Accurate Normal Modes Calculations of Protein Molecule by an Iterative Approach

Hyo Weon Jang

Department of Chemistry, Sunchon National University, Sunchon, Chonnam 540-742, Korea
Received June 26, 1998

The analysis of low-frequency normal modes of macromolecules involved in biological processes, is useful for studying the relations between multiple stable conformations of molecules like proteins. Upon binding a substrate, the protein molecule (an enzyme, perhaps) may undergo large conformational changes involving distant parts of the molecule which may not be near the active site, but still may be important in understanding the details of the enzyme catalytic action. Such collective motions are often characterized by long (nanosecond or beyond) time-scale, and it may be possible to identify one or a few low-frequency normal modes which would connect those conformations in some cases.[1,2] This kind of theoretical analysis may be useful for getting information about the motion of various parts of the molecule during the conformational changes.

One of the major challenges in this type of analysis, lies in the difficulty of computing the vectors and frequencies of normal modes themselves. It arises from the large dimension of the force constant matrix (3 N where N is the number of atoms in the molecule). Mathematically, the normal modes are eigenvectors of force constant matrix in mass-weighted Cartesian coordinates (so called, Hessian matrix) and the squares of normal mode frequencies are the corresponding eigenvalues. If the matrix and auxiliary vectors needed are too large to fit into the core memory of the computer available, or if it would take too much CPU time to find the eigensolutions, in applying direct computation methods, one has to resort to either some iterative methods which would yield exact eigensolutions (if fully converged), or other one-step methods to get approximate solutions which would be accurate enough for the analysis. The goals of those numerical methods are to reduce the core memory requirement and/or CPU time for computation, perhaps sacrificing accuracies in the numerical results, and/or obtaining only part of the eigensolutions. The latter point is not really a disadvantage in practical applications. Since the conformational changes usually accompany with low-frequency normal modes, only such eigensolutions are needed for the analysis.[2]

In this work, we present a modification of an existing iterative approach used in quantum chemistry and apply it for the above type of problem. In particular, we test it for a natural protein molecule, ubiquitin. Other iterative approaches include "effective Hamiltonian theory", "diagonalization in a mixed basis", and so on.[1-4] Similar large dimensional matrix eigenvalue problems appear as "configuration interaction problems" in quantum chemistry. Davidson and others, developed certain class of methods to yield several low-energy configuration interaction eigen-solutions some years ago.[5,6] In Davidson's approach, a small number of efficient basis vectors (which would hopefully span the target eigenvector) are generated or updated, and then the original matrix is re-expressed in this reduced basis set at each stage of iteration. The diagonalization of the resulting small-size matrix produces new approximate eigensolutions. If these solutions are not converged enough, the next step of iteration proceeds with new updated reduced basis vectors and values, combined with the current approximate solutions.

We represent the eigenvalue problem in matrix-vector terms as

$$KC = EC, \tag{1}$$

where K is the Hessian matrix, C is the eigenvector and E is the corresponding eigenvalue. We need certain initial trial solutions (say, C_0's and E_0's) to initiate the iteration process. Reasonable trial solutions may be obtained by partitioning the whole matrix into a certain block form and diagonalizing the diagonal blocks. The partitioning utilizes the structural characteristics of protein molecules. In this scheme, each diagonal block may correspond to a certain structural unit of the protein molecule like a group of amino acids in peptide sequence. The off-diagonal block elements tend to be small compared to the diagonal ones because the forces between the distant units should be smaller than otherwise. Indeed, several workers presented numerical methods which yield approximate low-frequency normal modes by keeping the low-frequency eigenvectors for each diagonal block and combining them to form a small-size reduced basis set for re-expressing the whole Hessian matrix. The subsequent diagonalization of this reduced Hessian matrix yields approximate normal modes and frequencies in low-frequency regime. Such approximate solutions may serve as initial trial solutions in the present work.[4,7-9] Even more simply, the free translation/rotation vectors for each unit may serve as basis vectors C_0 if combined for all units (the number of basis vectors would be 6 times the number of units). The subsequent diagonalization of this reduced matrix yields approximate solutions suitable for the initial trial solutions in the present iterative approach.

