# Comparative Study of Various Normal Mode Analysis Techniques Based on Partial Hessians

AN GHYSELS,[1] VERONIQUE VAN SPEYBROECK,[1] EWALD PAUWELS,[1] SARON CATAK,[1]
BERNARD R. BROOKS,[2] DIMITRI VAN NECK,[1] MICHEL WAROQUIER[1]
[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, Maryland 20892*

*Received 26 March 2009; Revised 7 June 2009; Accepted 2 July 2009*
*DOI 10.1002/jcc.21386*
*Published online 7 October 2009 in Wiley InterScience (www.interscience.wiley.com).*

**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 article, 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, that the MBH not only reproduces localized and global modes but also serves as an analysis tool of the spectrum, and that 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.

 J Comput Chem 31: 994–1007, 2010

**Key words:** Hessian; partial Hessian; NMA, normal modes; vibrational modes; mobile block Hessian; MBH; vibrational subsystem analysis; VSA; PHVA; RTB; partial optimization; LAO binding protein; cinchona

## 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 multilevel 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

Additional Supporting Information may be found in the online version of this article.

***Correspondence to:*** A. Ghysels; e-mail: an.ghysels@ugent.be

Contract/grant sponsors: Fund for Scientific Research-Flanders (FWO), Research Board of Ghent University (BOF), Belgian program on Interuniversity Attraction Poles (IAP).

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 $C_\alpha$ 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 coworkers 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 nonoptimized 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 $C_\alpha$ 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),[23–25] 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. A variant of VSA that neglects all mass of the environment will also 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 article, 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 "methods" section a short outline is given of the basic concepts for PHVA, MBH and VSA. In "Assessment of the Methods" section 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 "Guidelines for Selecting Partial Hessian Methods" section, we propose guidelines for the selection of the appropriate NMA method.

## Methods

The harmonic approximation is the underlying assumption in a normal mode analysis. The potential energy surface near a reference point $\boldsymbol{r}^0$ (a vector consisting of the $3N$ coordinates of the $N$ atoms) is expanded in a Taylor series up to second order

$$V(\boldsymbol{r}) \approx V(0) + G^T(\boldsymbol{r} - \boldsymbol{r}^0) + \frac{1}{2}(\boldsymbol{r} - \boldsymbol{r}^0)^T H(\boldsymbol{r} - \boldsymbol{r}^0) \qquad (1)$$

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

$$G_i = \left(\frac{\partial V}{\partial r_i}\right)_0 \qquad (2)$$

$$H_{i,j} = \left(\frac{\partial^2 V}{\partial r_i \partial r_j}\right)_0 \qquad (3)$$

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 $\boldsymbol{r}^0$. 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

$$M\Delta\ddot{\boldsymbol{r}} = -H\Delta\boldsymbol{r} \qquad (4)$$

where $\Delta \boldsymbol{r} = \boldsymbol{r} - \boldsymbol{r}^0$ and $M$ is the $3N \times 3N$ diagonal mass matrix. These are the equations of a system of $3N$ coupled harmonic oscillators. The solutions $\boldsymbol{r}(t) = \boldsymbol{r}^0 + v\cos(\omega t)$ are found from the generalized eigenvalue problem

$$Hv = \omega^2 Mv \tag{5}$$

where $\omega^2$ 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),[11,13] 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 $H_{\mathrm{E}}$ and $M_{\mathrm{E}}$ of the Hessian and mass matrix that correspond with the non-fixed (equilibrated) atoms,

$$H_{\mathrm{E}} v = \omega^2 M_{\mathrm{E}} v. \tag{6}$$

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 $p_{b\alpha}$ ($\alpha = 1, \ldots, 6$) of a translation/rotation group.[26] The instantaneous positions $\boldsymbol{r}_A(t)$ of the atoms $A$ in a block are then generated by applying a common transformation with parameters $p_b$ to the reference positions $\boldsymbol{r}_A^0$,

$$\boldsymbol{r}_A(t) \;=\; \boldsymbol{g}\big(\boldsymbol{r}_A^0, p_b(t)\big), \quad \forall A \in b. \tag{7}$$

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

$$\tilde{V}(\{p_b\}) = V\left(\left\{\boldsymbol{g}\big(\boldsymbol{r}_A^0, p_{b(A)}\big)\right\}\right), \tag{8}$$

where $V(\{\boldsymbol{r}_A\})$ 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:

$$\tilde{H} v = \omega^2 \tilde{M} v \tag{9}$$

with

$$\tilde{H}_{b\alpha,b'\alpha'} = \left(\frac{\partial^2 V}{\partial p_{b\alpha}\partial p_{b'\alpha'}}\right)_0 \tag{10}$$

$$\tilde{M}_{b\alpha,b'\alpha'} = \sum_i m_i \left(\frac{\partial r_i}{\partial p_{b\alpha}}\right)_0 \left(\frac{\partial r_i}{\partial p_{b'\alpha'}}\right)_0. \tag{11}$$

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

$$\begin{pmatrix} H_{\mathrm{ss}} & H_{\mathrm{se}} \\ H_{\mathrm{es}} & H_{\mathrm{ee}} \end{pmatrix} \begin{pmatrix} v_{\mathrm{s}} \\ v_{\mathrm{e}} \end{pmatrix} = \omega^2 \begin{pmatrix} M_{\mathrm{s}} & 0 \\ 0 & M_{\mathrm{e}} \end{pmatrix} \begin{pmatrix} v_{\mathrm{s}} \\ v_{\mathrm{e}} \end{pmatrix}. \tag{12}$$

