

RESEARCH ARTICLE

Protein conformational transitions explored by a morphing approach based on normal mode analysis in internal coordinates

Byung Ho Lee¹, Soon Woo Park¹, Soojin Jo², Moon Ki Kim^{1,3*}

1 School of Mechanical Engineering, Sungkyunkwan University, Suwon, South Korea, **2** Department of Physics and Institute of Basic Science, Sungkyunkwan University, Suwon, South Korea, **3** Sungkyunkwan Advanced Institute of Nanotechnology (SAINT), Sungkyunkwan University, Suwon, South Korea

* mkkim1212@skku.edu**OPEN ACCESS**

Citation: Lee BH, Park SW, Jo S, Kim MK (2021) Protein conformational transitions explored by a morphing approach based on normal mode analysis in internal coordinates. PLoS ONE 16(11): e0258818. <https://doi.org/10.1371/journal.pone.0258818>

Editor: Wenfei Li, Nanjing University, CHINA

Received: June 14, 2021

Accepted: October 5, 2021

Published: November 4, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0258818>

Copyright: © 2021 Lee et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are available at <http://bioengineering.skku.ac.kr/kosmos/download.php>.

Funding: BHL, SWP, and MKK are supported by Basic Science Research Program through the

Abstract

Large-scale conformational changes are essential for proteins to function properly. Given that these transition events rarely occur, however, it is challenging to comprehend their underlying mechanisms through experimental and theoretical approaches. In this study, we propose a new computational methodology called internal coordinate normal mode-guided elastic network interpolation (ICONGENI) to predict conformational transition pathways in proteins. Its basic approach is to sample intermediate conformations by interpolating the interatomic distance between two end-point conformations with the degrees of freedom constrained by the low-frequency dynamics afforded by normal mode analysis in internal coordinates. For validation of ICONGENI, it is applied to proteins that undergo open-closed transitions, and the simulation results (i.e., simulated transition pathways) are compared with those of another technique, to demonstrate that ICONGENI can explore highly reliable pathways in terms of thermal and chemical stability. Furthermore, we generate an ensemble of transition pathways through ICONGENI and investigate the possibility of using this method to reveal the transition mechanisms even when there are unknown metastable states on rough energy landscapes.

1 Introduction

Max Perutz and John Kendrew first determined the three-dimensional (3D) structures of hemoglobin and myoglobin in the 1960s, which laid the foundation for the field of structural biology [1–3]. Since then, numerous experiment-based studies have been performed to reveal structural information of macromolecules, resulting in more than 183,000 atomic-level structures in the Protein Data Bank (PDB) archive [4]. In addition, the vast array of information has demonstrated that regulated conformational changes are of crucial importance for proteins to perform their biological functions, which has led to increasing awareness of the need to probe these large transitions. Indeed, various experimental techniques such as nuclear magnetic resonance spectroscopy [5], small-angle X-ray scattering [6], and single-molecule spectroscopy [7] have been widely utilized to analyze the dynamic behavior of proteins. However,

National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (No. 2019R1F1A1042633) and by the Ministry of Education (No. 2021R1A6A1A03039696). SJ is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2020R111A1A01055405). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

obtaining experimental information on the conformational changes of proteins is a longstanding challenge due to not only the intrinsic properties of the transition events with short-lived intermediate conformations, but also several technical limitations like sample preparation, system size, and time scale [8, 9].

Aside from the experimental studies, computational methods have played a key role in better understanding the functionally relevant dynamics of proteins that are difficult to capture through experimental approaches. Especially, molecular dynamics (MD) simulation, which samples conformational states in atomic detail by calculating interatomic forces using molecular mechanics force fields, has become one of the most powerful and popular tools [10, 11]. However, despite its successful applications in numerous studies, MD simulation has intrinsic limitations in exploring the large-scale conformational changes: the simulated systems easily get trapped in stable or metastable states and rarely cross high-energy barriers toward functional states, even on millisecond time scales. Recently, various MD strategies, such as development of special-purpose supercomputers [12, 13] and enhanced sampling methods [14–16], have contributed greatly to improving the performance of MD simulation, but the time-scale limitation remains to be resolved.

As an alternative approach to overcome the issue of computational complexity, normal mode analysis (NMA) has received much attention because it provides an efficient way to elucidate the intrinsic dynamics of proteins that are related to the global transitions [17–20]. NMA calculation is based on harmonic approximation of the potential energy function, and the resulting mode shapes are valid only near an equilibrium state. In other words, this method has inherent limitations in directly predicting conformational transitions that require inharmonic movements over energy barriers. Therefore, various methods combining NMA with other computational techniques have been developed to explore effective transition pathways between two end-point conformations [21–24].

In this study, we propose a new NMA-based pathway generation method called internal coordinate normal mode-guided elastic network interpolation (ICONGENI), an improved technique over the normal mode-guided elastic network interpolation (NGENI) [25]. The fundamental concept of both methods is to obtain intermediate conformations comprising a transition pathway by iteratively calculating displacement vectors to minimize error between the simulated intermediates and the targeted ones. In this process, NMA calculation is required to represent the displacement vectors as linear combinations of the lowest normal mode shapes and makes a critical difference between the two techniques: it is achieved in internal coordinates (IC-NMA) and Cartesian coordinates (CC-NMA) in ICONGENI and NGENI, respectively. CC-NMA has been widely used in studying protein dynamics due to the inherent nature of Cartesian coordinates (CCs): high computational efficiency and intuitive expression of protein dynamics but has the disadvantage of producing the mode shapes having unrealistic distortions like bond length stretching and bond angle bending [26]. On the other hand, IC-NMA has a distinctive advantage in describing the large-scale transitions in proteins. Internal coordinates (ICs) inherently facilitate the separation of the torsion angles from the others, so that IC-NMA can be performed in torsion angle space. Given that the conformational changes are dominantly influenced by the variations in the torsion angles, not in the bond lengths and the bond angles that are nearly rigid, this strategy enables it to produce chemically relevant mode shapes preventing the unrealistic distortions of bond lengths and bond angles and extending the validity of the harmonic approximation in calculation [26–28]. In terms of computational complexity, IC-NMA is less efficient than CC-NMA because it has extra calculations involved in transformation either from CCs to ICs or from ICs to CCs (see section 2.2 for further details), but this is not a critical issue because both methods can be performed at the personal computer level. In other words, IC-NMA is more suitable than CC-NMA for

describing and exploring the conformational changes in proteins, which provides an insight into the development of ICONGENI.

For validation of ICONGENI, we first demonstrate the superiority of IC-NMA in predicting large-scale conformational transitions by comparing the performance of ICONGENI to that of NGENI where the pathway is explored using CC-NMA [25]. Both methods are applied to two proteins: *E. coli* adenylate kinase (ADK) and *E. coli* ribose-binding protein (RBP). The comparative analyses of the distributions of ICs and the potential energies of the resulting pathways show that the transition pathways simulated by ICONGENI have higher thermal and chemical stability than those by NGENI. However, its efficient manner of computing intermediate structures has intrinsic limitations in exploring large-scale transitions on complex energy landscapes. To address this issue, ICONGENI generated a pathway ensemble for ADK dependent on the number of normal modes used in these simulations and characterized the ensemble in interdomain angle space, demonstrating that ICONGENI can explore plausible pathways on complex free energy landscapes.

2 Materials and methods

2.1 Protein structural information

To explore transition pathways of proteins through ICONGENI and NGENI, two end-point structures of each protein should be used as reference information. In the ADK case, the open and closed structures are chain A in PDB entry 4AKE (4AKE:A) [29] and 1AKE:A [30], respectively. In the RBP case, the open and closed structures are 1BA2:A [31] and 2DRI:A [32], respectively. In addition, we used several experimental intermediate structures of ADK whose PDB code 1ZIN, 1ZIO, 1ZIP [33], and 1DVR [34] to experimentally evaluate the ICONGENI simulation results. Because these intermediate structures have similar conformations, but different sequences with the reference structures (*i.e.*, 4AKE:A and 1AKE:A), homology modeling was implemented using Modeller v9.25 [35]. In detail, 10 candidate models of the intermediate structures were constructed by using their 3D conformations (as templates) and the 2D sequences of the reference structures (as target proteins), and the best models for each template were selected based on DOPE score [36]. The selected structures were refined by energy minimization (500 steps of conjugate gradient) using the CHARMM36m force field [37].

2.2 Internal coordinate normal mode-guided elastic network interpolation (ICONGENI)

The ultimate goal of ICONGENI is to predict pathways for large-scale conformational changes of macromolecules based on structural information on two end-point (*i.e.*, initial and final) conformations. To do this, valid displacement vectors can be obtained iteratively through ICONGENI, leading to determination of the consecutive intermediate conformations that comprise the pathways. A brief explanation of its algorithm is as follows.

First, IC-NMA is required prior to any procedure because we assume that the displacement vectors are linear combinations of low-frequency normal mode vectors. Next, the cost function is constructed to compute the degree of difference in interatomic distance between the resulting and desired conformations. Using the cost function, we devise compromise solutions (*i.e.*, a series of weights assigned to the normal modes of the displacement vectors) between minimizing the values of the cost function and constraining the degrees of freedom (DOFs) of the structural dynamics with some low-frequency mode shapes. By repeating these cycles, promising transition pathways can be predicted. A detailed explanation of ICONGENI will be introduced in the following subsections.

2.2.1 Elastic network model (ENM). ICs characterize molecular geometry using bond lengths, bond angles, and torsion angles terms to facilitate a better understanding of the

structural dynamics of molecules. While the backbone bond lengths and bond angles are nearly fixed in molecular systems, some torsion angles (ϕ and ψ angles in the protein conformation) can vary, and their dynamics exert a strong influence on the large-scale conformational changes. To effectively describe molecular systems in ICs, we use a coarse-grained modeling method, the elastic network model (ENM), wherein specific atoms of a protein backbone (*i.e.*, N , C_{α} , and C) are sampled and linked by a unit spring constant. The spring constant k matrix is defined as

$$k_{i,j} = \begin{cases} 0; & d > d_{cutoff} \\ 1; & d \leq d_{cutoff} \end{cases} \tag{1}$$

where $k_{i,j}$ is a binary spring constant between atoms i and j , d is an actual distance between them, and d_{cutoff} is a cutoff distance set to be 12 Å [19].

2.2.2 Normal mode analysis in internal coordinates (IC-NMA). For IC-NMA calculation, the second derivatives of the kinetic and potential energy functions are required. The main strategy is to calculate the functions with respect to CCs and then to transform them into those with respect to ICs. All calculations of IC-NMA will be described in this section with respect to previous works [38–41].

2.2.2.1 Transformation from Cartesian to internal coordinates. To obtain the second derivatives of the kinetic and potential energy functions, the first derivatives of CCs with respect to ICs must be defined. Only torsion angles are regarded as variables for the calculation, while the bond lengths and bond angles are considered to be fixed. The following derivation of the first derivatives will be based on some assumptions. First, the Eckart condition is assumed to separate the internal and external motions because the ICs cannot express external motions [42]. For the position vector \mathbf{r}_i of atom i , the condition can be satisfied by the following equations:

$$\sum_i m_i d\mathbf{r}_i = 0 \tag{2}$$

$$\sum_i m_i \mathbf{r}_i^{ref} \times d\mathbf{r}_i = 0 \tag{3}$$

where m_i and \mathbf{r}_i^{ref} are the mass and fixed position vector of atom i , respectively, in the reference conformation.

