PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

A network of dynamically conserved residues deciphers the motions of maltose transporter

Suryani Lukman and Guy H. Grant*

Department of Chemistry, University of Cambridge, Cambridge, United Kingdom

ABSTRACT

The maltose transporter of *Escherichia coli* is a member of the ATP-binding cassette (ABC) transporter superfamily. The crystal structures of maltose transporter MalK have been determined for distinct conformations in the presence and absence of the ligand ATP, and other interacting proteins. Using the distinct MalK structures, normal mode analysis was performed to understand the dynamics behavior of the system. A network of dynamically important residues was obtained from the normal mode analysis and the analysis of point mutation on the normal modes. Our results suggest that the intradomain rotation occurs earlier than the interdomain rotation during the maltose-binding protein (MBP)-driven conformational changes of MalK. We inquire if protein motion and functional-driven evolutionary conservation are related. The sequence conservation of MalK was analyzed to derive a network of evolutionarily important residues. There are highly significant correlations between protein sequence and protein dynamics in many regions on the maltose transporter MalK, suggesting a link between the protein evolution and dynamics. The significant overlaps between the network of dynamically important residues and the network of evolutionarily important residues form a network of dynamically conserved residues.

Proteins 2009; 76:588–597.

Key words: ATP-binding cassette; conformational change; elastic network model; evolutionary information analysis; normal mode analysis.

INTRODUCTION

ATP-binding cassette (ABC) transporters constitute the largest transporter gene family and function ubiquitously from bacteria to humans.[1] Using the ATP-derived energy, ABC transporters transport various substrates or allocrites, ranging from ions and small organic molecules to carbohydrates and proteins, across membranes. In humans, ABC transporters are involved in cystic fibrosis, Tangier disease, and drug resistance.[2]

Most ABC transporters possess two transmembrane domains (TMDs) and two cytosolic nucleotide-binding domains (NBDs). Prokaryotic ABC transporters include a periplasmic component that binds specific substrates with high affinity.[3] The NBDs contain the classical nucleotide-binding Walker A and Walker B motifs, a conserved LSGGQ signature motif, a Q-loop mediating the interactions between the TMDs and the NBDs, and a D-loop.

Distinct crystal structures of ABCs in the maltose transporter MalK,[4] vitamin B12 importer BtuD,[5] and metal-chelate importer HI1470[6] reveal that the NBDs exist in at least three different states: open, semi-open, and closed. The availability of such crystal structures provides a gateway to understand the conformational changes concomitant with substrate transport and ATP hydrolysis in ABC transporters. In this study, we focus on the maltose transporter, which is involved in maltose/maltodextrin import. It comprises a periplasmic maltose-binding protein (MBP or MalE), two integral membrane proteins, MalF and MalG, and two copies of their cytoplasmic ABC MalK. The MalK monomer has two domains, an N-terminal 235-residue nucleotide binding domain (NBD) and a C-terminal 135-residue regulatory domain (RD). Crystal structures of MalK existing in different conformations, from open to closed (ATP-bound) conformations, are available.[3,4]

Molecular dynamics simulation on isolated dimeric MalK structures indicates that the simulation time of 15 ns is insufficient for the open conformation to evolve to the fully closed conformation.[7] To provide another perspective on the understanding of ABC transporters dynamics, we performed normal mode analysis (NMA), which has been used to study protein motions.[8,9] Although protein motions involve anharmonicity,[8] the low frequency normal modes are capable of describing functionally significant conformational rearrangements. Low-frequency normal modes have directional similarities to large-amplitude fluctuations in molecular dynamics simulations[10] and provide details concerning the important interdomain motions and displacements of flexible intradomain components.[11] Comparisons

Grant sponsor: Agency for Science, Technology, and Research, Singapore.
*Correspondence to: Guy H. Grant, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB21EW, United Kingdom. E-mail: ghg24@cam.ac.uk.
Received 16 October 2008; Revised 8 December 2008; Accepted 23 December 2008
Published online 20 January 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.22372

between two protein structural data of distinct conformations indicate that low-frequency normal modes overlap with real conformational changes.[12]

NMA has been used to study proteins such as citrate synthase,[13] lysozyme,[14] polymerase,[15] matrix metalloproteinase,[16] and multimer proteins such as hemoglobin,[17] aspartate transcarbamylase,[18] tryptophan synthase,[19] HIV-1 reverse transcriptase,[20] chaperonin GroEL,[11,21] chorismate mutase,[22] as well as ribosome.[23,24] Focusing on the protein motions, many of these studies have enhanced our understandings on the relationships between protein motions and functions. In NMA, proteins are analyzed as a large set of coupled harmonic oscillators. NMA using all-atom empirical potentials is computationally expensive and can conceal the global changes, especially for large protein systems.[25] However, NMA for large biomolecules can be performed with a simplified representation of potential energy; such a method is termed an elastic network model (ENM).[15,26] Large-scale motion in proteins can be described using simple pairwise Hookean potentials with a single spring-constant parameter.[26] Following this, the one-dimensional coarse-grained Gaussian network[27] and the three-dimensional anisotropic network[25] models were developed. The Gaussian network model has been fully reviewed in the Chapter 3 of Ref. 9. In all these ENMs, the nodes of the network represent atoms of amino acid residues and the edges are springs linking all nodes within a cutoff distance. ENM has been used to predict accurately temperature B-factors in these proteins: crambin, trypsin inhibitor, ribonuclease A, and T4 lysozyme.[27] ENM has also been successfully used to study the ligand entry and binding in the retinol-binding protein[25] and the global ratchet-like motions of ribosome.[28] Motivated by these, we used an ENM based on a Hookean harmonic potential with a single force constant to investigate the dynamics in MalK maltose transporter and the residues crucial in the conformational changes. These residues form a network of dynamically important residues.