Projection onto the $s$-subspace yields

$$\big(H_{\mathrm{ss}} + H_{\mathrm{se}}\big(\omega^2 M_{\mathrm{e}} - H_{\mathrm{ee}}\big)^{-1} H_{\mathrm{es}}\big) v_{\mathrm{s}} = \omega^2 M_{\mathrm{s}} v_{\mathrm{s}}. \tag{13}$$

For the low frequency modes one can expand to first order $(\omega^2 M_{\mathrm{e}} - H_{\mathrm{ee}})^{-1} \approx -H_{\mathrm{ee}}^{-1} - \omega^2 H_{\mathrm{ee}}^{-1} M_{\mathrm{e}} H_{\mathrm{ee}}^{-1}$ and derive the VSA normal mode equation as

$$\big(H_{\mathrm{ss}} - H_{\mathrm{se}} H_{\mathrm{ee}}^{-1} H_{\mathrm{es}}\big) v = \omega^2 \big(M_{\mathrm{s}} + H_{\mathrm{se}} H_{\mathrm{ee}}^{-1} M_{\mathrm{e}} H_{\mathrm{ee}}^{-1} H_{\mathrm{es}}\big) v. \tag{14}$$

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] The matrix on the left side of eq. (14) is called the effective Hessian of

the subsystem and was previously discussed by Hinsen et al.[27] in connection with the derivation of an effective pair force constant for protein $C_\alpha$ carbons starting from an all-atom description.

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

$$\left(H_{ss} - H_{se}H_{ee}^{-1}H_{es}\right)v = \omega^2 M_s v. \quad (15)$$

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.

## 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 or localized modes*: Modes that are concentrated in a region of space. 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≡N, C−H, O−H, and N−H stretches.[28] 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.[29–32] In a more extended biosystem, a local mode can for instance refer to the motion of a complete side chain.[33]

*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 suffies 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[34,35] 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).[2,36,37]

### *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.[38] The LAO binding protein was simulated using a molecular mechanics description. The open and closed conformations (2lao, respectively 1lst, in the protein database) were optimized in absence of a solvent with the PARAM22 force field in CHARMM.[39]

The masses, positions and Hessian served as input for the add-on post-processing code[40] written in the programming language Python.[41] 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 $\{|\nu_i\rangle\}$.

### *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.[28] The number of N−H absorption peaks depends on the degree of substitution: two for a primary amine ($NH_2$−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 $CH_2$−$CH_2$−$CH_3$).

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

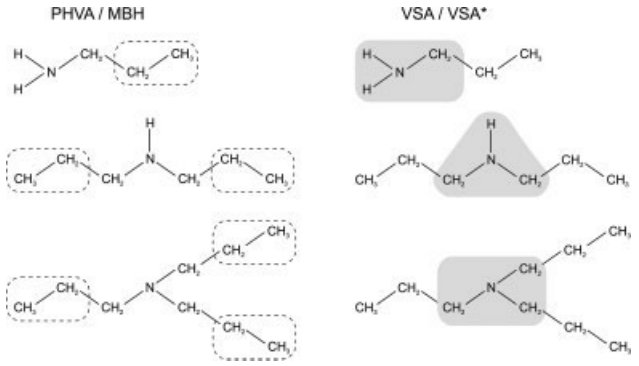


**Figure 1.** Amines — Partitioning for three amines: n-propylamine, dipropylamine, tripropylamine. (1) For PHVA and MBH, the system is divided in free atoms and blocks (dashed line). In PHVA, the presence of two or three blocks of infinite mass is equivalent to one combined block of infinite mass. (2) For VSA and VSA*, the system is divided into subsystem (grey region) and environment.

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 1 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 $|\langle \nu^{\text{full}} | \nu^{\text{approx}} \rangle|^2$ 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 1, 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.

### *Analysis of Spectra with Partial Hessian Methods: Cinchona Alkaloids*

#### *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,[42] chiral auxiliaries in heterogeneous catalysis[43] and chiral catalysts for asymmetric induction in homogeneous transformations.[44] 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 $OsO_4$,[45] where a pyridazine linked bis-dihydroquinidine $(DHQD)_2PYDZ$ derivative (Fig. 2) is used as an enzyme-like binding pocket.[46]

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),

**Table 1.** Amines – High Frequency Spectrum (above 3150 $cm^{-1}$) of Three Amines: n-propylamine, Dipropylamine, Tripropylamine (R is $CH_2-CH_2-CH_3$).

| | $NH_2-R$ | | $NH-R_2$ | $N-R_3$ |
|---|---|---|---|---|
| Method | (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 (*92.9*) | 3617 (*98.8*) | 3526 (*73.4*) | – |

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.