Next, let θ_α be a torsion angle around chemical bond α and domains A and B be the two domains divided by bond α as shown in Fig 1. Then, the two domains are regarded as rigid bodies, based on which Eq (2) can be rewritten as

$$M_A d\mathbf{r}_A + M_B d\mathbf{r}_B = 0 \tag{4}$$

where $M_A = \sum_i^A m_i$, $M_B = \sum_i^B m_i$, and \mathbf{r}_A and \mathbf{r}_B are the position vectors of the center of mass of domains A and B , respectively.

If $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$ are the rotation vectors of domains A and B , respectively, their relative rotation vectors $\boldsymbol{\omega}_{AB}$ and $\delta\mathbf{r}_i$ are defined as follows:

$$\boldsymbol{\omega}_{AB} = d\theta_\alpha \mathbf{e}_\alpha = \boldsymbol{\omega}_B - \boldsymbol{\omega}_A \tag{5}$$

$$d\mathbf{r}_i = \begin{cases} d\mathbf{r}_A + \boldsymbol{\omega}_A \times (\mathbf{r}_i^{ref} - \mathbf{r}_A^{ref}); & i \in A \\ d\mathbf{r}_B + \boldsymbol{\omega}_B \times (\mathbf{r}_i^{ref} - \mathbf{r}_B^{ref}); & i \in B \end{cases} \tag{6}$$

where $\mathbf{e}_\alpha = \frac{\mathbf{r}_{i(z)} - \mathbf{r}_{i(z)-1}}{\|\mathbf{r}_{i(z)} - \mathbf{r}_{i(z)-1}\|}$.

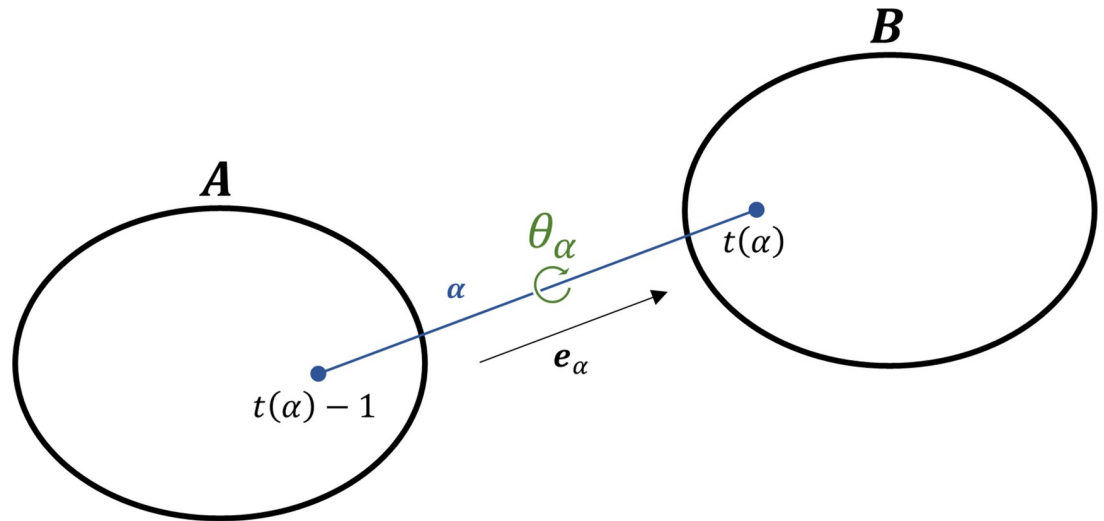


Fig 1. Schematic of a molecular system composed of two rigid bodies A and B with a chemical bond α . The relative displacement between the two bodies can be defined by the torsion angle θ_α around the bond α . If a bond α links atoms $(i-1)$ to i , $t(\alpha)$ designates i .

<https://doi.org/10.1371/journal.pone.0258818.g001>

Using Eq (5), $d\mathbf{r}_i$ in domain B (corresponding to the second equation of Eq (6)) can be rewritten as

$$d\mathbf{r}_i = d\mathbf{r}_A + \boldsymbol{\omega}_A \times (\mathbf{r}_i^{ref} - \mathbf{r}_A^{ref}) + d\theta_\alpha \mathbf{e}_\alpha \times (\mathbf{r}_i^{ref} - \mathbf{r}_{t(\alpha)}^{ref}); \quad i \in B \tag{7}$$

From Eqs (6) and (7),

$$d\mathbf{r}_A + \boldsymbol{\omega}_A \times (\mathbf{r}_{t(\alpha)}^{ref} - \mathbf{r}_A^{ref}) = d\mathbf{r}_B + \boldsymbol{\omega}_B \times (\mathbf{r}_{t(\alpha)}^{ref} - \mathbf{r}_B^{ref}) \tag{8}$$

Subsequently, using Eqs (3) and (6) and the concept of angular momentum,

$$M_A \mathbf{r}_A \times (d\mathbf{r}_A - \boldsymbol{\omega}_A \times \mathbf{r}_A^{ref}) + M_B \mathbf{r}_B \times (d\mathbf{r}_B - \boldsymbol{\omega}_B \times \mathbf{r}_B^{ref}) + \mathbf{I}_A \boldsymbol{\omega}_A + \mathbf{I}_B \boldsymbol{\omega}_B = 0 \tag{9}$$

where the inertia tensors are given by $\mathbf{I}_A = \sum_i^A m_i \|\mathbf{r}_i^{ref}\|^2 \mathbf{E}_3 - \mathbf{r}_i^{ref} (\mathbf{r}_i^{ref})^T$ and $\mathbf{I}_B = \sum_i^B m_i \|\mathbf{r}_i^{ref}\|^2 \mathbf{E}_3 - \mathbf{r}_i^{ref} (\mathbf{r}_i^{ref})^T$. \mathbf{E}_3 is the 3x3 identity matrix.

From Eqs (4), (5), and (8),

$$d\mathbf{r}_A - \boldsymbol{\omega}_A \times \mathbf{r}_A^{ref} = d\theta_\alpha \mathbf{e}_\alpha \times \left\{ \left(1 - \frac{M_A}{M}\right) \mathbf{r}_{t(\alpha)}^{ref} + \frac{M_A}{M} \mathbf{r}_A^{ref} \right\} \tag{10}$$

where $M = M_A + M_B$.

Using Eqs (9) and (10), $\boldsymbol{\omega}_A$ is expressed as

$$\boldsymbol{\omega}_A = -d\theta_\alpha [\mathbf{I}^{-1} \{ M_A \mathbf{r}_A^{ref} \times (\mathbf{e}_\alpha \times \mathbf{r}_{t(\alpha)}^{ref}) + (\mathbf{I} - \mathbf{I}_A) \mathbf{e}_\alpha \}] \tag{11}$$

where $\mathbf{I} = \mathbf{I}_A + \mathbf{I}_B$.

By substituting Eqs (10) and (11) into Eq (6), the derivative of CCs with respect to ICs is

$$\frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} = \begin{cases} \mathbf{e}_\alpha \times \left\{ \left(1 - \frac{M_A}{M} \right) \mathbf{r}_{t(\alpha)}^{ref} + \frac{M_A}{M} \mathbf{r}_A^{ref} \right\} - \mathbf{I}^{-1} \left\{ M_A \mathbf{r}_A^{ref} \times (\mathbf{e}_\alpha \times \mathbf{r}_{t(\alpha)}^{ref}) + (\mathbf{I} - \mathbf{I}_A) \mathbf{e}_\alpha \right\} \times \mathbf{r}_i^{ref}; & i \in A \\ -\mathbf{e}_\alpha \times \left\{ \left(1 - \frac{M_B}{M} \right) \mathbf{r}_{t(\alpha)}^{ref} + \frac{M_B}{M} \mathbf{r}_B^{ref} \right\} + \mathbf{I}^{-1} \left\{ M_B \mathbf{r}_B^{ref} \times (\mathbf{e}_\alpha \times \mathbf{r}_{t(\alpha)}^{ref}) + (\mathbf{I} - \mathbf{I}_B) \mathbf{e}_\alpha \right\} \times \mathbf{r}_i^{ref}; & i \in B \end{cases} \quad (12)$$

Finally, these equations can be rewritten in the form of matrix-vector multiplication:

$$\frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} = \begin{cases} (\mathbf{P}_i \ \mathbf{E}_3) \begin{bmatrix} \mathbf{I}^{-1} \mathbf{I}_B & \mathbf{I}^{-1} (\mathbf{P}_B)^T \\ \frac{\mathbf{P}_B}{M} & \frac{M_B}{M} \mathbf{E}_3 \end{bmatrix} \mathbf{d}_\alpha; & i \in A \\ -(\mathbf{P}_i \ \mathbf{E}_3) \begin{bmatrix} \mathbf{I}^{-1} \mathbf{I}_A & \mathbf{I}^{-1} (\mathbf{P}_A)^T \\ \frac{\mathbf{P}_A}{M} & \frac{M_A}{M} \mathbf{E}_3 \end{bmatrix} \mathbf{d}_\alpha; & i \in B \end{cases} \quad (13)$$

where $\mathbf{P}_i = \mathbf{r}_i \times = \begin{bmatrix} 0 & -z_i & y_i \\ z_i & 0 & -x_i \\ -y_i & x_i & 0 \end{bmatrix}$, $\mathbf{P}_A = \sum_i^A m_i \mathbf{P}_i$, $\mathbf{P}_B = -\mathbf{P}_A$, and $\mathbf{d}_\alpha = [\mathbf{e}_\alpha \ \mathbf{e}_\alpha \times \mathbf{r}_{t(\alpha)}]^T$.

2.2.2.2 Construction of the second derivative of kinetic energy. As shown in Fig 2, the kinetic energy T of the molecular system in ICs can be calculated using two torsion angle variables θ_α and θ_β as follows:

$$T = \frac{1}{2} \sum_i m_i \dot{\mathbf{r}}_i^2 = \frac{1}{2} \sum_{\alpha, \beta} \underbrace{\left(\sum_i m_i \frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} \cdot \frac{\partial \mathbf{r}_i}{\partial \theta_\beta} \right)}_{M_{\alpha, \beta}} \dot{\theta}_\alpha \dot{\theta}_\beta \quad (14)$$

where $M_{\alpha, \beta}$ is the second derivative of kinetic energy for bonds α and β .

