. The question arises whether this reduction has a significant influence on the reaction rate constant. The thorough validation in Ref. [42] demonstrates that the normal modes disappearing when defining fixed blocks, have little influence on the chemical kinetics. The MBH method opens a lot of perspectives in predicting accurate kinetic parameters in chemical reactions of extended molecular systems where the standard full Hessian procedure fails.

Test cases for the validation

Ref. [42] is devoted to the validation of MBH as an adequate method to predict rate constants. We have compared the MBH and PHVA predictions k^{MBH} and k^{PHVA} with the full Hessian benchmark values k , for a variety of reactions and with various block choices.

The test set included a prototype substitution reaction, a hydrogen transfer reaction, as well as several radical addition reactions, since these have a localized reactive site. Several types of blocks were considered: they could include the reactive center (which at first view looks as a surprising choice), be directly connected with the reactive center by a single bond, or be separated by more than one bond. Also the case of multiple blocks was considered, and is illustrated with a more extended aminophosphonate system in Section 3.3.2, for which explicit solvent molecules are taken into account.

Benchmark values

Benchmark structures, frequencies and rate constants were generated with a full geometry optimization at a high level of theory (DFT/B3LYP/6-311g**) with tight convergence criteria such that the residual gradients on the PES are negligibly small. Consequently a frequency calculation was carried out at the same level of theory for the whole molecular system. These equilibrium geometries permitted to calculate the reaction rate with the full Cartesian Hessian frequencies.

Transition state theory (TST) has been proven to be very useful to determine the reaction rate constants [44–47]. It supposes that the transition state or activated complex is in equilibrium with the reactants, although, strictly speaking this hypothesis is not valid since the transition state corresponds to a saddle point

rather than a minimum on the PES. Within this assumption the rate constant is completely determined by the microscopic partition functions and the reaction barrier at 0 K.

Within the harmonic oscillator approximation, the $3N$ degrees of freedom of a N -atom system can be decoupled into three groups of independent motions – 3 translational, 3 rotational and $3N - 6$ vibrational motions – that all contribute to the total molecular partition function q :

$$q = q_{trans} q_{rot} q_{vib}. \quad (3.4)$$

Each vibration with frequency ν_i gives a contribution

$$q_{(v_i)} = \frac{e^{-h\nu_i/2k_B T}}{1 - e^{-h\nu_i/k_B T}} \quad (3.5)$$

(with h the Planck constant, k_B the Boltzmann constant, T the temperature), to the total vibrational partition function

$$q_{vib} = \prod_i q_{(v_i)} \quad (3.6)$$

For a unimolecular reaction $A \rightarrow A^\ddagger \rightarrow B$ or $A \rightarrow A^\ddagger \rightarrow B + C$ (with \ddagger indicating the activated complex) the rate constant k is given by:

$$k(T) = \frac{k_B T}{h} \frac{q(\ddagger)/V}{q(A)/V} e^{-\Delta E_0/k_B T} \quad (3.7)$$

V is the volume and ΔE_0 represents the molecular energy difference at 0 K between the activated complex and the reactants. The transition state frequency is assumed not to be included in the partition function $q(\ddagger)$ of the activated complex. k is expressed in units s^{-1} .

For a bimolecular reaction $A + B \rightarrow (AB)^\ddagger \rightarrow C$ or $A + B \rightarrow (AB)^\ddagger \rightarrow C + D$ the expression for the rate constant becomes:

$$k(T) = \frac{k_B T}{h} \frac{q(\ddagger)/V}{q(A)/V q(B)/V} e^{-\Delta E_0/k_B T} \quad (3.8)$$

expressed in units $\text{m}^3 \text{mol}^{-1} \text{s}^{-1}$.

MBH and PHVA rate constants

The partially optimized geometries are obtained as follows. First one optimizes the system at a low level of theory (HF/STO-3g) to find a plausible starting

structure. Then the rigid blocks are introduced and the system is partially optimized at a high level of theory (DFT/B3LYP/6-311g**), while keeping the rigid blocks fixed at their initial internal geometry. Only the position and orientation (six degrees of freedom) of each block are optimized. We remind that PHVA is always limited to a single block, whereas MBH is very suitable to treat multiple blocks. Next Hessian elements were calculated using the same high level of theory. The standard full Hessian frequency analysis would give unphysical results due to the residual forces present in the partially optimized structures. Instead the PHVA or MBH normal modes Eqs. (2.2) and (2.32) are constructed, as these yield physical frequencies. Obviously, the same rigid blocks are chosen as those considered in the precedent partial optimization.

When MBH (PHVA) is used instead of the full Hessian NMA, the vibrational partition function q_{vib} in relation (3.4) is constructed with the MBH (PHVA) frequencies instead of the full Hessian frequencies:

$$q^{MBH} = q_{trans} q_{rot} q_{vib}^{MBH} \quad (3.9)$$

$$q^{PHVA} = q_{trans} q_{rot} q_{vib}^{PHVA} \quad (3.10)$$

Results

The chosen test set of chemical reactions allowed an exhaustive investigation of the influence on the rate constant of the position of the rigid block, the block's mass, its distance to the reactive center and the stoichiometry of the reaction. Several conclusions could be drawn from the analysis in Ref. [42].

Overall agreement. The overall agreement of the MBH rate constants with the benchmark values is remarkably good. The reaction rate constants are reproduced to within a reasonable factor of two. This observation holds for a variety of reactions: for unimolecular and bimolecular reactions, for radical and non radical reactions, and for heavy or small block masses. The deviation is within acceptable limits and is smaller than corrections induced by the level of theory [49], internal rotations [18, 20], tunnel effects and other factors [50, 51].

The agreement of the MBH predictions with the benchmark implies that the contribution of the omitted normal modes, inherent to the MBH method, is of the same magnitude for both the transition states and reactants. Apparently the omitted modes are not essential in the determination of the rate constant. These rather unimportant modes are localized in the fixed block or spread out over the fixed block and the optimized region. The more interesting modes

are located in the optimized region that contains the active site, and are well reproduced by the MBH approach. The coupling of the MBH modes with the modes localized in the fixed block is left out in the model, but a logical choice of the blocks makes this coupling irrelevant for the rate constant.

Bond-distance-rule. When a block is chosen too close to the active site, the coupling between MBH modes and the omitted modes is not always irrelevant anymore, leading to deviating MBH estimates. As a general rule, it is recommended not to bring the block region too close to the active site. A distance of at least one bond between reactive center and block border should be respected; this is the bond-distance-rule.

Electronic energy. The agreement between MBH and benchmark results also implies the canceling of errors in the electronic energy. The ground-state configuration of a partially optimized system is obviously less bound than the fully optimized system. However, the energy increase is mostly compensated by a similar increase of the binding energy of the TS, hence resulting in an almost equal reaction barrier.

Kinetic parameters. The kinetic parameters A and E_a , determined with a linear fit to the Arrhenius plot within the temperature range 300–700 K, have also been studied. Activation energies E_a remain almost unaffected as could be expected. Potential deviations of k^{MBH} originate from the pre-exponential factor A , which is mainly determined by the vibrational contribution to the partition function. In the test cases, the pre-exponential factor A is well reproduced by MBH, which is consistent with the results for the k values.

Delocalized systems. When the block's border crosses a bond that is part of a delocalized system, the k constants are badly reproduced by MBH, even when the bond-distance-rule is respected. Benzene rings, for instance, should therefore be considered as a whole block, rather than letting the block's border cross a bond of the ring.

Multiplication rule for multiple blocks. When multiple blocks a and b are introduced, the resulting rate constant can be deduced from the cases where only the a block or the b are present. Inspection of the MBH values obtained

with the individual blocks *a* and *b*, learns that the global effect of multiple blocks is mostly given by the following multiplication rule:

$$\frac{k_{a-b}^{MBH}}{k} \approx \frac{k_a^{MBH}}{k} \times \frac{k_b^{MBH}}{k} \quad (3.11)$$

This seems to be true for the forward and backward reaction.

Modified PHVA: PHVA*. The results of the MBH have been compared to those of the PHVA approach as well and, based on theoretical considerations, a modified PHVA method was presented (PHVA*). Acceptable reaction rate estimates can be obtained with PHVA* due to an additional factor depending on the mass and inertia tensor of the reactants and the TS.

Conclusion. The above discussion validates the use of the MBH model to predict the rate constant on an accurate level. A plausible choice of the fixed block is the only essential ingredient a potential user of MBH should take into account to get adequate predictions of chemical kinetics. MBH is computationally attractive, makes quantum chemical calculations feasible in extended molecular systems and preserves the true reaction mechanism.

3.3.2 Illustration: modeling a reaction in solvent

Finally we have tested the concept of multiple MBH blocks on a more realistic and more extended example, where several explicit solvent molecules are taken into account. Blocks can be chosen within the reacting molecule and/or the solvent molecules can be treated as blocks. We have chosen the cyclization of functionalized aminophosphonates, as a representative reaction occurring in an organic solvent (reaction R11 is the 11-th reaction of the test set discussed in Ref. [42], see Fig 3.2).

The studied system consists of the aminophosphonate anion together with a sodium counterion and solvated in dimethylether solvent molecules (DME). The latter were taken as model molecules for tetrahydrofuran. Three cases are considered: the reaction in absence of explicit solvent molecules (R11), in presence of one DME (R11 + 1 DME), and in presence of two DME (R11 + 2 DME). The benchmark values of *k*, *A*, *E_a* and ΔE_0 are given in Table 3.4 for *T* = 300 K. The ratios (for *k* and *A*) and differences (for energies) given between brackets indicate the effect of the solvation. The presence of one or two solvent

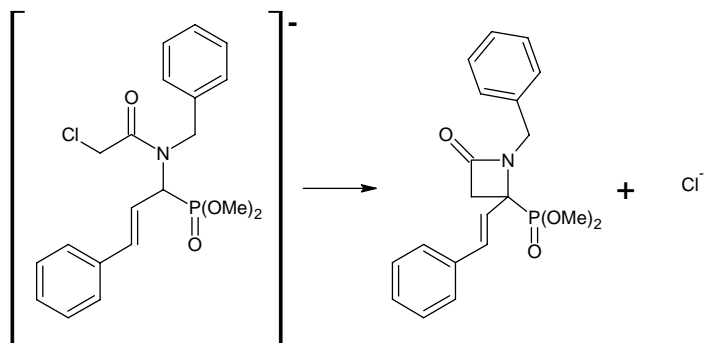


Figure 3.2: Reaction R11.

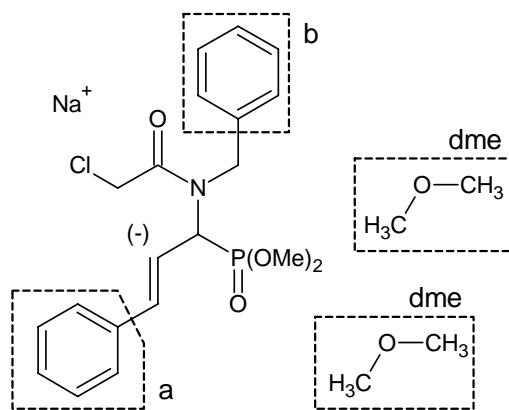


Figure 3.3: Definition of blocks, reaction R11.

	R11	R11 + 1 DME		R11 + 2 DME	
k	7.62E-15	6.79E-14	(8.91)	2.09E-12	(274.63)
A	2.10E+13	8.43E+12	(0.40)	5.64E+13	(2.69)
E_a	157.73	149.99	(-7.73)	146.19	(-11.54)
ΔE_0	159.85	151.78	(-8.07)	148.23	(-11.63)

Table 3.4: Benchmark results for reaction R11 without, with 1, and with 2 DME. Rate constant at 300 K and kinetic parameters fitted in temperature range 300–700 K. k and A in s^{-1} , energies in kJ/mol. Ratios (k and A) and differences (E_a and ΔE_0) between solvated and non-solvated values are given between brackets.

block	R11	R11 + 1 DME	R11 + 2 DME
a	1.71	1.62	1.43
b	0.98	0.97	0.88
dme	-	1.07	-
b-dme	-	1.04	-
dme-dme	-	-	0.94
b-dme-dme	-	-	0.83

Table 3.5: Ratio k^{MBH}/k of the MBH predictions with respect to the benchmark for reaction R11 without, with 1, and with 2 DME, calculated at 300 K. Several blocks choices are taken up.

molecules indeed increases the reaction rate constant by a factor 8.91 or 274.63 respectively with respect to the non-solvated situation.

The relevant question is whether the MBH model is capable of reproducing the enhancement of k due to the solvent. The several block choices are depicted in Figure 3.3, including the case of blocks within the reactant (a, b), as well as blocks consisting of solvent molecules (dme). Table 3.5 shows the ratios between the MBH estimates and the benchmark values of the rate constant. Block a is clearly not a good choice, which is easily understood when noting that the block's border cuts through a delocalized bond. Therefore possible combinations of a with blocks b or dme are not considered in the table. Block b and block dme on the other hand are excellent block choices: since the ratios are close to 1.0, the k enhancement 1:8.91:274.63 as reported by the benchmark is maintained and the MBH is thus clearly capable of reproducing the solvation effect. Multiple block combinations such as b-dme, dme-dme, and b-dme-dme reproduce the rate constant very well, which is in agreement with the multiplication rule as stated in Eq. (3.11).

Resuming, the multiple MBH has proven to be extremely useful and effective in predicting reaction rates, both with blocks belonging to the reactants and TS, or with blocks coinciding with solvent molecules.

3.4 Low frequency modes

3.4.1 MBH with adjoined blocks to describe the low frequency modes of proteins

By the introduction of blocks in the vibrational analysis, the internal degrees of freedom within a block are eliminated from the NMA. The extension to adjoined blocks in addition eliminates the three relative translational degrees of the blocks. Blocks can only move with respect to each other by rotating about an axis through the adjoining atom(s). Since the low frequency modes correspond typically with such angle bending motions, the adjoined version of the MBH is expected to perform well in reproducing the lowest frequency modes. This is indeed the case, as was shown in Ref. [16] for a series of test cases (see Table 3.1). As an example, we here consider the low frequency modes of peptide chains. Several partitionings are proposed and illustrated with crambin, a small protein with 46 amino acid residues.

Block choices for proteins

Block choices will be labeled by the share number s_A . For instance, in block choice $[C_\alpha]^1$ of Figure 3.4, the C_α carbon belongs to 1 block.

A protein chain can be seen as a sequence of rigid components, i.e. the peptide units, because *the peptide bond* (C=O-NH) is known to be a stable planar unit. It is straightforward to introduce blocks that include the entire peptide bond. Peptides have the ability to form different secondary structure elements due to the conformational flexibility of the backbone. There are essentially two backbone degrees of freedom for amino acid residue: the dihedral angles ϕ (C-N- C_α -C) and ψ (N- C_α -C-N). The angles describe rotations about the N- C_α bond and C_α -C bond respectively. The proposed block choices consider the peptide bond as a rigid block, but keep the ϕ and ψ angles as variables.

Figure 3.4 gives an overview of the considered plausible partitioning schemes for peptide chains. The smallest block choice corresponds to considering consecutive peptide bonds (C=O-NH) as rigid blocks, labeled by $[C_\alpha]^0$. In case $[C_\alpha]^1$, a complete residue – peptide bond plus side chain – is considered as a block. This corresponds to the RTB approach introduced by Durand et al. and extensively tested by Tama et al. on a large variety of proteins [14, 15]. From a methodological point of view it corresponds to the MBH approach introduced in our earlier papers [13, 35, 42]. It is also possible to combine the adjoined blocks with atoms not included in any block. This is useful for the calculation of localized modes of interest, e.g. in the chemically active part. Such block choices have not been considered here.

The adjoined MBH approach allows to connect the various blocks by one or more adjoining atoms. As a first approximation the peptide units may be adjoined by *one adjoining atom*: the C_α atom. This further reduces, by three, the total number of degrees of freedom for each considered block. Several options are possible to treat the side chain and the H atom connected to the C_α carbon, such as $[C_\alpha]^2$, $[C_\alpha]^{3a}$, $[C_\alpha]^4$, and $[C_\alpha]^{3b}$ in Figure 3.4. Note that the actual block choice for the endings of the chain has little influence for long peptide chains.

Finally the subsequent blocks may share *two adjoining atoms* as shown at the bottom of Figure 3.4. Each introduced block shares the C_α -N bond or the C_α -C with the block connecting the subsequent peptide units. As a result, the only degrees of freedom that are left over, are the internal rotations about the common C_α -N and C_α -C bonds, which are exactly the ϕ and ψ dihedral angles. This block choice is labeled by $[C_\alpha\text{-N}]^2+[C_\alpha\text{-C}]^2$, referring to the common C_α -N and C_α -C bonds.

By simply choosing adjoined blocks, as in $[C_\alpha\text{-N}]^2+[C_\alpha\text{-C}]^2$, one is able to select the dihedral angles as variables. The MBH model is thus useful to do NMA

with variables of interest, without the need for extra computational implementations for the variables construction (i.e. no complicated transformation to internal coordinates required). As for the geometry optimization, only the ϕ and ψ dihedral angles have to be force free; other internal coordinates do not necessarily have to be optimized in the MBH framework. From a computational point of view, the reduction of the Hessian size is impressive. Only two degrees of freedom (ϕ, ψ) are considered per residue. A protein with N_{res} residues easily has over $30N_{res}$ degrees of freedom, but the adjoined blocks with hinges reduces this number to $2N_R$.

Alanine dipeptide

The block choices were first tested on blocked (or capped) alanine dipeptide (N-acetyl-L-alanine-N'-methylamide) which is commonly studied as a prototype of nonglycine/nonproline protein backbones. The analysis of the alanine test case by means of the cumulative square overlap of Eq. (3.2) showed that grouping the atoms of the side chain into an adjoined block is a suitable choice to reproduce the lowest frequency modes, as is the case in $[C_\alpha]^{3a}$ and $[C_\alpha]^{4}$. Moreover the side chain- C_α -H bending and the C_α -H stretch are irrelevant for the lower spectrum, such that the hydrogen atom can be included in the side chain block (block choice $[C_\alpha]^{3b}$).

3.4.2 Illustration: crambin

As an example the well known protein crambin (Protein Data Bank: 1CCN) is chosen, as depicted in Figure 3.5. The crambin molecule contains three disulfide bonds and displays β -strand, β -turn, and helical elements of protein secondary structure. The structure of crambin in water as determined by 2D-NMR spectroscopy [52] was taken as a starting point for the geometry optimization (in absence of the solvent). Crambin has 46 residues, 648 atoms and hence the full NMA would result in 1944 frequencies. It is our aim to reduce the number of frequencies drastically by performing an MBH analysis on the system, but still to preserve the essential lowest frequency normal modes.

The first block choice $[C_\alpha]^0$ considers the peptide groups (C=O-NH) as non-adjoined blocks, which is realizable with the previous MBH implementation. The square overlaps (not shown) revealed that the peptide units indeed move as nearly rigid groups, even when some of the residues are heavier and the peptide chain is longer than in the alanine dipeptide case.

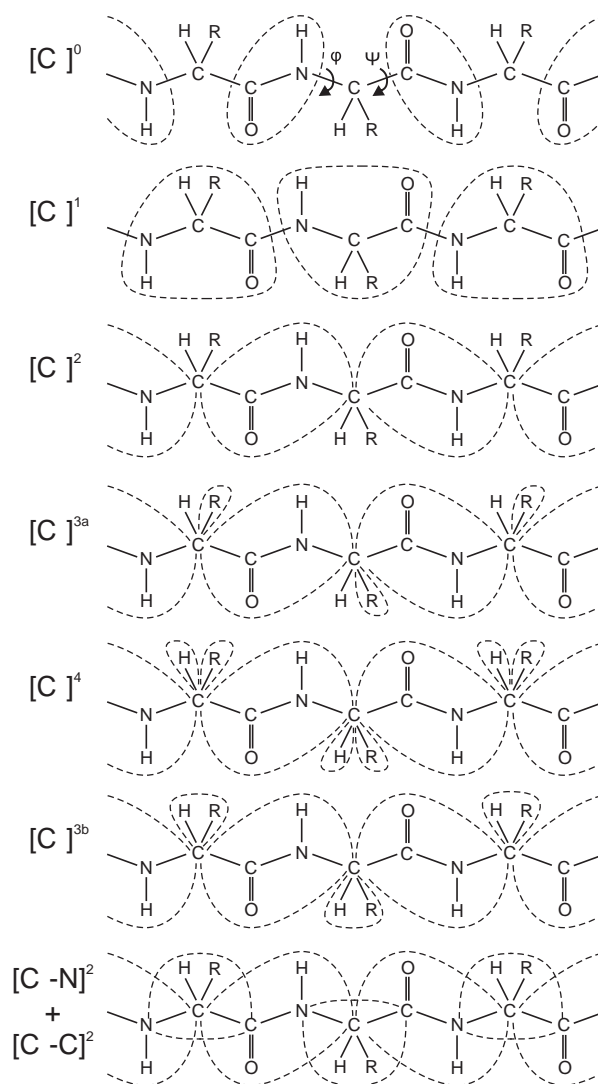


Figure 3.4: Partitioning schemes for peptide chains. Non-adjointed blocks: $[C_\alpha]^0$ and $[C_\alpha]^1$. Blocks adjoined by one adjoining atom: $[C_\alpha]^2$, $[C_\alpha]^{3a}$, $[C_\alpha]^4$, $[C_\alpha]^{3b}$. Blocks adjoined by two adjoining atoms: $[C_\alpha-N]^2 + [C_\alpha-C]^2$. Dihedral angles ϕ and ψ representing the two backbone degrees of freedom are also given in the top figure.

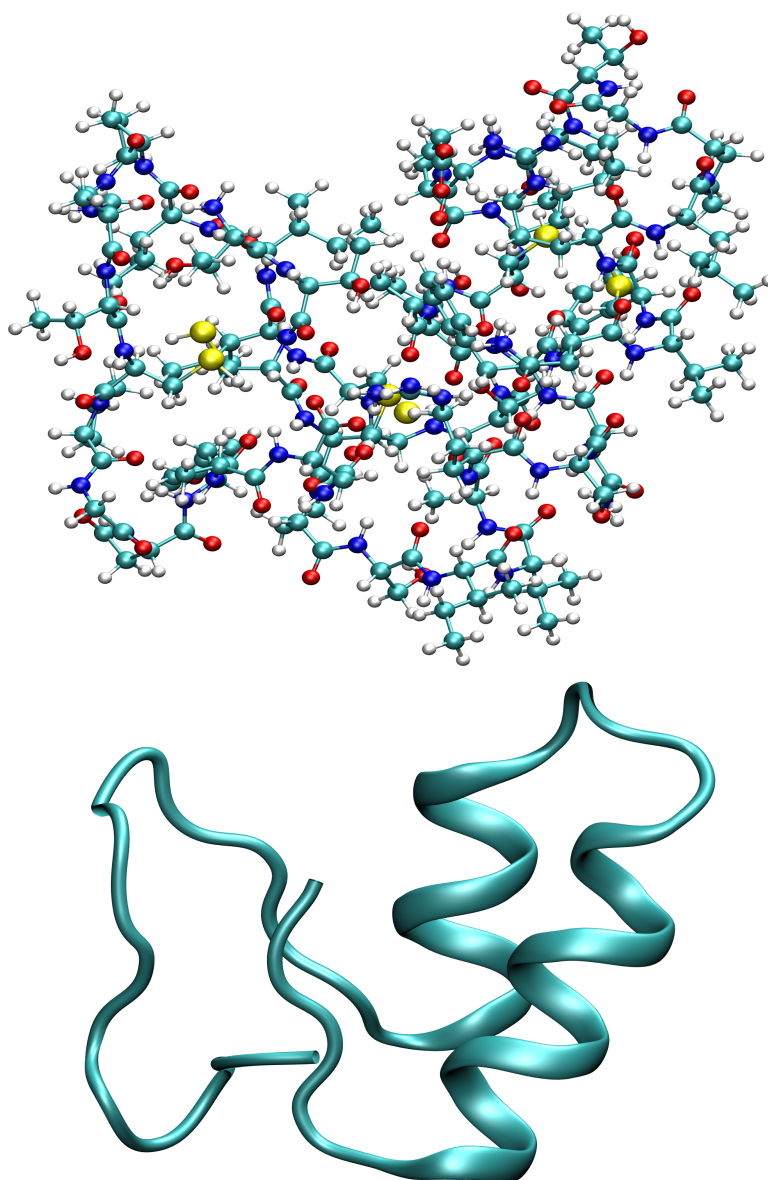


Figure 3.5: Crambin protein (Protein Data Bank: 1CCN), all-atom (top) and backbone (bottom) representation.

Can a block choice with adjoined blocks accurately describe the low frequency modes? To answer this question, three block choices from the list in Figure 3.4 are proposed, which gave already promising results in the alanine test case:

- (1) case $[C_\alpha]^{3b}$ groups the side chain and the H atom in a adjoined block,
- (2) case $[C_\alpha]^1$ is the scheme introduced by Tama et al.,
- (3) case $[C_\alpha-N]^2+[C_\alpha-C]^2$ uses the hinge type connection.

Figure 3.6 compares the low frequency part of the benchmark (full Hessian) spectrum with the MBH results. A grayscale indicates the square overlap between the modes. A dot on the diagonal would indicate an exact frequency estimation, and the darker the dot, the better the overlap. A huge amount of frequencies (1944) lies in a limited frequency interval – between 0 and 4000 cm^{-1} – and therefore the square overlaps seldom reach high values because of degeneracy. The cumulative square overlap P_j does not depend on the degeneracy as it sums over all MBH frequencies. Figure 3.7 plots the values P_j for the benchmark modes below 50 cm^{-1} . It indicates how well a benchmark mode is still represented by all MBH modes.

Block choice $[C_\alpha]^{3b}$ overestimates the frequencies because of the stiffening of the system, but the cumulative overlap shows that reproduction of the benchmark modes is very good in the low frequency range: all but one benchmark modes below 50 cm^{-1} are reproduced for at least 90% by the MBH modes, which is largely sufficient. Block choice $[C_\alpha]^{3b}$ differs from block choice $[C_\alpha]^4$ by the treatment of the H atom connected to the C_α carbon. Elimination of the C_α -H stretch without constraining the side chain- C_α -H angle, gives a somewhat more flexible model of the backbone and case $[C_\alpha]^4$ is expected to be a more accurate approximation. However, the effect is minimal in the lower spectrum and becomes only manifest in the range $> 250\text{ cm}^{-1}$ (not shown), because the H atom motions interact very little with the backbone modes. Hence, when one is interested in the lowest frequencies, the side chain- C_α -H angle bending is not essential and a partitioning as in case $[C_\alpha]^{3b}$, where all motions of the hydrogen atom are omitted, is sufficient.

The introduction of rigid blocks for the peptide chain implies a serious reduction of degrees of freedom. In general, a peptide chain with N_{res} residues can be partitioned according to the $[C_\alpha]^{3b}$ block choice in a chain of N_{res} peptide blocks connected by the C_α atoms and N_{res} side chain blocks adjoined to the chain by the C_α atoms as well. With three degrees of freedom per block, the total number becomes $6N_{res}$. E.g. for crambin the MBH method results in 276 modes which is 14% of the initial 1944 variables.

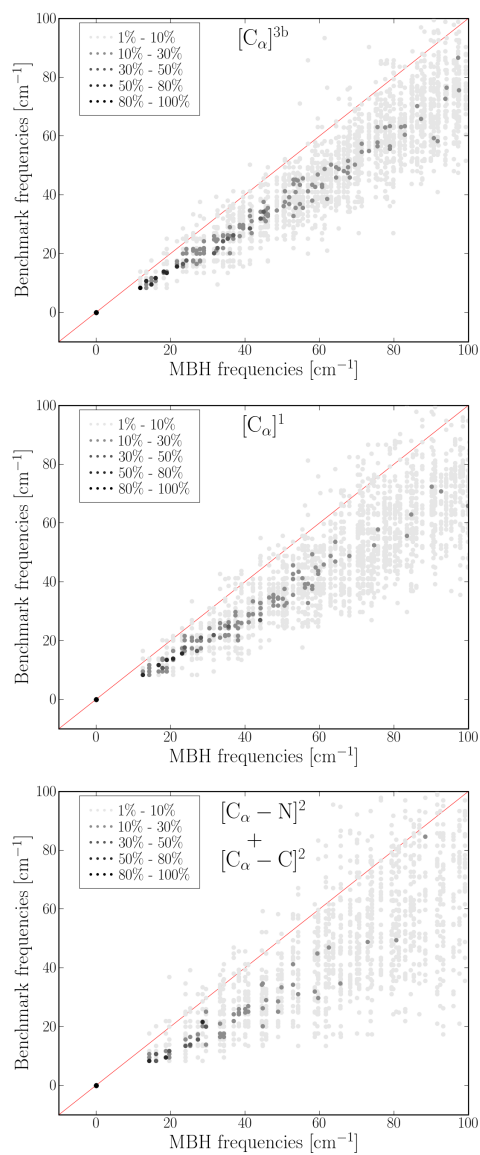


Figure 3.6: Crambin – Square overlap (in %) between MBH modes and benchmark (full Hessian) modes in the 0 – 100 cm^{-1} frequency range for block choices $[C_\alpha]^{3b}$, $[C_\alpha]^1$ and $[C_\alpha - N]^2 + [C_\alpha - C]^2$.

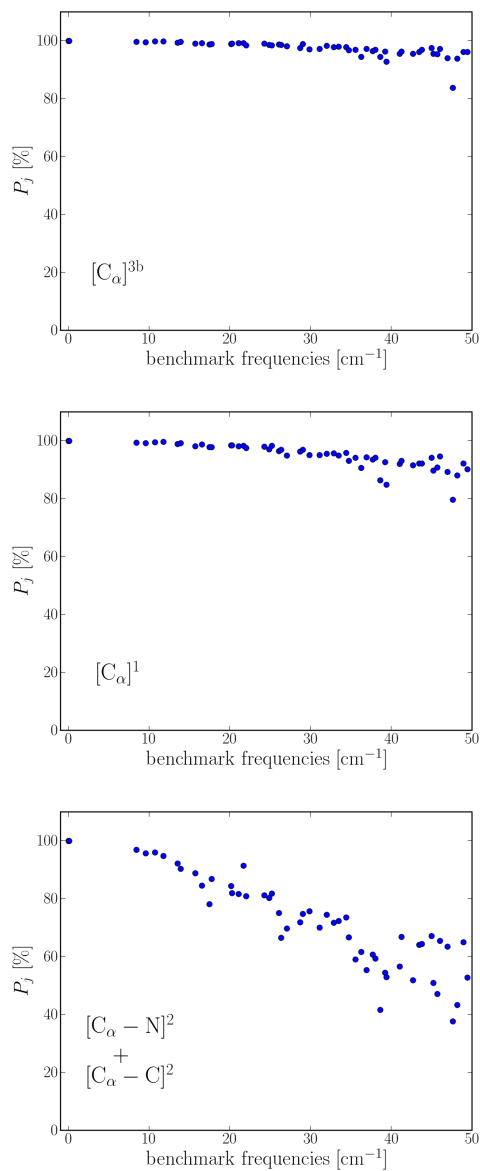


Figure 3.7: Crambin – Cumulative square overlap P_j (in %) for benchmark modes in the $0 - 50 \text{ cm}^{-1}$ range for block choices $[C_\alpha]^{3b}$, $[C_\alpha]^1$ and $[C_\alpha - N]^2 + [C_\alpha - C]^2$.

Comparison to RTB

A similar reduction was obtained by Tama et al. [15] with blocks consisting of several (n) peptide plus side chain units. The blocks are not linked to each other, leading to a total number of $6N_{res}/n$ degrees of freedom. For comparison, we reproduced the overlap results in case of one single peptide plus side chain unit per block ($n = 1$). This partitioning, block choice $[C_\alpha]^1$ in Figure 3.4, gives the same reduction of modes as the $[C_\alpha]^{3b}$ partitioning with adjoined blocks. The overlap and cumulative overlap of $[C_\alpha]^1$ also perform well, though slightly worse than the MBH approach with adjoined blocks. Thus for the same degree of simplification the introduction of adjoined blocks leads to a slight improvement of the partitioning proposed by Tama et al.

Tama et al. observed a linear relationship between the benchmark and the MBH frequencies, as explained in Section 3.1.2:

$$v_{mbh} = d \cdot v_{bench}. \quad (3.12)$$

The factor d is obtained by a linear least squares fit when comparing the lowest, say n_{low} , benchmark modes with the n_{low} lowest MBH modes. For cases $[C_\alpha]^{3b}$ and $[C_\alpha]^1$ we find $d = 1.41$ and 1.49 respectively, when the lowest modes (below 40 cm^{-1}) are taken into account.

The overestimation of the frequencies, i.e. $d > 1$, is due to the stiffening of the system. Applying the block constraints limits the motions to a hypersurface on the potential energy surface, which has higher curvatures than the total potential energy surface. In other words, the constraints make the system stiffer because the motion of blocks as a whole is more hindered than it would be if the atoms in the block were allowed to relax. On the other hand, the inertia of the blocks may lower the frequencies, because a block moving as a whole has more inertia than the atoms moving individually. The frequency change is thus governed by the balance between stiffening and inertia effects. The systematic frequency overestimation indicates that in the lower spectrum of peptide chains the stiffening is the dominating effect for the considered block choices.

Selecting dihedral angles

The number of modes can be further reduced by using the partitioning scheme with hinges. The dihedral basis $[C_\alpha-N]^2 + [C_\alpha-C]^2$ as illustrated in Figure 3.4 has been applied to crambin and reduces the number of modes to 96 which is as low as 4.9% of the initial number of variables. The results are again compared with the benchmark (full Hessian) frequencies in Figures 3.6–3.7.

The overlap plot indicates that less modes are reproduced and that the frequency overestimation increases: $d = 1.85$. Still the cumulative overlap P_j shows that 8 modes are reproduced by over 90%, while 20 modes are reproduced by over 80%. Visualization of the modes showed that the lowest MBH modes indeed represent large amplitude motions where large parts of the chain move as a whole. The cumulative overlap decays rapidly when the frequency increases. In practice, of course, one usually investigates only the motions in the lowest five frequency modes for the purpose of analyzing conformational changes. Hence it is clear that the simplification of block choice $[C_\alpha-N]^2+[C_\alpha-C]^2$ still captures the essentials of the lowest modes, and can be used as an analysis tool that focuses on the lower spectrum.

Conclusion

An important application of MBH with adjoined blocks is the calculation of the lower spectrum of macromolecules. Especially for proteins, the MBH provides an efficient method to focus on the modes involved in conformational changes. Several block choices were examined for their ability to reproduce the low frequencies and modes. Among the adjoined block schemes, the $[C_\alpha]^{3b}$ block choice, where peptide units and side chains are grouped into blocks with C_α as the adjoining atom, performs well in reproducing the modes of relevance while reducing the number of modes considerably. In the $[C_\alpha-N]^2+[C_\alpha-C]^2$ block choice, where only dihedral angles are retained as degrees of freedom, the reduction of the dimensionality is very large. The resulting modes are less accurate, but still reasonable. The MBH method with adjoined blocks therefore appears to be an appropriate analysis tool of the lower spectrum of peptide chains.

4 Conclusion and perspectives

4.1 Conclusion

In Chapter 2, the MBH equations were derived in terms of internal coordinates and block parameters. The four MBH schemes mentioned at the end of this chapter were implemented in Python, such that a thorough validation of the MBH with a variety of test cases could be performed. In Chapter 3, several features of the MBH were highlighted by means of three examples: ethanol, an aminophosphonate in solvent, and crambin. The main conclusion is that, with a suitable block choice, MBH can reproduce localized modes, reaction rate constants, and low frequency modes, even when the geometry is partially optimized.

The choice of the blocks is crucial for the functionality of MBH and depends on what the real study object is. For a chemical reaction, the block borders should keep a distance of at least one bond to the reactive center in order to reproduce accurate reaction kinetics. In peptide chains, the peptide bond can be treated as a rigid block, but the ϕ and ψ dihedral angles should be free variables in order to capture the modes involved in conformational change.

At first sight the MBH is similar to existing NMA methods, such as the PHVA, the RTB, and united atom models. However, the MBH differs from the PHVA because of the correct description of the mass as pointed out in Section 3.2. In addition, the MBH can be considered as an improved, more general version of the RTB. Whereas RTB is limited to nonlinear fully optimized blocks, the MBH blocks may have a non-optimized internal geometry. MBH also accommodates linear blocks and adjoined blocks, such that any topology of blocks can be treated with MBH. Finally, the MBH seems related to the united atom concept

in force fields, since groups of atoms are treated there also as a single entity [53]. However, in spite of this resemblance, the MBH is essentially different. In the MBH blocks each atom keeps its identity and continues to contribute individually to e.g. moments of inertia, Hessian elements, steric hindrance, etc. Coarse-grained or united atom methods reduce the number of atoms and the initial all-atom potential energy surface is approximated by a parameterized PES of lower dimension. The MBH on the other hand does not simplify the potential energy surface, but freezes certain degrees of freedom when performing the vibrational analysis.

Another recent NMA model which reduces the dimension of the Hessian, is the vibrational subsystem analysis (VSA) method, originally developed to probe vibrational free energies [17, 54]. A comparative study of various NMA methods based on a Hessian of reduced dimension, including MBH, PHVA and VSA, has been accomplished very recently [43] in collaboration with prof. B. R. Brooks of the Laboratory of Computational Biology (NIH) in Bethesda (Maryland). We found that PHVA is capable of reproducing localized modes and that MBH can reproduce both localized and global modes. VSA is most suited for the frequency/mode reproduction of the lower spectrum. In partially optimized structures, PHVA and MBH can still yield physical frequencies. Moreover, by varying the size of the blocks, MBH can be used as an analysis tool for identifying specific modes in the spectrum.

Concluding, the MBH is a newly developed NMA method which combines a good reproduction of frequencies/modes with a reasonable computational performance. With a block choice adapted to the application, most common simulation problems can be addressed with a focus on the lower spectrum or on the characteristic frequencies.

4.2 Perspectives

Critical note

Despite the success of MBH, it should be kept in mind that NMA remains an approximate method. Since it is limited to a quadratic description of the PES, it misses all information from the derivatives of higher order. Moreover, extended systems have a complex PES with multiple minima, while NMA is based on a single reference point. Therefore NMA (and MBH) should best be regarded as an analysis tool which is complementary to other simulation techniques such as principal component analysis (PCA), quasi-harmonic analysis, Monte Carlo or molecular dynamics.

Computational performance

MBH has the potential to become a widely used analysis tool. However, there is certainly room for implementation improvements in order to increase the computational performance. A first item is the partial geometry optimization, which is implemented efficiently in QM program packages such as Gaussian03, but not yet in MM code such as CHARMM. At present, CHARMM is able to constrain coordinates during the geometry optimization by means of penalty functions, which may cause numerical inaccuracy in case of a large number of restraints. A more favorable scheme would update the coordinates in each cycle of the energy minimization according to predefined basis vectors, namely the rotation/translation-vectors of the MBH blocks.

A second item is the direct calculation of second derivatives with respect to block parameters, as stated on p. 35, instead of calculating the complete Cartesian Hessian and reducing its size afterwards. For the calculation of derivatives based on finite differences, this requires the generation of a series of geometries where the atoms of a block are displaced as a whole by translation/rotation, rather than displacing every single atom of the block individually in three directions. In this way, the number of energy (or gradient) evaluations would decrease drastically. This approach is followed in the implementation that will be available in the next release of ADF.

When analytical MM derivatives are used, the appropriate derivatives of each energy contribution of the force field should be properly implemented. In fact, it is not the derivative calculation which is the limiting factor in force field simulations, but the storage and diagonalization of the Hessian. Instead of avoiding the use of Cartesian derivatives, it would be most interesting to implement an on-the-fly algorithm which stores the Cartesian Hessian elements only temporarily in memory. The block derivatives are constructed and stored on-the-fly, which circumvents the storage of the complete Cartesian Hessian. Note that such an implementation was available with the VIBBLOCK command in CHARMM (module 'BNM'); however its poor documentation makes the keyword impractical at present.

For the calculation of analytical QM second derivatives, the dimension of the coupled perturbed Hartree-Fock (CPHF) equations should be reduced. Instead of three equations for each atom, corresponding to the perturbations caused by the displacement in three directions, a new set of perturbations can be designed, i.e. the six rotations/translations of an MBH block. Also for QM/MM simulations, a similar approach should be followed, for the motion of blocks in the MM region as well as for those in the QM region. Currently such an implementation is in development in collaboration with the Laboratory

of Computational Biology (Bethesda, Maryland) for the Q-Chem/CHARMM interface.

Further extensions

To increase the applicability of MBH, it is desirable to develop an implementation of MBH (and all variants) in periodic codes. CP2K, a continuation of the Car-Parrinello Molecular Dynamics (CPMD) code, is more and more used for performing molecular simulations of extended biological systems in a solvent or simulations of crystals. The periodic boundary conditions cause specific difficulties in deriving normal modes with the MBH scheme. More precisely, the periodic assumption implies that a block in the reference unit cell is repeated in every image of the cell. The question should be asked whether such a block choice still yields physical frequencies. It is my aim to implement and validate the periodic version of MBH with a molecule in flexible solvent and in a more rigid porous structure. If successful, this extension of MBH can be used e.g. to calculate vibrational spectra of metal-organic frameworks (MOFs) of which the spectra are very complex and hard to interpret, and where the introduction of blocks should give considerable more insight.

Furthermore the block concept of the MBH could be combined with the PHVA concept or the VSA concept. Combining the MBH with the PHVA model, one could keep part of the system immobile, and at the same time introduce mobile blocks. This MBH-PHVA model is expected to be very suitable to describe an environment that has both rigid and flexible parts, e.g. a reaction in solvent taking place at a surface, or mobile guest molecules moving in a rigid porous structure. A hybrid MBH-VSA model combines the block concept of MBH with the subsystem/environment concept of VSA. After the introduction of blocks in the environment, one could integrate over the block parameters instead of over the Cartesian coordinates. It is expected that this model will be useful for the study of the coupling between local and global motions, for instance of proteins in a solvent.

Finally, it is also interesting to investigate the relation between MBH and the widely used ENM and other coarse-grained (CG) models. In principle, it must be possible to use MBH in the design of advanced ENM and CG models.

Future applications

In the near future, the current version of MBH, will be applied to a range of interesting molecular systems. A first example is the influenza virus

hemagglutinine, a protein where the lowest modes most probably represent domain motions, possibly linked to biological function. Another example is amylose, a polysaccharide with $\alpha(1 \rightarrow 4)$ linkage. The lowest modes of the helical conformation may lead to estimates of macroscopic mechanical properties, such as the elasticity modulus of Young. As already mentioned, extensions of the MBH will open the way to even more complex systems such as proteins in solvent and metal-organic frameworks.

Bibliography

- [1] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.*, 103(2):227–249, 1976.
- [2] X. Assfeld and J. L. Rivail. Quantum chemical computations on parts of large molecules: The ab initio local self consistent field method. *Chem. Phys. Lett.*, 263(1-2):100–106, 1996.
- [3] J. L. Gao, P. Amara, C. Alhambra, and M. J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102(24):4714–4721, 1998.
- [4] Y. K. Zhang, T. S. Lee, and W. T. Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.*, 110(1):46–54, 1999.
- [5] H. M. Senn and W. Thiel. QM/MM methods for biological systems. In *Atomistic Approaches in Modern Biology: from Quantum Chemistry to Molecular Simulations*, volume 268 of *Top. Curr. Chem.*, pages 173–290. Springer-Verlag Berlin, Berlin, 2007.
- [6] M. M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77(9):1905–1908, 1996.
- [7] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2(3):173–181, 1997.
- [8] K. Hinsen. Analysis of domain motions by approximate normal mode calculations. *Prot. Struct. Funct. Genet.*, 33(3):417–429, 1998.
- [9] S. Q. Jin and J. D. Head. Theoretical investigation of molecular water-adsorption on the Al(111) surface. *Surf. Sci.*, 318(1-2):204–216, 1994.

- [10] M. D. Calvin, J. D. Head, and S. Q. Jin. Theoretically modelling the water bilayer on the Al(111) surface using cluster calculations. *Surf. Sci.*, 345(1-2):161–172, 1996.
- [11] H. Li and J. H. Jensen. Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.*, 107:211–219, 2002.
- [12] N. A. Besley and K. A. Metcalf. Computation of the amide I band of polypeptides and proteins using a partial Hessian approach. *J. Chem. Phys.*, 126(3):035101, 2007.
- [13] A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier. Vibrational modes in partially optimized molecular systems. *J. Chem. Phys.*, 126(22):224102, 2007.
- [14] P. Durand, G. Trinquier, and Y. H. Sanejouand. New approach for determining low-frequency normal-modes in macromolecules. *Biopolymers*, 34(6):759–771, 1994.
- [15] F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct. Funct. Genet.*, 41(1):1–7, 2000.
- [16] A. Ghysels, V. Van Speybroeck, E. Pauwels, D. Van Neck, B. R. Brooks, and M. Waroquier. Mobile block Hessian approach with adjoined blocks: an efficient approach for the calculation of frequencies in macromolecules. *J. Chem. Theory Comput.*, 2009. In press.
- [17] H. L. Woodcock, W. J. Zheng, A. Ghysels, Y. H. Shao, J. Kong, and B. R. Brooks. Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. *J. Chem. Phys.*, 129(21):214109, 2008.
- [18] V. Van Speybroeck, D. Van Neck, and M. Waroquier. Ab initio study of radical addition reactions: addition of a primary ethylbenzene radical to ethene (I). *J. Phys. Chem. A*, 104(46):10939–10950, 2000.
- [19] V. Van Speybroeck, P. Vansteenkiste, D. Van Neck, and M. Waroquier. Why does the uncoupled hindered rotor model work well for the thermodynamics of n-alkanes? *Chem. Phys. Lett.*, 402(4-6):479–484, 2005.
- [20] P. Vansteenkiste, D. Van Neck, V. Van Speybroeck, and M. Waroquier. An extended hindered-rotor model with incorporation of Coriolis and vibrational-rotational coupling for calculation partition functions and derived quantities. *J. Chem. Phys.*, 124(4):044314, 2006.

- [21] I. Kolossvary and C. McMartin. On the degeneracy of the Hessian matrix. *J. Math. Chem.*, 9(4):359–367, 1992.
- [22] J. Goldstone, A. Salam, and S. Weinberg. Broken symmetries. *Phys. Rev.*, 127(3):965–970, 1962.
- [23] A. Tachibana and K. Fukui. Differential geometry of chemically reacting systems. *Theor. Chim. Acta*, 49(4):321–347, 1978.
- [24] David J. Wales. Potential energy surfaces and coordinate dependence. *J. Chem. Phys.*, 113:3926–3927, 2000.
- [25] R.L. Murry, J. T. Fourkas, Li Wu-Xiong, and T. Keyes. Molecular coordinates for instantaneous normal mode calculations. I Coordinate dependence. *J. Chem. Phys.*, 110:10410–10422, 1999.
- [26] J. D. Head. Computation of vibrational frequencies for adsorbates on surfaces. *Int. J. Quantum Chem.*, 65(5):827–838, 1997.
- [27] J. D. Head and Y. Shi. Characterization of Fermi resonances in adsorbate vibrational spectra using cluster calculations: Methoxy adsorption on Al(111) and Cu(111). *Int. J. Quantum Chem.*, 75(4-5):815–820, 1999.
- [28] J. D. Head. A vibrational analysis with Fermi resonances for methoxy adsorption on Cu(111) using ab initio cluster calculations. *Int. J. Quantum Chem.*, 77(1):350–357, 2000.
- [29] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.
- [30] B. Brooks and M. Karplus. Normal-modes for specific motions of macromolecules - application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. U.S.A.*, 82(15):4995–4999, 1985.
- [31] <http://www.charmm.org>.
- [32] G. Li and Q. Cui. A coarse-grained normal mode approach for macromolecules. *Biophys. J.*, 83:2457, 2002.
- [33] G. Li and Q. Cui. Analysis of functional motions in Brownian molecular machines with an efficient block normal mode approach: Myosin-II and Ca²⁺-ATPase. *Biophys. J.*, 86:743, 2004.

- [34] W. F. Vangunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.*, 34(5):1311–1327, 1977.
- [35] A. Ghysels, D. Van Neck, and M. Waroquier. Cartesian formulation of the mobile block Hessian approach to vibrational analysis in partially optimized systems. *J. Chem. Phys.*, 127(16):164108, 2007.
- [36] <http://www.python.org>.
- [37] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [38] <http://cp2k.berlios.de>.
- [39] <http://www.scm.com>.
- [40] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O’Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. Advances in methods and algorithms in a

- modern quantum chemistry program package. *Phys. Chem. Chem. Phys.*, 8(27):3172–3191, 2006.
- [41] A. Ghysels, D. Van Neck, V. Van Speybroeck, B. R. Brooks, and M. Waroquier. Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach. *J. Chem. Phys.*, 130:084107, 2009.
- [42] A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier. Calculating reaction rates with partial Hessians: Validation of the mobile block Hessian approach. *J. Chem. Theory Comput.*, 4(4):614–625, 2008.
- [43] A. Ghysels, Brooks B. R., V. Van Speybroeck, D. Van Neck, and M. Waroquier. Comparative study of various normal mode analysis techniques based on partial Hessians. *J. Chem. Theory Comput.*, 2009.
- [44] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3:107, 1935.
- [45] M. G. Evans and M. Polanyi. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.*, 31:875, 1935.
- [46] K. J. Laidler. *Chemical Kinetics*. Harper Collins Publishers, Inc., 1987.
- [47] D. A. Mc Quarrie and J. D. Simon. *Physical Chemistry - a molecular approach*. University Science Books, Sausalito, California, 1997.
- [48] J. L. Gao and D. G. Truhlar. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.*, 53:467–505, 2002.
- [49] V. Van Speybroeck, K. Van Cauter, B. Coussens, and M. Waroquier. Ab initio study of free-radical polymerizations: Cost-effective methods to determine the reaction rates. *ChemPhysChem*, 6(1):180–189, 2005.
- [50] M. K. Sabbe, M. Saeys, M. F. Reyniers, G. B. Marin, V. Van Speybroeck, and M. Waroquier. Group additive values for the gas phase standard enthalpy of formation of hydrocarbons and hydrocarbon radicals. *J. Phys. Chem. A*, 109(33):7466–7480, 2005.
- [51] M. K. Sabbe, A. G. Vandeputte, M. F. O. Reyniers, V. Van Speybroeck, M. Waroquier, and G. B. Marin. Ab initio thermochemistry and kinetics for carbon-centered radical addition and beta-scission reactions. *J. Phys. Chem. A*, 111(34):8416–8428, 2007.

- [52] A. M. J. J. Bonvin, R. Boelens, and R. Kaptein. Direct nuclear overhauser effect refinement of crambin from 2-dimensional nmr data using a slow-cooling annealing protocol. *Biopolymers*, 34(1):39–50, 1994.
- [53] L. J. Yang, C. H. Tan, M. J. Hsieh, J. M. Wang, Y. Duan, P. Cieplak, J. Caldwell, P. A. Kollman, and R. Luo. New-generation Amber united-atom force field. *J. Phys. Chem. B*, 110(26):13166–13176, 2006.
- [54] W. J. Zheng and B. R. Brooks. Probing the local dynamics of nucleotide-binding pocket coupled to the global dynamics: Myosin versus kinesin. *Biophys. J.*, 89(1):167–178, 2005.

Appendix

Papers

Paper I

Vibrational modes in partially optimized molecular systems

A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck and
M. Waroquier

Journal of Chemical Physics 126 (22), 224102 (2007)

Copyright 2007 by American Institute of Physics

Vibrational modes in partially optimized molecular systems

A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier^{a)}
*Center for Molecular Modeling, Laboratory of Theoretical Physics, Ghent University, Proeftuinstraat 86,
B-9000 Gent, Belgium*

(Received 27 February 2007; accepted 12 April 2007; published online 12 June 2007)

In this paper the authors develop a method to accurately calculate localized vibrational modes for partially optimized molecular structures or for structures containing link atoms. The method avoids artificially introduced imaginary frequencies and keeps track of the invariance under global translations and rotations. Only a subblock of the Hessian matrix has to be constructed and diagonalized, leading to a serious reduction of the computational time for the frequency analysis. The mobile block Hessian approach (MBH) proposed in this work can be regarded as an extension of the partial Hessian vibrational analysis approach proposed by Head [Int. J. Quantum Chem. **65**, 827 (1997)]. Instead of giving the nonoptimized region of the system an infinite mass, it is allowed to move as a rigid body with respect to the optimized region of the system. The MBH approach is then extended to the case where several parts of the molecule can move as independent multiple rigid blocks in combination with single atoms. The merits of both models are extensively tested on ethanol and di-*n*-octyl-ether. © 2007 American Institute of Physics.
[DOI: [10.1063/1.2737444](https://doi.org/10.1063/1.2737444)]

I. INTRODUCTION

The applications of molecular modeling nowadays focus more and more on extended systems, in which numerous atoms are involved. Examples are polymer chains, supramolecular assemblies, systems embedded in a solvent, or (macro)molecules adsorbed within porous materials. For the description of the electronic part of the system, one usually resorts to a hybrid model, in which the chemically active part where bonds may be formed or broken is described at a quantum mechanical level, whereas the outer region is described at a lower molecular mechanics level. The previous models are often referred to as quantum mechanical/molecular mechanical methods.¹⁻⁴ The basic idea originated from the observation that normally only a portion of the atoms in a reaction is directly involved in bond breaking and forming or in changing of bond order. These methods can be used either in a cluster or a periodic based approach.

In many cases, only part of the system is optimized to restrict the computational cost or to prevent unphysical deformations of the border of periodic systems. Good examples are reactions occurring in porous materials, where the border of the system is kept fixed during the geometry optimization to prevent unphysical deformations due to the neglect of the whole periodic structure. Another example is the simulation of defects embedded in a crystal lattice. In such cases, it is common practice to cut out a cluster around the defect and to keep the cluster border atoms fixed during the geometry optimization.⁵ After such a constrained geometry optimization, residual forces remain on the fixed border atoms and the partially optimized structure corresponds to a nonequilibrium state. The usual normal mode analysis (NMA) equations can

be applied to this nonequilibrium configuration using the full Hessian of all atoms in the structure, i.e., the matrix of second derivatives of the potential energy surface with respect to all nuclear coordinates. Such a procedure, however, has some serious defects. The Hessian will have only three eigenvalues equal to zero instead of six, implying that the rotational invariance of the potential surface is not manifest. In addition, spurious imaginary normal frequencies appear, suggesting that the partially optimized system resides in a transition state, even when this is obviously not the case.

Whereas for the determination of the energy the partitioning into chemically active and passive areas is common practice, it is far less applied for the determination of the Hessian of such extended systems. The calculation of the Hessian is one of the most expensive steps in the calculation of free energies, so a partitioning scheme for the Hessian would also seriously reduce the computational cost.

Within this respect, the work of Head and co-workers is especially interesting.^{6,7} They introduced a strategy by which only the frequencies of part of a chemical system are computed. In 2002, Li and Jensen introduced the name partial Hessian vibrational analysis (PHVA) and extended the method for the calculation of vibrational enthalpy and entropy changes for chemical reactions.⁸ Recently, Besley and Metcalf applied the partial Hessian approximation to calculate the amide I band of polypeptides and proteins.⁹ Within this methodology, however, the normal modes are calculated for the system with the fixed atoms frozen at their reference positions as if they were given an infinite mass and only the relaxed atoms can participate in the small amplitude vibrations. Head and co-workers have also developed a more sophisticated partial Hessian method,¹⁰⁻¹² where frequencies are corrected in lowest order perturbation theory for the cou-

^{a)}Author to whom correspondence should be addressed. Electronic mail: michel.waroquier@ugent.be

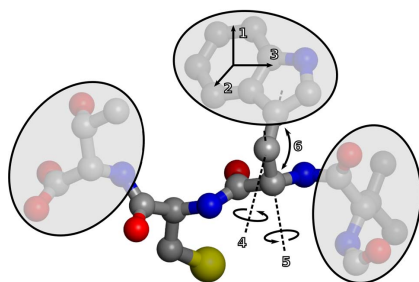


FIG. 1. Schematic representation of the basic idea behind the MBH method. The shaded blocks symbolize the parts of the molecule of which the internal geometry is kept fixed during the partial geometry optimization. In the MBH approach, they are described as rigid bodies with six degrees of freedom (translations and rotations).

pling between PHVA modes and modes in the fixed part. This method has been very useful for the calculation of localized vibrations of adsorbates on surfaces.

In this paper we propose an extended version of the PHVA method, in which the atoms that were kept fixed during the optimization can participate in small amplitude vibrations of the system, with the restriction that only coherent movements as a single block are allowed. The block of the fixed atoms can rotate and translate as a rigid body, while the internal geometry of the block is kept fixed. This model is referred hereafter to as the mobile block Hessian (MBH) approach. A schematic representation of the basic idea behind the MBH method is given in Fig. 1.

It can be expected that both methods would give similar results in those cases where the nonoptimized part of the system is quite rigid. The example of a defect in a crystal lattice is a case in point. For systems in which the molecular environment is more flexible, such as reactions occurring in solvents or polymer chains, the motions of the surroundings cannot *a priori* be neglected. It needs to be investigated to what extent the two methods coincide for such applications and whether such a partial optimization can give an accurate description of the localized modes in the optimized region. Localized modes are characterized by the fact that during the small amplitude vibrations, changes in the geometry only occur in a restricted region of the molecular system. In order to investigate this question, one should be able to construct the normal modes and frequencies of the partially optimized system in a rigorous way and compare them with frequencies of the fully optimized system.

The computational cost of the PHVA and MBH methods is similar, and the NMA equations can be rephrased in terms of a Hessian of reduced dimension in both cases. This leads to a significantly reduced load compared to a full Hessian frequency calculation. Related with a less expensive treatment of the Hessian is the work of Lin *et al.*¹³ They have proposed efficient methods for calculating the Hessian in the optimization procedure of the multiconfigurational molecular mechanics method. The partial Hessian vibrational analysis and also the mobile block Hessian approach could be useful

in combination with such methods as they could additionally speed up the computationally expensive task of determining the Hessian for reactions on extended systems.

In Sec. II the problem of determining normal modes in nonequilibrium systems is discussed. The partial Hessian vibrational analysis is revised, and a detailed theoretical derivation is given of the mobile block Hessian approach, which we propose as an extension of the PHVA method for extended systems which are quite flexible. In Sec. III the two methods are compared for the example of the ethanol molecule. This very simple molecule has been chosen as a test case, as it enables one to address various phenomenological issues of the PHVA and MBH methods. The MBH approach can be easily extended to the case where several parts of the molecule can move as independent rigid bodies in combination with single atoms. The framework of this “multiple” mobile block Hessian approach is worked out in Sec. IV and is applied to di-*n*-octyl-ether in Sec. V. Finally, the results are summarized in Sec. VI, and future applications of the MBH model are discussed.