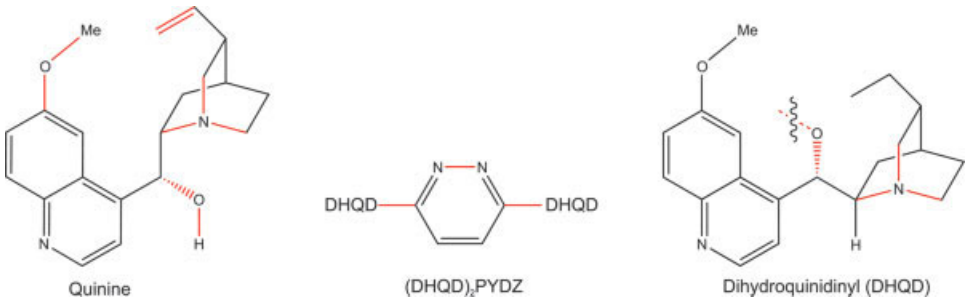


**Figure 2.** Structures of the investigated cinchona alkaloids: quinine and a bis-cinchona derivative, $(DHQD)_2PYDZ$, which consists of two dihydroquinidinyls (DHQD) linked by a pyridazine (PYDZ). The investigated bond stretches are highlighted in red.

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 $C(sp^2)-O$ and $C(sp^3)-O$),
3. the localized modes in the pyridazine in the linker region, the N−N stretch in particular.

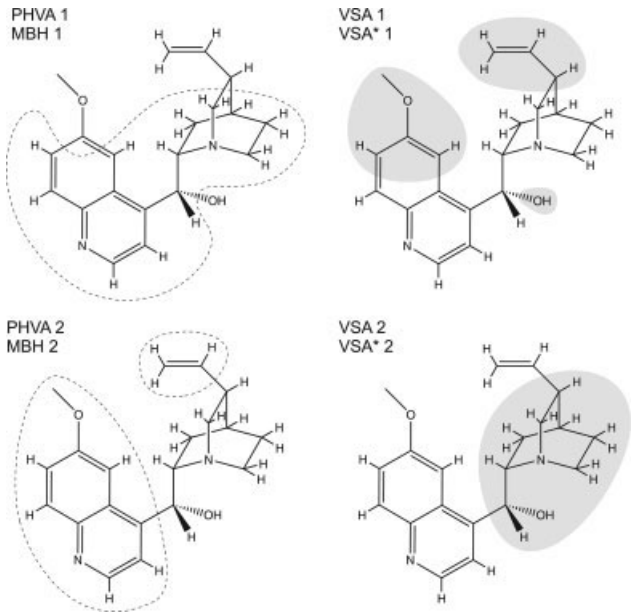


**Figure 3.** Quinine — Partitions for the partial Hessian methods. (1) For PHVA or MBH, the system is divided in free atoms and blocks (dashed line). In PHVA, the presence of two blocks of infinite mass is equivalent to one combined block of infinite mass. Specifically for MBH (not PHVA), all remaining 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.

*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 subsytem. The frequencies and modes obtained with the full Hessian calculation serve as a benchmark, as given in the first column of Table 2. The approximate frequency with highest square overlap $|\langle\nu^{\text{full}}|\nu^{\text{approx}}\rangle|^2$ (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 2), which is a consequence of the nonorthogonality of the modes (see Section). 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.

**Table 2.** 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.

| 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)* |

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 $|\langle\nu^{\text{full}}|\nu^{\text{approx}}\rangle|^2$ with the benchmark mode (in %), those in italics for VSA* are only indicative.

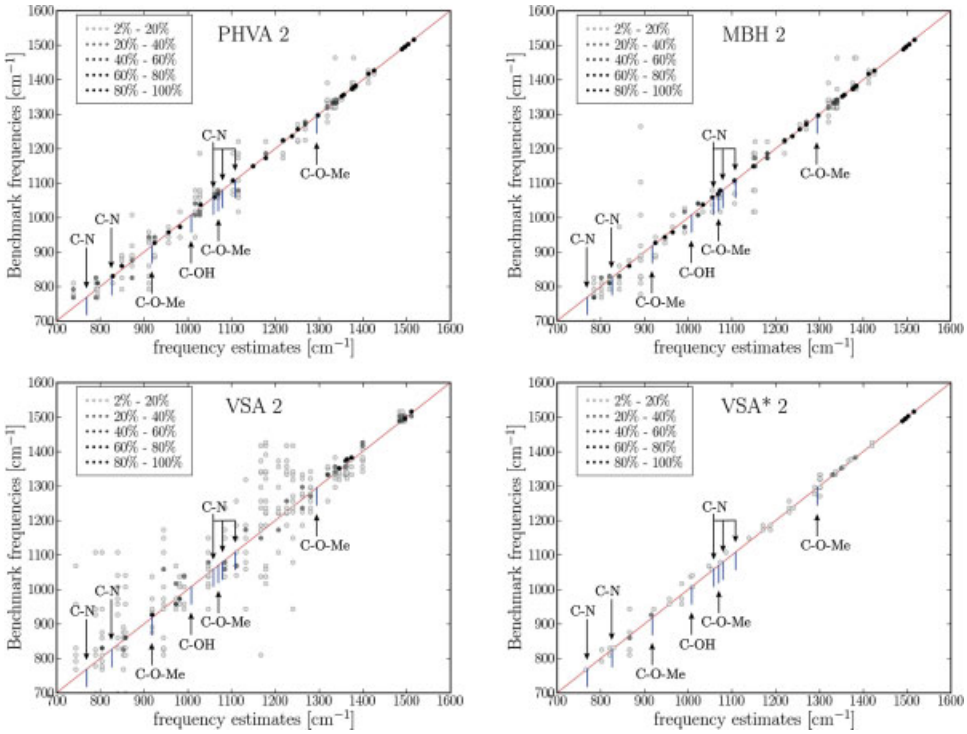


**Figure 4.** Quinine — Benchmark (full Hessian) frequencies plotted against the approximate PHVA/MBH/VSA/VSA* frequencies, in cm$^{-1}$, according to the schemes PHVA2/MBH2/VSA2/VSA*2 of Figure 3. A grayscale indicates the square overlap $|\langle\nu^{\text{full}}|\nu^{\text{approx}}\rangle|^2$. The frequencies of Table 3 are marked with an arrow.

