PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Predicting protein conformational motions using energetic frustration analysis and AlphaFold2

Xingyue Guan[a,b], Qian-Yuan Tang[c], Weitong Ren[b], Mingchen Chen[d], Wei Wang[a,1], Peter G. Wolynes[e,1], and Wenfei Li[a,b,1]

Affiliations are included on p. 10.

Contributed by Peter G. Wolynes; received May 30, 2024; accepted July 16, 2024; reviewed by Pratyush Tiwary and Bin Zhang

Proteins perform their biological functions through motion. Although high throughput prediction of the three-dimensional static structures of proteins has proved feasible using deep-learning-based methods, predicting the conformational motions remains a challenge. Purely data-driven machine learning methods encounter difficulty for addressing such motions because available laboratory data on conformational motions are still limited. In this work, we develop a method for generating protein allosteric motions by integrating physical energy landscape information into deep-learning-based methods. We show that local energetic frustration, which represents a quantification of the local features of the energy landscape governing protein allosteric dynamics, can be utilized to empower AlphaFold2 (AF2) to predict protein conformational motions. Starting from ground state static structures, this integrative method generates alternative structures as well as pathways of protein conformational motions, using a progressive enhancement of the energetic frustration features in the input multiple sequence alignment sequences. For a model protein adenylate kinase, we show that the generated conformational motions are consistent with available experimental and molecular dynamics simulation data. Applying the method to another two proteins KaiB and ribose-binding protein, which involve large-amplitude conformational changes, can also successfully generate the alternative conformations. We also show how to extract overall features of the AF2 energy landscape topography, which has been considered by many to be black box. Incorporating physical knowledge into deep-learning-based structure prediction algorithms provides a useful strategy to address the challenges of dynamic structure prediction of allosteric proteins.

protein folding | energy landscapes | deep-learning | multiple sequence alignment | structure prediction

Significance

How proteins make conformational motions is not only a fundamental biophysical question, but also critically important for drug design and other practical applications. Using deep machine learning, AlphaFold2 (AF2) and RoseTTAFold can predict protein static structure quite routinely. Here, we provide a strategy to resolve conformational motions by integrating AF2's prediction capability with energy landscape analysis of protein allosteric motions. We show that localized energetic frustration, which quantifies the energy landscape features of allosteric proteins, can be used to empower AF2 to successfully predict alternative structures and pathways of allosteric transitions.

Large-scale conformational motions of proteins are essential for many biological processes, such as catalysis, signal transduction, and transport. Natural proteins have evolved their sequences to shape unique energy landscapes, which not only determine the corresponding three-dimensional native structures, but also dictate their functionally relevant motions (1). Finding the key local features of the protein energy landscape controlling the folding and functional motions gives insights into intricate relationships between sequence, structure, and functional dynamics (2). Many of these key features can also be inferred from evolutionary patterns embedded in the sequences of a protein family (3–8). The successes of AlphaFold2 (AF2) (9) and RoseTTAFold (10) in directly generating structure from sequence were made possible by harnessing the evolutionary data.

AF2 leverages state-of-the-art transformer architectures to predict protein structures from sequences with near experimental accuracy (9, 11). AF2 was specialized to provide a single predicted structure, but it provides a basis for much wider applications. Recent research has started to delve into the capabilities of AF2 in predicting alternative conformations, dynamics, functions, and mutational effects of proteins, fully utilizing the wealth of evolutionary information derived from multiple sequence alignment (MSA) (12–21). Jussupow and Kaila have shown that there exists a strong correlation between the statistical prediction scores given by AF2 and the intrinsic conformational fluctuations of proteins sampled by molecular dynamics (MD) simulations (13). Roney and Ovchinnikov also showed that AF2 has effectively learned an information-based energy function that can be used to rank the quality of output protein predictions (14). These studies highlight the untapped potential of AF2 in harnessing evolutionary data from MSAs to provide insights into the energy landscapes of proteins. There have also been successes at steering AF2 to predict alternative structures by using shallow MSAs and other pipelines. These studies show that reducing the MSA information, either by restricting the MSA depth or by masking the amino acid identity at certain positions

Author contributions: X.G., Q.-Y.T., P.G.W., and W.L. designed research; X.G. and W.R. performed research; X.G. and M.C. contributed new reagents/analytic tools; X.G., Q.-Y.T., W.R., M.C., W.W., P.G.W., and W.L. analyzed data; and X.G., W.W., P.G.W., and W.L. wrote the paper.

Reviewers: P.T., University of Maryland at College Park; and B.Z., Massachusetts Institute of Technology.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: wangwei@nju.edu.cn, pwolynes@rice.edu, or wfli@nju.edu.cn.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2410662121/-/DCSupplemental.

Published August 20, 2024.

of the sequences, acts as a perturbation to the predicted results of AF2 allowing one to obtain alternative structures (16, 18, 19, 22, 23). More recently, Wayment-Steele and coworkers showed that AF2 can be used to predict multiple conformational states with high confidence by clustering the MSAs with high sequence similarity(17). By combining AF2 structure prediction, all-atom enhanced sampling MD, and induced-fit docking, Gu et al. developed a workflow named AF2RAVE-Glide, which can successfully generate the metastable conformations for three different kinases (24). These works pave the way for using AF2 to predict conformational dynamics. In addition to these AF2-based pipelines, other strategies that utilize improved predictions of residue–residue distance distribution and contact maps to predict protein multiple conformations also show successes (25–27). Applying energy filtering to the generated conformational ensemble can further improve the predictions as shown by Audagnotto et al. using the trRosetta pipeline (25, 28). A diffusion-based generative model trained with MD simulation trajectories and force-field information developed by Zheng and coworkers can successfully generate diverse and functionally relevant structures for several proteins (29).

AF2 has often been perceived as a complex "black box" whose underlying processes are largely unexplored. Our understanding of the physical principles governing protein folding and allosteric motions however has already reached a level of maturity during the past decades. The principle of minimal frustration reveals how kinetic and thermodynamic constraints are encoded in the sequence of foldable proteins (30, 31). Evolution has molded a globally funnel-shaped energy landscape, but the exceptions to the perfect funnel usually encode function. Local frustration at the residue level pinpoints the linchpins of protein allosteric motions (2, 32–34). Enzyme catalytic pathways also rely on the specifically sculptured spatial distribution pattern of local frustration (2, 35, 36). Ideally, one would like to use these physical ideas to predict and design protein functional dynamics. The great success of AF2 and RoseTTAFold in directly generating protein three-dimensional structures inspired us to ask: Is it possible to combine such data-driven methods with available physical ideas about the energy landscape to directly predict large-scale protein conformational motions without carrying out molecular dynamics simulations?

By exploiting AF2's ability to use evolutionary data from MSAs in this paper, we introduce a method that synergizes the structure prediction prowess of AF2 with the analysis of protein energetic frustration to predict protein conformational motions (Fig. 1). The concept of frustration, which highlights the role of conflicting interactions in protein folding, dynamics, and function, provides the key to unlocking this deeper capability of AF2. By harnessing available energetic frustration patterns of allosteric proteins extracted based on established biophysical tools (37), we aim to extend the capabilities of AF2 beyond static structure prediction to include the dynamic aspects of allosteric proteins, thereby improving our understanding of the genotype–phenotype mapping. The present method not only successfully predicts alternative conformations of some typical allosteric proteins, but also provides information on the dynamic pathway of the predicted conformational motions. This computational approach combining AF2 and biophysical rules will allow efficient in-depth investigation of protein dynamics, potentially relevant to therapeutic strategies.

Results

Allosteric Feature of Protein Energy Landscape. Although to achieve robust protein folding globally requires a funnel-shaped energy landscape with relatively few energetic conflicts in the native structures (30, 31), allosteric proteins often show significant regions of high local frustration. These frustrated regions are essential for functionally relevant conformational motions (2, 32). The local frustration index characterizes how energetically favorable a given contact between two residues is in a native structure compared to all other possible random contacts made in alternative configurations. The protein regions with minimally frustrated interactions are often relatively rigid. In comparison, the highly frustrated regions can adopt different configurations and therefore are potentially more flexible. The local frustration index can be readily calculated using the Frustratometer developed by Ferreiro et al. (37). Adenylate kinase (AdK), a multidomain protein composed of the LID domain, the NMP domain, and the core domain (Fig. 2*A* and *SI Appendix*, Fig. S1) (38), has become a paradigmatic system for the study of allostery. It has already been used to illustrate the interplay between localized energetic frustration, conformational motions, and catalytic functions (33, 39–43). AdK plays a vital functional role in all kingdoms of life by reversibly catalyzing the phosphate transfer between adenosine triphosphate (ATP) and adenosine monophosphate to produce two adenosine diphosphate molecules (44), thereby maintaining the cellular ATP level. The kinase exhibits large-amplitude conformational transitions between open and closed conformations during the enzymatic cycle (*SI Appendix*, Fig. S1) involving several intermediate structures (45). It has been shown that local energetic frustration plays a crucial role in sculpting the catalytic pathways by promoting specific conformational motions (33–36, 39, 42). In this paper, we will also employ this enzyme as a model to show how one can predict protein conformational motions with AF2 by making use of the local features of energy landscape that are quantified by the local energetic frustration scores.