II. BACKGROUND AND THEORETICAL DEVELOPMENT

A. Normal modes in nonequilibrium configurations

Consider a molecule with N masses m_A ($A=1, \dots, N$), the positions of which are described by Cartesian coordinates $\mathbf{r}_A = \{r_{A\mu}\}_{\mu=x,y,z}$, with respect to a space-fixed frame. The energy of the system reads

$$E = \frac{1}{2} \sum_{A\mu} m_A \dot{r}_{A\mu}^2 + V(\{\mathbf{r}_A\}), \quad (1)$$

where $\dot{r}_{A\mu}$ is a time derivative and V is the potential energy. Expanding V around a reference configuration $\{r_{A\mu}^0\}$ gives

$$V(\{\mathbf{r}_A\}) = V_0 + \sum_{A\mu} \left(\frac{\partial V}{\partial r_{A\mu}} \right)_0 \Delta_{A\mu} + \frac{1}{2} \sum_{A\mu, B\nu} \left(\frac{\partial^2 V}{\partial r_{A\mu} \partial r_{B\nu}} \right)_0 \Delta_{A\mu} \Delta_{B\nu} + \dots \quad (2)$$

in terms of the displacement coordinates $\Delta_{A\mu} = r_{A\mu} - r_{A\mu}^0$. By collecting the coordinates in one vector $x_i = r_{A\mu}$, with $i \equiv A\mu = 1, \dots, 3N$, one can expand the energy in matrix form as

$$E - V_0 = \frac{1}{2} \Delta^T M \Delta + \Delta^T G + \frac{1}{2} \Delta^T H \Delta \quad (3)$$

up to second order in the displacements $\Delta_i = x_i - x_i^0$. In Eq. (3), the mass matrix M is a diagonal matrix containing the masses, G is the gradient vector defined by the gradients of the potential energy at the reference point, and H is the Hessian or second-derivative matrix in the reference point. The NMA equation determining the eigenmodes v and normal frequencies $\lambda^{1/2}$ reads

$$Hv = \lambda Mv, \quad (4)$$

representing a generalized eigenvalue problem.

Solution schemes of these equations are implemented in various standard *ab initio* molecular modeling packages, but if the whole molecular system is not in its equilibrium state, the reference configuration is not gradient-free and some of the resulting frequencies are completely unphysical. An additional problem is that a discrimination of the unphysical frequencies from the physical values is not obvious.

We want to make two comments in the situation of non-equilibrium configurations (gradient vector $G \neq 0$). First, one can show that the eigenvalues of Eq. (4) are coordinate-dependent: a second-order expansion of the potential energy expressed in curvilinear coordinates, which are nonlinearly related to the Cartesian coordinates, leads to different normal mode frequencies. This phenomenon is “well known” and is related to the difference between ordinary and covariant derivatives in a nontrivial metric space.¹⁴ Its importance and unpleasant consequences have recently been emphasized by several authors.^{15,16}

A second comment deals with the invariance of the potential energy surface under overall rotations. When $G \neq 0$, the Cartesian Hessian will only generate three zero eigenvalues (those related to the global translation). The three eigenvalues associated with the global rotation are different from zero,¹⁷ and the absence of these Goldstone modes¹⁸ is due to the use of the second-order expansion in rectilinear Cartesian coordinates, which breaks the rotational symmetry of V . This defect is simply cured by taking coordinates that respect the symmetries of V , i.e., internal coordinates.

With any choice of $3N-6$ internal coordinates $\{\theta_j\}$, and a body frame whose origin lies at the center of mass $\mathbf{r}_{\text{c.m.}}$, the energy in Eq. (1) can be rewritten in a standard way¹⁹⁻²¹ as the sum of the potential energy $W(\{\theta_j\})$ expressed in internal coordinates and kinetic energy terms arising from the center-of-mass motion, global rotation, Coriolis coupling, and internal motions:

$$E = \frac{1}{2} \mathcal{M} \dot{\mathbf{r}}_{\text{c.m.}}^2 + \frac{1}{2} \boldsymbol{\omega}^T \bar{\mathbf{I}} \boldsymbol{\omega} + \boldsymbol{\omega} \sum_I \mathbf{A}_I \dot{\theta}_I + \frac{1}{2} \sum_{IJ} B_{IJ} \dot{\theta}_I \dot{\theta}_J + W(\{\theta_j\}). \quad (5)$$

Here \mathcal{M} is the total mass and $\boldsymbol{\omega}$ the angular velocity vector of the body frame. The inertial tensor $\bar{\mathbf{I}}$, the Coriolis coupling \mathbf{A}_I between the body-frame rotation and internal velocity $\dot{\theta}_I$, and the B_{IJ} matrix are all functions of the internal coordinates.

The potential energy $W(\{\theta_j\})$, expanded to second order about a reference geometry $\{\theta_j^0\}$, reads

$$W(\{\theta_j\}) = V_0 + \sum_I \left(\frac{\partial W}{\partial \theta_I} \right)_0 \Delta_I + \frac{1}{2} \sum_{IJ} \left(\frac{\partial^2 W}{\partial \theta_I \partial \theta_J} \right)_0 \Delta_I \Delta_J. \quad (6)$$

The set of new coordinates now consists of the center-of-mass coordinates $\mathbf{r}_{\text{c.m.}}$, three angles to specify the orientation of the body frame (e.g., Euler angles), and the $3N-6$ internal coordinates θ_j . The NMA equations using internal coordinates show a block structure:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & H^{(ii)} \end{pmatrix} \begin{pmatrix} \mathbf{v}^{(c)} \\ \mathbf{v}^{(r)} \\ \mathbf{v}^{(i)} \end{pmatrix} = \lambda \begin{pmatrix} M^{(cc)} & 0 & 0 \\ 0 & M^{(rr)} & M^{(ri)} \\ 0 & (M^{(ri)})^T & M^{(ii)} \end{pmatrix} \begin{pmatrix} \mathbf{v}^{(c)} \\ \mathbf{v}^{(r)} \\ \mathbf{v}^{(i)} \end{pmatrix}, \quad (7)$$

where $\mathbf{v}^{(c)}$, $\mathbf{v}^{(r)}$, and $\mathbf{v}^{(i)}$ have dimensions 3, 3, and $3N-6$, respectively, and the mass matrix entries read $M_{\mu\nu}^{(cc)} = \mathcal{M} \delta_{\mu\nu}$, $M_{\mu\nu}^{(rr)} = I_{\mu\mu}(\{\theta_j^0\})$, $M_{\mu\nu}^{(ri)} = A_{\mu i}(\{\theta_j^0\})$, and $M_{ij}^{(ii)} = B_{ij}(\{\theta_j^0\})$. If the gradient is nonzero, the NMA equations in Eq. (7) will give different eigenvalues from those of Eq. (4). In particular, the eigenvalue equations (7) always generate six zero eigenvalues due to its construction, as the Hessian has vanishing matrix elements in the (c) and (r) subspaces, while in Eq. (4) the presence of six zero eigenvalues is not ensured.

In a completely general situation it is, in principle, not possible, due to the coordinate dependence, to define meaningful normal modes in a nonequilibrium point. However, the situations of interest for the present paper are not completely general but arise from physical considerations.

We consider cases where the reference point is obtained by optimizing the energy with respect to a *subset* of coordinates, keeping the remainder fixed during the optimization. The fixed coordinates correspond, e.g., to the geometry of a part of the molecule that is expected to influence the localized mode of interest only slightly. The system corresponding to the subset of coordinates that have been optimized is actually in equilibrium, even though the complete gradient is nonzero.

A detailed theoretical analysis of the coordinate (in)dependence and symmetry properties of the NMA equations in nonequilibrium systems, as well as the link between descriptions using Cartesian and internal coordinates, will be given in a separate publication. In this paper we focus on the comparison of two practical methods for treating such cases.

B. Partial Hessian vibrational analysis

One assumes a certain reference structure $\{\mathbf{r}_{A_F}^0\}$ for part of the molecule consisting of N_F atoms (labeled A_F, B_F, \dots). Keeping $\{\mathbf{r}_{A_F}^0\}$ fixed in space, the positions of the remaining $N_E = N - N_F$ atoms (labeled A_E, B_E, \dots) are optimized, resulting in an “equilibrium” configuration $\{\mathbf{r}_{A_E}^0\}$. Obviously, the full gradient is nonzero, since $(\partial V / \partial \mathbf{r}_{A_E})_0 = 0$, but $(\partial V / \partial \mathbf{r}_{A_F})_0 \neq 0$. Solutions of the NMA equations taking into account the full Hessian and mass matrix will therefore lead to unphysical results.

In a first model, known as partial Hessian vibrational analysis, one assumes that the fixed atoms A_F do not participate in the small amplitude vibrations. Hence their displacements $\Delta_{A_F\mu}$ and velocities $\dot{\Delta}_{A_F\mu}$ are set to zero. This is consistent with a situation in which infinite masses m_{A_F} are associated with the fixed atoms. With this assumption, the second-order energy of Eq. (3) reduces to

$$\begin{aligned}
E - V_0 &= \frac{1}{2} \sum_{A_E \mu} m_{A_E} \dot{\Delta}_{A_E \mu}^2 \\
&+ \frac{1}{2} \sum_{A_E \mu; B_E \nu} \left(\frac{\partial^2 V}{\partial \Delta_{A_E \mu} \partial \Delta_{B_E \nu}} \right)_0 \Delta_{A_E \mu} \Delta_{B_E \nu} + \dots \\
&= \frac{1}{2} \dot{\Delta}'^T M' \dot{\Delta}' + \frac{1}{2} \Delta'^T H' \Delta', \quad (8)
\end{aligned}$$

where Δ' , $\dot{\Delta}'$, M' , and H' are the displacements, velocities, mass matrix, and Hessian restricted to the reduced $3N_E$ -dimensional subsystem of the nonfixed atoms with mass m_{A_E} . Note that the reduced gradient ($G'=0$) vanishes in this subspace, so the $3N_E$ system is in equilibrium at the reference configuration and provides coordinate-independent normal modes.

In practice, one simply needs to disregard the rows and columns related to the coordinates of the fixed atoms in the full (Cartesian) Hessian and mass matrix. The NMA equation reduces to a $3N_E \times 3N_E$ generalized eigenvalue problem:

$$\sum_{B_E \nu} \left(\frac{\partial^2 V}{\partial r_{A_E \mu} \partial r_{B_E \nu}} \right)_0 v_{B_E \nu} = \lambda m_{A_E} v_{A_E \mu}. \quad (9)$$

The reduced mass matrix M' is positive definite, and the presence of negative solutions λ in Eq. (9) gives indication that the reduced system resides in a transition state rather than a minimum. The number of zero eigenvalues depends on the remaining symmetry of the reduced system. Usually no zeros will occur, since the fixed atoms act as an external field breaking translational and rotational invariance for the nonfixed atoms.

C. The mobile block Hessian approach

In the MBH approach, the fixed atoms are allowed to participate in the small amplitude vibrations by moving as a rigid body (block). This is a physically different situation from the previous approach, since one now takes into account the finite masses of the fixed atoms $\{m_{A_F}\}$.

This approach is easily implemented by a suitable choice of the $3N-6$ internal coordinates $\{\theta_i\}$, which is always possible. In the Z -matrix formalism, for instance, one can determine the first N_F atoms by $3N_F-6$ Z coordinates (distances, angles, and dihedral angles), and the next N_E atoms can consecutively be described by $3N_E$ Z coordinates.

The $3N_F-6$ internal coordinates, describing the geometry of the fixed atoms A_F , are labeled $\{\theta_{i_F}\}$, and the remaining $3N_E$ internal coordinates are labeled $\{\theta_{i_E}\}$. The imposed reference structure $\{\mathbf{r}_{A_F}^0\}$ of the fixed atoms then determines the values of the $3N_F-6$ internal coordinates, $\{\theta_{i_F}^0\}$. Optimizing the energy with fixed $\{\theta_{i_F}^0\}$ and varying $\{\theta_{i_E}\}$ give the reference structure $\{\theta_{i_E}^0\}$. Again, the full gradient is nonzero, since $(\partial W / \partial \theta_{i_E})_0 = 0$ but $(\partial W / \partial \theta_{i_F})_0 \neq 0$. As a result, the NMA equation (7) with the full Hessian will not give correct frequencies (though one does have six zero eigenvalues).

Allowing the fixed block to move as a rigid body, keeping its internal geometry, can be imposed by putting the displacements Δ_{i_F} and corresponding velocities $\dot{\Delta}_{i_F}$ equal to

zero in the second-order expansion of Eq. (5). The part of the molecule described by the $3N_E$ internal coordinates $\{\theta_{i_E}\}$ has been relaxed during the geometry optimization and hence is in equilibrium. The gradient term in Eq. (5) thereby vanishes.

The corresponding NMA equations are obtained by omitting the rows and columns related to the $\{\theta_{i_F}\}$ variables from the Hessian and mass matrix in Eq. (7). The resulting reduced eigenvalue problem of dimension $3N_E+6$ still has six zero eigenvalues corresponding to overall translation and rotation. These can be decoupled in the usual way¹⁹⁻²¹ by a congruent transformation, $\tilde{\mathbf{v}}^{(c)} = \mathbf{v}^{(c)}$, $\tilde{\mathbf{v}}^{(r)} = \mathbf{v}^{(r)}$, and $\tilde{\mathbf{v}}^{(r')} = \mathbf{v}^{(r')} + (M^{(rr)})^{-1} M^{(ri)} \mathbf{v}^{(i)}$, yielding the final $3N_E$ -dimensional NMA equation,

$$H^{(ii)} \tilde{\mathbf{v}}^{(i)} = \lambda [M^{(ii)} - (M^{(ri)})^T (M^{(rr)})^{-1} M^{(ri)}] \tilde{\mathbf{v}}^{(i)}. \quad (10)$$

The primed matrices do not contain the components related to $\{\theta_{i_F}\}$. The transformed mass matrix in the right hand side of Eq. (10) is positive definite and the presence of negative eigenvalues λ will now unambiguously indicate that the reduced system is in a transition state.

D. Discussion: PHVA versus MBH

Both PHVA and MBH models offer a vibrational analysis in which only a subblock of the Hessian matrix is diagonalized to produce vibrational frequencies for partially optimized systems. In both descriptions, the molecular system is composed of a rigid body with a number N_F of fixed atoms and N_E atoms that are free to relax in a partial geometry optimization. Conceptually, the difference between PHVA and MBH lies in the use of near-infinite masses in the former approach for the atoms in the rigid body, while in the MBH model, reduced masses are used.

A more transparent and fundamental comparative study between the PHVA and MBH models can be made when using a specific choice of internal coordinates: for the set of $3N_E$ internal coordinates $\{\theta_{i_E}\}$ describing the geometry of the nonfixed atoms, one can take the Cartesian coordinates $\{\mathbf{r}_{A_E}^{\text{RB}}\}$ in a frame attached to the rigid block (RB) (e.g., the frame constructed by the principal axes of the fixed atoms). The potential energy then becomes

$$W(\{\theta_{i_F}, \theta_{i_E}\}) \equiv V(\{\mathbf{r}_{A_F}^{\text{RB}}, \mathbf{r}_{A_E}^{\text{RB}}\}). \quad (11)$$

Since the $\{\mathbf{r}_{A_E}^{\text{RB}}\}$ are constant in the fixed geometry of the block, one has

$$\left(\frac{\partial^2 W}{\partial \theta_{i_E} \partial \theta_{j_E}} \right)_0 \equiv \left(\frac{\partial^2 V}{\partial r_{A_E \mu} \partial r_{A_E \nu}} \right)_0. \quad (12)$$

Here it is assumed that at the reference point the frame attached to the rigid body coincides with the space-fixed frame, which can always be done. The reduced Hessian $H^{(ii)}$ of MBH now exactly coincides with the Cartesian H' of the PHVA. This gives evidence that the discrepancy between PHVA and MBH only lies in a different approach of handling the mass matrix.

The difference between both approaches is best illustrated by the example of two masses m_1 and m_2 moving in one dimension and joined by a spring characterized by a

spring constant K . Following the method of Sec. II B (PHVA), the mass m_1 is considered fixed in space and the normal frequency of the system (vibration of mass m_2) is given by $\omega^2 = K/m_2$. In the method of Sec. II C (MBH), on the other hand, the mass m_1 is allowed to participate in the vibration and one simply has the normal frequency of the free spring, $\omega^2 = K/\mu$, with μ the reduced mass. In the limit of $m_1 \gg m_2$, the reduced mass tends to m_2 and the eigenvalues of the two approaches converge to the same value.

Generally, the partial Hessian vibrational analysis model can be considered as a limit case of the mobile block Hessian approach. When the mass and the inertial moments of the rigid body increase with respect to the N_E free masses, the mass matrix of the MBH in the limit becomes equal to the mass matrix of the PHVA. Thus the higher the fixed atom masses, the more the two models converge to each other.

A more stringent comparison is obtained by inspecting the mass term taken up in Eq. (10). Following the various mass definitions given in Sec. II A, one sees that the fixed atom masses m_{A_f} only occur in the matrix $M^{(rr)}$, which coincides with the inertial matrix, while $M^{(ri)}$ is independent of the m_{A_f} . The larger the number of atoms in the fixed part of the molecule, the more the total mass of the rigid body increases. As mentioned above, only the elements of the matrix $M^{(rr)}$ are influenced, and the second term in the right hand side of Eq. (10) will tend to zero. With the specific choice of internal coordinates as stated in Eq. (11), it is clear that the PHVA frequencies form the limits of the MBH frequencies.

Another (non-negligible) aspect concerns the computational cost of both models. Since both models require the numerical evaluation of the reduced Hessian H' or $H'^{(ii)}$, no difference is expected in computing time. The calculation of the corresponding mass matrices is also straightforward. Since most of the required computation time goes to the evaluation of the second derivatives of the potential energy surface, the reduction of the dimension of the full Hessian to $3N_E$ is essential to get a serious reduction of the computation time.

The applicability of both methods depends on the type of the simulated system under consideration. It can be expected that the MBH method is better suited to describe molecules in the gas phase, because they move quasifreely in space and can be considered as isolated systems. Assigning an infinite mass to the nonoptimized part of the system may be too crude an approximation. The potential energy surface should be invariant under global translation and rotation, and these global modes should effectively decouple from the internal vibrations.

However, when modeling a lattice using the cluster *in vacuo* approach and fixing the border atoms, the situation is different. The cluster cannot move freely and there is no reason why the description of the cluster should be translationally or rotationally invariant. In fact, the border atoms are more or less pinned down at their positions by the presence of the surrounding infinite lattice, which was, however, left out of the simulation cluster. So it is quite a realistic picture to assume that the border atoms are really fixed by external

forces generated by the surrounding lattice. In other words, the rigid body of fixed border atoms may be assumed to have a fixed position and orientation, and thus, the PHVA approach is the appropriate way to calculate eigenvalues.

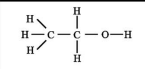

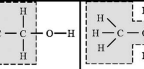

In what follows, both the PHVA and MBH methods are numerically validated and benchmarked against the normal modes extracted from a fully optimized geometry. Independent of the specific application under consideration, some general remarks are useful. Inherent to the PHVA and MBH methods is the introduction of a rigid part of the system that is not optimized, and thus, the number of PHVA and MBH frequencies is always smaller than the number of benchmark frequencies. The benchmark modes generated by atoms belonging to the rigid body are *a priori* not reproduced. Additionally, modes where the displacements are spread out over nonfixed as well as fixed atoms are expected to be very badly described by both the PHVA and MBH. On the other hand, normal modes that are completely localized in the relaxed molecular region with nearly vanishing fixed atom displacements, or normal modes where the fixed atoms move collectively with respect to the optimized region, can be expected to result from the MBH approach.

Note that Head and co-worker investigated the coupling between the PHVA modes and the omitted modes occurring in the fixed atom region.¹⁰⁻¹² It may be possible to apply similar methods to the MBH modes.

III. APPLICATION TO THE ETHANOL MOLECULE

The PHVA and MBH methods are, in principle, developed to evaluate the frequencies in a large molecular system which cannot be optimized entirely at a high level of theory as the size increases, but in which only part (the active site) of the molecule is optimized at the high level and the remaining part at a substantially lower level of theory. However, in order to understand the advantages and disadvantages of both methods, it is instructive to validate them first on a small molecule, where the exact frequencies are readily available and can be compared with those predicted by PHVA and MBH in partially optimized geometries. To meet this purpose, we have chosen ethanol containing a well localized O-H stretch. The entire molecule has been optimized at a so-called high level [B3LYP/6-31+g(d)], and this optimized geometry will be used further on as the reference. Frequencies of all present normal modes are obtained by solving the mass-weighted eigenvalue problem [Eq. (4)] of the full 27×27 Hessian, and are tabulated in the first column of Table I. They coincide with the values that are obtained from the standard analytical frequency calculation in GAUSSIAN03,²² and will serve as benchmark values for further comparative studies. An exact treatment should generate six eigenvalues exactly equal to zero, corresponding to the global translation and rotation. In practice, the values differ slightly from zero (varying between -1 and 8 cm^{-1}). Translational frequencies are sensitive to numerical errors in the construction of the Hessian. Rotational frequencies are, in addition, affected by the small residual forces due to the finite convergence criteria. The effect of the almost zero fre-

TABLE I. Normal mode frequencies (in cm^{-1}) of ethanol derived from the benchmark geometry, which corresponds to the geometry optimization obtained at B3LYP/6-31+g(d). The rigid body is composed of the atoms in the shaded region. In the left column, translational and rotational frequencies from the full Hessian calculation are plotted before and after projection. Vibrational frequencies are not affected by this projection. The PHVA and MBH frequencies were ordered according to the maximum overlap with the benchmark modes.

									
Full	Projected	PHVA	MBH	PHVA	MBH	PHVA	MBH	PHVA	MBH
-1	0	-	-	-	-	72	-	59	-
-1	0	-	-	-	-	242	-	94	-
0	0	-	-	-	-	678	-	211	-
1	0	-	-	-	-	-	-	355	-
2	0	-	-	-	-	-	-	-	-
8	0	-	-	-	-	-	-	-	-
internal rotation:									
244		-	-	-	-	-	249	-	244
295		275	289	299	289	279	297	292	295
C-C-O stretch:									
417		-	-	-	491	-	458	-	419
mixed modes:									
825		-	-	272	807	823	992	1024	831
902		-	-	-	-	-	1011	-	906
1037		-	-	-	-	-	-	748	1040
1104		-	-	-	1031	-	-	1058	1111
1185		-	-	-	-	-	-	-	1195
1270		1164	1208	1224	1245	1210	1218	1226	1277
1305		-	-	715	-	1254	1269	1262	1308
1419		-	-	-	-	-	-	-	-
1461		-	-	-	-	1317	1378	1445	1452
1505		-	-	-	-	-	-	-	-
1521		-	-	-	-	-	-	-	-
C-H stretch:									
1544		-	-	-	-	1501	1534	1541	1541
2995		-	-	245	-	2928	2977	2995	2995
3021		-	-	-	-	2877	3015	3023	3023
3049		-	-	-	-	-	-	-	-
3116		-	-	-	-	-	-	-	-
3124		-	-	-	-	-	-	-	-
O-H stretch:									
3756		3645	3708	3755	3756	3755	3756	3756	3756

quencies on the other 21 frequencies is negligible here: projecting out the overall translation and rotation, which is implemented in most of the program packages such as GAUSSIAN03, gives six eigenmodes exactly equal to zero and does not affect the vibrational frequencies. Some of the

modes have a clear interpretation: the two lowest frequencies (244 and 295 cm^{-1}) correspond to internal rotations of the methyl top and the hydroxyl group, while the highest (3756 cm^{-1}) is associated with the highly localized O-H stretching mode.

We applied the PHVA and MBH methods to two different cases: first to the fully optimized structure and then to several partially optimized structures. The PHVA equations (9) are constructed with the submatrix H' of the Hessian that contains the second derivatives of the potential energy with respect to the Cartesian coordinates of the free atoms. The MBH frequencies corresponding to Eq. (10) are calculated, following the discussion in Sec. II D, with the same submatrix H' [see Eq. (12)] but with a different mass matrix.

The ordering of the calculated frequencies in the tables is determined by the maximum of the square of the overlap $|\langle \nu_{\text{bench}} | \nu \rangle|^2$ between the benchmark mode with frequency ν_{bench} and the calculated mode with frequency ν . For modes without a pronounced maximum overlap, this is of course rather arbitrary.

A. PHVA and MBH applied to the equilibrium structure

Frequencies are calculated for the fully optimized geometry, while the part of the Hessian chosen to be included in the vibrational analysis is varied. In this case, there are no remaining residual forces and the difference between the two models (PHVA or MBH) can easily be studied. The results are given in Table I for various rigid body sizes. The shaded box indicates the part of the molecule that is not included in the calculation of the Hessian. For instance, the second column of the table reports the three frequencies corresponding to the modes generated by a rigid body and one single atom that can vibrate. The only nonfixed atom in the vibrational analysis is the hydrogen atom of the hydroxyl group. In the third column, the whole hydroxyl group can vibrate, while in the last column we display the situation with only the methyl group fixed at its reference geometry.

Only three PHVA and MBH modes are calculated in the case where only the H atom of the hydroxyl group is taken into account (second column of Table I). Visualization of the three modes revealed that they all qualitatively correspond to a normal mode of the benchmark frequency spectrum, but that their values are underestimated (benchmark values are 295, 1270, and 3756 cm^{-1}). In the PHVA method, the underestimation is even more pronounced due to the use of infinite masses. This is a general conclusion: the PHVA fails in accurately reproducing the benchmark values, especially in the small and medium frequency regions, while the quantitative agreement of the MBH predictions of localized modes—taking place in the nonfixed region—with the benchmark values is manifestly present. The agreement is even very close in the last case, where the rigid body consists of a fixed methyl group. The O–H stretch mode is always present in the frequency spectrum, only in the case of one relaxed atom (the hydrogen) is the obtained frequency slightly underestimated.

Looking at lower frequencies, the MBH approach gives consistently better results than the PHVA, because the reduced mass effect is taken into account. As the total mass of the fixed block decreases, it is obvious that the PHVA induces spurious low frequency modes of the order of 60–100 cm^{-1} , corresponding to translations/rotations of the nonfixed atoms in the field of their environment (the fixed

atoms). Summarizing, when working with molecules in the gas phase, the use of the MBH model is highly recommended as soon as the total mass of the fixed block becomes of the same order as the total mass of the relaxed atoms.

B. PHVA and MBH applied to partially optimized structures

We introduce partially optimized geometries and verify whether the relevant calculated frequencies in the active site can be reproduced by both the PHVA and MBH models in an accurate way with respect to the benchmark frequencies. Initially, the ethanol molecule has been optimized at the lower HFO/STO-3g level. Then the system was partially optimized at the higher B3LYP/6-31+g(d) level while keeping the atoms of the rigid block fixed at their initial HF/STO-3g positions. Results are collected in Table II. The atoms belonging to the shaded box were not optimized at the high level but were kept fixed at their low level geometries. For each case, the frequencies resulting from a full Hessian [calculated at B3LYP/6-31+g(d)] diagonalization, i.e., from a standard normal mode analysis in Gaussian, as well as the normal modes resulting from the PHVA and MBH methods are tabulated. The standard frequency analysis always gives a number of spurious imaginary frequencies, as could be expected since residual forces on the nonoptimized atoms disturb the evaluation of the frequencies. In addition, the positive frequencies deviate substantially from the benchmark values given in Table I. This means that the normal frequencies generated by standard procedures in program packages such as GAUSSIAN03,²² in molecules whose atomic positions have not been optimized at the same level of theory, are far from being accurate.

By applying the PHVA or the MBH model, the unphysical imaginary frequencies disappear, but the resulting frequencies, however, differ significantly between the two methods. The MBH results converge rapidly to the benchmark values, highlighting the efficiency of the proposed MBH model. A striking resemblance is even observed in the last column where the fixed body is restricted to the ending methyl group. The low frequency spectrum of the PHVA method, however, deviates largely from the benchmark values. This is entirely due to the reduced mass effect, inducing spurious unphysical modes.

While the full Hessian frequencies are very sensitive to the exact molecular geometry, it is remarkable that PHVA and MBH frequencies from the partially optimized geometry are very close to those obtained with the benchmark geometry with the same block size. This indicates that the PHVA and MBH models are less sensitive to the exact internal geometry of the fixed block. This could be important for future applications in large systems.

It is also interesting to compare the thermodynamical quantities associated with the vibrational part of the partition function of the system. In Fig. 2 the vibrational contribution to the entropy S and the free enthalpy G are given for the different partially optimized ethanol configurations at $T = 298.15$ K. The values calculated with the benchmark frequencies are also indicated. It is clear that reducing the num-

TABLE II. Normal mode frequencies (in cm^{-1}) of ethanol derived on the basis of partially optimized geometries at the B3LYP/6-31+ $g(d)$ level of theory. The rigid body is composed of atoms in the shaded region and its geometry is originated from a geometry optimization of the whole molecule at the low HF/STO-3 g level. Benchmark frequencies are given in the left column for comparison. The PHVA and MBH frequencies were ordered according to the maximum overlap with the benchmark modes.

bench	Full PHVA MBH	Full PHVA MBH	Full PHVA MBH	Full PHVA MBH	Full PHVA MBH
-1	-258 - -	-264 - -	-254 71 -	-251 60 -	-
-1	-130 - -	-132 - -	-80 243 -	-90 95 -	-
0	-62 - -	-75 - -	-61 678 -	-66 212 -	-
1	-1 - -	-2 - -	-1 - -	-1 355 -	-
2	-1 - -	0 - -	1 - -	1 - -	-
8	1 - -	1 - -	1 - -	1 - -	-
244	91 - -	42 245 -	124 - 248	124 - 246	-
295	294 281 295	284 296 285	289 280 298	289 294 297	-
417	403 - -	396 - 490	397 - 458	397 - 419	-
825	760 - -	758 272 813	793 824 992	793 746 839	-
902	873 - -	877 - -	880 - -	883 - 909	-
1037	1021 - -	1025 - 1032	1025 - 1013	1029 1025 1041	-
1104	1070 - -	1096 - -	1097 - -	1098 1058 1116	-
1185	1147 - -	1156 - -	1177 - -	1177 - 1201	-
1270	1257 1161 1205	1261 1224 1244	1266 1210 1218	1267 1225 1280	-
1305	1267 - -	1277 715 -	1298 1252 1268	1300 1262 1313	-
1419	1384 - -	1385 - -	1388 - -	1390 - -	-
1461	1422 - -	1429 - -	1452 1315 1376	1454 - -	-
1505	1489 - -	1488 - -	1488 - -	1488 1444 1452	-
1521	1505 - -	1506 - -	1507 - -	1508 - -	-
1544	1527 - -	1527 - -	1544 1502 1535	1544 1542 1542	-
2995	3140 - -	3137 - -	2995 2877 2978	2994 2995 2995	-
3021	3162 - -	3162 - -	3023 2929 3016	3022 3023 3023	-
3049	3177 - -	3173 - -	3162 - -	3162 - -	-
3116	3242 - -	3243 - -	3243 - -	3242 - -	-
3124	3253 - -	3252 - -	3251 - -	3251 - -	-
3756	3751 3640 3703	3755 3755 3755	3756 3756 3756	3756 3756 3756	-

ber of modes by introducing the rigid block affects both the vibrational entropy and free enthalpy. However, the MBH values tend consistently to the benchmark value when the block size decreases, whereas the PHVA values do not. This is to be expected, since PHVA assumes infinite mass for the fixed block and for the present application the MBH is physically more relevant.

IV. EXTENSION: MULTIPLE MOBILE BLOCKS

The application field of MBH can be easily extended to multiple mobile blocks. Suppose the molecule is decomposed into K rigid blocks and N_E freely relaxed atoms. The whole molecule is optimized at a low level of theory determining the position of each atom in each of the rigid blocks. In a second step, a higher level of theory is used and one optimizes the positions of the N_E free atoms, as well as the positions and orientations of the rigid blocks, keeping their original internal geometry. As a consequence, residual forces

remain between the fixed atoms within a block, thus the full gradient is nonzero, and a normal frequency analysis along the lines of Eq. (7) with the full Hessian will therefore produce unphysical frequencies.

In the extended method, one assumes that each block is allowed to participate in the small amplitude vibrations, moving as a rigid body. To implement this approach, a suitable choice of internal coordinates is necessary, e.g., the Z -matrix formalism. It is, in fact, sufficient to choose the numbering of the atoms in the Z -matrix construction such that the atoms in each fixed block are numbered consecutively.

As in Sec. II C, the geometry of each block b is described by $3N_{F,b}-6$ internal coordinates $\{\theta_{F,b}^i\}$. The imposed reference structure $\{\mathbf{r}_{A,F,b}^0\}$ of the fixed atoms of block b then determines the values of the $3N_{F,b}-6$ internal coordinates $\{\theta_{F,b}^i\}$. The remaining $3N_E+6(K-1)$ internal coordinates will

be labeled $\{\theta_{i_E}\}$. Optimizing the energy with fixed $\{\theta_{i_{F,b}}^0\}$ and varying $\{\theta_{i_E}\}$ yield the reference structure $\{\theta_{i_E}^0\}$.

To impose the fixed internal geometry of each block during the vibrational analysis, the displacements $\Delta_{i_{F,b}}$ and corresponding velocities $\Delta_{i_{F,b}}$ are set equal to zero in the second-order expansion of Eq. (6). The gradient term in Eq. (6) thereby vanishes, so the reduced system of $3N_E+6(K-1)$ internal coordinates $\{\theta_{i_E}\}$ is in equilibrium.

The corresponding NMA equations are obtained by omitting the rows and columns related to the $\{\theta_{i_{F,b}}^0\}$ variables from the Hessian and mass matrix in Eq. (7). The resulting reduced eigenvalue problem of dimension $3N_E+6K$ has six zero eigenvalues corresponding to overall translation and rotation. These can be decoupled as in the mobile block Hessian approach with a congruent transformation similar to Eq. (10).

The discussion of this multiple MBH approach is similar to the single MBH with only one rigid body. First, the disturbing negative eigenvalues due to residual forces are eliminated, and imaginary frequencies will only occur if the reduced system of relaxed atoms and K blocks is actually in a transition state. The presence of six zeros shows that the translational and rotational invariance are respected in the second-order expansion of the potential energy of the reduced system. Concerning the mass effect, the finite mass of each block is taken into account. Moreover, the method provides a reduction in computer time, because for each block b with $N_{F,b}$ atoms, only derivatives with respect to six coordinates instead of $3N_{F,b}$ have to be calculated.

Of course, the main question is whether the multiple MBH approach is capable of reproducing the normal mode frequencies of the fully optimized benchmark structure. Generally, two types of benchmark modes are well reproduced by the multiple MBH in analogy with the single MBH ver-

sion: (i) modes localized in the relaxed part of the molecule (ii) and modes where the fixed atoms move as a whole. Some other modes present in the benchmark, on the other hand, are obviously not reproduced by the multiple MBH. These are modes in which the internal geometry of the rigid blocks is changed. Such modes are not localized in the chemically active part of the molecule and are therefore not relevant for the present study. Modes where the displacements are spread out over the relaxed as well as fixed atoms are mostly not reproduced in the multiple MBH model, except those where the atoms of the rigid bodies move coherently [modes of type (ii)]. Some examples of such modes are shown in Fig. 1. For instance, the rigid bodies can rotate about the single bonding axis connecting them with the active part of the molecule or they can contribute to a stretching, a bending motion, etc.

For practical purposes, it is important to select the freely relaxed region of the molecule attentively. One of the main advantages of this extension to multiple MBH is the computational profit as a consequence of the reduced number of Hessian matrix elements to be evaluated, but unfortunately this option is not (yet) present in most of the standard program packages.

V. APPLICATION TO DI-*n*-OCTYL-ETHER

A suitable example to validate the multiple MBH approach is di-*n*-octyl-ether ($C_8H_{17}-O-C_8H_{17}$). Ethers are known to have a C–O stretching band that falls in the fingerprint region at $1050-1260\text{ cm}^{-1}$. This vibration is a localized mode and very characteristic for all ethers. The ability of the various models in reproducing this stretching band is a strong validation for the MBH method.

The geometry of the molecule is first fully optimized at the B3LYP/6-31+g(*d*) level of theory, and a normal mode

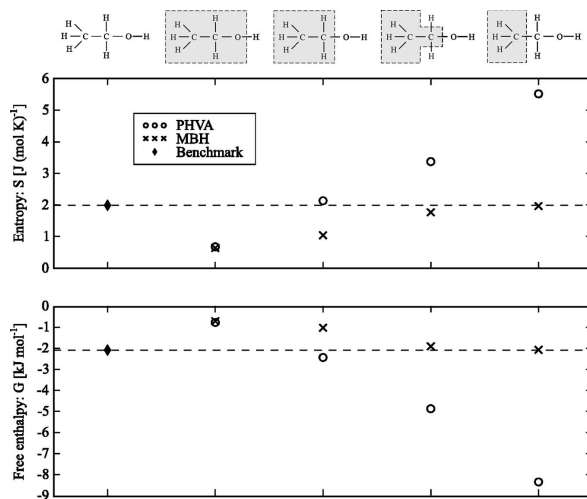


FIG. 2. Vibrational contribution to the entropy S and the free enthalpy G calculated with PHVA (\circ) and MBH (\times) frequencies are given for the different partially optimized ethanol configurations at $T=298.15\text{ K}$. Benchmark values are indicated by the dashed lines. The fixed block in the MBH calculation consists of the atoms in the shaded box.

TABLE III. Frequencies $\lambda^{1/2}$ (in cm^{-1}) of di-*n*-octyl-ether of various partially optimized configurations defined in Fig. 3 are compared with the benchmark frequencies of the fully optimized geometry (left column). Three approaches are used: the full Hessian calculation (Full), the PHVA method, and the MBH approach. The size of the rigid bodies is defined by the configuration label. The Full/PHVA/MBH frequencies are ordered according to the maximum overlaps $|\langle \nu_{\text{bench}} | \nu \rangle|^2$ with benchmark eigenmodes, which are given by the values between parentheses (in %).

Bench	Configuration 8			Configuration 5			Configuration 2		
	Full	PHVA	MBH	Full	PHVA	MBH	Full	PHVA	MBH
0	0 (100)	- (-)	-1 (100)	0 (100)	29 (33)	-1 (100)	0 (100)	12 (56)	-1 (100)
0	0 (100)	- (-)	0 (100)	0 (100)	- (-)	-1 (100)	0 (100)	134 (73)	0 (100)
0	0 (100)	- (-)	0 (100)	0 (100)	- (-)	0 (100)	0 (100)	22 (58)	0 (100)
1	-8 (95)	- (-)	1 (100)	7 (92)	- (-)	1 (100)	9 (88)	228 (17)	0 (100)
1	-12 (96)	- (-)	0 (100)	-12 (96)	- (-)	1 (100)	-9 (98)	- (-)	-1 (100)
4	-156 (61)	- (-)	4 (97)	-124 (29)	- (-)	4 (99)	- (-)	34 (67)	3 (100)
11	9 (65)	- (-)	16 (95)	27 (54)	- (-)	12 (99)	34 (53)	228 (17)	11 (100)
19	-15 (98)	- (-)	24 (98)	-15 (99)	50 (40)	19 (100)	-9 (99)	- (-)	19 (100)
27	-154 (40)	- (-)	42 (67)	-124 (32)	- (-)	20 (99)	-36 (29)	63 (49)	27 (100)
33	53 (37)	- (-)	64 (54)	-52 (43)	- (-)	35 (98)	64 (34)	34 (59)	33 (100)
51	36 (96)	- (-)	114 (78)	37 (95)	132 (44)	55 (99)	40 (97)	53 (60)	51 (100)
53	-104 (34)	215 (30)	283 (52)	-53 (41)	- (-)	60 (94)	37 (53)	- (-)	53 (100)
65	- (-)	- (-)	- (-)	57 (50)	57 (39)	79 (86)	- (-)	63 (42)	65 (100)
84	-96 (25)	- (-)	- (-)	61 (32)	79 (30)	101 (71)	75 (66)	- (-)	84 (100)
92	- (-)	- (-)	- (-)	90 (48)	101 (33)	117 (63)	- (-)	95 (78)	92 (100)
96	84 (98)	261 (22)	- (-)	86 (97)	- (-)	114 (93)	87 (98)	102 (72)	96 (100)
113	77 (32)	- (-)	- (-)	90 (34)	- (-)	- (-)	109 (51)	95 (51)	113 (100)
129	90 (57)	- (-)	- (-)	118 (76)	- (-)	362 (54)	153 (34)	133 (87)	130 (100)
133	110 (41)	244 (50)	300 (47)	116 (52)	144 (48)	- (-)	132 (68)	122 (62)	133 (100)
138	137 (100)	- (-)	180 (88)	138 (100)	- (-)	145 (99)	138 (100)	276 (56)	138 (100)
150	- (-)	- (-)	- (-)	139 (83)	228 (44)	362 (51)	169 (53)	149 (62)	150 (100)
154	143 (98)	- (-)	- (-)	144 (99)	- (-)	198 (79)	145 (99)	167 (71)	154 (100)
160	127 (44)	- (-)	- (-)	152 (80)	- (-)	149 (54)	158 (98)	157 (68)	160 (100)
162	128 (46)	- (-)	- (-)	150 (80)	130 (36)	143 (55)	162 (73)	154 (76)	162 (100)
178	160 (83)	125 (29)	143 (41)	175 (95)	230 (57)	169 (78)	182 (82)	178 (95)	178 (100)
220	209 (99)	- (-)	336 (27)	210 (99)	243 (55)	209 (56)	212 (99)	233 (83)	220 (100)
...									
1149	1146 (99)	1146 (94)	1147 (95)	1148 (100)	1148 (100)	1148 (100)	1149 (100)	1149 (100)	1149 (98)
1461	1457 (99)	1442 (83)	1442 (83)	1461 (100)	1461 (100)	1461 (100)	1461 (100)	1461 (100)	1461 (100)
1549	1547 (98)	1545 (88)	1545 (88)	1549 (100)	1548 (100)	1548 (100)	1549 (100)	1549 (100)	1549 (100)
2954	2953 (100)	2953 (99)	2953 (99)	2954 (100)	2954 (100)	2954 (100)	2954 (100)	2954 (100)	2954 (100)
2965	2965 (99)	2965 (99)	2965 (99)	2965 (100)	2965 (100)	2965 (100)	2965 (100)	2965 (100)	2965 (100)
2983	2983 (99)	2984 (98)	2984 (98)	2983 (100)	2983 (100)	2983 (100)	2983 (100)	2983 (100)	2983 (100)
2985	2985 (99)	2986 (98)	2986 (98)	2985 (100)	2985 (100)	2985 (100)	2985 (100)	2985 (100)	2985 (100)

analysis at the same level, constructed with analytical second derivatives, provides the whole spectrum of frequencies, which serve as benchmark values. Some eigenfrequencies are listed in the left column of Table III. The full Hessian frequency analysis on this completely optimized geometry generates six (almost) zero eigenvalues, as it should be. In the *low* frequency spectrum, typical modes are found that involve large parts of the molecule. In this class, bending and torsionlike modes are found where large massive blocks of the molecule are involved. Also internal rotations and collective accordionlike motions are present. For instance, the three lowest benchmark frequencies (11, 19, and 27 cm^{-1}) are identified as the relative rotation/torsion of the two octyl chains around the three axes of inertia of the molecule. In view of later application in which the frequencies are used as input for the construction of the partition function in the harmonic oscillator approximation, this low frequency spectrum is extremely important as they give the largest contri-

bution to the vibrational partition function. In the *high* frequency range ($\sim 3000 \text{ cm}^{-1}$), localized C–H stretches are found. Some specific modes are interesting as they are localized near the central O atom: 2985, 2983, 2965, and 2954 cm^{-1} correspond to C–H stretching modes—symmetrically or antisymmetrically—of the CH_2 moieties attached to the middle oxygen, 1461 cm^{-1} is mainly a bending of these two CH_2 units with respect to each other and 1149 cm^{-1} is the typical C–O stretch for ethers.

In what follows, the multiple MBH model is applied to a set of partially optimized structures at the B3LYP/6-31+*g*(*d*) level of theory, while part of the molecule is kept fixed at the geometry optimized at the lower HF/STO-3*g* level. The central part of the molecule is always allowed to relax. Two nonoptimized blocks are systematically taken into account, located at both sides of the central part. They differ in the length of the chain that is considered in the nonoptimized

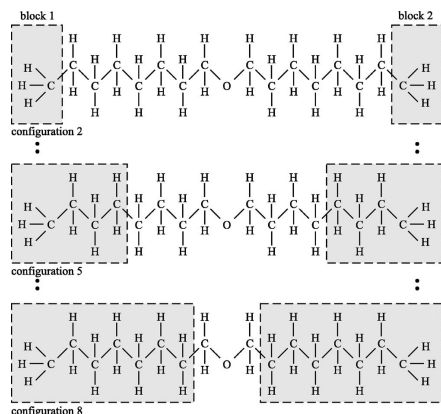


FIG. 3. Specification of the various configurations of di-*n*-octyl-ether with rigid bodies indicated as shaded regions. Atoms in shaded boxes are fixed at HF/STO-3*g* positions during the partial geometry optimization at the B3LYP/6-31+*g*(*d*) level.

block, as schematically depicted in Fig. 3. The influence of varying chain length on the reproduction of the frequency spectrum will be investigated. The fully optimized case is referred to as configuration 1 (the benchmark geometry), whereas in configuration 8, two heptyl chains are constrained during the optimization.

The low energy part of the spectra obtained by diagonalizing the full Cartesian Hessian (as would be done in the common normal mode analysis) is represented in Fig. 4. The same deficiencies related to nonequilibrium geometries as in the ethanol example are noticed: the appearance of non-negligible negative (imaginary) frequencies of the order of -225 cm^{-1} and the lack of three eigenvalues equal to zero belonging to the global rotation. Moreover, the number of spurious negative frequencies can vary depending on the number of atoms belonging to the fixed block and is thus

unpredictable. The repercussion of the negative frequencies is mainly found in the low frequency modes, where large parts of the molecule, including some nonoptimized atoms, are involved. Values for a selected number of modes are given in Table III. The reproduction of the three lowest eigenfrequencies is very poor and even negative values are found. It is, nevertheless, remarkable that the previously mentioned localized modes are very well reproduced even within the presence of some negative frequencies. For instance, the localized C–O band, situated at 1149 cm^{-1} , is well reproduced, probably owing to the fact that only optimized atoms are involved in this internal motion.

This numerical example illustrates the necessity to strictly reduce the dimension of the Hessian, following the procedure explained in the (multiple) MBH approach, in order to get rid of the disturbing negative frequencies. The low frequency part of the resulting spectra is plotted in Fig. 5 for the different configurations. Thereby all negative eigenvalues are eliminated, six eigenvalues are equal to zero as could be expected, and the other frequencies are all physically significant. Figure 5 allows qualitative comparison between the MBH normal mode frequencies and the benchmark values. Obviously, the MBH results belonging to the largest fixed blocks (configuration 8) differ substantially from the benchmark values. As the number of relaxed atoms increases, the agreement improves (configuration 2). Quantitative values are given for a selected number of modes in Table III, where the MBH modes were sorted according to their resemblance to the benchmark modes (see caption of table). The PHVA method was applied as well, combining the two individual fixed blocks into one fixed block, as shown in Fig. 6. The PHVA method is capable of reproducing localized modes, but frequencies in the lower spectrum are very poorly reproduced or are absent. Apparently, the MBH model is able to reproduce accurately not only the previously mentioned localized modes, but more importantly, the low frequency modes have much more realistic values compared to the full Hessian values or the PHVA values.

A more pronounced study can be made by evaluating the overlap between the calculated multiple MBH modes and

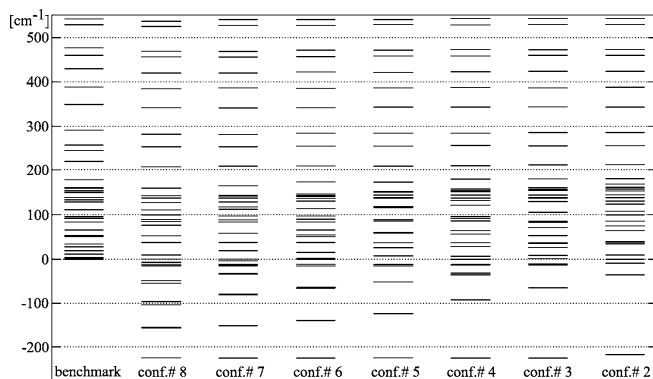


FIG. 4. Lowest frequencies $\lambda^{1/2}$ (in cm^{-1}) of di-*n*-octyl-ether based on the full Cartesian Hessian belonging to the various partially optimized configurations defined in Fig. 3. Partial optimization at the B3LYP/6-31+*g*(*d*) level. Plot on the left displays the exact normal mode frequencies (full geometry optimization) that serve as benchmark.

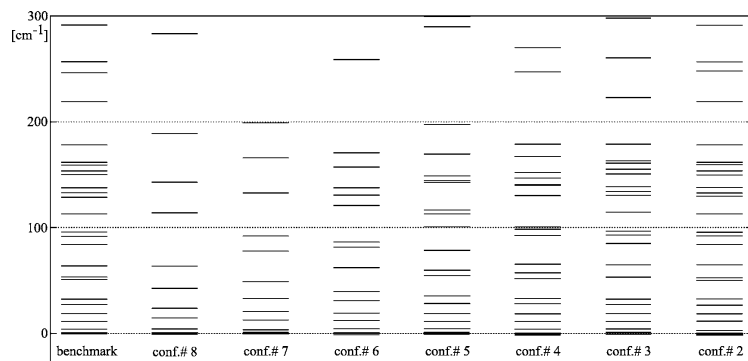


FIG. 5. Lowest frequencies $\lambda^{1/2}$ (in cm^{-1}) of di-*n*-octyl-ether based on the multiple MBH model belonging to the various partially optimized configurations defined in Fig. 3. Partial geometry optimization at the B3LYP/6-31+*g*(*d*) level. Plot on the left displays the exact normal mode frequencies (full geometry optimization) that serve as benchmark for the other plots where two rigid bodies (defined by the configuration label) are taken into account in the frequency analysis.

benchmark modes. The square of the overlap $|\langle \nu_{\text{bench}} | \nu_{\text{MBH}} \rangle|^2$ gives a measure of how strong the benchmark mode with frequency ν_{bench} is involved in the specific MBH mode with frequency ν_{MBH} . Due to the completeness of the basis of benchmark modes, the sum of these numbers over all benchmark frequencies is equal to 1. The results for configurations 8, 5, and 2 are given in a scatter plot in Fig. 7. All values above a certain limit (we take 20% throughout this work) are indicated by a circle, and the darker the fill intensity of the circle, the larger the magnitude of the overlap. The most ideal case is a black filled circle on the diagonal, as it means that the multiple MBH model has reproduced the benchmark mode in an excellent way. If the strength of a specific MBH mode is spread out over various benchmark modes, a larger scattering of circles is noticed off the diagonal, which is indeed the case for configuration 8.

By enlarging the optimized region, the discrepancy with the benchmark almost disappears. In fact, already from configuration 5 results have essentially converged (see Fig. 7). By retaining only the ending methyl groups in the fixed boxes (configuration 2), all relevant frequencies are excellently reproduced. In the low frequency region (below 1500 cm^{-1}), they are identical within a margin of 1 cm^{-1} . Only a few nonrelevant modes that are localized in the methyl tops are absent.

In Table III the maximum overlap $|\langle \nu_{\text{bench}} | \nu \rangle|^2$ between benchmark modes $|\nu_{\text{bench}}\rangle$ and the selected full Hessian,

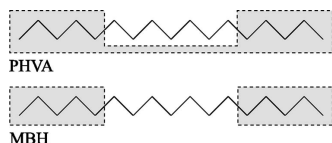


FIG. 6. The PHVA method implies the introduction of one block. For the multiple MBH method, two rigid blocks were used.

PHVA, and MBH modes $|\nu\rangle$ is added between brackets. The overlap for the localized modes is consistently excellent. The overlap for the full Hessian or PHVA modes is quite poor for the low frequencies, whereas the MBH modes show very reasonable overlaps even in this region.

VI. SUMMARY AND CONCLUSIONS

In this paper a new method, referred to as the MBH approach, was introduced to calculate the normal modes of partially optimized molecular systems. The MBH approach is an extension of the previously introduced PHVA method in the sense that the nonoptimized regions of the system are allowed to move as rigid blocks with respect to the optimized part of the system. Both the MBH and PHVA methods eliminate the imaginary frequencies that result from the common full Hessian normal mode analysis on a partially optimized structure. In cases where the surrounding nonoptimized part of the system effectively are kept immobile by external forces at its reference position, such as in a lattice, the PHVA and MBH methods perform equally well, but for a flexible surrounding medium the MBH method is more appropriate. The various methods were outlined in two examples, i.e., ethanol and di-*n*-octyl-ether. It was found that the localized modes in the optimized part of the system are always well reproduced, irrespective of the applied method. Even a full Hessian normal mode analysis gives quite accurate values for the frequencies of localized modes. However, for normal modes that involve a larger part of the molecular system, the MBH method performs better, since the nonoptimized blocks can move coherently with respect to the chemically active part of the system.

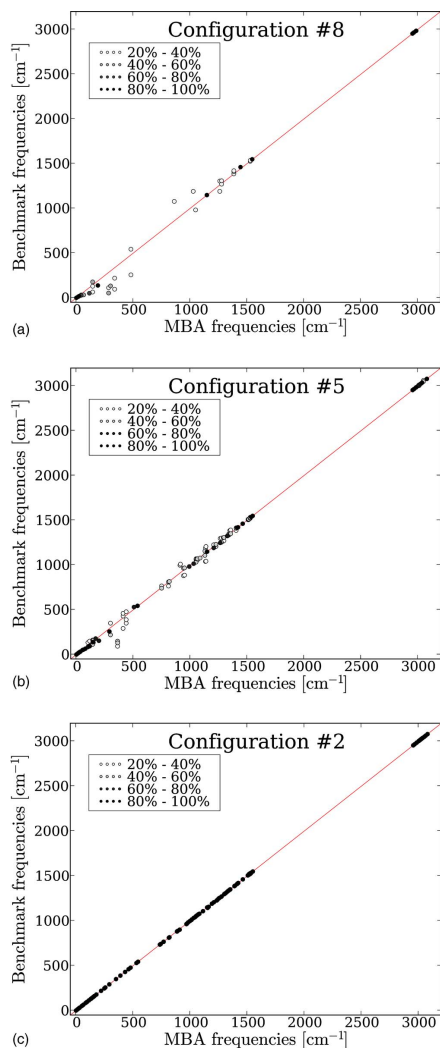


FIG. 7. Square of the overlap between the MBH normal modes and the benchmark normal mode frequencies of di-*n*-octyl-ether. The sum of the benchmarks is always normalized to 1 for each MBH frequency.

The MBH method does not only eliminate spurious negative frequencies but implicates also a serious reduction of the computational cost for large molecular systems, as the calculation of the Hessian is the most expensive part after a geometry optimization. Molecular modeling focuses more and more on these extended molecular systems, and hence, such efficient techniques are indispensable. The multiple MBH model looks highly suited for use in systems in which the molecular environment is rather flexible, such as reaction in solvents. Most of the solvent molecules can be regarded as rigid bodies moving in all directions with respect to the optimized central part of the system. Their participation to the normal modes can be simulated by the MBH model. The application of the MBH approach for the calculation of partition functions and derived quantities will be further investigated in the future.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research-Flanders and the Research Board of Ghent University. This work was partly performed within the framework of the SBO-BIPOM program of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

- ¹A. Warshel and M. Levitt, *J. Mol. Biol.* **103**, 227 (1976).
- ²X. Assfeld and J. L. Rivail, *Chem. Phys. Lett.* **263**, 100 (1996).
- ³J. L. Gao, P. Amara, C. Alhambra, and M. J. Field, *J. Phys. Chem. A* **102**, 4714 (1998).
- ⁴Y. K. Zhang, T. S. Lee, and W. T. Yang, *J. Chem. Phys.* **110**, 46 (1999).
- ⁵F. Stevens, H. Vrielinck, V. Van Speybroeck, E. Pauwels, F. Callens, and M. Waroquier, *J. Phys. Chem. B* **110**, 8204 (2006).
- ⁶S. Q. Jin and J. D. Head, *Surf. Sci.* **318**, 204 (1994).
- ⁷M. D. Calvin, J. D. Head, and S. Q. Jin, *Surf. Sci.* **345**, 161 (1996).
- ⁸H. Li and J. Jensen, *Theor. Chem. Acc.* **107**, 211 (2002).
- ⁹N. A. Besley and K. A. Metcalf, *J. Chem. Phys.* **126**, 7 (2007).
- ¹⁰J. D. Head, *Int. J. Quantum Chem.* **65**, 827 (1997).
- ¹¹J. D. Head and Y. Shi, *Int. J. Quantum Chem.* **75**, 815 (1999).
- ¹²J. D. Head, *Int. J. Quantum Chem.* **77**, 350 (2000).
- ¹³H. Lin, J. Z. Pu, T. V. Albu, and D. G. Truhlar, *J. Phys. Chem. A* **108**, 4112 (2004).
- ¹⁴A. Tachibana and K. Fukui, *Theor. Chim. Acta* **49**, 321 (1978).
- ¹⁵D. J. Wales, *J. Chem. Phys.* **113**, 3926 (2000).
- ¹⁶R. Murry, J. T. Fourkas, L. Wu-Xiong, and T. Keyes, *J. Chem. Phys.* **110**, 10410 (1999).
- ¹⁷I. Kolossvary and C. McMartin, *J. Math. Chem.* **9**, 359 (1992).
- ¹⁸J. Goldstone, A. Salam, and S. Weinberg, *Phys. Rev.* **127**, 965 (1962).
- ¹⁹V. Van Speybroeck, D. Van Neck, and M. Waroquier, *J. Phys. Chem. A* **104**, 10939 (2000).
- ²⁰V. Van Speybroeck, P. Vansteenkiste, D. Van Neck, and M. Waroquier, *Chem. Phys. Lett.* **402**, 479 (2005).
- ²¹P. Vansteenkiste, D. Van Neck, V. Van Speybroeck, and M. Waroquier, *J. Chem. Phys.* **124**, 044314 (2006).
- ²²M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.

Paper II

Cartesian formulation of the Mobile Block Hessian approach to vibrational analysis in partially optimized systems

A. Ghysels, D. Van Neck and M. Waroquier

Journal of Chemical Physics 127 (16), 164108 (2007)

Copyright 2007 by American Institute of Physics

Cartesian formulation of the mobile block Hessian approach to vibrational analysis in partially optimized systems

A. Ghysels, D. Van Neck, and M. Waroquier^{a)}*Center for Molecular Modeling, Laboratory of Theoretical Physics, Ghent University, Proeftuinstraat 86, B-9000 Gent, Belgium*

(Received 8 June 2007; accepted 30 August 2007; published online 23 October 2007)

Partial optimization is a useful technique to reduce the computational load in simulations of extended systems. In such nonequilibrium structures, the accurate calculation of localized vibrational modes can be troublesome, since the standard normal mode analysis becomes inappropriate. In a previous paper [A. Ghysels *et al.*, *J. Chem. Phys.* **126**, 224102 (2007)], the mobile block Hessian (MBH) approach was presented to deal with the vibrational analysis in partially optimized systems. In the MBH model, the nonoptimized regions of the system are represented by one or several blocks, which can move as rigid bodies with respect to the atoms of the optimized region. In this way unphysical imaginary frequencies are avoided and the translational/rotational invariance of the potential energy surface is fully respected. In this paper we focus on issues concerning the practical numerical implementation of the MBH model. The MBH normal mode equations are worked out for several coordinate choices. The introduction of a consistent group-theoretical notation facilitates the treatment of both the case of a single block and the case of multiple blocks. Special attention is paid to the formulation in terms of Cartesian variables, in order to provide a link with the standard output of common molecular modeling programs. © 2007 American Institute of Physics. [DOI: 10.1063/1.2789429]

I. INTRODUCTION

Partial optimization is of interest to model extended molecular systems where a full optimization is still computationally too expensive. Examples include polymer chains,¹ supramolecular assemblies, systems embedded in a solvent or (macro)molecules adsorbed within porous materials,² etc. The reference point is obtained by optimizing the geometry with respect to a subset of internal coordinates, keeping the remainder fixed during the optimization. The fixed internal coordinates, e.g., correspond to the fixed geometry of a part of the molecule, which is not expected to influence the more interesting region—the active site—of the molecule, and may even be calculated at a lower level of theory.