We are motivated to further extend the current approaches. The fact that the sequences of MalK NBDs are highly conserved as compared to the diverse sequences of TMDs[29] has motivated us to include sequence analysis to understand the evolutionary influence on protein dynamics. Sequence analysis to predict a set of energetically coupled residues due to amino acids coevolution[30] has proven effective to understand and design proteins, for example the statistical information obtained from the sequence analysis of WW domain enables the sequence design of properly folded proteins.[31] Besides determining protein tertiary structures, what other information can sequence analysis tell us? Are the motions of ABC transporters and functionally driven evolutionary conservation related? Quantification of such a relationship in terms of residues is useful for drug design and protein engineering. Previous studies employing ENM are related to our studies. An extended ENM has been used together with molecular mechanics to elucidate the conformational changes of RNA polymerase in the functional cycle of transcription.[32] Conserved residues involved in the efficient open/closed transition of polymerases have been successfully identified in details from two structures in open/closed conformations.[33] The availability of multiple structural data of MalK allows us to go beyond identifying the conserved residues in conformational changes. Different regions of MalK exhibit more influences on the dynamics than other regions, depending on the distinct conformational transitions. This study demonstrates the capacity of the integration of sequence analysis and ENM to dissect conformational changes occurring in multimers and in multiple steps, from open via semi-open to closed conformations.

MATERIALS AND METHODS

The starting structures for dynamics analysis were the structures of *Escherichia coli* MalK in different dimeric conformations: open (1Q1E), semi-open (1Q1B), closed (1Q12),[4] and in-complex (2R6G).[3] The open and semi-open conformations are nucleotide-free, the closed conformation has ATP bound at each of its two active sites. The in-complex conformation consists of two MalKs, MalF, MalG, and MBP (also known as MalE).

Evolutionary information analysis

Homologous sequences from the non-redundant database were retrieved using the PSI-BLAST (www.ncbi.nlm.nih.gov/blast) with E cutoff = 10. Multiple sequence alignments (MSA) on 500 sequences were initially performed using the ClustalX2.0.4. Alignments were then manually adjusted using structure-based sequence alignment. Partial sequences containing <60% of the number of residues in the query protein were removed and the alignment was truncated to include only positions present on the query sequence.

To focus on evolutionary relationships among amino acid positions in the homologous proteins, the MSA is ensured to be statistically large and diverse enough. For a well-sampled MSA, random addition or elimination sequences should not significantly change the distribution at amino acid positions. In other words, the MSA has reached a state of statistical equilibrium in sequence space. Sufficiently diverse MSA should contain several positions on sequences with amino acid distributions equivalent to the mean in all natural proteins. For a processed MSA that represents evolutionary conservation, the information content is given by

$$S_n = \log 20 + \sum_{aa=1\ldots20} p^n_{aa} \log\left(p^n_{aa}\right), \qquad (1)$$

where p^n_{aa} is probability of substitution by amino acid aa at position n.[33] A high S_n means that a residue at position n is highly conserved. Residues of high S_n within a protein form a network of evolutionarily important residues.

Normal mode analysis

An elastic network model (ENM) in ElNémo,[34] which is based on a Hookean harmonic potential with a single force constant, was used to calculate pairwise atomic interactions that are within a cutoff distance R_c.[26] R_c accounts for van der Waals radii and decay of interactions with distance. The potential energy based on the ENM,

$$E_{\mathrm{ENM}} = \frac{C}{2} \sum_{d^0_{12} < R_c} (d_{12} - d^0_{12})^2, \tag{2}$$

where d_{12} is the distance between the dynamical coordinates of the atoms 1 and 2, d^0_{12} is the distance between the initial coordinates of the atoms 1 and 2 as given in the PDB structure. The strength of the potential C is assumed to be the same for all interacting pairs. The energy function, E_{ENM} is a minimum for any chosen configuration of any protein system, hence no minimization is required before NMA.

If the atoms are undergoing thermal fluctuations along each normal mode, the amount of their oscillations around their specified position can be quantified as B-factor,

$$B^{\mathrm{NMA}}_{\mathrm{factor}} = \frac{8\pi^2}{3} \times \frac{k_B T}{m_i} \sum_i^{3N-6} \frac{A^2_{ik}}{\omega^2_k}, \tag{3}$$

where k_B is the Boltzmann constant, T is the absolute temperature, m_i is the atomic mass of atom i, A_{ik} is the amplitude of the displacement of atom i in the kth normal mode, ω_k is the frequency of the kth normal mode.

Using the data from two or more of different conformations, a direction of conformational change can be identified. The eigenvectors of the lowest frequency normal modes were used to compute the similarity degree between the direction of a conformational change and the direction given by normal modes considered. To represent the similarity as overlap,

$$O_k = \frac{\left| \sum^{3N} A_{ik} \left(r^1_i - r^2_i \right) \right|}{\left| \sum^{3N} A^2_{ik} \sum^{3N} \left(r^1_i - r^2_i \right)^2 \right|^{\frac{1}{2}}}, \tag{4}$$

where r^1_i and r^2_i are the coordinates of atom i in the conformation 1 and 2, respectively.[13] If $O_k = 1$, the direction given by the kth normal mode is identical to that of conformational change. O_k has been proved remarkably successful in describing observed large-scale motions for a range of proteins.[12] To measure how well the first k modes together can represent a conformational change, we further defined the cumulative overlap,

$$\mathrm{CO} = \left(\sum O^2_k \right)^{\frac{1}{2}}. \tag{5}$$

Analysis of point mutation on normal modes

For a given mode, each residue, represented by its Cα, was deformed at sequence position n along the kth normal mode with a small amplitude δα. Upon deformation, the resulting changes in the strength of spring potential C between the Cα of residue at sequence position n and other atoms will result in the change of elastic energy, δE_{ENM} in accordance to the Eq. (2). The deformation was experimentally achieved by point mutations. The response to the deformation in the frequency of the kth normal mode was calculated using

$$\delta\eta_{kn} = v^T_k \cdot \delta H \cdot v_k, \tag{6}$$

where v_k is the eigenvector of the kth normal mode and δH is the Hessian matrix of δE_{ENM}.[33] If the kth normal mode is related to a biological function of the protein's conformational change, residues with high $\delta\eta_{kn}$ are likely to be dynamically important and potentially dynamic hot-spot residues. They form a network of dynamically important residues.