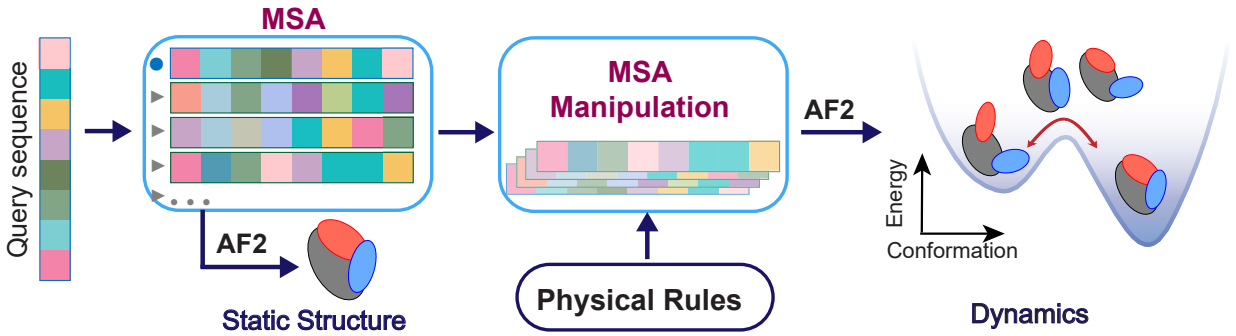

Fig. 1. Schematic illustration of physical rule dictated protein dynamics prediction with AF2.

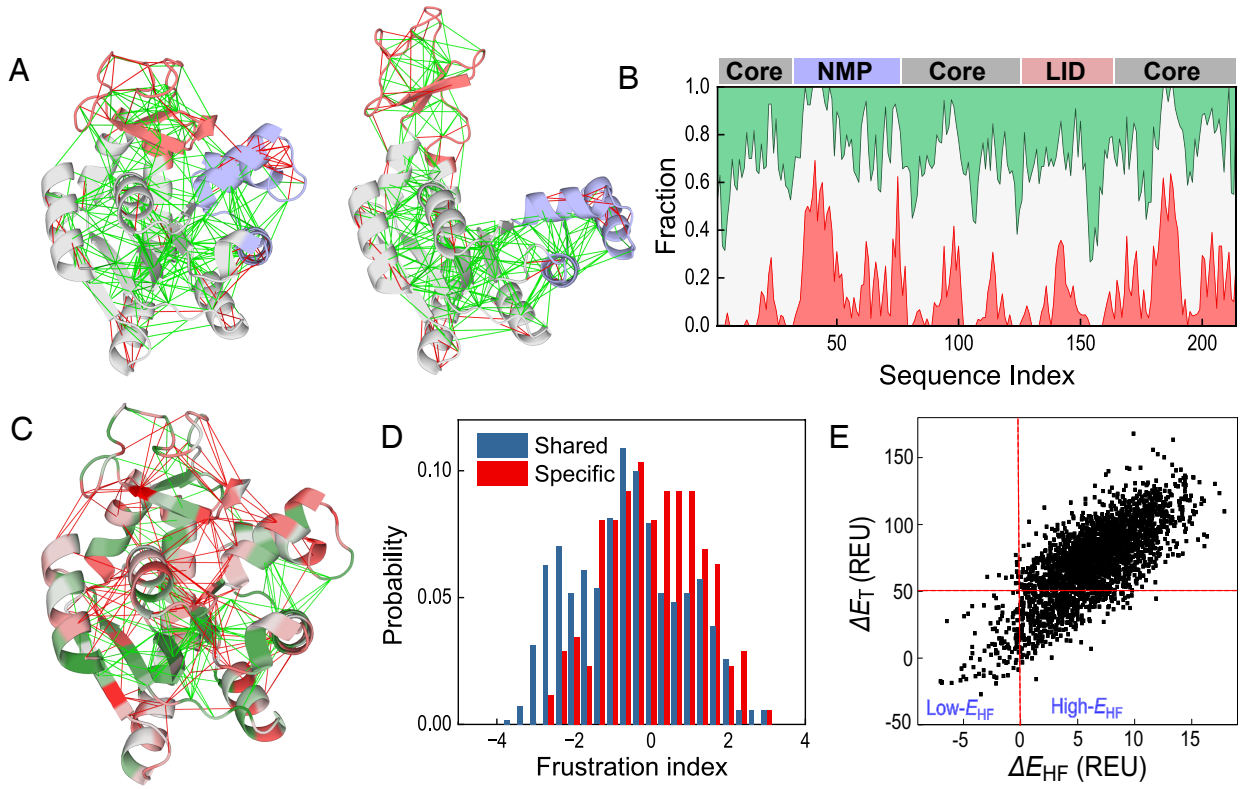

Fig. 2. Localized energetic frustration of AdK. (*A*) Cartoon representation of the three-dimensional structure of AdK at closed (Protein Data Bank (PDB) code: 1AKE) and open (PDB code: 4AKE) states. The LID domain, NMP domain, and core domain are colored in red, blue, and gray, respectively. The contacts with minimal frustration (green lines) and high frustration (red lines) are indicated. (*B*) The density fraction of highly frustrated contacts (red) and minimally frustrated contacts (green) around a 5 Å sphere of the Cα of each residue. (*C*) Cartoon representation of the three-dimensional structure of AdK at closed state showing the contacts with high F_{ij}^s (red) and low F_{ij}^s (green). The residues are colored according to the residue-wise frustration index F_i. The residue-wise frustration becomes higher with the color changing from green to red. (*D*) Distribution of F_{ij}^s for the shared contacts and the specific contacts. Larger F_{ij}^s values indicate higher frustrations. (*E*) MSA sequences of AdK are projected onto the energy space formed by total energy ΔE_T and the energy of highly frustrated contacts ΔE_{HF}. Here, ΔE_T and ΔE_{HF} represent the energy differences relative to the reference values.

Fig. 2*A* illustrates the local frustration patterns in the closed and open structural forms of AdK. The pairwise frustration index F_{ij} calculated by means of the Frustratometer (37) were used. As also observed in a previous paper by Ferreiro et al. (34), the interiors of the AdK domains are enriched in minimally frustrated interactions (green lines), forming densely connected networks. In contrast, the interactions at the domain interfaces and the hinge regions are often highly frustrated (red lines). The density fraction of the highly frustrated contacts around the residues of AdK shows discontinuous patches along the sequence (Fig. 2*B*). To more clearly demonstrate the key features of the frustration patterns, we also calculated smoothed frustration scores $F_{ij}^s = -(F_i + F_j)$, with F_i and F_j being the residue-wise frustration index given by the Frustratometer. Therefore, high F_{ij}^s contacts correspond to the interface of the highly frustrated sites. We see that the domain interfaces of the closed structure of AdK are enriched in contacts with high F_{ij}^s values (Fig. 2*C*, red lines), which seems to be an important energetic feature facilitating the domain motions. We further calculated the distributions of the smoothed frustration scores for two classes of contacts, i.e., the shared contacts and specific contacts. Here, we classify the native contacts as "shared contacts" if they were fully formed in both structural forms and as "specific contacts" if the contact only exists in one of the structures. These specific contacts need to break during the conformational change with domain motions and often are located at the domain interfaces of the closed structure only. Residue pairs (i, j) with C$_\alpha$ distances less than 10 Å and sequence separation larger than 8 were considered in classifying the shared and specific contacts and in calculating F_{ij}^s. As shown in Fig. 2*D*, the specific contacts tend to be more frustrated than the shared contacts, consistent with previous observations on the connection between frustration and allostery (33, 34). Surveys of allosteric proteins show that this feature of the protein energy landscape observed for AdK is general for other allosteric proteins (33, 34). Interestingly, a recent work by Raisinghani et al. showed that the low-populated inactive conformation of ABL kinase is featured by presence of large high-frustration residue clusters, which tends to introduce difficulty for structure prediction with AF2 (46).