After partial optimization, the system is still in a global nonequilibrium state, and severe difficulties arise when normal mode analysis is applied in a standard fashion.^{3–6} Several approaches have been developed for the determination of normal modes in such partially optimized geometries. As a general rule, one should restrict the normal mode analysis to the degrees of freedom that have been optimized, since the dynamical system corresponding to the subset of optimized internal coordinates is in equilibrium.

A method based on a subblock of the Cartesian Hessian was first introduced by Head and co-workers^{7,8} and further developed by Li and Jensen.⁹ Within this methodology, hereafter referred to as the partial Hessian vibrational analysis (PHVA), the normal modes are calculated for the system with the fixed atoms frozen at their reference positions as if they were given an infinite mass, and only the relaxed atoms

can participate in the vibrations. An improved version of the PHVA was investigated by Head^{10–12} but this method requires the knowledge of an additional off-diagonal block of the Hessian.

Recently, we presented the mobile block Hessian approach¹³ (MBH) which has an equal computational load compared to the plain PHVA. In the MBH method, the group of fixed atoms is considered as a rigid block, which is allowed to participate in the small amplitude vibrations with the restriction that the internal geometry of the block remains unchanged. As a result, the unphysical effects related with nonequilibrium structures are avoided. The extension was made to the multiple MBH where several rigid blocks are introduced in the vibrational analysis. In Ref. 13, the MBH method was analyzed in a few test applications on small molecules, for which comparison with an exact treatment is feasible. The results indicated that the MBH approach is capable of reproducing localized modes, but also gives consistently better results for the low frequency modes with respect to the PHVA method. Also, the derived thermodynamical quantities such as entropy and free enthalpy are better reproduced when calculated with MBH frequencies instead of PHVA frequencies.

An important issue is the implementation in existing *ab initio* and molecular mechanics programs. In Ref. 13, the MBH equations are formulated in a set of internal coordinates. However, some standard packages only give the full Hessian in Cartesian coordinates as output, and the transformation between different sets of coordinates can be cumbersome. It is therefore of interest to analyze the derivation of the MBH normal mode equations in both internal and Carte-

^{a)}Electronic mail: michel.waroquier@ugent.be

sian coordinates, and to provide explicit formulae for the case where only Cartesian Hessian elements are available after the partial optimization. The introduction of some group-theoretical aspects will result, at the end of the discussion, in a rather simple calculation scheme for the MBH approach.

Apart from the physical content of the method, the computational cost of its implementation is of great importance. Partial optimization reduces the computational cost in finding the (approximate) reference geometry. It is then desirable that the subsequent frequency analysis is less time consuming as well, by a reduction in the required number of second derivatives. In this respect the PHVA is highly economical, as it only needs a simple submatrix of the Cartesian Hessian. The MBH as formulated in Ref. 13 needs an equivalent number of second derivatives, but requires a submatrix of the Hessian in internal coordinates when multiple blocks are involved. Special attention will therefore be paid to the required number of second derivatives in internal as well as in Cartesian coordinates.

The paper is organized as follows. In Sec. II we start with the introduction of some group-theoretical concepts which will be used extensively throughout the paper. Section III treats the MBH model with one single block. The formulation in internal coordinates is revised and it is shown that it is always possible to express the MBH equations using only the derivatives with respect to the Cartesian coordinates of the relaxed atoms. The same analysis is applied in Sec. IV to the multiple MBH model with an arbitrary number of blocks. In Sec. V the practical calculation scheme is summarized. Finally, in Sec. VI, we discuss the computational profit of the MBH approach compared to a full Cartesian Hessian calculation.

II. SOME GROUP-THEORETICAL CONCEPTS

Consider a molecule with N masses m_A , $A=1, \dots, N$. The positions are described by Cartesian coordinates $\mathbf{r}_A \equiv \{r_{A\mu}\}_{\mu=x,y,z}$ with respect to a space-fixed frame. The energy of the system reads

$$E = \frac{1}{2} \sum_{A\mu} m_A \dot{r}_{A\mu}^2 + V(\{\mathbf{r}_A\}), \quad (1)$$

where $\dot{r}_{A\mu}$ is a time derivative and V is the potential energy. In the MBH approach, one can select parts of the molecule (blocks) of which the internal geometry is kept fixed, but which can move as rigid bodies (i.e., rotate and translate). The analysis of MBH in Cartesian coordinates is greatly facilitated by introducing some (elementary) group-theoretical concepts, which will be used extensively in the following sections. We first consider the relevant transformation group in three-dimensional space that leaves the potential energy invariant. This six-parameter group consists of combinations of translations and rotations, taking coordinates \mathbf{r} to their translated/rotated positions \mathbf{r}' ,

$$\mathbf{r}' \rightarrow \mathbf{r}'_\mu = \mathbf{g}_\mu(\mathbf{r}, p). \quad (2)$$

Here, p stands for the six parameters p_α , $\alpha=1, \dots, 6$. The composition of two transformations can be written as

$$\mathbf{g}(\mathbf{g}(\mathbf{r}, p), p') = \mathbf{g}(\mathbf{r}, \Phi(p, p')), \quad (3)$$

where $\Phi(p, p')$ are the parameters of the composition (i.e., the group multiplication law).

The precise parameterization of the group is arbitrary, but it is convenient to take $p_\alpha=0$ as the identity (no translation or rotation). As an example, one can take

$$\mathbf{g}(\mathbf{r}, p) = \sum_{\mu} p_{\mu} \mathbf{e}_{\mu} + \hat{R}_x(p_4) \hat{R}_y(p_5) \hat{R}_z(p_6) \mathbf{r}, \quad (4)$$

where \mathbf{e}_{μ} , for $\mu=x, y, z$, are unit vectors along the x, y, z directions, and $\hat{R}_{\mu}(\phi)$ is a rotation around the μ axis over an angle ϕ , e.g., $\mathbf{r}' = \hat{R}_z(\phi) \mathbf{r}$ has components

$$x' = x \cos \phi - y \sin \phi; \quad y' = x \sin \phi + y \cos \phi; \quad z' = z. \quad (5)$$

The invariance properties of the potential energy surface are now simply expressed as

$$V(\{\mathbf{r}_A\}) = V(\{\mathbf{g}(\mathbf{r}_A, p)\}). \quad (6)$$

Taking the first- and second-order derivatives of Eq. (6) at a reference point $\{\mathbf{r}_A^0\}$ and at the identity ($p=0$), one finds

$$\frac{\partial}{\partial p_{\alpha}} \rightarrow 0 = \sum_{A\mu} G_{A\mu} D_{A\mu}^{(\alpha)}, \quad (7)$$

$$\begin{aligned} \frac{\partial^2}{\partial p_{\alpha} \partial p_{\alpha'}} \rightarrow 0 = & \sum_{A\mu, A'\mu'} H_{A\mu, A'\mu'} D_{A\mu}^{(\alpha)} D_{A'\mu'}^{(\alpha')} \\ & + \sum_{A\mu} G_{A\mu} C_{A\mu}^{(\alpha\alpha')}, \end{aligned} \quad (8)$$

$$\frac{\partial^2}{\partial p_{\alpha} \partial r_{A\mu}} \rightarrow 0 = \sum_{A'\mu'} H_{A\mu, A'\mu'} D_{A'\mu'}^{(\alpha)} + \sum_{\mu'} G_{A\mu'} \left(\frac{\partial D_{A\mu'}^{(\alpha)}}{\partial r_{A\mu}^0} \right), \quad (9)$$

in terms of the Cartesian gradient $G_{A\mu} = (\partial V / \partial r_{A\mu})_0$ and Hessian matrix $H_{A\mu, A'\mu'} = (\partial^2 V / \partial r_{A\mu} \partial r_{A'\mu'})_0$ at the reference point. The vectors $D^{(\alpha)}$ and $C^{(\alpha\alpha')}$ have components

$$D_{A\mu}^{(\alpha)} = \frac{\partial g_{\mu}}{\partial p_{\alpha}}(\mathbf{r}_A^0, 0), \quad (10)$$

$$C_{A\mu}^{(\alpha\alpha')} = \frac{\partial^2 g_{\mu}}{\partial p_{\alpha} \partial p_{\alpha'}}(\mathbf{r}_A^0, 0), \quad (11)$$

and are listed in Table I for the parameterization defined in Eq. (4).

Equation (9) implies that at an equilibrium point (where $G=0$) the $D^{(\alpha)}$ are six zero eigenvectors of the Hessian. At a nonequilibrium point, this is no longer the case, except when $(\partial D_{A\mu}^{(\alpha)} / \partial r_{A\mu}^0) = 0$. This is indeed the case for the translational subgroup (see Table I). So when the gradient is nonzero, the full Cartesian Hessian still has three zero eigenvectors from the translational invariance, but in general no zero eigenvectors related to the rotational invariance of V .

TABLE I. Derivatives of the transformation g for the parametrization defined in Eq. (4). The Levi-Civita symbol $\epsilon_{\lambda\mu\nu}$ equals 1 (-1) if $\lambda\mu\nu$ is a cyclic (anticyclic) permutation of xyz and zero otherwise.

α	$D_{A\mu}^{(\alpha)}$	α	$\frac{\partial D_{A\mu}^{(\alpha)}}{\partial r_{A\nu}^0}$	α	α'	$C_{A\mu}^{(\alpha\alpha')}$
1	$\delta_{\mu x}$	1	0	1	1-6	0
2	$\delta_{\mu y}$	2	0	2	1-6	0
3	$\delta_{\mu z}$	3	0	3	1-6	0
4	$\sum_{\lambda} \epsilon_{\lambda\mu\nu} r_{A\lambda}^0$	4	$\epsilon_{\nu\mu x}$	4	4	$\delta_{\mu x}^0 r_{A4}^0 - r_{A\mu}^0$
5	$\sum_{\lambda} \epsilon_{\lambda\mu\nu} r_{A\lambda}^0$	5	$\epsilon_{\nu\mu y}$	5	5	$\delta_{\mu y}^0 r_{A5}^0 - r_{A\mu}^0$
6	$\sum_{\lambda} \epsilon_{\lambda\mu\nu} r_{A\lambda}^0$	6	$\epsilon_{\nu\mu z}$	6	6	$\delta_{\mu z}^0 r_{A6}^0 - r_{A\mu}^0$
				5	5	$\delta_{\mu y}^0 r_{A5}^0 - r_{A\mu}^0$
				6	6	$\delta_{\mu z}^0 r_{A6}^0 - r_{A\mu}^0$
				6	6	$\delta_{\mu z}^0 r_{A6}^0 - r_{A\mu}^0$

III. MOBILE BLOCK HESSIAN: CASE OF A SINGLE BLOCK

A. Blocks with fixed geometry

For simplicity, we first consider the case where one MBH block is considered, consisting of N_F atoms which will be labeled $\{F\}$. The internal geometry of the block has been fixed (e.g., using a lower-level theoretical method) and is described by the configuration $\{\mathbf{r}_F^0\}$. Keeping the latter fixed, one can optimize the positions of the remaining $N_E = N - N_F$ atoms, which will be labeled $\{E\}$. This requires a partial optimization,

$$\frac{\partial V}{\partial r_{E\nu}^0}(\{\mathbf{r}_F^0\}, \{\mathbf{r}_E^0\}) = 0, \quad \forall E\nu. \quad (12)$$

The solution $\{\mathbf{r}_E^0\}$ determines the total reference structure of the system $\{\mathbf{r}_A^0\} = \{\mathbf{r}_F^0, \mathbf{r}_E^0\}$.

Obviously, the total system is not in equilibrium since $G_{E\mu} = (\partial V / \partial r_{E\mu}^0)_0 = 0$, but the gradients $G_{F\mu} \neq 0$ are in general nonzero. However, the MBH system with fixed internal geometry of the block has a reduced number, $3N_E + 6$, of degrees of freedom (6 coming from translation/rotation of the rigid-body motion of the block, and $3N_E$ of the remaining atoms $\{E\}$). The reduced MBH system is in an equilibrium state, and normal mode analysis can be applied without problems. In particular, the symmetry of the potential surface will give rise to six normal modes at zero energy, corresponding to global translations and rotations.

B. MBH normal modes in internal coordinates

In Ref. 13, the MBH normal mode equations were derived using internal coordinates, as this is the most transparent. The choice of $3N - 6$ internal coordinates $\{\theta_j\}$ is not completely arbitrary: in the present case of a single MBH block, one should arrange that $3N_F - 6$ of the internal coordinates $\{\theta_j\}$ describe the internal geometry of the atoms F in the block. The remaining $3N_E$ internal coordinates will be labeled $\{\theta_j\}$. Such an arrangement is always possible, e.g., using Z matrix internal coordinates with the atoms of the block numbered consecutively, as indicated in Fig. 1.

The fixed geometry of the block determines $\{\theta_j^0\}$. If the potential energy in internal coordinates is given by $W(\{\theta_j\})$, finding the MBH equilibrium geometry now requires a partial optimization,

$$\frac{\partial W}{\partial \theta_{jE}}(\{\theta_j^0\}, \{\theta_{jE}^0\}) = 0, \quad \forall \theta_{jE}, \quad (13)$$

and the solution $\{\theta_j^0\}$ determines the total reference structure $\{\theta_j^0\} = \{\theta_j^0, \theta_{jE}^0\}$.

With a body frame whose origin is at the center of mass $\mathbf{r}_{c.m.}$, the energy of Eq. (1) can be rewritten in terms of the internal coordinates $\{\theta_j\}$, the three components of $\mathbf{r}_{c.m.}$, and three angles specifying the orientation of the body frame, as

$$E = \frac{1}{2} \mathcal{M} \mathbf{r}_{c.m.}^2 + \frac{1}{2} \omega \cdot \bar{\mathbf{I}} \cdot \omega + \omega \cdot \sum_I \mathbf{A}_I \dot{\theta}_I + \frac{1}{2} \sum_{IJ} B_{IJ} \dot{\theta}_I \dot{\theta}_J + W(\{\theta_j\}). \quad (14)$$

Here \mathcal{M} is the total mass and ω the angular velocity vector of the body frame. The inertial tensor $\bar{\mathbf{I}}$, the Coriolis coupling \mathbf{A}_I between the body-frame rotation and internal velocity $\dot{\theta}_I$, and the B_{IJ} matrix are all functions of the internal coordinates.

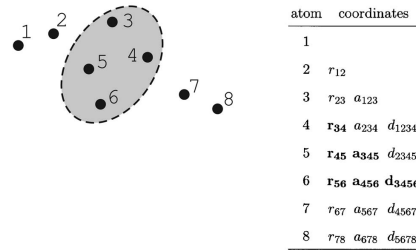


FIG. 1. System of eight atoms with a single block. In the Z -matrix construction of internal coordinates, the atoms of the block (3-6) should be numbered consecutively. Notation r is used for distances between atoms, a for angles, and d for dihedral angles. The Z coordinates indicated in boldface describe the internal geometry of the block and are kept fixed during the vibrational analysis.

Normal mode equations are generally obtained by expanding the energy in Eq. (1) up to quadratic terms in the displacements $\Delta_I = \theta_I - \theta_I^0$ and the velocities $\dot{\Delta}_I$, ω , and $\dot{\mathbf{r}}_{\text{c.m.}}$. In the MBH approach, the block is allowed to move as a rigid body during the small amplitude motion, with fixed geometry. This can now be simply expressed by setting the displacements $\Delta_I = 0$ and velocities $\dot{\Delta}_I = 0$. The quadratic expansion of the MBH potential energy, e.g., becomes

$$\begin{aligned} W(\{\theta_j\}) &\approx V_0 + \frac{1}{2} \sum_{I \in E} \left(\frac{\partial^2 W}{\partial \Delta_I \partial \Delta_I} \right)_0 \Delta_I \Delta_I \\ &= V_0 + \frac{1}{2} \sum_{I \in E} H_{I \in E}^{(ii)} \Delta_I \Delta_I. \end{aligned} \quad (15)$$

The corresponding MBH normal mode equations are just the standard ones in internal coordinates, but with the columns and rows corresponding to the fixed θ_F omitted from the Hessian and mass matrix,

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & H^{(ii)} \end{pmatrix} \begin{pmatrix} v^{(c)} \\ v^{(r)} \\ v^{(i)} \end{pmatrix} = \omega^2 \begin{pmatrix} M^{(cc)} & 0 & 0 \\ 0 & M^{(rr)} & M^{(ri)} \\ 0 & (M^{(ri)})^T & M^{(ii)} \end{pmatrix} \begin{pmatrix} v^{(c)} \\ v^{(r)} \\ v^{(i)} \end{pmatrix}, \quad (16)$$

where $v^{(c)}$, $v^{(r)}$, and $v^{(i)}$ have dimensions 3, 3, and $3N_E$, respectively, and the mass matrix entries read $M_{\mu\nu}^{(cc)} = \mathcal{M} \delta_{\mu\nu}$, $M_{\mu\nu}^{(rr)} = I_{\mu\nu}(\{\theta_j^0\})$, $M_{\mu\nu}^{(ri)} = A_{\mu\nu E}(\{\theta_j^0\})$, and $M_{\mu\nu}^{(ii)} = B_{I \in E}(\{\theta_j^0\})$. The MBH eigenvalue problem of dimension $3N_E + 6$ still has six zero eigenvalues corresponding to overall translation and rotation. These can be decoupled in the usual way¹⁴⁻¹⁶ by a congruent transformation, $\tilde{v}^{(c)} = v^{(c)}$, $\tilde{v}^{(i)} = v^{(i)}$, $\tilde{v}^{(r)} = v^{(r)} + (M^{(rr)})^{-1} M^{(ri)} v^{(i)}$, yielding the final $3N_E$ -dimensional normal mode equation,

$$H^{(ii)} \tilde{v}^{(i)} = \omega^2 [M^{(ii)} - (M^{(ri)})^T (M^{(rr)})^{-1} M^{(ri)}] \tilde{v}^{(i)}. \quad (17)$$

If one has access to selected Hessian matrix elements in the chosen internal coordinates, this is obviously the most straightforward way of implementing MBH. However, the Hessian is commonly given in Cartesian coordinates, and it is of interest to analyze the elements of $H^{(ii)}$ in terms of these. At first sight, one may think that *all* Cartesian Hessian elements are needed (including those involving the atoms $\{F\}$ of the fixed block) for the evaluation of $H^{(ii)}$, since a displacement Δ_E can change the position or orientation of the block with respect to a space-fixed frame. In fact, it is always possible to use the global translational/rotational invariance of the potential surface to rewrite $H^{(ii)}$ in terms of the Cartesian Hessian matrix elements of the relaxed atoms $\{E\}$.

This was already noted in Ref. 13, where a particular set of internal coordinates $\{\theta_E\}$ was used (the Cartesian coordinates of the atoms $\{E\}$ in a frame attached to the rigid block). We now show that this holds for the most general choice of internal coordinates $\{\theta_j\}$, provided that $3N_F - 6$ of them determine the geometry of the block.

Suppose that for each shape $\{\theta_j\}$ one has selected a configuration $\{\mathbf{r}_A^{\text{bf}}(\{\theta_j\})\}$ in coordinate space. This corresponds to the selection of a body frame.¹⁷ Then the additional global translational/rotational coordinates can be defined as the parameters p of the symmetry transformation needed to let the actual configuration in the space-fixed frame coincides with the body-frame configuration,

$$\mathbf{r}_A = \mathbf{g}(\mathbf{r}_A^{\text{bf}}(\{\theta_j\}), p), \quad (18)$$

and the potential energy becomes

$$V(\{\mathbf{r}_A\}) = V(\{\mathbf{g}(\mathbf{r}_A^{\text{bf}}(\{\theta_j\}), p)\}) = V(\{\mathbf{r}_A^{\text{bf}}(\{\theta_j\})\}) \equiv W(\{\theta_j\}). \quad (19)$$

When describing the reduced (MBH) system, the $\{\theta_F^0\}$ are fixed. It is then always possible to attach the body frame to the fixed MBH block. As a consequence, the body-frame coordinates of the $\{F\}$ atoms are constant,

$$\frac{\partial \mathbf{r}_F^{\text{bf}}(\{\omega_F^0\}, \{\theta_E\})}{\partial \theta_E} = 0, \quad (20)$$

and the Hessian $H^{(ii)}$ in Eq. (15) simply becomes

$$H_{I \in E}^{(ii)} = \sum_{E', \mu', E'', \mu''} \left(\frac{\partial^2 V}{\partial r_{E', \mu'} \partial r_{E'', \mu''}} \right)_0 \left(\frac{\partial r_{E', \mu'}^{\text{bf}}}{\partial \theta_I} \right)_0 \left(\frac{\partial r_{E'', \mu''}^{\text{bf}}}{\partial \theta_I} \right)_0, \quad (21)$$

where only Cartesian matrix elements involving the $\{E\}$ atoms are needed in the summation.

C. Cartesian formulation of single-block MBH using group coordinates

In the following section, we develop an alternative formulation of the MBH normal mode equations, suitable when only Cartesian (Hessian and gradient) information is available. It has the advantage that no system of internal coordinates and body frame needs to be specified. This avoids the evaluation of the body-frame coordinate derivatives in Eq. (21), and facilitates the automatization of the input for the numerical code. The method was used in the practical calculations of Ref. 13, as it can be very easily extended to the case of multiple blocks (see Sec. IV).

The movement of the block as a rigid body can be described in a natural way by treating the six group parameters in Eq. (2) as dynamical variables, such that the instantaneous position of each atom F in the block is given by the result of a common translation/rotation of the reference position $\{\mathbf{r}_F^0\}$ in the block,

$$\mathbf{r}_F(t) = \mathbf{g}(\mathbf{r}_F^0, p(t)). \quad (22)$$

The velocities of the N_F atoms in the block are

$$\dot{r}_{F\mu} = \sum_{\alpha} \frac{\partial g_{\mu}}{\partial p_{\alpha}}(\mathbf{r}_F^0, p) \dot{p}_{\alpha}. \quad (23)$$

In order to link the MBH normal mode equations in Eq. (17) with the Cartesian Hessian, we will therefore choose as variables the six parameters p_{α} , combined with the $3N_E$ Car-

tesian coordinates $\{\mathbf{r}_E\}$ of the remaining atoms. The potential energy expressed in the new coordinates becomes

$$\tilde{V}(p, \{\mathbf{r}_E\}) = V(\{\mathbf{g}(\mathbf{r}_F, p)\}, \{\mathbf{r}_E\}). \quad (24)$$

Obviously, the partially optimized MBH reference structure $\{\mathbf{r}_A^0\}$, determined in Sec. III A, corresponds to $p=0$ and $\{\mathbf{r}_E^0\}$ in the present coordinates. At this reference point, the first derivatives of \tilde{V} vanish,

$$\tilde{G}_\alpha = \left(\frac{\partial \tilde{V}}{\partial p_\alpha} \right)_0 = \sum_{F\mu} G_{F\mu} D_{F\mu}^{(\alpha)} = 0, \quad (25)$$

$$\tilde{G}_{E\mu} = \left(\frac{\partial \tilde{V}}{\partial r_{E\mu}} \right)_0 = G_{E\mu} = 0, \quad (26)$$

as can also be seen from Eqs. (7) and (12).

Global translational/rotational invariance can now be expressed by applying the same transformation P to both the block and to the atoms $\{E\}$,

$$\tilde{V}(p, \{\mathbf{r}_E\}) = \tilde{V}(\Phi(p, P), \{\mathbf{g}(\mathbf{r}_E, P)\}). \quad (27)$$

In order to calculate derivatives of Eq. (27) one needs an elementary property of the group multiplication law $\Phi(p, P)$ near the identity. Since $\Phi(p, 0)=p$ and $\Phi(0, P)=P$, one has

$$\frac{\partial \Phi_\beta}{\partial p_\alpha}(0, 0) = \delta_{\alpha, \beta} = \frac{\partial \Phi_\beta}{\partial P_\alpha}(0, 0). \quad (28)$$

Calculating the derivatives of Eq. (27) at the reference point and at $P=0$, one now finds

$$\frac{\partial^2}{\partial P_\alpha \partial p_{\alpha'}} \rightarrow \tilde{H}_{\alpha, \alpha'} + \sum_{E\mu} \tilde{H}_{E\mu, \alpha'} D_{E\mu}^{(\alpha)} = 0, \quad (29)$$

$$\frac{\partial^2}{\partial P_\alpha \partial r_{E\mu}} \rightarrow \tilde{H}_{\alpha, E\mu} + \sum_{E'\mu'} \tilde{H}_{E\mu, E'\mu'} D_{E'\mu'}^{(\alpha)} = 0, \quad (30)$$

where we used the obvious notation

$$\tilde{H}_{\alpha, \alpha'} = \left(\frac{\partial^2 W}{\partial p_\alpha \partial p_{\alpha'}} \right)_0, \quad (31)$$

$$\tilde{H}_{\alpha, E\mu} = \left(\frac{\partial^2 W}{\partial p_\alpha \partial r_{E\mu}} \right)_0, \quad (32)$$

$$\tilde{H}_{E\mu, E'\mu'} = \left(\frac{\partial^2 W}{\partial r_{E\mu} \partial r_{E'\mu'}} \right)_0. \quad (33)$$

It is now easy to check from Eqs. (29) and (30) that the six vectors $v^{(\alpha)}$, with components

$$v_{\alpha'}^{(\alpha)} = \delta_{\alpha, \alpha'}; \quad v_{E\mu}^{(\alpha)} = D_{E\mu}^{(\alpha)}, \quad (34)$$

are eigenvectors of the Hessian \tilde{H} with zero eigenvalue.

The normal mode equations in the group coordinates are of dimension $3N_E+6$ and read in standard form,

$$\tilde{H}v = \omega^2 \tilde{M}v, \quad (35)$$

with \tilde{M} the corresponding mass matrix, to be derived from the kinetic energy. Expanding around the total reference structure $\{\mathbf{r}_A^0\}$, the kinetic energy [see Eq. (23)] is

$$T = \frac{1}{2} \sum_{F\mu} m_F \left(\sum_\alpha D_{F\mu}^{(\alpha)} p_\alpha \right)^2 + \frac{1}{2} \sum_E m_E \dot{\mathbf{r}}_E^2. \quad (36)$$

This determines the mass matrix components,

$$\tilde{M}_{\alpha, \alpha'} = \sum_{F\mu} m_F D_{F\mu}^{(\alpha)} D_{F\mu}^{(\alpha')}, \quad (37)$$

$$\tilde{M}_{\alpha, E\mu} = 0, \quad (38)$$

$$\tilde{M}_{E\mu, E'\mu'} = \delta_{E, E'} \delta_{\mu, \mu'} m_E. \quad (39)$$

Note that the mass matrix is block diagonal in the block structure induced by the present choice of the six coordinates p_α and $3N_E$ coordinates $r_{E\mu}$.

The six zero eigenvectors in Eq. (34) can now be decoupled from the intrinsic normal modes, using congruent transformations to affect a simultaneous block diagonalization of \tilde{H} and \tilde{M} . The required transformation matrices are then given by

$$T_1 = \begin{pmatrix} 1_{6 \times 6} & 0_{6 \times d} \\ x & 1_{d \times d} \end{pmatrix} = (u \ w); \quad T_2 = \begin{pmatrix} 1_{6 \times 6} & y \\ 0_{d \times 6} & 1_{d \times d} \end{pmatrix}, \quad (40)$$

where the dimensions of the identity and zero submatrices are explicitly indicated, and with the dimension $d=3N_E$. When the first transformation matrix is determined by the submatrix $x_{E\mu, \alpha} = D_{E\mu}^{(\alpha)}$, the first six columns of T_1 become the zero eigenvectors in Eq. (34), collected in the matrix u . The transformed Hessian,

$$T_1^T \tilde{H} T_1 = \begin{pmatrix} 0_{6 \times 6} & 0_{6 \times d} \\ 0_{d \times 6} & \tilde{H}' \end{pmatrix}, \quad (41)$$

has zero at its first six rows and columns, while the remainder is left unchanged,

$$\tilde{H}'_{E\mu, E'\mu'} = \tilde{H}_{E\mu, E'\mu'}. \quad (42)$$

The second transformation leaves the Hessian matrix in Eq. (41) intact, and can be used to block diagonalize the mass matrix. This requires the choice

$$y_{\alpha, E\mu} = - \sum_{\alpha'} [S^{-1}]_{\alpha, \alpha'} D_{E\mu}^{(\alpha')} m_E, \quad (43)$$

or, in matrix notation,

$$y = -S^{-1} u^T \tilde{M} w. \quad (44)$$

The matrix S has components

$$S_{\alpha, \alpha'} = \sum_{A\mu} m_A D_{A\mu}^{(\alpha)} D_{A\mu}^{(\alpha')}, \quad (45)$$

and contains the information (mass and inertial tensor) for rigid-body motion of the global system in the reference configuration $\{\mathbf{r}_A^0\}$. One can now verify that the transformed

Hessian and mass matrix have a decoupled form,

$$\begin{aligned} T_2^T T_1^T \tilde{H} T_1 T_2 &= \begin{pmatrix} 0_{6 \times 6} & 0_{6 \times d} \\ 0_{d \times 6} & \tilde{H}' \end{pmatrix}, \\ T_2^T T_1^T \tilde{M} T_1 T_2 &= \begin{pmatrix} S & 0_{6 \times d} \\ 0_{d \times 6} & \tilde{M}' \end{pmatrix}, \end{aligned} \quad (46)$$

and the final normal mode equations for the $3N_E$ intrinsic modes are

$$\tilde{H}' v' = \omega^2 \tilde{M}' v', \quad (47)$$

where the correct mass matrix is given by

$$\tilde{M}'_{E\mu, E'\mu'} = \delta_{E, E'} \delta_{\mu, \mu'} m_E - \sum_{\alpha\alpha'} m_E m_{E'} D_{E\mu}^{(\alpha)} [S^{-1}]_{\alpha, \alpha'} D_{E'\mu'}^{(\alpha')}, \quad (48)$$

or, in matrix notation,

$$\tilde{M}' = w^T \tilde{M} w - w^T \tilde{M} u S^{-1} u^T \tilde{M} w. \quad (49)$$

We still need to express \tilde{H} in terms of Cartesian quantities. A direct evaluation starting from Eq. (24) yields

$$\tilde{H}_{\alpha, \alpha'} = \sum_{F\mu, F'\mu'} H_{F\mu, F'\mu'} D_{F\mu}^{(\alpha)} D_{F'\mu'}^{(\alpha')} + \sum_{F\mu} G_{F\mu} C_{F\mu}^{(\alpha\alpha')}, \quad (50)$$

$$\tilde{H}_{\alpha, E\mu} = \sum_{F'\mu'} H_{E\mu, F'\mu'} D_{F'\mu'}^{(\alpha)}, \quad (51)$$

$$\tilde{H}_{E\mu, E'\mu'} = H_{E\mu, E'\mu'}. \quad (52)$$

In agreement with the discussion in Sec. III B, it is clear that only the Cartesian Hessian matrix elements of the $\{E\}$ atoms are needed for the normal mode equations in Eq. (47). In fact, for a single MBH block, one can even show that the other Hessian elements $\tilde{H}_{\alpha, E\mu}$ and $\tilde{H}_{\alpha, \alpha'}$ can be rewritten solely in terms of the $H_{E\mu, E'\mu'}$. This is indeed the case, since Eqs. (8) and (9) imply that zero is obtained when the summations over the $\{F\}$ atoms in Eqs. (50) and (51) are extended over *all* atoms $\{A\}$. As a consequence,

$$\tilde{H}_{\alpha, \alpha'} = \sum_{E\mu, E'\mu'} H_{E\mu, E'\mu'} D_{E\mu}^{(\alpha)} D_{E'\mu'}^{(\alpha')}, \quad (53)$$

$$\tilde{H}_{\alpha, E\mu} = - \sum_{E'\mu'} H_{E\mu, E'\mu'} D_{E'\mu'}^{(\alpha)}. \quad (54)$$

IV. MOBILE BLOCK HESSIAN: CASE OF MULTIPLE BLOCKS

A. Blocks with fixed geometry in internal coordinates

We now discuss the extension to the case where multiple blocks with fixed geometry are present, each of which can move as a rigid body. The blocks will be labeled $b = 1, \dots, K$, where block b contains N_{F_b} atoms labeled $\{F_b\}$. As before, the remaining N_E atoms (which can move freely) are labeled $\{E\}$, and $N_E + \sum_b N_{F_b} = N$, where N is the total number of atoms in the molecule.

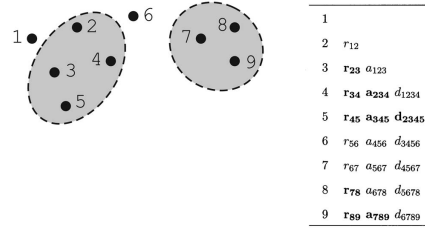


FIG. 2. System of nine atoms with two MBH blocks. In the Z-matrix construction, the atoms of each block (2–5 and 7–9) should be numbered consecutively. The Z coordinates in boldface are kept fixed during the vibrational analysis.

Again, the most transparent derivation of the MBH approach proceeds through the introduction of $3N-6$ internal coordinates $\{\theta_i\}$. One now should take care that the internal geometry of the atoms in each block b is described by $3N_{F_b}-6$ internal coordinates $\{\theta_{F_b}\}$. The fixed geometry of the blocks is known beforehand, and given by $\{\theta_{F_b}^0\}$. The remaining $3N_E+6(K-1)$ coordinates $\{\theta_{E_i}\}$ are determined by partial optimization of the potential energy $W(\{\theta_i\})$:

$$\frac{\partial W}{\partial \theta_{E_i}}(\{\theta_{F_b}^0\}, \{\theta_{E_i}\}) = 0, \quad \forall \theta_{E_i}, \quad (55)$$

and the solution $\{\theta_{E_i}^0\}$ determines the total reference structure $\{\theta_i^0\} = \{\theta_{F_b}^0, \theta_{E_i}^0\}$.

We emphasize that a suitable set of internal coordinates can always be constructed. In the Z-matrix formalism, e.g., one should only take care that the numbering of the atoms occurs consecutively in each block, as indicated in the example of Fig. 2.

The same reasoning as in Sec. III B leads to the conclusion that the normal mode equations of the multiple-block MBH system are identical to the single-block equations in Eq. (16) or (17), except that now the rows and columns corresponding to all the fixed $\{\theta_{F_b}\}$ should be omitted from the Hessian and mass matrix.

If one has access to the selected Hessian matrix elements $(\partial^2 W / \partial \theta_{E_i} \partial \theta_{E_j})_0$ in internal coordinates, this is the most economical method (in terms of Hessian elements evaluations). If only Cartesian Hessian matrix elements are available, we will show that it is possible to generalize the description of Sec. III C in terms of group coordinates, thereby avoiding the construction of internal coordinates altogether.

B. Cartesian formulation of multiple-block MBH using group coordinates

The Cartesian formulation for one single block can easily be extended to multiple blocks in a straightforward way. Most of the expressions (22)–(54) remain unaffected apart from an additional index b running over the various blocks.

The potential energy \tilde{V} now becomes a function of the $6K$ block parameters $\{p_b\}$, combined with the Cartesian co-

ordinates $\{\mathbf{r}_E\}$ of the remaining atoms. For completeness, we mention that one can make a distinction between a “normal” block and a block of collinear atoms, where the latter only requires five group parameters instead of six. For the sake of brevity, this distinction is not made here. Moreover, the final normal mode equations apply also in the presence of a “collinear block,” except that one additional zero eigenvalue appears corresponding to rotation over the symmetry axis of the block.

The global translational/rotational invariance of \tilde{V} is expressed by applying the same transformation P to all the blocks and the atoms $\{E\}$,

$$\tilde{V}(\{p_b\}, \{\mathbf{r}_E\}) = \tilde{V}(\{\Phi(p_b, P)\}, \{\mathbf{g}(\mathbf{r}_E, P)\}). \quad (56)$$

The ensuing consequences for the Hessian are readily derived by calculating the derivatives of Eq. (56) at the reference point and at $P=0$:

$$\frac{\partial^2}{\partial P_a \partial p_{b' \alpha'}} \rightarrow \sum_b \tilde{H}_{b\alpha, b' \alpha'} + \sum_{E\mu} \tilde{H}_{E\mu, b' \alpha'} D_{E\mu}^{(\alpha)} = 0, \quad (57)$$

$$\frac{\partial^2}{\partial P_a \partial r_{E\mu}} \rightarrow \sum_b \tilde{H}_{b\alpha, E\mu} + \sum_{E' \mu'} \tilde{H}_{E\mu, E' \mu'} D_{E' \mu'}^{(\alpha)} = 0, \quad (58)$$

using obvious notations for $\tilde{H}_{b\alpha, b' \alpha'}$, $\tilde{H}_{b\alpha, E\mu}$, and $\tilde{H}_{E\mu, E' \mu'}$ similar to those of Eqs. (31)–(33).

We see that the extension to multiple blocks is very straightforward: an extra index b indicates the block under consideration. In the kinetic energy expression [see Eq. (36)], an obvious summation over the multiple blocks b is introduced. It determines the mass matrix elements $\tilde{M}_{b\alpha, E\mu}$, $\tilde{M}_{E\mu, E' \mu'}$, and $\tilde{M}_{b\alpha, b' \alpha'}$. The first two are similar to Eqs. (38) and (39), while the last one is diagonal in the block index b ,

$$\tilde{M}_{b\alpha, b' \alpha'} = \delta_{b, b'} \sum_{F_b \mu} m_{F_b} D_{F_b \mu}^{(\alpha)} D_{F_b \mu}^{(\alpha')}. \quad (59)$$

The normal mode equations are formally identical to Eq. (35), and are of dimension $6K+3N_E$. Inspection of Eqs. (57) and (58) indicates the form of the six eigenvectors $v^{(\alpha)}$ with zero eigenvalue, which express the translational/rotational invariance, as in Eq. (34). Elimination of these six zero modes $v^{(\alpha)}$ proceeds in much the same way as in the case of a single block. We (arbitrarily) select one of the blocks, B , and label this as the first block. The required transformation matrices T_1 and T_2 are then again given by Eq. (40) with the dimension $d=6(K-1)+3N_E$. The submatrix x in the first transformation has dimension $d \times 6$ and must be such that the first six columns of T_1 are the eigenvectors with zero eigenvalues. This requires $x_{b' \alpha', B\alpha} = \delta_{\alpha, \alpha'}$ for $b'=2, \dots, K$ (or equivalently $b' \neq B$) and $x_{E\mu, B\alpha} = D_{E\mu}^{(\alpha)}$. The transformed Hessian $T_1^T \tilde{H} T_1$ now has zero at its first six rows and columns, while the remainder is left unchanged. The second transformation T_2 leaves $T_1^T \tilde{H} T_1$ intact. Block diagonalization of the mass matrix is obtained with the choice

$$y_{B\alpha, b' \alpha'} = - \sum_{\alpha'} [S^{-1}]_{\alpha, \alpha'} \sum_{F_{b' \mu}} D_{F_{b' \mu}}^{(\alpha)} D_{F_{b' \mu}}^{(\alpha')} m_{F_{b' \mu}}, \quad (60)$$

for $b' \neq B$,

$$y_{B\alpha, E\mu} = - \sum_{\alpha'} [S^{-1}]_{\alpha, \alpha'} D_{E\mu}^{(\alpha')} m_{E\mu}, \quad (61)$$

or in matrix notation $y = -S^{-1} u^T \tilde{M} w$, where the 6×6 matrix S is given by Eq. (45). One can now verify that the transformed Hessian and mass matrix have the same decoupled form as in Eq. (46). The final normal mode equations for the d intrinsic modes are again formally identical to Eq. (47). The correct mass matrix is given by $(b, b' \neq B)$

$$\tilde{M}'_{b\alpha, b' \alpha'} = \tilde{M}_{b\alpha, b' \alpha'} - \sum_{\beta\beta'} [S^{-1}]_{\beta, \beta'} \tilde{M}_{b\alpha, \beta\beta'} \tilde{M}_{\beta\beta', b' \alpha'}, \quad (62)$$

$$\tilde{M}'_{b\alpha, E\mu} = 0, \quad (63)$$

$$\tilde{M}'_{E\mu, E' \mu'} = \delta_{E, E'} \delta_{\mu, \mu'} m_E - \sum_{\alpha\alpha'} m_E m_{E'} D_{E\mu}^{(\alpha)} [S^{-1}]_{\alpha, \alpha'} D_{E' \mu'}^{(\alpha')}. \quad (64)$$

In matrix notation, these expressions equal Eq. (49).

Note that the reduced Hessian \tilde{H}' in the intrinsic normal mode equations has identical components as \tilde{H} , but the sub-space involving the block B has been eliminated. It is again straightforward to express \tilde{H} in terms of Cartesian quantities, with $\tilde{H}_{b\alpha, E' \mu'}$ and $\tilde{H}_{E\mu, E' \mu'}$ as in Eqs. (51) and (52), and

$$\begin{aligned} \tilde{H}_{b\alpha, b' \alpha'} &= \sum_{F_b \mu, F_{b'} \mu'} H_{F_b \mu, F_{b'} \mu'} D_{F_b \mu}^{(\alpha)} D_{F_{b'} \mu'}^{(\alpha')} \\ &+ \delta_{b, b'} \sum_{F_b \mu} G_{F_b \mu} C_{F_b \mu}^{(\alpha\alpha')}. \end{aligned} \quad (65)$$

It is now clear that, in order to evaluate the reduced Hessian \tilde{H}' , one only needs the Cartesian Hessian and gradient involving the blocks $b \neq B$, as well as the atoms $\{E\}$. In practice, it is therefore advantageous to pick the largest block, as the block B that is eliminated.

As a final remark, we note that it is possible to rewrite the Cartesian gradient term in Eq. (65) in favor of Cartesian Hessian elements. Using the expressions for the zero eigenvectors of \tilde{H} , one finds

$$\begin{aligned} \sum_{F_b \mu} G_{F_b \mu} C_{F_b \mu}^{(\alpha\alpha')} &= - \sum_{b'} \sum_{F_{b' \mu}, F_{b' \mu}'} H_{F_{b' \mu}, F_{b' \mu}'} D_{F_{b' \mu}}^{(\alpha)} D_{F_{b' \mu}'}^{(\alpha')} \\ &- \frac{1}{2} \sum_{F_b \mu, E\nu} H_{F_b \mu, E\nu} (D_{F_b \mu}^{(\alpha)} D_{E\nu}^{(\alpha')} \\ &+ D_{F_b \mu}^{(\alpha')} D_{E\nu}^{(\alpha)}), \end{aligned} \quad (66)$$

where the right-hand side has been symmetrized in α and α' , to reduce numerical errors. In practice, it is more convenient to use Eq. (65) since Eq. (66) requires *all* Cartesian Hessian elements, and the previous computational advantage of eliminating the elements of a selected block is lost.

V. PRACTICAL PROCEDURE FOR IMPLEMENTING MBH

For clarity, we summarize here the practical procedure for the MBH scheme for partially optimized systems. The example of Fig. 2 will be used to illustrate the procedure.

- (1) *Block selection.* Choose blocks and number the atoms of the blocks consecutively. In Fig. 2, two blocks are defined: atoms 2–5 and atoms 7–9.
- (2) *Partial optimization.* Optimize the system while keeping the internal geometry of the blocks fixed. In Fig. 2, the internal coordinates in boldface are kept fixed during the optimization [this option is, e.g., provided in the modeling package GAUSSIAN03 (Ref. 18)]. Note that one can as well keep the atoms of block 1 at their absolute positions, while internal coordinates should be fixed for the other blocks. This is equivalent, as we are interested in the partially optimized *internal* geometry and not the absolute space frame coordinates. In Fig. 2, for instance, the positions \mathbf{r}_2 – \mathbf{r}_5 and the internal coordinates r_{78} , r_{89} , a_{789} of the second block can be held constant.
- (3) *Calculation of second derivatives.* The second derivatives at the partially optimized (reference) structure should be calculated at the same level of theory as used in the partial optimization procedure.
 - If the Hessian in internal coordinates is available, construct $H^{(ii)}$ of Eq. (16) by omitting the rows and columns of the Hessian corresponding to fixed internal coordinates. The MBH normal modes follow from solving Eq. (17). Note that it is of computational profit to calculate, if possible, only the Hessian elements corresponding to the $\{\theta_{iE}\}$ coordinates.
 - If only the Cartesian Hessian (and gradient) is available, first calculate the coefficients $D_{A\mu}^{(a)}$ and $C_{A\mu}^{(aa')}$ using Table I. Construct \tilde{M} using Eqs. (38), (39), and (59). Construct \tilde{H} using Eqs. (51), (52), and (65). The MBH normal modes follow from solving Eq. (35). Note that one can obtain a reduction in the number of required Cartesian Hessian matrix elements by considering the submatrix \tilde{H}' of \tilde{H} , in which the six rows/columns involving one block B are omitted. The reduction is considerable if the block B is large. The corresponding mass matrix \tilde{M}' is given by Eqs. (62)–(64). The MBH normal modes then follow from solving Eq. (47).
 - If it is possible to calculate second derivatives in arbitrary coordinates (e.g., using finite difference approximation), one may choose the set $\{p_b, \mathbf{r}_{iE}\}$ that includes the group parameters for the blocks, and calculate \tilde{H} or \tilde{H}' directly.

VI. CONCLUSION

Partially optimized systems require an adapted vibrational analysis, since the standard normal mode analysis ap-

plied to nonoptimized structures yields unphysical results. Therefore, the MBH approach was developed, which introduces rigid blocks in the vibrational analysis. This paper concentrates on the formulation of the MBH model in three sets of coordinates: internal coordinates, Cartesian coordinates, and a set including group parameters. A practical calculation scheme for all cases was summarized in the previous section, to allow a simple implementation in existing simulation packages.

Concerning the computational cost of the MBH model, the first reduction in computer time is, of course, the partial optimization instead of an expensive full optimization. For a system with N atoms, treated in terms of K blocks and N_E freely moving atoms, only $d=6(K-1)+3N_E$ coordinates have to be optimized. When one has the freedom to choose which Hessian elements are calculated in the frequency analysis, the MBH approach is able to reduce the number of Hessian elements significantly as well.

In internal coordinates, one needs the second derivatives with respect to d coordinates $\{\theta_{iE}\}$ only, where usually $d \ll 3N$, or the MBH approach is much less expensive than the full Hessian calculation.

A very natural description of the fixed block dynamics was obtained by introducing group parameters for the six translational/rotational degrees of freedom of each block. This also avoids the explicit construction of internal coordinates. The complete set of $d+6$ coordinates $\{p_b, \mathbf{r}_{iE}\}$ consists of the group parameters $\{p_b\}$ and the Cartesian coordinates $\{\mathbf{r}_{iE}\}$ of the relaxed atoms. Since the translational/rotational invariance is fully respected, six modes with zero eigenfrequency are found which can be decoupled easily from the intrinsic modes. These symmetries also allow to eliminate the group parameters of one arbitrary block B . As a result, the normal mode equations now require the second derivatives with respect to d variables $\{p_b, \mathbf{r}_{iE}\}_{b \neq B}$.

The formulation of the MBH approach in Cartesian coordinates is of great practical importance. The implementation is perfectly possible but, at first sight, the computational profit of the MBH method is lost, since all Cartesian Hessian elements seem to be needed. Nevertheless, even in Cartesian coordinates, a certain computer time reduction can be realized. Using symmetry arguments, it is possible to eliminate the Cartesian components related to the atoms of one arbitrary block B . By choosing this block B to be the biggest one, the largest reduction in computer time is obtained: only second derivatives with respect to $3N-3N_{F_B}$ variables $\{r_{A\mu}\}_{A \neq B}$ are required. Note that in the case of a single block ($K=1$), this implies that the MBH equations are constructed uniquely with Cartesian Hessian elements of the free atoms, and are computationally equally demanding as the PHVA method.

We remark that the MBH model aims at describing selected modes of interest, for which one assumes that they do not involve changes in the geometry of the fixed blocks. What still missing is the coupling between the normal modes of interest and those that were eliminated by fixing the block geometry. An interesting extension in the future would be to estimate this coupling in perturbation theory, using methods similar to the ones described in Refs. 10–12.

ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research—Flanders and the Research Board of Ghent University.

- ¹N. A. Besley and K. A. Metcalf, *J. Chem. Phys.* **126**, 035101 (2007).
- ²D. Lesthaeghe, G. Delcour, V. Van Speybroeck, G. Marin, and M. Waroquier, *Microporous Mesoporous Mater.* **96**, 350 (2006).
- ³A. Tachibana and K. Fukui, *Theor. Chim. Acta* **49**, 321 (1978).
- ⁴R. Murry, J. T. Fourkas, L. Wu-Xiong, and T. Keyes, *J. Chem. Phys.* **110**, 10410 (1999).
- ⁵D. J. Wales, *J. Chem. Phys.* **113**, 3926 (2000).
- ⁶I. Kolossvary and C. McMartin, *J. Math. Chem.* **9**, 359 (1992).
- ⁷S. Q. Jin and J. D. Head, *Surf. Sci.* **318**, 204 (1994).
- ⁸M. D. Calvin, J. D. Head, and S. Q. Jin, *Surf. Sci.* **345**, 161 (1996).
- ⁹H. Li and J. H. Jensen, *Theor. Chem. Acc.* **107**, 211 (2002).
- ¹⁰J. D. Head, *Int. J. Quantum Chem.* **65**, 827 (1997).
- ¹¹J. D. Head and Y. Shi, *Int. J. Quantum Chem.* **75**, 815 (1999).
- ¹²J. D. Head, *Int. J. Quantum Chem.* **77**, 350 (2000).
- ¹³A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, *J. Chem. Phys.* **126**, 224102 (2007).
- ¹⁴V. Van Speybroeck, D. Van Neck, and M. Waroquier, *J. Phys. Chem. A* **104**, 10939 (2000).
- ¹⁵V. Van Speybroeck, P. Vansteenkiste, D. Van Neck, and M. Waroquier, *Chem. Phys. Lett.* **402**, 479 (2005).
- ¹⁶P. Vansteenkiste, D. Van Neck, V. Van Speybroeck, and M. Waroquier, *J. Chem. Phys.* **124**, 044314 (2006).
- ¹⁷R. G. Littlejohn and M. Reinsch, *Rev. Mod. Phys.* **69**, 213 (1997).
- ¹⁸M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant *et al.*, GAUSSIAN 03, Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.

Paper III

Calculating reaction rates with partial Hessians: validation of the Mobile Block Hessian approach

A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck and
M. Waroquier

Journal of Chemical Theory and Computation 4 (4), 614-625
(2008)

Copyright 2008 by American Chemical Society

**Calculating Reaction Rates with Partial Hessians:
Validation of the Mobile Block Hessian Approach**

A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier*

*Center for Molecular Modeling, Ghent University, Proeftuinstraat 86,
B-9000 Gent, Belgium*

Received October 24, 2007

Abstract: In an earlier paper, the authors have developed a new method, the mobile block Hessian (MBH), to accurately calculate vibrational modes for partially optimized molecular structures [*J. Chem. Phys.* **2007**, *126* (22), 224102]. The proposed procedure remedies the artifact of imaginary frequencies, occurring in standard frequency calculations, when parts of the molecular system are optimized at different levels of theory. Frequencies are an essential ingredient in predicting reaction rate coefficients due to their input in the vibrational partition functions. The question arises whether the MBH method is able to describe the chemical reaction kinetics in an accurate way in large molecular systems where a full quantum chemical treatment at a reasonably high level of theory is unfeasible due to computational constraints. In this work, such a validation is tested in depth. The MBH method opens a lot of perspectives in predicting accurate kinetic parameters in chemical reactions where the standard full Hessian procedure fails.

1. Introduction

Ab initio prediction of reaction rate constants of chemical reactions has a high computational cost, especially when large (bio)molecular systems are involved. An accurate description of chemical kinetics of reactions in gas phase is nowadays perfectly practicable for moderate-sized molecules, but once the molecular environment comes into play, one has to adapt the level of theory in such way to make the computation feasible.¹ This puts a heavy burden on the accuracy of the numerical results. Chemical kinetics in static approaches is still widely based on transition state theory (TST).^{2–5} Key parameters are the reaction energy barrier between the reactants and activated complex (the transition state) and the vibrational frequencies, which serve as an input in the partition functions, and their accurate computation is essential. In the molecular-statistical formulation of TST, they completely determine the equilibrium constant by the use of partition functions.

In the harmonic oscillator approximation, the molecular partition function is factorized in a translational, rotational, and vibrational contribution, where the latter is completely

determined by the eigenfrequencies. Frequencies are usually computed by a normal-mode analysis (NMA). This is the main bottleneck in ab initio predictions of chemical kinetics in large molecular systems, since frequency calculations are computationally very demanding even if analytical second derivatives are employed, rather than numerical ones. If a molecular mechanics (MM) force field is used instead of a quantum mechanics (QM) or hybrid (QM/MM)^{6–9} description, the frequency calculation becomes less problematic, though even at the full MM level other issues, such as the storage and manipulation of the huge Hessian matrices associated with very large systems, can become prohibitive in real applications. Anyway, chemical reactions inherently involve bond breaking and charge transfer; so, it is essential to provide a QM description for (at least) the reactive region and a full MM description is usually no option.

In addition, there are computational limitations in the geometry optimization of extended systems at a high level of theory (LOT). Very often one goes over to a partial optimization: the interesting region containing the active site is optimized at a high LOT, while the environment is kept fixed at a low LOT geometry. This approach permits one to obtain an ab initio description of the chemically active site

* Corresponding author. E-mail: michel.waroquier@UGent.be.

in large molecular systems, but at the same time, it creates several new problems. One of them is the extraction of accurate frequencies for the relevant vibrational modes. All partially optimized systems are nonequilibrium structures, and as a consequence of the residual gradients on the potential energy surface (PES), the standard full Hessian normal-mode analysis may show some unphysical results, e.g., spurious imaginary frequencies may appear. A frequency analysis in terms of a subset of coordinates that are optimized, i.e. a partial Hessian method, can avoid these problems.

The authors have succeeded recently in deriving a method that is able to calculate physical frequencies. The main idea is to group the atoms that were kept fixed during the partial optimization into one or more blocks that are able to move as rigid bodies with respect to the relaxed molecular part in the vibrational analysis.¹⁰ This mobile block Hessian (MBH) method has shown to be very efficient for an accurate evaluation of relevant frequencies of vibrational modes. The proposed procedure remedies the artifact of imaginary frequencies occurring in standard frequency calculations for partially optimized systems. In addition, only a subblock of the Hessian matrix has to be constructed and diagonalized, leading to a serious reduction of the computation time for the frequency analysis.

MBH can be regarded as an extension of the partial Hessian vibrational analysis approach (PHVA). Only part of the cartesian Hessian has been retained, excluding all the atoms of the passive site of the molecule that is kept fixed during the optimization. This methodology was first introduced and developed by Head and co-workers^{11–14} and was further investigated by Li and Jensen¹⁵ and Besley and Metcalf.¹⁶ It comes to giving an infinite mass to the fixed atoms so that they are frozen at their initial position. Only the relaxed atoms can participate in the small amplitude vibrations.

The novelty of MBH with respect to PHVA lies in the fact that, in the former, the finite mass of each block is taken into consideration in the NMA, instead of giving an infinite mass to the fixed atoms. Six degrees of freedom are attributed to each block to describe its position and orientation with respect to the fully optimized part, and the global translational/rotational invariance of the potential energy surface (PES) is fully respected. Moreover, the PHVA is always limited to the case of one immobile block with infinite mass, whereas in the MBH model, parts of the molecular system can be ranged in multiple blocks which can move as rigid bodies with respect to the relaxed part of the molecule. In ref 10, both PHVA and MBH methods are submitted to a tough comparative study, while in ref 17, attention is given to the practical implementation of the MBH model and the interface with molecular modeling program packages.

One of the main applications that can be deduced from the knowledge of accurate normal-mode frequencies, is the prediction of chemical kinetics, as already mentioned. By means of the partition functions and a molecular-statistical formulation of transition state theory, the reaction rate constant k of a chemical reaction can be determined.^{2–5} A somewhat different approach is proposed by the group of

Lin et al.¹⁸ Basic assumption is that the Hessian elements that involve only the atoms of the active site might be more critical than the other Hessian elements. The less critical elements are approximated following some interpolation procedure, mainly for elements at the nonstationary points on the potential energy surface which are not consistently constructed by the same level (dual level scheme). Other related papers suggest proper methods to predict accurate QM/MM kinetics by incorporating quantum mechanical effects by treating vibrational motions quantum mechanically and applying multidimensional tunneling approximations into reaction rate calculations.^{19,1} Recently, more sophisticated techniques concerning transition state theory have been developed including tunneling effects, quantum dynamical effects and multiple pathways (we refer to ref 20 for a review of all modern developments), but in view of the goal of this paper to validate the MBH approach in predicting kinetics, conventional TST largely suffices and tunneling and other effects will not be incorporated.

In principle, the expression of k includes all normal vibrational modes in reactants and activated complex. It is inherent to both MBH and PHVA approaches that the number of frequencies is always smaller than in a standard frequency calculation. The question arises whether this reduction has a significant influence on the reaction rate constant. Here lies the scope of this work: we will demonstrate that the normal modes which disappear when defining fixed blocks have little influence on the chemical kinetics. This work aims at promoting MBH as a suitable and highly efficient tool for predicting accurate chemical kinetics parameters in large extended molecular systems where the standard full Hessian procedure fails.

Applications of MBH are numerous. They can be classified in various categories:

- (i) Large biosystems consisting of thousands of atoms require a hybrid quantum mechanical/molecular mechanical (QM/MM) approach.^{6–9} The whole MM region can be taken up in one or multiple blocks.
- (ii) A cluster description of zeolites or other periodic systems, such as lattices, requires fixed positions of border atoms to prevent collapse of the molecule during optimization.^{21–23} This represents a particular situation of partial optimization.
- (iii) Reactions in solvents often require an approach with a chemical reactive site and various layers treated at different levels of theory (QM/MM or QM/QM'). The whole can even be circumvented by a bulk solvent described by a polarizable continuum model (PCM).^{24,25} In MBH, the various solvent molecules are all regarded as mobile blocks which can translate and rotate freely around the active site. Only the internal structure of each solvent molecule is held fixed.

The structure of the paper is as follows. In section 2, a short outline of the theoretical methodology is given, and in section 3, the computational details are summarized. Section 4 is devoted to the validation of MBH as an adequate method to predict rate constants. Different reactions with various block choices are taken up in the test set for validation with the benchmark values (full optimization before frequency calculation). The test set includes a prototype substitution reaction, a hydrogen transfer reaction, as well as several

radical addition reactions, since these have a localized reactive site. The effectiveness of the multiple MBH has been illustrated with a more extended aminophosphonate system in section 4.3, for which solvent molecules are taken into account. The results of the MBH have been compared to those of the PHVA approach as well, and based on theoretical considerations, a modified PHVA method is presented in section 4.4, hereafter referred to as PHVA*. Finally in section 5, some conclusions are drawn.

2. Theoretical Background

2.1. Partition functions. Within the harmonic oscillator approximation, the $3N$ degrees of freedom of a N -atom system can be decoupled into three groups of independent motions—3 translational, 3 rotational, and $3N - 6$ vibrational motions—that all contribute to the total partition function Q :

$$Q = q q_{\text{elec}} \quad (1)$$

where

$$q = q_{\text{trans}} q_{\text{rot}} q_{\text{vib}} \quad (2)$$

The translational partition function reads

$$q_{\text{trans}} = \left(\frac{2\pi M k_B T}{h^2} \right)^{3/2} V \quad (3)$$

M stands for the total mass of the system, T is the temperature, k_B is the Boltzmann constant, h is the Planck constant, and V is the volume. If I_1 , I_2 , and I_3 denote the moments of inertia of the system and σ is the symmetry number, the rotational partition function reads

$$q_{\text{rot}} = \frac{8\pi^2}{\sigma} \left(\frac{2\pi k_B T}{h^2} \right)^{3/2} \sqrt{I_1 I_2 I_3} \quad (4)$$

Each vibration with frequency ν_i gives a contribution

$$q_{(v_i)} = \frac{e^{-h\nu_i/2k_B T}}{1 - e^{-h\nu_i/k_B T}} \quad (5)$$

to the total vibrational partition function

$$q_{\text{vib}} = \prod_i q_{(v_i)} \quad (6)$$

The electrons also contribute to the partition function, but when the first electronic excitation energy is much greater than $k_B T$, the first and higher excited states are assumed to be inaccessible. If E_0 is the energy of the ground-state level, this assumption simplifies the electronic partition function to

$$q_{\text{elec}} = e^{-E_0/k_B T} \quad (7)$$

Note that the zero point energy contribution $e^{-h\nu_i/2k_B T}$ in the numerator of eq 5 is frequently left out from the vibrational partition function and incorporated in the electronic partition function.

