

# Normal mode analysis of macromolecular systems with the Mobile Block Hessian method

An Ghysels\*, Veronique Van Speybroeck\*, Dimitri Van Neck\*, Bernard R. Brooks† and Michel Waroquier\*

\*Center for Molecular Modeling, Ghent University, Technologiepark 903, 9052 Zwijnaarde, Belgium

†Laboratory for Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, 5636 Fisher's Ln, 20851 Rockville, MD, USA

**Abstract.** Until recently, normal mode analysis (NMA) was limited to small proteins, not only because the required energy minimization is a computationally exhausting task, but also because NMA requires the expensive diagonalization of a  $3N_a \times 3N_a$  matrix with  $N_a$  the number of atoms. A series of simplified models has been proposed, in particular the Rotation-Translation Blocks (RTB) method by Tama et al. for the simulation of proteins. It makes use of the concept that a peptide chain or protein can be seen as a subsequent set of rigid components, i.e. the peptide units. A peptide chain is thus divided into rigid blocks with six degrees of freedom each.

Recently we developed the Mobile Block Hessian (MBH) method, which in a sense has similar features as the RTB method. The main difference is that MBH was developed to deal with partially optimized systems. The position/orientation of each block is optimized while the internal geometry is kept fixed at a plausible – but not necessarily optimized – geometry. This reduces the computational cost of the energy minimization. Applying the standard NMA on a partially optimized structure however results in spurious imaginary frequencies and unwanted coordinate dependence. The MBH avoids these unphysical effects by taking into account energy gradient corrections. Moreover the number of variables is reduced, which facilitates the diagonalization of the Hessian.

In the original implementation of MBH, atoms could only be part of one rigid block. The MBH is now extended to the case where atoms can be part of two or more blocks. Two basic linkages can be realized: (1) blocks connected by one link atom, or (2) by two link atoms, where the latter is referred to as the hinge type connection. In this work we present the MBH concept and illustrate its performance with the crambin protein as an example.

**Keywords:** NMA, vibrational analysis, IR spectrum, reduced dimension, partial optimization

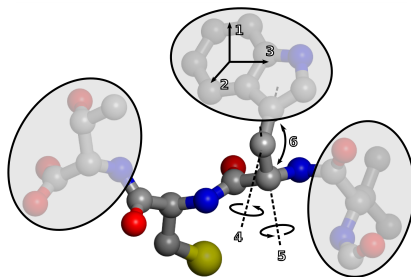
**PACS:** 33.20.Tp, 31.15

## INTRODUCTION

Normal mode analysis (NMA) is a useful technique for exploring the shape of the potential energy surface within the harmonic oscillator approximation. Conformational changes of macromolecules, for instance, are essential in the understanding of protein function and drug design, since the lowest frequency modes are proven to contribute most to a conformational change. In the higher frequency range, characteristic localized modes are used for the identification of functionalities.

The frequencies/modes are obtained from the diagonalization of the Hessian, which contains the second derivatives of the energy with respect to displacements of the nuclear coordinates. 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 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. Not only the calculation of the second derivatives but also the storage of the huge Hessian and its diagonalization become problematic. 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.

We have developed an approximate NMA method, which is able to zoom on the frequency range of interest [1]. The main concept of the Mobile Block Hessian (MBH) method is the introduction of blocks of atoms which move coherently during the small amplitude vibrations. Each block has three translational and three rotational degrees of freedom, as depicted in Fig. 1. The internal structure of a block need not be optimized, so a partial geometry optimization is sufficient to apply MBH, in contrast to the plain Rotation-Translation Block (RTB) method by Tama et



**FIGURE 1.** Schematic representation of the block concept of 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, as well as during the vibrational analysis. Nonlinear (linear) blocks have six (five) degrees of freedom each (3 translations and 3 rotations).

al.[2]. The block choice determines whether the low or high frequency modes, or the coupling between both groups, are reproduced. The block concept was recently extended to a variety of block choices, including linear blocks and blocks which may have atoms in common. This makes MBH a versatile NMA method which can also play the role of an analysis tool of complex spectra, e.g. to identify stretches or dihedral angle bending modes. In addition, the reduced number of degrees of freedom brings along a reduction in computational load by the smaller number of calculated derivatives and the smaller Hessian dimension, which is essential for advanced models such as hybrid QM/MM models. In the following sections the MBH equations will be constructed and its performance will be illustrated with an example from the field of proteins.

## METHODS

### Normal mode analysis

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

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

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

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

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  $r^0$ . If the stationary point is a minimum, the matrix  $H$  is positive semidefinite. The equations of motions of a molecule in a harmonic well are given by  $M\Delta\ddot{r} = -H\Delta r$  where  $\Delta r = r - r^0$  and  $M$  is the  $3N \times 3N$  diagonal mass matrix. These are the equations of a system of  $3N$  coupled harmonic oscillators. The solutions  $r(t) = r^0 + v \cos(\omega t)$  are found from the eigenvalue problem  $Hx = \omega^2 Mx$  where  $\omega^2$  and  $x$  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.

### Partially optimized structures: PHVA and MBH

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) [3, 4], 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 ‘E’) atoms,  $H_E x = \omega^2 M_E x$ .