To obtain a detailed picture of the highly conserved protein regions by examining the correlation between S_n and $\delta\eta_{kn}$, we used Pearson's correlation method and performed a coarse-grained average of S_n and $\delta\eta_{kn}$. Residues which are both evolutionarily important and dynamically important form a network of dynamically conserved residues, indicated by the high correlation between S_n and $\delta\eta_{kn}$.

RESULTS AND DISCUSSION

Sequence conservation in MalK

The evolutionary information analysis well captured the conserved motifs in the sequences of maltose transporters (Fig. 1). The Walker A (residues 36–43), the Q-loop (residues 80–84), the LSGGQ (residues 134–141), the D-loop (residues 162–165), and the Switch (residues 184–193) motifs have high S_n, reflecting their high conservation. Compared to these motifs, the Walker B (residues 154–158) are relatively less conserved. Other regions with moderately high S_n include the residues 50–53 and the loop 235–239. All these residues form a network of evolutionarily important residues.

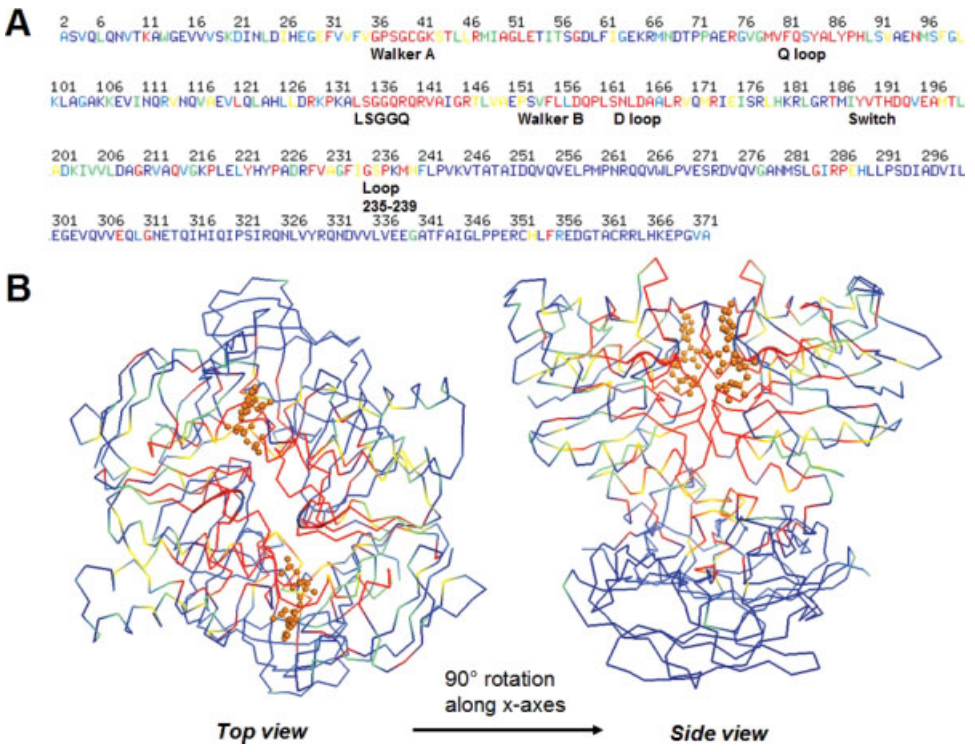

Figure 1

The S_n obtained from the evolutionary information analysis reflects the sequence conservation of the ABC transporters in general and the maltose transporter MalK in particular. The S_n is color-represented from red (corresponding to high conservation) to blue (corresponding to low conservation). (**A**) The sequence of MalK is presented with conserved motifs and important regions of high S_n values. (**B**) The color scheme is the same as that of (A) and for both images. *Top view* is seen from the plasma membrane side. *Side view* is after a rotation along *x*-axes for 90°.

The nucleotide-binding Walker A and Walker B motifs are seen in many ATPases, whereas the LSGGQ motif is unique to the ABC family.[4] Some of the conserved motifs and important regions with high S_n are far apart in the sequence space (Fig. 1). For example, the Walker A and the LSGGQ signature motifs are 99-residue apart in the sequence space, but the Walker A motif of a subunit comes closer to the LSGGQ motif of another subunit in the closed, ATP-bound conformation. Among the residues with high S_n crucial for the ATP stabilization, the R129 and the Q138 residues are engaged in van der Waals interaction with the ATP ribose.[4] Residues in the Walker A motif of a subunit and the LSGGQ motif of another subunit are involved in electrostatic and van der Waals interactions with the ATP phosphate groups. Among the phosphate-interacting residues with high S_n in the Walker A motif, the residue T44 interacts with the α-phosphate, the residues G39, C40, G41, K42, and S43 interact with the β-phosphate, and the residues S38, G39, and K42 interact with the γ-phosphate. Among the phosphate-interacting residues with high S_n in the LSGGQ motif, the residues S135, G136, and G137 interact with the γ-phosphate. These confirm the role of the evolutionary-important residues in nucleotide binding. Note that the evolutionary-important residues may serve distinct functions, such as ligand binding, enzymatic activity, and structural fold maintenance. In this study, we focus on the conserved residues that play significant roles in conformational changes.