Ab initio molecular calculations can be used to generate the molecular properties required for the evaluation of the above partition functions, such as the geometry (for the moments of inertia I_i), the Hessian matrix (for the vibrational frequencies ν_i), and the electronic ground-state energy E_0 .

2.2. Reaction Rate Constant within Conventional Transition State Theory (TST). Transition state theory has been proved to be very useful to determine the reaction rate constants.²⁻⁵ It supposes that the transition state or activated complex is in equilibrium with the reactants, although, strictly speaking, this hypothesis is not valid since the transition state corresponds to a saddle point rather than a minimum on the PES. Within this assumption the rate constant is completely determined by the microscopic partition functions and the reaction barrier at 0 K.

For a unimolecular reaction, $A \rightarrow A^\ddagger \rightarrow B$ or $A \rightarrow A^\ddagger \rightarrow B + C$ (with the \ddagger superscript indicating the activated complex) the rate constant k is given by:

$$k(T) = \frac{k_B T}{h} \frac{q(\ddagger)/V}{q(A)/V} e^{-\Delta E_0/k_B T} \quad (8)$$

ΔE_0 represents the molecular energy difference at 0 K between the activated complex and the reactants. The transition state frequency is assumed not to be included in the partition function $q(\ddagger)$ of the activated complex. k is expressed in units s^{-1} .

For a bimolecular reaction $A + B \rightarrow (AB)^\ddagger \rightarrow C$ or $A + B \rightarrow (AB)^\ddagger \rightarrow C + D$, the expression for the rate constant becomes

$$k(T) = \frac{k_B T}{h} \frac{q(\ddagger)/V}{q(A)/V q(B)/V} e^{-\Delta E_0/k_B T} \quad (9)$$

expressed in units of cubic meters per mole second.

2.3. MBH and PHVA. Partially optimized geometries are nonequilibrium structures. The usual normal-mode analysis (NMA) equations $H\nu = \omega^2 M\nu$, with H being the full cartesian Hessian and M the Cartesian diagonal mass matrix, could be solved to obtain the frequencies, but this procedure shows some serious defects. The Hessian H is the second derivative matrix of the potential energy with respect to all the Cartesian coordinates. At nonequilibrium geometries, it has only three zero-eigenvalues instead of six, implying that the rotational invariance of the potential energy surface is not manifest anymore.²⁶ Spurious imaginary frequencies appear. Moreover, the eigenvalues of the Hessian depend on the choice of coordinates.^{27,28}

In the partial Hessian vibrational analysis (PHVA),^{13,15} these defects are surmounted by giving the fixed atoms an infinite mass. The normal mode equations are then restricted to the relaxed atoms only, by taking a submatrix of the Hessian and the mass matrix:

$$H_E \nu = \omega^2 M_E \nu \quad (10)$$

The mobile block Hessian (MBH) model has been proposed recently by the authors¹⁰ as an improvement of the PHVA. In the MBH model the fixed part is considered as a rigid body that is allowed to participate in the small amplitude vibrations, thus taking into account the finite mass of the fixed block. The spurious frequencies and the coordinate dependence are avoided since the system composed of optimized atoms plus block is in equilibrium. Relying on the global translational and rotational invariance, it is possible¹⁰ to write the single block MBH normal mode equations in terms of the same submatrix H_E of the Hessian,

Validation of the MBH Approach

while the corresponding mass matrix is adapted because of the finite block mass:

$$H_E v' = \omega'^2 \tilde{M}' v' \quad (11)$$

with

$$\tilde{M}' = M_E - M_E D_E S^{-1} D_E^T M_E \quad (12)$$

The matrix D_E is constructed in terms of the coordinates of the free atoms with respect to a space fixed frame. The matrix S contains the information on the mass distribution, i.e. the total mass and the moments of inertia of the molecule. Details can be found in the Appendix and more extensively in ref 17.

The usefulness and applicability of the MBH approach are seriously increasing in case of extension to several mobile blocks. The multiple MBH takes into account the finite mass of each block, by including six parameters per block describing its position and orientation into the NMA equations and by mass weighting with the appropriate block mass and moments of inertia.

The multiple MBH method is for instance extremely useful when simulating chemical reactions in a solvent. Solvent molecules can easily be associated to rigid blocks with a fixed internal structure. They can move freely with respect to each other and with respect to the active site of the molecule.

At first sight the MBH is similar to the united atom concept in force fields, since groups of atoms are treated there also as a single entity.²⁹ However, in spite of this resemblance, the MBH is essentially different. In the MBH blocks each atom keeps its identity and continues to contribute individually to, e.g., moments of inertia, Hessian elements, steric hindrance, etc. Coarse-grained or united atom methods reduce the number of atoms and the initial all-atom potential energy surface is approximated by a parametrized PES of lower dimension. The MBH on the other hand does not simplify the potential energy surface but freezes certain degrees of freedom when performing the vibrational analysis.

3. Computational Details

In order to validate both MBH and PHVA methods in their performance in reproducing accurate chemical kinetics, we compare the MBH and PHVA predictions for the reaction rate constant with benchmark values k .

Benchmark structures and frequencies are generated with a full geometry optimization at a high level of theory (DFT/B3LYP/6-311 g**) with tight convergence criteria such that the residual gradients on the PES are negligibly small. Consequently, a frequency calculation is carried out at the same level of theory for the whole molecular system. These equilibrium geometries permit to calculate the reaction rate with the full cartesian Hessian frequencies.

In a first analysis, frequencies and rate constants are calculated for the fully optimized geometry, while the block size is varied in the vibrational analysis. For each reaction under study, we take into consideration various choices of fixed blocks, or, various submatrices H_E of the Hessian. The normal mode equations, eqs 10 and 11, are constructed and

J. Chem. Theory Comput., Vol. 4, No. 4, 2008 617

solved using the same geometry, and thereby, any perturbation resulting from geometry differences is excluded in this particular treatment. This comparative study is thus highly appropriate to investigate the influence on the rate constants of exclusion of parts of the Hessian in the frequency calculation, i.e. limiting the NMA to a partial Hessian.

In a second analysis, partial geometry optimization is performed and consequently followed by a frequency calculation. For the MBH model, the position/orientation (six degrees of freedom) of each block are optimized, in contrary to PHVA, where the atoms in the single block are kept fixed in space. Therefore, a partial optimization with multiple blocks produces a better structure than one with a single block. We remind that PHVA is always limited to a single block, whereas MBH is very suitable to treat multiple blocks.

The partial optimization is performed as follows. First, one optimizes the system at a low level of theory (HF/STO-3g) to find a plausible starting structure. Then the rigid blocks are introduced and the system is partially optimized at a high level of theory (DFT/B3LYP/6-311 g**), while keeping the rigid blocks fixed at their initial internal geometry. All calculations were carried out with the Gaussian03 software package.³⁰ Next, a frequency calculation is performed with the second derivatives of the potential energy using the same high level of theory. The standard full Hessian frequency analysis would give unphysical results due to the residual forces present in the partial optimized structures, as mentioned in the Introduction. Instead, the PHVA or MBH normal mode eqs 10 and 11 are constructed, as these yield physical frequencies. Obviously, the same rigid blocks are chosen as those considered in the precedent partial optimization.

A partial Hessian method such as the MBH or PHVA approach, however, reduces the number of calculated frequencies. The difference in the number of degrees of freedom between reactants and transition state determines the temperature dependence of the reaction rate, as can be easily seen by inspecting eqs 8 and 9. This difference should not change when introducing the MBH blocks. It is therefore obvious that the chosen blocks must consist of the same atoms in reactant(s) and transition state. Note that strictly speaking the internal rigid block geometry might differ between reactants and transition state, because of the first step, i.e. optimization at the low level of theory before the actual partial optimization.

Finally, we made a selection of various chemical reactions for the validation. Most of them are radical addition reactions, but also one prototype substitution reaction (S_N2) and the hydrogen abstraction of one of the ending carbons are included (R6 and R7, respectively). In these reactions the reactive site (the radical center) is well localized. We choose addition reactions of ethene to a large variety of radicals with different substituents. It enables us to select various types of blocks (large and heavy blocks, substituents with ring structure(s), etc) and to give some recommendations in choosing the fixed blocks and the relaxed molecular region.

An overview of the different reactions under study is depicted in Figure 1. The reactions are labeled as R1, R2, etc. and the reactants and products are numbered. The block choices in the reactants are indicated and labeled in Figures

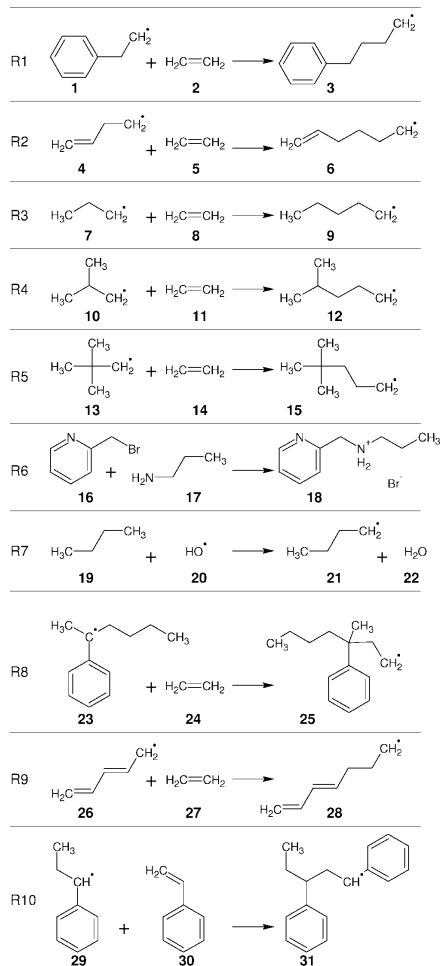


Figure 1. Overview of studied reactions.

2 and 3, and the transition state and product are assumed to contain the same block(s). Blocks can be classified in various types: they can include the reactive center (which at first view appears to be a surprising choice), they can directly be connected with the reactive center by a single bond, or they are separated by more than one bond. It is also possible to combine blocks to the case of multiple blocks. In bimolecular reactions, rigid blocks can be introduced in each of the two reactants; in the activated complex and product, they form multiple blocks, hindering in principle the application of the PHVA method. To illustrate, in reaction R10, one can combine blocks a and b in the description of the two

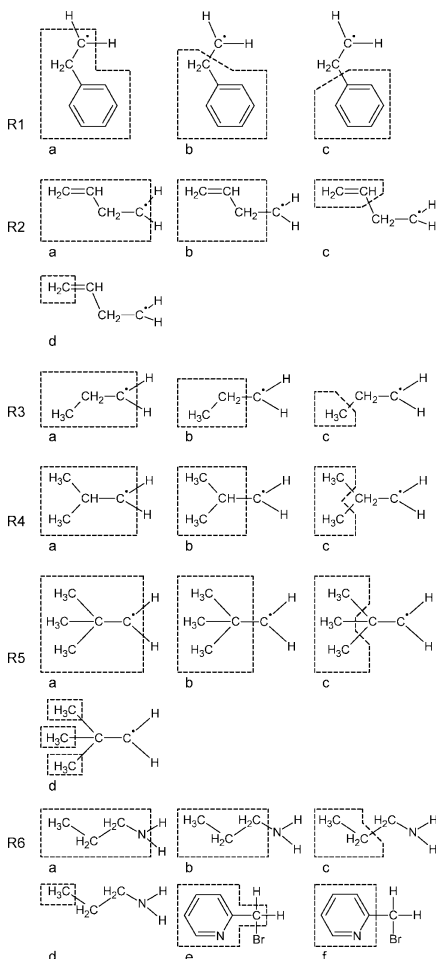


Figure 2. Numbering and choice of the different blocks. reactants. This case will be denoted as a–b and only makes sense when using MBH.

4. Discussion

4.1. MBH with a Single Block. In the MBH (PHVA) approach, the total partition function of eq 1 is used to calculate the reaction rate, but the vibrational partition function q_{vib} is constructed with the MBH (PHVA) frequencies:

$$q^{\text{MBH}} = q_{\text{trans}} q_{\text{rot}} q_{\text{vib}}^{\text{MBH}} \quad (13)$$

$$q^{\text{PHVA}} = q_{\text{trans}} q_{\text{rot}} q_{\text{vib}}^{\text{PHVA}} \quad (14)$$

Validation of the MBH Approach

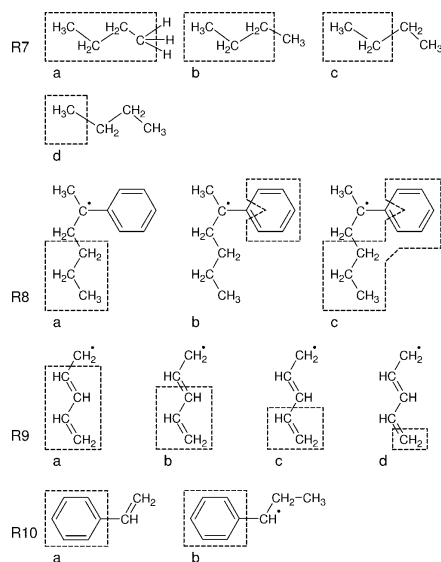


Figure 3. Numbering and choice of the different blocks—continuation.

The chosen test set of chemical reactions allows an exhaustive investigation of the influence on the rate constant of the position of the rigid block, the block's mass, its distance to the reactive center, and the stoichiometry of the reaction.

In Tables 1 and 2, the rate constants at $T = 300$ K are listed for the several reactions in units of cubic meters per mole second (bimolecular reactions) or inverse seconds (unimolecular reactions). In the first column, the benchmark values k , calculated with the full Hessian frequencies of the equilibrium structure, are tabulated for comparison. The benchmark is only available for the fully optimized structure and is calculated in absence of any block. The block size in the MBH or PHVA approach applied on a fully optimized structure is indicated by a, b, c etc. A prime is added if the geometry was obtained by partial optimization, e.g. a', b', c', etc.

In a first step, we concentrate on the results obtained with the fully optimized structures. In the next step, the influence of the partial optimization will be discussed.

As can be seen in Tables 1 and 2, the overall agreement of the MBH rate constants with the benchmark values is remarkably good. The reaction rate constants are reproduced to within a factor of 2, apart from a few cases in Table 2, which are discussed further. This observation holds for a variety of reactions: for unimolecular and bimolecular reactions, for radical and nonradical reactions, and for heavy or small block masses. The deviation is within acceptable limits and is smaller than corrections induced by the level of theory,³¹ internal rotations,^{32,33} tunnel effects, and other factors.^{34,35}

J. Chem. Theory Comput., Vol. 4, No. 4, 2008 619

The apparent agreement of the MBH predictions with the benchmark implies that the contribution of the omitted normal modes, inherent to the MBH method, is of the same magnitude for both the transition states and reactants. Apparently the omitted modes are not essential in the determination of the rate constant. These rather unimportant modes are localized in the fixed block or spread out over the fixed block and the optimized region. The more interesting modes are located in the optimized region that contains the active site, and are well reproduced by the MBH approach. The coupling of the MBH modes with the modes localized in the fixed block is left out in the model, but a logical choice of the blocks makes this coupling irrelevant for the rate constant.

When a block is chosen too close to the active site, the coupling between MBH modes and the omitted modes is not always irrelevant anymore. In reaction R1 the rigid block a includes the reactive center, and the border of block b crosses the bond connecting the radical center. The reaction rate constant k^{MBH} indeed overestimates the benchmark value. Block c is a better choice because it is not directly connected to the reactive site.

In some particular cases, e.g. reaction R4 with block a or b, the MBH approach reproduces k fairly well even with a direct bond between active site and block. However, one should not rely on such coincidences, and anyway, a more suitable choice of a block further away from the radical center still improves the rate estimate. As a general rule, hereafter referred to as the bond-distance rule, it is recommended not to bring the block region too close to the active site.

The mass of the rigid block does not play a crucial role in the validation of MBH in reproducing rate constants. This is best illustrated by comparing reactions R1 and R2. In R1, the fixed block c contains a phenyl group, while block c in R2 consists of an ethyl group. Results are comparable for both the forward and reverse reactions.

When we finally consider the results of the partial optimization, it is clear that the effect is rather moderate. We concentrate on the forward reaction R1 for a detailed study (Table 3). The partial optimization affects the geometry, because the rigid block conserves its initial internal geometry. This will cause differences with the benchmark geometry. In this simple example, this induces quite slight changes (some C–C distances are increased by 0.03 Å), but in more complex systems, the low level of theory geometry and partial optimized geometry may differ substantially. Or, the full optimization at the low level of theory should give a plausible internal geometry for the blocks, but the exact position/orientation of the blocks and the positions of the relaxed atoms are less important, since these are optimized during the consecutive partial optimization at the high level of theory, giving a plausible geometry of the whole system.

The ground-state configuration of a partially optimized system is obviously less bound than the fully optimized system. However, the energy increase of 2 kJ/mol, noticed in the ethylbenzene radical, is mostly compensated by a similar increase of the binding energy of the TS, hence resulting in an almost equal reaction barrier. For instance,

Table 1. Calculated Rate Constants at $T = 300$ K, for Reactions R1–R4 of the Test Set^a

reaction	forward					backward				
	k	block	k^{PHVA}/k	$k^{\text{PHVA}'}/k$	k^{MBH}/k	k	block	k^{PHVA}/k	$k^{\text{PHVA}'}/k$	k^{MBH}/k
R1	3.46E-02	a	5.36	1.23	1.36	1.86E-06	a	0.39	0.35	0.48
		b	7.44	1.71	1.74		b	0.91	0.83	0.83
		c	4.33	0.99	1.02		c	1.13	1.02	0.95
		a'	5.97	1.37	1.52		a'	0.37	0.34	0.46
		b'	7.58	1.73	1.76		b'	0.99	0.90	0.90
R2	2.85E-02	c'	4.21	0.96	0.99	c'	1.15	1.04	0.97	
		a	10.00	1.00	1.12	a	0.43	0.36	0.65	
		b	14.68	1.47	1.44	b	1.03	0.86	0.88	
		c	8.91	0.89	0.93	c	1.09	0.91	0.96	
		d	9.62	0.96	1.00	d	1.16	0.96	1.00	
		a'	8.31	0.84	0.94	a'	0.40	0.34	0.61	
		b'	10.83	1.09	1.07	b'	1.06	0.89	0.92	
R3	1.47E-02	c'	8.41	0.84	0.88	c'	1.07	0.89	0.96	
		d'	9.62	0.96	1.00	d'	1.16	0.96	1.00	
		a	14.68	0.98	1.23	a	0.44	0.35	0.69	
		b	19.84	1.32	1.33	b	1.00	0.81	0.85	
		c	14.78	0.98	0.99	c	1.18	0.95	0.99	
		a'	16.56	1.10	1.39	a'	0.41	0.33	0.65	
R4	1.99E-03	b'	21.08	1.40	1.41	b'	1.06	0.85	0.89	
		c'	14.73	0.98	0.99	c'	1.22	0.98	1.03	
		a	7.21	0.82	1.10	a	0.34	0.28	0.61	
		b	9.14	1.04	1.09	b	0.75	0.62	0.70	
		c	8.66	0.98	0.97	c	1.27	1.05	1.08	
		a'	7.52	0.86	1.14	a'	0.28	0.23	0.49	
		b'	8.52	0.97	1.01	b'	0.76	0.63	0.71	
		c'	8.57	0.97	0.96	c'	1.38	1.14	1.16	

^a The forward rate constants are expressed in units of cubic meters per mole second (bimolecular), and the backward rate constants are in units of inverse seconds (unimolecular). The benchmark value k is given for comparison. Rate constants k^{MBH} (k^{PHVA} , $k^{\text{PHVA}'}$) are calculated with the MBH (PHVA, PHVA') frequencies, for several block choices. The ratios reflect the influence of the MBH (PHVA, PHVA') treatment with respect to the benchmark value. A block without a prime indicates a fully optimized structure, and a block with a prime indicates a partially optimized structure.

in the case study, a suitable choice of the fixed block (block c') predicts a reaction barrier that is hardly different (by 0.04 kJ/mol) from the benchmark value (see Table 3). Significant changes of reaction barriers alter the reaction rate constant to a large extent, but apparently the various reactions R1–R5 of the test set give no indication of this behavior, if one respects sufficient distance between the fixed blocks and the reactive site.

In Table 3, the kinetic parameters A and E_a , determined within the temperature range 300–700 K, are also given. Activation energies remain almost unaffected as could be expected. Potential deviations of k^{MBH} originate from the pre-exponential factor, which is mainly determined by the vibrational contribution to the partition function.

The above discussion validates the use of the MBH model to predict the rate constant on an accurate level. A plausible choice of the fixed block is the only essential ingredient a potential user of MBH should take into account to get adequate predictions of chemical kinetics. MBH is computationally attractive, makes quantum chemical calculations feasible in extended molecular systems, and preserves the true reaction mechanism.

4.2. MBH with Multiple Blocks. The ability of MBH to choose multiple blocks freely moving but conserving their internal structure makes it a powerful tool to a broad range of applications. This is demonstrated in reactions R5, R8, and R10 of the test set, and the results are tabulated in Tables 1 and 2.

In reaction R5, the effect of multiple blocks compared to a single block (block choice d versus c) is moderate. Reaction R8 describes a more complex system. Two individual blocks a and b can be merged to one solid block c, or they can be considered as two mobile blocks a–b. Here, the multiple MBH implies a significant improvement with respect to the single block treatment c. Block c yields ratios 3.86 and 0.44 for the forward and backward reaction, respectively, while the multiple blocks a–b give values of 1.67 and 0.76. Inspection of the MBH values obtained with the individual blocks a and b shows that the global effect of multiple blocks is mostly given by the following multiplication rule:

$$\frac{k_{a-b}^{\text{MBH}}}{k} \approx \frac{k_a^{\text{MBH}}}{k} \times \frac{k_b^{\text{MBH}}}{k} \quad (15)$$

This seems to be true for the forward and backward reaction.

A third example is reaction R10 where we choose a block in each reactant. The TS will then contain two blocks, which are treated within the multiple MBH. The overall factor is indeed fairly well reproduced by the multiplication rule (14). At least it gives an indication on the global error induced by the presence of multiple blocks. A plausible block choice is always of importance to keep the error within the limits. Unphysical block choices are for example b in R8 and c in R9. In both cases the block's border crosses a bond that is part of a delocalized system, and therefore, the k^{MBH} ratios are badly reproduced, even when the bond-distance rule is respected.

Validation of the MBH Approach

J. Chem. Theory Comput., Vol. 4, No. 4, 2008 621**Table 2.** Calculated Rate Constants at $T = 300$ K, for Reactions R5–R10 of the Test Set^a

reaction	forward					backward				
	k	block	k^{PHVA}/k	$k^{\text{PHVA}'}/k$	k^{MBH}/k	k	block	k^{PHVA}/k	$k^{\text{PHVA}'}/k$	k^{MBH}/k
R5	1.72E-03	a	5.46	0.84	1.01	2.94E-06	a	0.42	0.36	0.60
		b	6.57	1.01	1.04		b	0.91	0.77	0.85
		c	6.28	0.96	0.99		c	1.03	0.88	0.95
		d			1.04		d			0.98
		a'	6.12	0.94	1.13		a'	0.43	0.36	0.61
		b'	6.47	0.99	1.03		b'	1.07	0.91	0.99
		c'	6.46	0.99	1.01		c'	1.19	1.01	1.08
R6	9.93E-12	a	1060.01	1.74	1.76	1.41E-13	a	0.51	0.49	0.69
		b	1002.56	1.64	1.71		b	0.81	0.78	0.78
		c	794.07	1.30	1.33		c	0.89	0.86	0.85
		d	657.28	1.08	1.10		d	1.11	1.08	1.00
		e	2.61	0.29	0.85		e	0.17	0.16	0.56
		f	8.70	0.96	1.00		f	0.83	0.80	0.80
					1.16		d'			1.23
R7	2.85E+07	a	4.15	0.43	0.95	1.24E-04	a	10.67	2.03	1.73
		b	4.66	0.49	0.99		b	8.37	1.59	1.56
		c	4.86	0.51	1.05		c	5.50	1.04	1.02
		d	4.79	0.50	1.03		d	5.41	1.03	1.02
		a'	5.06	0.53	1.16		a'	16.69	3.19	2.71
		b'	5.22	0.55	1.11		b'	9.74	1.84	1.82
		c'	5.03	0.52	1.08		c'	5.58	1.06	1.04
R8	8.04E-11	d'	4.85	0.51	1.04	d'	5.48	1.04	1.04	
		a	2.14	1.14	1.11	a	1.31	1.25	0.99	
		b	3.01	1.61	1.50	b	0.89	0.84	0.77	
		c	5.91	3.15	3.86	c	0.48	0.46	0.44	
R9	1.83E-09	a-b	1.67		1.67	a-b			0.76	
		a	21.89	2.53	2.64	a	1.60	1.31	2.27	
		b	14.00	1.62	1.50	b	0.99	0.81	1.29	
		c	14.40	1.66	1.63	c	2.29	1.87	5.14	
R10	3.02E-06	d	10.24	1.18	0.98	d	0.69	0.57	2.04	
		a	43.51	0.89	0.96					
		b	30.18	1.34	1.59					
		a-b			1.53					

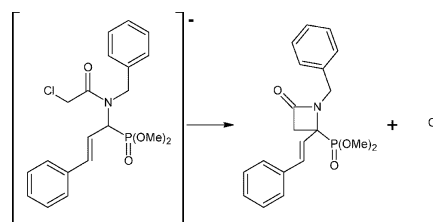
^a See the footnote of Table 1 for more details.**Table 3.** MBH for Reaction R1 with Fully and Partially Optimized Structures^a

	full optim			partial optim		
	a	b	c	a'	b'	c'
k^{MBH}/k	1.36	1.74	1.02	1.52	1.76	0.99
A^{MBH}/A	1.29	1.59	1.03	1.38	1.70	1.01
$E_a^{\text{MBH}} - \Delta E_a$	-0.14	-0.21	0.01	-0.22	-0.02	+0.06
$\Delta E_a^{\text{MBH}} - \Delta E_0$	0	0	0	-0.13	+0.13	+0.04

^a The rate constant is given at 300 K, and kinetic parameters are fitted in the temperature range 300–700 K. k and A are in cubic meters per mole second, and energies are in kilojoules per mole. Benchmark values: $k = 3.46 \times 10^{-2} \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$, $A = 67.22 \times 10^2 \text{ m}^3 \text{ mol}^{-1} \text{ s}^{-1}$, $E_a = 36.53 \text{ kJ/mol}$, $\Delta K_0 = 24.66 \text{ kJ/mol}$.

4.3. MBH for Modeling Solvents. Finally, we have tested the concept of multiple blocks on a more realistic and more extended example, where several explicit solvent molecules are taken into account in the computation. In this case, blocks can be chosen within the reacting molecule and/or the solvent molecules can be treated as blocks, and moreover, the influence of solvent species on both reaction barrier and frequencies, i.e. pre-exponential factor, can be tested.

We have chosen the cyclization of functionalized aminophosphonates, as a representative reaction occurring in an organic solvent (reaction R11, see Figure 4). The choice of this reaction was inspired by a recent combined experimental and theoretical study on the formation of β -lactams by some

**Figure 4.** Reaction R11.

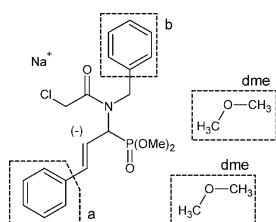
of the authors.³⁶ It was found that, starting from an ambient allylic anion, ring closure occurred exclusively by 4-ring formation, without any trace of 6-ring lactams. At that time, pre-exponential factors were not calculated due to the high computational cost. This is thus an ideal example to validate the approach.

The studied system consists of the aminophosphonate anion together with a sodium counterion and solvated in dimethyl ether solvent molecules (DME). The latter were taken as model molecules for tetrahydrofuran. Three cases are considered: the reaction in the absence of explicit solvent molecules (R11), in the presence of one DME (R11 + 1DME), and in the presence of two DMEs (R11 + 2DME). The benchmark values of k , A , E_a , and ΔE_0 are given in Table 4 for $T = 300$ K. The ratios (for k and A) and differences

Table 4. Benchmark Results for Reaction R11 without and with 1 and 2DME^a

	R11	R11 + 1DME		R11 + 2DME	
<i>k</i>	7.62E-15	6.79E-14	(8.91)	2.09E-12	(274.63)
<i>A</i>	2.10E+13	8.43E+12	(0.40)	5.64E+13	(2.69)
<i>E_a</i>	157.73	149.99	(-7.73)	146.19	(-11.54)
ΔE_0	159.85	151.78	(-8.07)	148.23	(-11.63)

^a The rate constant is given at 300 K, and kinetic parameters are fitted in the temperature range 300–700 K. *k* and *A* are in inverse seconds, and energies are in kilojoules per mole. Ratios (*k* and *A*) and differences (*E_a* and ΔE_0) between solvated and nonsolvated values are given between brackets.

**Figure 5.** Definition of blocks, reaction R11.**Table 5.** Calculated at 300 K^a

block	R11	R11 + 1DME	R11 + 2DME
a	1.71	1.62	1.43
b	0.98	0.97	0.88
dme		1.07	
b-dme		1.04	
dme-dme			0.94
b-dme-dme			0.83

^a Several blocks choices are taken up.

(for energies) given between brackets indicate the effect of the solvation. The presence of one or two solvent molecules indeed increases the reaction rate constant by a factor of 8.91 or 274.63, respectively, with respect to the nonsolvated situation.

The relevant question is whether the MBH model is capable of reproducing the enhancement of *k* due to the solvent. Several block choices are depicted in Figure 5, including the case of blocks within the reactant (a, b), as well as blocks consisting of solvent molecules (dme). Table 5 shows the ratios between the MBH estimates and the benchmark values of the rate constant. Block a is clearly not a good choice, which is easily understood when noting that the block's border cuts through a delocalized bond. Therefore, possible combinations of a with blocks b or dme are not considered in the table. Block b and block dme on the other hand are excellent block choices: since the ratios are close to 1.0, the *k* enhancement 1:8.91:274.63 as reported by the benchmark is maintained and the MBH is thus clearly capable of reproducing the solvation effect. Multiple block combinations such as b-dme, dme-dme, and b-dme-dme reproduce the rate constant very well, which is in agreement with the multiplication rule as stated in eq 14. Resuming, the multiple MBH has proven to be extremely useful and effective in predicting reaction rates, both with blocks

belonging to the reactant or with blocks coinciding with solvent molecules.

4.4. PHVA and PHVA*. Conceptually, the difference between MBH and PHVA is mainly a mass effect. In the MBH, the finite mass of the blocks is taken into consideration, while in the PHVA approach, infinite masses are associated with the atoms in the rigid body. As a result, an extension to multiple blocks has no physical meaning in PHVA. When two blocks with infinite mass are present within one molecule, the system of free atoms and blocks will behave as if the two blocks were one big block with infinite mass. The case of one block in each of the reactants of a bimolecular reaction must also be excluded. The transition state itself would have two blocks with infinite mass. Thereby, six degrees of freedom describing the relative position and orientation of the two blocks will be lost in the transition state, leading to a completely wrong temperature dependence of the reaction rate constant. From a physical point of view, it is also hard to imagine how two reactants, each containing a block with infinite mass, could ever approach each other to form the transition state. The following discussion is therefore limited to the case of a single block with fixed geometry.

In contrast to the MBH, the PHVA cannot be extended to treat multiple blocks. PHVA is thus only applicable within the single block approximation. An overview of the various PHVA reaction rates in Tables 1 and 2 shows that unimolecular reactions are reasonably well described using PHVA frequencies. On the other hand, bimolecular reaction rates are poorly reproduced and significant deviations are noticed. The systematic overestimation of the reaction rate finds its origin in the appearance of spurious low frequency modes in the PHVA approach. A profound investigation of these spurious modes reveals that they represent slow translation/rotationlike movements of the whole group of free atoms. This collective motion encompasses a lot of mass, explaining why (through the mass weighting in the NMA analysis) these frequencies are low. They give a significant contribution to the vibrational partition functions, while the translational/rotational degrees of freedom, however, are already taken into account in the total partition function. The larger the total mass of the free atoms with respect to the mass of the fixed block, the more pronounced is this double counting. Hence, in unimolecular reactions, the enhancement of the vibrational partition functions due to this double counting effect is nearly similar for reactant and transition state, and the enhancement factor is canceled (see eq 8). In bimolecular reactions on the other hand, the double counting is much more prominent for the transition state than for the reactants, thus leading to an overestimated reaction rate.

In order to prevent this double counting effect, we present a corrected version of the PHVA method. In Figure 6, the ratio $q_{\text{vib}}^{\text{MBH}}/q_{\text{vib}}^{\text{PHVA}}$ between the MBH and PHVA vibrational partition functions for the reactants, TS, and products of reactions R1–R6 is plotted against a mass related factor *t* given by

$$t = \sqrt{\frac{M_F^3 I_{F1} I_{F2} I_{F3}}{M^2 I_1 I_2 I_3}} \quad (16)$$

Validation of the MBH Approach

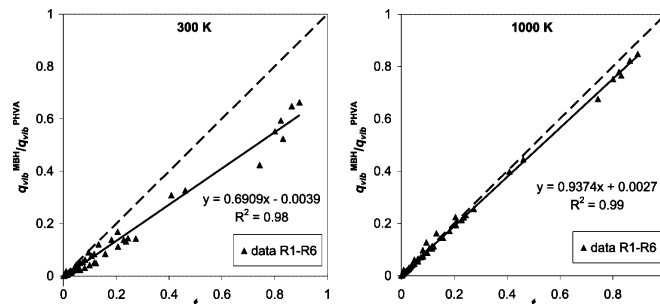
J. Chem. Theory Comput., Vol. 4, No. 4, 2008 623

Figure 6. Ratio $q_{\text{vib}}^{\text{MBH}}/q_{\text{vib}}^{\text{PHVA}}$ for reactants, TS, and products of reactions R1–R6 plotted against the mass related factor t at 300 and 1000 K. The linear regression line (full) is fitted to the data with the least-squares method. The diagonal (dashed line) is added for comparison.

where M is the total mass and I_i ($i = 1, 2, 3$) are the moments of inertia of the molecule, while M_F and I_{Fi} ($i = 1, 2, 3$) are the total mass and moments of inertia of the fixed block. For higher temperatures, an almost linear behavior is observed.

$$\frac{q_{\text{vib}}^{\text{MBH}}}{q_{\text{vib}}^{\text{PHVA}}} \approx t \quad (17)$$

On the basis of eq 17, we now propose the following corrected PHVA partition function, hereafter referred to as PHVA*,

$$q^{\text{PHVA}^*} = q_{\text{trans}} q_{\text{rot}} q_{\text{vib}}^{\text{PHVA}} \sqrt{\frac{M_F^3 I_{F1} I_{F2} I_{F3}}{M^3 I_1 I_2 I_3}} \quad (18)$$

It is not surprising that the ratio $q_{\text{vib}}^{\text{MBH}}/q_{\text{vib}}^{\text{PHVA}}$ depends only on mass properties, because the essential difference between the MBH and PHVA approach is a reduced mass effect. The plot in Figure 6 is so overwhelming that a mathematical proof of eq 16 should be on hand and that a close similarity between PHVA* and MBH predictions for the reaction rate constants is expected. The latter is indeed confirmed by the corrected PHVA* estimates taken up in Tables 1 and 2. Concluding, PHVA* and MBH perform equally well, but the main advantage of MBH, i.e. enabling the extension the procedure to multiple blocks, still holds.

In the following, a mathematical derivation of eq 17 will be presented. In the high temperature limit ($k_B T \gg h\nu$, $\forall \nu$), it is possible to relate $q_{\text{vib}}^{\text{PHVA}}$ and $q_{\text{vib}}^{\text{MBH}}$ as a simple expression containing ratios of the masses and moments of inertia. The contribution of a vibration with frequency ν to the partition function is given by eq 5 and can be approximated by $k_B T/h\nu$ if the temperature is high with respect to the vibrational temperature $h\nu/k_B$. Since the number of MBH and PHVA frequencies is equal, the ratio is independent of temperature.

$$\frac{q_{\text{vib}}^{\text{MBH}}}{q_{\text{vib}}^{\text{PHVA}}} = \frac{\prod_{\nu} \nu^{\text{PHVA}}}{\prod_{\nu} \nu^{\text{MBH}}} \quad (19)$$

The product of frequencies coincides with the square root of the determinant of the matrix in the NMA normal mode equations. In the PHVA model (eq 10), this is

$$\prod_{\nu} \nu^{\text{PHVA}} = \sqrt{\det(M_E^{-1/2} H_E M_E^{-1/2})} \quad (20)$$

while in the MBH model (eq 11), this is

$$\prod_{\nu} \nu^{\text{MBH}} = \sqrt{\det(\tilde{M}^{-1/2} H_E \tilde{M}^{-1/2})} \quad (21)$$

The ratio becomes

$$\frac{q_{\text{vib}}^{\text{MBH}}}{q_{\text{vib}}^{\text{PHVA}}} = \sqrt{\frac{\det \tilde{M}}{\det M_E}} \quad (22)$$

We now introduce the matrix S , defined in the Appendix, which contains the mass information of the complete system. Eigenvalues are the total mass M and the moments of inertia I_i . Similarly we introduce the matrix S_F for the fixed atoms, with eigenvalues M_F and I_{Fi} . Using the properties described in ref 17, the ratio can be rewritten (see the Appendix):

$$\frac{q_{\text{vib}}^{\text{MBH}}}{q_{\text{vib}}^{\text{PHVA}}} = \sqrt{\frac{\det S_F}{\det S}} \quad (23)$$

which is equivalent to expression 17 and which proves the PHVA* correction factor of eq 18.

Numerically, we find that eq 18 is not only valid for high temperatures but that its validity holds quite well for lower temperatures (300 K); see Figure 6.

An interesting property is that the mass related factor t of eq 16 is also equal to the ratio of the translational/rotational partition functions of the fixed block versus global molecule

$$\sqrt{\frac{M_F^3 I_{F1} I_{F2} I_{F3}}{M^3 I_1 I_2 I_3}} = \frac{q_{F,\text{trans}} q_{F,\text{rot}}}{q_{\text{trans}} q_{\text{rot}}} \quad (24)$$

The subscript F refers to the fixed block atoms. Thus, an alternative formulation of the PHVA* approach is presented, where only the vibrational partition function is taken into account. The total translational and rotational partition

function of the molecule are omitted as if it were to avoid the double counting effect.

$$q^{\text{PHVA}^*} = q_{\text{vib}}^{\text{PHVA}} \quad (25)$$

Expression 25 is different from the one proposed in eq 17, but it amounts to the same result when calculating reaction rates. The factor $M_F^{3/2}$ is the same in both TS and reactants because the same block atoms are chosen, and thus, this factor is canceled in the numerator and denominator in the expression of k . The factor $\sqrt{I_{F1}I_{F2}I_{F3}}$ might slightly differ between TS and reactants, if the internal geometry of the blocks is not completely identical for the TS and reactants, but in good approximation, it is canceled as well. The factor $M^{3/2}$ cancels with the translational partition function and $\sqrt{I_1I_2I_3}$ with the vibrational partition function. Therefore, eq 25 will lead to (almost) identical results as eq 17.

5. Conclusion

In this work, the MBH method has been shown to act as an accurate method for the prediction of chemical kinetics in large extended molecular systems. In contrast to the PHVA approach, the MBH method also performs fairly well in bimolecular reactions. An adapted version of PHVA is presented correcting for the double counting effect of global rotation and translation inherent to the PHVA method. The surplus value of MBH with regard to PHVA* lies in the flexibility of MBH to introduce multiple rigid blocks which are freely moving with respect to each other but keeping their initial internal structure. This facility gives a lot of new perspectives in predicting chemical kinetics in very complex systems, where the introduction of one single fixed block is a too crude approximation. Partial optimization is necessary to make quantum chemical computations feasible. The possibility to introduce multiple blocks, each still having six degrees of freedom, makes an accurate reproduction of kinetics to the possibilities.

Most promising application field of MBH would be the description of chemical reactions in a solvent. Each solvent molecule may be regarded as a fixed block, keeping its internal structure, but still enabling to translate/rotate freely with respect to the chemically active part of the system. All ab initio program packages can be used on the condition that the built-in optimization routine allows constraints on internal degrees of freedom. The computational advantage of the MBH method can be exploited when the program package has the ability to calculate partial Hessians. If both features are implemented, MBH could be regarded as a groundbreaking model in the treatment of complex reactions where environment plays a crucial role.

Acknowledgment. This work is supported by the Fund for Scientific Research—Flanders and by the BOF funds of Ghent University.

Appendix

Consider a molecule with N masses m_A , $A = 1, \dots, N$. The positions are described by Cartesian coordinates $r_A \equiv \{r_{A\mu}\}_{\mu=x,y,z}$, with respect to a space-fixed frame. We will treat

the case of one MBH block, consisting of N_F atoms. The remaining $N_E = N - N_F$ atoms are in equilibrium due to the partial optimization. An index E (F) will be used to indicate quantities where only the free (fixed) atoms are considered.

We will focus on the normal mode equations for PHVA and for MBH and, in particular, on the difference in the mass matrices, in order to study the transition from eq 22 to eq 23. The PHVA mass matrix is simply given by M_E [see ()]. The original (not yet transformed to (11)) MBH normal mode equations read

$$\tilde{H}v = \lambda \tilde{M}v \quad (26)$$

where \tilde{M} and \tilde{H} are the MBH mass matrix and Hessian [see ref 17].

Define now a $3N \times 6$ matrix D with components

$$D_{A\mu,\alpha} = \begin{cases} \delta_{\mu,x} & \alpha = 1 \\ \delta_{\mu,y} & \alpha = 2 \\ \delta_{\mu,z} & \alpha = 3 \\ \sum_i \epsilon_{i\mu x} & \alpha = 4 \\ \sum_i \epsilon_{i\mu y} & \alpha = 5 \\ \sum_i \epsilon_{i\mu z} & \alpha = 6 \end{cases} \quad (27)$$

With M as the diagonal $3N \times 3N$ mass matrix, the matrix $S = D^T M D$ is introduced, and similarly, $S_F = D_F^T M_F D_F$. The MBH mass matrix is then given by the block diagonal matrix

$$\tilde{M} = \begin{pmatrix} S_F & 0_{6 \times d} \\ 0_{d \times 6} & M_E \end{pmatrix} \quad (28)$$

with $d = 3N_E$. The normal mode equations are transformed by simultaneous block diagonalization of \tilde{H} and \tilde{M} . The required transformation matrices are given by

$$T_1 = \begin{pmatrix} 1_{6 \times 6} & 0_{6 \times d} \\ x & 1_{d \times d} \end{pmatrix}; \quad T_2 = \begin{pmatrix} 1_{6 \times 6} & y \\ 0_{d \times 6} & 1_{d \times d} \end{pmatrix} \quad (29)$$

with $x = D_E$ and $y = -S^{-1} D_E^T M_E$. The transformed MBH mass matrix and Hessian directly lead to eq 11:

$$T_2^T T_1^T \tilde{H} T_1 T_2 = \begin{pmatrix} 0_{6 \times 6} & 0_{6 \times d} \\ 0_{d \times 6} & H_E \end{pmatrix}, \quad T_2^T T_1^T \tilde{M} T_1 T_2 = \begin{pmatrix} S & 0_{6 \times d} \\ 0_{d \times 6} & \tilde{M}' \end{pmatrix} \quad (30)$$

with $\tilde{M}' = M_E - M_E D_E S^{-1} D_E^T M_E$. Or, the relevant mass matrix is \tilde{M}' for MBH.

Since by construction $\det T_1 = \det T_2 = 1$, it is obvious that the following relations between determinants hold:

$$\det \tilde{M} = \det S_F \det M_E \quad (31)$$

$$\det (T_2^T T_1^T \tilde{M} T_1 T_2) = \det \tilde{M} = \det S \det \tilde{M}' \quad (32)$$

or

$$\frac{\det \tilde{M}'}{\det M_E} = \frac{\det S_F}{\det S} \quad (33)$$

This proves the transition between eqs 22 and 23.

References

- (1) Gao, J. L.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467–505.
- (2) Eyring, H. *J. Chem. Phys.* **1935**, *3*, 107.

Validation of the MBH Approach

- (3) Evans, M. G.; Polanyi, M. *Trans. Faraday Soc.* **1935**, *31*, 875.
- (4) Laidler, K. J. *Chemical Kinetics*; Harper Collins Publishers, Inc.: New York, 1987, 87–138.
- (5) Mc Quarrie, D. A.; Simon, J. D. *Physical Chemistry - a molecular approach*; University Science Books: Sausalito, CA, 1997; pp 1075–1079.
- (6) Warshel, A.; Levitt, M. J. *Mol. Biol.* **1976**, *103* (2), 227–249.
- (7) Assfeld, X.; Rivail, J. L. *Chem. Phys. Letters* **1996**, *263* (1–2), 100–106.
- (8) Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102* (24), 4714–4721.
- (9) Zhang, Y. K.; Lee, T. S.; Yang, W. T. *J. Chem. Phys.* **1999**, *110* (1), 46–54.
- (10) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. *J. Chem. Phys.* **2007**, *126* (22), 224102.
- (11) Jin, S. Q.; Head, J. D. *Surf. Sci.* **1994**, *318* (1–2), 204–216.
- (12) Calvin, M. D.; Head, J. D.; Jin, S. Q. *Surf. Sci.* **1996**, *345* (1–2), 161–172.
- (13) Head, J. D. *Int. J. Quantum Chem.* **1997**, *65* (5), 827–838.
- (14) Head, J. D. *Int. J. Quantum Chem.* **2000**, *77* (1), 350–357.
- (15) Li, H.; Jensen, J. H. *Theor. Chem. Acc.* **2002**, *107*, 211–219.
- (16) Besley, N. A.; Metcalf, K. A. *J. Chem. Phys.* **2007**, *126* (3), 035101.
- (17) Ghysels, A.; Van Neck, D.; Waroquier, M. *J. Chem. Phys.* **2007**, *127*, 164108.
- (18) Lin, H.; Pu, J. Z.; Albu, T. V.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108* (18), 4112–4124.
- (19) Garcia-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2001**, *114* (22), 9953–9958.
- (20) Fernandez-Ramos, A.; Miller, J. A.; Klippenstein, S. J.; Truhlar, D. G. *Chem. Rev.* **2006**, *106* (11), 4518–4584.
- (21) Stevens, F.; Vrielinck, H.; Van Speybroeck, V.; Pauwels, E.; Callens, F.; Waroquier, M. *J. Phys. Chem. B* **2006**, *110* (16), 8204–8212.
- (22) Lesthaeghe, D.; Delcour, G.; Van Speybroeck, V.; Marin, G.; Waroquier, M. *Microporous Mesoporous Mater.* **2006**, *96*, 350–356.
- (23) Lesthaeghe, D.; De Sterck, B.; Van Speybroeck, V.; Marin, G. B.; Waroquier, M. *Angew. Chem., Int. Ed.* **2007**, *46* (8), 1311–1314.
- (24) Cui, Q. *J. Chem. Phys.* **2002**, *117* (10), 4720.
- (25) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105* (8), 2999–3093.
- (26) Tachibana, A.; Fukui, K. *Theor. Chim. Acta* **1978**, *49* (4), 321–347.
- (27) Murry, R.; Fourkas, J. T.; Wu-Xiong, L.; Keyes, T. *J. Chem. Phys.* **1999**, *110*, 10410–10422.
- (28) Wales, D. J. *J. Chem. Phys.* **2000**, *113*, 3926–3927.
- (29) Yang, L. J.; Tan, C. H.; Hsieh, M. J.; Wang, J. M.; Duan, Y.; Cieplak, P.; Caldwell, J.; Kollman, P. A.; Luo, R. *J. Phys. Chem. B* **2006**, *110* (26), 13166–13176.
- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (31) Van Speybroeck, V.; Van Cauter, K.; Coussens, B.; Waroquier, M. *Chemphyschem* **2005**, *6* (1), 180–189.
- (32) Van Speybroeck, V.; Van Neck, D.; Waroquier, M. *J. Phys. Chem. A* **2000**, *104* (46), 10939–10950.
- (33) Vansteenkiste, P.; Van Neck, D.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Phys.* **2006**, *124* (4), 044314.
- (34) Sabbe, M. K.; Saeys, M.; Reyniers, M. F.; Marin, G. B.; Van Speybroeck, V.; Waroquier, M. *J. Phys. Chem. A* **2005**, *109* (33), 7466–7480.
- (35) Sabbe, M. K.; Vandeputte, A. G.; Reyniers, M. F. O.; Van Speybroeck, V.; Waroquier, M.; Marin, G. B. *J. Phys. Chem. A* **2007**, *111* (34), 8416–8428.
- (36) Van Speybroeck, V.; Moonen, K.; Hemelsoet, K.; Stevens, C. V.; Waroquier, M. *J. Am. Chem. Soc.* **2006**, *128* (26), 8468–8478.

CT7002836

Paper IV

MFI Fingerprint: how pentasil-induced IR bands shift during zeolite nanogrowth

D. Lesthaeghe, P. Vansteenkiste, T. Verstraelen, A. Ghysels,
C. E. A. Kirschhock, J. A. Martens, V. Van Speybroeck and
M. Waroquier

Journal of Physical Chemistry C 112, 9186-9191 (2008)

Copyright 2008 by American Chemical Society

MFI Fingerprint: How Pentasil-Induced IR Bands Shift during Zeolite Nanogrowth

David Lesthaeghe,[†] Peter Vansteenkiste,[‡] Toon Verstraelen,[‡] An Ghysels,[‡]
Christine E. A. Kirschhock,[‡] Johan A. Martens,[‡] Veronique Van Speybroeck,^{*,†} and
Michel Waroquier[†]

Center for Molecular Modeling, Ghent University, Proeftuinstraat 86, 9000 Ghent, Belgium, and Centre for Surface Chemistry and Catalysis, Katholieke Universiteit Leuven, Kasteelpark Arenberg 23, 3001 Heverlee, Belgium

Received: December 7, 2007; Revised Manuscript Received: March 20, 2008

Silicalite-1 zeolite exhibits a characteristic pentasil framework vibration around 540–550 cm⁻¹. In the initial stages of zeolite synthesis, however, this band is observed at much higher wavenumbers: literature shows this vibration to depend on particle size and to shift over 100 cm⁻¹ with increasing condensation. In this work, the pentasil vibration frequency was derived from theoretical molecular dynamics simulations to obtain the correct IR band assignments for important nanoparticles. The IR spectroscopic fingerprint of oligomeric five-ring containing precursors proposed in the literature was computed and compared with experimental data. Our theoretical results show that, while isolated five-membered rings show characteristic vibrational bands around 650 cm⁻¹, the combination of five-membered rings in the full MFI-type structure readily generates the bathochromic shift to the typical pentasil vibration around 550 cm⁻¹. As opposed to what was previously believed, the IR band does not shift gradually as nanoparticle size increases, but it is highly dependent on the specific way structural units are added. The most important feature is the appearance of an additional band when double five-membered rings are included, which allows for a clear distinction between the key stages of early zeolite nucleation. Furthermore, the combination of the simulated spectra with the experimental observation of this spectral feature in nanoparticles extracted from silicalite-1 clear solutions supports their structured nature. The theoretical insights on the dependency of pentasil vibrations with the degree of condensation offer valuable support toward future investigations on the genesis of a zeolite crystal.

Introduction

The literature of IR studies on the early stages of zeolite synthesis is exclusively experimental and severely fragmented. We will, therefore, first give a structured overview of the existing literature before addressing the specific issues that arise.

Zeolites are nanoporous aluminosilicate materials, often used in industrial applications for their unique properties in both catalysis and molecular separation.¹ The development of various new synthesis procedures in the past 2 decades has made it possible to design a wide variety of microporous, mesoporous, and hierarchical materials, with varying pore sizes and channel structures. As the number of new materials increases, the field is expanding to all kinds of new applications, e.g., controlled release, sensors, optics, and electronic components.² In parallel with the expansion of the nanoporous materials family, fundamental understanding of the molecular and supramolecular mechanisms of the polymerization of specific porous silicate materials is also steadily improving. However, even though significant advances have been made in understanding the role of template molecules, there is still a glaring lack of insight into the details of elementary steps in silica organization. The process of new material discovery and upscaling of syntheses would benefit greatly from submicroscopic insight into the discrete steps, from initial nucleation to full crystal growth.

A landmark discovery in zeolite synthesis was the clear-solution technique of silicalite-1,^{3,4} which has the MFI topology and is the full silica version of the industrially important aluminosilicate ZSM-5.⁵ The clear-solution mixture is composed of tetrapropylammonium hydroxide ((TPA)OH), water, and tetraethyl orthosilicate (TEOS) at a rather low concentration. When this solution is heated for a few hours, it forms sub-micrometer-sized high-quality crystals of silicalite-1, with its channel intersections occupied by the TPA cations. Even though other synthesis procedures for silicalite-1 exist, the clear-solution technique provides a major advantage for experimental researchers, as silicate particles can be readily extracted and freeze-dried. Furthermore, the clear-solution technique allows for the use of all kinds of *in situ* diagnostics that require optically transparent and dilute media, such as, for example, dynamic light scattering (DLS). Because the clear-solution synthesis technique is also highly reproducible, it makes analysis of the initial synthesis steps much more amenable for fundamental research in zeolite nucleation and crystal growth.

During the clear-solution synthesis of silicalite-1, a stable suspension of nanometer-sized silica particles has been unequivocally observed, prior to the observation of zeolite crystals characterized by Bragg scattering of X-rays.^{6,7} However, the precise structure and role of these nanoparticles during the first steps has been highly debated ever since. Various nanosized structure models have been proposed, ranging from amorphous silicate particles to highly defined framework fragments resembling Lego building blocks.⁸ The growth process along which these nanoparticles assemble is also widely debated: do structured silicate nanoparticles self-assemble piece-by-piece in a

* To whom correspondence should be addressed. Telephone: +32 9 264 65 58. Fax: +32 9 264 66 97. E-mail: veronique.vanspeybroeck@ugent.be.

[†] Ghent University.

[‡] Katholieke Universiteit Leuven.

MFI Fingerprint

multicenter aggregation process.^{9,8–10} Or does zeolite growth occur more in analogy to an Ostwald ripening process, where small nanoparticles sacrifice monomers to feed steadily growing larger particles?^{11,12} To add further complexity to this issue, it has even been suggested that one mechanism may dominate the earlier stages, while the other takes over later on.¹³

The major problem in identification of these first stages is that, even with the clear-solution technique, experimental characterization of these embryonic nanoparticles remains difficult.¹⁴ Zeolite nucleation and growth occurs at length scales just above the NMR window but also just below the diffraction regime.^{11,15} Furthermore, the template (TPA)OH and the co-solvent EtOH give a very high Raman yield with the consequence that the comparably weak Raman signals of the silicate are unobservable. Because of a lack of other suitable tools, Fourier transform infrared spectroscopy (FTIR) has been one of the key methods used to investigate the solid samples.^{16,17} The most important fingerprint in these studies has always been the observed band at approximately 540–550 cm^{-1} , which demonstrates the presence of condensed five-membered structures.^{9,18,19} This band is considered to be the spectroscopic signature of MFI-type zeolite,^{20,21} but it is also a typical feature of similar structures built from five-membered rings, like the MOR and FER frameworks. Most importantly, this absorbance band is completely absent for amorphous silica particles, which makes its presence highly sensitive to the locally structured nature of the material. In combination with a neighboring band around 450 cm^{-1} , the ratio of band intensities is often used as an approximate assessment of the degree of crystallinity of the observed material.

A handful of papers have pointed out a special evolution of this signature band: as MFI-structured nanoparticles grow, the band red-shifts from initially higher values to the frequency associated with fully crystalline MFI structures.^{22–24} In-situ IR spectra of (TPA)OH–TEOS–H₂O mixtures initially display IR absorption around 650 cm^{-1} , shifting and broadening to 600 cm^{-1} with time, as shown in Figure 1A.²² Kirschhock et al.²² managed to isolate two different populations of particles and to record their IR absorption spectra ex situ. The smaller species was named the “precursor”, containing between 33 and 36 Si atoms. The larger species was called the silicalite “nanoslab” and is built from the aggregation of 12 precursor species, containing at least 400 Si atoms. Comparison of the IR spectra of the precursor, the nanoslab, and crystalline silicalite-1 revealed an aggregation-induced red shift. More specifically, for the precursor sample, Kirschhock et al. observed an absorption band around 590 cm^{-1} , while for the sample of nanoslabs this band was red-shifted over 20 cm^{-1} . An additional shift was observed after these slabs condensed into Bragg-crystalline silicalite-1 structure, where the band was located around the expected 550 cm^{-1} , illustrated in Figure 1B. Other researchers have observed similar results but have not gone so far as identifying the nanoparticles. Hsu et al.²⁴ found that young precursor-type species show a broad band at 570 cm^{-1} , and as the aging time increased, the band red-shifted toward 550 cm^{-1} and became narrower and more intense. In a similar study Serrano and van Grieken observed a broad vibration band around 560 cm^{-1} in extracted subcolloidal particles from clear-solution synthesis.²³ Titanium-substituted specimens have been shown to exhibit similar spectroscopic changes depending on particle size.²⁵

Even though it has been only occasionally exploited to date, this IR shift seems to be a powerful tool toward identification of intermediates in the synthesis stage. However, as several

J. Phys. Chem. C, Vol. 112, No. 25, 2008 9187

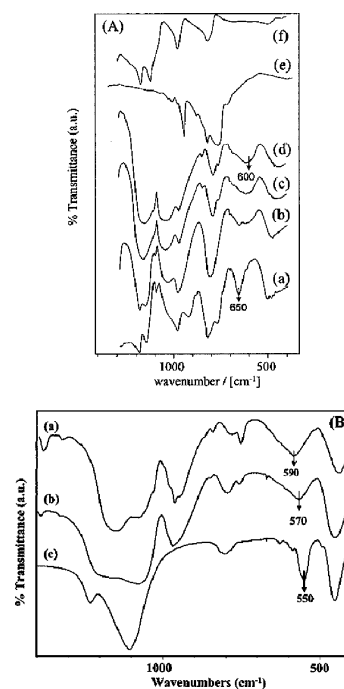


Figure 1. (A) In situ IR spectra of (TPA)OH–TEOS mixture after time intervals of (a) 1, (b) 10, (c) 20, and (d) 46 min, respectively, and (e) TEOS and (f) (TPA)OH. (B) IR spectra of extracted solids: (a) precursor, (b) nanoslab, and (c) micrometer size silicalite-1. Reproduced with permission from ref 22. Copyright 1999 American Chemical Society.

important questions still need to be answered, measuring the shift is not yet as widespread as measuring the intensity of the MFI fingerprint IR band in final materials. Can such a shift really be assigned to nanoparticle assembly? At what stage or stages does a shift occur and why?