Predicting Alternative Structures of Allosteric Proteins with AF2 by Frustration Filtering. For *Escherichia coli* AdK, with the input of the full set of MSA sequences, AF2 returns a predicted structure having a closed conformation, highly similar to the corresponding crystal structure (*SI Appendix*, Fig. S1). The energetic frustrations calculated based on this predicted structure also exhibit the nearly same patterns as those of the crystal structure. We will next demonstrate that such local frustration information can be used to force AF2 to predict protein conformational motions. Considering that the MSA sequences containing coevolutionary information provide hints for structure prediction, we therefore use the energetic frustration

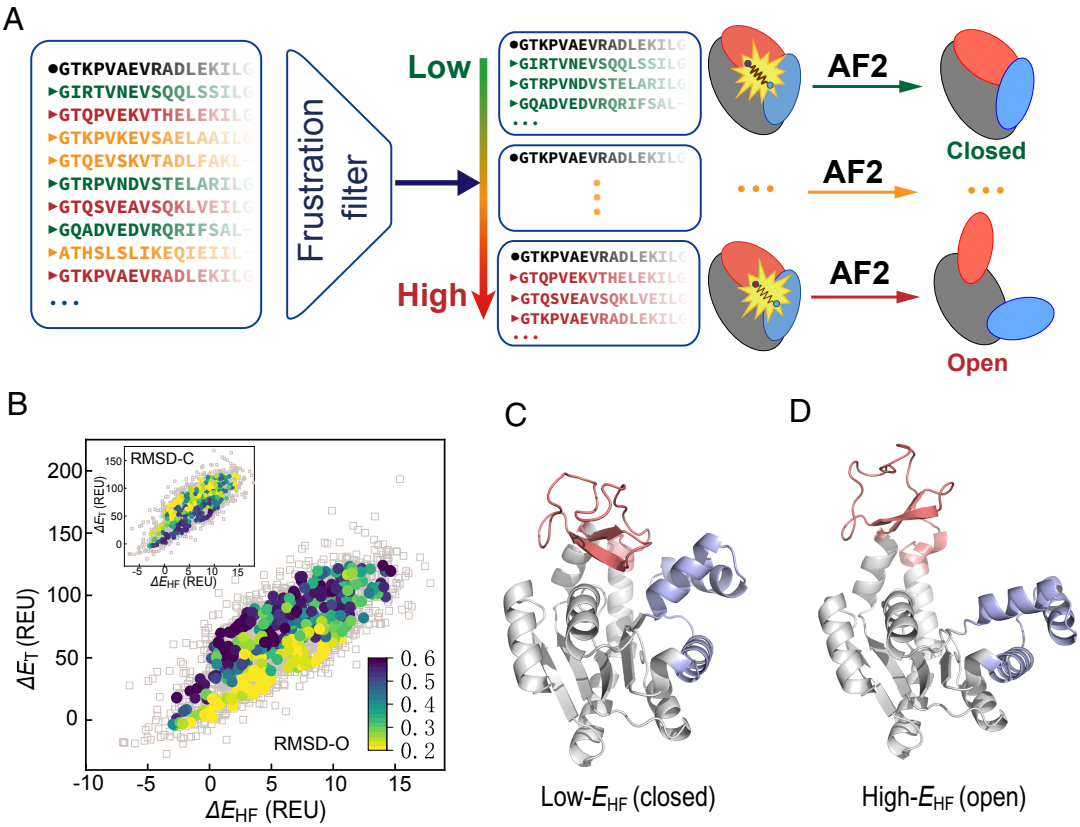

Fig. 3. Predicting alternative structures of AdK by frustration-filtering strategy. (*A*) The workflow of the frustration-filtering strategy in predicting protein conformational motions. (*B*) RMSD with respect to the reference open structure (RMSD-O, PDB code: 4AKE) for the predicted structures with the input MSA sequences taken from different locations of the two-dimensional energy space. *Inset*: the RMSD with respect to the reference closed structure (RMSD-C, PDB code: 1AKE). (*C* and *D*) Cartoon representations of the predicted three-dimensional structures of AdK with the low E_{HF} (*C*) and high E_{HF} (*D*) MSA sequences.

profiles of the MSA sequences on a given reference structure to encourage AF2 to generate alternative structures.

As previously mentioned, the conformational transition from a closed to an open state of AdK involves the disruption of specific contacts, which are often characterized by high frustration. We can direct AF2 to model the structural dynamics associated with the reconfiguration of these high-frustration sites. We first thread the MSA sequences onto the closed structure of AdK and calculate the total energies for all the native contacts (E_T) as well as for the highly frustrated contacts (E_{HF}) using the Rosetta energy function (*Materials and Methods*) (47). The MSA sequences can then be classified based on the two-dimensional "evolutionary" reaction coordinates, ΔE_T and ΔE_{HF} (Fig. 2*E*). These describe the overall stability and extent of local frustration of the closed structure for a given sequence relative to those of the reference *E. coli* AdK sequence. The sequences with higher ΔE_{HF} tend to have more pronounced local frustration for these highly frustrated sites, which are often involved in the specific contacts (Fig. 2*D*). It is reasonable to assume that the sequences demonstrating lower energies for these specific contacts will take on a closed conformation in AF2 structure prediction (Fig. 3*A*). On the contrary, sequences that exhibit higher contact energies for the specific contacts are likely to favor an open structure, as these contacts are destabilized. Consequently, we see that we can use AF2 to generate closed and open structures by manipulating the MSA as informed by characterizing the local energetic frustration.

Based on this physical picture of frustration filtering, we introduced an energy threshold approach for MSA sequence subsampling and prediction of alternative conformations of allosteric proteins. We separately sampled sequences from specific different regions of the two-dimensional energy space and provided only these sequences, along with the input query sequence, to AF2 for structural predictions. From each region, 50 MSA sequences are used for structure prediction. Remarkably, AF2 predominantly returns the closed structures for the MSA sequences with low ΔE_{HF} and returns open structures for those with high ΔE_{HF}, in each case with high confidence (Fig. 3*B* and *SI Appendix*, Figs. S2 and S3). Particularly, when the subsampled MSA sequences with the lowest ΔE_{HF} (500 sequences) are used, AF2 returns a nearly fully closed structure having RMSD = 0.80 Å with respect to the crystal structure at closed state (PDB code: 4AKE). Likewise, when only the MSA sequences with high ΔE_{HF} and low ΔE_T (50 sequences) are used, AF2 yields an open structure having RMSD = 1.63 Å from the crystal structure of the open state (PDB code: 4AKE) (Fig. 3 *C* and *D*). Correspondingly, the MSA sequences generating the open structures tend to have higher ΔE_{HF} values and decreased contact energies of the specific contacts (*SI Appendix*, Figs. S4 and S5). We see that the information of local energetic frustration can be utilized to force AF2 to predict meaningful alternative structures for allosteric proteins. As a control, we also conducted additional test calculations following a similar pipeline but with the frustration scores shuffled. When the frustration scores are shuffled, no correspondence between the MSA energetics and the conformations of the predicted structures is observed (*SI Appendix*, Fig. S5), demonstrating again the crucial role of the local energetic frustration in determining meaningful alternative states.

Characterizing the Pathways of Protein Conformational Motions. In this subsection, we introduce an approach for visualizing the conformational motions with AF2 by examining how the conformations of predicted structures change when we sample different organized sets of sequences from the MSA pools. Starting from the MSA pools that produce the above fully closed conformation and the nearly open conformation of AdK, we then randomly resample the sequences using new probability weights w for the closed pool and $(1 - w)$ for the open pool, respectively (Fig. 4*A*). These linearly combined MSA sequences were then used to generate predicted structures with AF2. By changing the weight parameter w, which is independent of ΔE_{HF}, one generates different MSA sequence sets with varying relative proportions of the high-ΔE_{HF} and low-ΔE_{HF} sequences. Interestingly, we find AF2 gives rise to a continuous structural transition between the closed and the open conformation as one gradually changes the relative weight w (Fig. 4 *A* and *B*). Not only are the fully closed and open structures produced, but other conformations having different degrees of opening of the LID and NMP domains are generated. As the relative weight $(1 - w)$ of the high-frustration sequences increases, the predicted structure tends to have a more open conformation. These results demonstrate the high sensitivity of AF2's prediction to local energetic frustration features, again suggesting the crucial role of the physical energy landscape in dictating the AF2 structure prediction.