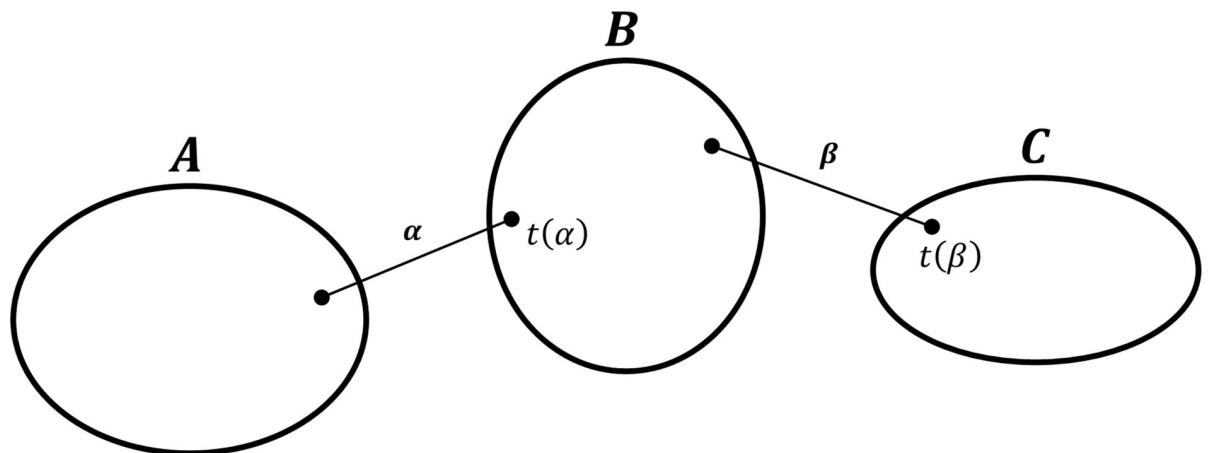


Fig 2. Schematic of a molecular system composed of three rigid bodies A, B, and C with two chemical bonds α and β . If the bond α links atoms $(i-1)$ to i , $t(\alpha)$ designates i .

<https://doi.org/10.1371/journal.pone.0258818.g002>

Then, using Eq (13), $M_{\alpha,\beta}$ can be defined as

$$\begin{aligned} \mathbf{M}_{\alpha,\beta} &= (\mathbf{M}_{\alpha,\beta})_A + (\mathbf{M}_{\alpha,\beta})_B + (\mathbf{M}_{\alpha,\beta})_C \\ &= (\mathbf{d}_\alpha)^T \begin{pmatrix} \mathbf{I}_A \mathbf{I}^{-1} \mathbf{I}_C + \frac{(\mathbf{P}_A)^T \mathbf{P}_C}{M} & \mathbf{I}_A \mathbf{I}^{-1} (\mathbf{P}_C)^T + \frac{M_C}{M} (\mathbf{P}_A)^T \\ \mathbf{P}_A \mathbf{I}^{-1} \mathbf{I}_C + \frac{M_A}{M} \mathbf{P}_C & \mathbf{P}_A \mathbf{I}^{-1} (\mathbf{P}_C)^T + \frac{M_A M_C}{M} \mathbf{E}_3 \end{pmatrix} \mathbf{d}_\beta \end{aligned} \quad (15)$$

where

$$\begin{aligned} (\mathbf{M}_{\alpha,\beta})_A &= \sum_i^A m_i \frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} \cdot \frac{\partial \mathbf{r}_i}{\partial \theta_\beta} \\ &= (\mathbf{d}_\alpha)^T \begin{pmatrix} (\mathbf{I} - \mathbf{I}_A) \mathbf{I}^{-1} & \frac{-(\mathbf{P}_A)^T}{M} \\ -\mathbf{P}_A \mathbf{I}^{-1} & \frac{M - M_A \mathbf{E}_3}{M} \end{pmatrix} \begin{pmatrix} \mathbf{I}_A & (\mathbf{P}_A)^T \\ \mathbf{P}_A & M_A \mathbf{E}_3 \end{pmatrix} \begin{pmatrix} \mathbf{I}^{-1} \mathbf{I}_C & \mathbf{I}^{-1} (\mathbf{P}_C)^T \\ \frac{\mathbf{P}_C}{M} & \frac{M_C}{M} \mathbf{E}_3 \end{pmatrix} \mathbf{d}_\beta, \end{aligned}$$

$$(\mathbf{M}_{\alpha,\beta})_B = -(\mathbf{d}_\alpha)^T \begin{pmatrix} \mathbf{I}_A \mathbf{I}^{-1} & \frac{(\mathbf{P}_A)^T}{M} \\ \mathbf{P}_A \mathbf{I}^{-1} & \frac{M_A}{M} \mathbf{E}_3 \end{pmatrix} \begin{pmatrix} \mathbf{I}_B & (\mathbf{P}_B)^T \\ \mathbf{P}_B & M_B \mathbf{E}_3 \end{pmatrix} \begin{pmatrix} \mathbf{I}^{-1} \mathbf{I}_C & \mathbf{I}^{-1} (\mathbf{P}_C)^T \\ \frac{\mathbf{P}_C}{M} & \frac{M_C}{M} \mathbf{E}_3 \end{pmatrix} (\mathbf{d}_\beta)^T,$$

and

$$(\mathbf{M}_{\alpha,\beta})_C = (\mathbf{d}_\alpha)^T \begin{pmatrix} \mathbf{I}_A \mathbf{I}^{-1} & \frac{(\mathbf{P}_A)^T}{M} \\ \mathbf{P}_A \mathbf{I}^{-1} & \frac{M_A}{M} \mathbf{E}_3 \end{pmatrix} \begin{pmatrix} \mathbf{I}_C & (\mathbf{P}_C)^T \\ \mathbf{P}_C & M_C \mathbf{E}_3 \end{pmatrix} \begin{pmatrix} \mathbf{I}^{-1} (\mathbf{I} - \mathbf{I}_C) & -\mathbf{I}^{-1} (\mathbf{P}_C)^T \\ -\frac{\mathbf{P}_C}{M} & \frac{M - M_C}{M} \mathbf{E}_3 \end{pmatrix}.$$

2.2.2.3 Construction of the second derivative of potential energy. Basically, the potential energy of a molecular system is defined based on interatomic interactions (e.g., van der Waals bond and covalent bond potentials). With the aid of a simple assumption that one rigid-body domain is fixed instead of the Eckart condition, we can simply reformulate $\frac{\partial \mathbf{r}_i}{\partial \theta_\alpha}$ against Eq (13). If domain A is fixed in Fig 1, the following equations are satisfied:

$$\frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} = \begin{cases} 0; & i \in A \\ -(\mathbf{P}_i \mathbf{E}_3) \mathbf{d}_\alpha; & i \in B \end{cases} \quad (16)$$

The second derivatives of potential energy in ICs can be derived from those in CCs as

$$\mathbf{H}_{\alpha,\beta} = \frac{\partial^2 V}{\partial \theta_\alpha \partial \theta_\beta} = \sum_{i,j} \frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} \cdot \underbrace{\frac{\partial^2 V}{\partial \mathbf{r}_i \partial \mathbf{r}_j}}_{\mathbf{H}_{i,j}^c} \cdot \frac{\partial \mathbf{r}_j}{\partial \theta_\beta} \quad (17)$$

where V is the potential energy function, and $\mathbf{H}_{\alpha,\beta}$ ($\mathbf{H}_{i,j}^c$) is the second derivative of potential energy for the ICs θ_α and θ_β (for the CCs \mathbf{r}_i and \mathbf{r}_j).

H_{ij}^c is defined simply by the following function of interatomic distances [19]:

$$H_{ij}^c = \frac{\partial^2 V}{\partial \mathbf{r}_i \partial \mathbf{r}_j} = - \frac{k_{ij}}{\|\mathbf{r}_i - \mathbf{r}_j\|^2} (\mathbf{r}_i - \mathbf{r}_j) (\mathbf{r}_i - \mathbf{r}_j)^T \tag{18}$$

Then, supposing that domain B is fixed in the diagram of Fig 2 and using Eqs (16) and (18), $H_{\alpha,\beta}$ can be formulated as

$$H_{\alpha,\beta} = (\mathbf{d}_\alpha)^T \left\{ \sum_i^A \sum_j^C \frac{k_{ij}}{\|\mathbf{r}_{ij}\|^2} \begin{bmatrix} \mathbf{r}_i \times \mathbf{r}_j \\ \mathbf{r}_i - \mathbf{r}_j \end{bmatrix} \begin{bmatrix} (\mathbf{r}_i \times \mathbf{r}_j)^T & (\mathbf{r}_i - \mathbf{r}_j)^T \end{bmatrix} \right\} \mathbf{d}_\beta \tag{19}$$

Using Eqs (15) and (19), we can construct the equation of motion with respect to ICs. For a molecular system having n torsion angle dynamics,

$$M\ddot{\boldsymbol{\theta}} + H\boldsymbol{\theta} = 0 \tag{20}$$

where $\boldsymbol{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_n]$, $M = \begin{bmatrix} M_{1,1} & \dots & M_{1,n} \\ \vdots & \ddots & \vdots \\ M_{n,1} & \dots & M_{n,n} \end{bmatrix}$, and $H = \begin{bmatrix} H_{1,1} & \dots & H_{1,n} \\ \vdots & \ddots & \vdots \\ H_{n,1} & \dots & H_{n,n} \end{bmatrix}$.

We can obtain normal modes (*i.e.*, pairs of eigenvalues and eigenvectors) by solving a generalized eigenvalue problem for Eq (20). Next, an extra calculation is required to transform the resulting eigenvectors in ICs to those in CCs. Therefore, the final form of the normal mode vectors can be determined by the following equation [41]:

$$\Delta \mathbf{r}_{i,k} = \sum_\alpha \frac{\partial \mathbf{r}_i}{\partial \theta_\alpha} \Delta \theta_{\alpha,k} \tag{21}$$

where $\Delta \mathbf{r}_{i,k}$ ($\Delta \theta_{\alpha,k}$) is the k^{th} normal mode vector of atom i (of bond α) with respect to CCs (ICs).

2.2.3 Construction of the cost function. The cost function is defined as a function of error in interatomic distances between the simulated conformations and the desired ones [25, 43]:

$$C(\boldsymbol{\delta}) = \frac{1}{2} \sum_{ij} k_{ij} \{ \|\mathbf{r}_i + \boldsymbol{\delta}_i - \mathbf{r}_j + \boldsymbol{\delta}_j\| - q_{ij} \}^2 \tag{22}$$

where $\boldsymbol{\delta}_i$ is the displacement vector of atom i describing conformational changes, and q_{ij} is the desired distance between atoms i and j .

q_{ij} in a target intermediate can be determined through linear interpolation between the two end-point conformations as

$$q_{ij} = (1 - s) \|\mathbf{r}_i^0 - \mathbf{r}_j^0\| + s \|\mathbf{r}_i^1 - \mathbf{r}_j^1\| \tag{23}$$

where \mathbf{r}_i^0 and \mathbf{r}_i^1 represent the position vectors of atom i for the initial and final conformations, respectively. s is a proportional representation of the location of an intermediate to be simulated on the pathway when the total length of the pathway is set to 1.

In accordance with the strategy of ICONGENI, $\boldsymbol{\delta}_i$ can be represented as the linear combination of a set of low-frequency normal mode vectors for reference conformations:

$$\boldsymbol{\delta}_i = w_1 \Delta \mathbf{r}_{i,1} + w_2 \Delta \mathbf{r}_{i,2} + \dots + w_m \Delta \mathbf{r}_{i,m} \tag{24}$$

where m denotes the number of low-frequency normal modes used in the simulation, and w_n is the weight of the n^{th} normal mode vector.