To continue the iteration, we divide the whole Hessian matrix into two parts as

$$K = D + V, \tag{2}$$

then the eigenvalue equation can be cast into

$$C = -(D - E)^{-1}VC, \tag{3}$$

where D and V are specified below. If we substitute C_0 and E_0 for C and E, respectively, on the right hand side of Eq.

(3), the resulting C on the left hand side would be a new vector unless the trial solutions are exact. The D in Eq. (2) may be conveniently chosen as block-diagonal form to facilitate the inversion operation in Eq. (3). In this case, V would be complemental matrix composed of off-diagonal blocks.

In the original configuration interaction application by Davidson, a single eigenvector is targeted and sought in one iteration series. He suggested to start with k trial vectors C_0's if the target state is the k-th from the bottom of the eigenvalue spectrum, and at each iteration step, add one newly generated vector prescribed by a more sophisticated equation related with Eq. (3). In the present work, we target many eigensolutions in one iteration series. First, starting with M trial solutions, we generate M new vectors by applying Eq. (3) for each trial solution. The resulting 2 M vectors are orthonormalized for convenience in representing the Hessian matrix and for eliminating any linearly dependent vector component. These new 2 M (or fewer based on the degree of linear dependence) basis vectors are used to represent the Hessian matrix, which we diagonalize to obtain the next updated M trial solutions. This procedure continues until the trial solutions are converged.

The test molecule, ubiquitin, is composed of 746 atoms and 76 residues and its conformation is energy-minimized using CHARMM force field.[10] The shape is similar to a bucket with the wall composed of an α-helix and a β-sheet as well as several connecting strands. At this minimum-energy conformation, the lowest normal mode frequencies are found to be 6.737 cm^{-1}, 8.595 cm^{-1}, 8.974 cm^{-1}, 9.958 cm^{-1}, and so on (six free translation/rotation eigenvectors have essentially zero frequencies), by the direct diagonalization of the whole Hessian matrix.

To set up the initial trial solutions, we partition the molecule into 14 structural units by grouping several residues, each containing between 47 and 65 atoms. The number of target low-frequency solutions are chosen as 100. Next, we combine 6 free translation/rotation vectors and additional 8 or 9 low-frequency eigenvectors for each diagonal block. The resulting 206 vectors are orthonormalized by Graham-Schmidt process before representing the whole Hessian matrix in this reduced basis set. (The same partitioning is used for D and V when applying Eq. (3) in later steps.) The reduced Hessian matrix $\tilde{K}$ is represented as

$$\tilde{K} = C^t \cdot K \cdot C, \tag{4}$$

where the dimension of $\tilde{K}$ is 206, and that of K is 2238, and C is the rectangular matrix composed of 206 orthonormal basis vectors arranged in columnwise. The diagonalization of $\tilde{K}$ yields 206 approximate eigensolutions. This finishes the initial trial solution setup step. By applying Eq. (3) for the lower-frequency 100 approximate eigenvectors and eigenvalues, mode by mode, (6 zero-frequency vectors excluded); we generate total 100 new vectors. They are augmented with 6 zero-frequency mode vectors and the previous lower-frequency 100 vectors. They serve as the 206 or fewer new reduced orthonormal basis vectors for Hessian matrix in the next step. We repeat the above procedures in iterative manner until convergence is achieved.

We compare the present iterative approach results against the direct diagonalization ones below (note that we ignore six zero-frequency solutions in comparison). The convergence of each eigenvalue is denoted by the magnitude of relative error ε as

$$\varepsilon = \left| \frac{\text{approximate value} - \text{exact value}}{\text{exact value}} \right|,$$

and for the corresponding eigenvector as

$$\gamma = 1 - \frac{C^t \cdot K \cdot C}{(C^t \cdot K^t \cdot K \cdot C)^{1/2}},$$

which was used by Perahia and Mouawad.[1] The γ value is between 0 and 1, and it should approach zero as the approximate eigenvector C converges to exact one. In Figure 1(a), we show the convergence behavior of frequencies during the iteration. It requires 5 iteration steps to achieve 5% ε error tolerance for the average error of the 100 target modes; 6 steps for the largest error; and 4 steps for the first non-zero frequency mode. All 100 modes are converged with $\varepsilon < 10^{-4}$ after 61 steps, by then the largest ε error is 0.97×10^{-4} for the 100th mode (whose frequency is 50.522 cm^{-1}). In Figure 1(b), we show the parallel

