A Coarse-Grained Normal Mode Approach for Macromolecules: An Efficient Implementation and Application to Ca²⁺-ATPase

Guohui Li and Qiang Cui

Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, Madison, Wisconsin 53706 USA

ABSTRACT   A block normal mode (BNM) algorithm, originally proposed by Tama et al., (Proteins Struct. Func. Genet. 41:1–7, 2000) was implemented into the simulation program CHARMM. The BNM approach projects the hessian matrix into local translation/rotation basis vectors and, therefore, dramatically reduces the size of the matrix involved in diagonalization. In the current work, by constructing the atomic hessian elements required in the projection operation on the fly, the memory requirement for the BNM approach has been significantly reduced from that of standard normal mode analysis and previous implementation of BNM. As a result, low frequency modes, which are of interest in large-scale conformational changes of large proteins or protein-nucleic acid complexes, can be readily obtained. Comparison of the BNM results with standard normal mode analysis for a number of small proteins and nucleic acids indicates that many properties dominated by low frequency motions are well reproduced by BNM; these include atomic fluctuations, the displacement covariance matrix, vibrational entropies, and involvement coefficients for conformational transitions. Preliminary application to a fairly large system, Ca²⁺-ATPase (994 residues), is described as an example. The structural flexibility of the cytoplasmic domains (especially domain N), correlated motions among residues on domain interfaces and displacement patterns for the trans-membrane helices observed in the BNM results are discussed in relation to the function of Ca²⁺-ATPase. The current implementation of the BNM approach has paved the way for developing efficient sampling algorithms with molecular dynamics or Monte Carlo for studying long-time scale dynamics of macromolecules.

INTRODUCTION

Normal mode analysis is a powerful approach for analyzing the structural and dynamical features of macromolecules (Brooks et al., 1988; Case, 1994). Although it is approximate because only the harmonic motion of the system around a single potential minimum is taken into account, qualitative estimates can be made for many properties (Brooks et al., 1995) of macromolecules such as the magnitude of atomic fluctuations, displacement covariance matrix, vibrational entropy (Tidor and Karplus, 1994), and neutron scattering spectra (Smith et al., 1990). Several recent studies have demonstrated that low-frequency normal modes are implicated in large-scale conformational transitions in allosteric proteins such as GroEL chaperonin (Ma and Karplus, 1998) and aspartate transcarbamylase (Thomas et al., 1996, 1999). The fact that these conformational changes occur typically on the time-scale of microseconds or longer and therefore are not accessible to standard molecular dynamics simulations makes the insights available from normal mode analysis very valuable. The low-frequency modes from normal mode analysis have also been used as basis vectors for approximate molecular dynamics simulations (Space et al., 1993; Zhang and Schlick, 1994) or refinement of x-ray (Kidera et al., 1992) or NMR data (Brüschweiler and Case, 1994). Moreover, quantum me-

chanical effects of nuclear motions on molecular properties (e.g., NMR order parameter (Brüschweiler, 1992)) can be included straightforwardly in the normal mode framework, whereas it is very difficult to do so with molecular dynamics (Voth, 1996; Makri, 1999).

The standard normal mode analysis requires the calculation of a mass-weighted second derivative matrix (hessian) followed by the diagonalization of that matrix. Accordingly, there are a number of bottlenecks associated with its application to large systems that contain more than ten thousand atoms. First of all, straightforward calculations of the hessian scale quadratically as the system size, although the scaling can be made linear using a cutoff scheme for nonbonded intermolecular interactions (Steinbach and Brooks, 1994). For long-range interactions such as electrostatics forces, it is possible, in principle, to extend techniques such as fast multipole algorithms (Greengard and Rokhlin, 1987) to make the evaluation of hessian scale favorably with the system size, although this has not been done to our knowledge. The second challenge concerns the storage of the hessian matrix, which formally scales as $(3N)^2/2$ (note that the hessian is a symmetric matrix), in which N is the number of atoms in the system. Thus, the size of the hessian quickly exceeds the memory of a typical workstation as the number of atoms increases, e.g., for a ten thousand-atom system, storage of the hessian requires more than 4 gigabytes of memory or disk-space. For large systems, however, the hessian is expected to be very sparse, especially when a finite cutoff is used for the intermolecular interactions. The storage problem, therefore, could be resolved by keeping only the nonzero matrix elements and the corresponding indices in the memory. The third challenge is related to the

Address reprint requests to Qiang Cui, Department of Chemistry and Theoretical Chemistry Institute, University of Wisconsin, Madison, 1101 University Avenue, Madison, WI 53706. Tel.: 608-262-9801; Fax: 608-262-4782; E-mail: cui@chem.wisc.edu.

diagonalization of the hessian, which scales as N^3 with standard diagonalization algorithms such as the Jacobi transform (Press et al., 1992). If only a small number of eigenvalues and eigenvectors are of interest, iterative diagonalization approaches with more favorable scaling such as block Lanczos algorithms can be used (Mouawad et al., 1993; Parlett, 1980; Marques and Sanejouand, 1995).

An alternative avenue for performing normal mode analysis for large molecules is to use a coarse-grained approach. This is particularly useful when only the low-frequency modes are of interest, such as in the analysis of allosteric effects in large proteins or protein complexes (Ma and Karplus, 1998; Thomas et al., 1996, 1999). This is because these low-frequency motions are typically delocalized throughout the system and involve mainly collective movements of residues. A promising approach along this line is the so-called block normal mode method (BNM, or as rotation translation basis) proposed by Sanejouand and co-workers (Tama et al., 2000). In this method, the system is divided into a number of blocks, and each block may contain one or more groups of atoms (such as an amino acid or a secondary structure element). The hessian matrix for the entire system is then projected into a subspace spanned by the translational and rotational motions of these blocks. Only the resulting projected hessian, which is much smaller in dimension than the full atomic hessian, has to be diagonalized. Therefore, the BNM method dramatically reduces the computational cost for the diagonalization step, which can be rate limiting for large systems. In the original implementation (Tama et al., 2000), the atomic hessian for the entire system is first constructed and then stored on disk for the subsequent projection, which is not practical for systems with more than ten thousand atoms due to large demands for disk space or memory. In the current work, we made a more efficient implementation of the BNM approach, in which the projected hessian is constructed directly and the required atomic hessian elements are computed on the fly and never have to be stored. This opens up the exciting possibility of carrying out (approximate) normal mode analysis for extremely large systems that include more than ten thousand atoms on standard workstations; coarse-grained molecular dynamics or Monte Carlo calculations that rely on normal mode eigenvectors can also benefit from the current implementation.

In the Method and Implementation section, we briefly review the algorithms of the BNM approach and describe our implementation. In the Test Calculations and Application to Ca^{2+}-ATPase section, we examine the quality of the eigenvalues and eigenvectors from BNM calculations by comparing the results with the standard normal mode analysis for a number of small proteins and nucleic acids. We then apply the BNM analysis to a fairly large system, Ca^{2+}-ATPase, which contains ~ 1000 residues, as an illustration. Finally, a few conclusions are summarized in the Conclusion section.

METHOD AND IMPLEMENTATION

In this section, we first briefly review the formulation of the BNM approach (Tama et al., 2000) and then describe our implementation in CHARMM (Brooks et al., 1983).

Methods

As described in previous work (Tama et al., 2000), a large system is first coarse-grained by partitioning the structure into n_b "blocks" (Fig. 1). Depending on the aim of the calculation (see Conformational Properties for Small Systems section), each block may correspond to one or several (protein/nucleic acid) residues or could be as large as a secondary structure element; the philosophy is rather similar to that adopted in the coarse-graining molecular dynamics approach, MB(O)ND (Chun et al., 2000). In the subsequent calculations, only the translational and rotational (T/R) degrees of freedom for the blocks will be considered, i.e., the internal structure of the blocks is fixed. This is likely to be an acceptable approximation if only properties that are related to the low-frequency modes of the macromolecule are of interest, because these modes are typically highly delocalized and involve little variations in the internal structure of blocks.

Once a partition scheme is chosen, a hessian matrix in the subspace (H^{sub}) spanned by the T/R basis vectors of the blocks can be constructed.

$$H^{sub} = P^T H P \qquad (1)$$

in which H is the standard mass-weighted atomic hessian matrix, and P is the projection matrix from the full atomic space ($3N_a$; N_a is the number of atoms) to the T/R subspace ($6n_b$). The P matrix is defined in terms of the infinitesimal T/R eigenvectors (Wilson et al., 1980) of the blocks

$$\begin{cases} P^\mu_{J,jv} = \sqrt{m_j/M_J}\ \delta_{\mu v}; & \mu = 1, 2, 3 \\ P^\mu_{J,jv} = \sum_{\alpha\beta} (I_J)^{-1/2}_{(\mu-3),\alpha} \sqrt{m_j}\ (r_j - r^0_J)_\beta \varepsilon_{\alpha\beta v}; & \mu = 4, 5, 6 \end{cases} \qquad (2)$$

in which J and j label blocks and atoms, respectively, and μ labels the translation ($\mu = 1, 2, 3$) and rotation ($\mu = 4, 5, 6$) of each block; M_J, I_J, and r^0_J is the sum of mass, moment of inertia, and center of mass for block J, respectively. If a single atom is treated as a block (e.g., calcium ions in Ca^{2+}-ATPase, see the Application to Ca^{2+}-ATPase section), only the translational vectors have to be included.

With the projection operation in Eq. 1, the matrix needs to be diagonalized in the normal mode analysis is reduced from $3N_a \times 3N_a$ to $6n_b \times 6n_b$. Thus, the memory requirement is reduced by a factor of $(N_a/2n_b)^2$, and the number of operations required in the standard diagonalization procedure is reduced by a factor of $(N_a/2n_b)^3$. Even with a low degree of coarse-graining, e.g., with each block containing one amino acid, this corresponds approximately to a reduction factor of 25 for storage and 125 for diagonalization (assuming that each block contains 10 atoms on average in an extended-atom representation of proteins (MacKerell, 2001)); more dramatic gains are expected for higher degrees of coarse graining.

The projected hessian, H^{sub}, is then diagonalized to generate the vibrational frequencies ($\frac{1}{2\pi}\Lambda^{1/2}$) and eigenvectors ($U^{sub}$).

$$[U^{sub}]^T H^{sub} [U^{sub}] = \Lambda \qquad (3)$$

The eigenvectors in the T/R subspace (U^{sub}) can be expanded back to the atomic space (U) using the transpose of the projection matrix, P,