The second method is the Mobile Block Hessian (MBH) approach [1]. 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. 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 [5]. The instantaneous positions  $r_A(t)$  of the atoms  $A$  in a block are then generated by applying a common transformation with parameters  $p_b$  to the reference positions  $r_A^0$ ,

$$r_A(t) = g(r_A^0, p_b(t)), \quad \forall A \in b. \quad (3)$$

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 expressed in the set of  $\{p_{b\alpha}\}$  variables becomes

$$\tilde{V}(\{p_b\}) = V(\{g(r_A^0, p_{b(A)})\}), \quad (4)$$

where  $V(\{r_A\})$  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:  $\tilde{H}x = \omega^2 \tilde{M}x$ , with

$$\tilde{H}_{b\alpha, b'\alpha'} = \left( \frac{\partial^2 V}{\partial p_{b\alpha} \partial p_{b'\alpha'}} \right)_0 \quad (5)$$

$$\tilde{M}_{b\alpha, b'\alpha'} = \sum_i m_i \left( \frac{\partial r_i}{\partial p_{b\alpha}} \right)_0 \left( \frac{\partial r_i}{\partial p_{b'\alpha'}} \right)_0. \quad (6)$$

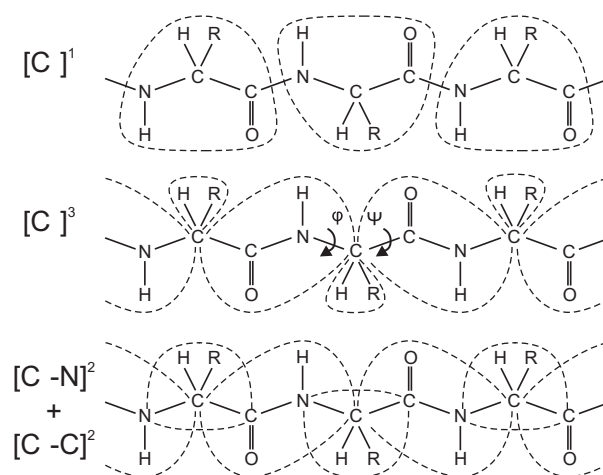
Recently, the MBH method has been extended to the case of adjoined blocks which have one or more atoms in common (adjoining atoms) [6, 7]. The adjoining atoms restrict the relative motion of the blocks, allowing greater flexibility to concentrate on the relevant motions.

## VALIDATION OF MOBILE BLOCK HESSIAN (MBH) METHOD

*Low frequency modes of proteins.* The MBH has been validated on a range of chemically diverse systems, with the number of atoms varying from 9 up to 20.000 atoms. In this paper, the subclass of protein systems is highlighted. Conformational changes of macromolecules are essential in the understanding of e.g. proteins and drug design. The theoretical prediction is far from trivial, especially for large molecules. In many cases, collective motions are present which occur on a timescale ( $\sim$  ms) that is too long to be accessible through molecular dynamics simulations. NMA has been proven succesful in exploring the potential energy surface (PES) within the harmonic oscillator approximation, because the lowest frequency modes contribute the most to a conformational change. We will show that the MBH is a computationally attractive method that selects modes from the lower spectrum.

*Block schemes for peptides.* In order to select an appropriate partitioning scheme for proteins, several block choices are tested as depicted in figure 2. The labels  $[C_\alpha]^k$  indicate to how many blocks the central carbon atom belongs. For instance, in block choice  $[C_\alpha]^3$  the carbon atom belongs to three blocks. Crambin, a polypeptide with 46 residues and 646 atoms, is taken as a test case.

The peptide bond (C=O-NH) is known to be fairly rigid, so peptide units are consistently treated as blocks. Block choice  $[C_\alpha]^1$  was proposed by Tama et al.[2]; it has no linked blocks. In the other block choices, the  $C_\alpha$  carbon



**FIGURE 2.** Three block choices for peptides. The dihedral angles  $\phi$  and  $\psi$  are responsible for the protein's backbone flexibility.

is the adjoining atom between the blocks. The approximate frequencies systematically overestimate the benchmark frequencies, since a constrained motion experiences higher resistance. The modes on the other hand are very well reproduced: the cumulative square overlap between  $[C_\alpha]^3$  and benchmark modes shows that the lowest 50 modes are reproduced by over 90%, which is an improvement with respect to the original scheme  $[C_\alpha]^1$  of Tama et al. Application of the hinge type connection in block choice  $[C_\alpha-N]^2 + [C_\alpha-C]^2$ , selects the internal rotations around the N- $C_\alpha$  and  $C_\alpha$ -C bonds (dihedral angles  $\phi$  and  $\psi$ ). Despite the rude approximation, the 20 lowest modes are reproduced by over 80%. As in practice mode following is only applied on the lowest five modes, this is again a satisfying result.

## CONCLUSION

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 a specific part of the spectrum. For instance, the crambin example illustrated how the extension of the MBH to linked blocks allows to reproduce the low frequency modes of proteins.

## ACKNOWLEDGMENTS

This work is supported by the Fund for Scientific Research - Flanders (FWO) and the Research Board of Ghent University (BOF).

## REFERENCES

1. A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, *J. Chem. Phys.* **126**, 224102 (2007).
2. F. Tama, F. X. Gadea, O. Marques, and Y. H. Sanejouand, *Proteins: Struct. Funct. Genet.* **41**, 1–7 (2000).
3. S. Q. Jin, and J. D. Head, *Surf. Sci.* **318**, 204–216 (1994).
4. H. Li, and J. H. Jensen, *Theor. Chem. Acc.* **107**, 211–219 (2002).
5. A. Ghysels, D. Van Neck, and M. Waroquier, *J. Chem. Phys.* **127**, 164108 (2007).
6. A. Ghysels, V. Van Speybroeck, E. Pauwels, D. Van Neck, B. R. Brooks, and M. Waroquier, *J. Chem. Theory Comput.* **5**, 1203–1215 (2009).
7. A. Ghysels, D. Van Neck, V. Van Speybroeck, B. R. Brooks, and M. Waroquier, *J. Chem. Phys.* **130**, 084107 (2009).