Normal modes and B-factors

We analyzed and compared the motions in four MalK conformations: open, semi-open, closed, and in-complex (PDB: 1Q1E, 1Q1B, 1Q12, and 2R6G, respectively) using normal mode analysis (NMA). The cutoff distance R_c (see Methods) does not have significant effect on frequency of slow normal modes (data not shown). Different R_c values have been used to account for pairwise interactions between all Cα atoms in previous studies: 7 Å,[27] 8 Å,[34] 10 Å.[15,21] To determine what R_c value to use, we correlated the NMA-derived B-factors, obtained from the 100 normal modes after the first six zero-frequency modes related to rotational and translational motions, to the the overall experimentally determined B-factors across different R_c in the all-atoms and the Cα models (Fig. 2). A high correlation indicates that the normal modes capture well the flexible characteristics of protein. Compared to the other conformations of MalK, the open conformation shows a relatively higher correla-

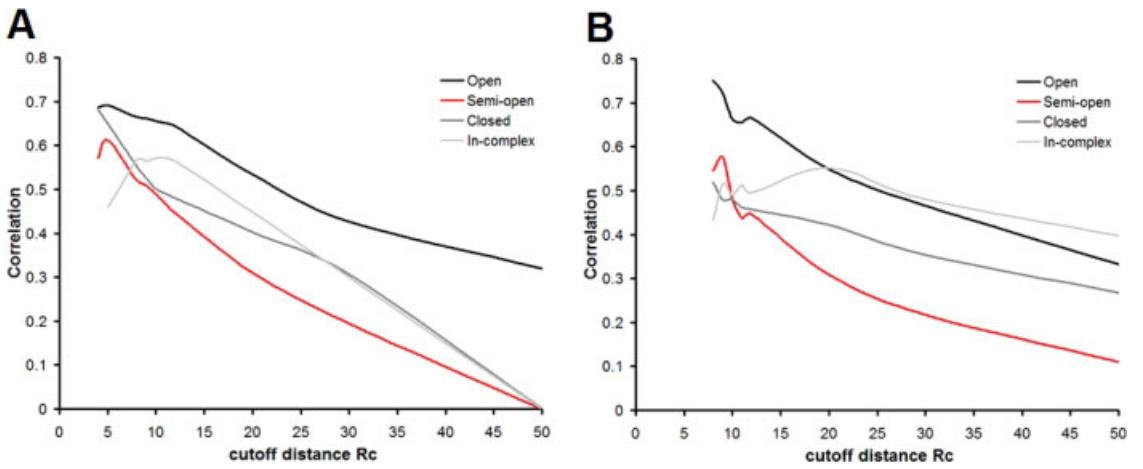

Figure 2

Correlations between NMA-derived and experimental B-factors for the different conformations of maltose transporter MalK across different R_c values. (**A**) All-atoms (**B**) Cα atoms. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tion in both all-atoms and Cα models. The optimum R_c is at shorter distances for all-atoms model than those for the Cα model. In DNA-dependent polymerases, the lowest and/or the second lowest frequency normal modes obtained from the open conformation have correlation coefficients greater than 0.5.[15] For enzymes such as ci-

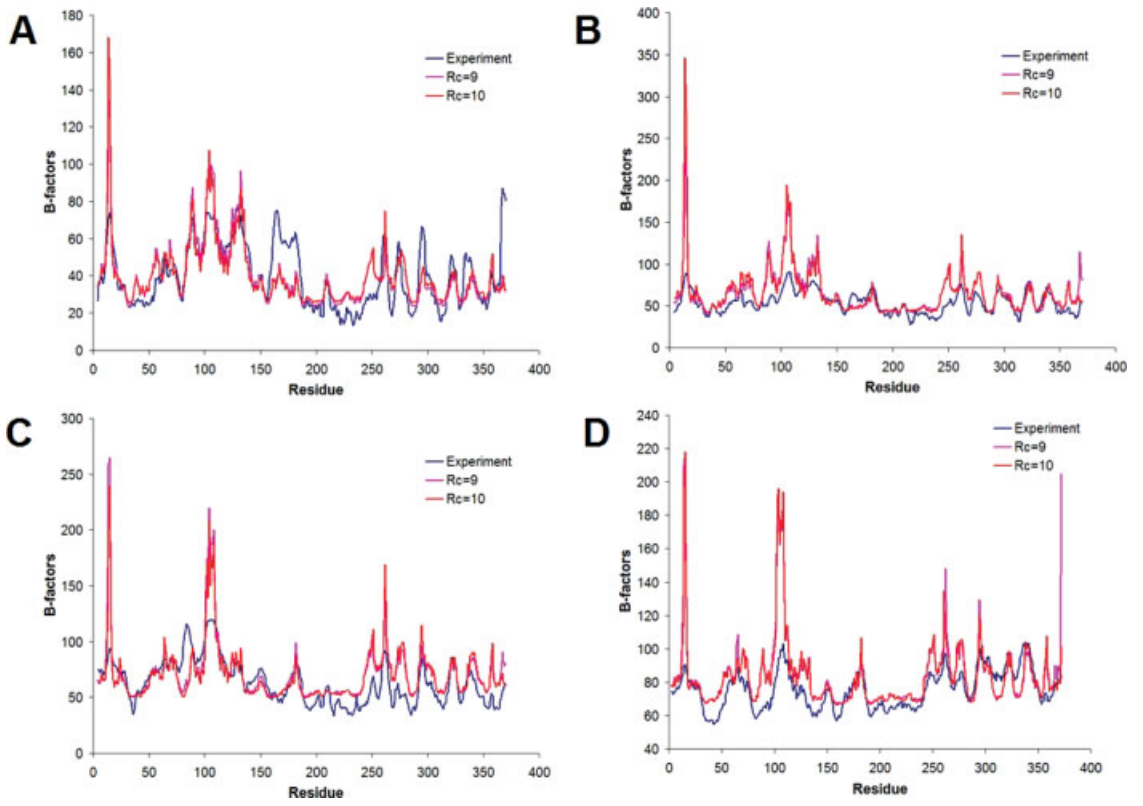

Figure 3

B-factors for different conformations of maltose transporter MalK derived from NMA using different cutoff distance R_c and from X-ray crystallography experiment. (**A**) Open conformation. (**B**) Semi-open conformation. (**C**) Closed conformation. (**D**) In-complex conformation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

trate synthase, triglyceride lipase, and thymidylate synthase, normal modes deduced from the open conformation compare better with the conformational changes than those obtained from other conformations.[12] For subsequent analysis, we used R_c = 9 Å as it gave high correlation in the Cα-atoms model and to facilitate the analysis of point mutations on normal modes.