$$U = P^T U^{sub} \qquad (4)$$

The structural and (approximate) dynamical properties of the macromolecules, such as root mean square (RMS) fluctuations and vibrational entropy
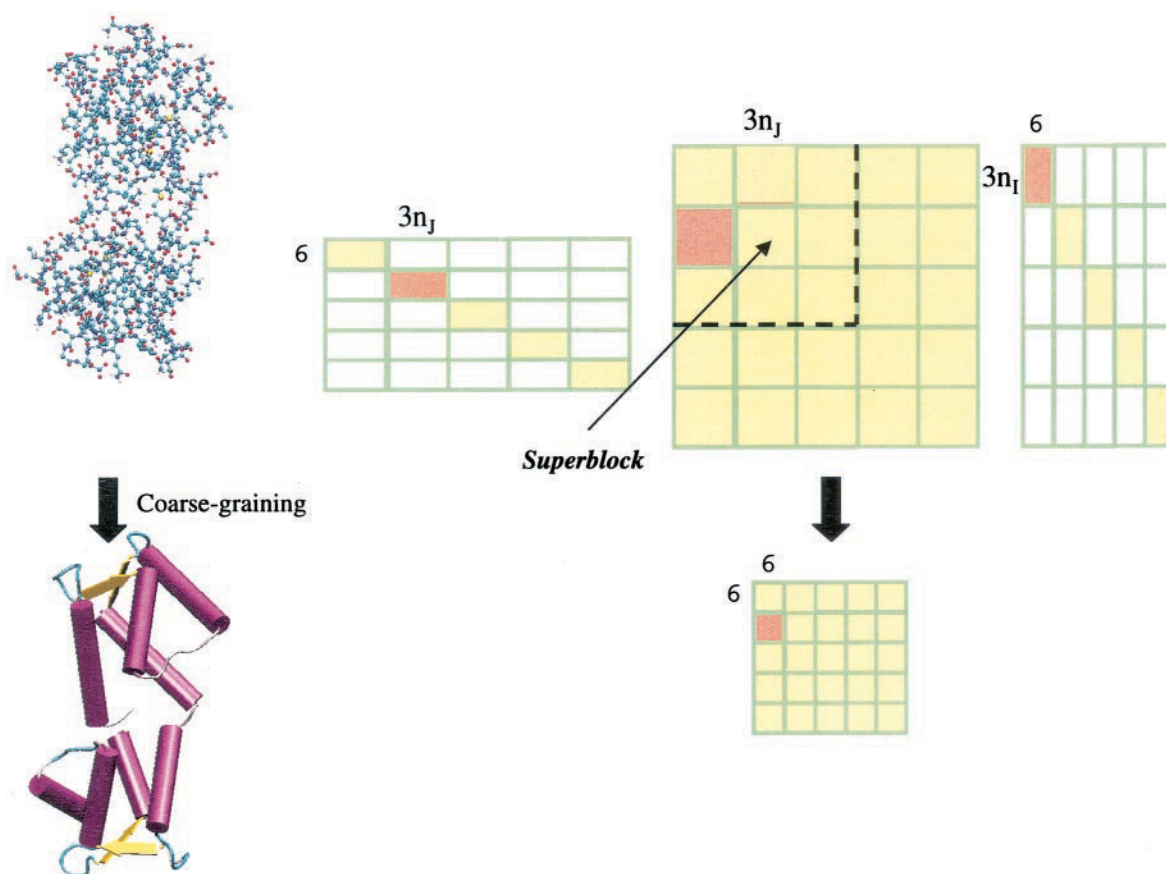
FIGURE 1 Illustrations for the basic idea of the BNM approach originally proposed by Sanejouand and co-workers (Tama et al., 2000) and the current implementation. The system is coarse grained from an all-atom representation into a collection of blocks, which can contain one or several residues. The full atomic hessian is then projected into the subspace spanned by the translation and rotation eigenvectors of the blocks, which leads to a projected hessian of a much smaller dimension. The current implementation uses "super blocks," with each super block consisting of several blocks; the atomic hessian elements for super blocks were first calculated on the fly, and then contracted with the projection matrix to obtain the projected hessian directly. In such a way, the full atomic hessian never has to be stored, and the algorithm is ideal for parallel computation.

can then be evaluated based on the eigenvalues and eigenvectors (see the Conformational Properties for Small Systems section). The eigenvalues (Λ) and eigenvectors (U) obtained in such a way are approximate eigenvalues and eigenvectors of the atomic hessian (H), and can be further improved in an iterative procedure in the effective Hamiltonian framework (Durand et al., 1994); this, however, has not been pursued in the current work.

It is interesting to briefly compare the BNM approach with other coarse-grained methods such as the "single parameter model" (Tirion, 1996; Bahar et al., 1997). In the latter model, a much simplified form of potential was found to produce atomic fluctuations in impressive agreements with x-ray data, which is consistent with the notion that the magnitude of fluctuation is highly correlated with the pattern of atomic packing (Halle, 2002). The advantage of the BNM approach is that it is based on a complete description of atomic interactions in the macromolecule, and therefore should provide more robust and realistic results for the structural and dynamical properties. The fact that the influence from the solvent is often neglected in reported BNM calculations is not a limitation of the BNM approach per se; the results can be improved by using implicit solvation models for the equilibrium effects (Lazaridis and Karplus, 2000; Roux and Simonson, 1999) and Langevian models for dynamical effects (Lamm and Szabo, 1986).

Implementation

In the original implementation of the BNM approach (Tama et al., 2000), the mass-weighted atomic hessian (H) is first constructed using CHARMM and then stored in memory or on disk. The projection (Eq. 1) is carried out next to obtain the hessian matrix in the T/R subspace (H^{sub}), which is finally diagonalized to obtain the eigenvalues and eigenvectors (Eqs. 2 and 3).

To improve the efficiency of the BNM approach such that it can be applied routinely to systems containing more than ten thousand or more atoms on regular work stations, we have made an implementation in which the atomic hessian matrix (H) is calculated on the fly and thus never has to be stored. This is made possible by a closer inspection of the projection matrix, P. Although it has a nominal dimension of $6n_b \times 3N_a$, it is a very sparse matrix. As shown in Eq. 2, it has only $6 \times 3n_J$ nonzero elements for the Jth block, in which n_J is the number of atoms in this block. Therefore, only a limited number of atomic hessian elements are required (see also Fig. 1) to obtain each element of the projected hessian (H^{sub}). In other words, a memory-efficient implementation of BNM is to construct the projected hessian directly, in which only the atomic hessian elements required for a given block-pair are computed on the fly, and they are discarded after contraction (Eq. 1) with the corresponding projection vectors (Eq. 2).

The efficiency of the current direct-approach can be further improved. For example, we note that each atom contributes to the projected hessian elements in more than one pair of blocks. In fact, atom k contributes to all the projected hessian elements for blocks that contain atoms within the cutoff distance from atom k. If one constructs the projected hessian by pairs of blocks, the same atomic second derivatives of energy have to be reevaluated many times because they are discarded after each block-pair of projected hessian is calculated. An efficient algorithm adopted in the current implementation is to consider "super-blocks," which simply contains a number of blocks (the number depends on the amount of available memory). The second derivatives of energy for all the atoms in each super block are constructed and then contracted with the projection matrix to obtain the contributions to the projected hessian elements (see Fig. 1), which avoids the repetitive evaluation of atomic second derivatives.

The implementation has been made to the VIBRAN module of CHARMM, and the ENERGY module involved in the atomic hessian evaluations has also been modified to maximize the efficiency for the construction of the projected hessian. The partition of system and the size of superblocks can be defined in a flexible manner by the user to fit the particular problem at hand. The code will be available in the future developmental version of CHARMM.

In short, the major contribution of the current work compared with the previous study of Tama and co-workers (Tama et al., 2000) is the direct implementation of the BNM approach. When the entire atomic hessian can fit into the computer memory, the current implementation does not provide a significant speedup than that of Tama et al., although a precise comparison is difficult because we do not have access to their implementation. For large systems where the entire hessian cannot be easily stored in memory, however, the current implementation makes a qualitative improvement and is required to make the normal mode analysis feasible. Compared with the standard normal mode approach, the speedup is highly significant even for small proteins. For instance, a full normal mode calculation took ~30 min for chemotaxis-Y protein (128 amino acids) on a 1.2-GHz Athlon, whereas a BNM calculation with one residue per block took less than a minute.

Future extensions

For systems that contain ten thousand atoms or so, the projected hessian is on the order of a hundred megabytes and therefore can be stored in memory for the subsequent diagonalization. For larger systems, such as the ribosome, the projected hessian itself can be extremely large if a low-level coarse-graining is done (from a practical point of view, a higher degree of coarse-graining such as secondary structure-based partitioning is appropriate for these large systems). In these cases, numerical techniques developed for standard normal mode analyses (such as DIMB) can be adopted. Moreover, we note that the BNM approach in the super-block form is ideally suited for parallel computations. The evaluation of projected hessian can be distributed to different nodes based on super-block combinations, and it is possible to optimize the algorithm such that the communication between different nodes is minimal. When symmetry is present in the system, such as in virus capsids, the symmetry-adapted approach can be applied to the BNM analysis (Simonson and Perahia, 1992; van Vlijmen and Karplus, 2001). Finally, we note that low-frequency modes form the basis of coarse-grained molecular dynamics approaches (Space et al., 1993) and x-ray or NMR structural refinement algorithms (Kidera et al., 1992; Brüschweiler and Case, 1994), efficient algorithms for generating such subspaces, like the current implementation of BNM, will be extremely useful in these context. These developments are currently underway and will be reported separately. By coupling BNM and numerical algorithms for sparse matrix diagonalization as well as parallel computing, meaningful normal mode or even molecular dynamics calculations can be carried out at a microscopic level for exceptionally large systems such as cellular components including the ribosome and microtubules.

TEST CALCULATIONS AND APPLICATION TO CA^{2+}-ATPASE

In this section, we first carry out BNM calculations for several small protein and nucleic acid systems, for which standard normal mode calculations can be readily done and compared with the BNM results. For properties (e.g., frequencies, eigenvector overlaps, and RMS fluctuations) that were examined in the original implementation (Tama et al., 2000), the comparison between BNM and standard normal mode (NM) results will only be discussed briefly, and analysis will be focused on other quantities that are also related to normal modes, such as vibrational entropy and displacement covariance matrix. Finally, we illustrate the BNM approach with the application to a fairly large system, the Ca^{2+}-ATPase, which contains 994 amino acids.

Conformational properties for small systems

To test the efficiency and accuracy of the BNM approach, we carried out calculations for two forms of calmodulin, which performs an important role in many signaling processes in many organisms (Crivici and Ikura, 1995, for their structures, see Fig. A5 in Supporting Material). The results were then compared with the standard (full) normal mode values. The calcium-free form is taken from a minimized average NMR structure (PDB code 1CFD (Kuboniwa et al., 1995)), and the calcium-bound form is taken from an x-ray structure at 1.7-Å resolution (PDB code 1CLL (Chattopadhyaya et al., 1992)). Because the major purpose here is to evaluate the performance of BNM approach in describing the intrinsic structural flexibility of calmodulin, no calcium ions were included in the calculation for either conformers. Selected calculations have also been carried out for a number of small proteins including crambin (PDB code 1CCN), chemotaxis-Y protein (PDB code 3CHY), and Human Immunodeficiency Virus protease (PDB code 1HHP), for which the results are summarized in Supporting Material. In addition, a nucleic-acid system was also briefly examined. A part of a ds-DNA, which is a segment in a DNA promoter sequence for T7-RNA polymerase (Von-Hippel et al., 1984; Cheetham and Steitz, 2000), was chosen. It contains 17 and 15 nucleotides, respectively; 13 nucleotides form the regular Watson-Crick base pairs, whereas the rest nucleotides were separated as part of the "transcription bubble" (Cheetham et al., 1999). This structure was chosen instead of a regular double-helical DNA due to the consideration that it has more structural variations and therefore serves as a better benchmark system for the BNM approach. The starting geometry was the T7-RNAP promoter complex (PDB code 1CEZ) from which the protein atoms were simply removed (see Fig. A3 in Supporting Material); we emphasize once again that the main purpose here is to compare the BNM approach with the regular NM analysis for vibrational calculations.

The proteins were described with the polar-hydrogen set of CHARMM parameters (Neria et al., 1996), and all calculations were carried out in vacuum. To approximately account for the solvation effect, the partial charges on the charged side-chains were scaled by an empirical factor of 0.3; in addition, the distance-dielectric constant was used (Northrup et al., 1980; Fisher et al., 1993). Such a combination was found to give a value of 1 to 3 kcal/mol for interactions between charged sidechains (e.g., salt bridges), which is close to the value in molecular dynamics simulations with explicit solvent (Simonson et al., 1997) and mutation experiments involving charged groups (Fersht, 1998). The electrostatic and van der Waals interactions were calculated using a cutoff scheme with an atom-based energy shift protocol at 8 Å (Steinbach and Brooks, 1994). A similar set-up was used for the DNA system, which was described with the CHARMM 27 force field for nucleic acids (MacKerell et al., 1995). Before the normal mode calculations, the structures were first energy minimized with 500 steps of steepest descent followed by adapted basis Newton-Rhapson until the RMS gradient is less than 10^{-6} kcal/(mol $\times$ Å); the typical RMS deviation from the x-ray structures is on the order of 1 Å. The BNM and standard NM calculations were then carried out on the minimized structures. We emphasize that in addition to the reduction in memory storage, the BNM approach is much faster than the standard NM calculations because the matrix that needs to be diagonalized becomes much smaller. For instance, a full normal mode calculation took ~28 min for camodulin (148 amino acids) on a 1.2-GHz Athlon, whereas a BNM calculation with one residue per block took less than a minute.