In addition to using the sequence remixing approach, we investigated how the predicted structures change with continuous increasing of the ΔE_{HF} values for the input MSA sequences. For this purpose, we designed a sequence sliding approach for the prediction of the protein motions, in which the MSA sequences in a sliding window (containing 50 sequences) were fed to the AF2 to generate the prediction structures. With the continuous sliding of the sequence window from the low-ΔE_{HF} end to the high-ΔE_{HF} end with the restraint of $\Delta E_T < 50$ Rosetta Energy Unit (REU) (arrowhead line in Fig. 4 *C*, *Inset*), AF2 also generates a continuous trajectory with the predicted structures gradually changing from the nearly closed conformation to the fully open conformation (Fig. 4*C*).

While the integration of AF2 with physical energy data can yield continuous transition pathways between two functional states for AdK, it is essential to determine whether the resulting pathways are physically meaningful. Fortunately, the pathways of conformational transitions for AdK have already been extensively characterized by experiments and molecular simulations, which can be used as a reference to assess the predicted pathways. We plotted the predicted structures along two reaction coordinates, i.e., "Angle NMP–Core" and "Angle LID–Core," which describe the opening angles of the LID and NMP domains, respectively, with respect to the core domain (Fig. 4*D*). As a comparison, we also plotted the experimentally resolved structures along the conformational transitions (*SI Appendix*, Text) (48). Interestingly, we can observe two major pathways. In one pathway, the LID domain opens partially first, followed by the full opening of the NMP domain and the further opening of the LID domain. In another pathway, the NMP domain opens largely first, followed by the full opening of the LID domain. One can see that the predicted pathways in this work align very well with the experimental structures captured along nearly the same pathways of conformational motions (Fig. 4*D*).

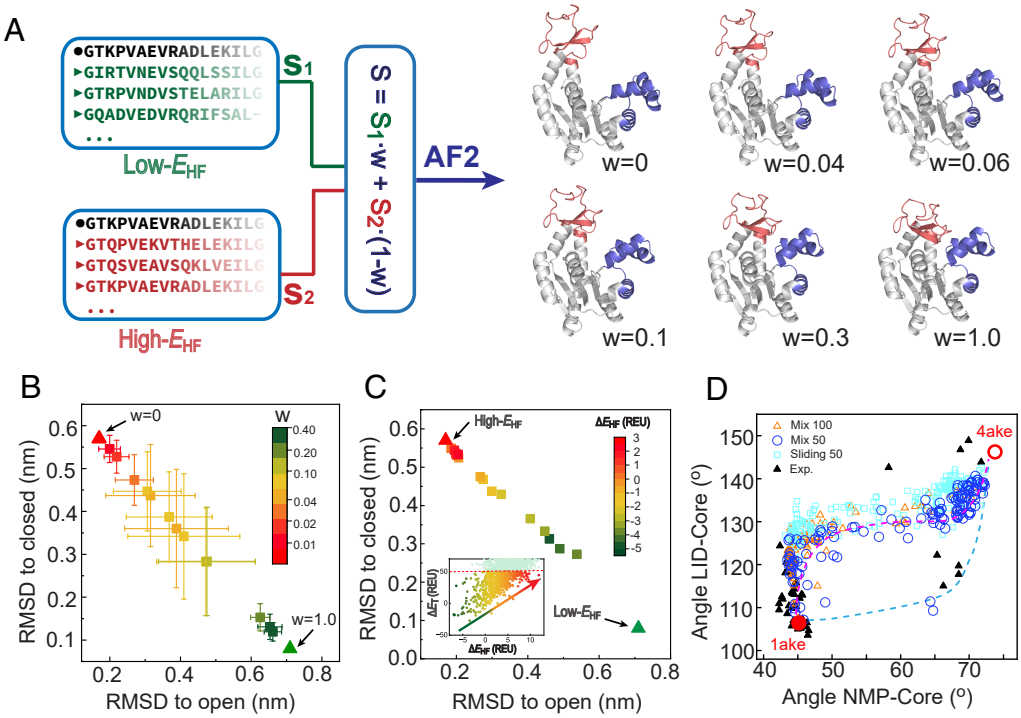

Fig. 4. Pathways of conformational change of AdK predicted with the frustration filtering strategy. (*A*) Working flow for generating pathways of conformational motions between closed and open states by mixing the MSA sequences of low-frustration region and high-frustration region with different weight w. The generated structures with different mixing weights were also shown. (*B*) Projection of the generated structures by the sequence mixing protocol along the RMSD with respect to the reference closed and open structures. (*C*) Projection of the generated structures with the sequence-sliding approach along the RMSD with respect to the reference closed and open structures. *Inset*: sequence-sliding scheme. (*D*) Projection of the generated structures along the reaction coordinates "Angle NMP–Core" and Angle LID–Core." Different colors represent the results with different structure generation protocols. For comparison, the closed (1ake) and open (4ake) conformations were highlighted for reference. Other experimental structures along the pathways of the conformational motions are shown by black triangles (data taken from ref. 48). Two major pathways were schematically indicated by dash lines.

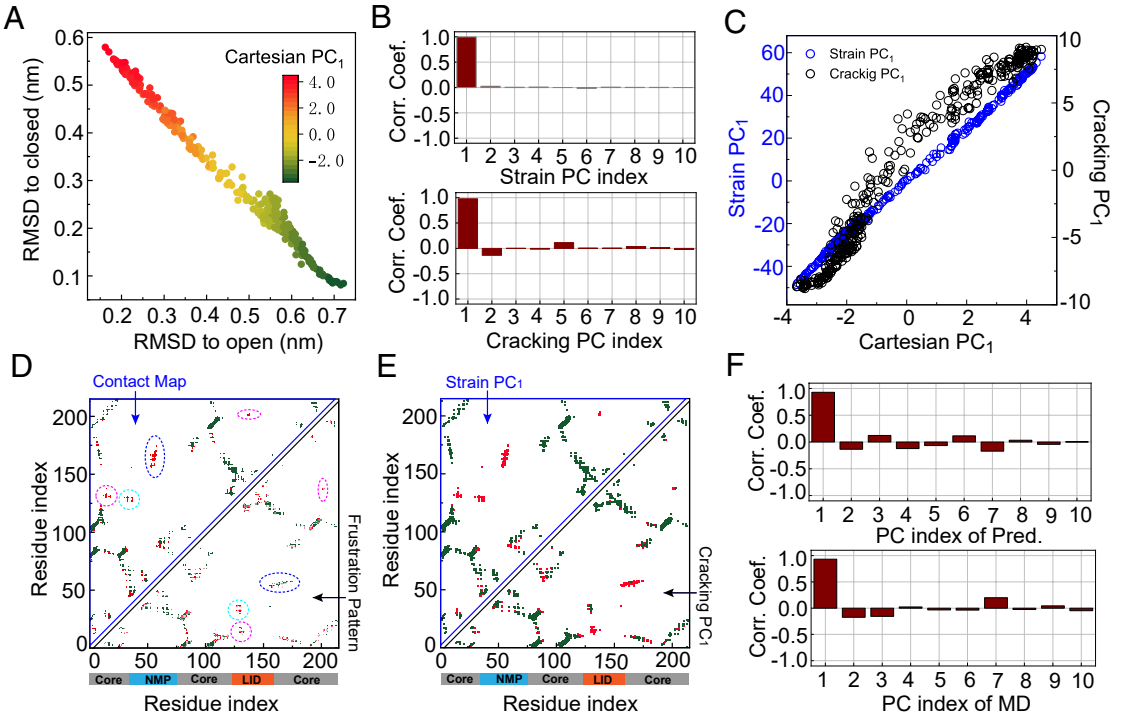

Fig. 5. Correlation between strain/cracking motions and the frustration pattern. (*A*) Projection of the predicted structures along the RMSDs with respect to the reference closed and open structures. The projections of the predicted structures along the Cartesian PC_1 were shown by different colors. (*B*) Correlation coefficients between the Cartesian PC_1 and different strain (cracking) PCs for the predicted structures. (*C*) Correlation plots of the predicted structures along the strain PC1 (cracking PC1) and Cartesian PC1. (*D*) Contact map (*Upper* triangle) and frustration pattern (*Lower* triangle) of AdK. The shared contacts (green) and state-specific contacts (red) were assigned by comparing the crystal structures at the closed and open states. The highly frustrated contacts and minimally frustrated contacts are color-coded with red and green, respectively. The specific contacts along the LID–NMP interface, NMP–Core interface, and LID–Core interface are circled by cyan, blue, and purple lines, respectively. (*E*) The strain PC_1 (*Upper* triangle) and the cracking PC_1 (*Lower* triangle) along the residue pairs. (*F*) *Upper*, correlation coefficients between the Cartesian PC1 of the predicted structures and different Cartesian PCs of the structures sampled by MD simulations; *Lower*, correlation coefficients between the Cartesian PC1 of the structures sampled by MD simulations and the different Cartesian PCs of the predicted structures.