Using Eq (24), Eq (22) can be rewritten as

$$C(\mathbf{W}) = \frac{1}{2} \sum_{i,j} k_{ij} \{ \|(\mathbf{r}_i - \mathbf{r}_j) + (\mathbf{V}_i - \mathbf{V}_j)\mathbf{W}\| - q_{ij} \}^2 \tag{25}$$

where $\mathbf{V}_i = [\Delta\mathbf{r}_{i,1} \ \Delta\mathbf{r}_{i,2} \ \dots \ \Delta\mathbf{r}_{i,m}] \in \mathbb{R}^{3 \times m}$, and $\mathbf{W} = [w_1 \ w_2 \ \dots \ w_m]^T \in \mathbb{R}^{m \times 1}$.

Then, to find the value of \mathbf{W} that minimizes $C(\mathbf{W})$, we simplify Eq (25) into the form of matrix-vector multiplication using a Taylor series expansion:

$$\begin{aligned} & \|(\mathbf{r}_i - \mathbf{r}_j) + (\mathbf{V}_i - \mathbf{V}_j)\mathbf{W}\| \\ & \approx \|\mathbf{r}_i - \mathbf{r}_j\| + \frac{(\mathbf{r}_i - \mathbf{r}_j)^T (\mathbf{V}_i - \mathbf{V}_j)\mathbf{W}}{\|\mathbf{r}_i - \mathbf{r}_j\|} + \frac{\mathbf{W}^T (\mathbf{V}_i - \mathbf{V}_j)^T \mathbf{A}(\mathbf{r}_i - \mathbf{r}_j) \{(\mathbf{V}_i - \mathbf{V}_j)\mathbf{W}\}}{2\|\mathbf{r}_i - \mathbf{r}_j\|} \end{aligned} \tag{26}$$

where $\mathbf{A}(\mathbf{r}_i - \mathbf{r}_j) = \mathbf{E}_3 - \frac{(\mathbf{r}_i - \mathbf{r}_j)(\mathbf{r}_i - \mathbf{r}_j)^T}{\|\mathbf{r}_i - \mathbf{r}_j\|^2}$.

Then, Eq (27) can be written in the form of matrix-vector multiplication:

$$C(\mathbf{W}) = \mathbf{W}^T \mathbf{P}^{(1)} \mathbf{W} + \mathbf{P}^{(2)} \mathbf{W} + \mathbf{P}^{(3)} \tag{27}$$

where

$$\mathbf{P}^{(1)} = \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij} (\mathbf{V}_i - \mathbf{V}_j)^T \left\{ \mathbf{E}_3 - q_{ij} \frac{\mathbf{A}(\mathbf{r}_i - \mathbf{r}_j)}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right\} (\mathbf{V}_i - \mathbf{V}_j),$$

$$\mathbf{P}^{(2)} = \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij} \left(2 - \frac{2q_{ij}}{\|\mathbf{r}_i - \mathbf{r}_j\|} \right) (\mathbf{r}_i - \mathbf{r}_j)^T (\mathbf{V}_i - \mathbf{V}_j),$$

and

$$\mathbf{P}^{(3)} = \frac{1}{2} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_{ij} \{ (\mathbf{r}_i - \mathbf{r}_j)^T (\mathbf{r}_i - \mathbf{r}_j) - 2q_{ij} \|\mathbf{r}_i - \mathbf{r}_j\| + q_{ij}^2 \}.$$

Finally, the optimal displacement vectors can be determined by solving for \mathbf{W} from the following equation:

$$\frac{\partial C(\mathbf{W})}{\partial \mathbf{W}} = 2\mathbf{P}^{(1)} \mathbf{W} + \mathbf{P}^{(2)} = 0 \tag{28}$$

The optimal displacement vector allows us to determine the intermediate conformation on the pathway. By repeating this calculation process using the simulated intermediate as a new reference, the transition pathways from initial to final conformations can be generated. The number of iteration steps was determined so that consecutive intermediate structures differ by a root-mean-square deviation (RMSD) of about 0.1 Å. Here, we set the number of iteration steps for the case of ADK (RBP) is set to 71 (62) because the RMSD value between the two end-point conformations is 7.16 Å (6.25 Å).

3. Results and discussion

3.1 Comparing the ICONGENI pathways to NGENI pathways

In this section, we discuss the effectiveness of ICONGENI by comparing the resulting pathways to those developed by NGENI [25] under the same conditions. Because the main

difference between the two techniques is the coordinate space in which the NMA is performed (*i.e.*, ICONGENI and NGENI are based on IC-NMA and CC-NMA, respectively), it is expected that this comparative analysis will demonstrate the superiority of IC-NMA in describing large deformations of proteins. We performed ICONGENI and NGENI for two proteins: ADK and RBP. ADK as a phosphotransferase enzyme catalyzing the reaction $\text{ATP} + \text{AMP} \rightleftharpoons 2 \text{ADP}$ is composed of three domains: CORE, NMP, and LID, and undergoes two pairs of hinge motions of NMP and LID relative to CORE to fulfill its function [44]. RBP as one of the periplasmic binding proteins binds ribose through a hinge motion of two domains, which enables cells to sense and transport the ligand [31]. To predict the transition from open state to closed state, we obtained their 3D structures from PDB; the open and closed structures of ADK are chain A in PDB entry 4AKE:A and 1AKE:A, respectively, and those of RBP are 1BA2:A and 2DRI:A, respectively. The DOFs of these simulations were set to be the 50 lowest normal modes considered empirically sufficient to simulate the conformational changes within the experimental resolution based on our previous study [25]. For better understanding, the transition pathways explored by ICONGENI of ADK and RBP are provided in **S1** and **S2 Movies**, respectively.

First, the convergence issue of the pathways was addressed. To assess geometric convergence, we measured the RMSDs of the consecutive structures comprising the paths with respect to the final conformation (*i.e.*, the closed structures) and judged that the paths satisfied the convergence condition if the RMSD values steadily decreased below certain thresholds that is selected as the smaller of the experimental resolutions of two reference structures (*i.e.*, open and closed structures). As shown in **Fig 3A**, the ADK pathways of the two techniques had similar graphs and converged below the value of the resolution. Similarly, their RBP pathways also got close to the final conformation at a level beyond the resolution (**Fig 3B**). This confirmed that both techniques had no problem in terms of pathway convergence when generating the transition pathways based on the DOFs of the 50 lowest normal modes.

Next, we investigated backbone bond length and bond angle distributions on the simulated pathways. Proteins undergo conformational changes mainly through variations of two types of backbone torsion angles: ϕ around the $N-C_{\alpha}$ bond and ψ around the $C_{\alpha}-C$ bond while variations of backbone bond lengths and bond angles are impractical during the transitions. The distributions of the bond lengths and bond angles provide key information to evaluate how well the proteins keep their molecular shape during large deformation. First, the backbone bond lengths are divided into three types: $N-C_{\alpha}$, $C_{\alpha}-C$, and $C-N$. According to the type, we calculated average (avg) and standard deviation (std) values over ICONGENI and NGENI pathways for two proteins (ADK and RBP) and analyzed their distributions using the experimental data as the reference avg and std values of the backbone bond length types (**Fig 4**) [45]. The avg values of the bond length in the ICONGENI paths were more concentrated around the corresponding experimental values than were those in the NGENI paths. Moreover, most std values of the bond lengths in the ICONGENI paths were distributed below the corresponding experimental values while many of those in the NGENI paths were higher than the experimental values. Subsequently, we investigated the backbone bond angle distributions (including $N-C_{\alpha}-C$, $C_{\alpha}-C-N$, and $C-N-C_{\alpha}$) of the simulated pathways in the same manner as described above for analysis of the bond length distributions. Similar to the results in the bond length distribution graphs, the avg values of the bond angles in the ICONGENI paths were densely distributed around the corresponding experimental values, and the std values were distributed close to zero compared to the NGENI pathways (**Fig 5**). These quantitative data commonly showed that the ICONGENI pathways were more likely to prevent unrealistic distortions in bond lengths and bond angles than were the NGENI pathways. On the other hand, there were exceptions to this principle, such as the small number of bond lengths in the