Vibrational frequencies and eigenvectors

Similar to previous benchmark calculations (Tama et al., 2000), the BNM frequencies correlate very well with the standard NM results for low-frequency modes. For example, with one residue per block, the BNM values correlate linearly ($R^2 \sim 0.997$) with the full NM results with a scaling factor of 1.8 for frequencies lower than 60 cm^{-1}. As the size of the block increases, the vibrational frequencies become consistently higher (Fig. 2 *a*); the scaling factor is 2.0 for two residues per block and 2.5 for three residues per block. With three residues per block, the BNM vibrational frequencies no longer scale linearly with the full NM values for frequencies higher than 20 cm^{-1}. As also found in the previous work (Tama et al., 2000), the scaling factors seem to be independent of the sequence and structure of the proteins (for more results for chemotaxis-Y protein, Human Immunodeficiency Virus protease, and crambin, see Supporting Materials), and thus can be taken as universal correction factors for BNM frequencies. This universality occurs presumably because a low-level coarse-graining (e.g., one residue per block) does not introduce effects that are sensitive to the overall structure of the macromolecule; rather,

the scaling-factors will depend mainly on the structural rigidity of each block. In this regard, it is interesting to note that for the promoter DNA strand (Fig. A3 in Supporting Material), the scaling factor for NM frequencies below 50 cm^{-1} is ~2.7 (Fig. A3b, $R^2 = 0.991$), which is much higher than the value of 1.8 found for proteins with one amino acid per block. This is consistent with the fact that the nucleotides on average have a larger number of atoms (~20) than amino acids (~10 with the extended atom representation).

To evaluate the quality of the BNM eigenvectors, we computed the maximal overlap for each standard normal mode with the BNM eigenvectors. For the few modes with the lowest frequencies, the maximal overlaps are larger than 0.8, indicating close agreements between the two sets of eigenvectors. As the frequency increases, however, the maximal overlap deviates quickly from unity and becomes as small as 0.2 even for modes with frequencies on the order of 30 cm^{-1}. Although the correspondence level between the BNM and full NM eigenvectors is not very high, it is sufficient in many applications (such as the subspace molecular dynamics method) that the low-frequency manifold is well described collectively. For that purpose, we projected the full NM eigenvectors onto the BNM eigenvectors, and computed the sum of the square of the expansion coefficients, which will be referred to as the "spanning coefficients" in the following discussions. The "spanning coefficients" reflect how well a subspace spanned by one set of eigenvectors is described by the other set. As shown in Fig. 2 *c*, for most modes with frequencies lower than 30 cm^{-1}, the "spanning coefficients" are larger than 0.8. This indicates that the low-frequency subspace is well described by the BNM eigenvectors, which paves the way for combining BNM and subspace molecular dynamics methods (Space et al., 1993). The qualitative behaviors for the overlaps and "spanning coefficients" are very similar for the several proteins and the promoter DNA studied here (for results, see Supporting Material).

RMS fluctuations and covariance matrix

Because the magnitude of atomic fluctuations is sensitive to the value of the vibrational frequencies ($\langle x_i^2 \rangle = \Sigma_j U_{ij} U_{ij} / \beta m_i \omega_j^2$), higher frequencies from BNM calculations imply reduced atomic fluctuations. However, we also realize that the major contributions of positional fluctuations originate from the low frequency modes, for which a nearly system-independent correction factor can be applied; it is 1.8 for proteins and 2.7 for nucleic acids according to the calculations above with one amino acid and one nucleotide assigned as a block, respectively. When this correction factor is taken into account, the BNM calculations gave RMS fluctuations very close to the full normal mode calculations, independent of the molecular systems (e.g., see Fig. 2 *d* and Fig. A4 in Supporting Material).

Moreover, the displacement covariance matrix is also fairly well reproduced by BNM calculations, especially with one
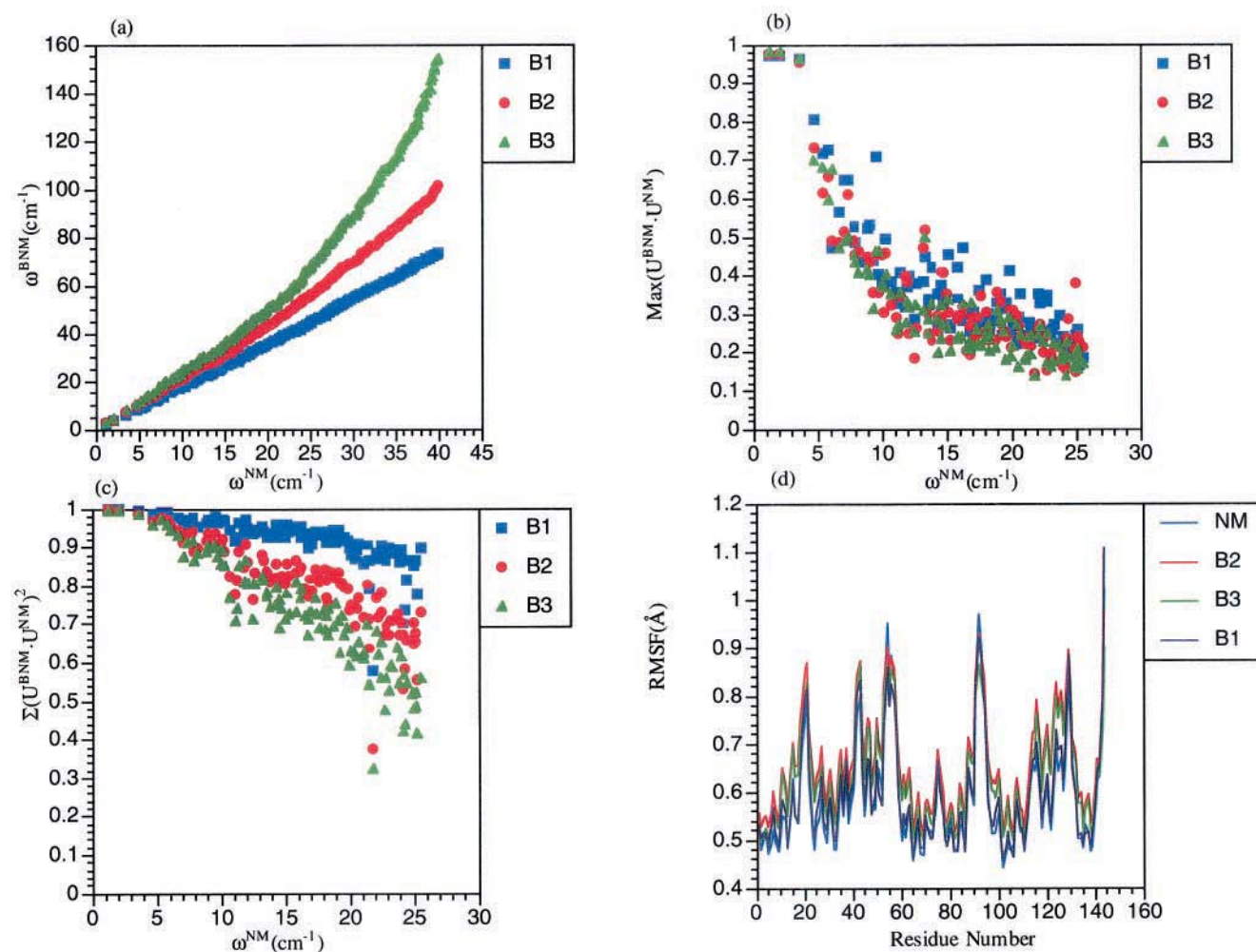
FIGURE 2   Comparisons of the vibrational frequencies, eigenvectors, and RMS fluctuations of Cα atoms from BNM and the standard NM calculations for the calcium-free form conformations of calmodulin (PDB code 1CFD) (for the corresponding results for the calcium-bound form, see Supporting Materials). (*a*) Vibrational frequencies (note the difference in scales for BNM and NM results); (*b*) maximal overlaps between an eigenvector from the NM analysis with BNM eigenvectors; (*c*) "spanning coefficients," which are defined as the sum of the squares of the expansion coefficients for an NM eigenvector in terms of BNM eigenvectors; (*d*) RMS fluctuation of Cα atoms. In each plot, results for each block containing one (B1), two (B2), and three (B3) amino acids are compared.

residue per block. Taking the results for the calcium-free form of calmodulin as an example, it is evident from Fig. 3, *a* and *b* that the two covariance matrices are quantitatively very similar. For example, both calculations predict substantial negative correlation between the residues around Pro-40 and these around Asp-90 and Glu-120. As the number of residues in each block increases, the overall correlation increases as one might expect. For example, long-range covariance elements (for Cα separated by more than 10 Å) on the order of 0.45 in the full NM calculations increase to ~0.6 to 0.7 with two or three residues per block (Fig. 3 *c*). The pattern in the covariance matrix, however, remains similar. As shown in Fig. 3 *d*, large covariance matrix elements occur between similar pair of residues in the full normal mode and BNM calculations, independent of block size.

Involvement coefficients and vibrational entropy

Normal mode calculation has been demonstrated to be of great value for characterizing large-scale conformational transitions in proteins (Ma and Karplus, 1998; Thomas et al., 1996, 1999). These motions are typically on the timescale of microseconds or longer and therefore are not directly accessible to the standard molecular dynamics simulations. Normal mode calculations are able to provide useful hints on the connection between the structural flexibility of proteins and their functional conformational transitions. Because this type of analysis concerns only the modes with very low frequencies, it is expected that BNM will give useful results. As an illustration, we briefly analyze the conformational transition between the two conformations of
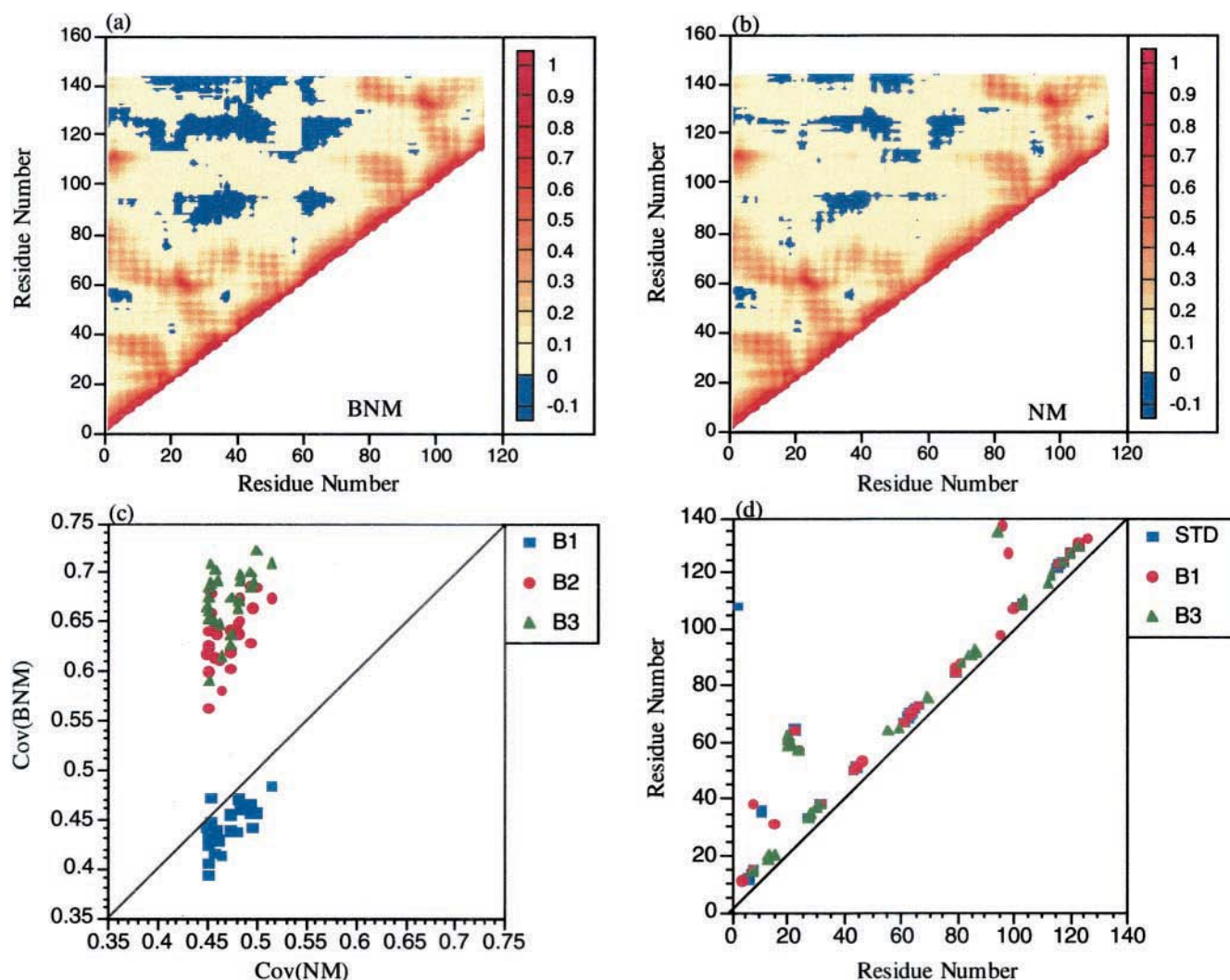
FIGURE 3   Comparisons of displacement covariance matrix from BNM and NM calculations for the Cα atoms in the calcium-bound form of calmodulin. (*a*) BNM result with one amino acid per block; (*b*) NM results; (*c*) correlation between BNM and NM results for covariance matrix elements larger than 0.45 in the full NM calculations; results for BNM calculations with one, two, and three amino acids per block are shown; note that as the number of residues included in each block increases, the covariance matrix elements become larger in magnitude, as expected. (*d*) Location of the 20 largest covariance matrix elements from NM, BNM (one or three residues per block). In *c* and *d*, we focus on long-range correlations, and thus selected points with the restriction that the Cα atoms are separated by 10 Å or further.