In this paper, we use modeling techniques to calculate the IR shifts for pentasil-induced bands from the initial stages of self-organization through to nanosized silicalite-1 crystals. By identifying a detailed link between IR band and crystallinity, this contribution aims to aid experimental scientists in this field. However, as this is a static approach on extracted solids, it does not elucidate the dynamics along which the nanoparticles are formed. The IR shift alone cannot, therefore, distinguish between precursor self-assembly or an Ostwald ripening process in which the zeolite crystal grows at the expense of other nanoparticles. However, since the insights provided in this paper will simplify the identification of intermediate structures and intermediate stages, these results will strongly assist future experimental studies on this issue.

Theoretical Basis. In this paper we make use of an in-house-developed force field for zeolites GCMFF_{SiOH} (version 0.2). This force field was calibrated at the post-Hartree–Fock MP2/6-311+g(d,p) level of theory, with the gradient curves method

(GCM).^{26,27} This is a novel technique that facilitates the development of transferable force-field models. It makes extensive use of regularization techniques²⁸ and of generic energy terms based on series expansions to obtain an optimal bias–variance tradeoff during the fitting procedure. The force-field parameters can be found in the Supporting Information. The IR spectrum was derived from molecular dynamics simulations based on this force field. During an initialization run of 2 ps, the molecular geometries were brought into equilibrium with a thermostat at 300 K, using a velocity-scaling algorithm. The equilibrated geometry was then used as the starting point for 10 consecutive NVT molecular dynamics simulations of 100 ps at a temperature of 300 K, using the Nosé–Hoover-chains method. At the beginning of each of the 10 simulations, the velocities were randomly sampled from the Maxwell distribution. The trajectories were divided into five intervals of 20 ps each, and the average IR spectrum was calculated from these 50 individual intervals.

Scaling of the theoretical results needs to be considered since, as was demonstrated by Scott and Radom,²⁹ the MP2 level of theory severely underestimates low-frequency vibrations. For full ab initio calculations, MP2 frequency scaling factors have been obtained, ranging from 1.01 to 1.05, depending on the basis set used.²⁹ As no such scaling factor has yet been obtained for the MP2-based classical force field, we suggest the use of 1.04, which is the ratio between the experimental band at 650 cm⁻¹ and our simulated result for small oligomers. It is reassuring to note that, even though just the initial stage was fitted to experimental data, the exact same scaling factor would independently be derived by fitting the final stages. For completeness, both the scaled and unscaled values are reported in Table 1. Figure 3 depicts the unscaled IR spectra, while in Figure 4 and in the remainder of this text we will always refer to the scaled values.

IR Spectrum. The IR adsorption cross-section is given by³⁰

$$\alpha(\omega) = \frac{4\pi^2\omega[1 - \exp(\hbar\beta\omega)]}{3\hbar cn} I(\omega)$$

where $\beta = k_B T$ and

$$I(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle d\bar{\mu}(0) \bar{\mu}(t) \rangle$$

In these expressions, k_B is the Boltzmann constant, c is the speed of light, n is the refractive index of the medium, \hbar is the reduced Planck constant, and $\langle \bar{\mu}(0) \bar{\mu}(t) \rangle$ is the time autocorrelation function of the dipole moment. Using standard manipulations, one obtains in the classical limit

$$\alpha(\omega) \sim \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \left| \int_0^{\tau} dt e^{-i\omega t} \frac{d\bar{\mu}(t)}{dt} \right|^2$$

where $\mu(t)$ is the time-dependent dipole moment from the molecular dynamics simulation. When the velocity and the charge of the atom i are given by $v_i(t)$ and q_i , respectively, the time derivative of the dipole moment can be approximated as

$$\frac{d\bar{\mu}(t)}{dt} = \sum_{i=1}^N q_i \bar{v}_i(t)$$

where N is the number of atoms.

Results and Discussion

In this section, we have grouped the results and discussion into several subsections corresponding to key synthesis stages.

TABLE 1: MFI Fingerprint Peak Position of Simulated IR Spectra for Complete Range of Structures^a

		theoretical band	scaled $\times 1.04$
small oligomers	5R	624	649
	8T (2 \times 5)	625	650
	11T (3 \times 5)	628	653
	7T (2 \times 5)	603	627
	8T (4 \times 5)	595	619
	12T (4 \times 5)	578	601
	10T (2 \times 5)	625	650
precursors	22T	578	601
		595	619
		553	575
growth along <i>a</i> -axis	36T	548	570
	2a	549	571
	2a'	545	567
growth along <i>b</i> -axis	3a	549	571
	2b	541	563
	2b'	543	565
	3b	540	562
	4b	538	560
growth along <i>c</i> -axis	4b'	540	562
	5b	539	561
	2c	538	560
	3c	536	557
	4c	536	557
half-slab	2b3c	531	552
	2b'3c	532	553
slab	4b3c	534	555
	4b'3c	534	555
block	2a2b2c	533	554
	2a2b'2c	533	554

^a For combining two precursors along the *a*- or the *b*-direction, two different connections are possible: a and a' or b and b', respectively. The frequency scaling factor of 1.04 has been applied to correct for the underestimation of the MP2-based force field.

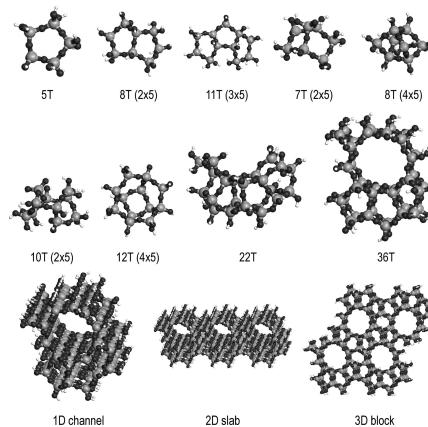


Figure 2. Nomenclature for the structures used in the simulations, ranging from small oligomers to an almost full MFI crystal. The figures were constructed with the in-house-developed Zeobuilder program.²⁷

Within each subsection we will discuss the theoretical results, followed immediately by an interpretation of the implications for that particular stage. Figure 2 gives an overview of the structures used (some of which are key components of the MFI structure and some of which might just be formed during the

MFI Fingerprint

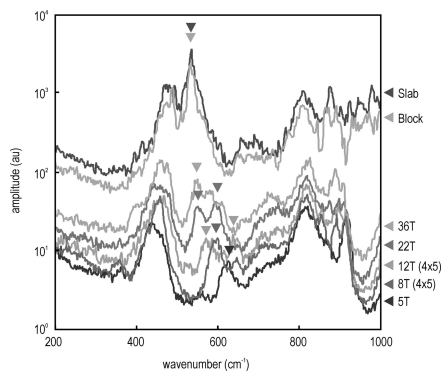


Figure 3. Several key simulated IR spectra, ranging from the isolated five-membered ring (5T), through the quadruple five-membered ring (8T), the pentacyclic dodecamer (12T), the 22T MFI-structured unit (22T), and the 36T MFI precursor (36T), to the 3×4 silicalite nanoslab (slab), and $2 \times 2 \times 2$ nanoblock (block).

early stages), accompanied by the nomenclature used in this paper. In Figure 3 several key IR spectra are shown, while the fingerprint IR bands for all structures are summarized in Table 1. Figure 4 presents a summarizing timeline (chronological order from right to left) in which each structure is linked to the position of the corresponding IR band.

Pentasil Ring and Small Oligomeric Units. The elementary building block in this study is the five-membered ring (5T), also called the pentasil ring, as shown in Figure 2. Combinations of this ring are deemed responsible for the typical IR-active vibrations observed at $540\text{--}550\text{ cm}^{-1}$.³¹ Our simulations of an isolated five-membered ring show a much higher theoretical band at 650 cm^{-1} (illustrated in Figure 3 and Table 1), matched to experimental observations.²² The value obtained is too high to correspond with the MFI fingerprint around $540\text{--}550\text{ cm}^{-1}$, which means that, during crystal growth, this value undergoes a significant red shift of almost 100 cm^{-1} .

By connecting pentasil rings, slightly larger building blocks can be constructed. One possible first step is the construction of two and three sideways annealed five-membered rings (called 8T (2×5) and 11T (3×5), respectively) by successively adding three T-atoms (as shown in Figure 2 and Figure 4). The resulting vibrational frequencies do not differ substantially from the single five-membered ring though and show a similar IR peak around 650 cm^{-1} (Table 1). The mere connection of five-membered rings, sharing just two silicon sites, has no major influence on the frequency of the IR band. This situation changes if, instead, a double five-membered ring is constructed in such a way that two adjacent sides are communal; i.e., three T-atoms are shared by the five-membered rings, forming an additional six-membered ring, referred to as the 7T (2×5) species in Figure 2. A sudden shift of approximately 20 cm^{-1} is observed (as shown in Table 1). An even more condensed structure is the silica octamer, referred to as 8T (4×5), in which four five-membered rings form a tiny cage-like structure. Here, opposing T-atoms in annealed double five-membered rings are connected by an extra bond, resulting in a highly symmetric species containing four identical Q^3 and four identical Q^2 Si atoms. This extremely condensed structure shows an even more significant shift in the infrared peak to a value of approximately 620 cm^{-1} .

J. Phys. Chem. C, Vol. 112, No. 25, 2008 9189

Increasing the number of five-membered rings can thus have two different effects on the frequency of the typical band assigned to pentasil vibrations: if the structures remain only loosely connected, sharing at most two silicon atoms between separate five-membered rings, the typical IR band around 650 cm^{-1} remains stationary and there is no observable shift. It would, therefore, be impossible to distinguish between these small oligomeric units from IR bands alone. However, when a similar number of five-membered rings are interconnected to form building blocks with a higher degree of condensation, a significant red shift over $20\text{--}50\text{ cm}^{-1}$ occurs. This effect is an important step toward interpretation of the experimentally observed red shift during early oligomeric stages of MFI synthesis,²² but it is only part of the full story. The simulated IR bands are still too far away from the characteristic MFI bands observed for larger precursor MFI structures and fully crystalline silicalite-1.

This picture undergoes a major change if the double five-membered ring units suggested by Jacobs et al.²⁰ are simulated. This species is referred to as 10T (2×5) in Figure 2. As the five-membered rings themselves do not share any oxygen bridges, the typical band around 650 cm^{-1} remains present. However, a second band appears at 600 cm^{-1} . Even though the five-membered rings do not share any adjacent sides, the appearance of a new peak provides an instant jump of approximately 50 cm^{-1} . This trend is also obvious for the silica dodecamer structure 12T (4×5), which was proposed as one early building unit in silicalite-1 precursor formation.³² This caplike species shows just a single strongly red-shifted band around 600 cm^{-1} .

MFI Precursors. The role of multiple IR bands becomes even more pronounced if several such units combine to form the small cages that are typically encountered in MFI. In this study we have focused on an MFI-structured 22T precursor, which forms a crucial part of the firm 10-ring channel walls. The simulations demonstrate how the multiple IR bands show up in our region of interest upon increasing degree of 5R condensation. As clearly shown in Figure 3 and Figure 4, the primary peak is observed at 620 cm^{-1} , after which it loses intensity and is joined by a new rising secondary band around 575 cm^{-1} . In a next step, the 22T precursor is extended to form the typical 10-membered ring (36T precursor). These essential features will finally lead to the straight and sinusoidal channels that are typical for the materials that have the MFI topology. Furthermore, the 36T structure is the actual precursor on which the aggregation model for zeolite synthesis, proposed by Kirschhock et al.,⁸ was based. The original peak around 620 cm^{-1} loses even more intensity and becomes difficult to distinguish. The new peak increases in intensity and undergoes an additional small red shift toward 570 cm^{-1} . However, as is shown in Figure 3, the IR spectrum of the 36T precursor could just as well be perceived as a single broad band around 600 cm^{-1} .

From our results, this seems to be the crucial stage in both the observed IR bands and the synthesis of silicalite-1. In the previous stages the IR band red-shifted gradually, depending on the way the five-membered rings were connected. However, these types of structures can still be expected to be found in amorphous silica particles as well as in many other zeolite structures. A major jump of 50 cm^{-1} occurs only once substantial characteristics of the true MFI framework topology are defined in the double five-membered rings that are present in nanosize precursors. This jump does not correspond to the typical gradual shift that seems to accompany nanogrowth but

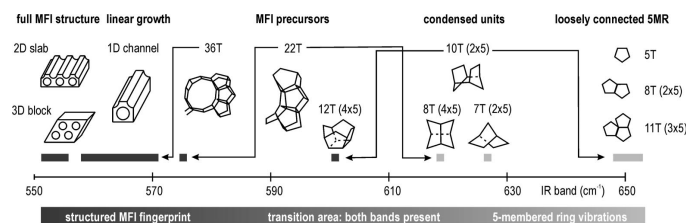


Figure 4. Different regions in silicalite-1 nanogrowth with their corresponding shift in IR band.

arises from the increased intensity of a new band smothering the higher-lying band. The presence of these two bands thus fully explains why Hsu et al.²⁴ observed a broad band around 570 cm^{-1} for young precursor-type particles. Kragten et al. also observed a broad weak-to-medium band at 565 cm^{-1} for nanoparticles, which they believed to be similar to an extremely weak and very broad band for amorphous silica in the same region.¹⁸ Ravishanker et al., however, succeeded in the observation of a split band, reporting a doublet of 555 and 570 cm^{-1} for crystalline nanosized silicalite-1 as compared to a single band at 550 cm^{-1} for micrometer-sized silicalite-1.³³

The theoretically simulated IR results thus not only correspond with earlier experimental work but also provide new insights into the fundamental nature of nanoparticle growth. As shown in Figure 4, the synthesis of small MFI precursors marks the boundary area between which two different IR bands are observed. Smaller particles will exhibit the vibrations typical for isolated five-membered rings, while larger more condensed structures in nanoparticles will show the MFI fingerprint even in the absence of Bragg crystallinity.

One-Dimensional Growth along Crystal Axes. Our simulations show that, once several of these precursors are attached to each other, the higher-lying band disappears completely, resulting in a single, much narrower band, in accordance with the observation of Hsu et al.²⁴ Furthermore, the 570 cm^{-1} band exhibits a further bathochromic shift. Quite remarkably, this shift is slightly dependent on the direction along which growth occurs (Table 1). Along the a -axis, the direction of the sinusoidal channels, only four connections are necessary for linking this type of precursor and hardly any shift occurs. For the growth along the b - and c -axes, six bonds are formed, which results in significant shifts of -10 cm^{-1} to 560 cm^{-1} . The shift appears to be slightly more pronounced for growth along the c -axis, which results in a very condensed network between two precursors. The number of connections and the resulting degree of condensation along the b -axis (i.e., direction of straight channels), on the other hand, is lower because of the presence of the channel crossings with the sinusoidal channels in the a -direction. When looking at linear growth by adding the precursors one-by-one, it is clear that the main shift already occurs when just two precursors are attached along the b - or c -axis. The effect of nanogrowth on the IR band diminishes rapidly if more building blocks are successively attached in the same direction.

Two-Dimensional and Three-Dimensional Growth to Full Crystal. When growth is considered in two dimensions, to create the nanoslabs as proposed by Kirschhock et al.,⁸ the IR band will shift to even lower values. Already for half a nanoslab (containing two 36T precursor units in the b -direction and three 36T precursor units along the c -axis), the IR band is located near 555 cm^{-1} . Enlargement to a full nanoslab does not

fundamentally alter the vibrational frequencies, and the value hovers around 555 cm^{-1} . Theoretically, a combination of 36T building blocks into three-dimensional blocks gives the maximal constraints on the lattice, yet, compared to the two-dimensional nanoslabs, no further shift is observed. The IR band shift of the MFI nanoblock remains at 555 cm^{-1} , close to the value observed in the fully periodic MFI lattice. From these results it seems that, from slabs and blocks onward, the silicalite-1 crystal is well-defined in terms of pentasil vibrations and the MFI fingerprint is observable at full strength.

On the basis of the available IR spectra of the silicalite-1 crystallization intermediate,^{22,33} there is some discrepancy between computed and experimentally determined FTIR signatures in the $540\text{--}600\text{ cm}^{-1}$ region. Extracted precursors and nanoslabs show IR absorption mainly around 590 cm^{-1} and 570 cm^{-1} ,²² respectively, while theoretically the absorption should already be almost entirely red-shifted to 555 cm^{-1} . Even $18\text{--}100\text{ nm}$ sized silicalite-1 nanophase material still shows a splitting of the pentasil vibration into 555 and 570 cm^{-1} wavenumbers.³³ This discrepancy can have several causes, both from the experimental and from the computational points of view. The samples might still contain a fraction of smaller entities leading to overlapped signals, or the precursor and larger units may not yet assume the ideal connectivity and degree of condensation assumed in the present computation. Computed values, on the other hand, should mainly be treated as a guide for a qualitative trend. Further investigation is needed to clarify these issues and better exploit the IR tool.

Conclusions

Because experimental characterization of silica nanoparticles is particularly difficult, shifts in IR bands of freeze-dried samples from clear solution have been used to define certain stages in the synthesis procedure of silicalite-1. In this paper, we have used molecular modeling techniques to verify whether such a shift should indeed be expected and to identify the position of the IR peaks for a complete range of structures. We have provided a strong foundation from a modeling viewpoint for the experimental observation of a red-shifting band, which starts out broad but narrows as the structure becomes more defined and condensed. This is not a continuous process as nanoparticle size increases, but it is highly dependent on the specific way monomers have been added. Since we have included simulation of the smallest possible five-membered rings, this shift spans a range of 100 cm^{-1} . However, the narrow MFI fingerprint peak around 550 cm^{-1} becomes strongly present only once the true MFI structure is defined in nanoslabs and nanoblocks. The strongest red shift occurs in the broad transition area, where small precursor-type particles exhibit both the typical five-ring vibrations and the MFI fingerprint band, caused by the vibrations

MFI Fingerprint

of a double five-membered ring as proposed by Jacobs et al.²⁰ Our simulations clearly reveal that pentasil rings at the particle boundary vibrate at higher wavenumbers than more embedded ones. Therefore, it can be expected that the IR spectrum will be sensitive to both particle size and morphology.

From these results it is clear that FTIR is a powerful tool in the characterization of silica nanoparticles. In combination with other techniques to determine size and shape, FTIR should be able to give a more defined picture of the initial stages in zeolite synthesis.

Acknowledgment. This work was sponsored by the Belgian government through an Interuniversity Attraction Pole program (IAP-PAI). The authors acknowledge support from the Fund for Scientific Research Flanders (FWO-Vlaanderen). The authors also acknowledge the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) for funding this strategic basic research (SBO).

Supporting Information Available: Force field expressions and parameters (pdf). This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Guisnet, M.; Gilson, J.-P. *Catalytic Science Series*, Vol. 3; Imperial College Press: London, 2002.
- (2) Davis, M. E. *Nature* **2002**, *417*, 813–821.
- (3) Schoeman, B. J.; Sterte, J.; Otterstedt, J. E. *Zeolites* **1994**, *14*, 110–116.
- (4) Persson, A. E.; Schoeman, B. J.; Sterte, J.; Otterstedt, J. E. *Zeolites* **1994**, *14*, 557–567.
- (5) Flanigen, E. M.; Bennett, J. M.; Grose, R. W.; Cohen, J. P.; Patton, R. L.; Kirchner, R. M.; Smith, J. V. *Nature* **1978**, *271*, 512–516.
- (6) Schoeman, B. J. *Microporous Mesoporous Mater.* **1998**, *22*, 9–22.
- (7) Ravishankar, R.; Kirschhock, C. E. A.; Knops-Gerrits, P. P.; Feijen, E. J. P.; Grobet, P. J.; Vanoppen, P.; De Schryver, F. C.; Mielhe, G.; Fuess, H.; Schoeman, B. J.; Jacobs, P. A.; Martens, J. A. *J. Phys. Chem. B* **1999**, *103*, 4960–4964.
- (8) Kirschhock, C. E. A.; Ravishankar, R.; Van Looveren, L.; Jacobs, P. A.; Martens, J. A. *J. Phys. Chem. B* **1999**, *103*, 4972–4978.
- (9) Watson, J. N.; Iton, L. E.; Keir, R. I.; Thomas, J. C.; Dowling, T. L.; White, J. W. *J. Phys. Chem. B* **1997**, *101*, 10094–10104.
- (10) de Moor, P.; Beelen, T. P. M.; van Santen, R. A. *J. Phys. Chem. B* **1999**, *103*, 1639–1650.
- (11) Auerbach, S. M.; Ford, M. H.; Monson, P. A. *Curr. Opin. Colloid Interface Sci.* **2005**, *10*, 220–225.
- (12) Provis, J. L.; Vlachos, D. G. *J. Phys. Chem. B* **2006**, *110*, 3098–3108.
- (13) Van Erp, T. S.; Caremans, T. P.; Kirschhock, C. E. A.; Martens, J. A. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1044–1051.
- (14) Schüth, F. *Curr. Opin. Solid State Mater. Sci.* **2001**, *5*, 389–395.
- (15) Wu, M. G.; Deem, M. W. *J. Chem. Phys.* **2002**, *116*, 2125–2137.
- (16) Jacobs, P. A.; Derouane, E. G.; Weitkamp, J. *J. Chem. Soc., Chem. Commun.* **1981**, 591–593.
- (17) Coudrier, G.; Naccache, C.; Védrine, J. C. *J. Chem. Soc., Chem. Commun.* **1982**, 1413–1415.
- (18) Kragten, D. D.; Fedeyko, J. M.; Sawant, K. R.; Rimer, J. D.; Vlachos, D. G.; Lobo, R. F.; Tsapatsis, M. *J. Phys. Chem. B* **2003**, *107*, 10006–10016.
- (19) Mohamed, R. M.; Aly, H. M.; El-Shahat, M. F.; Ibrahim, I. A. *Microporous Mesoporous Mater.* **2005**, *79*, 7–12.
- (20) Jacobs, P. A.; Beyer, H. K.; Valyon, J. *Zeolites* **1981**, *1*, 161–168.
- (21) Scarano, D.; Zecchina, A.; Bordiga, S.; Geobaldo, F.; Spoto, G.; Petrini, G.; Leofanti, G.; Padovan, M.; Tozzola, G. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 4123–4130.
- (22) Kirschhock, C. E. A.; Ravishankar, R.; Verspeurt, F.; Grobet, P. J.; Jacobs, P. A.; Martens, J. A. *J. Phys. Chem. B* **1999**, *103*, 4965–4971.
- (23) Serrano, D. P.; van Grieken, R. *J. Mater. Chem.* **2001**, *11*, 2391.
- (24) Hsu, C. Y.; Chiang, A. S. T.; Selvin, R.; Thompson, R. W. *J. Phys. Chem. B* **2005**, *109*, 18804–18814.
- (25) Ravishankar, R.; Kirschhock, C.; Schoeman, B. J.; De Vos, D.; Grobet, P. J.; Jacobs, P. A.; Martens, J. A. *Proceedings of the 12th International Zeolite Conference*, Vol. III; Treacy, M. M. J., Marcus, B. K., Bisher, M. E., Jiggins, J. B., Eds.; Materials Research Society: Warrendale, PA, 1999; p 1825.
- (26) Verstraelen, T.; Van Neck, D.; Ayers, P. W.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Theory Comput.* **2007**, *3*, 1420–1434.
- (27) Verstraelen, T.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Inf. Model.*, in press.
- (28) Wood, S. N. *J. R. Stat. Soc. B* **2000**, *62*, 413–428.
- (29) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502–16513.
- (30) Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1981**, *74*, 4872–4882.
- (31) Jansen, J. C.; Vandergaag, F. J.; Vanbekkum, H. *Zeolites* **1984**, *4*, 369–372.
- (32) Kirschhock, C. E. A.; Kremer, S. P. B.; Grobet, P. J.; Jacobs, P. A.; Martens, J. A. *J. Phys. Chem. B* **2002**, *106*, 4897–4900.
- (33) Ravishankar, R.; Kirschhock, C.; Schoeman, B. J.; Vanoppen, P.; Grobet, P. J.; Storck, S.; Maier, W. F.; Martens, J. A.; De Schryver, F. C.; Jacobs, P. A. *J. Phys. Chem. B* **1998**, *102*, 2633–2639.

JP711550S

Paper V

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

H. L. Woodcock, W. Zheng, A. Ghysels, Y. Shao, J. Kong and
B. R. Brooks

Journal of Chemical Physics 129 (21), 214109 (2008)

Copyright 2008 by American Institute of Physics

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

H. Lee Woodcock,^{1,a)} Wenjun Zheng,^{1,b)} An Ghysels,² Yihan Shao,³ Jing Kong,³ and Bernard R. Brooks¹

¹Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

²Center for Molecular Modeling, Ghent University, Proefuinststraat 86, B-9000 Gent, Belgium

³Q-Chem Inc., The Design Center, Suite 690, 5001 Baum Boulevard, Pittsburgh, Pennsylvania 15213, USA

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

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

I. INTRODUCTION

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

Normal mode analysis (NMA) methods have been proposed and utilized to extend the range of biomolecules researchers are capable of examining.⁵⁻⁹ This ability is predicated on the theory that harmonic dynamics are sufficient for understanding the collective motions of near-native states with low frequency normal modes dominating the qualitative

dynamics.^{7,10-14} This assertion has been shown to be a good approximation by numerous groups¹⁵⁻²³ and has led to the development of elastic network model (ENM) methods.²⁻⁴ These methods typically replace all-atom descriptions with C_α -centered harmonic potentials that utilize a single force constant to account for all pairwise interactions within some cutoff distance (R_c). For example, the energy of an ENM representation can be defined as

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

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

In the current work we present the vibrational subsystem analysis (VSA) method for coupling global motion to a local subsystem. This method is a partitioning scheme that separates (and integrates out) the motion of the environment from a user defined subsystem (see Methods section) while still allowing the environmental motion to perturb the local subsystem dynamics. It was originally applied to ENM models²⁵ but is now generalized and extended for all-atom representations and hybrid quantum mechanical/molecular mechanical (QM/MM) potentials. Below is a brief list of possible uses.

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

^{a)}Electronic mail: hlwood@nih.gov.

^{b)}Current address: Department of Physics, University at Buffalo, The State University of New York, Buffalo, New York 14260-1500, USA.

- include a mass term in the ENM-NMA procedure which allows for better coupling of timescales since the inertial terms in the environment are handled more appropriately.
- (2) Eliminating specific degrees of freedom without the detrimental effects of constraining the motion (i.e., making the system too rigid) or deleting part of the system (i.e., artificially increasing flexibility). A particularly useful illustration of this is in the “tip effect” (Ref. 27) which refers to small pieces of proteins that can protrude out of the main globular body (e.g., surface loops, unresolved density, etc.). Using the VSA method these problematic degrees of freedom can be integrated out of the environment while still being included as a perturbation to the subsystem of interest.
 - (3) Elimination of noise when computing harmonic vibration free energy of large biomolecular systems. The term noise is used to depict unwanted weak coupling between degrees of freedom in the environment and degrees of freedom in the subsystem that makes it difficult to calculate properties of interest. This is particularly important when performing QM/MM free energy calculations as the classical entropic effects from the environment can be included as a first order perturbation to the QM harmonic free energy. In addition, this can also be a very useful tool in describing how global motion of a protein can couple to the active site before and during biochemical reactions.
 - (4) Combining VSA with simulation approaches. For example, the use of quasiharmonic analysis to generate an effective force constant matrix from a detailed simulation can allow exploration of “hidden” couplings that ENM models cannot capture. In general, the VSA approach can be combined with numerous additional normal mode treatments; such as reduced basis methods (i.e., block normal mode method) to further reduce the number of degrees of freedom treated.²⁸
 - (5) Inclusion of very light or mass-less particles into NMA without the need for constraints or inclusion of unwanted high frequency heat capacity. This could be particularly useful as Drude oscillator approaches become a standard classical polarization technique.
 - (6) One of the most popular reaction path following techniques currently in use is the intrinsic reaction coordinate (IRC) method.^{29,30} The IRC method, which requires knowledge of the transition state, employs internal coordinates to map the minimum energy pathway connecting reactants and products. Although this is an effective approach, IRC mapping is typically not performed on high dimensional systems (i.e., proteins) for various reasons. One benefit of the VSA approach is that by integrating out the environment, rather than fixing or deleting it, second order methods that have previously been sparingly used (i.e., IRC) will again become available. Taking this in combination with the full QM/MM analytic second derivatives introduced in the current work QM/MM IRC calculations on biomolecular systems could easily gain mainstream use.

In Sec. II (Methods) we improve the original approach for vibrational subsystem analysis²⁶ and derive the formula for including environmental inertial effects in the VSA method in two different ways. Also described is the extension of the vibrational capabilities of CHARMM to utilize hybrid QM/MM potentials. In Sec. III (Results and Discussion) we give four examples of where VSA can be applied and discuss how the results correspond to full NMA; (1) VSA is applied to a simple harmonic string of beads and computed results are compared to analytically derived results. (2) VSA is applied to the torsional potential of butane with only carbon atoms defined as the subsystem. (3) VSA is applied to the cyclohexane chair to boat conversion with only carbon atoms defined as the subsystem. (4) QM (B3LYP/6-31G*) NMA, VSA, and block normal mode (BNM) methods are applied to a peptide radical rearrangement. Full NMA, VSA, and BNM results are compared and implications of the choice of subsystem is explored. Section IV (Conclusions) is an overview of the VSA method with results and future directions highlighted.

II. METHODS

A. The VSA method

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

$$2E_{\text{pot}} = x^T H x = \begin{bmatrix} x_s \\ x_e \end{bmatrix}^T H \begin{bmatrix} x_s \\ x_e \end{bmatrix} = x_s^T H_{ss} x_s + x_s^T H_{se} x_e + x_e^T H_{es} x_s + x_e^T H_{ee} x_e, \quad (2)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and consequently

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

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

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

B. Hybrid QM/MM analytic second derivatives

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

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

	QM-QM	QM-MM/MM-QM	MM-MM
$PH^{\nu\gamma}$	\checkmark	\checkmark	\checkmark
$P\ ^{xy}P$	\checkmark	\times	\times
$(PFP)^{S^{\nu\gamma}}$	\checkmark	\times	\times
$P^{\nu}H^{\gamma}$	\checkmark	\checkmark	\checkmark
$P^{\nu}\ ^{xy}P$	\checkmark	\checkmark	\times
$(PFP)^{\nu}S^{\gamma}$	\checkmark	\checkmark	\times
$\gamma^{\nu\gamma}$	\checkmark	\checkmark	\checkmark

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

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

$$E = \langle PH \rangle + \frac{1}{2} \langle P \| P \rangle + \gamma, \quad (21)$$

where P is the density matrix, H is the core Hamiltonian, and $\|$ are the antisymmetrized two-electron integrals over spin orbitals $\mu, \nu, \lambda, \sigma$, and γ is the nuclear repulsion energy. The Fock operator is then defined as

$$F = H + P \| \quad (22)$$

and the standard SCF can be written as

$$FPS = SPF, \quad (23)$$

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

$$\frac{\partial^2 E}{\partial x \partial y} = \langle PH^{\nu\gamma} \rangle + \frac{1}{2} \langle P \|^{xy} P \rangle + \langle P^{\nu} H^{\gamma} \rangle + \langle P^{\nu} \|^{xy} P \rangle - \langle (PFP)^{\nu\gamma} S^{\gamma} \rangle + \gamma^{\nu\gamma}. \quad (24)$$

At this point the coupled perturbed equations (CPHF for Hartree-Fock and CPKS for DFT) need to be solved. However, in contrast to standard QM theory the Hessian matrix has four blocks: QM-QM, QM-MM, MM-QM, and MM-MM with the QM-MM and MM-QM block being equivalent. At first glance solving the CP equations would appear to be impossible for very large (thousands of atoms) systems, but given the nature of classical point charge terms involving overlap matrix derivatives ($S^{\nu}, S^{\nu\gamma}$) and two-electron integral derivatives ($\|^{xy}, \|^{xy}$) can be dropped (Table I). This simplification, combined with improved techniques for solving CP equations³⁴ lead to the case where problems that were previously impossible to examine are now feasible.

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

In addition to implementing QM/MM analytic second derivatives, we also added eigenvector following routines to the Newton-Raphson minimization routines in CHARMM. This is done in a flexible manner that allows efficient minimization both in and out of a harmonic well. If the system

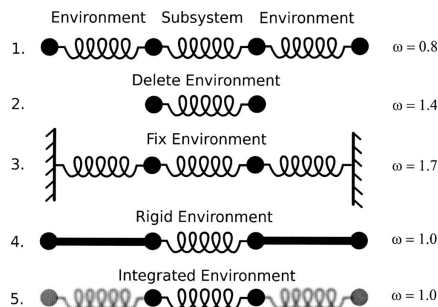


FIG. 1. Illustration and results of harmonic beads on strings. The subsystem is defined as the middle two beads with various treatments applied to the "environment." (1) Full treatment, (2) environment deleted, (3) environment fixed, (4) rigid environment, and (5) environment integrated out (i.e., VSA).

can be approximated harmonically then standard Newton-Raphson minimization occurs making use of the energy (E), gradient of the energy (∇E) and Hessian of the energy ($\nabla^2 E$). However, if the system cannot be approximated as a set of harmonic oscillators then a line search is performed along the specified eigenvector(s). In addition to the three pieces of information already gathered (vide supra) two additional gradient calculations are performed along the positive and negative directions of eigenvector(s). These five pieces of information are then fit to a third order polynomial ($f(x) = ax^3 + bx^2 + cx + d$) and solved via a linear least squares fit procedure.

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

III. RESULTS AND DISCUSSION

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

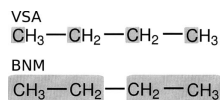


FIG. 2. Illustration of the schemes used to examine the butane anti to eclipsed rotation. The top scheme was employed with the VSA method with the bottom scheme being used with the BNM method.

A. Harmonic beads on springs

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

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

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

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

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

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

cluded [Eq. (18)] in all results and is clearly needed to account for the inertia of the environment. For example, the frequency of the VSA result would be 2.0 cm⁻¹ if inertial effects are ignored.

B. Torsional barrier of butane

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

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



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

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

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

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

C. Chair to boat inversion of cyclohexane

The chair to boat interconversion pathway of cyclohexane is examined using both the all atom and VSA models (Fig. 3). The subsystem employed includes only the ring carbons of cyclohexane which is a 67% reduction in system size. Additionally, we compute the ΔG_{298} between the four extreme points located along the pathway: (1) the global minimum chair conformer, (2) the global maximum/

transition state, (3) a local minimum twist-boat conformer, and (4) the boat conformer which is a local transition state. All transition states are located using the newly implemented saddle point finder. The free energies are defined relative to the global minimum (i.e., chair conformer) and are listed in Table IV.

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

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

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

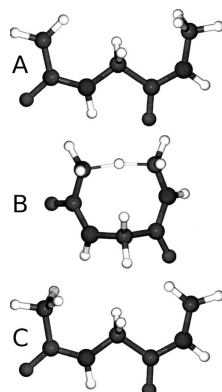


FIG. 4. Illustration of the reactant (a), transition state (b), and product (c) of the hydrogen transfer reaction in the glycine-glycine dipeptide radical.

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

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

D. Quantum mechanical peptide radical rearrangement

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

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

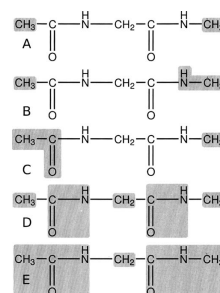


FIG. 5. Definition of subsystem and blocking schemes used to examine the hydrogen transfer reaction in the glycine-glycine dipeptide radical. Schemes (a), (b), and (c) were applied to the vibrational subsystem analysis method while schemes (d) and (e) were employed with the BNM method. See Table VI for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

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

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

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

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

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

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

IV. CONCLUSIONS

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

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

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

ACKNOWLEDGMENTS

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

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

Paper V

214109-9 Vibrational subsystem analysis

J. Chem. Phys. **129**, 214109 (2008)

- ³⁴M. Frisch, M. Head-Gordon, and J. Pople, *Chem. Phys.* **141**, 189 (1990).
³⁵A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, *J. Chem. Phys.* **126**, 224102 (2007).
³⁶A. Ghysels, D. Van Neck, and M. Waroquier, *J. Chem. Phys.* **127**, 164108 (2007).

- ³⁷A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier, *J. Chem. Theory Comput.* **4**, 614 (2008).
³⁸D. Moran, R. Jacob, G. P. F. Wood, M. L. Coote, M. J. Davies, R. A. J. O'Hair, C. J. Easton, and L. Radom, *Helv. Chim. Acta* **89**, 2254 (2006).

Paper VI

Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach

A. Ghysels, D. Van Neck, B. R. Brooks, V. Van Speybroeck and
M. Waroquier

Journal of Chemical Physics 130 (1), 084107 (2009)

Copyright 2009 by American Institute of Physics

Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach

A. Ghysels,¹ D. Van Neck,¹ B. R. Brooks,² V. Van Speybroeck,¹ and M. Waroquier^{1,a)}¹Center for Molecular Modeling, Ghent University, Proefuinstraat 86, B-9000 Gent, Belgium²Laboratory of Computational Biology, National Heart Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

(Received 11 November 2008; accepted 22 December 2008; published online 25 February 2009)

In a previous paper [Ghysels *et al.*, *J. Chem. Phys.* **126**, 224102 (2007)] the mobile block Hessian (MBH) approach was presented. The method was designed to accurately compute vibrational modes of partially optimized molecular structures. The key concept was the introduction of several blocks of atoms, which can move as rigid bodies with respect to a local, fully optimized subsystem. The choice of the blocks was restricted in the sense that none of them could be connected, and also linear blocks were not taken into consideration. In this paper an extended version of the MBH method is presented that is generally applicable and allows blocks to be adjoined by one or two common atoms. This extension to all possible block partitions of the molecule provides a structural flexibility varying from very rigid to extremely relaxed. The general MBH method is very well suited to study selected normal modes of large macromolecules (such as proteins and polymers) because the number of degrees of freedom can be greatly reduced while still keeping the essential motions of the molecular system. The reduction in the number of degrees of freedom due to the block linkages is imposed here directly using a constraint method, in contrast to restraint methods where stiff harmonic couplings are introduced to restrain the relative motion of the blocks. The computational cost of this constraint method is less than that of an implementation using a restraint method. This is illustrated for the α -helix conformation of an alanine-20-polypeptide. © 2009 American Institute of Physics. [DOI: 10.1063/1.3071261]

I. INTRODUCTION

The calculation of normal modes of not only an extended molecular system such as polypeptides and proteins but also polymer chains, supramolecular assemblies, systems embedded in a solvent, (macro)molecules adsorbed within porous materials, etc., is still a computationally intensive task that often goes beyond the capabilities of currently available computational hardware and software. The calculation of normal modes—often referred to as normal mode analysis (NMA)—is, however, essential for many purposes, ranging from the evaluation of chemical kinetics and thermodynamic properties such as entropies and free energies to the analysis of infrared and Raman spectra.^{1–6}

Frequency calculations need the construction of the full (all-atom) Hessian determined by the complete set of $3N_a$ degrees of freedom (with N_a the number of atoms). For large systems such as proteins, this can be beyond computational means. Even when feasible, the resulting huge amount of normal modes is often hard to analyze and interpret, and there is a need to filter out the relevant motions.

Another complicating factor in the NMA of extended systems is the fact that the geometry of these systems is often not fully optimized. An example of such a partial optimization is the simulation of reactions at surfaces, where the reactive center's geometry is optimized but the rest of the sur-

face is kept fixed during the optimization.⁷ The computation of the frequencies in such partially optimized systems is far from trivial. The main problem is that the global system is in nonequilibrium, and application of the standard NMA procedure is wrong, leading to wrong frequencies,⁸ which may even include a series of spurious imaginary frequencies.

One is frequently confronted with this kind of problems, and a proper solution is strongly needed. Appropriate methods have already been published recently; one of them is the mobile block Hessian (MBH) approach,⁸ although other related methods exist, which will be discussed later. The key concept of the MBH is the partitioning of the system into several blocks of atoms that move as rigid bodies during the vibrational analysis. Atoms that do not belong to any block are considered as free atoms. Forces on the atoms within the blocks may still be present because of the partial optimization; however, the position of the free atoms and the position and orientation of each block are assumed to be optimized.

In general, any system can be partitioned into blocks. Within each block the level of theory used for the optimization can differ and the atoms of some blocks may be nonoptimized. Blocks can be connected to each other by means of chemical bonds or nonbonding interactions or they can share one or two atoms. The usefulness and merit of this block concept have already been proven in some application fields: (i) The partition of a protein into blocks containing multiple amino-acid residues was shown to be very efficient for determining the normal modes of large proteins.^{9,10} (ii) In

^{a)}Author to whom correspondence should be addressed. Electronic mail: michel.waroquier@ugent.be.

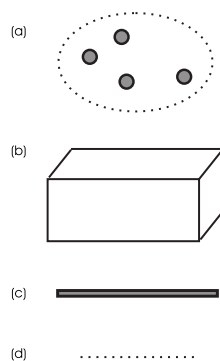


FIG. 1. Ingredients for a schematic depiction of a block partition of a molecule: an assembly of free atoms (a), and rigid nonlinear (b) or linear (c) blocks. A dotted line between two ingredients represents a chemical bond or nonbonding interaction (d).

chemical reactions only a fraction of the atoms is directly involved in bond breaking and forming or in a change in bond order, and the internal geometry of “spectator” chemical groups may be fixed.^{11–15}

A Cartesian formulation of the MBH approach was presented in Ref. 16, allowing a simple implementation in existing modeling packages. Up to now the MBH approach is already implemented in the last upgrade of CHARMM (Ref. 17) and Q-CHEM,¹⁸ and will be available too in the next release of ADF.¹⁹ The ability of MBH to accurately predict reaction rate constants has been validated in Ref. 13. It was shown there that, on condition of a suitable choice of the blocks, MBH includes the motions that are essential for a correct description of the reaction.

The choice of the blocks is crucial for the functionality of MBH and depends on what the real study object is. Normal modes usually involve the correlated motion of a number of atoms, but they can be divided into two main categories: the low-frequency modes where large parts of the molecule are involved moving in a collective way, and the local modes at higher frequency, localized in a specific part of the molecule, like the C=O stretch of the amide carbonyl group in polypeptides and proteins.⁴ Both categories have their practical usefulness in the prediction of chemical and physical properties, and in either case block choices can be introduced for an accurate reproduction of the relevant normal modes.⁸ The overlap with the full Hessian normal modes was used as a measure for the quality of the approximate MBH modes. Other measures are possible, for instance, by gauging how much an approximate normal mode deviates from an eigenvector of the full Hessian matrix.²⁰

Once the blocks have been fixed and optimized according to a certain level of theory, the internal motions in each block are frozen: blocks move as rigid bodies, but their relative motion is still free (apart from constraints imposed by common atoms). This is illustrated in Fig. 1, displaying the various ingredients entering possible partitions of a molecule

into blocks: (a) assemblies of free atoms, (b) linear and nonlinear rigid blocks containing atoms whose internal coordinates are frozen, and (c) dotted lines connecting the various bodies if no atom is shared by two or more blocks. This serves as a reminder that neighboring blocks without common atoms are not moving completely independently, but are connected by chemical bonds or nonbonding interactions. The ingredients of Fig. 1 can now be adjoined together in several ways, as specified in Fig. 2. For each topology a molecular example is given. The previous MBH method was restricted to a partitioning of the system following the scheme displayed in Fig. 2(a). This limits the block choices available in the existing MBH method: linear blocks were excluded and all blocks were assumed to be nonadjoined (i.e., connected by chemical bonds or nonbonding interactions only, with no adjoining atoms shared between the blocks).

The introduction of adjoining atoms in the partition scheme of a molecule, in particular, is a promising generalization of the MBH concept in systems with a large number of atoms because of the serious reduction in the degrees of freedom. One can distinguish between singly adjoined blocks connected by one common adjoining atom as displayed [see Figs. 2(b)–2(d)] and doubly adjoined or hinge blocks connected by two common adjoining atoms [see Fig. 2(e)]. Apart from global translation/rotation, two singly adjoined nonlinear blocks have three additional degrees of freedom (instead of six for nonadjoined blocks), while two hinge blocks have one degree of freedom corresponding to a hinge-bending motion. Multiple adjoining blocks can lead to a ring structure as shown in Fig. 2(f). All topologies can be constructed by combining all nonadjoined and adjoined (non)linear blocks.

The MBH method has been extended in a recent paper,²¹ to accommodate adjoining blocks sharing one or two atoms, as shown in Figs. 2(b)–2(f). It was found that the adjoined MBH model works well in extended systems where one is mainly interested in the low-frequency modes. For example, in peptide chains, the peptide unit is known to be quite rigid. The MBH approach with adjoining blocks then enables to connect various peptide units by the C_α atom. The main advantage of the adjoining blocks is that the number of degrees of freedom is further decreased, which is highly recommended in the case of extended systems with thousands of atoms.

The extension of the standard MBH to adjoining blocks was performed by introducing dummy atoms (see Sec. 2 of Ref. 21). An adjoining atom shared by several blocks is formally duplicated in each of those blocks. Adding a strong harmonic coupling between the dummies, the system behaves as if the blocks were linked. This technique was shown to work well, but has a significant disadvantage in the sense that the introduction of the dummy atoms leads to an extension of the dynamical system. Thereby the dimension of the Hessian temporarily increases in size, and for very large molecular systems this can be quite bothersome and time consuming. Moreover, the strong coupling between the dummies can cause numerical instabilities in the diagonalization of the Hessian for very large systems.

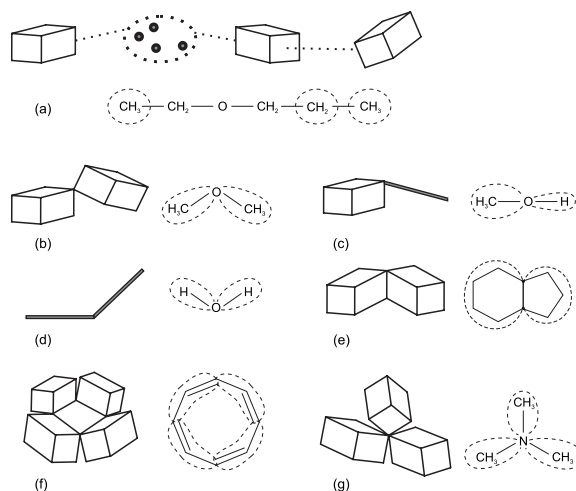


FIG. 2. With the ingredients of Fig. 1 various topologies can be constructed. In each case a molecular example is given. (a) An example of nonadjoined blocks. Nonlinear blocks have six; linear blocks have five degrees of freedom. [(b)–(d)] Some examples of adjoined blocks with one adjoining atom. (e) A hinge-type connection with blocks having two adjoining atoms gives rise to a hinge-bending motion with one rotational degree of freedom. (f) A ring structure consisting of singly adjoined blocks. (g) Example of three blocks with one adjoining atom, shared by the three blocks.

In order to overcome these shortcomings of the dummy atom concept, an exact and more direct treatment of the MBH with adjoined blocks is proposed in this paper. The main idea is to introduce block parameters for each block and then, in a second step, to impose the link constraints on the set of block parameters. The implementation avoids the extra degrees of freedom inherent to dummy atoms and is able to perform a MBH analysis for any arbitrary topology, even in the case of a ring structure of blocks.

Besides MBH there are other models in literature for the calculation of normal mode frequencies in extended systems. In particular, the partial Hessian vibrational analysis (PHVA) approach should be mentioned. It has been introduced by Head and co-workers^{22,7} and further investigated by Li and Jensen.¹² Normal modes are computed for the system with the fixed atoms frozen at their reference positions as if they were given an infinite mass and only the relaxed atoms can participate in the vibrations. An improved version of the PHVA was presented by Head,^{23,11,24} but this method requires the knowledge of an additional off-diagonal block of the Hessian. Another application of the partial Hessian approach is given by Besley and Metcalf.⁸ They compute a partial Hessian comprising the derivatives of the energy with respect to the movement of some selected atoms of relevance for their final goal. Their method has been applied successfully for the calculation of the amide I band in polypeptides and proteins. They limited the selection to the carbon and oxygen atoms of the backbone carbonyl groups and the amide nitrogens. NMA based on a simplified energy functional is also frequently used in biophysical chemistry.¹⁰ The Hessian matrix can also be reduced using group theoretical

methods based on the symmetry properties of the macromolecule. For highly symmetric systems this procedure is very appropriate and provides an excellent description of the low-frequency normal modes of proteins, which are linked to large-amplitude collective motions.²⁵ Once the symmetry becomes low and the molecular system involves heterogeneous distributions of different chemical groups, the concept of blocks has been introduced.^{26,27} The merit of the block normal mode (BNM) method as an effective coarse-grained NMA approach is that the cost of computation is dramatically reduced as compared to taking full account of intermolecular interactions. BNM is closely related to the rotations-translations of blocks (RTB) method developed by Tama *et al.*,⁹ which proved to be very efficient and adequate for NMA of proteins. The RTB method relies on the hypothesis that the low-frequency normal modes of proteins can be described as pure rigid-body motions of blocks of consecutive amino-acid residues. These frequencies seem to depend little on the size of the block (number of amino acids per block), strongly supporting the above hypothesis. By construction, RTB (or BNM) and MBH (in particular, the previous version) show large similarities: rigid blocks having six degrees of freedom (translation-rotation). The main difference lies in the fact that MBH is able to compute accurately normal modes in nonequilibrium (partially optimized) block systems, by taking into account gradient correction terms. These corrections are not involved in RTB (or BNM). In addition RTB (or BNM) is suitable for handling proteins, while the application field of MBH is much broader. Another method for NMA in large biomolecular systems is the vibrational subsystem analysis (VSA) method presented by Woodcock *et*

*al.*²⁸ This method is a partitioning scheme that separates (and integrates out) the motion of the environment from the user-defined subsystem. It is ideally suited for hybrid quantum mechanical/molecular mechanical potentials. A comparative study of the different approaches (PHVA, MBH, and VSA) will be presented in a forthcoming study.

In Sec. II we first present a more general formulation of the MBH, where the size of the blocks is arbitrary. A block can consist of one atom (corresponding to a free atom in the previous formulation of Ref. 8) or it may be a linear or nonlinear genuine block. Accordingly the block's number of degrees of freedom is three, five, or six. This view on the partition allows a transparent formulation of the MBH equations in terms of Cartesian quantities. The Cartesian gradient and Hessian are available in the output logfile of most simulation programs, such that the present MBH formulation is generally applicable. In Sec. III we then demonstrate that link constraints can be added in an elegant way. The only condition to be taken into account is that the geometry of the system should be partially optimized, such that the position/orientation of each block is optimized. In Sec. IV a practical numerical scheme resumes the steps in a general MBH analysis, indicating how MBH can be implemented. In Sec. V we report the results from the simulation of a α -helical peptide chain with 20 alanine residues.

II. ALL-BLOCK FORMULATION OF MBH

A. Transformations as dynamical variables

The MBH approach was introduced in previous papers^{8,16} as a flexible tool for performing vibrational analysis in the case of molecular systems where only a partial optimization is performed. Within the molecule, groups of neighboring atoms are identified as blocks, and the internal geometry (which need not be the equilibrium geometry) in each block is kept fixed during the motion. The different blocks have no atoms in common, but it is convenient to think of them as connected by chemical bonds or nonbonding interactions [see Fig. 2(a)].

The rigid-body motion of such a block b can be conveniently described by six parameters $p_{b\alpha}$ ($\alpha=1, \dots, 6$) of a translation/rotation group. The instantaneous positions $r_A(t)$ of the atoms A in the block are simply generated by applying a common transformation with parameters p_b to the reference positions r_A^0 ,

$$r_A(t) = g(r_A^0, p_b(t)), \quad \forall A \in b. \quad (1)$$

The parametrization of the transformation group is arbitrary, but we will stick to the convention introduced in Ref. 8 of successively rotating along the fixed z -, y -, and x -axes of a space-fixed frame, followed by a translation,

$$g(r, p) = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos p_4 & -\sin p_4 \\ 0 & \sin p_4 & \cos p_4 \end{pmatrix} \times \begin{pmatrix} \cos p_5 & 0 & \sin p_5 \\ 0 & 1 & 0 \\ -\sin p_5 & 0 & \cos p_5 \end{pmatrix} \begin{pmatrix} \cos p_6 & -\sin p_6 & 0 \\ \sin p_6 & \cos p_6 & 0 \\ 0 & 0 & 1 \end{pmatrix} \times \begin{pmatrix} x \\ y \\ z \end{pmatrix}. \quad (2)$$

The group variables p_b are used as the dynamical variables. Starting from a reference geometry that is optimized with respect to the p_b variables of the various blocks, the normal modes of this dynamical system can be calculated in a graceful way, while respecting the invariance under global rotations and translations of the molecular potential energy surface (PES).

In Ref. 16 the fixed-geometry blocks were treated separately from the remaining atoms, which could translate freely. This somewhat complicated the resulting expressions. In addition, linear blocks (which have only five degrees of freedom) were not considered explicitly. To remove these limitations, this section provides a unified framework solely in terms of block transformation parameters. This will facilitate, in the next stage, the development of the formalism in the presence of adjoining atoms.

We therefore assume that *all* the atoms in the molecule have been partitioned into blocks. The motion of the atoms in a block is still given by Eq. (1), but now three types of blocks are distinguished, depending on the number of its degrees of freedom d_b . The transformation $g(r_A^0, p_b)$ in Eq. (1) then depends on the number d_b of block variables as determined by the type of block.

- (1) For a normal block (containing at least three noncollinear atoms), $d_b=6$ and all parameters p_1, \dots, p_6 are needed.
- (2) For a single-atom block, $d_b=3$ and only the translational parameters p_1, p_2, p_3 are retained in the transformation (1).
- (3) For a linear block (containing two or more collinear atoms), $d_b=5$. The translational parameters p_1, \dots, p_3 and two of the rotational parameters are now retained in the block transformation (1). For such a linear block, we must exclude rotations about a particular axis (e.g., the μ -axis with $\mu=x, y,$ or z) of the space-fixed frame. Note that in this case the block transformations no longer form a group. However, for the purpose of deriving the normal modes it is sufficient that the transformation is twice differentiable and covers all possible positions of the linear block near its reference position. As we shall see, the latter condition is always fulfilled, unless the reference position of the linear block happens to be orthogonal to the μ -axis; this can be simply avoided by a suitable choice of μ , e.g., as the axis with largest projection $|r_{A\mu}^0 - r_{B\mu}^0|$ for two atoms in the linear block.

TABLE I. Coefficients $D^{(\alpha)}$ and $C^{(\alpha\alpha')}$ defined in Eq. (4), expressing the first and second derivatives of the transformation in Eq. (1). The $v_{b\beta}^{(\alpha)}$ coefficients are defined in Eq. (13). The Levi-Civita symbol $\epsilon_{\lambda\mu\nu}$ equals 1 (-1) if $\lambda\mu\nu$ is a cyclic (anticyclic) permutation of xyz and zero otherwise.

α	$D_{A\mu}^{(\alpha)}$	α	α'	$C_{A\mu}^{(\alpha\alpha')}$	d_b	α	β	$v_{b\beta}^{(\alpha)} = (\partial\Psi_{b\beta}/\partial P_\alpha)_0$
1	$\delta_{\mu x}$	1	1-6	0	3	1-6	1-3	$D_{A\beta}^{(\alpha)}$
2	$\delta_{\mu y}$	2	1-6	0	6	1-6	1-6	$\delta_{\alpha,\beta}$
3	$\delta_{\mu z}$	3	1-6	0	5	1-6, $\neq 3+\mu$	1-3	$\delta_{\alpha,\beta}$
4	$\sum_\lambda \epsilon_{\lambda\mu\nu} r_{A\lambda}^0$	4	4	$\delta_{\mu\nu} r_{A\lambda}^0 - r_{A\mu}^0$			4-6, $\neq 3+\mu$	$\delta_{\alpha,\beta}$
5	$\sum_\lambda \epsilon_{\lambda\mu\nu} r_{A\lambda}^0$	5		$\delta_{\mu\nu} r_{A\lambda}^0$		3+ μ	1-3	$\sum_{\gamma\delta} \epsilon_{\beta\gamma\delta} \frac{(r_{A'\beta\delta}^0 - r_{A'\beta\delta}^0)}{r_{A\mu}^0 - r_{B\mu}^0}$
6	$\sum_\lambda \epsilon_{\lambda\mu\nu} r_{A\lambda}^0$	6		$\delta_{\mu\nu} r_{A\lambda}^0$			4-6, $\neq 3+\mu$	$-\frac{(r_{A'\beta\delta}^0 - r_{A'\beta\delta}^0)}{r_{A\mu}^0 - r_{B\mu}^0}$
		5	5	$\delta_{\mu\nu} r_{A\lambda}^0 - r_{A\mu}^0$				
		6	6	$\delta_{\mu\nu} r_{A\lambda}^0$				
		6	6	$\delta_{\mu\nu} r_{A\lambda}^0 - r_{A\mu}^0$				

For a particular block, the allowed transformations on the coordinates of its atoms thus depend on a number d_b of block variables, determined by the type of block. Note that the reduced set of p -variables in the case of a linear block or free atom can be complemented with zeros for the missing variables in the transformation $\mathbf{g}(r_A^0, p_b)$ in Eq. (1). To avoid confusion, we will denote by \hat{p}_b this complemented set of six parameters.

B. Derivation of the normal mode equations

The PES expressed in the set of $\{p_{b\alpha}\}$ variables becomes

$$\tilde{V}(\{p_b\}) = V(\{\mathbf{g}(r_A^0, \hat{p}_{b(A)})\}), \quad (3)$$

where $V(\{r_A\})$ is the PES in Cartesian coordinates and $b(A)$ is the block to which atom A belongs.

Derivation of the normal modes proceeds using the same notations as in Ref. 16, i.e., $G_{A\mu} = (\partial V / \partial r_{A\mu})_0$ and $H_{A\mu, A'\mu'} = (\partial^2 V / \partial r_{A\mu} \partial r_{A'\mu'})_0$ for the gradient and Hessian of the PES in Cartesian coordinates, evaluated at the reference point, and

$$D_{A\mu}^{(\alpha)} = \frac{\partial g_\mu}{\partial p_\alpha}(r_A^0, 0), \quad C_{A\mu}^{(\alpha\alpha')} = \frac{\partial^2 g_\mu}{\partial p_\alpha \partial p_{\alpha'}}(r_A^0, 0) \quad (4)$$

for the first and second derivatives of the transformation (2) with respect to the p_α parameters, evaluated at the identity ($p_\alpha=0$). The $D^{(\alpha)}$ and $C^{(\alpha\alpha')}$ coefficients are tabulated in Table I.

The equilibrium position of the blocks is taken as the reference point, i.e., the gradients in the $p_{b\alpha}$ variables,

$$\tilde{G}_{b\alpha} = \left(\frac{\partial \tilde{V}}{\partial p_{b\alpha}} \right)_0 = \sum_{A \in b, \mu} G_{A\mu} D_{A\mu}^{(\alpha)} = 0, \quad (5)$$

all vanish. The Hessian in the $p_{b\alpha}$ variables then becomes

$$\begin{aligned} \tilde{H}_{b\alpha, b'\alpha'} &= \left(\frac{\partial^2 \tilde{V}}{\partial p_{b\alpha} \partial p_{b'\alpha'}} \right)_0 \\ &= \sum_{A \in b, \mu} \sum_{A' \in b', \mu'} H_{A\mu, A'\mu'} D_{A\mu}^{(\alpha)} D_{A'\mu'}^{(\alpha')} \\ &\quad + \delta_{b, b'} \sum_{A \in b, \mu} G_{A\mu} C_{A\mu}^{(\alpha\alpha')}. \end{aligned} \quad (6)$$

This form is identical to the Hessian derived in Eq. 65 of Ref. 16 except that now the indices α pertaining to a certain block b need not correspond to all six parameters of the full translation/rotation group, but can be restricted to five or three in order to accommodate linear blocks or single atoms. The same remark applies to the mass matrix [Eq. 59 in Ref. 16].

$$\tilde{M}_{b\alpha, b'\alpha'} = \delta_{b, b'} \sum_{A \in b, \mu} m_A D_{A\mu}^{(\alpha)} D_{A\mu}^{(\alpha')}, \quad (7)$$

and to the normal mode equations themselves, which will not be repeated here.