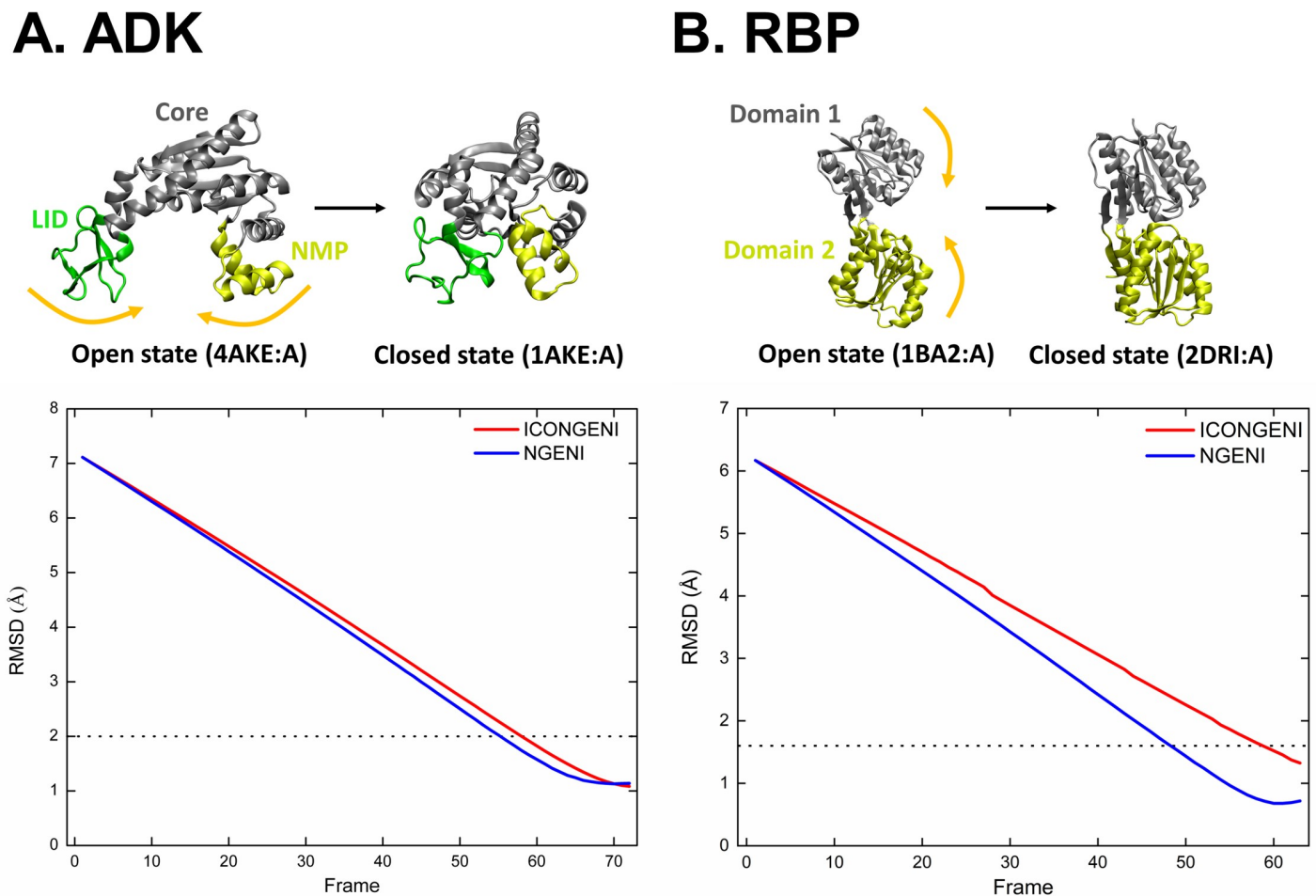


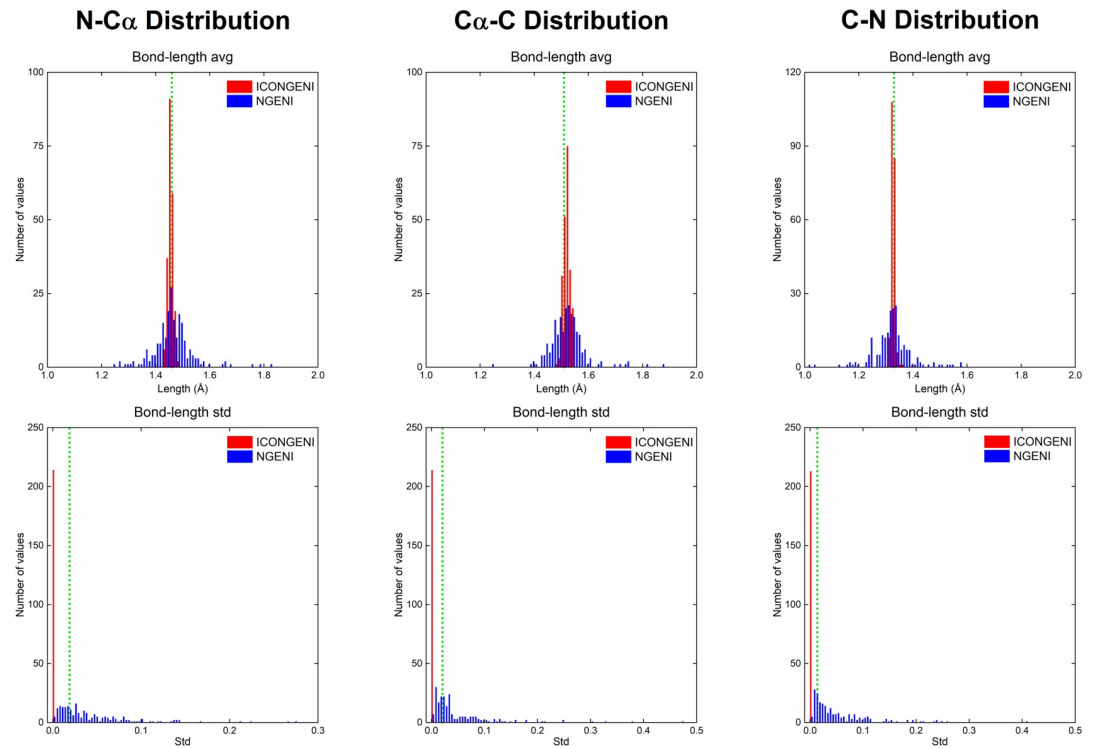
Fig 3. Conformational transition from open to closed states of (A) ADK and (B) RBP. Upper figures represent crystal structures of open and closed states of the proteins. ADK is composed of three domains: CORE (residues 1–29, 60–121, and 160–214), NMP (residues 30–59), and LID (residues 122–159). RBP is composed of Domain 1 (residues 1–103 and 236–264) and Domain 2 (residues 104–235 and 265–271). The lower graphs describe the convergence of simulated pathways of each protein by measuring changes in RMSD between predicted intermediates and the final structure. The results of ICONGENI and NGENI pathways are shown as red and blue lines, respectively. The black dotted lines represent the corresponding experimental resolution of the proteins.

<https://doi.org/10.1371/journal.pone.0258818.g003>

ICONGENI paths with irrational avg or std values that deviated farther from the corresponding experimental ones than did those of the NGENI path (denoted by pink circles in Fig 4B). We confirmed that all bond lengths that fell under these exceptions belonged to either the first or last residue of the proteins, which would imply that these unrealistic distortions were caused by the tip effect [41]. The tip effect is an inherent weakness of NMA (regardless of the type of coordinate system) in which the highly flexible parts in protein structures (*e.g.*, hanging loops and protruding ends) exhibit abnormal behavior in some of the lowest normal modes. IC-NMA may suffer more from the tip effect than did CC-NMA because the mode shapes in IC-NMA is intrinsically limited in describing the movements of either the first or last residue where any torsion angle cannot be defined, which can explain the exceptions in Fig 4B. However, this limitation of ICONGENI is not a critical issue in predicting the transition pathways because the distortions of the tip parts could be considered as local vibrations that has little effect on the global motions.

In the same context as investigating the bond length and bond angle distributions, the potential energies of the intermediate conformations comprising the resulting pathways were

A. ADK



B. RBP

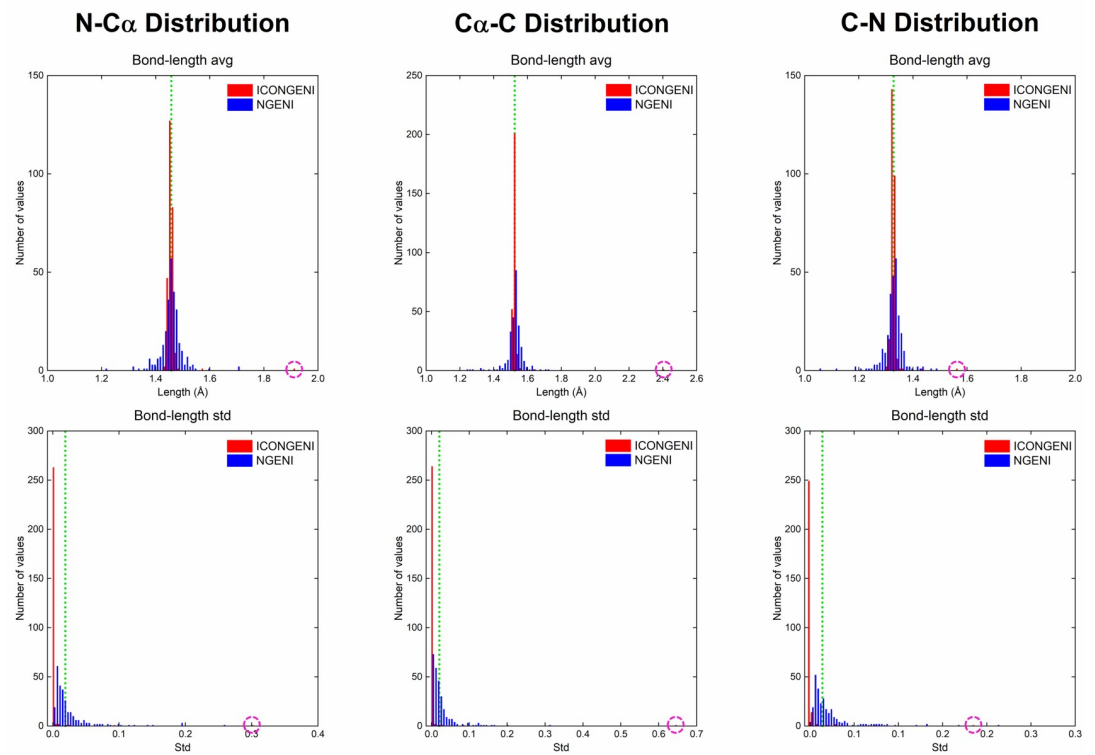


Fig 4. Bond length distribution of the transition pathways of (A) ADK and (B) RBP. Comparison of the distributions of the avg and std values of backbone bond lengths in ICONGENI (denoted by red) and NGENI (denoted by blue) pathways. Both methods explore the transition pathways based on the DOFs of the 50 lowest normal modes. The bond length distributions are measured for three backbone bond length coordinates: $N-C_{\alpha}$, $C_{\alpha}-C$, and $C-N$. The green lines represent the corresponding experimental values of the coordinates. The pink circles represent specific cases where the ICONGENI pathway has irrational values of bond length avg or std.

<https://doi.org/10.1371/journal.pone.0258818.g004>

calculated. Because both simulation methods were carried out by using a coarse-grained modeling method, non-backbone atoms and O atoms in the backbone from reference structures (4AKE:A and 1BA2:A for ADK and RBP, respectively) were grafted to all intermediates and the generated all atom models were energy minimized within CHARMM36m force field for 500 steps of conjugate gradient [37] to eliminate any steric clashes and inappropriate geometries. Next, we calculated the potential energies of all intermediates of the NGENI and ICONGENI pathways by using CHARMM36m force field to quantitatively evaluate how the corresponding transitions are stable. Fig 6 shows the difference of the potential energy of each frame between the NGENI and ICONGENI pathways. From the results, we confirmed that the ICONGENI pathways are generally more stable (*i.e.*, having lower potential energies) than the NGENI pathways. Furthermore, their energy gap increased as the pathways progress, which suggests that the qualitative difference between the pathways is increasingly noticeable in that the geometric errors are gradually accumulated when anharmonic transitions are explored by harmonic modes. Finally, these simulation results imply that the ICONGENI pathways are more reliable than the NGENI pathways in terms of thermal and chemical stability.

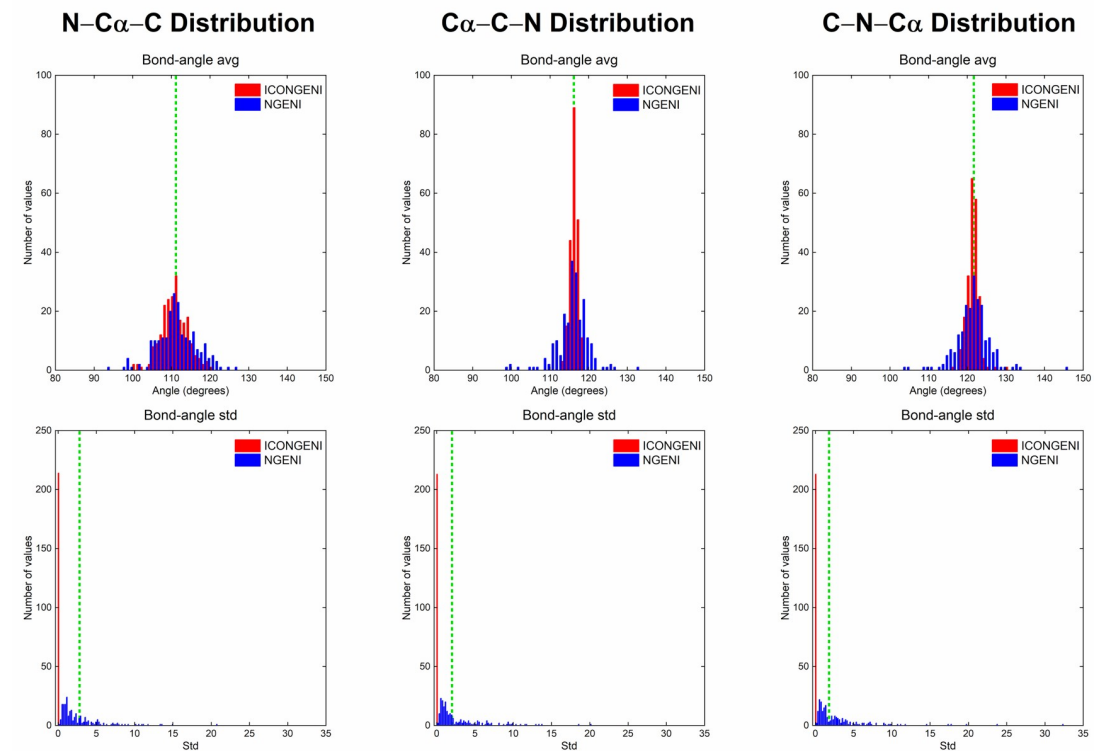
3.2 Predicting a transition pathway ensemble depending on a set of lowest normal modes

In the previous section, we tried to predict conformational transition pathways using ICONGENI with the 50 lowest normal modes considered sufficient to describe large deformation. Although the resulting pathways are shown to be reliable in terms of thermal and chemical stability, this study does not verify that ICONGENI can provide information on real transition trajectories. ICONGENI with the 50 lowest normal modes finds the deterministic and most effective pathways in terms of atomic displacements due to the intrinsic properties of the established cost function (see Section 2.3). In this section, we discuss the possibility that ICONGENI can predict plausible routes for conformational changes on complex energy landscapes by applying it to ADK of which transition mechanisms have been studied in numerous research works.