calmodulin, which is triggered in the cellular environment by calcium binding (Zhang et al., 1995; Finn et al., 1995; for more descriptions of the structures, see Supporting Materials). To quantify the evaluation of BNM results, involvement coefficients (*I.C.* in Fig. 4, Eq. 5a) and vibrational entropies (Eq. 5b) were calculated for these two conformations.

$$I_k = \frac{\bar{X}_1 - \bar{X}_2}{|\bar{X}_1 - \bar{X}_2|} \times U_k \qquad (5a)$$

$$S_v = \sum_i \left[\frac{\hbar \omega_i}{T(e^{\hbar \omega_i / k_B T} - 1)} - k_B \ln(1 - e^{-\hbar \omega_i / k_B T}) \right] \quad (5b)$$

in which $\bar{X}_1$ and $\bar{X}_2$ are the atomic positions for the two conformations

As shown in Fig. 4, *a* and *b*, only several low-frequency modes have large involvement coefficients in the standard NM calculations. This trend is similar with the eigenvectors for either conformation; there are four modes with involvement coefficient larger than 0.15 in 1CLL, and five in 1CFD. The result is in agreement with the previous observation that large-scale conformational transitions tend to be dominated by a few low frequency modes (Ma and Karplus, 1998; Thomas et al., 1996, 1999; Tama and Sanejouand, 2001). The involvement coefficients from the BNM calculations agree fairly well with the standard normal mode
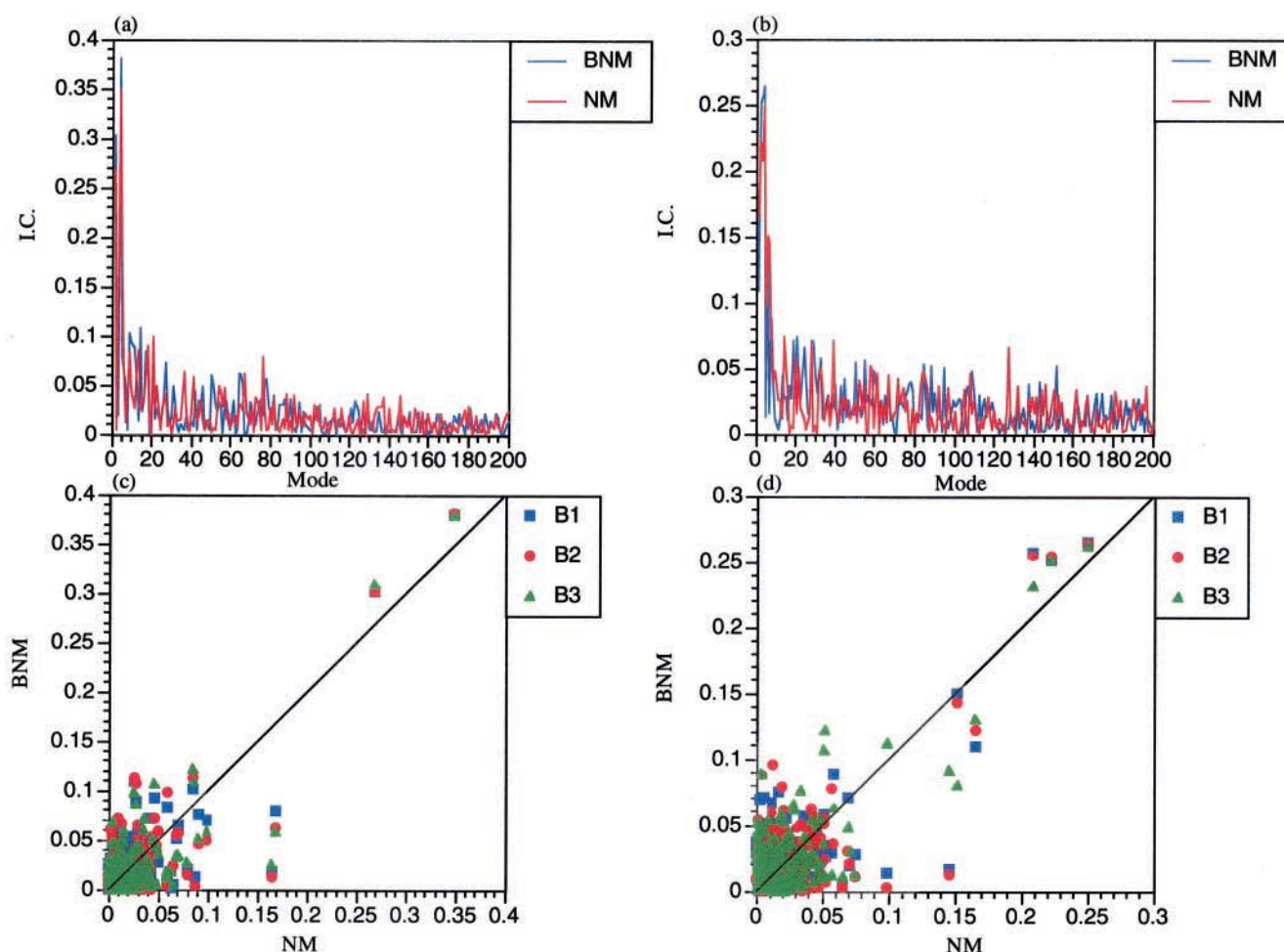
FIGURE 4   Comparisons of involvement coefficients (Involvement Coefficients and Vibrational Entropy section) from BNM and NM calculations for conformational transition between the two forms of calmodulin (with and without calcium ions bound; for their structures, see Supporting Materials). In *a* and *c*, eigenvectors for the calcium-bound form were used, whereas those for the calcium-free form were used in *b* and *d*. (*First row*) Involvement coefficients as a function of the mode index; (*second row*) correlation between the NM and BNM (one, two, or three residue(s) per block) results.

results, and the degree of agreement has a weak dependence on the block size (Fig. 4, *c* and *d*). This is consistent with the discussions in the previous subsection that the eigenvectors for the low-frequency modes are well described with the BNM approach.

As shown in Fig. A5 in the Supporting Material, the structure of the calcium-free form tends to be more flexible with the central helix broken into two pieces. As a result, full NM calculations predict that the vibrational entropy of the calcium-free from is substantially more positive; the $T\Delta S$ is 13.2 kcal/mol at 300 K. Because low frequency modes have the largest contributions to vibrational entropy, it is not surprising that BNM calculations give qualitatively correct results (6.6 kcal/mol), provided that the BNM frequencies are corrected by the scaling factor. Quantitatively, however, the BNM result is not satisfactory and only one-half of the entropic difference is reproduced. This is due to the fact that high frequency modes also make substantial

cumulative contributions. For instance, when the same number of modes as those in the BNM calculations are considered (864 instead of 4107 modes), the standard NM calculation gives a value of 7.8 kcal/mol, which is much closer to the BNM results.

The test calculations in this subsection demonstrated that BNM calculations can give vibrational frequencies, eigenvectors, as well as properties that are dominated by low-frequency modes (e.g., involvement coefficients) close to the standard (full) normal mode results, at a much lower computational cost. The agreement becomes less favorable when the number of residues included in each block increases, although many trends such as the pattern in the displacement covariance matrix and involvement coefficients remain satisfactory. For properties that include non-negligible contributions from the high-frequency modes, such as vibrational entropy, the BNM results can be only considered as qualitatively acceptable.
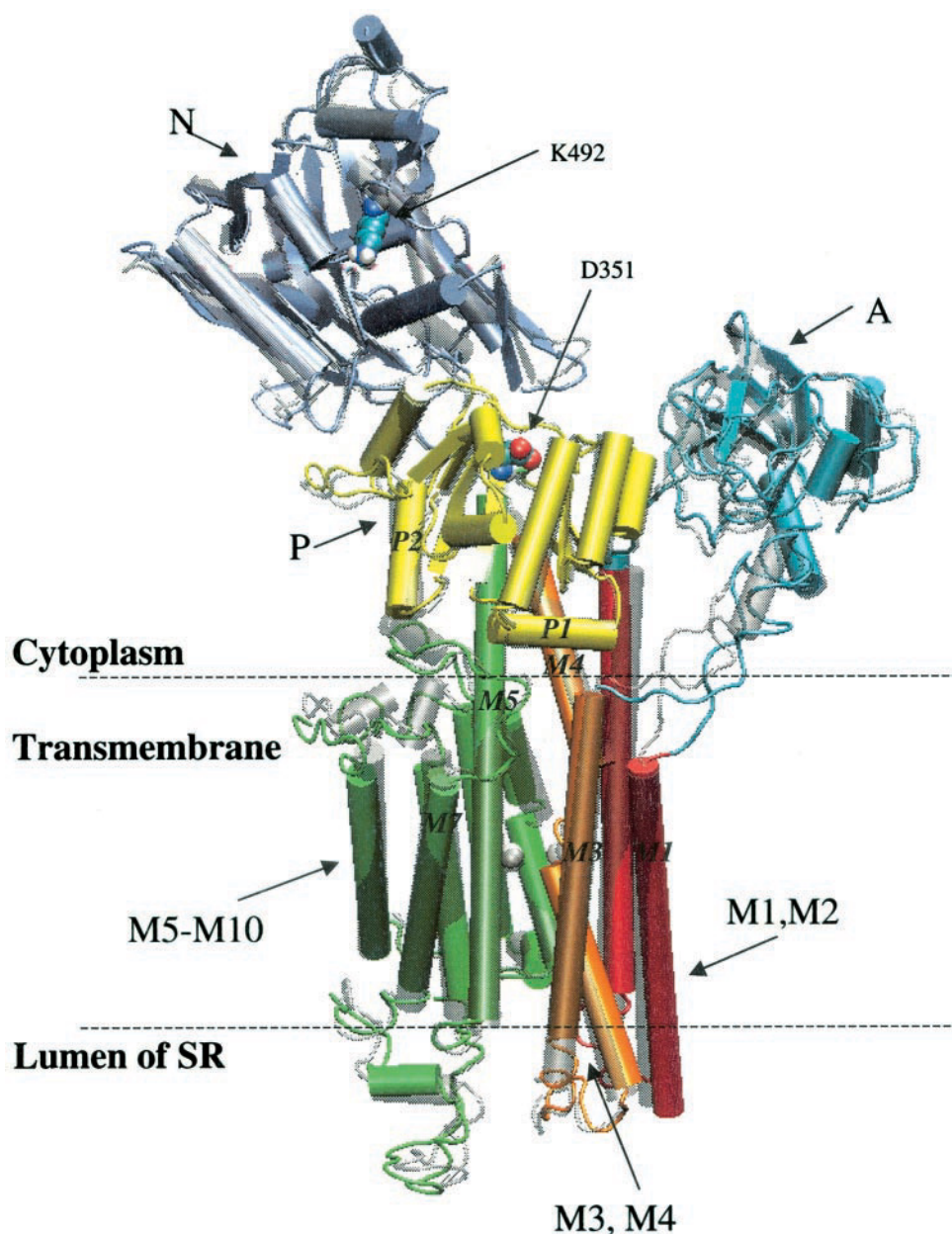
FIGURE 5 Overall structure of Ca^{2+}-ATPase according to the x-ray structure (Toyoshima et al., 2000). The cytoplasmic headpiece has three domains: A (activation), N (nucleotide binding), and P (phosphorylation), which are shown in light-blue, dark-blue, and yellow, respectively. The transmembrane region contain 10 helices, which are separated into three segments by sequence, which are shown in red (M1, M2), orange (M3, M4), and green (M5-M10). Two side-by-side calcium binding sites in the transmembrane region were found in the x-ray structure, which are shown as spheres. Superimposed on top of the x-ray structure as semitransparent gray is the structure obtained from energy minimization using CHARMM force field (see the Application to Ca^{2+}-ATPase section). Note that some coils become more ordered upon minimization and thus were identified as helices by the plotting program visual molecular dynamics (Humphrey et al., 1996).