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 $|\langle\nu^{\text{full}}|\nu^{\text{approx}}\rangle|^2$ 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.

### *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 $P_j = \sum_i |\langle\nu_i^{\text{mbh}}|\nu_j^{\text{full}}\rangle|^2$ 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

**Table 3.** Quinine — Which Modes Contribute Mostly to the O−H, C=C, C−O−Me, N−C, and C−OH Stretches?

| 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 |

Benchmark frequencies in $cm^{-1}$ are listed according to the highest contribution $100\% - P_j$ (in %).

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 3 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 4 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. As the two sides are not completely decoupled, the frequencies are not exactly degenerate.

**Table 4.** Bis-Cinchona — Benchmark Frequencies (in $cm^{-1}$) with the Highest Contribution (in %) to the six C−N Stretches in the Two Quinuclidines.

| 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 |

Contribution of each quinuclidine individually is also indicated.

**Table 5.** 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 $C(sp^2)$−O and the Two $C(sp^3)$−O Stretches.

| 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 5 lists the frequencies mostly contributing ($100\% - P_j$) to the four C−O stretches in the linker region of bis-cinchona: a pair of $C(sp^2)$−O and a pair of $C(sp^3)$−O stretches, directly connected to the pyridazine ring or separated by one bond, respectively. By calculating the contribution to the two $C(sp^2)$−O and $C(sp^3)$−O stretches separately, the frequency range of 1300-1480 $cm^{-1}$ could be assigned to $C(sp^2)$−O stretching, and the frequency range 970-1115 $cm^{-1}$ to $C(sp^3)$−O stretching. The latter is similar to the C−OH modes identified for quinine in Table 3.

For the identification of the modes that are mainly localized in the linker region, we grouped all 8 pyridazine atoms in one block. Table 6 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 (dihedral) angle bending motions. Frequency 1249 $cm^{-1}$ has been identified as the dominant N−N stretching mode.

**Table 6.** Bis-Cinchona — Benchmark Frequencies (in $cm^{-1}$) Which are Mainly Localized in the Pyridazine (pydz).

| 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 | – | – | – | – |

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.

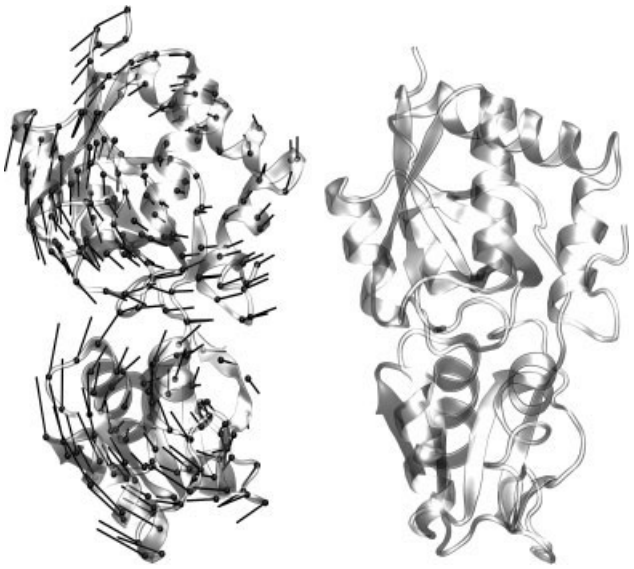

**Figure 5.** LAO — Representation of the open (left) and closed (right) conformation of the LAO binding protein. The motion of the first mode, visualized on the open form, corresponds to the hinge-bending motion between the open and closed conformation.

### *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.[47] 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".[48] Figure 5 illustrates the conformational change between the open, apo form[49] (pdb: 2lao) and the closed form of LAO in complex with lysine[47] (pdb: 1lst). 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.[37,50,51] 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:

$$\Delta \boldsymbol{r}' = \frac{M^{1/2}\Delta \boldsymbol{r}}{\Delta \boldsymbol{r}^T M \Delta \boldsymbol{r}}, \qquad (16)$$

where $\Delta \boldsymbol{r} = \boldsymbol{r}_{2\mathrm{lao}} - \boldsymbol{r}_{1\mathrm{lst}}$. However, since several atoms are ill-resolved in the X-ray diffraction experiments, $\Delta \boldsymbol{r}$ is determined with structures optimized with the CHARMM force field, as discussed in the computational details.

Table 7 lists the square overlap $Q_j = |\langle \nu_j | \Delta \boldsymbol{r}' \rangle|^2$ 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.

#### *Block Choices*

An overview of the block choices and the corresponding number of degrees of freedom is given in Table 8.