The B-factors from normal modes across different R_c are generally similar to those from experiment (Fig. 3). Both the N- and C- termini have high B-factors as they are inherently flexible parts of proteins and observed normal modes amplified such flexibility with B-factors much higher than those experimentally determined. We noted that B-factors account for not only temperature-dependent vibrations of the atoms as derived from the Eq. (3), but also other effects such as static disorders.[35] It should also be noted that the theory of B-factors unrealistically assumes independent atomic oscillations, in contrast to the coupled oscillations of NMA.

The NBD (residues 1–235) and RD (residues 236–370) of MalK have an α/β motif and an all-β fold, respectively. Residues with relatively low B-values are more localized and constrained, as they generally participate in secondary structures. In the open conformation, the NMA-derived B-factors for residues 160–187 are lower than the experimental ones, although they display similar fluctuation patterns. Residues 162–165 form the conserved D-loop. Residues 184–187 are parts of Switch region, which has His192 that, in the presence of ATP, forms a strong H-bond with the γ-phosphate of ATP stabilizing its position.[4] The value differences between NMA-derived and experimental B-factors for residues 160–187 may result from extrinsic factors such as temperature and lattice environments, and these effects are unabled to be portrayed by NMA. Residues 103–108 (in the NBD) and 221–255 (in the RD) display higher NMA-derived B-factors than the experimental ones. Residues 103–108 are at the interaction region with the lipid bilayers, and NMA is performed in vacuum in the absence of cell membrane, hence explaining the higher NMA-derived B-factors. Residues 235–255 form loops, hence they have more freedom to move.

In the semi-open conformation, residues 85–90, 101–112, 125–135 have higher NMA-derived B-factors [Fig. 3(B)]. Most of these residues do not belong to constrained secondary structure: residues 85–90, 101–105, and 125–134 are loops. These loop regions that move during the transition from the open to the closed conformation, have NMA-derived B-factors which are 2–3 times higher than the experimental B-factors.

The D-loop (residues 162–165) in the semi-open conformation and the the Q-loop (residues 80–84) in the closed conformation have lower NMA-derived B-factors than the experimental ones [Fig. 3(B, C)]. The Q-loop interacts with γ-phosphate through a H-bond with water. The Q-loop also interacts with the EAA-loops of MalF and MalG transmembrane proteins.[4] In the closed conformation, the experimental B-factors for the Q-loop corresponds to increased values, suggesting the flexible nature of the Q-loop. However, the NMA-derived B-factors are able to capture the well-ordered character of the Q-loop, even in the absence of MalF and MalG in NMA. In the intact complex containing two MalK as well as MalF and MalG, the Q-loop is well-ordered [Fig. 3(D)] and contains a β-strand.[3] The low B-factors may suggest a region of conservation in the MalK dimer, especially for loop regions which generally display less constraint in motion. In fact, the Q-loop and the D-loop are one of the critical conserved motifs.[36] In contrast to the high B-factors presented in the study on the monomer ABC ATPase domain of human TAP1, the transporter associated with antigen processing,[36] we infer that the presence of dimer present a more ordered environment in the conserved D-loops of MalK.

For the in-complex conformation, in which MalK dimers are analyzed together with the presence of MBP, MalF, and MalG, the NMA-derived B-factors generally show similar fluctuation patterns to those of experiment [Fig. 3(D)]. The Walker A motif, the Q-loop, the LSGGQ motif, and the D-loop display low B-factors.

Conformational changes and dynamically important residues

For the analysis of normal mode using R_c = 9, the cumulative overlaps for the first 10 nonzero modes are 0.707, 0.796, 0.720 for open to semi-open, semi-open to closed, and open to closed conformations, respectively [Fig. 4(A)]. The low-frequency modes are sufficient to represent the conformational changes in MalK, especially the first 10 nonzero modes in which more than 70% cumulative overlaps with conformational changes are achieved. The highest overlap among the first 10 nonzero modes is given by mode 9, with values of 0.503 (open to semi-open) and 0.612 (open to closed). For the semi-open to closed transition, the highest overlaps of 0.321 and 0.231 are given by mode 8 and mode 7, respectively. From the RMSD of mode 9 in the open conformation, and the RMSD of modes 7 and 8 in the semi-open conformation [Fig. 4(B)], the atomic displacements demonstrate similar patterns for both modes, especially the high RMSDs involving the residues in the NBDs. Among the critical conserved sequence motifs in MalK, the C-termini of the Q-loop (Q-loop is formed by the residues 80-84) experience an increase in RMSD as compared to previous neigboring residues in the three modes. In all of the three modes, peaks of RMSD are also observed in the the N-terminal of Walker B (residues 154–158), the Switch (residues 184–193), and the residues (50, 52, 73) that form a hydrophobic pocket to interact with MalF and MalG. There is a difference in the D-loop (residues 162–165) in which it has high RMSD in the mode 9 of open confor-

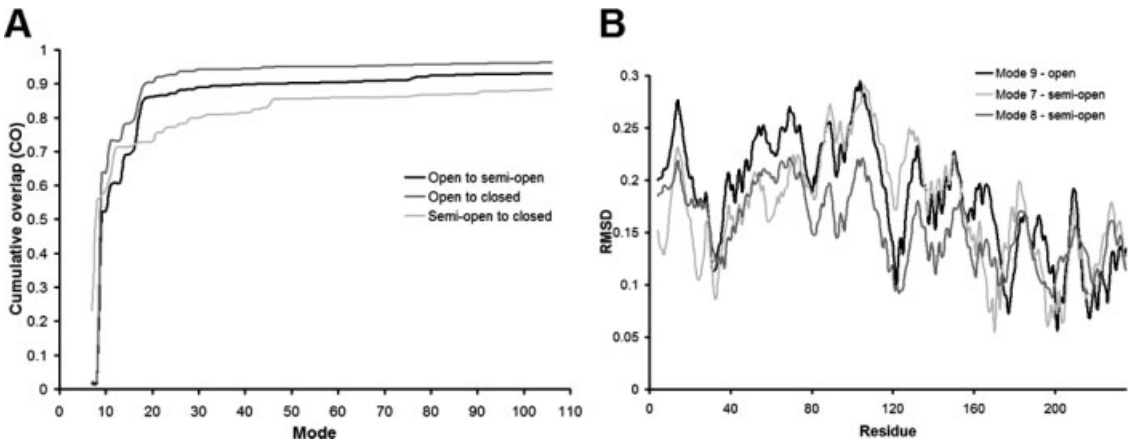

Figure 4

The normal modes capture the conformational changes. (**A**) The low-frequency modes are sufficient to represent the conformational changes in MalK. (**B**) The RMSDs of modes representing MalK conformational changes are given.

mation and low RMSD in the mode 7 of semi-open conformation.