Application to Ca^{2+}-ATPase

Ca^{2+}-ATPase, which is also known as the calcium pump (Møller et al. 1996), is a member of the P-type ATPase family (de Meis, 1981; Jencks, 1992). Its major function is to transport calcium ions across the membrane against a concentration gradient with the chemical energy provided by ATP binding and hydrolysis. Although much is known about its approximate working mechanisms through genetic (MacLennan et al., 1985; Aravind et al., 1998), biochemical (Jencks, 1989), mutagenesis (Maclennan et al., 1997), as well as electron microscopy studies (Toyoshima et al., 1993; Zhang et al., 1998), the possibilities of studying calcium transport at an atomic resolution (Lee and East,

2001) was made possible only recently by the report of a fairly high-resolution x-ray structure (2.6 Å) for sarcoplasmic reticulum Ca^{2+}-ATPase (Toyoshima et al., 2000). Consistent with previous electron microscopy results (Toyoshima et al., 1993; Zhang et al., 1998), the x-ray structure showed that Ca^{2+}-ATPase consists of a small transmembrane domain and a large cytoplasmic headpiece; the two are connected with a short stalk. As shown in Fig. 5, the transmembrane (M) domain contains 10 α-helices of different lengths, and the cytoplasmic part includes three separate domains termed A (activation), N (nucleotide binding), and P (phosphorylation). Domain N (residue 360–604) and P (residue 330–359 and residue 605–737) contains the ATP binding site and phosphorylation site (Asp-351), respectively, and domain A (residue 1–47 and residue 122–247) is somewhat more isolated from the rest. Two side-by-side calcium-binding sites occupied by two calcium ions in the middle of the transmembrane segment were found in the x-ray structure. One site is surround by helices M5, M6, and M8, whereas the other is on helix M4; the two sites are bridged by Asp-800 and a number of hydrogen bond networks that might be responsible for the co-operativity. No evidence for the existence of the lumenal binding sites was found, which is not inconsistent with the weak binding affinity expected for these sites (Jencks et al., 1993). Unlike the potassium channel (Doyle et al., 1998; Carbral et al., 2001; Zhou et al., 2001) or porin (Cowan, 1993), there is no obvious pathway leading from the cytoplasmic surface to the pair of calcium binding sites; one proposal involves helices M4 and M6 (Toyoshima et al., 2000), whereas the other concerns M1 (Lee and East, 2001). The release pathway for the calcium ions (or the pathway from the strong binding sites to the weaker lumenal binding sites) is also not clear from the x-ray structure, although M3 and M5 were thought to be involved (Toyoshima et al., 2000). Major conformational transitions, such as twisting of several transmembrane helices are expected to be involved in the calcium transport (Toyoshima et al., 2000; Lee and East, 2001), which is one of the reasons that the rate for Ca^{2+} transport is substantially lower than that for potassium transport through the potassium channel. A loop between the transmembrane and cytoplasmic segments, termed L67 (because it connects transmembrane helix 6 and 7), was proposed to be essential in coupling the calcium binding site to the phosphorylation as well as the ATP hydrolysis sites (Toyoshima et al., 2000).

To provide further insights into the structural-dynamical properties of the Ca^{2+}-ATPase (e.g., identifying intrinsic flexibility of different domains and correlated motions among them) and to illustrate the BNM approach for fairly large systems, we have calculated and analyzed low-frequency normal modes of Ca^{2+}-ATPase. The system contains 994 amino acids, which implies that a standard normal mode calculation would require more than 3 GB of memory just for the storage of the hessian matrix (without taking advantage of the sparsi-

ty). Although this amount of memory is accessible, the memory manipulation is already problematic with the standard CHARMM package on most platforms. An alternative is to use the iterative diagonalization procedure (e.g., DIMB in CHARMM), which does not have large memory requirements but takes many iterations to reach convergence. The BNM approach, by contrast, only requires 140 megabytes of memory for the storage of the projected hessian with one residue per block and does not require iteration for diagonalization. The entire BNM calculation for 5964 modes took less than 6 h on a 1.2-GHz Athlon machine.

The calculations were set up with a protocol similar to that used for the proteins in Conformation Properties for Small Systems section, i.e., the polar hydrogen set of CHARMM force field (Neria et al., 1996) was used to describe the protein and the two bound calcium ions, and charge-scaling plus distance-dependent dielectric constant were used to approximate the effect of the solvent. We realize that Ca^{2+}-ATPase is a membrane protein, and therefore should be arranged in a heterogeneous environment. The precise position of the protein in the membrane bilayer, however, has not been fully resolved. Considering the fact that lipid bilayer has a rather large fluidity, ignoring its effect on the normal modes may not be too problematic for the current purpose of analyzing the intrinsic structural properties of the Ca^{2+}-ATPase. The crystal structure (PDB code 1EUL) (Toyoshima et al., 2000) was first minimized in the presence of a set of gradually decreasing harmonic constraints, for the purpose of avoiding large structural distortions. The system was then subject to a few hundred steps of adapted basis Newton-Rhapson without any external constraints until the RMS gradient is lower than 10^{-2} kcal/(mol $\times$ Å); the final structure has a backbone RMS deviation of 1.3 Å compared with the x-ray structure. It is noted that several coils in the x-ray structure became somewhat more ordered upon energy minimization and was assigned to be short helices by visual molecular dynamics (Humphrey et al., 1996) in the minimized structure; this occurred at, for example, residue 230 to 240 in domain A and the lumenal end of transmembrane helix M2 (Fig. 5).

The BNM calculations were carried out at the minimized structure with one residue per block. The characters of a number of low-frequency modes were analyzed using the RMS atomic fluctuations ($\langle x_i^2 \rangle = U_{ik} U_{ik}/\beta m_i \omega_k^2$ for the kth mode), participation ratios, and visual inspection. The participation ratios have been used to characterize the degree of delocalization of the normal modes in liquid (Cho et al., 1994) and protein systems (Sagnella and Straub, 1999). Two quantities, R_k^I, R_k^{II} are defined for each normal mode, k; they are

$$R_k^I = \sum_{j}^{3N} U_{jk}^4 \tag{6a}$$

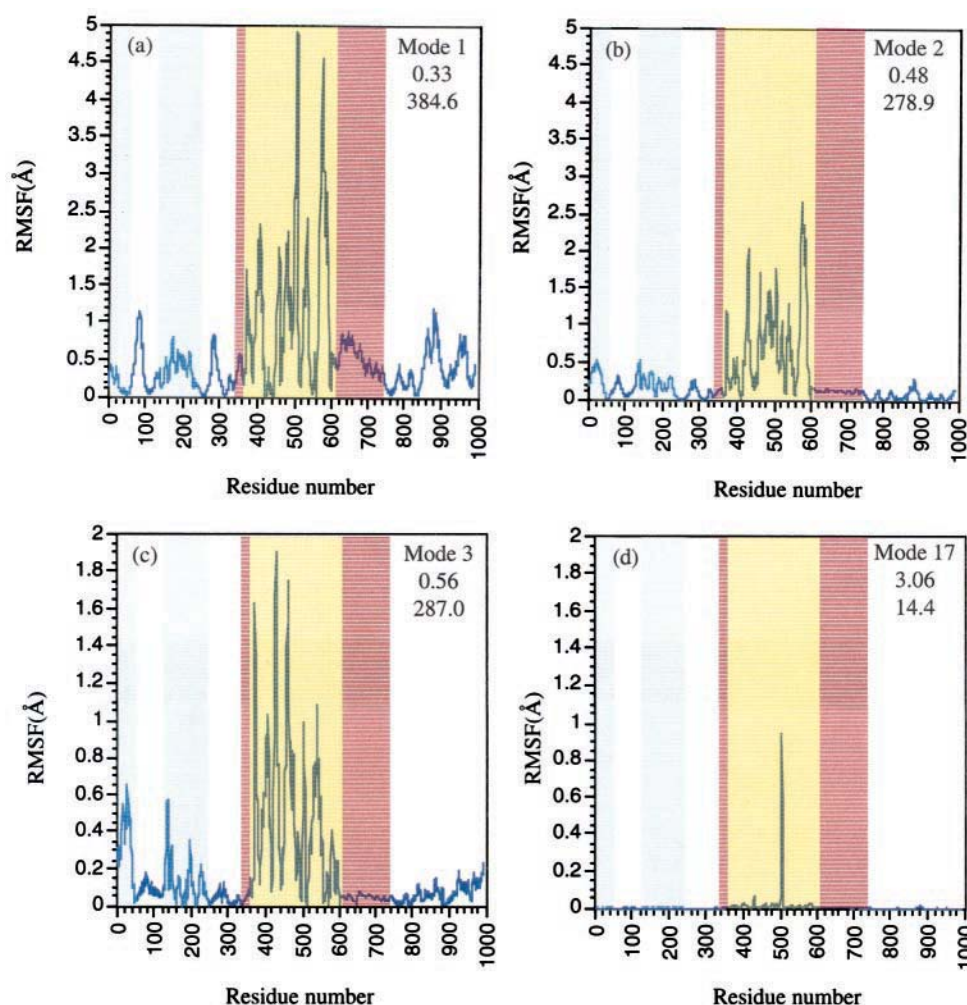$$R_k^{II} = \sum_{i}^{M_{res}} \left[\sum_{j}^{3N_i} U_{jk}^2 \right]^2 \tag{6b}$$

FIGURE 6   RMS fluctuations (Å) for the Cα atoms in Ca^{2+}-ATPase for a number of normal modes with large overall fluctuations. For each mode, the frequency and residue participation ratio (Eq. 6b) are also shown. Semitransparent color patterns were used to help identifying domains: light-blue, domain A; yellow, domain N; pink, domain P; white, transmembrane region and stalk.

in which N and N_i is the number of atoms in the whole system and in the ith residue, respectively; M_{res} is the number of residues in the system. The inverse of R_k^I indicates the number of atoms involved in the kth (generalized) normal mode, and the inverse of R_k^{II} corresponds approximately to the number of residues involved.