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 $[\mathrm{A}]^{s_A}$ 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 $[\mathrm{C}_\alpha]^1$ 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 $\mathrm{RTB}_1$. The RTB method was extensively tested by Durand et al. and later Tama et al.[19,20,37,52,53] They also considered block choices with more than one residue per block, denoted $\mathrm{RTB}_n$, with $n$ the number of residues in each block. This considerably reduces the degrees of freedom to $6N_{\mathrm{res}}/n$, with $N_{\mathrm{res}}$ 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 $[\mathrm{C}_\alpha]^3$, where peptide bonds and side chains are blocks connected by the adjoining atom $\mathrm{C}_\alpha$, and $[\mathrm{C}_\alpha{-}\mathrm{N}]^2 + [\mathrm{C}_\alpha{-}\mathrm{C}]^2$ which only selects the dihedral angles $\phi$ and $\psi$ as variables.

**Table 7.** LAO — Frequencies (in $\mathrm{cm}^{-1}$) of the First Normal Modes of the LAO Binding Protein, Calculated with the Full Hessian on the Open Form.

| Benchmark [$\mathrm{cm}^{-1}$] | $Q_j$ [%] |
|---|---|
| 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 |

For each mode, the square overlap $Q_j = |\langle \nu_j^{\mathrm{bench}} | \Delta \boldsymbol{r}' \rangle|^2$ (in %) with the "pac-man" difference vector is given.

**Table 8.** LAO — Overview of Applied Methods, the Number of Degrees of Freedom (DOF) and the Overestimation Factor $d$ Fitted to the Lowest 50 Frequencies.

| Method | DOF | $d$ | $P_{1,1}$ | $Q_1^{\text{approx}}$ |
|---|---|---|---|---|
| | MBH: RTB | | | |
| $RTB_1/[C_\alpha]^1$ | $6N_{res}$ | 1.70 | 93.7 | 77.7 |
| $RTB_2$ | $3N_{res}$ | 2.08 | 90.2 | 79.3 |
| $RTB_3$ | $2N_{res}$ | 2.41 | 85.7 | 80.9 |
| $RTB_4$ | $1.50N_{res}$ | 2.65 | 87.5 | 80.3 |
| $RTB_5$ | $1.20N_{res}$ | 2.90 | 87.6 | 79.9 |
| | MBH: Adjoined Blocks | | | |
| $[C_\alpha]^3$ | $6N_{res}$ | 1.55 | 95.6 | 76.0 |
| $[C_\alpha\text{-N}]^2+[C_\alpha\text{-C}]^2$ | $2N_{res}$ | 2.09 | 94.7 | 72.5 |
| | VSA | | | |
| $VSA(C_\alpha)$ | $3N_{res}$ | 1.04 | 100.0 | 66.1 |
| $VSA_2(C_\alpha)$ | $1.50N_{res}$ | 1.06 | 100.0 | 66.1 |
| $VSA_3(C_\alpha)$ | $N_{res}$ | 1.09 | 100.0 | 66.1 |
| $VSA_4(C_\alpha)$ | $0.75N_{res}$ | 1.12 | 100.0 | 66.1 |
| $VSA_5(C_\alpha)$ | $0.60N_{res}$ | 1.18 | 100.0 | 66.1 |
| | VSA* | | | |
| $VSA^*(C_\alpha)$ | $3N_{res}$ | 3.34 | *(99.9)* | *(66.1)* |
| $VSA_2^*(C_\alpha)$ | $1.50N_{res}$ | 4.55 | *(99.9)* | *(66.1)* |
| $VSA_3^*(C_\alpha)$ | $N_{res}$ | 5.50 | *(99.9)* | *(66.0)* |
| $VSA_4^*(C_\alpha)$ | $0.75N_{res}$ | 6.00 | *(99.8)* | *(65.0)* |
| $VSA_5^*(C_\alpha)$ | $0.60N_{res}$ | 6.75 | *(99.3)* | *(67.0)* |

The square overlap $P_{1,1} = |\langle \nu_1^{\text{approx}} | \nu_1^{\text{bench}} \rangle|^2$ between the first approximate and benchmark mode, as well as the contribution $Q_1^{\text{approx}} = |\langle \nu_1^{\text{approx}} | \Delta \boldsymbol{r}' \rangle|^2$ of the first approximate mode to the conformational change vector are given in %, those for VSA* are only indicative.

Note that instead of block choices based on experience, one could make use of automized routines to determine the rigid regions of the molecular system. The FIRST (Floppy Inclusion and Rigid Substructure Topography) program for instance divides the system into rigid blocks, possibly adjoined, based on a rigidity analysis.[54]

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.

### *Mode Reproduction*

The quality of the reproduction of a benchmark mode $|\nu_j^{\text{full}}\rangle$ by the approximate modes is measured by the cumulative square overlap $P_j = \sum_i |\langle \nu_i^{\text{approx}} | \nu_j^{\text{full}} \rangle|^2$, 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 $[C_\alpha\text{-N}]^2+[C_\alpha\text{-C}]^2$ 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 approximation — 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 8 by the square overlap $P_{1,1} = |\langle \nu_1^{\text{approx}} | \nu_1^{\text{full}} \rangle|^2$. Also, the square overlap $Q_1^{\text{approx}} = |\langle \nu_1^{\text{approx}} | \Delta \boldsymbol{r}' \rangle|^2$ 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.