C. Zero modes of the Hessian

Invariance of the PES under a global translation/rotation with (six) parameters P_α can be expressed as

$$\tilde{V}(\{p_b\}) = V(\{\mathbf{g}(r_A^0, \Phi(\hat{p}_{b(A)}, P))\}), \quad (8)$$

where $\Phi(p, p')$ represents the six parameters of the combined transformation $\mathbf{g}(\mathbf{r}, p, p') = \mathbf{g}(\mathbf{r}, \Phi(p, p'))$ (see Ref. 16).

For a particular block, $\Phi(\hat{p}_{b(A)}, P)$ represents a six-parameter transformation that takes the atoms of the block to new positions. Since any position of the block can be uniquely described by d_b parameters, the effect of this six-parameter transformation $\Phi(\hat{p}_{b(A)}, P)$ on the block atoms can always be replaced by an equivalent d_b -parameter transformation, $\Psi_b(p_b, P)$. The latter is completely determined by the equivalence relation

$$\nabla A \in b: \mathbf{g}(r_A^0, \hat{\Psi}_b(p_b, P)) = \mathbf{g}(r_A^0, \Phi(p_b, P)). \quad (9)$$

The final translation/rotation invariance of the PES can now be expressed as

$$\tilde{V}(\{p_b\}) = \tilde{V}(\{\Psi_b(p_b, P)\}). \quad (10)$$

Taking the second derivative of Eq. (10) with respect to $p_{b\alpha}$ and neglecting gradient terms on account of Eq. (5), one obtains

$$0 = \left(\frac{\partial^2 \tilde{V}}{\partial p_{b\alpha} \partial P_\beta} \right)_0 = \sum_{b', \alpha'} \tilde{H}_{b\alpha, b'\alpha'} \left(\frac{\partial \Psi_{b\alpha'}}{\partial p_{b\alpha}} \right)_0 \left(\frac{\partial \Psi_{b'\alpha'}}{\partial P_\beta} \right)_0. \quad (11)$$

Since $\Psi_b(p_b, 0) = p_b$ one always has $(\partial \Psi_{b\alpha'} / \partial p_{b\alpha})_0 = \delta_{\alpha, \alpha'}$, and Eq. (11) can be simplified as

$$0 = \sum_{b', \alpha'} \tilde{H}_{b\alpha, b'\alpha'} \left(\frac{\partial \Psi_{b'\alpha'}}{\partial P_\beta} \right)_0. \quad (12)$$

By inspection, Eq. (12) implies that the Hessian \tilde{H} has six zero eigenvalues with corresponding eigenvectors $v^{(\beta)}$, whose components are given by

$$v_{b\alpha}^{(\beta)} = \left(\frac{\partial \Psi_{b\alpha}}{\partial P_\beta} \right)_0. \quad (13)$$

The $v_{b\alpha}^{(\beta)}$ are tabulated in Table I. Most of the required first-order derivatives follow straightforwardly from Eq. (9) evaluated at $p_b=0$,

$$\nabla A \in b: \mathbf{g}(r_A^0, \hat{\Psi}_b(0, P)) = \mathbf{g}(r_A^0, P). \quad (14)$$

The only nontrivial case occurs for a linear block ($d_b=5$). Suppose the rotations around the μ -axis are excluded and we want to find $v^{(\beta)}$ for $\beta=3+\mu$. Then Eq. (14) implies that the positions of two atoms A and B of the linear block after an infinitesimal rotation about the μ -axis must be reproduced by a combination of infinitesimal translations and rotations about the two other axes. With (μ, κ, λ) in cyclical order, $(\delta a_\mu, \delta a_\kappa, \delta a_\lambda)$ the rotation angles about the respective space-fixed axes, and $(\delta t_\mu, \delta t_\kappa, \delta t_\lambda)$ the corresponding translation parameters, we can use a first-order expansion of Eq. (2) and express the equivalence of the positions of A and B under both transformations. This leads to three equations for atom A ,

$$\begin{aligned} \delta t_\kappa + \delta a_\lambda r_{A\mu}^0 &= -\delta a_\mu r_{A\lambda}^0, \\ \delta t_\lambda - \delta a_\kappa r_{A\mu}^0 &= \delta a_\mu r_{A\kappa}^0, \\ \delta t_\mu - \delta a_\lambda r_{A\kappa}^0 + \delta a_\kappa r_{A\lambda}^0 &= 0, \end{aligned} \quad (15)$$

and three analogous ones for atom B , which should be solved for the five unknown variables $\delta t_\mu, \delta t_\kappa, \delta t_\lambda, \delta a_\kappa,$ and δa_λ . Provided that $r_{A\mu}^0 \neq r_{B\mu}^0$ (or equivalently, that the linear block is not orthogonal to the μ -axis), there is a unique solution

$$\delta a_\lambda = -(\delta a_\mu) \frac{r_{A\lambda}^0 - r_{B\lambda}^0}{r_{A\mu}^0 - r_{B\mu}^0},$$

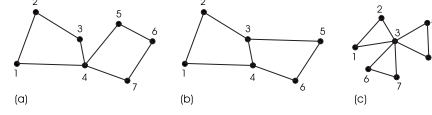


FIG. 3. Example of adjoining atoms. (a) Two blocks [1,2,3,4],[4,5,6,7] having one adjoining atom (4) in common. Keeping the position of [1,2,3,4] fixed, the relative motion of [4,5,6,7] is limited to rotations about an axis through the adjoining atom. (b) Two blocks [1,2,3,4],[3,4,5,6] having two adjoining atoms (3,4) in common (=hinge). Keeping the position of [1,2,3,4] fixed, the relative motion of [3,4,5,6] is limited to hinge-bending motions about the axis through the two adjoining atoms. (c) Three blocks [1,2,3],[3,4,5],[3,6,7] having one adjoining atom (3) in common. The share number of atom 3 is 3.

$$\begin{aligned} \delta a_\kappa &= -(\delta a_\mu) \frac{r_{A\kappa}^0 - r_{B\kappa}^0}{r_{A\mu}^0 - r_{B\mu}^0}, \\ \delta t_\lambda &= (\delta a_\mu) \frac{r_{A\mu}^0 r_{B\kappa}^0 - r_{A\kappa}^0 r_{B\mu}^0}{r_{A\mu}^0 - r_{B\mu}^0}, \\ \delta t_\kappa &= (\delta a_\mu) \frac{r_{A\lambda}^0 r_{B\mu}^0 - r_{A\mu}^0 r_{B\kappa}^0}{r_{A\mu}^0 - r_{B\mu}^0}, \\ \delta t_\mu &= (\delta a_\mu) \frac{r_{A\kappa}^0 r_{B\lambda}^0 - r_{A\lambda}^0 r_{B\kappa}^0}{r_{A\mu}^0 - r_{B\mu}^0}. \end{aligned} \quad (16)$$

This solution determines precisely the first-order derivatives of Ψ_b appearing in Eq. (13) and gives rise to the corresponding components of $v^{(\beta)}$ as listed in Table I.

It follows that the six zero-mode eigenvectors of the Hessian are analytically known for any given block structure of the molecule. Just as in Ref. 16 one has the option of decoupling them explicitly from the intrinsic normal modes, using congruent transformations on \tilde{H} and \tilde{M} .

III. MBH FOR SYSTEMS WITH LINK CONSTRAINTS

A. Formulating the link constraints

In this section the previous MBH method is generalized to adjoining blocks, a significant extension opening a lot of new perspectives (see Sec. I). From now on we assume that the relative motion of some of the blocks is no longer independent, but obeys some constraints. This can be described by allowing some atoms, henceforth called adjoining atoms, to belong to two (or more) blocks. Using adjoining atoms to connect blocks, a large variety of topologies becomes possible (see Fig. 2). In addition, some of the blocks may be linear and have one degree of freedom less than normal blocks. It is clear that highly complex situations can arise. It turns out, however, that the normal modes can still be handled easily within the MBH framework.

After the partitioning of the molecular system into blocks (nonadjoined-adjoined-free atoms) each atom A is characterized by the share number s_A (i.e., the number of blocks to which A belongs), and the corresponding block indices $b(A)_1, \dots, b(A)_{s_A}$. Some examples are schematically displayed in Fig. 3. The positions of the atoms are still gov-

erned by the transformations (1). However, the transformation parameters $\{p_b\}$ of the blocks are no longer independent variables when adjoining atoms are present. An adjoining atom belongs to more than one block, and the fact that the position of the adjoining atom must be the same when transformed according to any of these blocks imposes constraints among the parameters of the blocks to which the adjoining atom belongs. For an adjoining atom L , which is shared by s_L blocks $b(L)_1, b(L)_2, \dots, b(L)_{s_L}$, these constraints can be neatly expressed as

$$\mathbf{g}(\mathbf{r}_L^0, \hat{p}_{b(L)_1}) = \mathbf{g}(\mathbf{r}_L^0, \hat{p}_{b(L)_2}) = \dots = \mathbf{g}(\mathbf{r}_L^0, \hat{p}_{b(L)_{s_L}}). \quad (17)$$

When all adjoining atoms are considered, the set of constraints of the form in Eq. (17) therefore impose relationships between the group parameters $\{p_b\}$, identifying the allowed motions in the presence of adjoining atoms.

These relationships may also be viewed as defining an allowed hypersurface $p_b \equiv p_b(y)$ in $\{p_b\}$ space, which can be parametrized in terms of k independent variables y_π , $\pi = 1, \dots, k$. These y variables will be taken as the dynamical variables in the NMA. The number k of independent degrees of freedom is as yet unknown and depends on the number of adjoining atoms and the topological structure of the links between the various blocks. Constructing a complete functional dependence $p_b(y)$ can be hideously complicated. For the purpose of a vibrational analysis, however, we only need small excursions from the reference geometry. As we shall see, this simplifies things considerably.

B. Construction of the Hessian

The PES expressed in terms of the y coordinates becomes

$$W(y) = \tilde{V}(\{p_b(y)\}) = V(\{\mathbf{g}(\mathbf{r}_A^0, \hat{p}_{b(A)_i}(y))\}). \quad (18)$$

Because of the constraints in Eq. (17), it is immaterial which block $b(A)_i$ is chosen to describe the motion of an adjoining atom in the Cartesian PES on the right of Eq. (18): for simplicity we always take the first block $b(A)_1$.

We assume that the reference geometry has been optimized in the presence of the adjoining atom constraints, i.e. it is gradient-free for all the independent variables y_π .

$$0 = G'_\pi = \left(\frac{\partial W}{\partial y_\pi} \right)_0 = \sum_{ba} \tilde{G}_{ba} \left(\frac{\partial p_{ba}}{\partial y_\pi} \right)_0. \quad (19)$$

The Hessian at the reference geometry is easily calculated using Eq. (18),

$$H'_{\pi,\tau} = \left(\frac{\partial^2 W}{\partial y_\pi \partial y_\tau} \right)_0 = \sum_{ba} \sum_{b'a'} \tilde{H}_{ba,b'a'} \left(\frac{\partial p_{ba}}{\partial y_\pi} \right)_0 \left(\frac{\partial p_{b'a'}}{\partial y_\tau} \right)_0 + \sum_{ba} \tilde{G}_{ba} \left(\frac{\partial^2 p_{ba}}{\partial y_\pi \partial y_\tau} \right)_0. \quad (20)$$

The gradient and Hessian elements that appear in Eqs. (19) and (20) can be expressed in terms of the Cartesian ones using Eqs. (5) (not necessarily zero in this case) and (6). It is important, however, to realize that Eqs. (5) and (6) are only valid for a strict partition of the atoms, where each atom belongs to a unique block. In the present situation such a partition is done automatically by selecting the first block for each atom. The general expressions (valid for systems with or without adjoining atoms) therefore read

$$\tilde{G}_{ba} = \sum_{A \in b, \mu} G_{A\mu} D_{A\mu}^{(a)} \delta_{b,b(A)_1}, \quad (21)$$

$$\begin{aligned} \tilde{H}_{ba,b'a'} &= \sum_{A \in b, \mu} \sum_{A' \in b', \mu'} H_{A\mu, A'\mu'} D_{A\mu}^{(a)} D_{A'\mu'}^{(a')} \\ &\quad \times \delta_{b,b(A)_1} \delta_{b',b(A')_1} \\ &\quad + \delta_{b,b'} \sum_{A \in b, \mu} G_{A\mu} C_{A\mu}^{(aa')} \delta_{b,b(A)_1}. \end{aligned} \quad (22)$$

Solution of the unknown coefficients $(\partial p_{ba}/\partial y_\pi)_0$ and $(\partial^2 p_{ba}/\partial y_\pi \partial y_\tau)_0$ in Eqs. (19) and (20) is surprisingly easy. Using the y variables, the constraint equations in Eq. (17) can be re-expressed as identities. We have, for all Cartesian components μ and all adjoining atoms L , a sequence of $s_L - 1$ identities,

$$g_\mu(\mathbf{r}_L^0, \hat{p}_{b(L)_1}(y)) = g_\mu(\mathbf{r}_L^0, \hat{p}_{b(L)_2}(y)) \dots = g_\mu(\mathbf{r}_L^0, \hat{p}_{b(L)_{s_L}}(y)). \quad (23)$$

The first and second derivatives of Eq. (23) with respect to the y variables can be worked out using the chain rule,

$$\sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial p_{b(L)_1, \alpha}}{\partial y_\pi} \right)_0 = \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial p_{b(L)_2, \alpha}}{\partial y_\pi} \right)_0 \dots = \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial p_{b(L)_{s_L}, \alpha}}{\partial y_\pi} \right)_0, \quad (24)$$

$$\begin{aligned} \sum_{\alpha\alpha'} C_{L\mu}^{(\alpha\alpha')} \left(\frac{\partial p_{b(L)_1, \alpha}}{\partial y_\pi} \right)_0 \left(\frac{\partial p_{b(L)_1, \alpha'}}{\partial y_\tau} \right)_0 + \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial^2 p_{b(L)_1, \alpha}}{\partial y_\pi \partial y_\tau} \right)_0 &= \sum_{\alpha\alpha'} C_{L\mu}^{(\alpha\alpha')} \left(\frac{\partial p_{b(L)_2, \alpha}}{\partial y_\pi} \right)_0 \left(\frac{\partial p_{b(L)_2, \alpha'}}{\partial y_\tau} \right)_0 + \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial^2 p_{b(L)_2, \alpha}}{\partial y_\pi \partial y_\tau} \right)_0 \\ \dots &= \sum_{\alpha\alpha'} C_{L\mu}^{(\alpha\alpha')} \left(\frac{\partial p_{b(L)_{s_L}, \alpha}}{\partial y_\pi} \right)_0 \left(\frac{\partial p_{b(L)_{s_L}, \alpha'}}{\partial y_\tau} \right)_0 \\ &\quad + \sum_{\alpha} D_{L\mu}^{(\alpha)} \left(\frac{\partial^2 p_{b(L)_{s_L}, \alpha}}{\partial y_\pi \partial y_\tau} \right)_0, \end{aligned} \quad (25)$$

in terms of the familiar MBH coefficients $D_{L\mu}^{(\alpha)}$ and $C_{L\mu}^{(\alpha\alpha')}$ tabulated in Table I.

For any particular $\pi(=1, \dots, k)$, Eq. (24) expresses the fact that the first-order derivatives of the transformation parameters, $(\partial p_{b\alpha} / \partial y_\pi)_0 = x_{b\alpha}^{(\pi)}$, must be a solution $x^{(\pi)}$ to the homogeneous linear system of $3\Sigma_L(s_L-1)$ equations

$$Kx^{(\pi)} = 0, \quad (26)$$

where the matrix K has entries (with L specifying the adjoining atom; $\mu=1, \dots, 3$; and $i=1, \dots, s_L-1$),

$$K_{\mu Li, b\alpha} = \begin{cases} D_{L\mu}^{(\alpha)} & \text{if } b = b(L)_i \\ -D_{L\mu}^{(\alpha)} & \text{if } b = b(L)_{i+1} \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

The linear system (26) therefore immediately determines the number of independent y variables, as the dimension of the null space of the matrix K . It is convenient to choose for the $x_{b\alpha}^{(\pi)}$ an orthonormal basis of the null space of K , so that

$$\sum_{b\alpha} x_{b\alpha}^{(\pi)} x_{b\alpha}^{(\tau)} = \delta_{\pi, \tau}. \quad (28)$$

Note that if too many constraints are carelessly imposed on the relative motions of the blocks, the whole system may be “locked” and move as a single rigid body. This would be detected by a matrix K having a null space of dimension equal to 6.

For a particular combination $\pi\tau$, Eq. (25) then fixes the corresponding second-order derivatives of the transformation parameters, $(\partial^2 p_{b\alpha} / \partial y_\pi \partial y_\tau)_0 = x_{b\alpha}^{(\pi\tau)}$, as a solution $x^{(\pi\tau)}$ of the set of inhomogeneous equations,

$$Kx^{(\pi\tau)} = y^{(\pi\tau)}, \quad (29)$$

where the vector on the right has components

$$y_{\mu Li}^{(\pi\tau)} = \sum_{\alpha\alpha'} C_{L\mu}^{(\alpha\alpha')} [x_{b(L)_{i+1}\alpha}^{(\pi)} x_{b(L)_{i+1}\alpha'}^{(\tau)} - x_{b(L)_i\alpha}^{(\pi)} x_{b(L)_i\alpha'}^{(\tau)}]. \quad (30)$$

The solution $x^{(\pi\tau)}$ of Eq. (29) is not unique; it is determined up to an arbitrary vector in the null space of K , i.e.,

$$x^{(\pi\tau)'} = x^{(\pi\tau)} + \sum_{\rho} \lambda_{\rho} x^{(\rho)} \quad (31)$$

is also a solution for arbitrary λ_{ρ} . However, this does not lead to an ambiguity when evaluating the second term in the Hessian of Eq. (20),

$$\sum_{b\alpha} \tilde{G}_{b\alpha} x_{b\alpha}^{(\pi\tau)'} = \sum_{b\alpha} \tilde{G}_{b\alpha} \left[x_{b\alpha}^{(\pi\tau)'} + \sum_{\rho} \lambda_{\rho} x_{b\alpha}^{(\rho)} \right]. \quad (32)$$

On account of Eq. (19), the combination

$$\sum_{b\alpha} \tilde{G}_{b\alpha} x_{b\alpha}^{(\rho)} = 0 \quad (33)$$

vanishes in an equilibrium point, and the Hessian in Eq. (20) is independent of the choice of λ_{ρ} .

We conclude that the Hessian expressed in the y variables is completely determined by solving the linear systems in Eqs. (26) and (29). Numerically, this can be done efficiently by performing a singular value decomposition (SVD) of the matrix K . When determining the null space of K , a

singular value is considered being zero if it is smaller than a small threshold value to account for numerical inaccuracies. In practice, this never leads to ambiguities.

Note that many different parametrizations $p_b(y)$ of the allowed relative motions of the constrained MBH blocks can be constructed; however, near the reference geometry the only freedom left is the choice of a particular basis for the null space of K . One can indeed show that the final normal mode equations are independent of this choice.

C. Derivation of the normal mode equations

With the Hessian given by Eq. (20), the further derivation of the normal modes proceeds in a standard fashion. The instantaneous velocities of the atoms in the molecule are given by the defining property

$$\dot{r}_{A\mu} = \sum_{\alpha\pi} \frac{\partial g_{\mu}}{\partial p_{\alpha}} (r_A^0, \hat{p}_{b(A)}(y)) \left(\frac{\partial p_{b(A)\alpha}}{\partial y_{\pi}}(y) \right) \dot{y}_{\pi}. \quad (34)$$

Note that for a contribution of an adjoining atom to the kinetic energy, one can pick arbitrarily a block $b(A)$ to which A belongs; any other choice is equivalent on account of Eq. (24). The kinetic energy in the y variables, expanded to second order, therefore reads

$$T = \frac{1}{2} \sum_{A\mu} m_A \left(\sum_{\alpha\pi} D_{A\mu}^{(\alpha)} x_{b(A)\alpha}^{(\pi)} \dot{y}_{\pi} \right)^2 = \frac{1}{2} \sum_{\pi\tau} M'_{\pi\tau} \dot{y}_{\pi} \dot{y}_{\tau}. \quad (35)$$

The last identity in Eq. (35) defines the mass matrix M' corresponding to the use of y as the dynamical variables; its entries are given by

$$M'_{\pi\tau} = \sum_{A\mu} m_A \sum_{\alpha\alpha'} x_{b(A)\alpha}^{(\pi)} x_{b(A)\alpha'}^{(\tau)} D_{A\mu}^{(\alpha)} D_{A\mu}^{(\alpha')}. \quad (36)$$

The final normal mode equations are represented by the generalized eigenvalue problem of dimension k ,

$$H'v = \omega^2 M'v, \quad (37)$$

with ω the vibrational frequency and v the corresponding normal mode eigenvector.

D. Zero modes

If the *same* global rotation/translation P is applied to all the constrained blocks, the block parameters $p_b(y)$ change to new values, which evidently must also be on the allowed surface and are described by new values of the y variables. The connection can be made explicit by introducing a function $G(y, P)$ through

$$\Psi_b(p_b(y), P) = p_b(G(y, P)), \quad (38)$$

where Ψ_b was defined in Eq. (9).

The overall translational/rotational invariance of the potential surface can then be expressed as

$$W(y) = W(G(y, P)), \quad (39)$$

and must give rise to six eigenvectors with zero eigenvalue of the Hessian expressed in the y variables. The explicit form is easily found by taking the second-order derivative

$\partial^2/\partial y_\pi \partial P_\alpha$ of Eq. (39) at the reference configuration and at $P=0$,

$$0 = \sum_{\pi\rho} H'_{\pi\rho} \left(\frac{\partial G_\tau}{\partial y_\pi} \right)_0 \left(\frac{\partial G_\rho}{\partial P_\alpha} \right)_0 + \sum_{\tau} G'_\tau \left(\frac{\partial^2 G_\tau}{\partial y_\pi \partial P_\alpha} \right)_0. \quad (40)$$

The second term contains the gradient and vanishes automatically at the reference configuration, whereas the behavior of the G -function for small arguments leads to further simplifications. When the global transformation P in Eq. (38) is the identity ($P=0$), nothing changes, so $G_\tau(y,0)=y_\tau$ and

$$\left(\frac{\partial G_\tau}{\partial y_\pi} \right)_0 = \delta_{\pi,\tau}. \quad (41)$$

Substitution of the above results into Eq. (40) allows to recognize the presence of six eigenvectors $v^{(\alpha)}$ corresponding to global rotation/translation,

$$0 = \sum_{\rho} H'_{\pi\rho} v_\rho^{(\alpha)}, \quad (42)$$

where the components of the eigenvector are given by $v_\rho^{(\alpha)} = \partial G_\rho / \partial P_\alpha$.

In order to determine the analytic form of the eigenvectors one can consider a global transformation P applied to the reference geometry ($y=0$). According to Eq. (38) this leads to

$$\Psi_b(p_b(0), P) = \Psi_b(0, P) = p_b(G(0, P)). \quad (43)$$

Taking the first derivative $\partial/\partial P_\alpha$ at $P=0$ of the β component, one finds

$$\left(\frac{\partial \Psi_{b\beta}}{\partial P_\alpha} \right)_0 = \sum_{\tau} \left(\frac{\partial p_{b\beta}}{\partial y_\tau} \right)_0 \left(\frac{\partial G_\tau}{\partial P_\alpha} \right)_0 = \sum_{\tau} x_{b\beta}^{(\tau)} \left(\frac{\partial G_\tau}{\partial P_\alpha} \right)_0, \quad (44)$$

$$\beta = 1, \dots, d_b.$$

Multiplying with $x_{b\beta}^{(\rho)}$, summing over $b\beta$, and using the orthonormality properties (28) lead to

$$\left(\frac{\partial G_\rho}{\partial P_\alpha} \right)_0 = \sum_{b\beta} x_{b\beta}^{(\rho)} \left(\frac{\partial \Psi_{b\beta}}{\partial P_\alpha} \right)_0, \quad (45)$$

so that the eigenvectors are easily expressed in terms of known quantities.

IV. PRACTICAL IMPLEMENTATION OF THE MBH SCHEME

In this section the general MBH procedure is rephrased in matrix notation to facilitate a numerical treatment with standard linear algebra subroutines. The starting point is the Cartesian gradient and Hessian, as provided by most molecular modeling packages. If the package allows to extract the gradient and Hessian in user-defined coordinates, step 4 in the procedure below can be omitted.

- (1) **Choice of blocks and constraints.** Identify the blocks b in the molecular system, and their respective type and number d_b of transformation parameters. Assign the atoms to the various blocks, by listing for each atom A the blocks to which it belongs and the share number s_A .
- (2) **Partial optimization.** Optimize the energy while the

internal coordinates of each block are kept fixed. This can, e.g., be realized with the Z -coordinate formalism, as explained in Ref. 16, and determines the reference geometry $\{r_A^0\}$. For a linear block ($d_b=5$) one can now decide upon an axis ($\mu=x, y, z$) of the space-fixed frame and omit rotations around this axis, following the discussion in Sec. II A.

- (3) **Determination of the actual degrees of freedom.**

With $s=\sum_A 3(s_A-1)$ and $d=\sum_b d_b$, construct the $s \times d$ matrix K according to Eq. (27). Perform a SVD, $K=VSW^T$, where W is an orthogonal $d \times d$ matrix, S is a diagonal $s \times d$ matrix containing the singular values, and V is an orthogonal $s \times s$ matrix. The number k of zero singular values equals the number of actual degrees of freedom, i.e., the number of y_π variables defined in Sec. III A. In practice, a singular value is considered zero below a small threshold value to account for numerical inaccuracy. The k columns of W corresponding to the zero singular values can be taken as the orthonormal basis vectors $x^{(\pi)}$ in Eq. (26), spanning the null space of K ; they are collected in a $d \times k$ matrix X .

- (4) **Construction of the $(d \times 1)$ gradient \tilde{G} and of the $d \times d$ Hessian \tilde{H} and mass matrix \tilde{M} .**

These quantities correspond to using the d block parameters $p_{b\alpha}$ as dynamical variables, and can be obtained in terms of the Cartesian $(3N_a \times 1)$ gradient G and the $3N_a \times 3N_a$ Hessian H and mass matrix M . According to Eq. (21) they read $\tilde{G}=U^T G$, $\tilde{M}=U^T M U$, and $\tilde{H}=U^T H U + \tilde{R}$, where the $3N_a \times d$ transformation matrix U is defined as $U_{A\mu,b\alpha} = D_{A\mu}^{(\alpha)} \delta_{b,b(A)_1}$. In nonequilibrium, \tilde{H} receives a gradient correction \tilde{R} given by $\tilde{R}_{b\alpha,b'\alpha'} = \delta_{b,b'} \sum_{A \in b, \mu} G_{A\mu} C_{A\mu}^{(\alpha\alpha')} \delta_{b,b(A)_1}$. The coefficients $C^{(\alpha\alpha')}$ and $D^{(\alpha)}$ are tabulated in Table I.

- (5) **Construction of the $k \times k$ Hessian H' and mass matrix M' .**

These quantities correspond to using the genuine independent variables y_π , $\pi=1, \dots, k$, as dynamical variables. According to Eqs. (20) and (36) they read $M'=X^T \tilde{M} X$ and $H'=X^T \tilde{H} X + R'$. In nonequilibrium, H' receives a gradient correction R' given by $R'_{\pi,\tau} = \sum_{b\alpha} \tilde{G}_{b\alpha} x_{b\alpha}^{(\pi\tau)}$. The $d \times 1$ vectors $x^{(\pi\tau)}$ are obtained as a solution to the inhomogeneous linear system in Eq. (29). Since the SVD of the matrix K has already been constructed in step 3, this requires little additional effort.

- (6) **Calculate the k eigenvalues and eigenvectors of $M'^{-1/2} H' M'^{-1/2}$.** These yield the desired frequencies and normal modes of the linked block system.

V. NUMERICAL APPLICATION

As an example of the procedure developed in this paper we here report the results of a peptide chain consisting of 20 L-alanine residues (20 side chains, 21 peptide bonds), hereafter referred to as (Ala)₂₀. This oligomer is long enough to model the secondary structure of a protein: here we choose the α -helix conformation. CHARMM was used for the geometry optimization and the Hessian calculation.¹⁷ A postpro-

TABLE II. Results for the low-lying normal modes in (Ala)₂₀. The full Hessian frequencies (units of cm⁻¹, first column) are compared to MBH results using several block choices, as indicated in Fig. 4. Frequencies calculated with the constraint or dummy atoms method [columns labeled (c) and (d), respectively] are identical within numerical precision. The last three lines contain some computational details: number of frequencies, dimension of matrices to be diagonalized, dimension of K matrix [see Eq. (27)] if constraints are used. In the case of the dummy atom method, the bracketed number is the dimension of the Cartesian Hessian in Eq. (6).

Block choice constraint/dummy	Full	[C _α] ¹		[C _α] ³		[C _α] ⁴		[C _α -N] ² + [C _α -C] ²	
		c	d	c	d	c	d		
-0.0077	-0.0004	-0.0004	-0.0051	-0.0206	-0.0162	-0.0004	-0.0082		
-0.0041	-0.0002	-0.0002	-0.0042	-0.0004	-0.0144	-0.0002	-0.0074		
-0.0004	-0.0002	-0.0001	-0.0040	-0.0001	-0.0096	-0.0002	-0.0045		
0.0002	0.0003	0.0003	-0.0029	0.0002	-0.0062	0.0003	-0.0020		
0.0003	0.0003	0.0003	0.0030	0.0003	0.0031	0.0003	0.0083		
0.0030	0.0004	0.0004	0.0050	0.0018	0.0067	0.0004	0.0101		
4.0140	8.0850	6.6441	6.6441	6.5904	6.5904	7.1478	7.1479		
7.5011	9.6303	8.8830	8.8830	8.8300	8.8300	9.7947	9.7947		
9.3331	12.2977	11.5214	11.5214	11.4967	11.4967	12.2417	12.2418		
10.8304	14.3976	13.3146	13.3146	13.2574	13.2574	15.1729	15.1729		
13.0483	17.7756	15.6395	15.6395	15.5541	15.5542	17.4958	17.4958		
15.8202	19.3651	19.6403	19.6403	19.5654	19.5654	23.9893	23.9893		
18.6776	23.2107	23.0828	23.0828	22.9777	22.9777	27.0471	27.0471		
20.3250	27.0383	24.6705	24.6705	24.5536	24.5536	30.6317	30.6316		
24.6424	29.5497	30.9905	30.9905	30.8257	30.8257	36.2858	36.2858		
27.9384	35.2907	32.7529	32.7529	32.5852	32.5853	41.2541	41.2541		
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
Frequencies	609	120	126	166	166	46	46		
Dimension	609	120	(729) 246	166	(789) 346	46	(849) 246		
Dimension K	-	-	120 × 246	-	180 × 346	-	240 × 246		

cessing program written in Python²⁹ extracts the Hessian from the CHARMM output file and calculates frequencies and modes in three ways: (1) the standard NMA with the full Hessian, (2) the MBH without adjoined blocks, and (3) the MBH with adjoined blocks. The latter calculation is performed both with the constraint method according to the above procedure or with the restraint method using dummy atoms.

Table II gives an overview of the lowest calculated frequencies and some computational parameters. In the first column one finds the full Hessian frequencies as a reference. In the other columns blocks were used as depicted in Fig. 4; a block choice is labeled by the notation [C_α]^{s₁} with s₁ the share number of the C_α carbon atom. In block choice [C_α]¹, the MBH blocks are not adjoined to each other. Dummies nor link constraints are required for the vibrational analysis, and the former version of MBH can be used.

In block choice [C_α]³ the carbon atom connects three nonlinear blocks, in [C_α]⁴ the carbon atom connects one linear (the C-H bond) and three nonlinear blocks. In block choice [C_α-N]²+ [C_α-C]², the C_α-N bond and C_α-C bond are common between subsequent blocks, which illustrates the hinge-type connection. Those block choices are calculated with the adjoined version of the MBH, either by adding link constraints (c) or by introducing dummy atoms (d). Table II shows that the results are the same within meaningful numerical accuracy, hence both methods give equivalent results. The only difference to be noticed is in the lowest six frequencies, which should be zero on account of global ro-

tational and translational invariance. The deviation from zero is larger with the dummy atom approach because the introduced strong harmonic coupling tends to make the Hessian more ill conditioned and the numerical diagonalization less stable.

The introduction of blocks stiffens the systems, causing

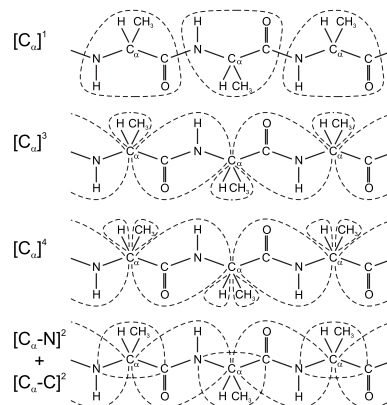


FIG. 4. Block choices for (Ala)₂₀. For clarity, only part of the chain is shown.

the frequencies to shift to higher values. As was observed by Tama *et al.*,⁹ this frequency shift is roughly linear and occurs in a rather systematic and predictable way. Comparing the nonadjoined $[C_\alpha]^1$ and the adjoined $[C_\alpha]^3$ block choice, we see that for a comparable number of frequencies (120 versus 126) the frequency overestimation is less for the block choice with adjoined blocks. Relaxation of the side chain-C-H angle as in $[C_\alpha]^4$ has little influence on the frequencies. In block choice $[C_\alpha-N]^2 + [C_\alpha-C]^2$ the large reduction in the number of frequencies (to only 8% of the original number) makes the overestimation higher.

The last lines of Table II contain some computational details for both methods. The number of frequencies depends on the number of blocks and how they are connected. When using the method with link constraints, the number of frequencies equals exactly the dimension (k) of the matrices to be diagonalized, i.e., the mass matrix and Hessian. The computational cost is dominated by the SVD of the $s \times d$ matrix K in Eq. (27), which scales as sd^2 (with $s < d$). In the restraint method, each dummy atom enlarges the dimension of the Cartesian Hessian in Eq. (6) by 3. Block choices $[C_\alpha]^3$, $[C_\alpha]^4$, and $[C_\alpha-N]^2 + [C_\alpha-C]^2$ have 40, 60, and 80 dummy atoms, respectively, and one has to manipulate considerably larger matrices. Finally a $d \times d$ matrix must be diagonalized. While the size of the present system is still rather modest, we do find that the constraint method has a more favorable overall computational cost, e.g., in block choice $[C_\alpha]^3$, the restraint method is a factor of 4 faster than in the dummy atom method.

The procedure of Sec. IV has also been tested for both fully and partially optimized structures and for a variety of topologies. Visualization of the modes showed that the blocks indeed vibrate as imposed by the link constraints. The test set (dimethylether, a series of alkanes, alanine dipeptide, crambin) is the same as those in Ref. 21, where the linked block dynamics were imposed with a restraint method using dummy atoms. While the frequencies were (virtually) identical in both methods, the present method has the advantage of being less computationally demanding. This advantage becomes more outspoken with increasing system size.

VI. CONCLUSION

In this paper the MBH model has been extended to any topology consisting of blocks (linear or nonlinear, unlinked or linked), as well as free atoms. The partitioning into blocks enables accurate frequency calculations of complex molecular systems due to the enormous reduction in the dimensionality of the problem (as the blocks have a rigid internal structure), and a further reduction can be obtained by imposing physically realistic link constraints between the blocks. Compared with other similar rotation/translation blocks (RTB/BNM) methods—frequently used in the polypeptide and protein world—the MBH routine incorporates gradient terms and thereby corrects for errors in the computed frequencies arising from partially optimized systems. Moreover, the option of introducing linkages between blocks is a new feature available in MBH but not in the other methods.

In the linked MBH scheme the blocks can move while

keeping their rigid internal structure and subject to the constraints that some of the blocks have common atoms. The geometry of this constrained system is optimized at a level of theory adapted to the size of the complex molecular system. As a result of the frozen internal structure of the blocks, as well as the presence of link constraints, the system is only partially optimized and the individual atoms are not force-free. Low frequencies correspond to collective motions and are the most interesting modes in the determination of physical and chemical properties of the system. But just these frequencies are the most sensitive to an improper handling of partially optimized systems.

The scheme proposed in this work is a general all-block formulation and may be regarded as a universal extension of the previous MBH procedure. In this way any complex molecular system can be partitioned, and in principle this should always be sufficient to make a frequency computation feasible. The new procedure makes use of a partial Hessian construction, and further matrix manipulations following known recipes. The procedure has been illustrated with the alanine-20-polypeptide.

One of the strong points of the scheme is the high flexibility, in the sense that any topology can be treated, even if it contains a mixture of free atoms, a branched chain of blocks, a ring structure of blocks, with or without linear blocks. Another important point is the absence of dummy atoms. As a consequence, the matrices to be diagonalized have considerably smaller dimensions than in the case where extra degrees of freedom are generated by the dummy atoms. Especially for large systems this is a big advantage of the constraint implementation of MBH with respect to the implementation with dummy atom restraints.

ACKNOWLEDGMENTS

A.G. is Aspirant of the Fund for Scientific Research—Flanders (FWO). This work is supported by the Fund for Scientific Research—Flanders (FWO) and by the Research Board of Ghent University (BOF).

- ¹J. P. Wang, J. X. Chen, and R. M. Hochstrasser, *J. Phys. Chem. B* **110**, 7545 (2006).
- ²R. Schweitzer, *Biophys. J.* **83**, 523 (2002).
- ³R. Schweitzer-Stenner, *J. Phys. Chem. B* **108**, 16965 (2004).
- ⁴N. A. Besley and K. A. Metcalf, *J. Chem. Phys.* **126**, 035101 (2007).
- ⁵L. Zhou and A. Siegelbaum, *Biophys. J.* **94**, 3461 (2008).
- ⁶G. J. Thomas, *Annu. Rev. Biophys. Biomol. Struct.* **28**, 1 (1999).
- ⁷M. D. Calvin, J. D. Head, and S. Q. Jin, *Surf. Sci.* **345**, 161 (1996).
- ⁸A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, *J. Chem. Phys.* **126**, 224102 (2007).
- ⁹F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, *Proteins: Struct., Funct., Genet.* **41**, 1 (2000).
- ¹⁰F. Tama and C. L. Brooks, *J. Mol. Biol.* **318**, 733 (2002).
- ¹¹J. D. Head and Y. Shi, *Int. J. Quantum Chem.* **75**, 815 (1999).
- ¹²H. Li and J. H. Jensen, *Theor. Chem. Acc.* **107**, 211 (2002).
- ¹³A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier, *J. Chem. Theory Comput.* **4**, 614 (2008).
- ¹⁴V. D. Dominguez-Soria, P. Calamini, and A. Goursot, *J. Chem. Phys.* **127**, 154710(2007).
- ¹⁵P. A. Molina and J. H. Jensen, *J. Phys. Chem. B* **107**, 6226 (2003).
- ¹⁶A. Ghysels, D. Van Neck, and M. Waroquier, *J. Chem. Phys.* **127**, 164108 (2007).
- ¹⁷B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- ¹⁸Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown,

- A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khaliullin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- ¹⁹For the latest release of ADF, visit <http://www.scm.com>.
- ²⁰C. Herrmann, J. Neugebauer, and M. Reiher, *J. Comput. Chem.* **29**, 2460 (2008).
- ²¹A. Ghysels, V. Van Speybroeck, D. Van Neck, E. Pauwels, B. R. Brooks, and M. Waroquier, "Mobile Block Hessian approach with adjointed blocks: An efficient approach for the calculation of frequencies in macromolecules," *J. Chem. Theory Comput.* (unpublished).
- ²²S. Q. Jin and J. D. Head, *Surf. Sci.* **318**, 204 (1994).
- ²³J. D. Head, *Int. J. Quantum Chem.* **65**, 827 (1997).
- ²⁴J. D. Head, *Int. J. Quantum Chem.* **77**, 350 (2000).
- ²⁵O. Marques and Y. H. Sanejouand, *Proteins: Struct., Funct., Genet.* **23**, 57 (1995).
- ²⁶G. Li and Q. Cui, *Biophys. J.* **83**, 2457 (2002).
- ²⁷G. Li and Q. Cui, *Biophys. J.* **86**, 743 (2004).
- ²⁸H. L. Woodcock, W. Zheng, A. Ghysels, Y. Shao, J. Kong, and B. R. Brooks, *J. Chem. Phys.* **129**, 214109 (2008).
- ²⁹For more information on the Python programming language, see <http://www.python.org>.

Paper VII

Mobile Block Hessian approach with adjoined blocks: An efficient approach for the calculation of frequencies in macromolecules

A. Ghysels, V. Van Speybroeck, D. Van Neck, E. Pauwels,
B. R. Brooks and M. Waroquier

Journal of Chemical Theory and Computation, in press (2009)

Copyright 2009 by American Chemical Society

JCTC

Journal of Chemical Theory and Computation

Mobile Block Hessian Approach with Adjoined Blocks: An Efficient Approach for the Calculation of Frequencies in Macromolecules

A. Ghysels,[†] V. Van Speybroeck,[†] E. Pauwels,[†] D. Van Neck,[†] B. R. Brooks,[‡] and
M. Waroquier^{*†}

Center for Molecular Modeling, Ghent University, Proeftuinstraat 86,
B-9000 Gent, Belgium, and Laboratory of Computational Biology, National Heart
Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892

Received November 11, 2008

Abstract: In an earlier work, the authors developed a new method, the mobile block Hessian (MBH) approach, to accurately calculate vibrational modes for partially optimized molecular structures [*J. Chem. Phys.* 2007, 126 (22), 224102.]. It is based on the introduction of blocks, consisting of groups of atoms, that can move as rigid bodies. The internal geometry of the blocks need not correspond to an overall optimization state of the total molecular structure. The standard MBH approach considers free blocks with six degrees of freedom. In the extended MBH approach introduced herein, the blocks can be connected by one or two adjoining atoms, which further reduces the number of degrees of freedom. The new approach paves the way for the normal-mode analysis of biomolecules such as proteins. It rests on the hypothesis that low-frequency modes of proteins can be described as pure rigid-body motions of blocks of consecutive amino acid residues. The method is validated for a series of small molecules and further applied to alanine dipeptide as a prototype to describe vibrational interactions between two peptide units; to crambin, a small protein with 46 amino acid residues; and to ICE/caspase-1, which contains 518 amino acid residues.

I. Introduction

Conformational changes of macromolecules can be probed by a variety of experimental techniques such as X-ray crystallography, NMR spectroscopy, and so on.^{1–4} The theoretical prediction of conformational flexibility is far from straightforward, especially for very large molecules such as proteins. In many cases, collective motions are present that occur on a time scale that is too long (on the order of milliseconds) to be accessible through molecular dynamics simulations.^{5,6} Normal mode analysis (NMA) has proven successful in representing domain and hinge-bending motions in proteins, and it has been shown that, for several systems, the lowest-frequency modes contribute the most to conformational changes.

Until recently, theoretical studies of dynamical properties by NMA were restricted to rather small proteins.⁷ This is partly due to the fact that a standard NMA needs to be preceded by an energy minimization. Because of the size of biological systems, this is already a computationally exhausting task. In addition, an NMA requires the diagonalization of a $3N_a \times 3N_a$ matrix (where N_a is the number of atoms). As the system becomes larger, this too becomes computationally very expensive.

A series of models has been proposed to simplify both the potential energy function⁸ and the normal mode equations. For a detailed description, we refer to a recent review by Tama.⁹ In particular, the method introduced by Durand et al.¹⁰ and later extensively tested by Tama et al.¹¹ on relatively large biosystems (up to citrate synthase with 856 residues) deserves attention. It makes use of the concept that a protein chain can be seen as a sequence of rigid components, i.e., the peptide units. The combination of rotation

* Corresponding author e-mail: Michel.Waroquier@UGent.be.

[†] Ghent University.

[‡] National Institutes of Health.

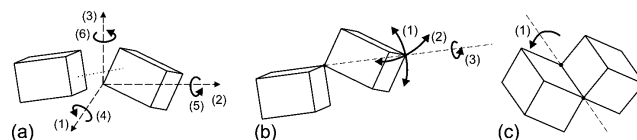


Figure 1. Degrees of freedom in the relative motion of (nonlinear) adjoining blocks. (a) Nonadjoined blocks have six degrees of freedom. The dotted line visualizes the bonding/nonbonding interaction between the blocks. (b) Blocks linked by one adjoining atom have three degrees of freedom. (c) Blocks linked by two adjoining atoms (hinge-type connection) have one degree of freedom.

and translation motions of these rigid blocks already gives a fair description of the lowest-frequency modes. For that reason, Tama et al. referred to this method as the RTB method, which stands for Rotations–Translations of Blocks.¹¹ In the approach developed by Durand et al.,¹⁰ the possibility exists to obtain, in a second step, the exact normal modes by an iterative procedure, using effective Hamiltonians that couple the RTB modes with the higher-frequency internal modes of each block. This approach, however, might be problematic for larger systems, as one deals with a highly degenerate low-frequency spectrum in this case. Tama et al. showed that a good approximation of the lowest-frequency spectrum could already be obtained by performing only the first step.

Recently, we developed the mobile block Hessian (MBH) approach, which, in a sense, very much resembles the RTB method developed by Durand et al., as each block is allowed to move with six degrees of freedom attributed to translation and rotation.^{12–14} The main difference between our original MBH method and the RTB method lies in the fact that the MBH method can also handle partially optimized systems. This kind of system also frequently occurs in other application fields, such as reactions taking place in porous materials or on surfaces. In that case, only part of the system is optimized, i.e., the chemically active part, whereas the rest is kept fixed to prevent unphysical deformations of the molecular environment. As a consequence of the partial optimization, the Hessian will have only three zero eigenvalues instead of six, implying that the rotational invariance of the potential surface is no longer manifest. Moreover, a set of spurious imaginary frequencies might appear.¹² In 2002, Li and Jensen introduced the partial Hessian vibrational analysis (PHVA) method, in which the fixed atoms are given an infinite mass during the frequency calculation and thus can no longer participate in the small-amplitude vibrations.¹⁵ This method was successfully applied by Besley and Metcalf to calculate the amide I band of polypeptides and proteins.¹⁶ More sophisticated schemes were introduced by Head and co-workers that allowed coupling between PHVA modes and modes in the fixed part.^{17–20}

In this work, we introduce an extension of our original MBH method, in which the blocks can be linked to each other by one or two common atoms (adjoining atoms), as schematically shown in Figure 1. This is important, for instance, in very large (bio)systems, where one would like to have the possibility of calculating specific normal modes at decreased computational cost, as the number of degrees of freedom is further reduced by the introduction of adjoining

atoms. If the blocks are linked by one atom, the block has three rotational degrees of freedom with respect to a previous block. If the blocks are linked by two atoms (creating a hinge-type connection), only one degree of freedom for the connected block remains. The linkage between blocks is a fundamental difference from the RTB method. Moreover, the MBH method is valid for partially optimized systems, even in the present adjoining version, extending the possible applications.

The newly introduced method is first tested on some basic test examples such as dimethylether and isobutane by assessing whether the system with reduced dimensionality can reproduce the original benchmark frequencies and associated normal modes. The hinge-type connection is tested for a series of *n*-alkanes and cyclohexane. Furthermore, the method is validated on a frequently investigated test example, alanine dipeptide, containing two peptide bonds. Finally, both connection schemes for the blocks are tested for crambin, a small protein with 46 amino acids. This system serves as a good test case because the MBH results can be compared to other methods that have also been tested on this polypeptide. To illustrate the ability of the MBH approach to treat quite large systems, the method is applied to ICE/caspase-1 with 518 amino acids.

II. Theoretical and Computational Approach

A. MBH Partitioning of the System. We distinguish among three different partitioning schemes, as illustrated in Figure 1. In the original implementation of the MBH method, atoms could only be part of one MBH block (a). In the current MBH approach, atoms can be part of two or more rigid blocks. Such a common atom, i.e., an adjoining atom, plays the role of a joint between the blocks. Two basic linkages can be realized with adjoining atoms: blocks connected by one adjoining atom (b) and blocks connected by two adjoining atoms (c). Unlinked blocks have six degrees of freedom in their relative motion, but blocks connected by an adjoining atom have only three (the three translational degrees of freedom are lost). The remaining three degrees of freedom correspond to rotations about an axis through the adjoining atom. If blocks are connected by two adjoining atoms to a previous block, the linkage is a hinge-type connection. Only one degree of freedom remains, corresponding to a rotation about the axis through the two adjoining atoms.

The hinge-type connection can correspond to some typical internal motions in molecules, as illustrated in Figure 2. The

Mobile Block Hessian Approach with Adjoined Blocks

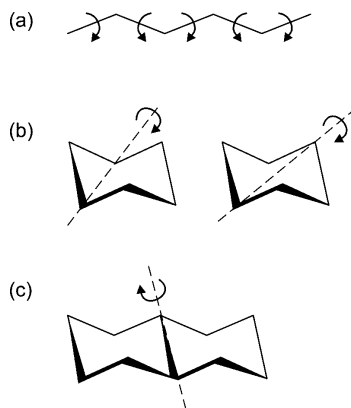


Figure 2. Three categories of motion characterized as involving hinge-type connections: (a) internal rotations, (b) ring inversion, (c) bending mode of two rings with respect to each other.

most important case (a) corresponds to internal rotations about a bond and occurs very frequently in chemistry. The hinge is determined by the atoms defining the rotational axis, i.e., by the chemical bond. Internal rotations in a molecule have frequently been studied in the literature, as they can show strong anharmonic motions.^{21–27} The second case (b) can be found in ring compounds, in which ring inversions can take place. The hinge is defined in this case by two atoms that are not directly connected by a chemical bond. The third case (c) is found in fused ring systems where two rings can bend with respect to each other. Here, the hinge corresponds to the single bond that is common to the two rings.

B. Computational Details. Preference was given to the use of one single molecular mechanics program package, even for the small test molecules. The aim of the work was not the accurate reproduction of experimental frequencies for the relevant modes, but a validation of the MBH method with respect to full Hessian (benchmark) NMA. All calculations were performed with the PARAM22/27 (proteins) and PARAM34 (alkanes, ethers) force fields of the CHARMM package.²⁸ First, the geometry was fully optimized (root-mean square of the gradient lower than 10^{-6} kcal mol⁻¹ Å⁻¹). For the dimethylether test case, two partially optimized geometries were generated as well using the SHAPE and restraints commands (RESD) from the CONS module. Next, the full Hessian was calculated in CHARMM and diagonalized with the VIBRAN module to obtain the benchmark (full Hessian) frequencies and modes.

C. Implementation. The extension of the MBH method to the case of adjoined blocks was implemented using a restraint technique and dummy atoms. This approach has the advantage that none of the MBH formulas in ref 13 need be changed. Only the Hessian must be complemented to impose the constraints. The procedure works for an arbitrary number of adjoining atoms connecting an arbitrary number of blocks,

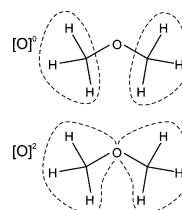


Figure 3. Dimethylether: Block choices.

so both linkages with one adjoining atom and linkages with two adjoining atoms can be treated.

Suppose an adjoining atom A belongs to s_A blocks. We can formally replace the adjoining atom by s_A dummy atom copies, A_1, \dots, A_{s_A} , at the same reference position. The mass of the dummy atoms is the original mass m_A divided by s_A . Each of the dummy atoms is assigned to one of the blocks. In this way, a strict partition (no adjoining atoms anymore) of the atoms over the blocks is restored, and the former MBH computational routines can be used.

We must still impose the fact that the dummy copies, in reality, describe the same adjoining atom A, i.e., during the motion, the coordinates $r_{A_1} = \dots = r_{A_{s_A}}$ should coincide. This can be done by adding to the potential energy harmonic terms of the form

$$V_A = \frac{1}{2} \kappa \sum_{i=2}^{s_A} |r_{A_i} - r_{A_1}|^2 \quad (1)$$

for each adjoining atom. Choosing the spring constant κ sufficiently large will ensure that, during the motion, the positions of the dummy copies (nearly) coincide; otherwise, a large energy penalty is generated. At the level of the normal modes, the introduction of the dummy copies will lead to more vibrations than in the true system. However, the addition of the V_A terms to the potential energy will result in a clear energy separation between the physical modes where the dummy atoms coincide and the unphysical modes involving stretches between dummy copies of the same atom. These unphysical modes are all at much higher energies and can be easily deleted from the frequency list.

The value of κ to be used depends on the following considerations: On one hand, κ must be in the decoupling regime, so that the frequency spectrum of the physical modes does not depend on its value. On the other hand, an exaggeratedly large value for κ can lead to numerical instability as the condition number of the Hessian matrix deteriorates. In practice, a value of about 10^7 hartree/(Bohr length)² was found to provide excellent results.

In the following discussion, the share number s_A is used to label the different block choices. For instance, in Figure 3, $[O]^2$ specifies the block choice where the oxygen atom belongs to two blocks.

III. Results and Discussion

A. Small Test Molecules for Blocks Linked by One Adjoining Atom. 1. Dimethylether. A first test case is

Table 1. Dimethylether: Frequency Calculations with the Fully Optimized Structure and Two Partially Optimized Structures^{a,b}

fully optimized					partially optimized [O] ⁰		partially optimized [O] ²	
full	blocks				full	blocks	full	blocks
	[O] ⁰		[O] ²			[O] ⁰		[O] ²
0	0		0		-28	0	0	0
0	0		0		0	0	0	0
0	0		0		0	0	0	0
0	0		0		0	0	59	0
0	0		0		12	0	156	0
0	0		0		58	0	193	0
177	177	(100)	178	(100)	199	176	250	175
244	245	(100)	245	(100)	257	248	323	232
478	479	(100)	497	(98)	482	483	496	480
898	931	(97)			897	932	975	
1023	1061	(94)			1027	1061	1111	
1096	1135	(91)			1094	1136	1118	
1149	1201	(86)			1145	1203	1363	
1149	1209	(86)			1156	1212	1405	
1204	1244	(87)			1216	1247	1430	
1437					1443		1449	
1442					1445		1471	
1442					1450		1576	
1460					1463		2099	
1583					1571		2100	
1597					1585		2176	
2850					2848		2849	
2850					2849		2914	
2912					2912		2916	
2913					2913		5423	
2915					2915		5434	
2916					2919		5644	

^a For each structure, the full Hessian (full) frequencies are compared with the MBH (blocks) frequencies, in cm^{-1} , including the frequencies due to global translation/rotation. ^b Maximum square overlap, $|\langle \nu^{\text{MBH}} | \nu^{\text{bench}} \rangle|^2$, given in parentheses (in %).

dimethylether. The full Hessian spectrum of the fully optimized structure has 21 normal modes in addition to the six global translational and rotational modes and serves as the benchmark (Table 1). The three lowest modes of dimethylether correspond to rotations of the methyl groups about an axis through the oxygen atom. Two block choices are considered as depicted in Figure 3. A natural choice is to put the methyl groups in two blocks, whereas the O atom in the center is a free atom not belonging to any block: [O]⁰. This block choice corresponds to the original MBH approach. The number of degrees of freedom can be further reduced by omitting the stretch between the oxygen atom and the methyl groups. This is realized in block choice [O]², where the oxygen atom is taken up in both blocks. Table 1 lists the MBH frequencies, with the maximum square overlap with the benchmark, $|\langle \nu^{\text{MBH}} | \nu^{\text{bench}} \rangle|^2$, given in parentheses.

The results from the full optimization are discussed first. The MBH frequencies for block choice [O]⁰ lie close to the benchmark, and the overlap values are larger than 86%. Only nine intrinsic normal modes remain, originating from the six degrees of freedom for each block, plus three translational degrees of freedom for the oxygen, minus the six global rotational and translational degrees of freedom. In the case of [O]², the oxygen atom acts as an adjoining atom between the two blocks. As a consequence, the C—O stretches no longer appear in the spectrum. Only three modes remain, which describe exactly the lowest normal modes with excellent overlaps of 98–100%. The value of the third lowest frequency has increased (497 cm^{-1}) because adjoining blocks

increase the stiffness of the system, thereby giving rise to higher frequencies.

The last four columns of Table 1 show frequencies calculated for partially optimized structures. These were obtained in the following way: First, an initial nonequilibrium geometry was generated by disturbing the positions of some of the atoms by 0.1 \AA from their global equilibrium positions. After the blocks had been defined (in this specific application, we considered [O]⁰ and [O]²), an energy minimization was performed with the internal coordinates within each block frozen. This led to a partially optimized structure. The individual atoms of the blocks might feel residual forces, but the positions and orientations of the blocks are optimized with respect to each other. For the thus-obtained partially optimized structure (which is different for each specific block choice), both a full Hessian calculation (column labeled “full”) and an MBH calculation (column labeled “blocks”) were carried out.

Partial optimization reduces the computation time for the geometry optimization; however, a standard full Hessian calculation on partially optimized structures might yield spurious low or even imaginary frequencies (indicated as negative values in Table 1). Also, unrealistically high frequencies can be detected in the vibrational spectrum (ca. 5000 cm^{-1}). Application of the MBH avoids such unphysical results. Moreover, the MBH frequencies from the partially optimized structures closely resemble those from the fully optimized structure, as can be seen in Table 1. The present MBH approach is therefore well-suited to treat partially

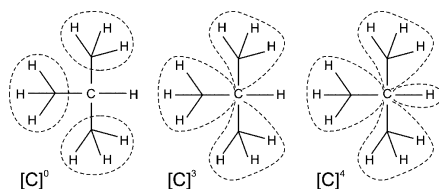


Figure 4. Isobutane: Block choices.

Table 2. Isobutane: Comparison of Benchmark (Full Hessian) and MBH Frequencies (cm^{-1}) for Different Block Choices^a

full	blocks					
	[C] ⁰		[C] ³		[C] ⁴	
248	248	(100)	248	(100)	248	(100)
263 × 2	263	(100)	263	(100)	263	(100)
361 × 2	363	(100)	373	(99)	373	(99)
428	430	(100)	456	(97)	458	(97)
795	800	(99)				
917 × 2	928	(98)				
959 × 2	977	(97)				
961	978	(97)				
1117	1133	(96)				
1139 × 2	1149	(97)				
1356 × 2	1364	(89)	1258	(51)	1258	(51)
1399 × 2						
1400						
1426						
1429 × 2						
1435 × 2						
1436						
2901 × 3						
2944	2947	(79)	2916	(77)		
2957						
2958 × 2						
2960 × 2						
2963						

^a Maximum (summed if degenerate modes) square overlap, $|\langle \nu_{\text{MBH}} | \nu_{\text{bench}} \rangle|^2$, given in parentheses (in %).

optimized structures. This simple test molecule illustrates that well-chosen adjoined blocks provide an economical way to reproduce interesting low-frequency normal modes.

2. *Isobutane*. As a second test case, we consider isobutane, where one can choose multiple adjoined blocks, all connected through the central carbon atom that serves as the adjoining atom. All investigated block choices are depicted in Figure 4. Case [C]⁰ corresponds to nonadjoined blocks, whereas cases [C]³ and [C]⁴ include linkages. Table 2 compares the MBH results with the benchmark (full Hessian) spectrum. Isobutane has 12 degenerate benchmark frequencies because of the symmetry of the molecule. When calculating the maximum square overlap, we have taken into account this degeneracy by summing the square overlaps, $\sum_i |\langle \nu_{\text{MBH}} | \nu_i^{\text{bench}} \rangle|^2$, over the modes ν_i^{bench} that belong to the same degenerate frequency ν^{bench} and picking the maximum summed square overlap.