We deformed amino acid residues for the mode 9 in the open conformation, the mode 7 and 8 semi-open conformations. The clusters of residues with high δη form the network of dynamically important residues (Fig. 5). Many of the residues in the network of dynamically important residues are parts of the critical conserved motifs in MalK, as given by the evolutionary information analysis (Fig. 1). These residues are important for allosteric functions of proteins, in particular their transition from open, via semi-open, to closed conformational states. In mode 9 of the open-conformation, the Q-loop and the D-loop are highly affected by deformation. This is followed by the high δη in the Walker A, LSGGQ, and Switch regions in the mode 8 of the semi-open conformations. The values of δη in the mode 7 of semi-open conformation are intermediate in-between those in the mode 9 and 8 of open and semi-open conformations respectively, for the Walker A, Q-loop, LSGGQ, D-loop, and Switch regions. We proposed that in the event of maltose-loaded MBP stimulation, the motions of Q-loop and the D-loops precede those of Walker A and LSGGQ which are involved in ATP nucleotide binding and the motion of the Switch region. This sequence of motions provides further evidence for the key concept that the conformational changes from the resting state of $MalFGK_2$ transporter (in which MalK dimers are in their open-conformation) to the catalytic state are driven by the binding of the maltose-loaded MBP to the MalF and MalG, instead of the binding of ATP to the MalK dimers. Similarly, the rate-limiting step, for the conversion of substrates to products in many enzymes, is the catalytic cycle conformational changes instead of the chemistry of active sites.[37,38] To dissect the motions within MalK dimers due to the MBP-triggered conformational changes, we analyzed each region involved.

Two helices from the helical subdomain (residues 88–157), the helix following the Walker A motif (residues 41–51), and the Q-loop form a cleft to dock the EAA-loops of the transmembrane MalF and MalG.[3] Inside the MalK cleft, the EAA-loops rotate in a manner similar to the motion of a ball-and-socket joint.[3] The high δη values of the Q-loop region in the open-conformation indicate its importance (Fig. 5). The Q-loop region also has high conservation as evidenced from its high S_n (Fig. 1). From the crystal structure of the $MalFGK_2$, it is revealed that the C-terminal tail of MalG inserts into the MalK dimer interface by forming three H-bonds with the backbone atoms of the Q-loop.[3] The C-terminal tail of MalG

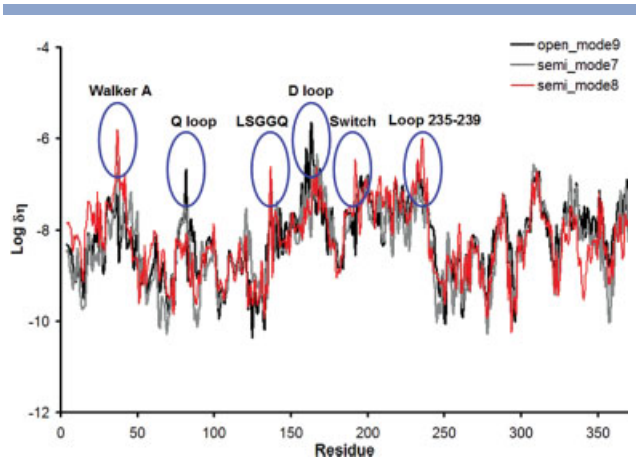

Figure 5

The δη obtained from the analysis of point mutation on mode 9 in the open conformation and on mode 7 and 8 in the semi-open conformation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

is not required for the closure of the MalK NBDs but may be crucial for the formation of the catalytic intermediate conformation of the entire *MalFGK*$_2$ transporter.[3,4]

ATP is postulated to bind first to the Walker A motifs instead of the LSGGQ motifs in the open conformation.[4] The Walker A and the LSGGQ motifs only come closer in the semi-open conformation (Fig. 5). Such motions are preceded by the optimization of the position of the helical subdomain relative to the RecA-like subdomain. The Q-loop, a crucial conserved motif in the helical subdomain, shows motion in the open-conformation (Fig. 5). The optimization of the position of the helical subdomain relative to the RecA-like subdomain is related to an intradomain rotation between the RecA-like and the helical subdomains. It is likely that the Q-loop, which connects the rigid-bodies of the helical and the RecA-like subdomains, is responsible for the intradomain rotation. As a result of this intradomain rotation, the helical subdomain moves inward toward the ATP-binding site.[4] Note that there is not only intradomain rotation but also interdomain rotation of the entire NBD relative to the RD about a hinge region situated in the loop (residues 235–239) connecting the NBD and RD. Loop 235–239 displays high S_n (Fig. 1) and high $\delta\eta$ (Fig. 5), especially in the mode 8 of the semi-open conformation. Based on these observations, we suggest that the intradomain rotation occurs earlier than the interdomain rotation during the MBP-driven conformational changes of MalK. To connect these findings with the previous analysis on the intact transporter,[3] the rotations of the EAA-loops in the transmembrane MalF and MalG implicate the intradomain rotations at the Q-loop. These are then followed by the interdomain rotation at the loop 235–239.