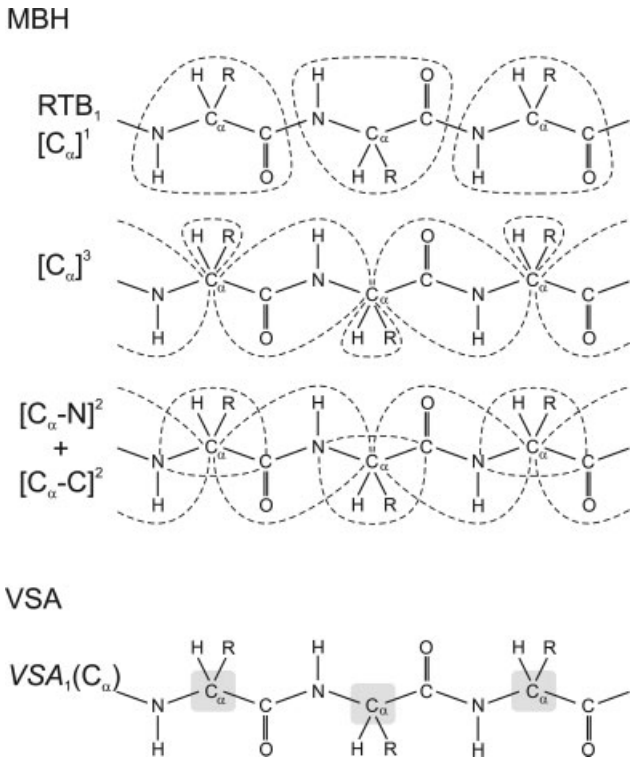


**Figure 6.** Partitioning schemes for proteins. (Top) For MBH, the system can be divided into non-adjoined or adjoined blocks (dashed lines). (Bottom) For VSA, the system is divided into a subsystem (grey region) and the environment.

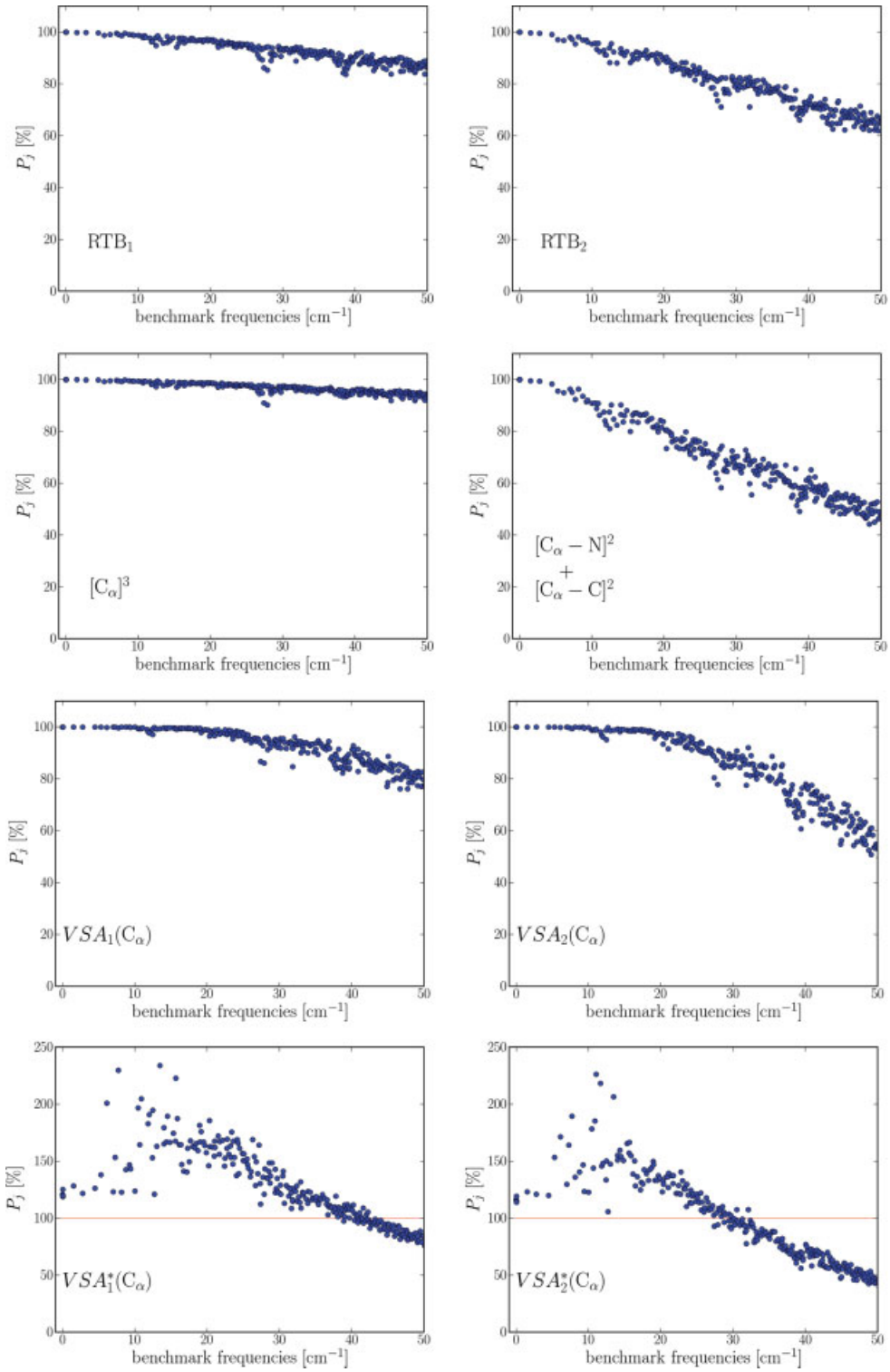


**Figure 7.** LAO — Cumulative square overlap $P_j$ (in %) with the LAO benchmark modes below 50 cm$^{-1}$ for various partial Hessian schemes. Note the difference in scale for VSA*. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