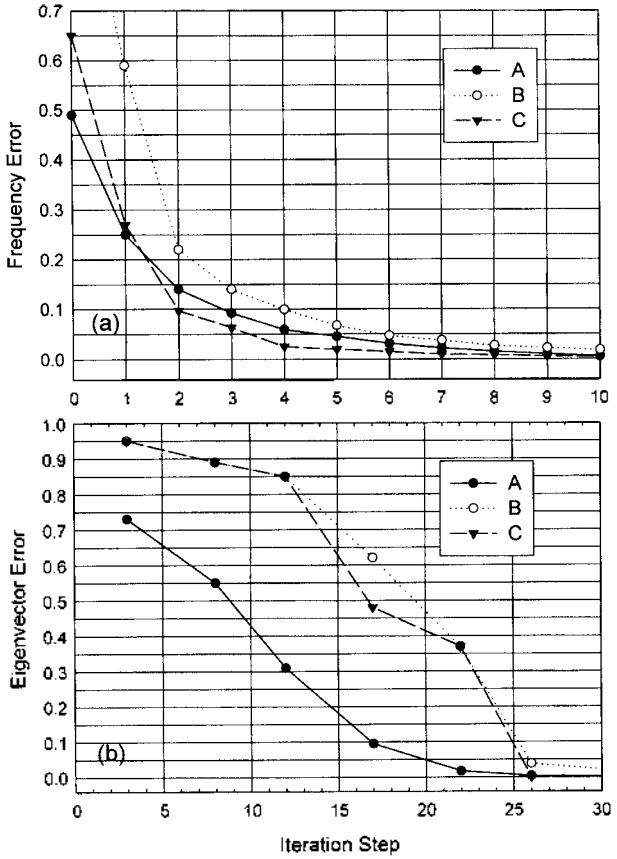


Figure 1. Convergence behavior of approximate eigensolutions during iteration. "A" denotes the average errors of all 100 target solutions, "B" for the largest error among the 100 target solutions, "C" for the lowest non-zero frequency mode. See text for more details. (a) Frequency errors ε for approximate normal mode frequencies as a function of iteration step. (b) Eigenvector errors γ for approximate normal mode vectors as a function of iteration step.

convergence behavior for normal mode vectors. As is well known, the eigenvectors are harder to be converged than the eigenvalues. It requires about 20 iteration steps to achieve 5% γ error tolerance for the average error of the 100 target modes; about 26 steps for the largest error; and about 26 steps for the first non-zero frequency mode. All 100 modes are converged with $\gamma < 3 \times 10^{-4}$, being the error of the 100th mode the largest after 61 steps.

As demonstrated above, the present iterative approach is able to yield exact normal mode solutions in an iterative manner through diagonalization of many small-size matrices (about one tenth of the whole matrix) in low-frequency regime, which would be useful in studying biological functions of protein molecules.

Acknowledgment. The author thanks Dr. Hon M. Chun, Prof. Martin Karplus, and Prof. John C. Light for helpful discussions and the University of Chicago for providing computing facilities.

References

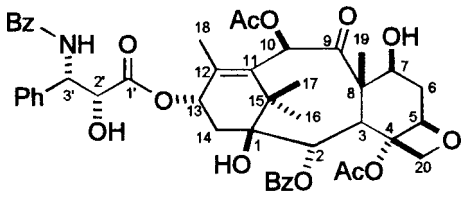
1. Perahia, D.; Mouawad, L. *Computers and Chemistry* **1995**, *19*, 241.
2. Durand, P.; Trinquier, G.; Sanejouand, Y.-H. *Biopolymers* **1994**, *34*, 759.
3. Durand, P. *Phys. Rev. A*, **1983**, *28*, 3184.
4. Mouawad, L.; Perahia, D. *Biopolymers* **1993**, *33*, 599.
5. Davidson, E. R. *J. Comput. Phys.* **1975**, *17*, 87.
6. Nesbet, R. K. *J. Chem. Phys.* **1965**, *43*, 311.
7. Jang, H. W. *Journal of Basic Science* (Sunchon National University) **1997**, *8*, 53.
8. Hao, M.-H.; Harvey, S. C. *Biopolymers* **1992**, *32*, 1393.
9. Hao, M.-H.; Scherage, A. *Biopolymers* **1994**, *34*, 321.
10. Brooks, B. M.; et al. *J. Comput. Chem.* **1983**, *4*, 187.