The first 20 modes are all below 4 cm^{-1} (after divided by the scaling factor of 1.8 found in previous work (Tama et al., 2000) and above discussions). Although most of them are very much delocalized with $1/R^{II}$ substantially larger than 200, a number of modes (e.g., mode 17) are rather localized and involve only tens of residues (Fig. 6 and Fig. A6 in Supporting Material). There is no obvious correlation between the magnitude of delocalization and the magnitude of fluctuations (Fig. A6, A7 in Supporting Material). For example, whereas the three lowest modes have very large fluctuations and involve on the order of 300 residues (mainly the N domain), a number of higher frequency modes

(e.g., mode 7 and 18) are more delocalized but have substantially smaller fluctuations. Among the four major structural domains (N, P, A, and M), the N domain seems to be the most flexible; all three lowest-frequency modes show large fluctuations on the order of 2 Å or more (Fig. 6); comparison of the two end structures along the normal coordinate of these modes with displacements correspond to the thermal energy at 300 K indicated translation and rotation of the N domain of 1 to 2 Å and 8° to 11°, respectively. Other domains have RMS fluctuations of 1 Å or so in the three lowest frequency modes, which correspond to translation and rotation of ~1 Å and several degrees. The x-ray B-factors (Toyoshima et al., 2000) suggest that both N and A domains are very flexible (also see below). The fact that the A domain is less flexible compared with the N domain in the current work might be related to the above observation that some coils become more ordered and thus more rigid in the minimized structure; whether this is due to the
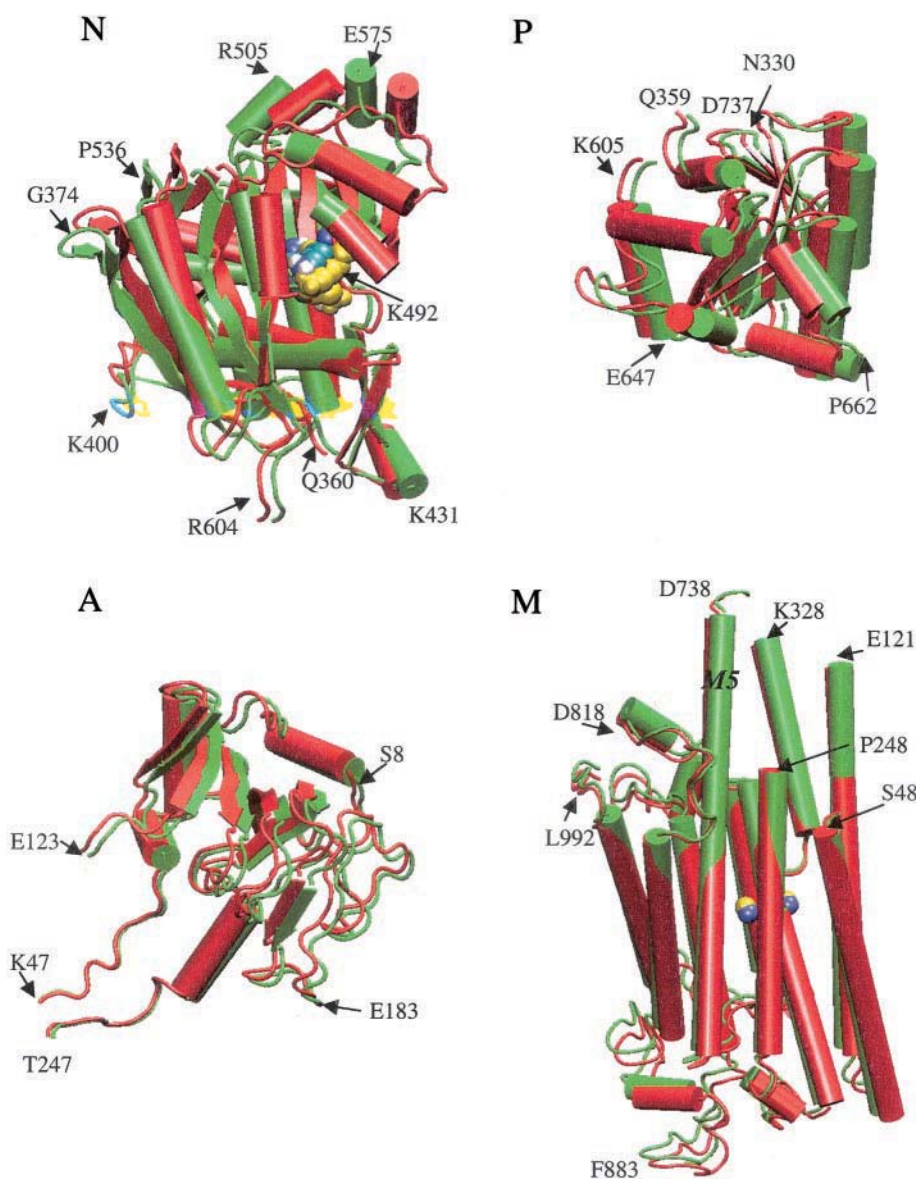
FIGURE 7 Displacements for different domains of Ca^{2+}-ATPase in the lowest frequency mode ($\omega = 0.33$ cm^{-1}). See Fig. 6 for the format of the structures. Note that the displacement of domain N, which is approximately a rotation about the N/P interface, is the most significant.

neglect of the solvent environment in the calculations is being investigated with a molecular dynamics simulation with an atomic description of the solvent and lipid bilayer (G. Li and Q. Cui, work in progress).

Among the modes examined, most involve concurrent movements of different domains according to the corresponding atomic fluctuations (Fig. 6 and Fig. A7 in Supporting Material). For example, mode 1 (Fig. 6 *a*) involves significant fluctuations in all domains, although those for domain N are substantially larger in magnitude. Both mode 2 and 3 (Fig. 6, *b* and *c*) involve domains N and A, whereas mode 4, 5, and 6 (Fig. A7 in Supporting Material) involve significant fluctua-

tions of domain A, M, and N. These reflect coupled motions among different domains (see below). For modes with higher frequencies, the fluctuations tend to be more localized to a certain domain or a small number of flexible residues in a domain, e.g., the coils in the lumenal side of the structure, such as residues F78 to T84 and E878 to C888, have large fluctuations in a number of modes (e.g., mode 1, 4, and 15) with frequencies between 1 and 4 cm^{-1}. Although fluctuations are generally larger for modes with lower frequencies (compare Fig. A6 and A7 in Supporting Material), a higher frequency mode can have large but localized fluctuations, e.g., mode 17 ($\omega = 3.06$ cm^{-1}) involves large RMS fluctuations on the order

of 1 Å for residue 502 to 506 in the nucleotide binding domain N (Fig. 6 *d*).

To further characterize the connection between structural flexibility and functions of Ca^{2+}-ATPase, we focus on the displacements involved in the lowest frequency mode ($\omega = 0.33$ cm^{-1}), which has considerably larger fluctuations (Fig. 6 *a*) compared with others. First, the two end structures along the normal coordinate of this mode with displacements corresponding to the thermal energy at 300 K are superimposed for different domains in Fig. 7. The overall motion of domain N can be qualitatively viewed as a rotation (~11°) about the two end-regions (Q360 and R604) as hinges, as reflected by the small RMS fluctuations (*a*; this region contains the invariant DPPR motif (Møller et al., 1996). The largest displacements occur in several segments including residues 568 to 587, 478 to 511, and 532 to 535; note that the list includes several residues in the ATP binding pocket, such as Lys-492 and Phe-487. The rotational pattern of domain N found in the current normal mode analysis is consistent with experiment (Toyoshima et al., 2000), which compared the high-resolution (2.6 Å) x-ray structure with bound calcium to the low-resolution density map (8 Å) for Ca^{2+}-ATPase without calcium but in the presence of decavanadate and dansyl-thapsigargin. It was found that to better fit the transmembrane domain in the two structures, the cytoplasmic headpiece has to be substantially reoriented in which domain N rotates by 20° to partially close the gap with the P domain (see below) (Toyoshima et al., 2000).

The fluctuations in the P domain are smaller in magnitude but more uniform (Fig. 6 *a*); the displacements involve significant translational motion (~1.3 Å) nearly perpendicular to the transmembrane domain (Fig. 7 *b*). The moving pattern of domain P is consistent with recent fluorescence energy transfer measurements (Baker et al., 1994) between certain residues on Ca^{2+}-ATPase and phospholipids, which showed that the position of these protein residues (Cys-344, Cys-670, Cys-674) relative to the membrane changed little upon calcium binding.

For domain A, the end of the loops that are connected to the transmembrane helices M1 and M3 (especially around K47 and T247) have rather small fluctuations, which suggests that they may serve as hinge points for the rotation inferred from comparing the Ca^{2+}-bound x-ray structure and the decavanadate bound low-density map (Toyoshima et al., 2000). Although a significant rotation (~90°) was found in that comparison, no tendency for large-scale rotation was observed in the current normal mode analysis; the two end structures for thermally excited first mode are rotated relative to each other by only 3° (Fig. 7 *c*). Superimposing several lowest normal modes, in particular mode 4, resulted in somewhat larger rotational motions of domain A. This difference between BNM calculation and experimental observations does not necessarily pose any conflict, because it is possible that domain A only undergoes large
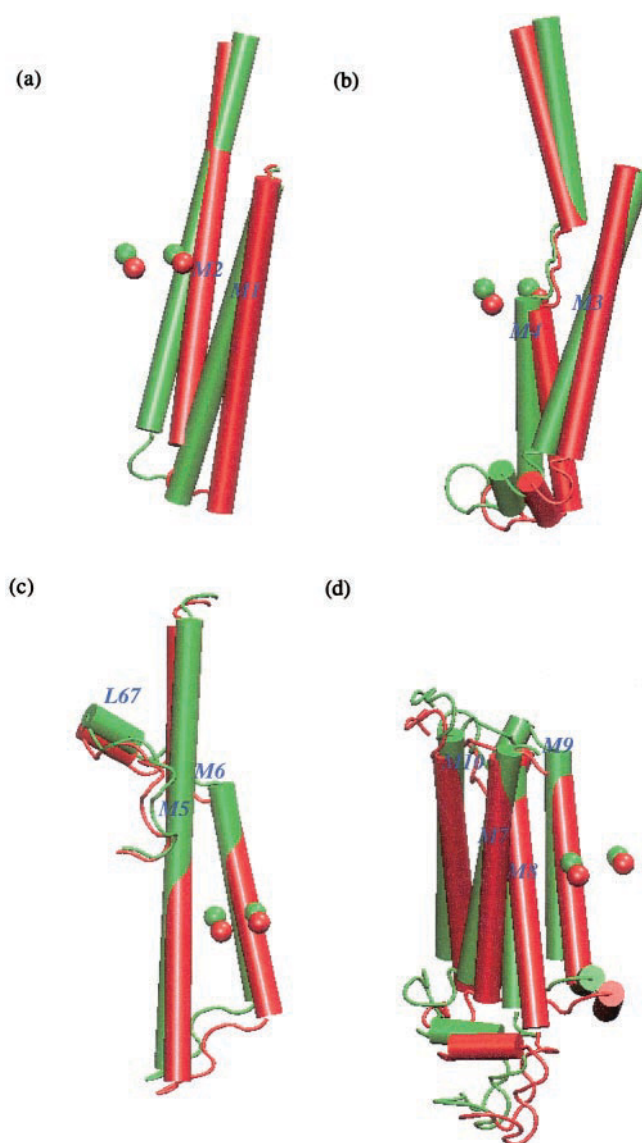

FIGURE 8 Similar to Fig. 7 *d*, i.e., the displacements of different helical regions in the transmembrane domain in the first normal mode. To better illustrate the displacements, a temperature of 1500 K was used rather than 300 K (which was used in all other figures) in generating the two end structures. The calcium ions are also shown in spheres.

scale rotation upon calcium binding and becomes less mobile afterwards.