*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,

$$\nu_{\text{approx}} = d.\nu_{\text{bench}}. \tag{17}$$

The overestimation factor $d$ is usually obtained by a least mean square fit in the low frequency range. The values in Table 8 were fitted with the lowest 50 frequencies, lying in the range 0–20 $\text{cm}^{-1}$.

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 $\text{RTB}_1$, $\text{RTB}_2$, $\text{RTB}_3$, and $\text{RTB}_5$ 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 $\text{RTB}_3$ and the $[\text{C}_\alpha\text{-N}]^2$+$[\text{C}_\alpha\text{-C}]^2$ 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 ($\text{C}_\alpha$, 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 $N_{\text{res}}$, as in $\text{VSA}_4(\text{C}_\alpha)$ and $\text{VSA}_5(\text{C}_\alpha)$, 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 $\text{VSA}^*_n(\text{C}_\alpha)$ partitions.

## 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 9.

***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.

**Table 9.** Concluding Table: Overview of Strong Points of Each Method.

| Property | PHVA | MBH | VSA | VSA* |
|---|---|---|---|---|
| Partially optimized | ✓ | ✓ | | |
| Localized modes | ✓ | ✓ | | (✓) |
| Global modes | | ✓ | ✓ | |

The VSA* estimates localized frequencies accurately but the modes lack a clear interpretation.

***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. 55. 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-adjoined and adjoined 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 neglection of the environment mass is in this case unrealistic such that VSA* yields unphysical low frequency modes.

***Computational Aspects***

The use of a partial Hessian method also offers the possibility to reduce the computational cost of normal mode analysis. The computational performance of partial Hessian methods is mainly determined by the number of second derivatives that have to be calculated, the disk space needed for storage, and the dimension of the matrices that have to be diagonalized.

In PHVA the dimension is restricted to that of the Hessian submatrix $H_{\text{E}}$ corresponding to the free atoms [see eq. (6)]. The smaller the number of free atoms, the more important the computational profit.

The MBH needs the construction of a partial Hessian $\tilde{H}$ with a dimension significantly smaller than that of the full Hessian [see eq. (9)], and hence the whole procedure together with the diagonalization step is much faster. To illustrate, all MBH modes corresponding with block choice $[\text{C}_\alpha]^3$ for the caspase-1 system with 8252 atoms (see ref. 22) could be calculated in 14% of the time needed by

CHARMM to calculate the lowest 20 frequencies only. The computational performance of MBH is most efficient if the program package under consideration is able to partially optimize the molecular system by freezing the internal degrees of freedom in each block during the energy minimization, and limiting the number of second derivatives to those with respect to the remaining block parameters. This feature will be available in the next release of ADF.[56] When this feature is not implemented (as in Gaussian03[38]), the full Hessian needs first to be calculated and stored before the transformation to the mobile block Hessian matrix $\tilde{H}$.

The VSA and VSA* are not developed for their computational performance. All full Hessian matrix elements need to be calculated for the construction of the effective Hessian [see eqs. (14 and 15)]. The latter also requires the inverse of the Hessian matrix $H_{ee}$ in the environment subspace. Solving the NMA equations then boils down to a generalized eigenvalue problem with subsystem dimension.

## 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 illustrate 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.