In addition, the predicted pathways are highly consistent with those generated by molecular simulations in this work and other previous works (*SI Appendix*, Fig. S6 and Text) (48, 49). It is worth mentioning that other alternative pathways were also generated in the previous molecular simulations, depending on the simulation methods and force fields. However, most of these previous studies reproduce these main representative pathways. The close resemblance between the predicted pathways and the reference pathways suggests that the combination of AF2 and frustration filter is not only able to precisely predict the alternative conformations, but also can provide physically reasonable pathways of the protein conformational transitions.

To better characterize the conformational dynamics predicted by the aforementioned method, we conducted a series of principal component analyses (PCA) based on the predicted structural ensemble. First, we conducted conventional Cartesian PCA, in which the Cartesian coordinates of residues in three-dimensional space were used in constructing the covariance matrix. Comparison between the Cartesian PC_1 (the first-principle component) and the RMSDs with respect to the open and closed structures suggest that the Cartesian PC_1 captures the majority of the conformational fluctuations (Fig. 5*A*). PCA's based on coordinates show the elastic response of the protein (50). More important are stages of "cracking" where regions become somewhat disordered (39). To detect these cracking motions we use the collective coordinates introduced by Potoyan et al. (51). We carried out therefore two kinds of PCA's, i.e., strain PCA and cracking PCA, for the predicted structural ensemble. In the strain PCA, the covariance matrix is constructed using the interresidue distances of contacts, allowing the derived principal components (PCs) to capture the collective stretching and contraction of the residue contacts. In the cracking PCA, a binary variable representing the rupture or formation of contacts was used to construct the covariance matrix, thus the derived PCs describe the cracking of structures. It is observed that both the strain PC_1 (Fig. 5 *B*, *Upper* and *C*) and the cracking PC_1 (Fig. 5 *B*, *Lower* and *C*) are well correlated with the Cartesian PC_1 and therefore the dominant mode of conformational fluctuations. Remarkably, both the strain PC_1 and the cracking PC_1 constructed based on the predicted structural ensemble captured well the specific contacts along the domain interfaces which need to be broken during the closed to open conformational changes (Fig. 5 *D* and *E*). Consistent with the above discussion, these specific contacts tend to be highly frustrated and localized as shown in the frustration pattern (Fig. 5 *D*, *Lower*). Compared to the LID–Core interface, there are also some specific contacts with low frustration along the NMP–Core interface. Such a result is in line with the observation that the predicted structures with open NMP and closed LID are less probable compared to the structures along the alternative pathway (Fig. 4*D*). In addition, the Cartesian PC_1s derived from the predicted structures and from the structural ensemble obtained from MD simulations are highly correlated (Fig. 5*F* and *SI Appendix*, Text). Overall, these PCA results suggest that the predicted structures reasonably describe the local cracking and deformation involved in the conformational motions of AdK.

Inferring the Overall Topography of AF2 Energy Landscape. A recent work showed that AF2 has learned an energy function with its optimum corresponding to the native structure (14). The coevolutionary information from MSA plays a strong role in locating the global minimum, so that the searching of the native structure in the subsequent optimization step becomes possible. However, the overall topography of the AF2 learned energy landscape is not so easy to extract and remains a black box. With the MSA manipulation guided by the information of the physical energy landscape, one can effectively seed the initial structures of the prediction along different locations of the AF2 energy landscape. By mapping the MSA inputs to the final AF2 predicted structures, it may be possible to infer the overall topography of the AF2 energy landscape.

In the above-mentioned energy threshold approach, AF2 predominantly generates the closed and open conformations, respectively, when the low-frustration sequences and high-frustration sequences with low ΔE_T were used in the MSA step (as illustrated in Fig. 3*B*). We then further investigated how the predicted structures change with the increase of ΔE_T for the input MSA sequences. We fixed the ΔE_{HF} value at ~5.0 REU and randomly sample sequences from different ΔE_T intervals. When the ΔE_T is less than a threshold value (~50 REU for AdK), AF2 predominantly returns an open conformation, which is in line with the above discussion that MSA sequences with high ΔE_{HF} tend to hint an open structure (Fig. 6*A*). Interestingly, when the ΔE_T exceeds the threshold value, both the closed conformation and open conformation can be sufficiently sampled with high probabilities. Particularly, at high E_T values, the closed conformation becomes dominant, although the open conformation can still be sampled. Such results may suggest that the closed state and open state of AdK in the AF2 energy landscape are separated by a high energy barrier (Fig. 6*B*). When the MSA inputs seed the initial structure of the AF2 prediction at the basins of the two states (with $\Delta E_T < 50$ REU), the optimization steps at the later stage can unambiguously find the optimum at the respective basins and cannot cross the energy barrier. However, when the sequences used in the MSA have high energies, further optimization either leads to the closed basin or open basin. Compared to the open basin, there is a higher probability of leading to the closed basin (Fig. 6*B*). These results may suggest that one can extract information on the overall topography of the AF2 energy landscape by integrating the physical energy landscape information with AF2 structure prediction.

Predicting Protein Conformational Changes with AF2 by Progressively Masking Highly Frustrated Sites. In the frustration-filtering strategy discussed above, the physical energy information was utilized to select homologous sequences in the MSA initialization step, by which we can generate alternative conformation and even the transition pathway with AF2. In a previous study, Stein and Mchaourab proposed an algorithm to model the alternative conformations with AF2 via the manipulation of the MSA columns (22). By mutating the residues at the interaction surface within the structure to alanine, they successfully generated alternative conformations for several proteins. One challenge of such a strategy is that prior knowledge is often needed for appropriately choosing those mutational sites in order to correctly generate the functionally relevant alternative structures. To address this issue, we explore the feasibility of utilizing physical energy information from frustration analysis to guide the manipulation of the MSA columns.

Protein energy landscape analysis suggests that native contacts specific to only a single conformational state will tend to be highly frustrated. This observation then suggests that masking the coevolutionary information from the highly frustrated sites of one conformational state (e.g., the closed state for AdK) may favor the prediction of the alternative conformational state (e.g., open state for AdK). To explore this idea, we first identified residues with high frustration according to the calculated density fraction of highly frustrated contacts (*Materials and Methods*). For these highly frustrated sites, we masked the amino acid identities and replaced them with a gap sign ("-") in the MSA (Fig. 7*A*). The MSA contains 50 sequences from the low-E_{HF} region and half of these sequences were randomly sampled for applying the site-masking. This masking process effectively reduces the coevolutionary constraints between the highly frustrated sites, potentially removing some of the bias to the existing structures in the structural prediction. We see that with the progressive masking of the highly frustrated sites, the predicted structure from AF2 for adenylate kinase changes from the closed conformation to the largely open conformation (Fig. 7 *A* and *B*). Depending on the precise criterion used in identifying the high-frustration sites, different ranges of the covered conformations are revealed. For example, when 10 sites with the highest frustration scores were progressively masked, the RMSD of the resulting predicted structures with respect to the closed structure (PDB code: 1AKE) range from 2.7 Å to 4.5 Å. In comparison, when 30 sites with the highest frustration scores are progressively masked, the RMSD

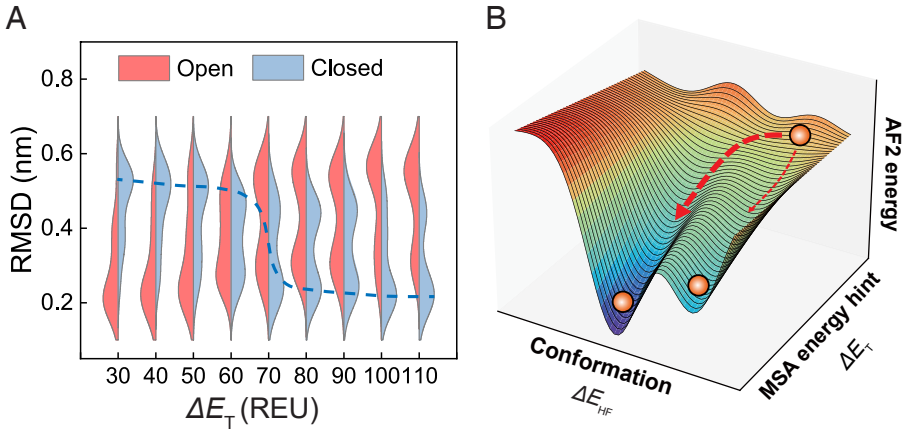

Fig. 6. Inferring the overall topography of AF2 learned energy landscape. (*A*) The distribution of RMSD with respect to the reference open structure (red) and closed structure (blue) at different ΔE_T with the ΔE_{HF} fixed at 5.0 REU. The dashed line schematically indicates the change in the peak position of the RMSD distributions. (*B*) Schematic illustration of the inferred overall topography of the AF2 energy landscape.