Case [C]⁰ with three methyl groups figuring as nonadjoined blocks yields good correspondence of the frequencies and also very good overlap of the modes (79% and higher). This is in full accordance with the previous MBH results reported in ref 12.

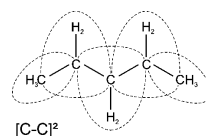


Figure 5. *n*-Alkanes: Block choice for pentane.

In the case of [C]³, where the three methyl blocks are linked to each other by including the C atom in each of the three methyl blocks, the four lowest-frequency modes are very well reproduced by 97–100%. Some frequencies have increased with respect to the benchmark values because of the stiffening of the system. The remaining frequencies in the medium- and high-frequency spectrum, i.e., 1258 and 2916 cm^{-1} , correspond to motions of the free H atom (not belonging to any of the blocks): bending with respect to the three methyl groups and stretching of the C–H bond. Especially for the bending motion, the overlap is reduced, as this motion is expected to couple with the internal motions in the methyl block (steric hindrance hydrogens).

Case [C]⁴ groups the H atom that was previously a free atom with the C atom into an additional linear adjoined block. Consequently, the C–H stretch is eliminated, whereas the rest of the frequencies and overlaps remain virtually unchanged. This illustrates the fact that the high-frequency stretch is not coupled to the lower-frequency spectrum. Such adiabatic behavior is the main justification for the use of constraints for the X–H bonds (N–H, O–H, etc.) in molecular dynamics simulations of extended (bio)systems (e.g., the SHAKE algorithm implemented in CHARMM²⁹). The C–H stretches occur at a much shorter time scale and are not essential for configurational changes and other characteristics. Similarly, the C–H stretches are not relevant for the lower spectrum of the NMA, thereby justifying block choice [C]⁴.

B. Small Test Molecules for Blocks Linked by Two Adjoining Atoms. 1. *n*-Alkanes. Internal rotations belong to the category of hinge-type connections. We illustrate this by considering subsequent internal rotations in *n*-alkanes, going from propane to decane. Figure 5 shows how each C–CH₂–C unit forms a block, such that subsequent blocks always have the C–C bond in common. The blocks at the ends of the chain consist of a C–CH₃ unit. This block choice can be denoted by [C–C]², pointing out that each C–C bond belongs to two blocks. The only degrees of freedom that are not fixed by the partitioning are the dihedral angles of the internal rotations about the C–C bonds (see also Figure 2a).

Table 3 compares the MBH frequencies to the lowest full Hessian frequencies, which serve as the benchmark. The number of frequencies is reduced substantially from $9n$ (full) to only $n - 1$ frequencies, for a C_{*n*}H_{2*n*+2} alkane. Still, the MBH frequencies resemble the benchmark fairly well, and only a slight frequency enhancement is noticed, despite the rather drastic approximation. The overlap between MBH and full modes is 99% and higher, which means the lower spectrum of the *n*-alkanes is indeed built up of dihedral angle bending motions. Several gaps are noticed in the spectrum,

Table 3. Alkanes: Benchmark (Full Hessian) Frequencies^a Compared to MBH Frequencies (cm⁻¹)^b

propane			butane			pentane			hexane		
full	[C-C] ²		full	[C-C] ²		full	[C-C] ²		full	[C-C] ²	
225	226	(100)	126	131	(100)	115	125	(100)	78	86	(99)
258	265	(99)	222	231	(99)	128	129	(100)	110	114	(99)
376			258	261	(100)	206			155		
...			293			244	249	(99)	161	167	(99)
			395			246	250	(100)	239	244	(99)
			...			356			249	254	(100)
						...			290		
									...		

heptane			octane			nonane			decane		
full	[C-C] ²		full	[C-C] ²		full	[C-C] ²		full	[C-C] ²	
61	69	(100)	46	52	(100)	38	43	(100)	30	35	(100)
89	91	(100)	79	81	(100)	68	70	(100)	62		
118			93			75			63	64	(100)
142	151	(99)	113	123	(99)	96	105	(99)	78	87	(100)
161	164	(100)	149	153	(99)	131	133	(100)	121	123	(100)
244	249	(99)	168	174	(100)	155	164	(99)	133	142	(99)
245	249	(100)	217			171	175	(100)	155		
265			243	248	(99)	187			165	169	(99)
277			246	250	(100)	210			171	177	(100)
...			246			244	244	(99)	193		
			363			245	250	(100)	244	244	(99)
			...			322			245	249	(100)
						...			272		
									...		

^a Only lowest listed. ^b Maximum square overlap, $(\langle \nu^{\text{MBH}} | \nu^{\text{bench}} \rangle)^2$, given in parentheses (in %).

e.g., the third frequency of pentane (206 cm⁻¹) is missing, but visualization of this mode reveals that it mainly involves C-C-C angle bending, which is impossible to describe with the [C-C]² block choice. Summarizing, the study of *n*-alkanes demonstrates that the MBH approach with hinges captures the low-frequency modes associated with dihedral angle bending.

2. *Cyclohexane*. The hinge-type connection can also be used to select the rotation about an axis that does not coincide with a bond, as shown in Figure 2b, for instance, to describe puckering modes. This application is tested on the chair conformation of cyclohexane, which contains in the lower vibrational spectrum six ring bending modes at 232, 375, 439, and 547 cm⁻¹. The 232 and 438 cm⁻¹ modes are doubly degenerate. These vibrational modes are schematically depicted in Figure 6a, for the hypothetical case that the molecule were planar. Modes 1-3 correspond to out-of-plane motions, whereas modes 4-6 correspond to (mainly) in-plane motions.³⁰ For the realistic nonplanar chair conformation of cyclohexane, the same terminology will be used but referring to the plane perpendicular to the 3-fold axis of the molecule. Various block choices were tested to investigate the extent to which the ring vibrations can be reproduced by the MBH approach (see Figure 6b). In the first case, [C1-C4]², the carbon ring is split into two adjoining blocks using the hinge-type connection. The blocks can rotate with respect to each other about the common C1-C4 axis. The C-H stretches are eliminated by considering each C-H pair as a linear block connected to the ring. Equivalent splittings (because of the symmetry) have also been considered, i.e., [C2-C5]² and [C3-C6]². In the second case, [C1-C3]² + [C3-C5]² + [C5-C1]², the ring is split in three blocks, subsequently

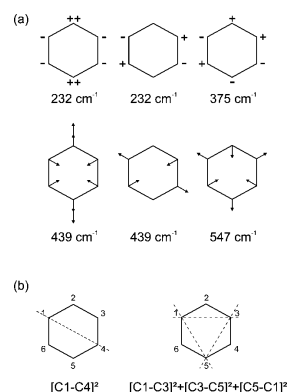


Figure 6. Planar representation of cyclohexane. (a) Six lowest-frequency bending modes. The first three are out-of-plane motions (\pm), and the next three are in-plane motions (arrows). (b) Block choices.

connected by one adjoining atom. This special case forms a ring structure of adjoining blocks.

The doubly degenerate mode at 439 cm⁻¹ can never be reproduced accurately with our block choices (maximum square overlap of no more than 16%) because these correspond to C-C-C angle bendings. From the schematic representation of the modes in Figure 6, it can be anticipated that the last block choice is most suited to describe the 375 and 547 cm⁻¹ modes. Both modes correspond to the flattening of the carbon ring. In the 375 cm⁻¹ mode,

alternating CH_2 units rotate in opposite directions perpendicular to the plane. For the 547 cm^{-1} mode, the CH_2 units move in phase, and some $\text{C}-\text{C}-\text{C}$ angle bending is involved. These modes can be best reproduced if the ring is split into three adjoined blocks (84% and 60%, respectively). One of the doubly degenerate 232 cm^{-1} modes coincides with the folding double motion and is indeed reproduced with the first block choice $[\text{C}1-\text{C}4]^2$ by 97%. The other 232 cm^{-1} mode represents a twist of the ring, which is a superposition of two folding double motions. Hence, the first block choice can only partially reproduce this mode (maximum square overlap of 74%). Splitting the ring into three blocks gives a better estimate of the mode, with a maximum square overlap of 84%.

This analysis illustrates how the nature of the complicated ring normal modes can be investigated with the MBH approach by examining the square overlaps. This is a useful feature, especially when the nature of the modes cannot be deduced uniquely by visualization. For the cyclohexane test case, it allowed us to confirm the nature of the six lowest bending modes.

C. Peptide Chains. 1. Block Choices. The next step is to investigate whether the adjoined-blocks MBH model is appropriate for simulating the low-frequency modes of peptide chains. The peptide bond is known to be a stable planar unit, and it is straightforward to introduce blocks that include the entire peptide bond. Peptides have the ability to form different secondary structure elements as a result of the conformational flexibility of the backbone. There are essentially two backbone degrees of freedom for the amino acid residue: the dihedral angles ϕ ($\text{C}-\text{N}-\text{C}_\alpha-\text{C}$) and ψ ($\text{N}-\text{C}_\alpha-\text{C}-\text{N}$). The angles describe rotations about the $\text{N}-\text{C}_\alpha$ and $\text{C}_\alpha-\text{C}$ bonds, respectively.

Figure 7 provides an overview of the considered plausible partitioning schemes for peptide chains. The smallest block choice corresponds to considering consecutive peptide bonds ($\text{C}=\text{O}-\text{NH}$) as rigid blocks, labeled by $[\text{C}_\alpha]^0$. In case $[\text{C}_\alpha]^1$, a complete residue—peptide bond plus side chain—is considered as a block. This corresponds to the RTB approach introduced by Durand et al. and extensively tested by Tama et al. on a large variety of proteins.^{10,11} From a methodological point of view, it corresponds to the MBH approach introduced in our earlier articles.^{12–14} It is also possible to combine the adjoined blocks with atoms not included in any block. This is useful for the calculation of localized modes of interest, e.g., in the chemically active part. Such block choices have not been considered in this work.

The adjoined MBH approach allows the various blocks to be connected by one or more adjoining atoms. As a first approximation, the peptide units can be linked by the C_α atoms. This further reduces, by three, the total number of degrees of freedom for each considered block. Several options are possible for treating the side chain and the H atom connected to the C_α carbon, such as $[\text{C}_\alpha]^2$, $[\text{C}_\alpha]^{3a}$, $[\text{C}_\alpha]^4$, and $[\text{C}_\alpha]^{3b}$ in Figure 7, which will be discussed later with respect to the alanine dipeptide test case. Note that the actual block choice for the ends of the chain has little influence for long peptide chains.

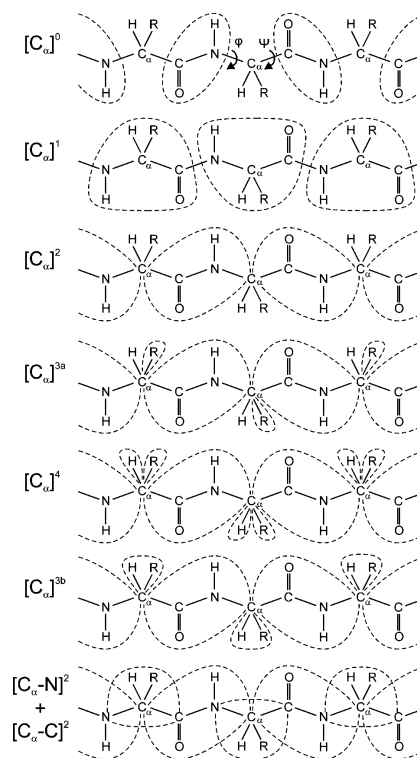


Figure 7. Partitioning schemes for peptide chains. Nonadjoined blocks: $[\text{C}_\alpha]^0$ and $[\text{C}_\alpha]^1$. Blocks linked by one adjoining atom: $[\text{C}_\alpha]^2$, $[\text{C}_\alpha]^{3a}$, $[\text{C}_\alpha]^4$, $[\text{C}_\alpha]^{3b}$. Blocks linked by two adjoining atoms: $[\text{C}_\alpha-\text{N}]^2 + [\text{C}_\alpha-\text{C}]^2$. Dihedral angles ϕ and ψ representing the two backbone degrees of freedom are also shown for the top chain.

Finally, the subsequent blocks can share two adjoining atoms as shown at the bottom of Figure 7. Each introduced block shares the $\text{C}_\alpha-\text{N}$ bond or the $\text{C}_\alpha-\text{C}$ with the block connecting the subsequent peptide units. As a result, the only degrees of freedom that are left over are the internal rotations about the common $\text{C}_\alpha-\text{N}$ and $\text{C}_\alpha-\text{C}$ bonds, which are exactly the ϕ and ψ dihedral angles. This block choice is labeled by $[\text{C}_\alpha-\text{N}]^2 + [\text{C}_\alpha-\text{C}]^2$, referring to the common $\text{C}_\alpha-\text{N}$ and $\text{C}_\alpha-\text{C}$ bonds.

By simply choosing adjoined blocks, as in $[\text{C}_\alpha-\text{N}]^2 + [\text{C}_\alpha-\text{C}]^2$, one is able to select the dihedral angles as variables. The MBH model is thus useful for NMA with variables of interest, without the need for extra computational implementations for the construction of the variables (i.e., no complicated transformation to internal coordinates is required). As for the geometry optimization, only the ϕ and ψ dihedral angles have to be force-free; other internal coordinates do not necessarily have to be optimized in the MBH

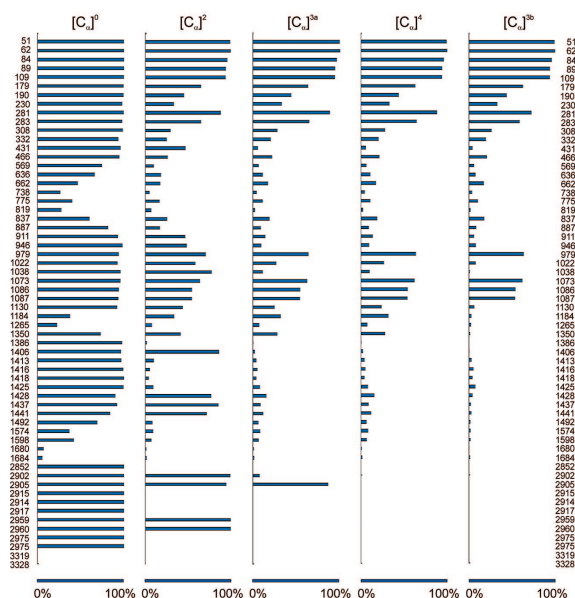


Figure 8. Alanine dipeptide: Cumulative square overlap P_j (%) for each non-zero-frequency benchmark mode.

framework. From a computational point of view, the reduction of the Hessian size is impressive. Only two degrees of freedom (ϕ , ψ) are considered per residue. A protein with N_R residues easily has over $30N_R$ degrees of freedom, but the adjoined blocks with hinges reduces this number to $2N_R$.

2. *Alanine Dipeptide.* Blocked (or capped) alanine dipeptide (N-acetyl-L-alanine-N'-methylamide) is commonly studied as a prototype of nonglycine/nonproline protein backbones, since it allows full sampling of the ϕ/ψ conformational space without the additional complexity of side-chain degrees of freedom.³¹ Numerous computational and experimental studies of alanine dipeptide have explored the thermodynamic,^{32,33} kinetic^{34–36} and spectroscopic properties.^{37,38} Alanine dipeptide takes several conformations in gas phase.^{39,40} In this study, the stable conformer with $\phi = -81^\circ$ and $\psi = 70^\circ$, usually referred to as $C7_{eq}$, was used.

To quantify how a benchmark normal mode, resulting from a full NMA analysis, corresponds to an MBH normal mode, the square overlap between the corresponding eigenvectors is again used. However, as the system size increases, many modes become degenerate or nearly degenerate. It is then more appropriate to use a cumulative square overlap P_j , defined as

$$P_j = \sum_i |(\mathbf{v}_i^{\text{MBH}} | \mathbf{v}_j^{\text{bench}})|^2 \quad (2)$$

which quantifies how well the benchmark mode $\mathbf{v}_j^{\text{bench}}$ can be reproduced by all MBH modes $\{\mathbf{v}_i^{\text{MBH}}\}$. The overlap does not necessarily become 100% because the MBH modes do not form a complete basis.

The efficiency and adequacy of the adjoined MBH method are best illustrated by examining the cumulative square overlap P_j for each non-zero-frequency benchmark mode, as can be seen in Figure 8. The block choices $[C_\alpha]^2$, $[C_\alpha]^{3a}$, $[C_\alpha]^4$, and $[C_\alpha]^{3b}$ from Figure 7 are considered to investigate how the side chain should be treated. To allow a full comparative study the results of block choice $[C_\alpha]^0$, where the peptide units are considered as nonadjoined blocks, are also shown in the first column of Figure 8. The 14 lowest-frequency modes and the high-frequency C–H stretches are reproduced by the MBH modes for at least 97%. Only some of the modes in the medium-frequency range are less well reproduced.

In the next four cases, blocks are linked to each other with the C_α atom as the adjoining atom. The ending methyl groups are also considered as blocks adjoined to the peptide unit blocks. In the case of $[C_\alpha]^2$, the side-chain atoms and the hydrogen atom are still free atoms. The frequencies are overestimated as a result of the stiffening of the system, but the cumulative square overlap indicates that the five lowest benchmark modes are still more than 90% reproduced. Case $[C_\alpha]^{3a}$ further reduces the number of frequencies by considering the side chain as an adjoined block as well (only the hydrogen atom is a free atom). The overlap in the lower spectrum is almost not influenced.

In case $[C_\alpha]^4$, the C_α –H stretch is eliminated by grouping the C_α and hydrogen atom in yet another adjoined block, such that the C_α atom links four blocks. As expected from the isobutane discussion, this C_α –H stretch is not coupled to the rest of the spectrum, which remains unchanged. Fixing

the angle between the C_α -H block and the side-chain block leads to case $[C_\alpha]^{3b}$ where the H atom is added to the side-chain block. Results largely resemble the previous $[C_\alpha]^4$ case, except for some medium frequencies around $\pm 1100\text{ cm}^{-1}$ associated with the side chain- C_α -H bending.

This example shows that grouping the atoms of the side chain into an adjoined block is a suitable choice to reproduce the lowest-frequency modes. Moreover the side-chain C_α -H bending and the C_α -H stretch are irrelevant for the lower spectrum, and the hydrogen atom can be included in the side-chain block (block choice $[C_\alpha]^{3b}$). The MBH method with linked blocks therefore appears to be an appropriate analysis tool for the lower spectrum of peptide chains.

3. *Crambin*. The MBH method with adjoined blocks is next applied to a longer, more representative peptide chain with a variety of amino acids. As an example, the well-known protein crambin (Protein Data Bank 1CCN) is chosen. The crambin molecule contains three disulfide bonds and displays β -strand, β -turn, and helical elements of protein secondary structure. The structure of crambin in water as determined by 2D NMR spectroscopy⁴¹ was taken as a starting point for the geometry optimization (in the absence of the solvent). Crambin has 46 residues, including 648 atoms, and hence, the full NMA would result in 1944 frequencies. It is our aim to reduce the number of frequencies drastically by performing an MBH analysis on the system, but still to preserve the essential lowest-frequency normal modes.

The first block choice $[C_\alpha]^0$ considers the peptide groups ($C=O-NH$) as nonadjoined blocks, which is realizable with the previous MBH implementation. The square overlaps (not shown) revealed that the peptide units indeed move as nearly rigid groups, even when some of the residues are heavier and the peptide chain is longer.

Subsequently, it is investigated whether blocks can be adjoined. Three block choices from the list in Figure 7 are proposed, of which the first already gave promising results in the previous section: (1) case $[C_\alpha]^{3b}$ groups the side chain and the H atom in an adjoined block, (2) case $[C_\alpha]^1$ is the scheme introduced by Tama et al., and (3) case $[C_\alpha-N]^2 + [C_\alpha-C]^2$ uses the hinge-type connection.

Figure 9 compares the low-frequency part of the benchmark (full Hessian) spectrum with the MBH results. A grayscale indicates the square overlap between the modes. A dot on the diagonal indicates an exact frequency estimation, and the darker the dot, the better the overlap. A huge number of frequencies (1944) lies in a limited frequency interval—between 0 and 4000 cm^{-1} —and therefore, the square overlaps seldomly reach high values because of degeneracy. The cumulative square overlap P_j does not depend on the degeneracy as it sums over all MBH frequencies. Figure 10 shows a plot of the values of P_j for the benchmark modes below 50 cm^{-1} . It indicates how well a benchmark mode is still represented by all MBH modes.

Block choice $[C_\alpha]^{3b}$ overestimates the frequencies because of the stiffening of the system, but the cumulative overlap shows that reproduction of the benchmark modes is very good in the low-frequency range: all but one of the benchmark modes below 50 cm^{-1} are reproduced for at least 90% by the MBH modes, which is largely sufficient. Block

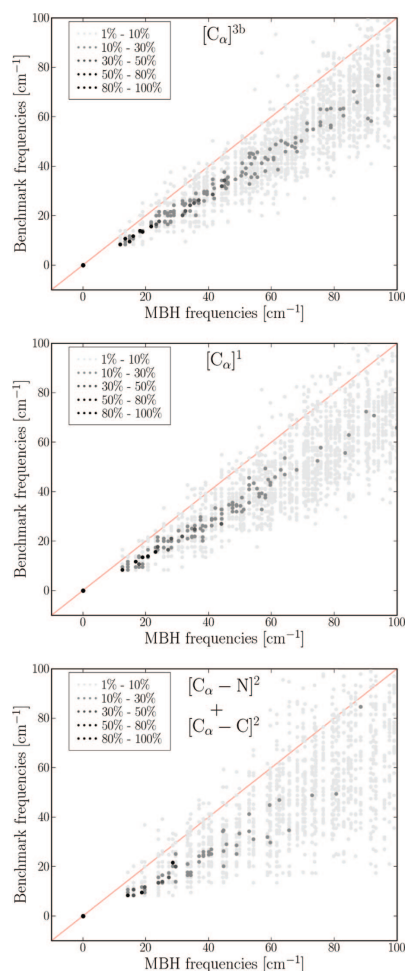


Figure 9. Crambin: Square overlap (%) between MBH modes and benchmark (full Hessian) modes in the 0 – 100 cm^{-1} frequency range for block choices $[C_\alpha]^{3b}$, $[C_\alpha]^1$, and $[C_\alpha-N]^2 + [C_\alpha-C]^2$.

choice $[C_\alpha]^{3b}$ differs from block choice $[C_\alpha]^4$ in the treatment of the H atom connected to the C_α carbon. Eliminating the C_α -H stretch without constraining the side chain C_α -H angle gives a somewhat more flexible model of the backbone, and case $[C_\alpha]^1$ is expected to be a more accurate approximation. However, the effect is minimal in the lower spectrum and becomes manifest only in the range above 250 cm^{-1} (not shown), because the H-atom motions interact very little with the backbone modes. Hence, when one is interested in the lowest frequencies, the side chain C_α -H angle bending

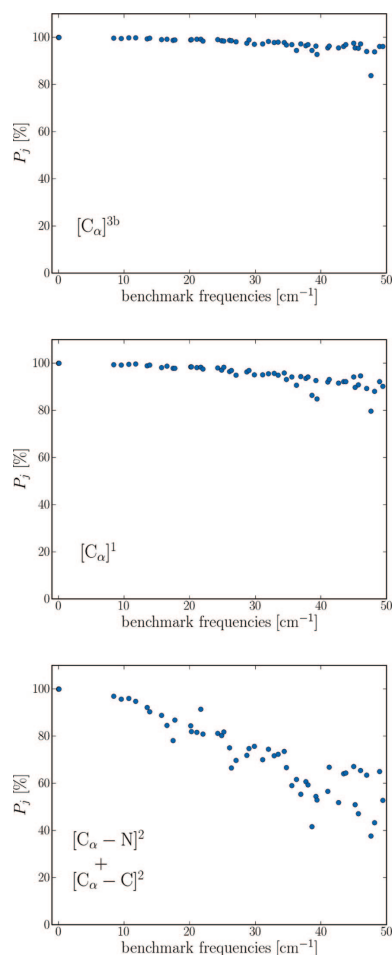


Figure 10. Crambin: Cumulative square overlap P_j (%) for benchmark modes in the 0–50 cm^{-1} range for block choices $[\text{C}_\alpha]^{3b}$, $[\text{C}_\alpha]^1$, and $[\text{C}_\alpha\text{--N}]^2 + [\text{C}_\alpha\text{--C}]^2$.

motion is not essential, and a partitioning as in case $[\text{C}_\alpha]^{3b}$, where all motions of the hydrogen atom are omitted, is sufficient.

The introduction of rigid blocks for the peptide chain implies a serious reduction of the number of degrees of freedom. In general, a peptide chain with N_{res} residues can be partitioned according to the $[\text{C}_\alpha]^{3b}$ block choice in a chain of N_{res} peptide blocks connected by the C_α atoms and N_{res} side-chain blocks linked to the chain by the C_α atoms as well. With three degrees of freedom per block, the total number becomes $6N_{\text{res}}$. For example, for crambin, the MBH

method results in 276 modes, which is 14% of the initial 1944 variables.

A similar reduction was obtained by Tama et al.¹¹ with blocks consisting of several (n) peptide-plus-side-chain units. The blocks are not linked to each other, leading to a total number of $6N_{\text{res}}/n$ degrees of freedom. For comparison, we reproduced the overlap results in the case of one single peptide-plus-side-chain unit per block ($n = 1$). This partitioning, block choice $[\text{C}_\alpha]^1$ in Figure 7, gives the same reduction of modes as the $[\text{C}_\alpha]^{3b}$ partitioning with adjoined blocks. The overlap and cumulative overlap of $[\text{C}_\alpha]^1$ also perform well, although somewhat worse than the MBH approach with adjoined blocks. Thus, for the same degree of simplification, the introduction of adjoined blocks leads to a slight improvement of the partitioning proposed by Tama et al.

Tama et al. observed a linear relationship between the benchmark and the MBH frequencies

$$\nu_{\text{MBH}} = d\nu_{\text{bench}} \quad (3)$$

The factor d was obtained by a linear least-squares fit when comparing the lowest, say, n_{low} benchmark modes with the n_{low} lowest MBH modes. For cases $[\text{C}_\alpha]^{3b}$ and $[\text{C}_\alpha]^1$, we find $d = 1.41$ and 1.49 , respectively, when the lowest modes (below 40 cm^{-1}) are taken into account. The overestimation of the frequencies, i.e., $d > 1$, is due to the stiffening of the system. Applying the block constraints limits the motions to a hypersurface on the potential energy surface, which has higher curvatures than the total potential energy surface. In other words, the constraints make the system stiffer because the motion of blocks as a whole is more hindered than it would be if the atoms in the block were allowed to relax. On the other hand, the inertia of the blocks can lower the frequencies, because a block moving as a whole has more inertia than the atoms moving individually. The frequency change is thus governed by the balance between stiffening and inertia effects. The systematic frequency overestimation indicates that, in the lower spectrum of peptide chains, the stiffening is the dominating effect for the considered block choices.

The number of modes can be further reduced by using the partitioning scheme with hinges. The dihedral basis $[\text{C}_\alpha\text{--N}]^2 + [\text{C}_\alpha\text{--C}]^2$ as illustrated in Figure 7 was applied to crambin and reduced the number of modes to 96, which is as low as 4.9% of the initial number of variables. The results are again compared with the benchmark (full Hessian) frequencies in Figures 9 and 10.

The overlap plot indicates that fewer modes are reproduced and the frequency overestimation increases: $d = 1.85$. Still, the cumulative overlap P_j shows that 8 modes are reproduced by over 90%, whereas 20 modes are reproduced by over 80%. Visualization of the modes shows that the lowest MBH modes indeed represent large-amplitude motions where large parts of the chain move as a whole. The cumulative overlap decays rapidly when the frequency increases. In practice, of course, one usually investigates only the motions in the lowest five frequency modes for the purpose of analyzing conformational changes. Hence, it is clear that the simplification of block choice $[\text{C}_\alpha\text{--N}]^2 + [\text{C}_\alpha\text{--C}]^2$ still captures the

Table 4. Caspase-1: Lowest 14 Nonzero Benchmark (Full Hessian) Frequencies Compared to the Lowest MBH Frequencies (cm^{-1}), by Means of the Square Overlap $|\langle \nu_{\text{MBH}} | \nu_{\text{bench}} \rangle|^2$ (%)^{a,b}

MBH	benchmark														
	3.4	3.6	3.9	4.8	5.4	5.4	5.7	5.7	5.8	5.8	6.0	6.5	6.6	7.3	
6.5	11		62	21											
7.3	25		61												
7.6		84													
8.8				33	46										
9.2						38					30	18			
9.4				12	11		42		11						
9.6								63							
10.2						38					34				
10.6												42			
11.0				24				13					11		
11.9									11			36			
12.1														63	
12.4										46					
12.6													11		
cum. sq. overlap	93	100	97	98	99	100	99	99	99	99	100	99	99	99	99

^a Square overlaps below 10% are not shown; square overlaps above 25% are highlighted to emphasize the diagonal. ^b Cumulative square overlap P_j (%) given for each benchmark frequency at the bottom of the table.

essentials of the lowest modes and can be used as an analysis tool that focuses on the lower spectrum.

4. *Caspase-1*. To demonstrate the ability to treat quite large systems, we will briefly discuss the case of the interleukin-1 β converting enzyme (ICE or caspase-1, Protein Data Bank 1ICE).⁴² This protein is a member of the caspase family of enzymes, which play a crucial role in apoptosis or programmed cell death. More specifically, caspase-1 proteolytically processes the interleukin-1 β precursor to its active state and, hence, contributes essentially to the immune response.⁴³

Caspase-1 is a tetramer, consisting of 518 residues and totaling 8252 atoms. A full Hessian diagonalization for the optimized structure would yield 24756 frequencies at a prohibitive expense. MBH with adjoined blocks considerably reduces the complexity to a fraction of the computational cost (less than 15%). For the calculation of the MBH frequencies, we made use of a recently developed method based on analytical constraints.⁴⁴ This method leads to the same numerical results as the dummy atom method described in section II but is much more efficient for extended molecular systems. The adjoined block choice $[C_{\alpha}]^{3b}$ was selected in view of its good performance in the crambin case. To show the potential of our MBH method, comparison was made with the 20 lowest exact full Hessian frequencies and modes, calculated with the DIAG option of CHARMM.

The six lowest frequencies correspond to the global translational and rotational modes. The next 14 nonzero modes lie in a narrow frequency range of 3.4–7.3 cm^{-1} , and their square overlaps with the lowest MBH frequencies, $|\langle \nu_{\text{MBH}} | \nu_{\text{bench}} \rangle|^2$, are given in Table 4. Square overlaps below 10% are not shown; square overlaps above 25% are highlighted. Because of the high degeneracy of the modes, the modes seem to be mixed up or to have changed order, but the highest square overlaps nevertheless lie close to the diagonal of the table. The cumulative square overlap P_j is

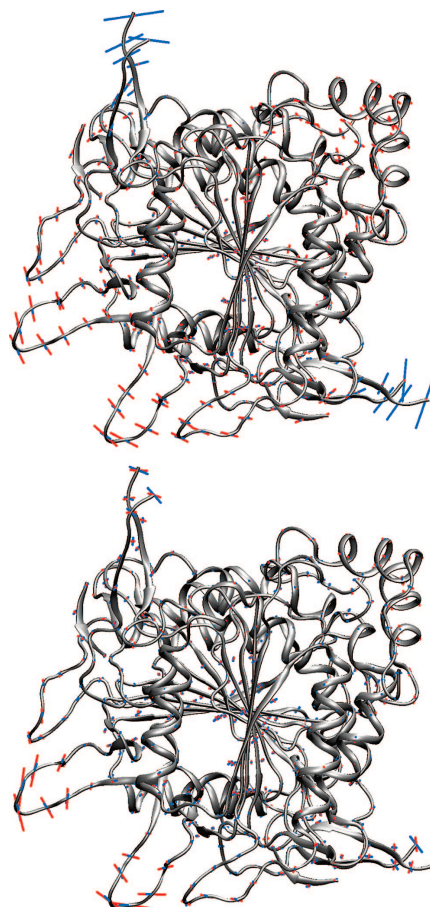


Figure 11. Caspase-1: Visualization of C_{α} movement in the MBH (red) and full (blue) normal modes. (top) First nonzero mode and (bottom) fourth nonzero mode clearly show movement of the caspase-1 subdomains with respect to each other.

given at the bottom of Table 4. Because P_j is consistently high with values of >93%, the MBH normal modes indeed represent adequate coordinates for the motions of the lower-frequency spectrum.

The first nonzero full Hessian mode is not so well described by the lowest MBH modes. As illustrated in Figure 11, this mode is mainly active in sections where the C-terminal and N-terminal regions of two polypeptide chains meet. Other modes, on the other hand, are much better reproduced and describe more coherent, global motions of the subdomains with respect to each other.

IV. Conclusions

The extension of the MBH method to the case of adjoined blocks is a computationally attractive method for analyzing modes from the lower-frequency spectrum. The computational profit offered by the MBH with adjoined blocks can be exploited at two levels. First, during the energy minimization, fewer coordinates need to be optimized. Second, during the vibrational analysis, one needs to calculate fewer second derivatives, and the much smaller size of the adapted Hessian speeds up the diagonalization.

Testing the adjoined MBH model on several small test systems revealed that frequencies can be somewhat overestimated because of the stiffening of the system, but in general, the low-frequency normal modes are very well reproduced. The block choice is obviously crucial for the quality of the MBH frequencies and modes. By varying the block sizes, the MBH can therefore be used as an analysis tool for characterizing modes. By excluding several atoms from the blocks, MBH can also analyze the coupling between those atoms (e.g., the reactive site) and the surrounding blocks (the environment). The main application of the MBH model with adjoined blocks, however, is in the calculation of the lower spectrum of macromolecules. Especially for proteins, the MBH approach provides an efficient method to focus on the modes involved in conformational changes. Several block choices were examined for their ability to reproduce the low frequencies and modes. Among the adjoined block schemes, the $[C_{\alpha}]^{2b}$ block choice, where peptide units and side chains are grouped into blocks with C_{α} as the adjoining atom, performs well in reproducing the modes of relevance while reducing the number of modes considerably. In the $[C_{\alpha}-N]^2 + [C_{\alpha}-C]^2$ block choice, where only dihedral angles are retained as degrees of freedom, the reduction of the dimensionality is very large. The resulting modes are less accurate, but still reasonable. The MBH method with adjoined blocks therefore appears to be an appropriate analysis tool for the lower spectrum of peptide chains.

Acknowledgment. A.G. is Aspirant of the Fund for Scientific Research - Flanders (FWO). This work was supported by the Fund for Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF), and the Belgian Program on Interuniversity Attraction Poles (IAP).

References

- (1) Xiao, M.; Reifenger, J. G.; Wells, A. L.; Baldacchino, C.; Chen, L. Q.; Ge, P. H.; Sweeney, H. L.; Selvin, P. R. *Nat. Struct. Biol.* **2003**, *10* (5), 402–408.
- (2) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7* (9), 740–743.
- (3) Saibil, H. R. *Nat. Struct. Biol.* **2000**, *7* (9), 711–714.
- (4) Rossmann, M. G.; Morais, M. C.; Leiman, P. G.; Zhang, W. *Structure* **2005**, *13* (3), 355–362.
- (5) Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15* (2), 151–156.
- (6) Schlick, T.; Barth, E.; Mandziuk, M. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 181–222.
- (7) Case, D. A. *Curr. Opin. Struct. Biol.* **1994**, *4* (2), 285–290.
- (8) Tirion, M. M. *Phys. Rev. Lett.* **1996**, *77* (9), 1905–1908.
- (9) Tama, F. *Protein Pept. Lett.* **2003**, *10* (2), 119–132.
- (10) Durand, P.; Trinquier, G.; Sanejouand, Y. H. *Biopolymers* **1994**, *34* (6), 759–771.
- (11) Tama, F.; Gadea, F. X.; Marques, O.; Sanejouand, Y. H. *Proteins: Struct. Funct. Genet.* **2000**, *41* (1), 1–7.
- (12) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. *J. Chem. Phys.* **2007**, *126* (22), 224102.
- (13) Ghysels, A.; Van Neck, D.; Waroquier, M. *J. Chem. Phys.* **2007**, *127* (16), 164108.
- (14) Ghysels, A.; Van Speybroeck, V.; Verstraelen, T.; Van Neck, D.; Waroquier, M. *J. Chem. Theory Comput.* **2008**, *4* (4), 614–625.
- (15) Li, H.; Jensen, J. H. *Theor. Chem. Acc.* **2002**, *107*, 211–219.
- (16) Besley, N. A.; Metcalf, K. A. *J. Chem. Phys.* **2007**, *126* (3), 035101.
- (17) Calvin, M. D.; Head, J. D.; Jin, S. Q. *Surf. Sci.* **1996**, *345* (1–2), 161–172.
- (18) Head, J. D. *Int. J. Quantum Chem.* **1997**, *65* (5), 827–838.
- (19) Head, J. D.; Shi, Y. *Int. J. Quantum Chem.* **1999**, *75* (4–5), 815–820.
- (20) Head, J. D. *Int. J. Quantum Chem.* **2000**, *77* (1), 350–357.
- (21) Heuts, J. P. A.; Gilbert, R. G.; Radom, L. *J. Phys. Chem.* **1996**, *100* (49), 18997–19006.
- (22) Van Speybroeck, V.; Van Neck, D.; Waroquier, M. *J. Phys. Chem. A* **2000**, *104* (46), 10939–10950.
- (23) Van Speybroeck, V.; Van Neck, D.; Waroquier, M. *J. Phys. Chem. A* **2002**, *106* (38), 8945–8950.
- (24) Van Speybroeck, V.; Vansteenkiste, P.; Van Neck, D.; Waroquier, M. *Chem. Phys. Lett.* **2005**, *402* (4–6), 479–484.
- (25) Lynch, V. A.; Mielke, S. L.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109* (44), 10092–10099.
- (26) Vansteenkiste, P.; Van Neck, D.; Van Speybroeck, V.; Waroquier, M. *J. Chem. Phys.* **2006**, *124* (4), 044314.
- (27) Pfaendtner, J.; Yu, X.; Broadbelt, L. J. *Theor. Chem. Acc.* **2007**, *118*, 881–898.
- (28) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4* (2), 187–217.
- (29) Vangunsteren, W. F.; Berendsen, H. J. C. *Mol. Phys.* **1977**, *34* (5), 1311–1327.
- (30) Pickett, H. M.; Strauss, H. L. *J. Am. Chem. Soc.* **1970**, *92* (25), 7281.
- (31) Feig, M. *J. Chem. Theory Comput.* **2008**, *4* (9), 1555–1564.
- (32) Smith, P. E. *J. Chem. Phys.* **1999**, *111* (12), 5568–5579.
- (33) Kwac, K.; Lee, K. K.; Han, J. B.; Oh, K. I.; Cho, M. *J. Chem. Phys.* **2008**, *128* (10), 105106.
- (34) Feig, M. *J. Chem. Theory Comput.* **2007**, *3* (5), 1734–1748.
- (35) Chekmarev, D. S.; Ishida, T.; Levy, R. M. *J. Phys. Chem. B* **2004**, *108* (50), 19487–19495.
- (36) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*

Paper VII

Mobile Block Hessian Approach with Adjoined Blocks

- (21), 6582–6594.
- (37) Kim, Y. S.; Wang, J. P.; Hochstrasser, R. M. *J. Phys. Chem. B* **2005**, *109* (15), 7511–7521.
- (38) Grdadolnik, J.; Grdadolnik, S. G.; Avbelj, F. *J. Phys. Chem. B* **2008**, *112* (9), 2712–2718.
- (39) Dadarlat, V. M. *Biophys. J.* **2005**, *89* (3), 1433–1445.
- (40) Solov'yov, I. A.; Yakubovich, A. V.; Solov'yov, A. V.; Greiner, W. *Phys. Rev. E* **2006**, *73* (2), 021916.
- (41) Bonvin, A.; Boelens, R.; Kaptein, R. *Biopolymers* **1994**, *34* (1), 39–50.
- J. Chem. Theory Comput.*, Vol. xxx, No. xx, XXXX **M**
- (42) Wilson, K. P.; Black, J. A. F.; Thomson, J. A.; Kim, E. E.; Griffith, J. P.; Navia, M. A.; Murcko, M. A.; Chambers, S. P.; Aldape, R. A.; Raybuck, S. A.; Livingston, D. J. *Nature* **1994**, *370* (6487), 270–275.
- (43) Chowdhury, I.; Tharakan, B.; Bhat, G. K. *Comp. Biochem. Physiol. B: Biochem. Mol. Biol.* **2008**, *151* (1), 10–27.
- (44) Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Brooks, B. R.; Waroquier, M. *J. Chem. Phys.* **2009**, *130*, 084107.

CT800489R

Paper VIII

Comparative study of various normal mode analysis techniques based on partial Hessians

A. Ghysels, V. Van Speybroeck, E. Pauwels, S. Catak,
B. R. Brooks, D. Van Neck, M. Waroquier

submitted to Journal of Computational Chemistry (2009)

**Comparative study of various normal mode analysis techniques
based on partial Hessians**

A. Ghysels¹, V. Van Speybroeck¹, E. Pauwels¹,
S. Catak¹, B. R. Brooks², D. Van Neck¹, M. Waroquier¹

¹ *Center for Molecular Modeling, Ghent University,
Proeftuinstraat 86, 9000 Gent, Belgium*

² *Laboratory of Computational Biology,
National Heart Lung and Blood Institute,
National Institutes of Health, Bethesda, MD 20892*

(Dated: March 26, 2009)

Abstract

Standard normal mode analysis becomes problematic for complex molecular systems, as a result of both the high computational cost and the excessive amount of information when the full Hessian matrix is used. Several partial Hessian methods have been proposed in the literature, yielding approximate normal modes. These methods aim at reducing the computational load and/or calculating only the relevant normal modes of interest in a specific application. Each method has its own (dis)advantages and application field but guidelines for the most suitable choice are lacking. We have investigated several partial Hessian methods, including the Partial Hessian Vibrational Analysis (PHVA), the Mobile Block Hessian (MBH), and the Vibrational Subsystem Analysis (VSA). In this paper we focus on the benefits and drawbacks of these methods, in terms of the reproduction of localized modes, collective modes, and the performance in partially optimized structures. We find that the PHVA is suitable for describing localized modes, the MBH not only reproduces localized and global modes but also serves as an analysis tool of the spectrum, and the VSA is mostly useful for the reproduction of the low frequency spectrum. These guidelines are illustrated with the reproduction of the localized amine-stretch, the spectrum of quinine and a bis-cinchona derivative, and the low frequency modes of the LAO binding protein.

Keywords: Hessian, partial Hessian, NMA, normal modes, vibrational modes, Mobile Block Hessian, MBH, Vibrational Subsystem Analysis, VSA, PHVA, RTB, partial optimization, LAO binding protein, cinchona

I. INTRODUCTION

Vibrational spectroscopy is an important technique for the structural characterization of (bio)molecules and materials. Infrared and Raman spectroscopic techniques not only probe the functional groups in the material, but also provide a unique “fingerprint” of the material due to the unique patterns of the absorption bands [1]. A frequently encountered problem in spectroscopy is the precise interpretation of the obtained experimental spectra. In this field theoretical predictions form an undeniable complement for the measured spectra [2]. The frequencies are obtained from normal mode analysis (NMA) resulting from the diagonalization of the full mass-weighted molecular Hessian matrix, which contains the second derivatives of the total potential energy with respect to the Cartesian nuclear coordinates. This requires the construction of the equilibrium potential energy surface (PES) by means of some energy minimization procedure. In extended molecular systems, like polypeptides and proteins, polymer chains, supramolecular assemblies, systems embedded in a solvent or (macro)molecules adsorbed within porous materials etc., this procedure poses two major problems. First, the size of the molecular systems can easily reach a few hundreds or several ten thousands of atoms, and full calculations of such large systems are computationally demanding if not impossible with accurate methods. Second, even if possible, such calculations provide a large amount of data that will be increasingly difficult to interpret.

The first problem can be addressed by using efficient computational methods such as linear-scaling algorithms, or by employing multi-level approaches, in which parts of the molecular system are described at a high quantum mechanical level of theory while the remaining part is treated with computationally less expensive methods such as molecular mechanics using classical force fields (QM/MM) [3–7]. For largely extended biomolecular systems, the description of the PES can be even further simplified by simply replacing the force field potential by a sum of Hookean pairwise potentials where the sum is restricted to atom pairs separated by less than a certain cutoff distance [8]. In this approach, the reference structure is by construction the minimum and a geometry optimization is not necessary. In a further simplification, the related Elastic Network Model (ENM) considers only pairs of the C_α backbone carbons of the protein [9, 10].

Another strategy, that gains more and more attention, is the partitioning of the system into fixed blocks, apart from some active relevant part in the molecule that remains freely

relaxed. In a first step the geometry of the complete system is optimized making use of a computationally cheap method, and in a second step, the system is partially optimized at a high level of theory, while keeping the internal geometry of the blocks fixed. The computational cost of the energy minimization is thus reduced, however, the system resides in a nonequilibrium state and application of the standard NMA procedure leads to wrong frequencies, which may be spurious and even imaginary. So far, two methods have been developed for the calculation of frequencies in partially optimized structures. In 2002, Li and Jensen introduced the Partial Hessian Vibrational Analysis (PHVA) method, in which the fixed atoms are given an infinite mass during the frequency calculation and thus can no longer participate in the small amplitude vibrations [11]. This method was successfully applied by Besley and Metcalf to calculate the amide I band of polypeptides and proteins [12]. More sophisticated schemes were introduced by Head and co-workers that allowed coupling between PHVA modes and modes in the fixed part [13–17]. We proposed recently an improved version, the Mobile Block Hessian (MBH) model [18], that takes into account the finite mass of the non-optimized atoms. The key concept is the partitioning of the system into several blocks of atoms, which move as rigid bodies during the vibrational analysis with only rotational and translational degrees of freedom. The Rotation-Translation Blocks (RTB) method developed by Durand and later Tama [19, 20] is closely related but can not yield physical frequencies in nonequilibrium points. The MBH has several variants according to the block choice and the way blocks are adjoined together [21, 22].

To solve the second problem regarding the analysis of the huge amount of data, the dimensionality of the NMA should be reduced, ideally by simultaneously decreasing the computational cost. For proteins one often chooses a model with one point mass per residue, typically located at the C_α position, which interacts through a short-range parametrized harmonic force field. The ENM belongs to this category of models. The PHVA and MBH do not alter the description of the PES but nevertheless reduce the number of normal modes, because the internal degrees of freedom of the atoms within a block are frozen. Another recently developed method is the Vibrational Subsystem Analysis (VSA) [23, 24], which partitions the system into a subsystem and an environment. Only modes initiated by subsystem motions are calculated, while the environment is assumed to follow the subsystem motions in an adiabatic way.

Summarizing, the literature offers a lot of partial Hessian methods where the dimension of

the Hessian matrix has been reduced by removing all non-relevant modes in the computation. Any reduction decreases the computational cost and/or the complexity in interpreting the data. Some of the models are also applicable to partially optimized systems. It is evident that each model has its specific advantages and disadvantages and that *a priori* no preference to a specific approach can be advised. It largely depends on the type of application. The main goal of this work is to put forward general guidelines regarding the most adequate method for a broad spectrum of applications. We distinguish three categories of methods: PHVA, MBH and VSA. The RTB concept constitutes a special case of the MBH and is not treated separately. Also a variant of VSA, that neglects all mass of the environment will be investigated. A common feature of all methods under investigation is the overall reduction of the dimensionality of the NMA, without altering the original (usually all-atom) description of the PES. The potential energy can be constructed in a pure quantum mechanical approach, or purely classically using force fields, or with hybrid models such as QM/MM, etc. The description of the PES has no impact in promoting one of the three methods in computing the Hessian matrix. In this paper we focus on the benefits and drawbacks of the three schemes in terms of the reproduction of localized modes (fingerprints for functional groups) and collective low frequency modes.

In Section II a short outline is given of the basic concepts for PHVA, MBH and VSA. In Section III some specific applications are selected to assess the performance of each method, including the reproduction of the localized amine-stretch, the analysis of the spectrum of quinine and a bis-cinchona derivative, and the low frequency modes of the LAO binding protein. In Section IV, we propose guidelines for the selection of the appropriate NMA method.

II. METHODS

The harmonic approximation is the underlying assumption in a normal mode analysis. The potential energy surface near a reference point \mathbf{r}^0 (a vector consisting of the $3N$ coordinates of the N atoms) is expanded in a Taylor series up to second order

$$V(\mathbf{r}) \approx V(0) + G^T(\mathbf{r} - \mathbf{r}^0) + \frac{1}{2}(\mathbf{r} - \mathbf{r}^0)^T H(\mathbf{r} - \mathbf{r}^0) \quad (1)$$

where G is a $3N$ -dimensional vector containing the gradient of the potential energy surface (PES), and H is the symmetric $3N \times 3N$ Hessian matrix containing the second derivatives of the PES, evaluated at the reference point

$$G_i = \left(\frac{\partial V}{\partial r_i} \right)_0 \quad (2)$$

$$H_{i,j} = \left(\frac{\partial^2 V}{\partial r_i \partial r_j} \right)_0 \quad (3)$$

The constant term $V(0)$ can be taken zero. Usually the reference point is a stationary point on the PES, i.e. $G = 0$ in \mathbf{r}^0 . If the stationary point is a minimum, the matrix H is positive semidefinite. If the stationary point is a saddle point, H has one negative eigenvalue (corresponding to the transition state frequency), or more if it is a transition state of higher order. The equations of motions of a molecule in a harmonic well are given by

$$M\Delta\ddot{\mathbf{r}} = -H\Delta\mathbf{r} \quad (4)$$

where $\Delta\mathbf{r} = \mathbf{r} - \mathbf{r}^0$ and M is the $3N \times 3N$ diagonal mass matrix. These are the equations of a system of $3N$ coupled harmonic oscillators. The solutions $\mathbf{r}(t) = \mathbf{r}^0 + v \cos(\omega t)$ are found from the generalized eigenvalue problem

$$Hv = \omega^2 Mv \quad (5)$$

where ω^2 and v are the eigenvalues and eigenvectors respectively. The frequency is given by $\nu = \omega/2\pi$. Global translations or rotations of the complete molecular system do not cause any energy increase on the PES. The invariance of the PES under these motions is manifest by the appearance of six zero frequencies, corresponding with global translations in three independent directions and global rotations about three independent axes. Linear systems (all atoms collinear) have five zero frequencies.

In nonequilibrium points (gradient $G \neq 0$) the standard NMA procedure has some serious defects. The curvature of the PES may be negative, and the Hessian H can have negative eigenvalues leading to imaginary frequencies. Although the PES is still invariant under the six global translations and rotations, the zero eigenvalues corresponding to global rotations may be lacking. Moreover, the frequencies depend on the choice of coordinates, due to the residual forces $-G_i$ on the atoms. One is frequently interested in situations where only the coordinates of a part of the total molecular system are optimized, while the rest of the atoms

are kept immobile during the energy minimization. Such partially optimized systems are in global nonequilibrium, and standard NMA cannot be applied. So far, two methods have been introduced to surmount the difficulties in case of a partially optimized system.

In the first, the partial Hessian vibrational analysis (PHVA) [11, 16], the atoms that were fixed during the energy minimization are given an infinite mass and do not move during the vibrational analysis. The normal mode equations are then restricted to the relaxed coordinates only. The NMA equations are adapted by simply taking the submatrices H_E and M_E of the Hessian and mass matrix that correspond with the non-fixed (equilibrated) atoms,

$$H_E v = \omega^2 M_E v. \quad (6)$$

The second method is the mobile block Hessian (MBH) approach [18]. Here, the fixed part is considered as a rigid body that can participate as a whole to the small amplitude vibrations. MBH is an improvement of the PHVA since it takes into account the finite mass of the mobile block. In fact, PHVA is embedded as a special case of MBH, as the limit situation where the mass of the fixed atoms becomes infinitely large. The MBH is also closely related to RTB/BNM but is more general as it incorporates corrections due to the forces on the fixed atoms in case of partial optimization. The MBH was extended to the case of multiple mobile blocks, where the relative positions and orientations of the blocks are included in the optimization. Six block parameters (three translational, three rotational) are assigned to each ‘normal’ block. The rigid-body motion of a block b can be described by block parameters $p_{b\alpha}$ ($\alpha = 1, \dots, 6$) of a translation/rotation group [25]. The instantaneous positions $\mathbf{r}_A(t)$ of the atoms A in a block are then generated by applying a common transformation with parameters p_b to the reference positions \mathbf{r}_A^0 ,

$$\mathbf{r}_A(t) = \mathbf{g}(\mathbf{r}_A^0, p_b(t)), \quad \forall A \in b. \quad (7)$$

Special blocks can consist of two or more collinear atoms, or of single atoms, and have five or three block parameters, respectively. The group variables p_b are used as the dynamical variables.

The potential energy surface (PES) expressed in the set of $\{p_{b\alpha}\}$ variables becomes

$$\tilde{V}(\{p_b\}) = V(\{\mathbf{g}(\mathbf{r}_A^0, p_{b(A)})\}), \quad (8)$$

where $V(\{\mathbf{r}_A\})$ is the PES in Cartesian coordinates, and $b(A)$ is the block to which atom A belongs.

The normal mode equations are then expressed in terms of all block parameters:

$$\tilde{H}v = \omega^2 \tilde{M}v \quad (9)$$

with

$$\tilde{H}_{b\alpha, b'\alpha'} = \left(\frac{\partial^2 V}{\partial p_{b\alpha} \partial p_{b'\alpha'}} \right)_0 \quad (10)$$

$$\tilde{M}_{b\alpha, b'\alpha'} = \sum_i m_i \left(\frac{\partial r_i}{\partial p_{b\alpha}} \right)_0 \left(\frac{\partial r_i}{\partial p_{b'\alpha'}} \right)_0. \quad (11)$$

Recently, the MBH method has been extended to the case of adjoined blocks which have one or more atoms in common (adjoining atoms) [21]. The adjoining atoms restrict the relative motion of the blocks, allowing greater flexibility to concentrate on the relevant motions.

A third method, specifically designed to reduce the number of normal modes, is the vibrational subsystem analysis (VSA). Only equilibrium structures can be treated in VSA. The system is divided into the region of interest (the subsystem s), and the molecular environment (e). This amounts to a partition of the atoms (and the atomic Cartesian coordinates) into s and e subspaces, and a corresponding block decomposition of the Hessian and mass matrix. The normal mode equations in Eq. (5) can be written as

$$\begin{pmatrix} H_{ss} & H_{se} \\ H_{es} & H_{ee} \end{pmatrix} \begin{pmatrix} v_s \\ v_e \end{pmatrix} = \omega^2 \begin{pmatrix} M_s & 0 \\ 0 & M_e \end{pmatrix} \begin{pmatrix} v_s \\ v_e \end{pmatrix}. \quad (12)$$

Projection onto the s -subspace yields

$$(H_{ss} + H_{se}(\omega^2 M_e - H_{ee})^{-1} H_{es}) v_s = \omega^2 M_s v_s. \quad (13)$$

For the low frequency modes one can expand to first order $(\omega^2 M_e - H_{ee})^{-1} \approx -H_{ee}^{-1} - \omega^2 H_{ee}^{-1} M_e H_{ee}^{-1}$ and derive the VSA normal mode equation as

$$(H_{ss} - H_{se} H_{ee}^{-1} H_{es}) v = \omega^2 (M_s + H_{se} H_{ee}^{-1} M_e H_{ee}^{-1} H_{es}) v. \quad (14)$$

This can be considered an adiabatic approximation, in which the environment follows every movement of the subsystem in order to remain gradient free during the normal mode vibration [24]. A simplified version of the VSA does not take into account the inertia of the environment and puts M_e equal to zero [23]. The mass correction in the right hand side of the above equations disappears, and the NMA equations become

$$(H_{ss} - H_{se} H_{ee}^{-1} H_{es}) v = \omega^2 M_s v. \quad (15)$$

This approximation is only acceptable if the environment consists of sufficiently light particles, such as hydrogen atoms, with respect to the atoms of the subsystem. This VSA version, hereafter referred to as VSA*, has the disadvantage that the resulting eigenvectors are not orthogonal with respect to the metric M of Cartesian space, hence the quantitative validation of the mode reproduction is ambiguous.

III. ASSESSMENT OF THE METHODS

Depending on the aim of the application, certain NMA methods are found to be more suitable than others. Also the choice of the fixed region (PHVA), the block choice (MBH) or the division in subsystem and environment (VSA, VSA*) are crucial in this respect. In general, one can be interested in the value of the frequencies or in the atomic motions involved in the normal mode vibrations themselves. One can distinguish between localized modes, global modes, and modes that involve both localized and global motions. The respective frequencies typically lie in the high, low and medium frequency range. We here further define the terms ‘local’ and ‘global’ modes.

Local modes. Modes that are concentrated in a region of space are referred to as *localized modes* or *local modes*. Only a small number of atoms participate in the vibration, while the motion of the rest of the atoms is negligible. The word ‘local’ should be interpreted in a *relative* sense: local modes involve a small number of atoms compared to the total system size. Since little mass (relatively to the total mass) is involved in the motion, a localized mode typically has a high frequency. Well-known examples are (with increasing frequency): the C-C, C-N and C-O stretches, O-H, C-H, N-H, C=C, C=N and C=O bending, C \equiv N, C-H, O-H and N-H stretches [26]. They help to recognize the functional groups of a compound. Another example is the imaginary frequency in a transition state, which is used in transition state theory to estimate the reaction rate constant [27–30]. In a more extended biosystem, a local mode can for instance refer to the motion of a complete side chain [31].

Global modes. Modes are considered to be global if a large number of atoms moves collectively, i.e. neighbouring atoms mostly move in a coherent way. Typically global modes have a low frequency because a lot of mass is involved, thereby lowering the frequency through the mass-weighting of the Hessian. Global modes are also referred to as *low frequency modes*; or *collective modes*, because domains of the molecule move with respect to

each other; or *large-amplitude modes*, because the energy cost for a low frequency is mode is small such that a small amount of energy suffices to realize a vibration with a large amplitude. Obvious global modes are the six global translations and rotations of the molecular system, which have zero frequency (no energy cost), and will not be discussed here. Low frequency modes typically involve deformation of (dihedral) angles, leading to bending and twisting modes. Examples are the accordion modes of alkane chains [32, 33] and the hinge bending and shear modes of proteins where two parts of the molecule move with respect to each other (angle bending or shear motion) [2, 34, 35].

A. Computational details

Focus lies in the reproduction of relevant local and low frequency modes by the models under study: PHVA, MBH, VSA and VSA*. In the test set that will be used to assess these NMA models, we include:

- (i) three amines with a different number of substituents (n-propylamine, dipropylamine, tripropylamine),
- (ii) two cinchona alkaloids (quinine, the bis-cinchona derivative (DHQD)₂PYDZ,
- (iii) the LAO binding protein.

The test molecules are very different in size, structure and composition and are more or less representative for the various types of molecules for which we want to propose useful guidelines suggesting the most suitable model. The amines, quinine, and bis-cinchona were simulated using a quantum mechanical description, at the MPW1B95/6-31+g(d), B3LYP/6-31+g(d,p) and B3LYP/6-31g* level of theory, respectively. The tight geometry optimization and Hessian calculation were performed with Gaussian03 [36]. The LAO binding protein was simulated using a molecular mechanics description. The open and closed conformations (2lao, respectively 1l1st, in the protein database) were optimized in absence of a solvent with the PARAM22 force field in CHARMM [37].