Synthesis and Biological Evaluation of C-2 Modified Taxol Analogs

Seok-Chan Kim*, Man-Sik Moon, Kwan-Min Choi, Sook-Jin Jun, Hyon-Kyong Kim, Dai-Il-Jung#, Kie-Seung Lee†, and Ki-Byung Chai†

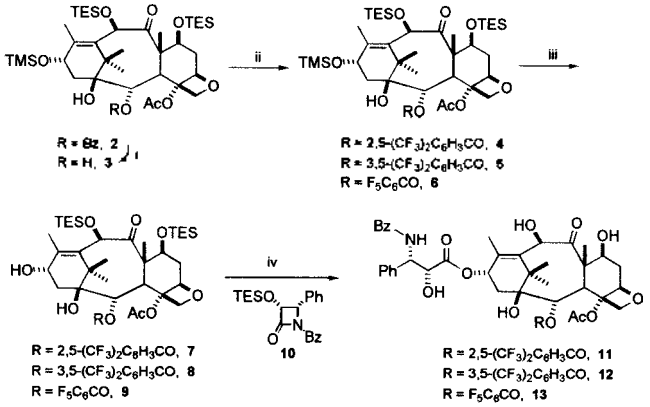
**Department of Chemistry, Kookmin University, Seoul 136-702, Korea*
#Department of Chemistry, Dong-A University, Pusan 604-714, Korea
†Department of Chemistry, Woosuk University, Chononbuk 565-701, Korea
‡Central Research Institute, Hanmi Pharm. Co., Ltd. Sampyong-dong, Boondang-ku, Kyonggi-do 463-400, Korea
Received July 7, 1998

Paclitaxel[1] (Taxol), **1** a complex diterpene isolated from the bark of the western yew[2] (*Taxus brevifolia*), is one of the most potent anticancer agent developed in the last decade for the treatment of ovarian and breast.[3] Of particular interest are structure-activity relationships associated with the highly functionalized terpenoid skeletal system. In general, the northern hemisphere functionalities C-7[4] and C-10[5] have little effect on biological activity, whereas functionalities to southern hemisphere C-2,[6] C-4,[7] and C-13 ester have a marked impact. Among these functionalities, we were interested in preparing new C-2 analogs and evaluating their activities.



Taxol, **1**

In a previous report,[8] we described a selective reduction of fully protected 10-deacetylbaccatin (III) **2** using Red-Al. We felt that intermediate **3** would be suitable for furnishing new C-2 ester analogs. In this communication, we outline the preparation of a series of C-2 modified new taxol analogs and their *in vitro* activity results. Attachment of new C-2 analogs has always been problematic because of easy formation of THF ring[9] between C-2 and C-20 in either basic or acidic media. Thus, upon treatment of diol **3** with 2 equivalents of LHMDS and substituted benzoyl chlorides at low temperature (−78 °C), new C-2 analogs **4**, **5**, and **6** were obtained. Partial desilylation afforded 7,10-TES baccatin derivatives **7**, **8**, and **9** which coupled[10] with optically active β-lactam **10** to give new C-2 ester analogs.



Reagents and Conditions : i) Red-Al, THF, 0 °C, (92.4%); ii) LHMDS, THF, -78 °C,then, 2,5-$(CF_3)_2C_6H_3COCl$, 4(87.5%), 3,5-$(CF_3)_2C_6H_3COCl$, 5(89.5%), F_5C_6COCl, 6(84.6%); iii) pyr., 48% HF, CH_3CN, 7(94.5%), 8(91.4%), 9(87.6%); iv) LHMDS, THF, -45 °C,10, then, pyr., 48% HF, CH_3CN, 11(80.1%), 12(78.5%), 13(72.7%)