First, we generated an ensemble of the transition pathways of ADK through ICONGENI depending on the number of lowest normal modes (from 5 to 100), and their convergence was measured by RMSD with the closed state, as in the previous section. As shown in Fig 7A, the fewer are the normal modes used in the simulation, the less likely it is that the corresponding pathway reaches the final conformation. In detail, the pathways using fewer than 25 normal modes do not satisfy the convergence condition (*i.e.*, their RMSD from the final conformation does not converge under the experimental resolution of ADK). This is not surprising given that the progression of pathways is influenced strongly by the DOFs describing the structural motions. In addition, this result suggests that it is necessary to focus on the “incomplete” pathways simulated with relatively few normal modes, as molecular systems usually explore seemingly inefficient routes of conformational transitions to arrive at functional states over several high-energy barriers.

Although the detailed transition mechanisms of ADK remain to be elucidated, previous experimental and theoretical studies have proposed several pathways via the NMP-closing/

A. ADK



B. RBP

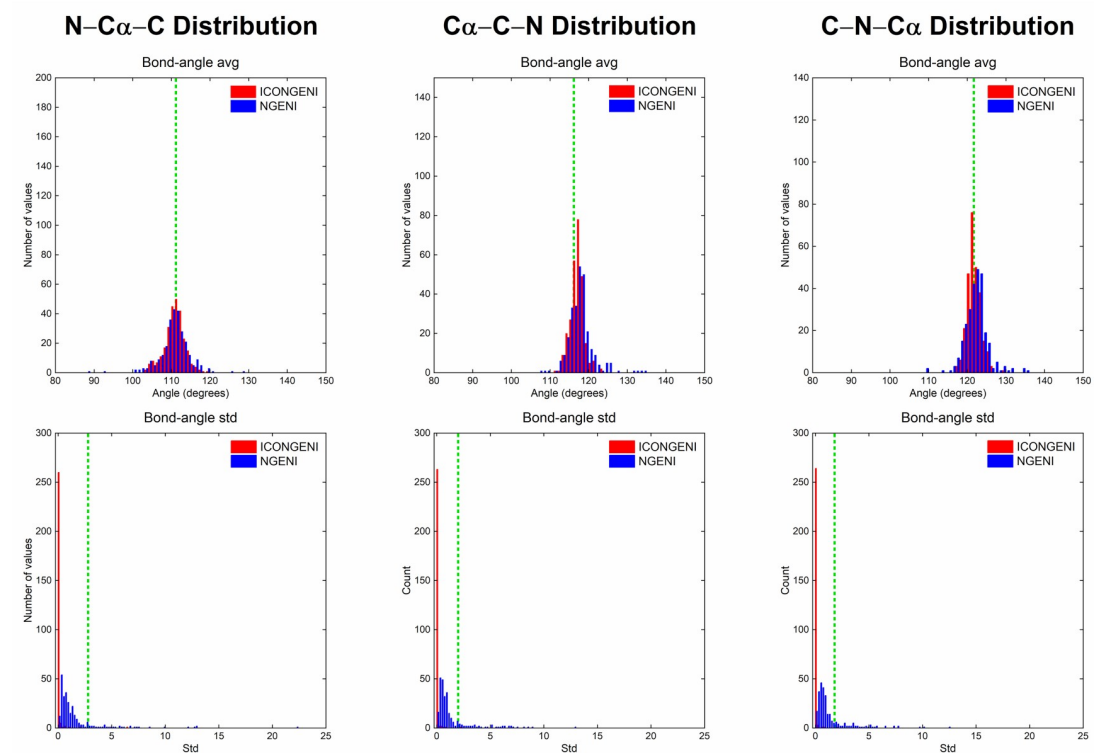
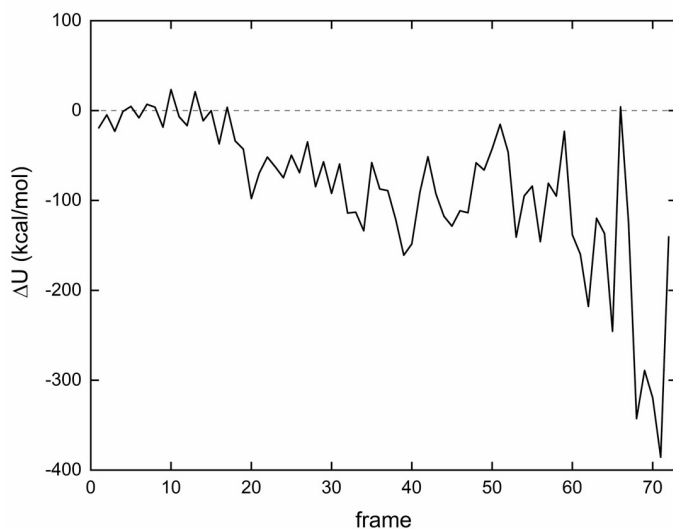


Fig 5. Bond angle distribution of the transition pathways of (A) ADK and (B) RBP. Comparison of the distributions of the avg and std values of backbone bond angles in ICONGENI (denoted by red) and NGENI (denoted by blue) pathways. Both methods explore the transition pathways based on the DOFs of the 50 lowest normal modes. The bond angle distributions are measured for backbone bond angle coordinates: $N-C_{\alpha}-C$, $C_{\alpha}-C-N$, and $C-N-C_{\alpha}$. The green lines represent the corresponding experimental values of the coordinates.

<https://doi.org/10.1371/journal.pone.0258818.g005>

LID-opening (NMP_C) state or the NMP-opening/LID-closing (NMP_O) state [33, 34, 46–52]. In other words, the large-scale transition of ADK is characterized by interdomain hinge motions of NMP and LID relative to CORE. To delineate the NMP and LID movements on the ICONGENI pathways, we projected them onto interdomain angle space with the NMP-CORE angle (θ_{NMP}) and the LID-CORE angle (θ_{LID}). θ_{NMP} (θ_{LID}) is defined by the centers of mass of the backbone including N , C_{α} , and C in residues 115–125, 90–100, and 35–55 (179–185, 115–125, and 125–153) in the notation used in a previous study [51]. In addition, we used some experimental structures for cross-validation with our simulation results. 4AKE:A and 1AKE:A defines the two end-point conformations of the transition. The crystal structures whose PDB codes are 1ZIN, 1ZIO, and 1ZIP [33] and those whose PDB code is 1DVR [34] approximate the NMP_C state and the NMP_O state, respectively. On $\theta_{NMP}-\theta_{LID}$ space, the pathway ensemble has a tendency: the larger is the number of normal modes used in the simulation, the straighter are the resulting pathways to the closed state (Fig 7B). This is not surprising given that the established cost function is designed to produce the most efficient and direct paths within the given DOFs. The straight path on the interdomain angle space refers to the trajectory at which NMP and LID simultaneously open during the transition but is not favorable in terms of the free energy landscapes [47, 48]. As the number of modes used in the simulation decreased to less than 35, the resulting pathways tended to closely approach the NMP_C state. When using significantly fewer normal modes (less than 10) for simulations, the corresponding pathways described the transition toward the more extreme NMP_C state than did the

A. ADK



B. RBP

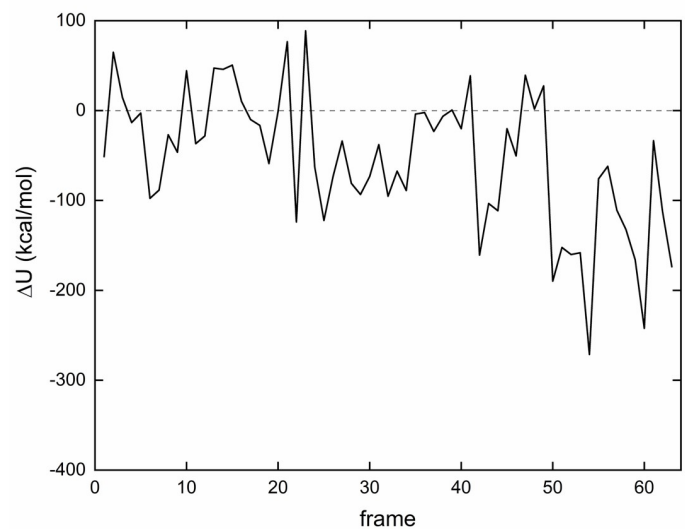


Fig 6. The difference between potential energies of ICONGENI and NGENI pathways for (A) ADK and (B) RBP. The difference of the potential energy $\Delta U = U_{\text{ICONGENI}} - U_{\text{NGENI}}$. Before calculating the potential energies, all simulated intermediate structures were transformed into all atom models based on corresponding reference structures (4AKE:A (1BA2:A) for ADK (RBP)) and were energy minimized using CHARMM36m force field for 500 steps of conjugate gradient.