The masses, positions and Hessian served as input for the add-on post-processing code [38] written in the programming language Python [39]. The code includes five models for the vibrational analysis: (1) the standard NMA with the full Hessian, (2) the PHVA, (3) the MBH, (4) the VSA, (5) the VSA*. The user can define blocks or specify the subsystem atoms.

The mode reproduction is quantitatively validated with the (cumulative) square overlap. The resemblance between two mass-weighted normalized modes $|\nu_i\rangle$ and $|\nu_j\rangle$ is measured by the square overlap $|\langle\nu_i|\nu_j\rangle|^2$ between the modes, a number that varies between 0 (no overlap) and 100% (a perfect match). The cumulative square overlap $P_j = \sum_i |\langle\nu_i|\nu_j\rangle|^2$ indicates how well each mode $|\nu_j\rangle$ is reproduced by the set of modes $\{|\nu_i\rangle\}$.

B. Reproduction of localized modes and frequencies: the amine stretch

The N-H stretch in amines is a well-known textbook example of a localized stretch, with conspicuous N-H absorption bands at 3150–3700 cm^{-1} , well above the C-H stretches [26]. The number of N-H absorption peaks depends on the degree of substitution: two for a primary amine ($\text{NH}_2\text{-R}$), one for a secondary amine (NH-R_2), and no peak for a tertiary amine (N-R_3). Here we consider n-propylamine, dipropylamine and tripropylamine respectively (R is $\text{CH}_2\text{-CH}_2\text{-CH}_3$).

The question is how well a partial Hessian technique can still reproduce the localized N-H mode. Figure 1 indicates how the fixed blocks have been chosen for PHVA and MBH and how the subsystem has been specified for use in VSA and VSA*. The amine group and the neighbouring methylene group(s) are chosen to be part of the free atoms (PHVA/MBH) or of the subsystem (VSA/VSA*). The rest of the atoms is considered as being far enough to be taken up in a block or in the environment. Table I gives an overview of the frequencies above the highest C-H stretch, where the full Hessian frequencies and modes serve as the benchmark values. For each approximate method, the table includes the mode with the maximum square overlap $|\langle\nu^{full}|\nu^{approx}\rangle|^2$ with the benchmark modes.

For all methods, no frequencies are found in the upper part of the spectrum of tripropylamine. For n-propylamine and dipropylamine, it is clear that PHVA and MBH reproduce the N-H stretches (frequencies as well as modes) exactly. This result shows that for high frequency stretches, the introduction of blocks in large parts of the molecule has negligible influence on the localized mode/frequency, provided the blocks are rationally chosen. The VSA however underestimates the N-H frequencies with a slight decrease in overlap. Indeed, a localized mode at a high frequency occurs at a short time scale. The environment is forced to follow the subsystem’s movements instantaneously, even if it is barely coupled by a small stretching or angle bending force. Since this additional motion of the environment increases

the mass involved in the mode, the motion has more inertia and shifts to a lower frequency. Moreover the eigenvector of the normal mode is affected. The VSA* omits the inertia of the environment, thus avoiding the frequency shift of the mode, and the frequency estimates indeed come very close to the benchmark frequencies. The square overlap between the VSA* and benchmark modes, in italics in Table I, is only indicative. Since VSA* modes are non-orthogonal in Cartesian space, they do not represent decoupled harmonic oscillators. As a consequence of this non-decoupled behaviour of the modes, a VSA* mode can even contain a global translational and rotational component.

C. Analysis of spectra with partial Hessian methods: cinchona alkaloids

1. Quinine and bis-cinchona

To proof the concept of partial Hessian methods to analyze complex frequency spectra, we have chosen to study two cinchona alkaloids. Cinchona alkaloids are an important class of natural chiral compounds, well-known for their therapeutic efficacy, especially against malaria, and their widespread use in chiral molecular recognition. Cinchona alkaloids and their derivatives have proven to be one of the most useful catalysts to date, owing to their stability, the variety of reactions they catalyze and the availability of both enantiomeric antipodes at a reasonable price. They have been successfully employed as chiral resolving agents [40], chiral auxiliaries in heterogeneous catalysis [41] and chiral catalysts for asymmetric induction in homogeneous transformations [42]. The family of cinchona alkaloids consists of two pseudoenantiomeric pairs, cinchonine-cinchonidine and quinine-quinidine (Fig. 2). They are composed of two relatively rigid entities, an aromatic quinoline ring and an aliphatic quinuclidine ring, coupled together by two carbon-carbon single bonds. The *Open(3)* conformation, which is known to be populated by 60-70% in apolar solvents, has been adopted in the current study.

Bis-cinchona catalysts consist of two alkaloids linked by a symmetric aromatic bridge – a para substituted heterocycle. One of the most notable is the metal-cinchona complex employed in the Sharpless enantioselective dihydroxylation of terminal olefins by OsO₄ [43], where a pyridazine linked bis-dihydroquinidine (DHQD)₂PYDZ derivative (Fig. 2) is used as an enzyme-like binding pocket [44].

The systems under study – quinine and bis-cinchona – have 48 and 106 atoms, resulting in 144 and 318 full Hessian modes, respectively. This is a low number of modes in terms of Hessian storage and diagonalization. However, the calculation of the QM Hessian matrix elements themselves is already a computationally intensive task. The use of partial Hessian methods should enable the reproduction of the specific frequencies of interest only. Moreover, the interpretation of the complex spectra is far from trivial. Partial Hessian techniques used as an analysis tool can give insight into the spectrum, as some modes are too complex to be identified by simple visualization.

For quinine, five local stretching modes are investigated, which are relevant to its reactivity or are characteristic frequencies of the functional groups:

- 1) the O-H stretch of the alcohol group,
- 2) the C=C stretch of the vinyl group,
- 3) both C-O stretches in the quinoline methoxygroup (denoted C-O-Me),
- 4) the C-O stretch of the alcohol group (denoted C-OH),
- 5) the three C-N stretches of the quinuclidine.

The stretches are highlighted in red in Figure 2. For bis-cinchona, we used partial Hessian methods to identify the following characteristic frequencies (also highlighted in Fig. 2):

- 1) the C-N stretches in the two quinuclidine units,
- 2) the C-O stretches in the linker region (denoted C(sp²)-O and C(sp³)-O),
- 3) the localized modes in the pyridazine in the linker region, the N-N stretch in particular.

2. *Reproduction of modes and frequencies*

The mode and frequency reproduction was tested with quinine. The blocks in PHVA and MBH are marked by a dashed line in Figure 3. All remaining X-H stretches are also kept fixed in the MBH schemes by considering them as linear blocks. Due to the concept of infinite mass for each block, it is inherent to PHVA to merge all blocks to one combined rigid immobile block. For the corresponding VSA schemes, which have a comparable number of frequencies, all atoms in the grey shaded region belong to the subsystem (Fig. 3).

For the reproduction of the O-H, C=C and C-O-Me stretches, schemes PHVA1/MBH1/VSA1/VSA*1 of Figure 3 are proper choices for blocks or subsystem. The frequencies and modes obtained with the full Hessian calculation serve as a benchmark,

as given in the first column of Table II. The approximate frequency with highest square overlap $|\langle \nu^{full} | \nu^{approx} \rangle|^2$ (in %) are shown in the table for the PHVA, MBH, VSA and VSA* method. Both PHVA and MBH reproduce the O-H alcohol and C=C vinyl stretches extremely well. The VSA however underestimates them, with a somewhat smaller overlap. The frequency estimates of VSA* are an improvement with respect to VSA, but mode overlaps are not as satisfactory (in italics in Table II), which is a consequence of the non-orthogonality of the modes (see Section II). The reproduction of the C-O-Me stretches is only moderate for all four methods. As a conclusion, the C-O-Me stretch is not as localized as expected due to coupling with the adjacent aromatic rings.

For the reproduction of the C-OH stretch and the motions of the quinuclidine, among which the C-N stretches, schemes PHVA2/MBH2/VSA2/VSA*2 (Fig. 3) are more suited. To obtain a good mode reproduction, the quinuclidine should not be divided into blocks, because the motions of the bicyclic system are expected to be coupled due to its geometrical structure. The modes related to the quinuclidine indeed lie in a broad frequency range of 750–1550 cm^{-1} . Figure 4 plots the approximate frequencies (PHVA/MBH/VSA/VSA*) as a function of the benchmark frequencies. The greyscale is a measure for the square overlap $|\langle \nu^{full} | \nu^{approx} \rangle|^2$ with the benchmark modes. Note that more than one dot per frequency is possible if a benchmark mode is distributed over many approximate modes (or *visa versa*). The arrows indicate the benchmark frequencies that mainly involve C-OH and C-N stretching. PHVA and MBH perform quite similarly and are both capable of reproducing the indicated (‘arrowed’) frequencies with acceptable accuracy. On the other hand, the VSA modes and frequencies largely deviate from the benchmark values, which is visibly clear from the scattering in the VSA overlap plot. The VSA* plot shows hardly any high values for the overlap, which is not surprising given the ambiguous interpretation of the non-orthogonal VSA* modes.

In conclusion, the PHVA and MBH are appropriate methods for the calculation of localized modes. The more localized the mode, the better is the mode/frequency reproduction. As for the VSA, localized modes are not as well described, which is consistent with the discussion of the amines in the previous section. The adiabatic assumption imposed by VSA is in general unrealistic for localized modes. An occasional good overlap and frequency reproduction can only be obtained with a much more extended subsystem than in PHVA/MBH and would point out that the vibrational energy is localized in the subsystem.

3. Identification of frequencies

An important practical question is to know what type of motion a benchmark mode represents. The identification through visualization of the mode is often a time consuming and inaccurate approach. We propose a methodology using MBH in order to determine in an efficient way which mode contributes most to a particular stretch. The two atoms involved in the stretching motion are considered as a linear block, and the MBH modes are calculated with this block choice. Inspection of the cumulative square overlap $P_j = \sum_i |\langle \nu_i^{mbh} | \nu_j^{full} \rangle|^2$ gives information on the importance of the stretch for each of the benchmark modes. If P_j is high, the mode $|\nu_j\rangle$ is well reproduced even if the stretch is fixed, indicating that the mode does not contain stretching. If P_j is low, the mode is poorly reproduced by a block choice that fixes the stretch, indicating that the mode is dominated by this stretching motion. The contribution to the stretch is thus measured by the value $100\% - P_j$. When fixing more than one degree of freedom, e.g. several stretches, the sum of all contributions can turn out to be more than 100%. The strong point of MBH, compared to the alternative technique of projection of the modes on a reduced basis set of vectors, is that MBH automatically filters out the global translational/rotational contributions of the motion of the two atoms.

We applied the proposed methodology to quinine in order to find out which benchmark modes are most responsible for the stretches under investigation. Five cases were considered: 1) one O-H linear block; 2) one C-O linear block in the alcohol group; 3) three C-N linear blocks in the quinuclidine; 4) two C-O linear blocks in C-O-Me; 5) one C=C linear block. Table III reveals that the O-H stretch is extremely localized, whereas the 1069 cm^{-1} mode is indeed only partially a C-O-Me stretch.

The same methodology was applied to bis-cinchona to analyze its complex spectrum. Table IV lists the frequencies with the highest contribution ($100\% - P_j$) to the six C-N stretches of the quinuclidines. The stretches are coupled with a large number of modes and have frequencies around 1050 cm^{-1} and 800 cm^{-1} , which is similar to the C-N frequencies of quinine. In addition, the contribution to the C-N stretches by each quinuclidine individually were calculated. They show that a mode that involves C-N stretches is mostly localized on one side of the molecule. Since the two sides are not completely decoupled, the frequencies are not exactly degenerate.

Table V lists the frequencies mostly contributing ($100\% - P_j$) to the four C-O stretches in

the linker region of bis-cinchona: a pair of C(sp²)-O and a pair of C(sp³)-O stretches, directly connected to the pyridazine ring or separated by one bond, respectively. By calculating the contribution to the two C(sp²)-O and C(sp³)-O stretches separately, the frequency range of 1300-1480 cm⁻¹ could be assigned to C(sp²)-O stretching, and the frequency range 970-1115 cm⁻¹ to C(sp³)-O stretching. The latter is similar to the C-OH modes identified for quinine in Table III.

For the identification of the modes that are mainly localized in the linker region, we grouped all 8 pyridazine atoms in one block. Table VI lists the frequencies with the highest contribution (100% - P_j) to the pyridazine motions. The modes have a wide frequency range, including two C-H stretches, typical aromatic ring frequencies, and frequencies in the fingerprint region. Further identification is possible by calculating the contributions to the 8 bond stretching vibrations: 1 N-N, 3 C-C, 2 C-N and 2 C-H stretches. This indicates for instance that mode 1602 is almost purely a stretching mode, while lower frequency modes such as 754 cm⁻¹ represent bending motions. Frequency 1249 cm⁻¹ has been identified as the dominant N-N stretching mode.

D. Reproduction of large amplitude modes and frequencies: the LAO binding protein

Partial Hessian techniques are also expected to be efficient for the reproduction of the low frequency modes, which we will investigate for the LAO binding protein. The lysine-arginine-ornithine binding protein (LAO) is a substrate-specific receptor that interacts with the periplasmic transport system to transport basic amino acids [45]. The protein is composed of two domains, interconnected by two short strands, which move towards each other in a hinge-bending motion when a ligand is trapped. The internal structure of the domains remains practically unaltered. This particular motion is often encountered in ligand-binding proteins and commonly referred to as “pac-man” or “venus flytrap” [46]. Figure 5 illustrates the conformational change between the open, apo form [47] (pdb: 2lao) and the closed form of LAO in complex with lysine [45] (pdb: 1l1st). It is well established that the lowest frequency mode resulting from standard or simplified NMA models on the open form strikingly resembles the characteristic pac-man motion [34, 48, 49]. This demonstrates that the conformational change required for biological function is intrinsic to the design of the protein.

The pac-man mode therefore presents an interesting test case for partial Hessian techniques.

LAO has 238 residues and 3649 atoms (at physiological pH), giving rise to 10947 frequencies in a NMA with the full Hessian. To gauge the similarity of each normal mode motion with the conformational difference between the open and closed crystallographic structures, the square overlap can be determined between the eigenvector of the former and the normalized, mass-weighted difference vector of the latter. This vector can be determined by:

$$\Delta\mathbf{r}' = \frac{M^{1/2}\Delta\mathbf{r}}{\Delta\mathbf{r}^T M \Delta\mathbf{r}}, \quad (16)$$

where $\Delta\mathbf{r} = \mathbf{r}_{2\text{la0}} - \mathbf{r}_{1\text{st}}$. However, since several atoms are ill-resolved in the X-ray diffraction experiments, $\Delta\mathbf{r}$ is determined with structures optimized with the CHARMM force field, as discussed in the computational details.

Table VII lists the square overlap $Q_j = |\langle \nu_j | \Delta\mathbf{r}' \rangle|^2$ between each j -th mode eigenvector and the pac-man difference vector. The first nonzero frequency mode clearly dominates (66.1%) and the overlap quickly drops to negligible contributions for higher frequencies. In the following, we investigate how well the MBH (RTB) and VSA (VSA*) partial Hessian techniques succeed in reproducing the character and frequencies of the full Hessian NMA modes, and the global, first mode in particular.

1. Block choices.

An overview of the block choices and the corresponding number of degrees of freedom is given in Table VIII.

- The PHVA is excluded *a priori* from the method list since all atoms participate in the lowest frequency mode – none of the domains is really immobile.

- The MBH partitionings in Figure 6 are labeled by the notation $[A]^{s_A}$ with s_A the share number, i.e. the number of blocks to which atom A belongs [22]. We distinguish between schemes without and with adjoined blocks:

Without adjoined blocks: The $[C_\alpha]^1$ block choice in fact corresponds to the RTB method, where each block includes a peptide unit and a side chain. We will refer to it as RTB_1 . The RTB method was extensively tested by Durand et al. and later Tama et al. [19, 20, 34, 50, 51]. They also considered block choices with more than one residue per block, denoted RTB_n , with n the number of residues in each block. This considerably reduces the degrees of

freedom to $6N_{res}/n$, with N_{res} the number of residues in the protein.

With adjoined blocks: In Ref. [21] we first introduced partitioning schemes with adjoined blocks. Promising schemes with a comparable number of degrees of freedom are $[C_\alpha]^3$, where peptide bonds and side chains are blocks connected by the adjoining atom C_α , and $[C_\alpha-N]^2+[C_\alpha-C]^2$ which only selects the dihedral angles ϕ and ψ as variables.

- The merits of the VSA for the reproduction of modes and frequencies of proteins have not yet been investigated. We propose to take up one atom per n residues into the subsystem, while all other atoms belong to the environment. This reduces the number of frequencies to $3N_{res}/n$ only. The VSA partitions are labeled $VSA_n(X)$, where X is the (type of) atom that belongs to the subsystem, and n indicates that one X atom is selected every n residues. In Figure 6 the subsystem is indicated with a shaded region for the $VSA_1(C_\alpha)$ scheme. Other choices can be made for the atom type in the subsystem. For instance, the backbone nitrogen or backbone carbon C ($\neq C_\alpha$) can be selected. However, this hardly changes the results.

- The same partition schemes were considered for VSA* as for VSA.

2. Mode reproduction.

The quality of the reproduction of a benchmark mode $|\nu_j^{full}\rangle$ by the approximate modes is measured by the cumulative square overlap $P_j = \sum_i |\langle \nu_i^{approx} | \nu_j^{full} \rangle|^2$, which varies between 0 and 100%. Figure 7 shows the overlap for the benchmark frequencies below 50 cm^{-1} for a selection of block choices. The results for other partial Hessian partitionings are taken up in the Supporting Information.

The RTB_1 and $[C_\alpha]^3$ scheme, both with $6N_{res}$ degrees of freedom, exhibit an excellent reproduction of the low frequency spectrum with values of over 80% for all of the 264 benchmark modes in the considered frequency range. The $[C_\alpha]^3$ partition performs slightly better, which is consistent with the findings for the crambin protein in Ref. [21]. With larger RTB blocks (e.g. RTB_2) or with the $[C_\alpha-N]^2+[C_\alpha-C]^2$ block choice (dihedral angles), the cumulative square overlap rapidly decreases with increasing frequency, but nevertheless, the lowest modes are well reproduced.

The VSA cumulative square overlaps show a different behavior. The lowest frequency benchmark modes are perfectly reproduced, which is a manifestation of the adiabatic ap-

proximation – the VSA is developed for slow motions. For smaller subsystem sizes, as e.g. in $VSA_2(C_\alpha)$, the frequency range grows smaller in which (almost) perfect reproduction is attained. For the VSA*, the cumulative square overlap exceeds 100% which is a direct consequence of the non-orthogonality of the eigenvectors.

The performance of each method to reproduce the first benchmark mode is quantified in Table VIII by the square overlap $P_{1,1} = |\langle \nu_1^{approx} | \nu_1^{full} \rangle|^2$. Also, the square overlap $Q_1^{approx} = |\langle \nu_1^{approx} | \Delta \mathbf{r}' \rangle|^2$ with the open-closed conformational change vector is given. Those of VSA* in italics are only indicative. For all block choices, the $P_{1,1}$ and Q_1 values are well above 66.1%, attesting that both MBH and VSA are appropriate partial Hessian methods for the calculation of the conformational change of LAO.

3. Frequency reproduction.

The introduction of MBH blocks stiffens the polypeptide chain because the atoms within a block cannot relax during the vibration. Tama et al. quantified the frequency overestimation by means of a factor d which is the slope of the plot of the first k approximate frequencies plotted against the first k benchmark (full Hessian) frequencies,

$$\nu_{approx} = d \cdot \nu_{bench}. \quad (17)$$

The overestimation factor d is usually obtained by a least mean square fit in the low frequency range. The values in Table VIII were fitted with the lowest 50 frequencies, lying in the range 0–20 cm^{-1} .

The RTB results are in agreement with the findings of Tama et al. [20], who found a value $d = 1.7, 2.1, 2.4$ and 3.0 for the RTB₁, RTB₂, RTB₃ and RTB₅ block choice, respectively. For a same reduction of degrees of freedom, an MBH scheme with adjoined blocks can slightly improve the quality of the frequencies. For instance, the RTB₃ and the $[C_\alpha-N]^2 + [C_\alpha-C]^2$ scheme both reduce the number of frequencies to $2N_{res}$, but the stiffening in the latter scheme is somewhat less.

The VSA partitions lead to remarkably low d factors, ranging from 1.04 to 1.18. Whereas MBH frequencies overestimate the full Hessian frequencies, though in a predictable manner [20], the VSA accurately reproduces the lowest frequencies. The selection of which backbone atom (C_α , C, or N) is taken up in the subsystem, appears to be of little importance (results

not shown). When the number of modes is reduced to less than N_{res} , as in $VSA_4(C_\alpha)$ and $VSA_5(C_\alpha)$, the overestimation d becomes higher than 10%. The reason is that several low frequency modes are missing, and the linear fit with the lowest 50 frequencies is no longer representative.

The strong performance of the VSA can be understood in terms of the adiabatic approximation on which the VSA concept is based. The VSA forces the environment atoms to follow the motions of the subsystem in an adiabatic way, which introduces no errors when slow motions (corresponding to the low frequencies) are concerned. The VSA indeed originates from a series development for small eigenvalues of the Hessian, as revised in the Methods section, and therefore performs best in the low frequency spectrum.

The VSA* frequency estimates are unsufficiently accurate. Indeed, the zero mass approximation $M_e = 0$ for the environment is unappropriate given the fact that the environment has a non-negligible mass in the proposed $VSA_n(C_\alpha)$ partitions.

IV. GUIDELINES FOR SELECTING PARTIAL HESSIAN METHODS

The analysis of the previous section enables us to formulate guidelines for selecting an appropriate partial Hessian method, summarized in Table IX.

Partial optimization. A first issue is the full or partial geometry optimization prior to the vibrational analysis. If the internal geometry of the blocks in the partitioning is not optimized, PHVA and MBH still yield physical frequencies. The VSA and VSA* however might result in spurious imaginary frequencies if the environment atoms are not optimized.

Localized modes. Localized modes are well reproduced by PHVA and MBH, if the motion is truly localized in the free atom region. It is advisable to leave one bond between the local mode and the block border. This was called the one-bond-distance rule in Ref. [52]. In VSA, the adiabatic approximation results in a systematic underestimation of the highest frequencies. By omitting the inertia of the environment, as is the case in the VSA*, the frequency estimates come very close to the benchmark frequencies, but the corresponding eigenvectors lack a clear interpretation.

Global modes. PHVA is not suitable for global modes description because PHVA treats part of the system as being immobile in space. The introduction of rigid mobile blocks in the non-adjointed and adjointed MBH automatically selects modes in which parts of the

system move coherently. The block choice determines the quality of the frequency and mode estimates, and for a good reproduction of the lower spectrum, it is sufficient to keep several ‘key’ degrees of freedom as variables such as the dihedral angles (internal rotations). As shown by the theoretical derivation, the VSA is most accurate in the lower spectrum, since the adiabatic assumption is more reasonable for slow motions. The lowest frequency/mode estimates have almost a perfect accuracy when selecting a ‘disperse’ subsystem, where the subsystem atoms are no neighboring atoms but rather distant in space. The total neglect of the environment mass is in this case unrealistic such that VSA* yields unphysical low frequency modes.

V. CONCLUSION

By reducing the dimension of the Hessian, less degrees of freedom are available and the frequencies/modes are affected. It is therefore necessary to select carefully the partial Hessian method that focuses on the desired information. We investigated the ability to reproduce accurate frequencies and modes for three methods: the PHVA, the MBH, the VSA (VSA*, a variant of VSA where the environment is considered as being mass-less). The amines and cinchona alkaloids test cases illustrates the localized modes, and the LAO binding protein the global modes. We found that PHVA is capable of reproducing localized modes and that MBH can reproduce both localized and global modes. VSA is most suited for the frequency/mode reproduction of the lower spectrum. VSA* improves the frequency estimates of localized modes with respect to VSA, but the eigenvectors lack an unambiguous interpretation. When the computational performance is essential for the choice of method, the VSA/VSA* should be avoided. In partially optimized structures, PHVA and MBH can still yield physical frequencies. With the guidelines, most common simulation problems can be addressed with a focus on the lower spectrum or on the characteristic frequencies. Moreover, by varying the size of the blocks, MBH can be used as an analysis tool to identify modes in complex spectra.

VI. ACKNOWLEDGEMENTS

A.G. is Aspirant of the Fund for Scientific Research - Flanders (FWO). E.P. is postdoctoral Fellow of the Fund for Scientific Research - Flanders (FWO). This work is supported by the Fund for Scientific Research - Flanders (FWO), the Research Board of Ghent University (BOF), and the Belgian program on Interuniversity Attraction Poles (IAP).

-
- [1] D. Lesthaeghe, P. Vansteenkiste, T. Verstraelen, A. Ghysels, C. E. A. Kirschhock, J. A. Martens, V. Van Speybroeck, and M. Waroquier. MFI fingerprint: How pentasil-induced IR bands shift during zeolite nanogrowth. *J. Phys. Chem. C*, 112(25):9186–9191, 2008.
 - [2] Q. Cui and I. Bahar. *Normal Mode Analysis: Theory and applications to biological and chemical systems*. Mathematical and Computational Biology Series. Chapman & Hall/CRC, Taylor & Francis Group, Boca Raton, FL, 2006.
 - [3] A. Warshel and M. Levitt. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.*, 103(2):227–249, 1976.
 - [4] X. Assfeld and J. L. Rivail. Quantum chemical computations on parts of large molecules: The ab initio local self consistent field method. *Chem. Phys. Lett.*, 263(1-2):100–106, 1996.
 - [5] J. L. Gao, P. Amara, C. Alhambra, and M. J. Field. A generalized hybrid orbital (GHO) method for the treatment of boundary atoms in combined QM/MM calculations. *J. Phys. Chem. A*, 102(24):4714–4721, 1998.
 - [6] Y. K. Zhang, T. S. Lee, and W. T. Yang. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *J. Chem. Phys.*, 110(1):46–54, 1999.
 - [7] H. M. Senn and W. Thiel. QM/MM methods for biological systems. In *Atomistic Approaches in Modern Biology: from Quantum Chemistry to Molecular Simulations*, volume 268 of *Top. Curr. Chem.*, pages 173–290. Springer-Verlag Berlin, Berlin, 2007.
 - [8] M. M. Tirion. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.*, 77(9):1905–1908, 1996.
 - [9] I. Bahar, A. R. Atilgan, and B. Erman. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.*, 2(3):173–181, 1997.

-
- [10] K. Hinsén. Analysis of domain motions by approximate normal mode calculations. *Prot. Struct. Funct. Genet.*, 33(3):417–429, 1998.
- [11] H. Li and J. H. Jensen. Partial Hessian vibrational analysis: the localization of the molecular vibrational energy and entropy. *Theor. Chem. Acc.*, 107:211–219, 2002.
- [12] N. A. Besley and K. A. Metcalf. Computation of the amide I band of polypeptides and proteins using a partial Hessian approach. *J. Chem. Phys.*, 126(3):035101, 2007.
- [16] S. Q. Jin and J. D. Head. Theoretical investigation of molecular water-adsorption on the Al(111) surface. *Surf. Sci.*, 318(1-2):204–216, 1994.
- [13] M. D. Calvin, J. D. Head, and S. Q. Jin. Theoretically modelling the water bilayer on the Al(111) surface using cluster calculations. *Surf. Sci.*, 345(1-2):161–172, 1996.
- [14] J. D. Head. Computation of vibrational frequencies for adsorbates on surfaces. *Int. J. Quantum Chem.*, 65(5):827–838, 1997.
- [17] J. D. Head and Y. Shi. Characterization of Fermi resonances in adsorbate vibrational spectra using cluster calculations: Methoxy adsorption on Al(111) and Cu(111). *Int. J. Quantum Chem.*, 75(4-5):815–820, 1999.
- [15] J. D. Head. A vibrational analysis with Fermi resonances for methoxy adsorption on Cu(111) using ab initio cluster calculations. *Int. J. Quantum Chem.*, 77(1):350–357, 2000.
- [18] A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier. Vibrational modes in partially optimized molecular systems. *J. Chem. Phys.*, 126(22):224102, 2007.
- [19] P. Durand, G. Trinquier, and Y. H. Sanejouand. New approach for determining low-frequency normal-modes in macromolecules. *Biopolymers*, 34(6):759–771, 1994.
- [20] F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins: Struct. Funct. Genet.*, 41(1):1–7, 2000.
- [21] A. Ghysels, V. Van Speybroeck, E. Pauwels, D. Van Neck, B. R. Brooks, and M. Waroquier. Mobile block Hessian approach with adjoined blocks: an efficient approach for the calculation of frequencies in macromolecules. *J. Chem. Theory Comput.*, 2009. In press.
- [22] A. Ghysels, D. Van Neck, V. Van Speybroeck, B. R. Brooks, and M. Waroquier. Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach. *J. Chem. Phys.*, 130:084107, 2009.
- [23] W. J. Zheng and B. R. Brooks. Probing the local dynamics of nucleotide-binding pocket

- coupled to the global dynamics: Myosin versus kinesin. *Biophys. J.*, 89(1):167–178, 2005.
- [24] H. L. Woodcock, W. J. Zheng, A. Ghysels, Y. H. Shao, J. Kong, and B. R. Brooks. Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit. *J. Chem. Phys.*, 129(21):214109, 2008.
- [25] A. Ghysels, D. Van Neck, and M. Waroquier. Cartesian formulation of the mobile block Hessian approach to vibrational analysis in partially optimized systems. *J. Chem. Phys.*, 127(16):164108, 2007.
- [26] R. J. Fessenden and J. S. Fessenden. *Organic chemistry, 4th ed.* Brooks/Cole Publishing Company, Belmont, California, 1990.
- [27] H. Eyring. The activated complex in chemical reactions. *J. Chem. Phys.*, 3:107, 1935.
- [28] M. G. Evans and M. Polanyi. Some applications of the transition state method to the calculation of reaction velocities, especially in solution. *Trans. Faraday Soc.*, 31:875, 1935.
- [29] K. J. Laidler. *Chemical Kinetics.* Harper Collins Publishers, Inc., 1987.
- [30] D. A. Mc Quarrie and J. D. Simon. *Physical Chemistry - a molecular approach.* University Science Books, Sausalito, California, 1997.
- [31] C. R. Jacob and M. Reiher. Localizing normal modes in large molecules. *J. Chem. Phys.*, 130(8):15, 2009.
- [32] P. Vansteenkiste, V. Van Speybroeck, G. B. Marin, and M. Waroquier. Ab initio calculation of entropy and heat capacity of gas-phase n-alkanes using internal rotations. *J. Phys. Chem. A*, 107(17):3139–3145, 2003.
- [33] K. Van Cauter, V. Van Speybroeck, P. Vansteenkiste, M. F. Reyniers, and M. Waroquier. Ab initio study of free-radical polymerization: Polyethylene propagation kinetics. *ChemPhysChem*, 7(1):131–140, 2006.
- [35] O. Marques and Y. H. Sanejouand. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins: Struct. Funct. Genet.*, 23(4):557, 1995.
- [34] F. Tama and Y. H. Sanejouand. Conformational change of proteins arising from normal mode calculations. *Protein Eng.*, 14(1):1–6, 2001.
- [36] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda,

- O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople. Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [37] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4(2):187–217, 1983.
- [38] <http://molmod.ugent.be/code>.
- [39] <http://www.python.org>.
- [40] N. M. Maier, S. Scheffzick, G. M. Lombardo, M. Feliz, K. Rissanen, W. Lindner, and K. B. Lipkowitz. Elucidation of the chiral recognition mechanism of cinchona alkaloid carbamate-type receptors for 3,5-dinitrobenzoyl amino acids. *J. Am. Chem. Soc.*, 124(29):8611–8629, 2002.
- [41] T. Mallat, E. Orglmeister, and A. Baiker. Asymmetric catalysis at chiral metal surfaces. *Chem. Rev.*, 107(11):4863–4890, 2007.
- [42] S. France, D. J. Guerin, S. J. Miller, and T. Lectka. Nucleophilic chiral amines as catalysts in asymmetric synthesis. *Chem. Rev.*, 103(8):2985–3012, 2003.
- [43] H. C. Kolb, M. S. VanNieuwenhze, and K. B. Sharpless. Catalytic asymmetric dihydroxylation. *Chem. Rev.*, 94(8):2483–2547, 1994.
- [44] E. J. Corey and M. C. Noe. Kinetic investigations provide additional evidence that an enzyme-like binding pocket is crucial for high enantioselectivity in the bis-cinchona alkaloid catalyzed asymmetric dihydroxylation of olefins. *J. Am. Chem. Soc.*, 118(2):319–329, 1996.
- [45] C. H. Kang, W. C. Shin, Y. Yamagata, S. Gokcen, G. F. Ames, and S. H. Kim. Crystal structure of the lysine-binding, arginine-binding, ornithine-binding protein (LAO) from salmonella-typhimurium at 2.7-Å resolution. *J. Biol. Chem.*, 266(35):23893–23899, 1991.
- [46] J. S. Sack, M. A. Saper, and F. A. Quiocho. Periplasmic binding-protein structure and function

- refined X-ray structures of the leucine isoleucine valine-binding protein and its complex with leucine. *J. Mol. Biol.*, 206(1):171–191, 1989.
- [47] B. H. Oh, J. Pandit, C. H. Kang, K. Nikaido, S. Gokcen, G. F. L. Ames, and S. H. Kim. 3-dimensional structures of the periplasmic lysine arginine ornithine-binding protein with and without a ligand. *J. Biol. Chem.*, 268(15):11348–11355, 1993.
- [48] O. Keskin, R. L. Jernigan, and I. Bahar. Proteins with similar architecture exhibit similar large-scale dynamic behavior. *Biophys. J.*, 78(4):2093–2106, 2000.
- [49] W. J. Zheng and B. R. Brooks. Modeling protein conformational changes by iterative fitting of distance constraints using reoriented normal modes. *Biophys. J.*, 90(12):4327–4336, 2006.
- [50] F. Tama and C. L. Brooks. The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus. *J. Mol. Biol.*, 318(3):733, 2002.
- [51] F. Tama. Normal mode analysis with simplified models to investigate the global dynamics of biological systems. *Protein Pept. Lett.*, 10(2):119–132, 2003.
- [52] A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck, and M. Waroquier. Calculating reaction rates with partial Hessians: Validation of the mobile block Hessian approach. *J. Chem. Theory Comput.*, 4(4):614–625, 2008.

method	NH ₂ -R		NH-R ₂	N-R ₃
	(1)	(2)	(1)	
benchmark	3526	3617	3526	-
PHVA	3526 (100.0)	3617 (100.0)	3526 (100.0)	-
MBH	3526 (100.0)	3617 (100.0)	3526 (100.0)	-
VSA	3518 (99.5)	3609 (99.5)	3467 (95.3)	-
VSA*	3525 (<i>92.9</i>)	3617 (<i>98.8</i>)	3526 (<i>73.4</i>)	-

TABLE I: Amines – High frequency spectrum (above 3150 cm⁻¹) of three amines: n-propylamine, dipropylamine, tripropylamine (R is CH₂-CH₂-CH₃). Frequencies (in cm⁻¹) are calculated with the full Hessian (benchmark), the PHVA, MBH, VSA and VSA*. The square overlap (in %) with the benchmark modes is given between brackets, those in italics for VSA* are only indicative.

benchmark		PHVA 1	MBH 1	VSA 1	VSA* 1
3821	O-H	3821 (100)	3821 (100)	3769 (97)	3819 (78)
1702	C=C	1698 (99)	1700 (99)	1627 (85)	1690 (49)
1069	C-O-Me	1059 (53)	1066 (54)	1133 (17)	984 (13)
1295	C-O-Me	1324 (38)	1332 (29)	1236 (28)	1264 (27)
918	C-O-Me	1059 (13)	1066 (12)	833 (15)	984 (8)

TABLE II: Quinine – PHVA, MBH, VSA and VSA* used to reproduce the O-H, the C=C and the C-O-Me stretches, according to the schemes PHVA1/MBH1/VSA1/VSA*1 of Figure 3. The first column contains the benchmark (full Hessian) frequencies in cm^{-1} and the mode type. The given approximate frequencies are those that have the maximum square overlap $|\langle \nu^{full} | \nu^{approx} \rangle|^2$ with the benchmark mode (in %), those in italics for VSA* are only indicative.

stretch	benchmark	contrib.
1 O-H	3821	100
1 C=C	1702	82
2 C-O-Me	1069	49
	1295	41
	918	15
1 C-OH	1108	33
	1008	15
3 C-N	1059	47
	1079	31
	826	29
	769	21
	1108	15

TABLE III: Quinine – Which modes contribute mostly to the O-H, C=C, C-O-Me, N-C and C-OH stretches? Benchmark frequencies in cm^{-1} are listed according to the highest contribution $100\% - P_j$ (in %).

benchmark	6 C-N	3 C-N	3 C-N
1072	36	0	36
1093	28	5	23
1075	27	27	0
1094	27	22	5
840	25	7	18
1059	17	17	0
824	16	13	3
1032	15	12	3
1031	15	3	12
1049	13	1	12

TABLE IV: Bis-cinchona – Benchmark frequencies (in cm^{-1}) with the highest contribution (in %) to the six C-N stretches in the two quinuclidines. Contribution of each quinuclidine individually is also indicated.

benchmark	4 C-O	benchmark	2 C(sp ²)-O	benchmark	2 C(sp ³)-O
1325	36	1325	35	1007	26
1478	27	1300	25	1017	19
1007	26	1478	25	1115	15
1300	25	1293	11	1111	14
1017	20				
1111	15				
1115	15				
1293	11				

TABLE V: Bis-cinchona – Benchmark frequencies (in cm^{-1}) with the highest contribution (in %) to the four C-O stretches in the linker region, in the two C(sp²)-O and the two C(sp³)-O stretches.

benchmark	pydz	8 bond	1 N-N	3 C-C	2 C-N	2 C-H
3240	100	100	-	-	-	100
3220	100	100	-	-	-	100
1602	97	89	-	76	34	-
972	95	0	-	-	-	-
754	94	0	-	-	-	-
1249	87	75	44	15	47	-
1159	83	16	6	8	-	-
1478	73	48	-	1	39	-
845	72	0	-	-	-	-
1652	69	61	6	46	9	-
1325	68	50	5	11	24	-
563	49	0	-	-	-	-

TABLE VI: Bis-cinchona – Benchmark frequencies (in cm^{-1}) which are mainly localized in the pyridazine (pydz). Contributions in %, values below 1% not shown. The contribution from the 8 bond stretches and the individual contributions (1 N-N, 3 C-C, 2 N-C, 2 C-H stretches) are also given.

benchmark [cm ⁻¹]	Q_j [%]
1.5	66.1
2.8	13.2
4.5	2.9
5.3	0.1
6.1	0.0
7.0	0.1
7.3	0.3
7.7	0.1
8.1	0.8
8.8	0.1

TABLE VII: LAO – Frequencies (in cm⁻¹) of the first normal modes of the LAO binding protein, calculated with the full Hessian on the open form. For each mode, the square overlap $Q_j = |\langle \nu_j^{bench} | \Delta \mathbf{r}' \rangle|^2$ (in %) with the “pac-man” difference vector is given.

method	DOF	d	$P_{1,1}$	Q_1^{approx}
MBH: RTB				
RTB ₁ /[C _α] ¹	6 N_{res}	1.70	93.7	77.7
RTB ₂	3 N_{res}	2.08	90.2	79.3
RTB ₃	2 N_{res}	2.41	85.7	80.9
RTB ₄	1.50 N_{res}	2.65	87.5	80.3
RTB ₅	1.20 N_{res}	2.90	87.6	79.9
MBH: adjoined blocks				
[C _α] ³	6 N_{res}	1.55	95.6	76.0
[C _α -N] ² + [C _α -C] ²	2 N_{res}	2.09	94.7	72.5
VSA				
VSA(C _α)	3 N_{res}	1.04	100.0	66.1
VSA ₂ (C _α)	1.50 N_{res}	1.06	100.0	66.1
VSA ₃ (C _α)	N_{res}	1.09	100.0	66.1
VSA ₄ (C _α)	0.75 N_{res}	1.12	100.0	66.1
VSA ₅ (C _α)	0.60 N_{res}	1.18	100.0	66.1
VSA*				
VSA*(C _α)	3 N_{res}	3.34	(99.9)	(66.1)
VSA ₂ *(C _α)	1.50 N_{res}	4.55	(99.9)	(66.1)
VSA ₃ *(C _α)	N_{res}	5.50	(99.9)	(66.0)
VSA ₄ *(C _α)	0.75 N_{res}	6.00	(99.8)	(65.0)
VSA ₅ *(C _α)	0.60 N_{res}	6.75	(99.3)	(67.0)

TABLE VIII: LAO – Overview of applied methods, the number of degrees of freedom (DOF) and the overestimation factor d fitted to the lowest 50 frequencies. The square overlap $P_{1,1} = |\langle \nu_1^{approx} | \nu_1^{bench} \rangle|^2$ between the first approximate and benchmark mode, as well as the contribution $Q_1^{approx} = |\langle \nu_1^{approx} | \Delta \mathbf{r}' \rangle|^2$ of the first approximate mode to the conformational change vector are given in %, those for VSA* are only indicative.

property	PHVA	MBH	VSA	VSA*
partially optimized	✓	✓		
localized modes	✓	✓		(✓)
global modes		✓	✓	

TABLE IX: Concluding table: overview of strong points of each method. The VSA* estimates localized frequencies accurately but the modes lack a clear interpretation.

List of figure captions

1. Amines – Partitioning for three amines: n-propylamine, dipropylamine, tripropylamine. (1) For PHVA and MBH, the system is divided in free atoms and blocks (dashed line). In PHVA, the presence of two blocks of infinite mass is equivalent to one combined block of infinite mass. (2) For VSA and VSA*, the system is divided into subsystem (grey region) and environment.
2. Structures of the investigated cinchona alkaloids: quinine and a bis-cinchona derivative, (DHQD)₂PYDZ, which consists of two dihydroquinidinyls (DHQD) linked by a pyridazine (PYDZ). The investigated bond stretches are highlighted in red.
3. Quinine – Partitions for the partial Hessian methods. (1) For PHVA or MBH, the system is divided in free atoms and blocks (dashed line). In PHVA, the presence of two blocks of infinite mass is equivalent to one combined block of infinite mass. Specifically for MBH (not PHVA), all remaining X-H bonds (C-H, O-H) are considered as linear blocks. (2) For VSA and VSA*, the subsystem is divided into subsystem (grey region) and environment.
4. Quinine – Benchmark (full Hessian) frequencies plotted against the approximate PHVA/MBH/VSA/VSA* frequencies, in cm⁻¹, according to the schemes PHVA2/MBH2/VSA2/VSA*2 of Fig. 3. A grayscale indicates the square overlap $|\langle \nu^{full} | \nu^{approx} \rangle|^2$. The frequencies of Table II are marked with an arrow.
5. LAO – Representation of the open (left) and closed (right) conformation of the LAO binding protein. The motion of the first mode, visualized on the open form, corresponds to the hinge-bending motion between the open and closed conformation.
6. Partitioning schemes for proteins. (Top) For MBH, the system can be divided into non-adjointed or adjointed blocks (dashed lines). (Bottom) For VSA, the system is divided into a subsystem (grey region) and the environment.
7. LAO – Cumulative square overlap P_j (in %) with the LAO benchmark modes below 50 cm⁻¹ for various partial Hessian schemes. Note the difference in scale for VSA*.

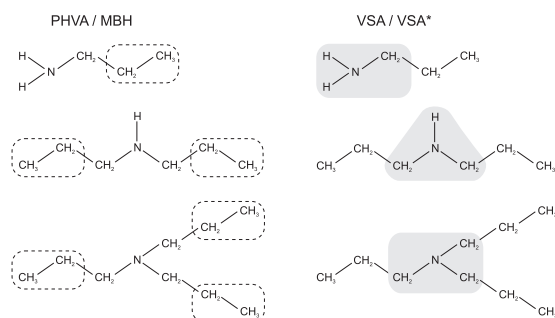


FIG. 1: Amines – Partitioning for three amines: n-propylamine, dipropylamine, tripropylamine. (1) For PHVA and MBH, the system is divided in free atoms and blocks (dashed line). In PHVA, the presence of two blocks of infinite mass is equivalent to one combined block of infinite mass. (2) For VSA and VSA*, the system is divided into subsystem (grey region) and environment.

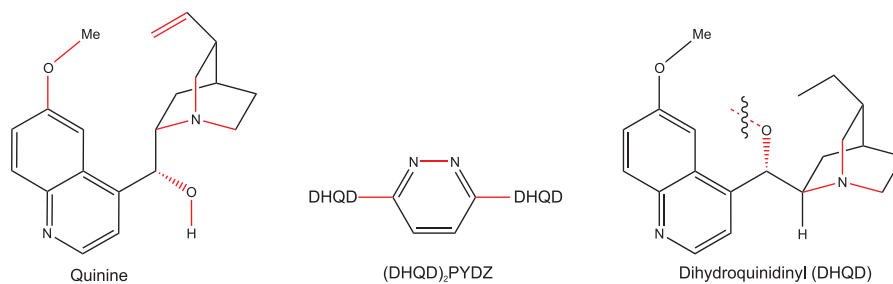


FIG. 2: Structures of the investigated cinchona alkaloids: quinine and a bis-cinchona derivative, (DHQD)₂PYDZ, which consists of two dihydroquinidinyIs (DHQD) linked by a pyridazine (PYDZ). The investigated bond stretches are highlighted in red.

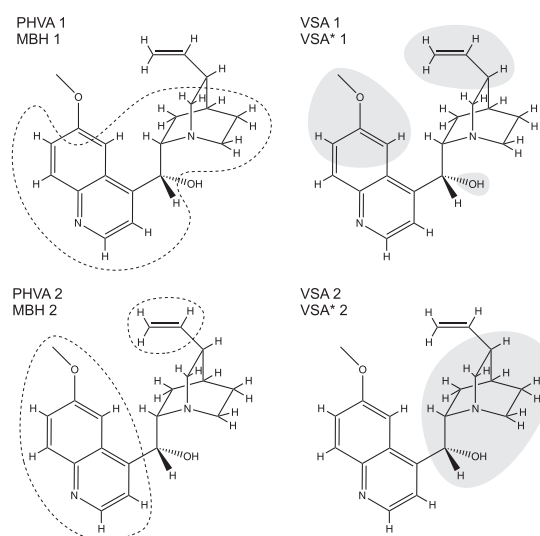


FIG. 3: Quinine – Partitions for the partial Hessian methods. (1) For PHVA or MBH, the system is divided in free atoms and blocks (dashed line). In PHVA, the presence of two blocks of infinite mass is equivalent to one combined block of infinite mass. Specifically for MBH (not PHVA), all remaining X-H bonds (C-H, O-H) are considered as linear blocks. (1) For VSA and VSA*, the subsystem is divided into subsystem (grey region) and environment.

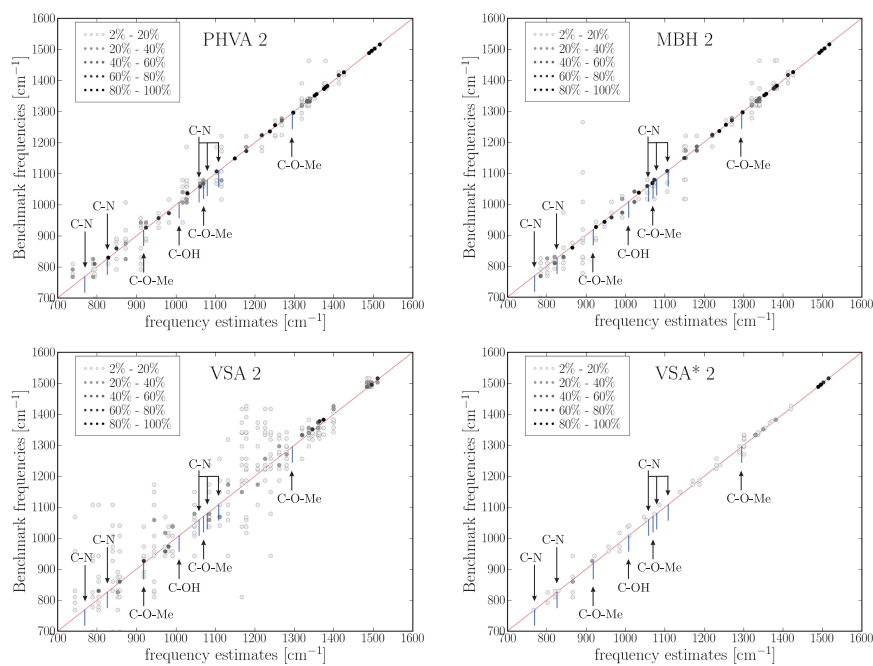


FIG. 4: Quinine – Benchmark (full Hessian) frequencies plotted against the approximate PHVA/MBH/VSA/VSA* frequencies, in cm^{-1} , according to the schemes PHVA2/MBH2/VSA2/VSA*2 of Fig. 3. A grayscale indicates the square overlap $|\langle \nu^{full} | \nu^{approx} \rangle|^2$. The frequencies of Table II are marked with an arrow.

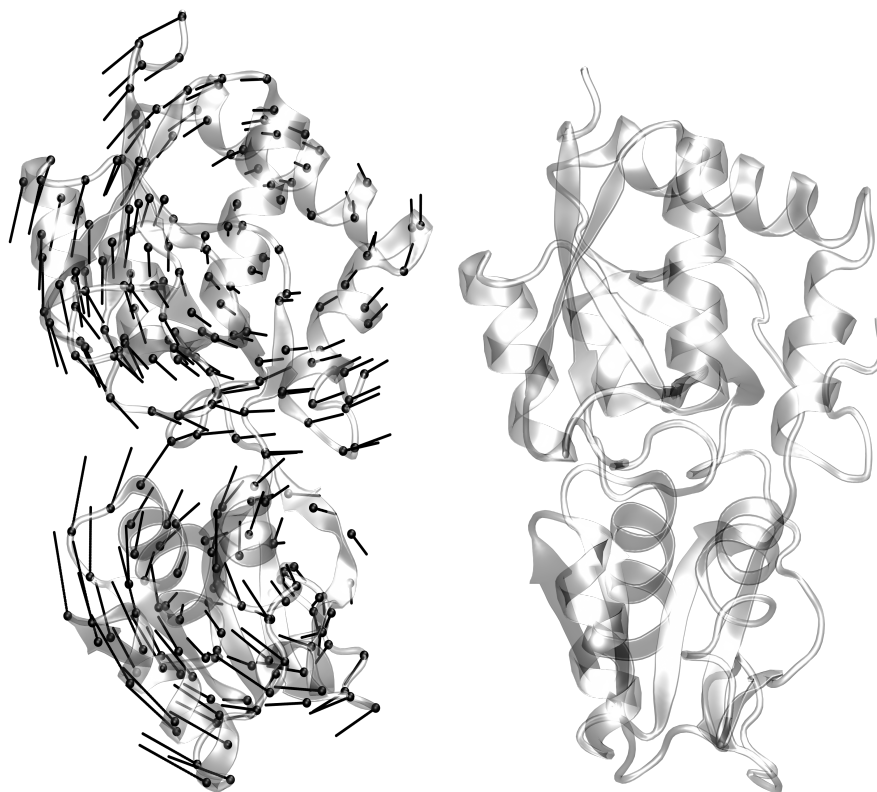


FIG. 5: LAO – Representation of the open (left) and closed (right) conformation of the LAO binding protein. The motion of the first mode, visualized on the open form, corresponds to the hinge-bending motion between the open and closed conformation.

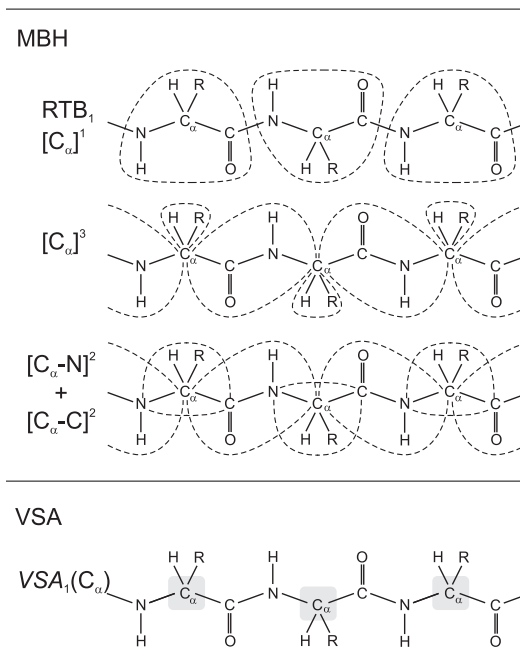


FIG. 6: Partitioning schemes for proteins. (Top) For MBH, the system can be divided into non-adjointed or adjointed blocks (dashed lines). (Bottom) For VSA, the system is divided into a subsystem (grey region) and the environment.

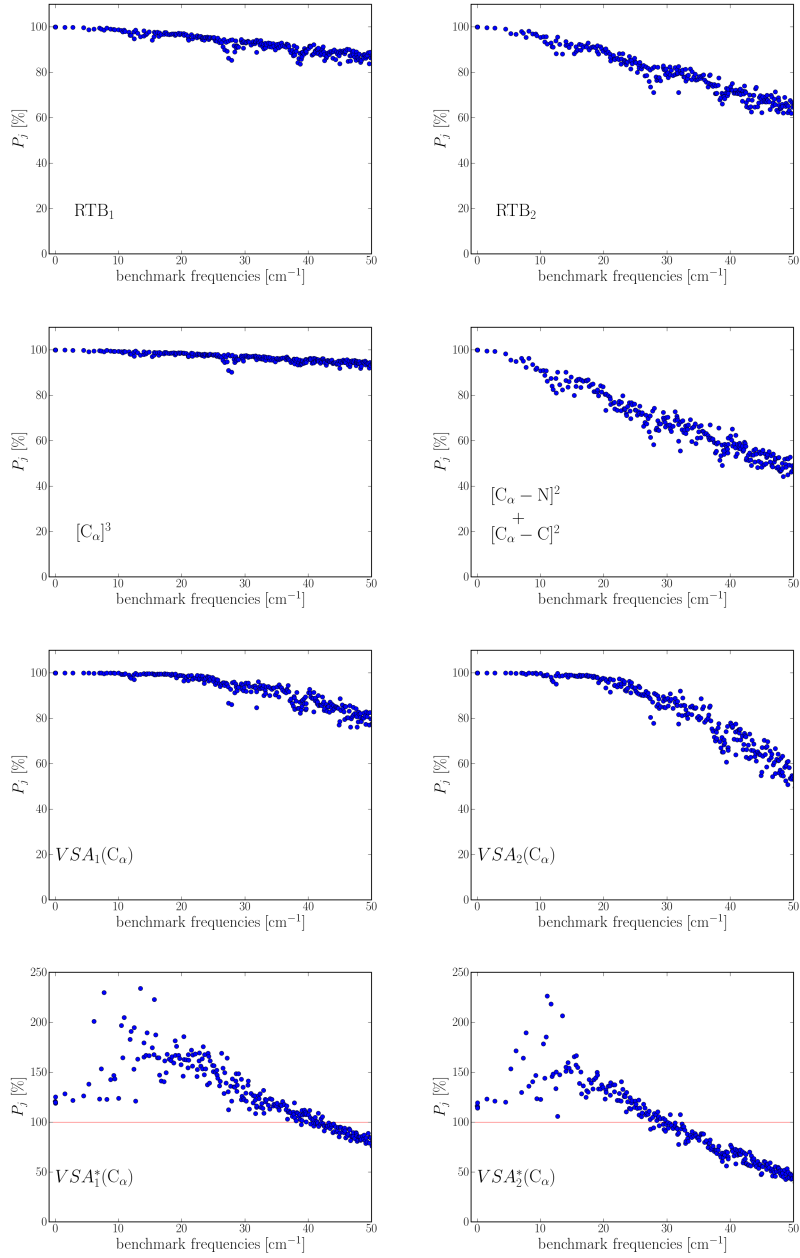


FIG. 7: LAO – Cumulative square overlap $P_{\mathcal{H}_1}$ (in %) with the LAO benchmark modes below 50 cm^{-1} for various partial Hessian schemes. Note the difference in scale for VSA*.

List of publications

Updated March 2009

Publications in international peer-reviewed journals

1. B. Van Tiggelen, D. Anache and A. Ghysels.
Role of mean free path in spatial phase correlation and nodal screening.
Europhys. Lett., 74 (6): 999-1005, 2006.
2. A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen and M. Waroquier.
Vibrational modes in partially optimized molecular systems.
J. Chem. Phys., 126(22): 224102, 2007.
3. A. Ghysels, D. Van Neck and M. Waroquier.
Cartesian formulation of the mobile block Hessian approach to vibrational analysis in partially optimized systems.
J. Chem. Phys., 127(16): 164108, 2007.
4. A. Ghysels, V. Van Speybroeck, T. Verstraelen, D. Van Neck and M. Waroquier.
Calculating reaction rates with partial Hessians: Validation of the mobile block Hessian approach.
J. Chem. Theory Comput., 4(4): 614-625, 2008.
5. D. Lesthaeghe, P. Vansteenkiste, T. Verstraelen, A. Ghysels, C. E. A. Kirschhock, J. A. Martens, V. Van Speybroeck and M. Waroquier.
MFI Fingerprint: how pentasil-induced IR bands shift during zeolite nanogrowth.
J. Phys. Chem. C, 112: 9186-9191, 2008.

List of publications

6. H. L. Woodcock, W. J. Zheng, A. Ghysels, Y. H. Shao, J. Kong and B. R. Brooks.
Vibrational subsystem analysis: A method for probing free energies and correlations in the harmonic limit.
J. Chem. Phys., 129(21): 214109, 2008.
7. A. Ghysels, D. Van Neck, V. Van Speybroeck, B. R. Brooks and M. Waroquier.
Normal modes for large molecules with arbitrary link constraints in the mobile block Hessian approach.
J. Chem. Phys., 130: 084107, 2009.
This article has been selected for the March 1, 2009 issue of *Virtual Journal of Biological Physics Research*. The Virtual Journal, which is published by the American Physical Society and the American Institute of Physics in cooperation with numerous other societies and publishers, is an edited compilation of links to articles from participating publishers, covering a focused area of frontier research. The Virtual Journal is accessible at <http://www.vjbio.org>.
8. A. Ghysels, V. Van Speybroeck, E. Pauwels, D. Van Neck, B. R. Brooks and M. Waroquier.
Mobile block Hessian approach with adjoined blocks: an efficient approach for the calculation of frequencies in macromolecules.
J. Chem. Theory Comput., in press, 2009.
9. A. Ghysels, V. Van Speybroeck, E. Pauwels, S. Catak, B. R. Brooks, D. Van Neck and M. Waroquier.
Comparative study of various normal mode analysis techniques based on partial Hessians.
J. Comput. Chem., submitted, 2009.

Oral contributions

1. Normal modes in partially optimized systems: the Mobile Block Hessian (MBH) approach.
An Ghysels, meeting with SCM, Amsterdam, The Netherlands, September 22, 2007.
2. Normal modes in partially optimized molecular structures: the mobile block Hessian (MBH) approach.
An Ghysels, group meeting at research group of B. R. Brooks, Laboratory of Computational Biology, NHLBI, NIH, Maryland, November 6, 2007.

3. Normal modes in partially optimized systems: the Mobile Block Hessian (MBH) approach.
An Ghysels, Workshop on Mechanistic Analysis of Biological Systems with Novel Computational models, Telluride, Colorado, March 24 - 28, 2008.
4. On the development of a partial vibrational analysis within a QM/MM approach.
An Ghysels, 3rd CMM 'Users Meet Developers' Workshop on QM/MM Simulations, Philadelphia, Pennsylvania, August 21 - 23, 2008.

Poster presentations

Research results were presented as poster presentations on several international conferences.