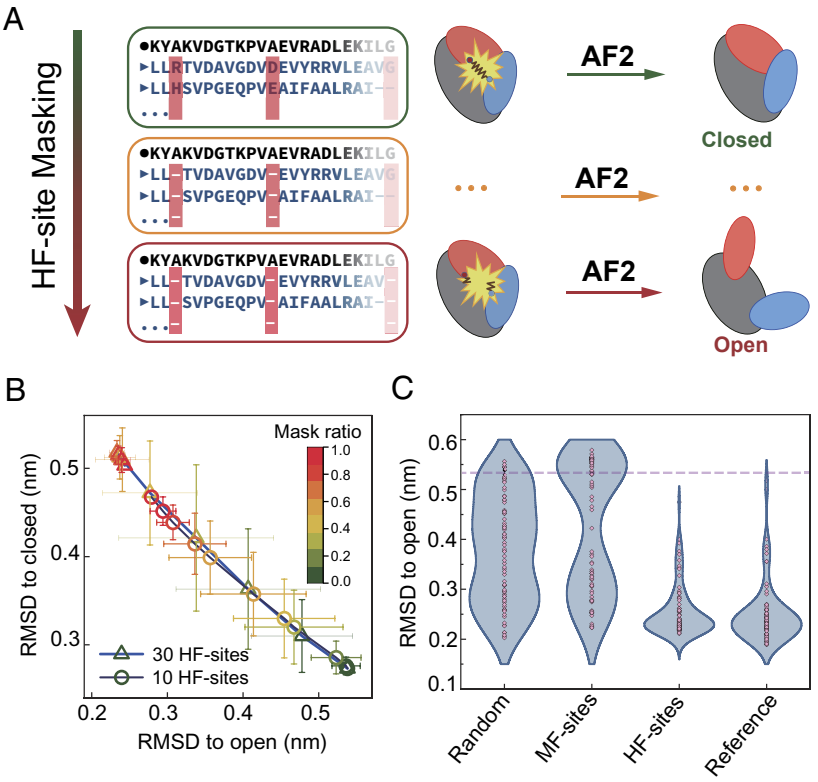

Fig. 7. Predicting conformational motions of AdK by the site-masking strategy. (*A*) Schematic illustration for the site masking strategy to generate conformational motions of AdK. (*B*) Projection of the generated structures by the site masking strategy along the RMSDs with respect to the reference closed and open structures. The error bars were calculated based on 20 independent AF2 predictions. (*C*) Distribution of the RMSDs with respect to the reference open structure for the predicted structures with different site-masking strategies, in which the masking sites were selected from the minimally frustrated sites (MF-sites), from the highly frustrated sites (HF-sites), and from the sites involved in specific contacts (Reference), respectively. For comparison, the results by randomly masking the same number of sites involved in shared contacts are also shown (Random). The dashed line corresponds to the case before site masking.

ranges further from 2.7 Å now to 5.5 Å, covering a larger range of the conformational space. The RMSD distribution of the predicted structures by masking the high-frustration sites can well reproduce the reference distribution, i.e., the RMSD distribution of the predicted structures found by masking the sites involved in the specific contacts (Fig. 7*C* and *SI Appendix*, Fig. S7). As controls, we also manipulated the MSA by masking the same number of the minimally frustrated sites and also by randomly masking the same number of the sites involved in the shared contacts. Both approaches failed to generate the reference distribution. These results again demonstrate that physical energy landscape information can be utilized to persuade AF2 to predict realistic protein motions. More detailed analysis shows that the conformations of predicted structures can be modulated by residue-resolved manipulation of MSA sequences (*SI Appendix*, Fig. S8), which may have significant implications for the prediction of the single-mutation effects on proteins with AF2.

Predicting Alternative Conformations for Other Allosteric Proteins. In the above discussions, we mainly focused on a model protein AdK. For this protein, the conformational motions can be pictured as domain motions, with negligible secondary structure changes. Recently, the AF-cluster method was successfully used to predict an alternative conformation of the protein KaiB, for which the conformational change involves fold switching between a ground state (G-state) and a fold-switch state (FS-state) (17). The standard AF2 protocol returns an FS-state structure. To further test the frustration-informed method, we employed the above-proposed strategy to KaiB. Following the frustration-filtering strategy, we separately fed the MSA sequences from different locations of the two-dimensional energy space to AF2 for structure predictions (Fig. 8 *A*–*D*). Overall, the predicted structures are closer to the reference FS-state structure (PDB code: 5JYT) for the low ΔE_{HF} MSA sequences while the algorithm led to the G-state structure (PDB code: 2QKE) when high ΔE_{HF} sequences were in MSA (Fig. 8 *B*–*D*). Correspondingly, the RMSD distributions of the predicted structures with the subcollections of the MSA sequences (with the size of 50) sampled from the "low-E_{HT}" region and the "high-E_{HT}" region are highly biased to the FS-state and G-state, respectively. In comparison, when the subcollections of the MSA sequences were sampled randomly, the AF2 dominantly gives the FS-state-like structures, and the G-state-like structures are rather rare (Fig. 8*E* and *SI Appendix*, Fig. S9). As expected, the MSA sequences generating the G-state structures tend to have higher ΔE_{HF} values (Fig. 8*F*). These results suggest that local energetic frustration information can also be used to predict protein motions involving more complex conformational changes. We also tested the above methodology for ribose-binding protein, which is another allosteric protein with large amplitude conformational motions for its biological function, and the overall correlation between the energetics of MSA sequences and the predicted structures can again be observed (*SI Appendix*, Fig. S10).