<https://doi.org/10.1371/journal.pone.0258818.g006>

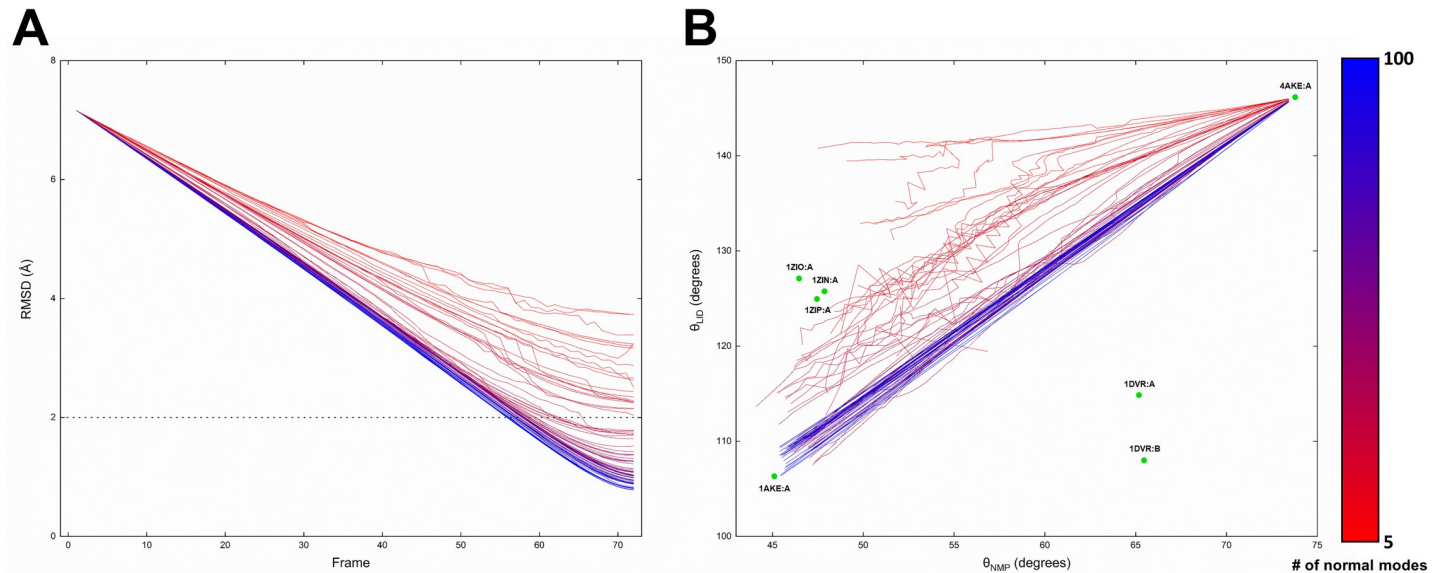


Fig 7. The pathway ensemble for ADK generated by ICONGENI. The ICONGENI transition pathways that make up the pathway ensemble are colored according to the lowest normal modes (from 5 to 100) used in the simulation (red to blue color scheme). (A) Convergence of the pathway ensemble. The RMSD values of each path relative to a final state are measured. The black dotted line represents the corresponding experimental resolution of ADK. (B) Projection of the pathway ensemble onto $\theta_{NMP}-\theta_{LID}$ space. The green points show the positions of experimental structures on $\theta_{NMP}-\theta_{LID}$ space. 4AKE:A and 1AKE:A indicate the open and closed states of ADK, respectively. 1ZIN:A, 1ZIO:A, and 1ZIP:A (1DVR:A, and 1DVR:B) indicate experimental structures at the NMP_C state (the NMP_O state).

<https://doi.org/10.1371/journal.pone.0258818.g007>

crystal structures approximating the NMP_C state (*i.e.*, 1ZIN:A, 1ZIO:A, and 1ZIP:A). This result implies that the vibrational features describing the dynamics of θ_{NMP} were preferentially arranged in the lowest normal modes, resulting from the flexibility between NMP and CORE is higher than that between LID and CORE. Therefore, we suggest that the open-to-closed transition via the NMP_C state is more plausible and reliable than that via the NMP_O state in terms of the vibrational characteristics of ADK. However, this result does not mean ICONGENI always returns a single candidate (*i.e.*, paths via the NMP_C state) of the transition paths. If the normal mode set as the system DOFs is determined under certain conditions, ICONGENI could explore transition pathways via the NMP_O state, which demonstrate that ICONGENI can explore multiple transition pathways compatible to several metastable states if information of the states is given (See more details in **S1 Text** and **S1 Fig**).

4 Conclusion

In this study, we introduced internal coordinate normal mode-guided elastic network interpolation (ICONGENI) as a theoretical method to explore the conformational transition pathways of proteins. By linearly interpolating the coarse-grained models of the two end-point states, ICONGENI defines virtual intermediate conformations of which the transition pathway is composed. Based on structural information, ICONGENI explores the optimal transition pathway (*i.e.*, the pathway minimizing a cost function showing the error between the simulated intermediates and the virtual ones). When iteratively obtaining the consecutive conformations describing the transition pathway, the key idea of the method is to represent the displacement vectors as a linear combination of lowest normal mode vectors produced by normal mode analysis in internal coordinates (IC-NMA). Given that IC-NMA can describe chemically relevant dynamics (suitable for describing large-scale transitions) compared to NMA in Cartesian coordinates (CC-NMA), this strategy enables the proposed method to explore reliable transition pathways in an efficient manner.

To evaluate the superiority of ICONGENI, we performed comparative studies of ICONGENI with our previous method based on CC-NMA (named NGENI). For two proteins: adenylate kinase (ADK) and ribose-binding protein (RBP), we predicted transition pathways through the two methods under the same conditions (using the 50 lowest normal modes as the system degrees of freedom). The distribution data of the bond lengths and bond angles of the resulting pathways confirmed that these coordinates remained highly stable in the ICONGENI pathways compared to those in the NGENI pathways (Figs 4 and 5). Furthermore, we also calculated the potential energies of the simulated pathways and identified the energies of the ICONGENI pathways were lower overall than those of the NGENI pathways (Fig 6). In conclusion, these results suggest that IC-NMA is suitable for representing realistic dynamics of the proteins, by extension, that ICONGENI could explore more reliable transition pathways than NGENI in terms of thermal and chemical stability.

Although ICONGENI using the degrees of freedom (DOFs) of the 50 or more lowest normal modes can provide a spatial understanding of conformational transitions, this approach is insufficient to explain the actual transition events on complex energy landscapes. To address this issue, we focused on a pathway ensemble for ADK simulated by ICONGENI. First, we confirmed that the more is the number of normal modes used in the simulation, the closer the initial structure is to the final one, which is not surprising because the number of normal modes directly indicates the DOFs to describe structural dynamics (Fig 7A). Next, we characterized the pathway ensemble by interdomain angles of ADK (*i.e.*, the NMP-CORE angle (θ_{NMP}) and the LID-CORE angle (θ_{LID})) and found that the deficient pathways (using less than 50 lowest normal modes) provided meaningful insights into the conformational transitions of ADK. When projecting the ensemble onto θ_{NMP} - θ_{LID} space, the deficient pathways showed the conformational transitions toward a metastable intermediate state (*i.e.*, the NMP-closing/LID-opening state) while the sufficient pathways (using more than 50 lowest normal modes) showed those directly to the final state (*i.e.*, the closed state) with unrealistic deformation (Fig 7B). Therefore, it is concluded that ICONGENI can explore meaningful transition pathways on complex energy landscapes.

The key role of computational approaches in investigating conformational transitions of proteins is to predict the trajectories that are beyond experimental capabilities. Our technique outlined here can shed light on the transition mechanisms in an efficient manner using only information on experimentally observed end-point structures. Furthermore, the simulation results strongly depend on a set of low-frequency normal modes as the system DOFs, enabling the method to generate a pathway ensemble based on dynamic characteristics and to provide low-energy paths. In this regard, our technique has the potential to find good candidates of unknown intermediate states on complex energy landscapes.

Supporting information

S1 Text. The ICONGENI simulations to explore ADK transition mechanisms via the NMP_O state.
(DOCX)

S1 Fig. The transition pathways for ADK via the NMP_O state generated by ICONGENI. Seven pathways in the vicinity of the NMP_O state are shown on θ_{NMP} - θ_{LID} space. The pathway named “diff. *a* degrees” means that it was explored by ICONGENI using the lowest normal modes that satisfy the condition: $\Delta\theta_{LID}-\Delta\theta_{NMP}>a$ (see details in S1 Text). The ADK crystal structures are taken as the references (indicated by green circles). 4AKE:A and 1AKE:A indicate the open and closed states of ADK, respectively. 1ZIN:A, 1ZIO:A, and 1ZIP:A (1DVR:A, and 1DVR:B) indicate experimental structures at the NMP_C state (the NMP_O state). The

pathway ensemble data (**Fig 7B**) is also included in this figure for comparison.
(TIF)

S1 Movie. The transition pathway for ADK simulated by ICONGENI based on the DOFs of the 50 lowest normal modes.

(WMV)

S2 Movie. The transition pathway for RBP simulated by ICONGENI based on the DOFs of the 50 lowest normal modes.

(WMV)

Author Contributions

Conceptualization: Byung Ho Lee, Soon Woo Park, Soojin Jo, Moon Ki Kim.

Data curation: Byung Ho Lee, Moon Ki Kim.

Formal analysis: Byung Ho Lee, Soon Woo Park.

Funding acquisition: Moon Ki Kim.

Investigation: Byung Ho Lee.

Methodology: Byung Ho Lee, Soon Woo Park, Moon Ki Kim.

Project administration: Moon Ki Kim.

Resources: Moon Ki Kim.

Software: Byung Ho Lee, Soon Woo Park.

Supervision: Moon Ki Kim.

Validation: Byung Ho Lee, Moon Ki Kim.

Visualization: Byung Ho Lee.

Writing – original draft: Byung Ho Lee.

Writing – review & editing: Byung Ho Lee, Soon Woo Park, Soojin Jo, Moon Ki Kim.

References

1. Kendrew JC, Dickerson RE, Strandberg BE, Hart RG, Davies DR, Phillips DC, et al. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. resolution. *Nature*. 1960; 185(4711):422–427. <https://doi.org/10.1038/185422a0> PMID: 18990802
2. Perutz MF, Rossmann MG, Cullis AF, Muirhead H, Will G, North, ACT. Structure of hæmoglobin: a three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature*. 1960; 185(4711):416–422. <https://doi.org/10.1038/185416a0> PMID: 18990801
3. Shi Y. A glimpse of structural biology through X-ray crystallography. *Cell*. 2014; 159(5):995–1014. <https://doi.org/10.1016/j.cell.2014.10.051> PMID: 25416941
4. Goodsell DS, Zardecki C, Di Costanzo L, Duarte JM, Hudson BP, Persikova I, et al. RCSB Protein Data Bank: Enabling biomedical research and drug discovery. *Protein Sci*. 2020; 29(1):52–65. <https://doi.org/10.1002/pro.3730> PMID: 31531901
5. Jensen MR, Zweckstetter M, Huang JR, Blackledge M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem Rev*. 2014; 114(13):6632–6660. <https://doi.org/10.1021/cr400688u> PMID: 24725176
6. Mertens HD, Svergun DI. Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch Biochem Biophys*. 2017; 628:33–41. <https://doi.org/10.1016/j.abb.2017.05.005> PMID: 28501583