For the transmembrane domain, large fluctuations occur mainly at the coils and loops in the lumenal region and the C-terminal region (Fig. 7 *d*). The helices are fairly rigid, as one might expect, and there are twisting motions in their relative arrangements. To illustrate the motions more clearly, end structures corresponding to thermal energy at 1500 K rather than 300 K are shown for different helical regions of the transmembrane domain in Fig. 8. It is interesting that the calcium binding residues, such as Asn-768, Glu-771, Thr-799, and Asp-800, all undergo relatively
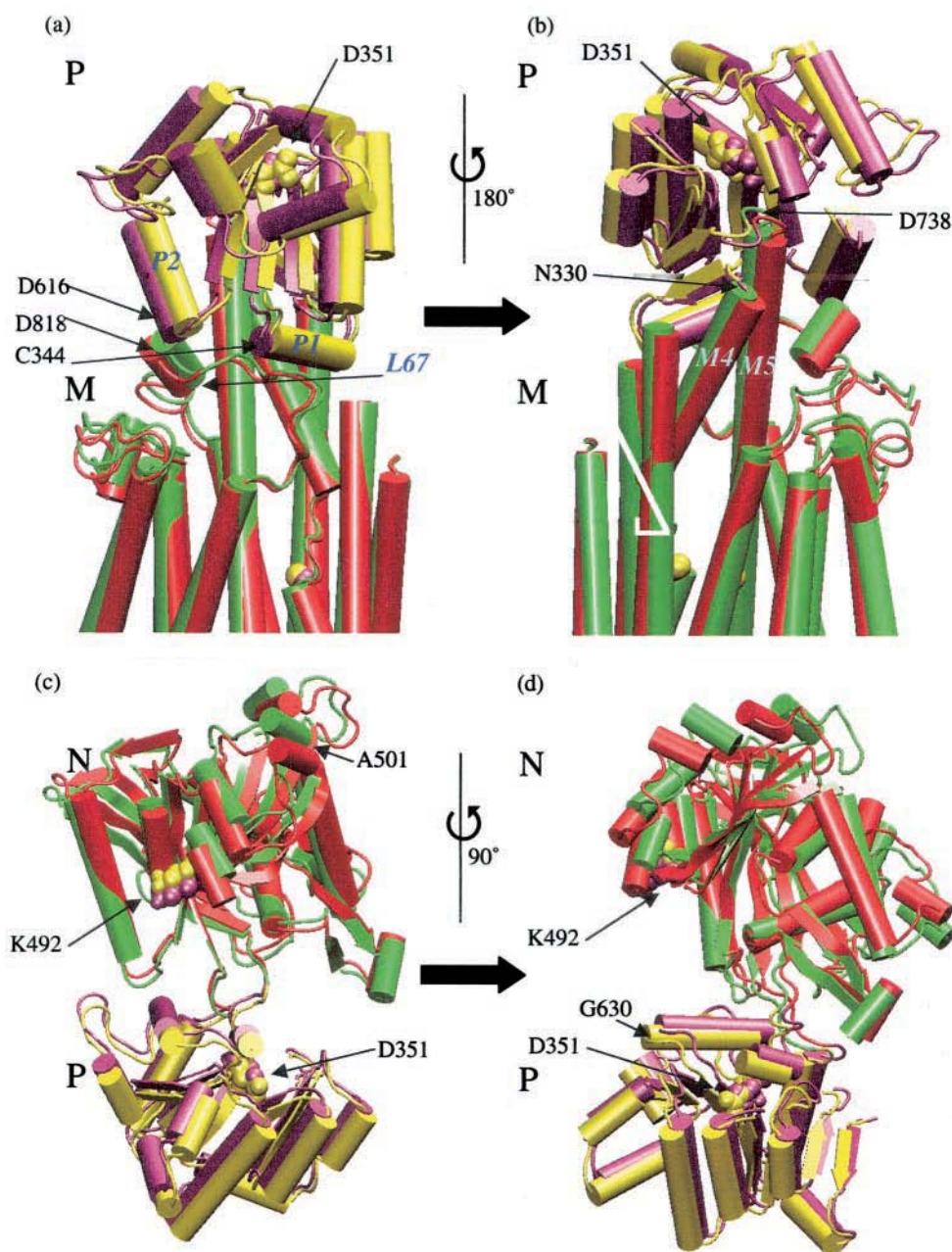
FIGURE 9 Similar to Fig. 7 but with more details on the P/M and N/P interface to reveal possible communication pathways among different domains reflected in the lowest frequency mode ($w = 0.33 \text{ cm}^{-1}$). Green and yellow color corresponds to one structure, whereas red and purple corresponds to the other. For the P/M interface, the orientations in *a* and *b* are related by a rotation of approximately 180° perpendicular to the transmembrane domain; for the P/N interface, the two orientations in *c* and *d* are related by a rotation ~90° perpendicular to the transmembrane domain.

small magnitude of fluctuations, which presumably can be related to their stability required for maintaining binding selectivity (Bernèche and Roux, 2001; Shen et al., 2002). Unlike other channels (Doyle et al., 1998; Carbral et al., 2001; Zhou et al., 2001; Bernèche and Roux, 2001), the static x-ray structure of Ca^{2+}-ATPase does not reveal any obvious transport or gating pathway. The normal mode results in Fig. 8, by contrast, showed clear tendency for the

transmembrane helices to "twist open," especially in regions between the calcium binding site to the lumenal side. The most striking examples include the lower M4 helix (Fig. 8 *b*), M2 and M8; displacements in M5 and M6 are relatively modest. Therefore, we expect that displacements of helix M2, M4, and M8 are essential for the release of calcium ions from the binding sites to lumen. We note that the magnitudes of movements shown in Fig. 8 are exagger-

ated by the higher temperature (1500 K) used in generating the structures, the intrinsic pattern in displacements, however, should remain valid. We also note that the current structure is in the transport-inactive form (E1′Ca$_2$ state in the overall transport scheme) (Lee and East, 2001; de Meis, 1981), and ATP binding, hydrolysis, and phosphorylation are expected to play an essential role in driving the system to the conformation that favors calcium transport (e.g., displacements of M2, M4, and M6). As to the calcium binding from cytoplasm, two pathways have been proposed; one proposal involves M4 and M6 (Toyoshima et al., 2000), whereas the other involves M1 (Lee and East, 2001). According to the normal mode results in Fig. 8, the N terminal of M1 has rather small displacements, whereas the upper parts of both M2 (Fig. 8 *a*) and M4 (Fig. 8 *b*) move significantly in the first mode. In addition, M4 contains the interesting titritable residue, Glu-309, which is in the coil region connecting the two helices in M4 and is close to the calcium binding sites; we note that calcium binding is regulated by pH (Henderson et al., 1994), which is consistent with the involvement of an ionizable group. Moreover, although Glu-58 in M1 is close to the calcium ions, mutation studies (Clarke et al., 1989) found little effect on calcium binding when Glu-58 and Asp-59 were mutated to an Ala. Thus, it seems that M4 is more likely to be involved in a gating motion that regulates the binding of calcium ions from cytoplasm to the transmembrane domain.

Finally, we examine the change in the interface between different domains in the lowest frequency mode. As shown in Fig. 9, *a* and *b*, the motion of domain P and M are correlated at many positions. First, helix P2 is in close contact with a segment (I816 to D818) in the loop (L67) connecting M6 and M7, and they move in a highly correlated fashion; otherwise, they would have collided into each other (color-coded purple and green in Fig. 9). Helix P1 is also close to L67, although the magnitude of motion seems to be smaller (Fig. 9). Because the displacement of L67 will alter the orientations of M6 and M7, which are involved in calcium binding, the interaction between P1, P2, and L67 is essential to the long-range (~50 Å) communication between the calcium binding site and phosphorylation that occurs in the P domain (e.g., Arg-604 on P2 is hydrogen-bonded to Gly-354 on the phosphorylation loop). Moreover, Fig. 9 *b* shows that residues connecting domain P to M4 and M5, e.g., residues around N330 and D738, have substantial correlated displacements. Because M4 and M5 include calcium-binding sites, these residues on the domain interface form another communication pathway between phosphorylation and calcium binding.

The interface between domain N and P, by contrast, changes very little in the normal mode, which is consistent with the fluorescent energy transfer measurements (Baker et al., 1994). Nevertheless, the rotational motion of the N domain and the translation character of the P domain bring closer the ATP binding site in the former and the phosphor-

ylation site in the latter by approximately 1 Å when the lowest frequency mode is thermally excited at 300 K. Because they are 25 Å apart in the x-ray structure, but have to be spatially close in the active form for the phosphorylation to occur, the structural flexibility revealed by the normal mode analysis is essential for the proper functioning of Ca^{2+}-ATPase. Even in the presence of the ATP analogue, 2′,3′-*O*-(2,4,6-trinitrophenyl)-AMP, the relative positions of domain N and P is quite similar to the nucleotide-free form (Toyoshima et al., 2000), which indicates that the presence of the γ-phosphate is essential for driving these two domains together; this is consistent with the results from the current normal mode analysis based on the nucleotide-free structure, which showed only small relative closure of the two domains.

CONCLUSION

Normal mode analysis is a useful approach for gaining insights into the structural and dynamical properties of macromolecules. The application of normal mode analysis to large systems that contain more than ten thousand atoms has been difficult due to the demands of hessian storage and matrix diagonalization. Although several numerical schemes have been developed in previous works to circumvent these bottlenecks, permitting interesting studies on several large systems such as aspartate transcarbamylase (Thomas et al., 1996, 1999) and citrate synthase (Mouawad and Perahia, 1993), physically motivated approximate models are also useful. When only the low-frequency modes, which are useful in characterizing large-scale cooperative motions in large systems, need to be determined, coarse-grained models can be used because these modes typically involve delocalized vibrations. Based on previous work of Sanejouand and co-workers (Tama et al., 2000), we have made an efficient implementation of one such method, the BNM approach, into the simulation package CHARMM. In BNM (Tama et al., 2000), the system is coarse-grained into a collection of blocks that each contains one or more residues, and the full atomic hessian is then projected into the subspace spanned by the translation and rotation of blocks. As a result, the matrix that needs to be diagonalized becomes much smaller, which can lead to a dramatic reduction in the computational cost. Compared with other methods developed for low-frequency modes (Tirion, 1996; Bahar et al., 1997), the BNM approach has the advantage that a full description of atomic interactions (potentially the effect of solvents with implicit solvent models) is included and thus should be more robust in general. In the current work, by taking advantage of the sparseness of the projection matrix, the block hessian is constructed in a direct manner, in which the required atomic hessian elements are only calculated on the fly and never have to be stored. As a result, the memory requirement for the BNM approach becomes much reduced and can be readily applied to large systems that contain

more than ten thousand atoms on regular workstations or personal computers. Coupled with more efficient numerical techniques for sparse matrices and parallel computation, extremely large systems such as the ribosome, can potentially be analyzed with BNM. Moreover, the BNM approach can be coupled with molecular dynamics or Monte Carlo methods that focus on the low-frequency subspace of macromolecules to develop efficient sampling techniques for studying long-time scale processes such as large-scale conformational transitions of molecular motors.

After the implementation, the BNM approach was tested with a number of small protein and nucleic acid systems. Properties such as vibrational frequencies, normal mode eigenvectors, RMS fluctuations, displacement covariance matrix, involvement coefficients, and vibrational entropy have been determined and compared with the standard normal mode results. It was found that the BNM results are quantitatively satisfactory for most properties that involve mainly low frequency modes, whereas the agreement with standard normal mode analysis can be only considered as qualitative for properties that have substantial contributions from high frequency modes (such as vibrational entropy).

As an illustration of the BNM approach for large systems, we have applied it to study the low-frequency modes of Ca^{2+}-ATPase, which contains 994 residues. Instead of requiring more than 3 gigabytes of memory, the BNM calculation with one residue per block (for 5964 modes) only used several hundred megabytes of memory and took less than 6 h on an 1.2-GHz Athlon. The BNM analysis revealed the structural flexibility of the cytoplasmic domains, among which domain N seems to be most mobile. Hinge regions were identified for the rotational type of motions of domain N and A, which are on the interface between N/P and A/M domains, respectively. Correlated motions were found between two helices in domain P (P1 and P2) and the loop (L67) connecting transmembrane helices M6 and M7, as well as among residues on the interface between domain P and transmembrane helices M4 and M5. These correlated motions are expected to be essential for the communication between the calcium-binding site in the transmembrane domain and the phosphorylation site in domain P, which are separated by $\sim$50 Å. The patterns in the motion of transmembrane helices suggest that helices M2, M4, and M8 are likely to be involved in calcium release to lumen, whereas helix M4 seems to be essential to the calcium binding from cytoplasm. The magnitude of their motions, however, is fairly small based on normal mode results at 300 K, which is consistent with the scheme that calcium binding and release are regulated by the binding and hydrolysis of ATP as well as phosphorylation (Lee and East, 2001; Toyoshima et al., 2000).

Finally, it is interesting to compare the current work to the direct approaches in electronic structure calculations pioneered by Almlof et al. (1982). The seemingly straightforward idea has made a profound impact on the electronic structure community. It is hoped that the direct implementation of BNM made in the current work will make normal mode analysis a routine tool for analyzing the structural and dynamical properties of very large molecules.

NOTE ADDED IN PROOF

After the submission of the present work, we have further optimized the efficiency of the code by using a sparse-matrix diagonalization routine. The new code is very efficient for large systems, especially when only a small number of modes are required. For Ca^{2+}-ATPase, for example, computing the 50 lowest frequency modes took less than 30 min on a 1.2 GHz Athlon.

REFERENCES

Almlof, J., K. Faegri, and K. Korsell. 1982. Principles for a direct SCF approach to LCAO-MO *ab initio* calculations. *J. Comput. Chem.* 3:385–399.

Aravind, L., M. Y. Galperin, and E. V. Koonin. 1998. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.* 23:127–129.

Bahar, I., A. R. Atilgan, and B. Erman. 1997. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Folding Des.* 2:173–181.

Baker, K. J., J. M. East, and A. G. Lee. 1994. Localization of the hinge region of the Ca^{2+}-Atpase of sarcoplasmic-reticulum using resonance energy-transfer. *Biochim. Biophys. Acta.* 1192:53–60.