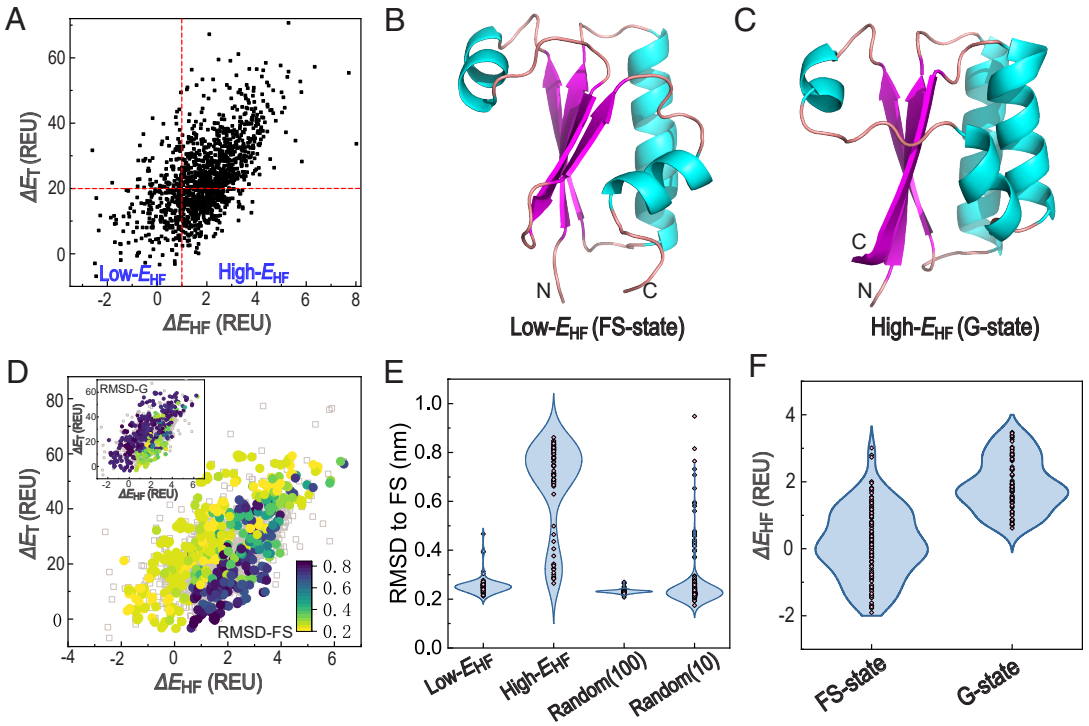

Fig. 8. Prediction of the alternative structure of KaiB with the frustration filtering strategy. (*A*) MSA sequences are projected onto the energy space formed by total energy ΔE_T and the energy of highly frustrated contacts ΔE_{HF}. (*B*) RMSD with respect to the reference fold-switch structure (RMSD-FS, PDB code: 5JYT) for the predicted structures with the input MSA sequences taken from different locations of the two-dimensional energy space. *Inset*: the RMSD with respect to the reference ground-state structure (RMSD-G, PDB code: 2QKE). (*C* and *D*) Cartoon representations of the predicted three-dimensional structures of KaiB with the low-frustration MSA sequences (*C*) and high-frustration MSA sequences (*D*), respectively. (*E*) Violin plot showing the distributions of RMSDs with respect to the reference FS-state structure for the predicted structures with the input MSA sequences from "low-E_{HF}" and "high-E_{HF}" regions, respectively. For comparison, the results for randomly selected MSA sequences with the sizes of 100 and 10 are also shown. (*F*) Violin plot showing the distributions of ΔE_{HF} for the MSA sequences generating FS-state structure and G-state structure, respectively.

Discussion and Conclusion

Great progress on the sequence–structure–function relationship has been made through the recent advent of AF2, which usually can predict three-dimensional structures with near experimental accuracy based solely on sequences. Nevertheless, to function, proteins often need to move. Therefore, predicting protein conformational motions is also fundamental to fully addressing the sequence-function relationship. In its default mode, AF2 only provides the static structure of a protein. On the other hand, energy landscape analysis of allosteric proteins has taught us the key physical rules and sequence design principles governing protein conformational motions. Directly predicting protein motions based on such physical rules using even coarse-grained simulation techniques often encounters computational challenges. In this work, we have shown that by combining AF2 and local energetic frustration scores, it is possible to directly predict the alternative structures of allosteric proteins and even the pathways of protein conformational motions from their sequences.

In this paper, we have put forward two strategies to extract information about large-scale protein motions, i.e., frustration-filtering of the MSA and frustrated site-masking. For the frustration-filtering strategy, we ranked the MSA sequences according to the contact energies of the highly frustrated native contacts. Using a model protein, AdK, as an illustration, we showed that the predicted structure by AF2 changes from a fully closed conformation to a fully open conformation with the increasing destabilization of the highly frustrated sites. Following this strategy, we developed energy threshold, sequence remixing, and sequence sliding approaches to predict the alternative conformations and to characterize the pathways of protein conformational motions. For the frustrated site-masking strategy, AF2 generates alternative structures through the progressive masking of the highly frustrated residue sites. We see that combining integrative data-driving and physics-based methodologies can generate physically meaningful pathways of conformational motions. We expect that these physical principle–guided integrative strategies will be useful for addressing the explainability issue in AF2-based structure predictions and for rational design of proteins by fine-tuning the relative distributions of different conformations. In addition to predicting protein motions, this work also provides useful implications for inferring the overall topography of the AF2 learned protein energy landscape, which has often been considered a black box, by mapping the energy space of the MSA sequences to the conformational space of the AF2 predicted structures.

Although the capability of the methodology integrating energetic frustration analysis and AF2 in predicting protein conformational motions has been clearly demonstrated for three paradigm proteins, further improvements to the implementation of the current methodology are expected in order to achieve high-throughput predictions. First, the performance of the methodology proposed in this work depends on the accuracy of the physical energy function used to calculate the frustration scores. Consequently, the calculated frustration scores may have large noise and therefore may not always fully capture the realistic frustration pattern encoded by the physical energy landscape. It is also worth noting that the proteins studied in this work involve large-amplitude conformational changes featured by low-dimensional dynamics. It seems likely, however, that many

allosteric proteins may perform their biological functions by small-amplitude conformational fluctuations, which would then involve more complex contributions of different conformational modes. In such cases, more accurate energy models and frustration calculations may be essential for correctly predicting the conformational motions. Even so, the results of this work demonstrate that the currently available energy functions and frustration quantification are already able to provide significant hints in forcing AF2 to return physically relevant predictions on protein conformational motions. Second, for many allosteric proteins, ligand binding could be the key driving force for the conformational motions. However, in the calculations of the frustration index and the AF2 structure prediction pipeline, the contribution of ligand binding is not explicitly considered. As a consequence, the alternative conformation for these proteins may not be easily captured by using the current methodology. Including the ligand binding information in the energetic analysis and structure prediction may further improve the prediction accuracy. A recent work by Qiao et al. showed that including the ligand binding information in a deep generative model NeuralPLexer, which is developed for predicting protein–ligand complex structures, can provide improved prediction of the representative structure pairs of allosteric proteins with large conformational changes (52). We expect that with further improvements in physical energy functions, frustration quantification, and ligand binding modeling, the current methodology of integrating deep-learning-based structure prediction algorithms and energy landscape ideas will provide more accurate predictions of the protein conformational motions, which may inspire progress in the rational design of functional proteins and therapeutic strategies.

Materials and Methods

Structure Prediction with AF2 Dictated by Physical Energy Landscape. The MSA and AF2 structure prediction were performed using ColabFold (53). The MSA sequences were further filtered by using the DBSCAN method, so that the final MSA sequences have high similarity to each other and all the MSA sequences are within a given distance from the query sequence (*E. coli* AdK sequence).

The frustration index and the density fraction of highly frustrated contacts around the residues of AdK were computed with Frustratometer 2 based on the AF2 prediction structure (or available crystal structure) and the query sequence (37). The frustration index for a contact formed between residue pair (i,j) is defined as the Z-score of the corresponding contact energy in the native structure compared with the distribution of decoy energies. The decoy is constructed by randomizing amino acid identities and local structures for the residue pair (i,j). The associative memory Hamiltonian water-mediated potential energy function is used in Frustratometer 2 for calculating the frustration index (54). The residue-wise frustration index is calculated in a similar way. The MSA sequences were then threaded onto this structure. The side chains of the threaded structure were relaxed by using Rosetta (with REF2015 force field) (47). The total energy of the structure for each threading sequence was calculated as $\Delta E_{\mathrm{T}} = \sum_i^N (E_i - E_{i,\mathrm{ref}})$. Similarly, the total energy of the highly frustrated contacts was calculated as $\Delta E_{\mathrm{HF}} = \sum_i^N D_{i,\mathrm{HF}} \times (E_i - E_{i,\mathrm{ref}})$. Here, $D_{i,\mathrm{HF}}$ represents the density fraction of the highly frustrated contacts within 5 Å around the residue i given by Frustratometer (37) (Fig. 2*B*). E_i and $E_{i,\mathrm{ref}}$ represent the energies of the site i for a given MSA sequence and for the reference query sequence, respectively, which were calculated as $E_i = \sum_j^N E_{ij}$ with E_{ij} being the contact energy between residues i and j given by Rosetta. N is the total residue number of a protein. By feeding the MSA sequences with different total energy (ΔE_{T}) and frustration extent (ΔE_{HF}) to AF2 together with the query sequence, we obtained a predicted structure dictated by physical energy landscape information. More detailed descriptions of the methods are given in *SI Appendix*, Text.

Data Analysis and Structure Visualization. We performed several types of PCA to characterize the conformational motions of AdK. In addition to the conventional Cartesian PCA, we also performed strain PCA and cracking PCA as discussed in previous work (51). In strain PCA, the covariance matrix was constructed based on the fluctuations of mutual distance (d_{ij}) between the C_α atoms of two residues for the predicted structures. In cracking PCA, the covariance matrix was constructed based on the fluctuations of a two-value observable for a residue pair that takes 0 ($d_{ij} > 10$ Å) or 1 ($d_{ij} \leq 10$ Å).

The three-dimensional structures were visualized by PyMOL (55). In the visualization of the frustration pattern, we calculated the smoothed frustration score $F_{ij}^S = -(\mathrm{F}_i + \mathrm{F}_j)$ for each pair of residues i and j forming a contact, where F_i and F_j are the residue-wise frustration indices of the two residues. Contacts with F_{ij}^S higher (lower) than 0.8 (−1.2) were shown by red and green lines in the cartoon illustration (Fig. 2*C*) and contact map (Fig. 5 *D*, *Lower*).

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*. The example code can be found in Github (56).