The switch region has significant motion as determined from the deformation analysis in the semi-open conformation (Fig. 5). This is in agreement with our hypothesis that the motion of the switch region is later than those of the Q-loop and the D-loop. The Switch region possesses His 192 which functions to polarize the water attachment for ATP hydrolysis. This is only likely to happen after the two NBDs come closer.

The network of dynamically conserved residues

To understand the relationship between protein dynamics and protein evolution, we correlated the $\delta\eta$ to sequence conservation represented by S_n (Fig. 6). For the open conformation as represented by the mode 9, the average Pearson's correlations are 0.69, 0.58, 0.52, 0.49, and 0.59 for residues in the Q-loop, the LSGGQ, the D-loop, the Switch, and the loop 235–239 regions, respectively. For the semi-open conformations as represented by the mode 7 and 8, the average Pearson's correlations are 0.22 and 0.27, 0.46 and 0.45, 0.20 and 0.28, 0.59 and 0.59, 0.71 and 0.69, for residues in the Q-loop, the LSGGQ,

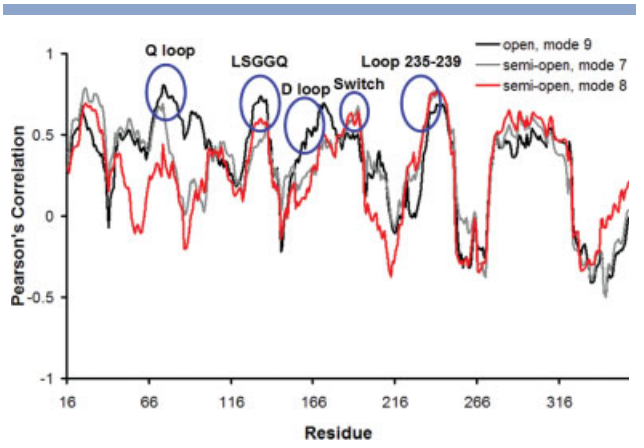

Figure 6

Pearson's correlation calculated between $\delta\eta$ and S_n obtained from the analysis of point mutations on normal modes and the analysis of the evolutionary information, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

the D-loop, the Switch, and the loop 235–239 region, respectively. The residues of these regions with significant correlation between the $\delta\eta$ and S_n form the network of dynamically conserved residues. The Q-loop, the LSGGQ motif, and the D-loop have higher Pearson's correlations in the open than those in the semi-open conformation, whereas the Switch region and the loop 235–239 have higher Pearson's correlations in the semi-open than those in the open conformation. These results provide evidence on the linkage between protein evolution and dynamics.

It is also interesting to note that some of the clusters in the network of dynamically conserved residues are both spatially distant in sequence and apart in 3D. For example, the Q-loop is 49, 69, 77, 99, 150 residues apart from the LSGGQ, Walker B, D-loop, Switch, and loop 235–239 motifs, respectively. In 3D of a MalK monomer, the distances from the LSGGQ motif to Walker A, Q-loop, Walker B, D-loop, Switch, and Loop 235–239 regions in the open conformation are 27.92, 14.67, 18.86, 18.06, 21.96, and 36.01 Å, respectively. The respective distances in the closed conformations are 24.75, 12.66, 19.04, 10.16, 21.38, and 31.90 Å. For the MalK homodimer, the distances between two Walker B from two interacting chains are 32.95, 37.93, and 43.88 Å in the closed, semi-open, and open conformations, respectively. These suggest the importance of long-range communication among residues in allosteric mechanism.

Mutational evidence for the residues in the network of dynamically conserved residues also suggest the crucial roles that these MalK residues play in maltose transport. The individual mutations of K42N (within the Walker A region) and H192N (within the Switch region), result in significant reduction in maltose transport and ATPase activities.[39] Another mutation along the LSGGQ sequence is Q140L, which inhibits the maltose transport

by trapping the MalK in ATP-bound conformation.[40] For a residue next to the Q-loop, the A85M substitution suppresses EAA-loop mutation,[41] suggesting its role in interacting with the EAA-loops of MalF and MalG. Since the intradomain rotation at the Q-loop are followed by the interdomain rotation at the loop 235–239, it will be interesting to see the phenotypes of loop 235–239 mutants and their maltose transport profiles. The conserved glutamate at position 159, a residue immediately following the Walker B motif, was suggested to polarize the water molecule that attacks the bond between the β- and γ-phosphates of ATP.[42] The E159Q mutation eliminates the catalytic carboxylate and results in no ATPase activity.[40]

There are also regions of residues with high S_n, but low δη-values. We propose that these residues are conserved for purposes other than dynamics and conformational changes. Further work is required to illuminate their roles in the functions of MalK maltose transporter.

CONCLUSION

Our approach, using an integration of sequence, structure, and normal mode analysis, can selectively determine the residues involved in the protein conformational changes. We also established evidence on the the relationship between protein dynamics (studied using normal mode analysis) and protein evolution (as represented by the sequence conservation). A network of residues with high degree of dynamical- and sequence-conservation is involved in the large-scale interdomain and intradomain movements that facilitate conformational changes in the MalK maltose transporter, adding further to our understanding on the ATP-binding cassette transport system, in particular maltose-specific one. The current approach is limited to proteins with resolved structures. For future studies, we plan to extend the method to include secondary structure analysis and comparison methods beyond sequence analysis.

ACKNOWLEDGMENTS

The authors thank Erik Lindahl for providing the source code to calculate the δη.