Bernèche, S., and B. Roux. 2001. Energetics of ion conduction through the K^+ channel. *Nature.* 414:73–77.

Brooks, B. R., R. E. Burccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

Brooks, B. R., D. Janezic, and M. Karplus. 1995. Harmonic-analysis of large systems: I. Methodology. *J. Comput. Chem.* 16:1522–1542.

Brooks, C. L. III, and M. Karplus Pettitt. 1988. Proteins: a theoretical perspective of dynamics, structure and thermodynamics. *Adv. Chem. Phys.* 71:1–249.

Brüschweiler, R. 1992. Normal-modes and NMR order parameters in proteins. *J. Am. Chem. Soc.* 114:5341–5344.

Brüschweiler, R., and D. A. Case. 1994. Collective NMR relaxation model applied to protein dynamics. *Phys. Rev. Lett.* 72:940–943.

Carbral, J. M., Y. Zhou, and R. MacKinnon. 2001. Energetic optimization of ion conduction rate by the K^+ selective filter. *Nature.* 414:37–42.

Case, D. A. 1994. Normal-mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* 4:285–290.

Chattopadhyaya, R., W. E. Meador, A. R. Means, and F. A. Quiocho. 1992. Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.* 228: 1177–1192.

Cheetham, G. M. T., D. Jeruzalmi, and T. A. Steitz. 1999. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature.* 399:80–83.

Cheetham, G. M. T., and T. A. Steitz. 2000. Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr. Opin. Struct. Biol.* 10:117–123.

Cho, M., G. R. Fleming, S. Saito, I. Ohmine, and R. M. Stratt. 1994. Instantaneous normal-mode analysis of liquid water. *J. Chem. Phys.* 100:6672–6683.

Chun, H. M., C. E. Padilla, D. N. Chin, M. Watanabe, V. I. Karlov, H. E. Alper, K. Soosaar, K. M. Blair, O. M. Baker, L. S. D. Caves, R. Nagle, D. N. Haney, B. L. Farmer. 2000. MBO(N)D: a multi-body method for long-time molecular dynamics simulations. *J. Comput. Chem.* 21: 159–184.

Clarke, D. M., K. Maruyama, T. W. Loo, E. Leberer, G. Inesi, and D. H. MacLennan. 1989. Functional consequences of glutamate, aspartate, glutamine, and asparagine mutations in the stalk sector of the Ca^{2+}-Atpase of sarcoplasmic-reticulum. *J. Biol. Chem.* 264:11246–11251.

Cowan, S. W. 1993. Bacterial porins: lessons from 3 high-resolution structures. *Curr. Opin. Struct. Biol.* 3:501–507.

Crivici, A., and M. Ikura. 1995. Molecular and structural basis of target recognition by calmodulin. *Annu. Rev. Biophys. Biomol. Struct.* 24: 85–116.

de Meis, L. 1981. The Sarcoplasmic Reticulum. Wiley, New York.

Doyle, D. A., J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. Mackinnon. 1998. The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science.* 280:69–77.

Durand, P., G. Trinquier, and Y. Sanejouand. 1994. New approach for determining low-frequency normal-modes in macromolecules. *Biopolymers.* 34:759–771.

Fersht, A. 1998. Enzyme Structure, Mechanism and Protein Foldings, 3rd Ed. W. H. Freeman, New York.

Finn, B. E., J. Evenäs, T. Drakenberg, J. P. Waltho, E. Thulin, and S. Forsèn. 1995. Calcium-induced structural-changes and domain autonomy in calmodulin. *Nat. Struct. Biol.* 2:777–783.

Fisher, S., S. Michnick, and M. Karplus. 1993. A mechanism for rotamase catalysis by the fk506 binding-protein (FKBP). *Biochemistry.* 32: 13830–13837.

Greengard, L., and V. Rokhlin. 1987. A fast algorithm for particle simulations. *J. Comput. Phys.* 73:325–348.

Halle, B. 2002. Flexibility and packing in proteins. *Proc. Natl. Acad. Sci. U. S. A.* 99:1274–1279.

Henderson, I. M. J., A. P. Starling, M. Wictome, J. M. East, and A. G. Lee. 1994. Binding of Ca^{2+} to the $(Ca^{2+}-Mg^{2+})$-ATPase of sarcoplasmic reticulum kinetic studies. *Biochem. J.* 297:625–636.

Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

Jencks, W. P. 1989. How does a calcium-pump pump calcium. *J. Biol. Chem.* 264:18855–18858.

Jencks, W. P. 1992. On the mechanism of atp-driven Ca^{2+} transport by the calcium Atpase of sarcoplasmic-reticulum. *Ann. N. Y. Acad. Sci.* 671: 49–57.

Jencks, W. P., T. Yang, D. Peisach, and J. Myung. 1993. Calcium-ATPase of sarcoplasmic-reticulum has 4 binding-sites for calcium. *J. Biochem.* 32:7030–7034.

Kidera, A., K. Inaka, M. Matsushima, and N. Go. 1992. Normal mode refinement: crystallographic refinement of protein dynamic structure: II. Application to human lysozyme. *J. Mol. Biol.* 225:477–486.

Kuboniwa, H., N. Tjandra, S. Grzesiek, H. Ren, C. B. Klee, and A. Bax. 1995. Solution structure of calcium-free calmodulin. *Nat. Struct. Biol.* 2:768–776.

Lamm, G., and A. Szabo. 1986. Langevin modes of macromolecules. *J. Chem. Phys.* 85:7334–7348.

Lazaridis, T., and M. Karplus. 2000. Effective energy functions for protein structure prediction. *Curr. Opin. Struct. Biol.* 10:139–145.

Lee, A. G., and J. M. East. 2001. What the structure of a calcium pump tells us about its mechanism. *Biochem. J.* 356:665–683.

Ma, J., and M. Karplus. 1998. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. *Proc. Natl. Acad. Sci. U. S. A.* 95: 8502–8507.

MacKerell, Jr., A. D. 2001. Computational Biochemistry and Biophysics, O. M. Becker, A. D. Mackerell, Jr., B. Roux, and M. Watanabe, editors. Marcel Dekker, Inc. New York.

MacKerell, Jr., A. D., J. Wiórkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic-acids. *J. Am. Chem. Soc.* 117:11946–11975.

MacLennan, D. H., C. J. Brandl, B. Korczak, and N. M. Green. 1985. Amino-acid-sequence of a $Ca^{2+}-Mg^{2+}$-dependent ATPase from rabbit muscle sarcoplasmic-reticulum, deduced from its complementary-DNA sequence. *Nature.* 316:696–700.

MacLennan, D. H., W. J. Rice, and N. M. Green. 1997. The mechanism of Ca^{2+} transport by sarco(endo)plasmic reticulum Ca^{2+}-ATPases. *J. Biol. Chem.* 272:28815–28818.

Makri, N. 1999. Time-dependent quantum methods for large systems. *Annu. Rev. Phys. Chem.* 50:167–191.

Marques, O., and Y. Sanejouand. 1995. Hinge-bending motion in citrate synthase arising from normal mode calculations. *Proteins.* 23:557–560.

Møller, J. V., B. Juul, and M. de Maire. 1996. Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochim. Biophys. Acta.* 1286:1–51.

Mouawad, L., and D. Perahia. 1993. Diagonalization in a mixed basis: a method to compute low-frequency normal-modes for large macromolecules. *Biopolymers.* 33:599–611.

Neria, E., S. Fischer, and M. Karplus. 1996. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* 105:1902–1921.

Northrup, S. H., M. R. Pear, J. A. McCammon, and M. Karplus. 1980. Molecular dynamics of ferrocytochrome c. *Nature.* 286:304–305.

Parlett, B. N. 1980. The Symmetric Eigenvalue Problem. Prentice-Hall, Englewood Cliffs, NJ.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes, 2nd Ed. Cambridge University Press, New York.

Roux, B., and T. Simonson. 1999. Implicit solvent models. *Biophys. Chem.* 78:1–20.

Sagnella, D. E., and J. E. Straub. 1999. A study of vibrational relaxation of B-state carbon monoxide in the heme pocket of photolyzed carboxymyoglobin. *Biophys. J.* 77:70–84.

Shen, Y., Y. Kong, and J. Ma. 2002. Intrinsic flexibility and gating mechanism of the potassium channel KcsA. *Proc. Natl. Acad. Sci. U. S. A.* 99:1949–1953.

Simonson, T., G. Archontis, and M. Karplus. 1997. Continuum treatment of long-range interactions in free energy calculations: application to protein-ligand binding. *J. Phys. Chem. B.* 41:8349–8362.

Simonson, T., and D. Perahia. 1992. Normal-modes of symmetrical protein assemblies: application to the tobacco mosaic-virus protein disk. *Biophys. J.* 61:410–427.

Smith, J., K. Kuczera, and M. Karplus. 1990. Dynamics of myoglobin: comparison of simulation results with neutron-scattering spectra. *Proc. Natl. Acad. Sci. U. S. A.* 87:1601–1605.

Space, B., H. Rabitz, and A. Askar. 1993. Long-time scale molecular-dynamics subspace integration method applied to anharmonic crystals and glasses. *J. Chem. Phys.* 99:9070–9079.

Steinbach, P. J., and B. R. Brooks. 1994. New spherical-cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* 15:667–683.

Tama, F., F. X. Gadea, O. Marques, and Y. Sanejouand. 2000. Building-block approach for determining low-frequency normal modes of macromolecules. *Proteins Struct. Funct. Genet.* 41:1–7.

Tama, F., and Y. Sanejouand. 2001. Conformational change of proteins arising from normal mode calculations. *Prot. Eng.* 14:1–6.

Thomas, A., M. J. Field, and D. Perahia. 1996. Analysis of the low-frequency normal modes of the R state of aspartate transcarbamylase and a comparison with the T state modes. *J. Mol. Biol.* 261:490–506.

Thomas, A., K. Hinsen, M. J. Field, and D. Perahia. 1999. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. *Proteins Struct. Funct. Genet.* 34:96–112.

Tidor, B., and M. Karplus. 1994. The contribution of vibrational entropy to molecular association: the dimerization of insulin. *J. Mol. Biol.* 238: 405–414.

Tirion, M. M. 1996. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *Phys. Rev. Lett.* 77:1905–1908.

Toyoshima, C., M. Nakasako, H. Nomura, and H. Ogawa. 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 angstrom resolution. *Nature.* 405:647–655.

Toyoshima, C., H. Sasabe, and D. L. Stokes. 1993. 3-dimensional cryo-electron microscopy of the calcium-ion pump in the sarcoplasmic-reticulum membrane. *Nature.* 362:467–471.

van Vlijmen, H. W. T., and M. Karplus. 2001. Normal mode analysis of large systems with icosahedral symmetry: application to (Diala-nine)(60) in full and reduced basis set implementations. *J. Chem. Phys.* 115:691–698.

Von-Hippel, P. H., D. G. Bear, W. D. Morgan, and J. A. McSwiggen. 1984. Protein nucleic-acid interactions in transcription-A molecular analysis. *Annu. Rev. Biochem.* 53:389–446.

Voth, G. 1996. Path-integral centroid methods in quantum statistical mechanics and dynamics. *Adv. Chem. Phys.* 93:135–218.

Wilson, Jr., E. B., J. C. Decius, and P. C. Cross. 1980. Molecular Vibrations. Dovers, New York.

Zhang, G., and T. Schlick. 1994. The Langevin/Implicit-Euler/Normal-mode scheme for molecular-dynamics at large time steps. *J. Chem. Phys.* 101:4995–5012.

Zhang, M., T. Tanaka, and M. Ikura. 1995. Calcium-induced conformational transition revealed by the solution structure of apo calmodulin. *Nat. Struct. Biol.* 2:758–767.

Zhang, P., C. Toyoshima, K. Yonekura, K., N. M. Green, and D. L. Stokes. 1998. Structure of the calcium pump from sarcoplasmic reticulum at 8-angstrom resolution. *Nature.* 392:835–839.

Zhou, Y., J. M. Cabral, A. Kaufman, and R. MacKinnon. 2001. Chemistry of ion coordination and hydration revealed by a K^+ channel-FAB complex at 2.0 angstrom resolution. *Nature.* 414:43–48.