## References


1. Lesthaeghe, D.; Vansteenkiste, P.; Verstraelen, T.; Ghysels, A.; Kirschhock, C. E. A.; Martens, J. A.; Van Speybroeck, V.; Waroquier, M. J Phys Chem C, 2008, 112, 9186.
2. Cui, Q.; Bahar, I. Mathematical and Computational Biology Series; Chapman & Hall/CRC, Taylor & Francis Group: Boca Raton, FL, 2006.
3. Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227.
4. Assfeld, X.; Rivail, J. L. Chem Phys Lett 1996, 263, 100.
5. Gao, J. L.; Amara, P.; Alhambra, C.; Field, M. J. J Phys Chem A 1998, 102, 4714.
6. Zhang, Y. K.; Lee, T. S.; Yang, W. T. J Chem Phys 1999, 110, 46.
7. Senn, H. M.; Thiel, W. QM/MM methods for biological systems. In Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations, Volume 268 of Topical Current Chemistry Springer-Verlag Berlin: Berlin, 2007; pp. 173–290.
8. Tirion, M. M. Phys Rev. Lett 1996, 77, 1905.
9. Bahar, I.; Atilgan, A. R.; Erman, B. Fold Des 1997, 2, 173.
10. Hinsen, K. Protein Struct Funct Genet 1998, 33, 417.
11. Li, H.; Jensen, J. H. Theor Chem Acc 2002, 107, 211.
12. Besley, N. A.; Metcalf, K. A. J Chem Phys 2007, 126, 035101.
13. Jin, S. Q.; Head, J. D. Surf Sci 1994, 318, 204.
14. Calvin, M. D.; Head, J. D.; Jin, S. Q. Surf Sci 1996, 345, 161.
15. Head, J. D. Int J Quantum Chem 1997, 65, 827.
16. Head, J. D. Shi, Y.; Int J Quantum Chem 1999, 75, 815.
17. Head, J. D. Int J Quantum Chem 2000, 77, 350.
18. Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Verstraelen, T.; Waroquier, M. J Chem Phys 2007, 126, 224102.
19. Durand, P.; Trinquier, G.; Sanejouand, Y. H. Biopolymers 1994, 34, 759.
20. Tama, F.; Gadea, F. X.; Marques, O. Sanejouand, Y. H.; Proteins: Struct Funct Genet 2000, 41, 1–7.
21. Ghysels, A.; Van Speybroeck, V.; Pauwels, E.; Van Neck, D.; Brooks, B. R.; Waroquier, M. J Chem Theory Comput 2009, 5, 1203.
22. Ghysels, A.; Van Neck, D.; Van Speybroeck, V.; Brooks, B. R.; Waroquier, M. J Chem Phys 2009, 130, 084107.
23. Zheng, W. J.; Brooks, B. R. Biophys J 2005, 89, 167.
24. Woodcock, H. L.; Zheng, W. J.; Ghysels, A.; Shao, Y. H.; Kong, J.; Brooks, B. R. J Chem Phys 2008, 129, 214109.
25. Hafner, J.; Zheng, W. J Chem Phys 2009, 130, 194111.
26. Ghysels, A.; Van Neck, D.; Waroquier, M. J Chem Phys 2007, 127, 164108.
27. Hinsen, K.; Petrescu, A. J.; Dellerue, S.; Bellissent-Funel, M. C.; Kneller, G. R. Chem Phys 2000, 261, 25.
28. Fessenden, R. J.; Fessenden, J. S. Organic Chemistry, 4th ed; Brooks/Cole Publishing Company; Belmont, California, 1990.
29. Eyring, H. J Chem Phys 1935, 3, 107.
30. Evans, M. G.; Polanyi, M. Trans Faraday Soc 1935, 31, 875.
31. Laidler, K. J. Chemical Kinetics, 3rd edition, Harper and Row, New York, 1987.
32. Mc Quarrie, D. A.; Simon, J. D. Physical Chemistry — A Molecular Approach. University Science Books, Sausalito: California, 1997.
33. Jacob, C. R.; Reiher, M. J Chem Phys 2009, 130, 15.
34. Vansteenkiste, P.; Van Speybroeck, V.; Marin, G. B.; Waroquier, M. J Phys Chem A 2003, 107, 3139.
35. Van Cauter, K.; Van Speybroeck, V.; Vansteenkiste, P.; Reyniers, M. F.; Waroquier, M. ChemPhysChem 2006, 7, 131.
36. Marques, O.; Sanejouand, Y. H. Proteins: Struct Funct Genet 1995, 23, 557.
37. Tama, F.; Sanejouand, Y. H. Protein Eng 2001, 14, 1.
38. 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.
39. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
40. http://molmod.ugent.be/code.

41. http://www.python.org.
42. Maier, N. M.; Schefzick, S.; Lombardo, G. M.; Feliz, M.; Rissanen, K.; Lindner, W.; Lipkowitz, K. B. J Am Chem Soc 2002, 124, 8611.
43. Mallat, T.; Orglmeister, E.; Baiker, A. Chem Rev 2007, 107, 4863.
44. France, S.; Guerin, D. J.; Miller, S. J.; Lectka, T. Chem Rev 2003, 103, 2985.
45. Kolb, H. C.; VanNieuwenhze, M. S.; Sharpless, K. B. Chem Rev 1994, 94, 2483.
46. Corey, E. J.; Noe, M. C. J Am Chem Soc 1996, 118, 319.
47. Kang, C. H.; Shin, W. C.; Yamagata, Y.; Gokcen, S.; Ames, G. F.; Kim, S. H. J Biol Chem 1991, 266, 23893.
48. Sack, J. S.; Saper, M. A.; Quiocho, F. A. J Mol Biol 1989, 206, 171.
49. Oh, B. H.; Pandit, J.; Kang, C. H.; Nikaido, K.; Gokcen, S.; Ames, G. F. L.; Kim, S. H. J Biol Chem 1993, 268, 11348.
50. Keskin, O.; Jernigan, R. L.; Bahar, I. Biophys J 2000, 78, 2093.
51. Zheng, W. J.; Brooks, B. R. Biophys J 2006, 90, 4327.
52. Tama, F.; Brooks, C. L. J Mol Biol 2002, 318, 733.
53. Tama, F. Protein Pept Lett 2003, 10, 119.
54. Gohlke, H.; Thorpe, M. F. Biophys J 2006, 91, 2115.
55. Ghysels, A.; Van Speybroeck, V.; Verstraelen, T.; Van Neck, D.; Waroquier, M. J Chem Theory Comput 2008, 4, 614.
56. http://www.scm.com.