REFERENCES

1. Higgins C. ABC transporters: from microorganisms to man. Annu Rev Cell Biol 1992;8:67–113.
2. Dean M, Rzhetsky A, Allikmets R. The human ATP-binding cassette (ABC) transporter superfamily. Genome Res 2001;11:1156–1166.
3. Oldham M, Khare D, Quiocho F, Davidson A, Chen J. Crystal structure of a catalytic intermediate of the maltose transporter. Nature 2007;450:515–521.
4. Chen J, Lu G, Lin J, Davidson A, Quiocho F. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. Mol Cell 2003;12:651–661.
5. Locher K, Lee A, Rees D. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. Science 2002;296: 1091–1098.
6. Pinkett H, Lee A, Lum P, Locher K, Rees D. An inward-facing conformation of a putative metal-chelate-type ABC transporter. Science 2007;315:373–377.
7. Oloo E, Fung E, Tieleman D. The dynamics of the MgATP-driven closure of MalK, the energy-transducing subunit of the maltose ABC transporter. J Biol Chem 2006;281:28397–28407.
8. McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press; 1987.
9. Cui Q, Bahar I. Normal mode analysis: theory and applications to biological and chemical systems. Boca Raton, FL: CRC Press; 2006.
10. Amadei A, Linssen A, Berendsen H. Essential dynamics of proteins. Proteins 1993;17:412–425.
11. Ma J, Karplus M. The allosteric mechanism of the chaperonin GroEL: a dynamic analysis. Proc Natl Acad Sci USA 1998;95:8502–8507.
12. Tama F, Sanejouand Y-H. Conformational change of proteins arising from normal mode calculations. Protein Eng Des Sel 2001;14:1–6.
13. Marques O, Sanejouand Y-H. Hinge-bending motion in citrate synthase arising from normal mode calculations. Proteins 1995;23:557–560.
14. Brooks B, Karplus M. Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. Proc Natl Acad Sci USA 1985;82:4995–4999.
15. Delarue M, Sanejouand Y-H. Simplified normal mode analysis of conformational transitions in DNA-dependent polymerases: the elastic network model. J Mol Biol 2002;320:1011–1024.
16. Floquet N, Marechal J-D, Badet-Deniset M-A, Robert CH, Dauchez M, Perahia D. Normal mode analysis as a prerequisite for drug design: application to matrix metalloproteinases inhibitors. FEBS Lett 2006;580:5130–5136.
17. Mouawad L, Perahia D. Motions in hemoglobin studied by normal mode analysis and energy minimization: evidence for the existence of tertiary T-like, quaternary R-like intermediate structures. J Mol Biol 1996;258:393–410.
18. Thomas A, Field M, Mouawad L, Perahia D. Analysis of the low frequency normal modes of the T-state of aspartate transcarbamylase. J Mol Biol 1996;257:1070–1087.
19. Bahar I, Jernigan R. Cooperative fluctuations and subunit communication in tryptophan synthase. Biochemistry 1999;38:3478–3490.
20. Bahar I, Erman B, Jernigan R, Atilgan A, Covell D. Examination of collective motions in HIV-1 reverse transcriptase. Examination of flexibility and enzyme function. J Mol Biol 1999;285:1023–1037.
21. Zheng W, Brooks B, Thirumalai D. Allosteric transitions in the chaperonin GroEL are captured by a dominant normal mode that is most robust to sequence variations. Biophys J 2007;93: 2289–2299.
22. Kong Y, Ma J, Karplus M, Lipscomb W. The allosteric mechanism of yeast chorismate mutase: a dynamic analysis. J Mol Biol 2006; 356:237–247.
23. Tama F, Valle M, Frank J, Brooks CL, III. Dynamic reorganization of the functionally active ribosome explored by normal mode analysis and cryo-electron microscopy. Proc Natl Acad Sci USA 2003;100:9319–9323.
24. Kurkcuoglu O, Doruker P, Sen T, Kloczkowski A, Jernigan R. The ribosome structure controls and directs mRNA entry, translocation and exit dynamics. Phys Biol 2008;5:46005.
25. Atilgan A, Durell S, Jernigan R, Demirel M, Keskin O, Bahar I. Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 2001;80:505–515.
26. Tirion M. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. Phys Rev Lett 1996;77:1905–1908.
27. Bahar I, Atilgan A, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 1997;2:173–181.

28. Wang Y, Rader A, Bahar I, Jernigan R. Global ribosome motions revealed with elastic network model. J Struct Biol 2004;147:302–314.
29. Holland I, Blight M. ABC-ATPases, adaptable energy generators fuelling transmembrane movement of a variety of molecules in organisms from bacteria to humans. J Mol Biol 1999;293:381–399.
30. Lockless S, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–299.
31. Socolich M, Lockless S, Russ W, Lee H, Gardner K, Ranganathan R. Evolutionary information for specifying a protein fold. Nature 2005;437:512–518.
32. Wynsberghe A, Li G, Cui Q. Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle. Biochemistry 2004;43:13083–13096.
33. Zheng W, Brooks B, Doniach S, Thirumalai D. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. Structure 2005;13:565–577.
34. Suhre K, Sanejouand Y-H. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. Nucleic Acids Res 2004;32:W610–W614.
35. Frauenfelder H, Petsko G, Tsernoglou D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. Nature 1979;280:558–563.
36. Gaudet R, Wiley D. Structure of the ABC ATPase domain of human TAP1, the transporter associated with antigen processing. EMBO J 2001;20:4964–4972.
37. Watt E., Shimada H, Kovrigin E, Loria J. The mechanism of rate-limiting motions in enzyme function. Proc Natl Acad Sci USA 2007;104:11981–11986.
38. Henzler-Wildman K, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, Pozharski E, Wilson M, Petsko G, Karplus M, Hübrer C, Kern D. Intrinsic motions along an enzymatic reaction trajectory. Nature 2007;450:838–844.
39. Davidson A, Sharma S. Mutation of a single MalK subunit severely impairs maltose transport activity in *Escherichia coli*. J Bacteriol 1997;179:5458–5464.
40. Daus M, Landmesser H, Schlosser A, Müller P, Herrmann A, Schneider E. ATP induces conformational changes of periplasmic loop regions of the maltose ATP-binding cassette transporter. J Biol Chem 2006;281:3856–3865.
41. Mourez M, Hofnung M, Dassa E. Subunits interactions in ABC transporters. A conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. EMBO J 1997;16:3066–3077.
42. Hung L-W, Wang I, Nikaido K, Liu P-Q, Ames G, Kim S-H. Crystal structure of the ATP-binding subunit of an ABC transporter. Nature 1998;396:703–707.