7. Schuler B, Soranno A, Hofmann H, Nettels D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Ann Rev Biophys*. 2016; 45:207–231. <https://doi.org/10.1146/annurev-biophys-062215-010915> PMID: 27145874
8. Uversky VN, Dunker AK. Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. *Anal Chem*. 2012; 84(5):2096–2104. <https://doi.org/10.1021/ac203096k> PMID: 22242801
9. Grant BJ, Gorfe AA, McCammon JA. Large conformational changes in proteins: signaling and other functions. *Curr Opin Struct Biol*. 2010; 20(2):142–147. <https://doi.org/10.1016/j.sbi.2009.12.004> PMID: 20060708
10. Hollingsworth SA, Dror RO. Molecular dynamics simulation for all. *Neuron*. 2018; 99(6):1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011> PMID: 30236283
11. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. *Nat Struct Biol*. 2002; 9(9):646–652. <https://doi.org/10.1038/nsb0902-646> PMID: 12198485
12. Shaw DE, Grossman JP, Bank JA, Batson B, Butts JA, Chao JC, et al. Anton 2: raising the bar for performance and programmability in a special-purpose molecular dynamics supercomputer. In: SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis. IEEE; 2014. pp. 41–53. <https://doi.org/10.1109/SC.2014.9>
13. Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. Biomolecular simulation: a computational microscope for molecular biology. *Annu Rev Biophys*. 2012; 41:429–452. <https://doi.org/10.1146/annurev-biophys-042910-155245> PMID: 22577825
14. Orellana L. Large-scale conformational changes and protein function: breaking the in silico barrier. *Front Mol biosci*. 2019; 6:117. <https://doi.org/10.3389/fmolb.2019.00117> PMID: 31750315
15. Pietrucci F. Strategies for the exploration of free energy landscapes: unity in diversity and challenges ahead. *Rev Phys*. 2017; 2:32–45. <https://doi.org/10.1016/j.revip.2017.05.001>
16. Ikebe J, Umezawa K, Higo J. Enhanced sampling simulations to construct free-energy landscape of protein–partner substrate interaction. *Biophys Rev*. 2016; 8(1):45–62. <https://doi.org/10.1007/s12551-015-0189-z> PMID: 28510144
17. Gō N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci USA*. 1983; 80(12):3696–3700. <https://doi.org/10.1073/pnas.80.12.3696> PMID: 6574507
18. Case DA. Normal mode analysis of protein dynamics. *Curr Opin Struct Biol*. 1994; 4(2):285–290. [https://doi.org/10.1016/S0959-440X\(94\)90321-2](https://doi.org/10.1016/S0959-440X(94)90321-2)
19. Kim MK, Chirikjian GS, Jernigan RL. Elastic models of conformational transitions in macromolecules. *J Mol Graph Model*. 2002; 21(2):151–160. [https://doi.org/10.1016/s1093-3263\(02\)00143-2](https://doi.org/10.1016/s1093-3263(02)00143-2) PMID: 12398345
20. Cui Q, Bahar I. Normal mode analysis: theory and applications to biological and chemical systems. 1st ed. UK: Chapman and Hall/CRC; 2006.
21. Tekpinar M, Zheng W. Predicting order of conformational changes during protein conformational transitions using an interpolated elastic network model. *Proteins*. 2010; 78(11):2469–2481. <https://doi.org/10.1002/prot.22755> PMID: 20602461
22. Gur M, Madura JD, Bahar I. Global transitions of proteins explored by a multiscale hybrid methodology: application to adenylate kinase. *Biophys J*. 2013; 105(7):1643–1652. <https://doi.org/10.1016/j.bpj.2013.07.058> PMID: 24094405
23. Uyar A, Kantarci-Carsibasi N, Haliloglu T, Doruker P. Features of large hinge-bending conformational transitions. Prediction of closed structure from open state. *Biophys J*. 2014; 106(12):2656–2666. <https://doi.org/10.1016/j.bpj.2014.05.017> PMID: 24940783
24. Saldaño TE, Freixas VM, Tosatto SC, Parisi G, Fernandez-Alberti S. Exploring Conformational Space with Thermal Fluctuations Obtained by Normal-Mode Analysis. *J Chem Inf Model*. 2020; 60(6):3068–3080. <https://doi.org/10.1021/acs.jcim.9b01136> PMID: 32216314
25. Lee BH, Seo S, Kim MH, Kim Y, Jo S, Choi MK, et al. Normal mode-guided transition pathway generation in proteins. *PloS One*. 2017; 12(10):e0185658. <https://doi.org/10.1371/journal.pone.0185658> PMID: 29020017
26. Frezza E, Lavery R. Internal Coordinate Normal Mode Analysis: A Strategy to Predict Protein Conformational Transitions. *J Phys Chem B*. 2019; 123(6):1294–1301. <https://doi.org/10.1021/acs.jpcc.8b11913> PMID: 30665293
27. Lopéz-Blanco JR, Garzón JI, Chacón P. iMod: multipurpose normal mode analysis in internal coordinates. *Bioinformatics*. 2011; 27(20):2843–2850. <https://doi.org/10.1093/bioinformatics/btr497> PMID: 21873636

28. Kitao A, Hayward S, Gō N. Comparison of normal mode analyses on a small globular protein in dihedral angle space and Cartesian coordinate space. *Biophys Chem.* 1994; 52(2):107–114. [https://doi.org/10.1016/0301-4622\(94\)00070-0](https://doi.org/10.1016/0301-4622(94)00070-0) PMID: 17020826
29. Müller CW, Schlauderer GJ, Reinstein J, Schulz GE. Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding. *Structure.* 1996; 4(2):147–156. [https://doi.org/10.1016/s0969-2126\(96\)00018-4](https://doi.org/10.1016/s0969-2126(96)00018-4) PMID: 8805521
30. Müller CW, Schulz GE. Structure of the complex between adenylate kinase from *Escherichia coli* and the inhibitor Ap5A refined at 1.9 Å resolution: A model for a catalytic transition state. *J Mol Biol.* 1992; 224(1):159–177. [https://doi.org/10.1016/0022-2836\(92\)90582-5](https://doi.org/10.1016/0022-2836(92)90582-5) PMID: 1548697
31. Björkman AJ, Mowbray SL. Multiple open forms of ribose-binding protein trace the path of its conformational change. *J Mol Biol.* 1998; 279(3):651–664. <https://doi.org/10.1006/jmbi.1998.1785> PMID: 9641984
32. Björkman AJ, Binnie RA, Zhang H, Cole LB, Hermodson MA, Mowbray SL. Probing protein-protein interactions. The ribose-binding protein in bacterial transport and chemotaxis. *J Biol Chem.* 1994; 269(48):30206–30211. [https://doi.org/10.1016/S0021-9258\(18\)43798-2](https://doi.org/10.1016/S0021-9258(18)43798-2) PMID: 7982928
33. Berry MB, Phillips GN Jr. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap5A, Mg²⁺ Ap5A, and Mn²⁺ Ap5A reveal an intermediate lid position and six coordinate octahedral geometry for bound Mg²⁺ and Mn²⁺. *Proteins.* 1998; 32(3):276–288. [https://doi.org/10.1002/\(SICI\)1097-0134\(19980815\)32:3<276::AID-PROT3>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0134(19980815)32:3<276::AID-PROT3>3.0.CO;2-G) PMID: 9715904
34. Schlauderer GJ, Proba K, Schulz GE. Structure of a mutant adenylate kinase ligated with an ATP-analogue showing domain closure over ATP. *J Mol Biol.* 1996; 256(2):223–227. <https://doi.org/10.1006/jmbi.1996.0080> PMID: 8594191
35. Webb B, Sali A. Comparative Protein Structure Modeling Using Modeller. *Curr Protoc Bioinformatics.* 2016; 54:5.6.1–5.6.37. <https://doi.org/10.1002/cpbi.3> PMID: 27322406
36. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* 2006; 15(11): 2507–2524. <https://doi.org/10.1110/ps.062416606> PMID: 17075131
37. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods.* 2017; 14(1):71–73. <https://doi.org/10.1038/nmeth.4067> PMID: 27819658
38. Noguti T, Gō N. Dynamics of Native Globular Proteins in Terms of Dihedral Angles. *J Phys Soc Jpn.* 1983; 52(9):3283–3288. <https://doi.org/10.1143/JPSJ.52.3283>
39. Braun W, Yoshioki S, Gō N. Formulation of static and dynamic conformational energy analysis of biopolymer systems consisting of two or more molecules. *J Phys Soc Jpn.* 1984; 53(9):3269–3275. <https://doi.org/10.1143/JPSJ.53.3269>
40. Kamiya K, Sugawara Y, Umeyama H. Algorithm for normal mode analysis with general internal coordinates. *J Comput Chem.* 2003; 24(7):826–841. <https://doi.org/10.1002/jcc.10247> PMID: 12692792
41. Lu M, Poon B, Ma J. A new method for coarse-grained elastic normal-mode analysis. *J Chem Theory Comput.* 2006; 2(3):464–471. <https://doi.org/10.1021/ct050307u> PMID: 21760758
42. Eckart C. Some studies concerning rotating axes and polyatomic molecules. *Phys Rev.* 1935; 47(7):552–558. <https://doi.org/10.1103/PhysRev.47.552>
43. Kim MK, Jernigan RL, Chirikjian GS. Efficient generation of feasible pathways for protein conformational transitions. *Biophys J.* 2002; 83(3):1620–1630. [https://doi.org/10.1016/S0006-3495\(02\)73931-3](https://doi.org/10.1016/S0006-3495(02)73931-3) PMID: 12202386
44. Schulz GE, Müller CW, Diederichs K. Induced-fit movements in adenylate kinases. *J Mol Biol.* 1990; 213(4):627–630. [https://doi.org/10.1016/S0022-2836\(05\)80250-5](https://doi.org/10.1016/S0022-2836(05)80250-5) PMID: 2162964
45. Engh RA, Huber R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst.* 1991; A47(4):392–400. <https://doi.org/10.1107/S0108767391001071>
46. Maragakis P, Karplus M. Large amplitude conformational change in proteins explored with a plastic network model: adenylate kinase. *J Mol Biol.* 2005; 352(4):807–822. <https://doi.org/10.1016/j.jmb.2005.07.031> PMID: 16139299
47. Jana B, Adkar BV, Biswas R, Bagchi B. Dynamic coupling between the LID and NMP domain motions in the catalytic conversion of ATP and AMP to ADP by adenylate kinase. *J Chem Phys.* 2011; 134(3):035101. <https://doi.org/10.1063/1.3516588> PMID: 21261390
48. Wang Y, Gan L, Wang E, Wang J. Exploring the dynamic functional landscape of adenylate kinase modulated by substrates. *J Chem Theory Comput.* 2013; 9(1):84–95. <https://doi.org/10.1021/ct300720s> PMID: 26589012
49. Lin CY, Huang JY, Lo LW. Deciphering the catalysis-associated conformational changes of human adenylate kinase 1 with single-molecule spectroscopy. *J Phys Chem B.* 2013; 117(45):13947–13955. <https://doi.org/10.1021/jp4019537> PMID: 24134437

50. Kong J, Li J, Lu J, Li W, Wang W. Role of substrate-product frustration on enzyme functional dynamics. *Phys Rev E*. 2019; 100(5):052409. <https://doi.org/10.1103/PhysRevE.100.052409> PMID: 31869999
51. Beckstein O, Denning EJ, Perilla JR, Woolf TB. Zipping and unzipping of adenylate kinase: atomistic insights into the ensemble of open \leftrightarrow closed transitions. *J Mol Biol*. 2009; 394(1):160–176. <https://doi.org/10.1016/j.jmb.2009.09.009> PMID: 19751742
52. Oshima H, Re S, Sugita Y. Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. *J Chem Theory Comput*. 2019; 15(10):5199–5208. <https://doi.org/10.1021/acs.jctc.9b00761> PMID: 31539245