ACKNOWLEDGMENTS. The authors thank Shoji Takada for helpful discussion. This work was supported by the National Natural Science Foundation of China (Grant Nos. 11974173, 11934008, and 12305052), the grant of Wenzhou Institute, University of Chinese Academy of Sciences (WIUCASQD2021010, WIUCASQD2023015), and Hong Kong Research Grant Council (No. 22302723). P.G.W. acknowledges support for the Center for Theoretical Biological Physics from NSF grant PHY-2019745 and the support of the Bullard-Welch Chair from Grant C-0016 from the Welch Foundation to Rice. We also thank the support of High Performance Computing Center of Nanjing University and High Performance Computing Center of Wenzhou Institute, University of Chinese Academy of Sciences.

Author affiliations: [a]Department of Physics, National Laboratory of Solid State Microstructure, Nanjing University, Nanjing 210093, China; [b]Wenzhou Key Laboratory of Biophysics, Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China; [c]Department of Physics, Hong Kong Baptist University, Kowloon Tong, Hong Kong Special Administrative Region 999077, China; [d]Changping Laboratory, Beijing 102206, China; and [e]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005

1. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The energy landscapes and motions of proteins. *Science* **254**, 1598–1603 (1991).
2. D. U. Ferreiro, E. A. Komives, P. G. Wolynes, Frustration in biomolecules. *Q. Rev. Biophys.* **47**, 285–363 (2014).
3. M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing. *Proc. Natl. Acad. Sci. U.S.A* **106**, 67–72 (2009).
4. F. Morcos *et al*., Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U.S.A* **108**, E1293–E1301 (2011).
5. D. S. Marks *et al*., Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, e28766 (2011).
6. S. Ovchinnikov, H. Kamisetty, D. Baker, Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014).
7. F. Morcos, B. Jana, T. Hwa, J. N. Onuchic, Coevolutionary signals across protein lineages help capture multiple protein conformations. *Proc. Natl. Acad. Sci. U.S.A* **110**, 20533–20538 (2013).
8. G. Uguzzoni *et al*., Large-scale identification of coevolution signals across homo-oligomeric protein interfaces by direct coupling analysis. *Proc. Natl. Acad. Sci. U.S.A* **114**, E2662–E2671 (2017).
9. P. Bryant, G. Pozzati, A. Elofsson, Improved prediction of protein-protein interactions using AlphaFold2. *Nat. Commun.* **13**, 1265 (2022).
10. M. Baek *et al*., Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
11. J. Jumper *et al*., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
12. R. Nussinov, M. Zhang, Y. Liu, H. Jang, AlphaFold, allosteric, and orthosteric drug discovery: Ways forward. *Drug Discovery Today* **28**, 103551 (2023).
13. A. Jussupow, V. R. Kaila, Effective molecular dynamics from neural network-based structure prediction models. *J. Chem. Theory Comput.* **19**, 1965–1975 (2023).
14. J. P. Roney, S. Ovchinnikov, State-of-the-art estimation of protein model accuracy using AlphaFold. *Phys. Rev. Lett.* **129**, 238101 (2022).
15. J. M. McBride *et al*., AlphaFold2 can predict single-mutation effects. *Phys. Rev. Lett.* **131**, 218401 (2023).
16. D. Del Alamo, D. Sala, H. S. Mchaourab, J. Meiler, Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **11**, e75751 (2022).

17. H. K. Wayment-Steele *et al*., Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625**, 832–839 (2024).
18. D. Sala, F. Engelberger, H. Mchaourab, J. Meiler, Modeling conformational states of proteins with AlphaFold. *Curr. Opin. Struct. Biol.* **81**, 102645 (2023).
19. G. Monteiro da Silva, J. Y. Cui, D. C. Dalgarno, G. P. Lisi, B. M. Rubenstein, High-throughput prediction of protein conformational distributions with subsampled AlphaFold2. *Nat. Commun.* **15**, 2464 (2024).
20. A. Schlessinger, M. Bonomi, Exploring the conformational diversity of proteins. *eLife* **11**, e78549 (2022).
21. L. Heo, M. Feig, Multi-state modeling of G-protein coupled receptors at experimental accuracy. *Proteins Struct. Funct. Bioinf.* **90**, 1873–1885 (2022).
22. R. A. Stein, H. S. Mchaourab, SPEACH_AF: Sampling protein ensembles and conformational heterogeneity with AlphaFold2. *PLoS Comput. Biol.* **18**, e1010483 (2022).
23. J. R. Riccabona *et al*., Assessing AF2's ability to predict structural ensembles of proteins. bioRxiv [Preprint] (2024). https://doi.org/10.1101/2024.04.16.589792 (Accessed 1 May 2024).
24. X. Gu, A. Aranganathan, P. Tiwary, Empowering AlphaFold2 for protein conformation selective drug discovery with AlphaFold2-rave. arXiv [Preprint] (2024). https://doi.org/10.1101/2024.04.16.589792 (Accessed 1 May 2024).
25. M. Audagnotto *et al*., Machine learning/molecular dynamic protein structure prediction approach to investigate the protein conformational ensemble. *Sci. Rep.* **12**, 10018 (2022).
26. J. Li, L. Wang, Z. Zhu, C. Song, Exploring the alternative conformation of a known protein structure based on contact map prediction. *J. Chem. Inf. Model.* **64**, 301–315 (2023).
27. M. Hou *et al*., Protein multiple conformation prediction using multi-objective evolution algorithm. *Interdiscip. Sci.: Comput. Life Sci.*, 10.1007/s12539-023-00597-5 (2024).
28. J. Yang *et al*., Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U.S.A* **117**, 1496–1503 (2020).
29. S. Zheng *et al*., Predicting equilibrium distributions for molecular systems with deep learning. *Nat. Mach. Intell.* **6**, 558–567 (2024).
30. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct. Funct. Bioinf.* **21**, 167–195 (1995).
31. J. N. Onuchic, Z. Luthey-Schulten, P. G. Wolynes, Theory of protein folding: The energy landscape perspective. *Annu. Rev. Phys. Chem.* **48**, 545–600 (1997).
32. D. U. Ferreiro, J. A. Hegler, E. A. Komives, P. G. Wolynes, Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 19819–19824 (2007).
33. W. Li, P. G. Wolynes, S. Takada, Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc. Natl. Acad. Sci. U.S.A* **108**, 3504–3509 (2011).
34. D. U. Ferreiro, J. A. Hegler, E. A. Komives, P. G. Wolynes, On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U.S.A* **108**, 3499–3503 (2011).
35. Y. Zhang *et al*., Frustration and the kinetic repartitioning mechanism of substrate inhibition in enzyme catalysis. *J. Phys. Chem. B* **126**, 6792–6801 (2022).
36. W. Li, J. Wang, J. Zhang, S. Takada, W. Wang, Overcoming the bottleneck of the enzymatic cycle by steric frustration. *Phys. Rev. Lett.* **122**, 238102 (2019).
37. R. G. Parra *et al*., Protein frustratometer 2: A tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res.* **44**, W356–W360 (2016).
38. C. Müller, G. Schlauderer, J. Reinstein, G. E. Schulz, Adenylate kinase motions during catalysis: An energetic counterweight balancing substrate binding. *Structure* **4**, 147–156 (1996).
39. O. Miyashita, J. N. Onuchic, P. G. Wolynes, Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc. Natl. Acad. Sci. U.S.A* **100**, 12570–12575 (2003).
40. Q. Lu, J. Wang, Single molecule conformational dynamics of adenylate kinase: Energy landscape, structural correlations, and transition state ensembles. *J. Am. Chem. Soc.* **130**, 4772–4783 (2008).
41. K. A. Henzler-Wildman *et al*., Intrinsic motions along an enzymatic reaction trajectory. *Nature* **450**, 838–844 (2007).
42. P. C. Whitford, S. Gosavi, J. N. Onuchic, Conformational transitions in adenylate kinase: Allosteric communication reduces misligation. *J. Biol. Chem.* **283**, 2042–2048 (2008).
43. H. Y. Aviram *et al*., Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions. *Proc. Natl. Acad. Sci. U.S.A* **115**, 3243–3248 (2018).
44. P. Dzeja, A. Terzic, Adenylate kinase and amp signaling networks: Metabolic monitoring, signal communication and body energy sensing. *Int. J. Mol. Sci.* **10**, 1729–1772 (2009).
45. K. A. Henzler-Wildman *et al*., A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916 (2007).
46. N. Raisinghani *et al*., Integration of a randomized sequence scanning approach in AlphaFold2 and local frustration profiling of conformational states enable interpretable atomistic characterization of conformational ensembles and detection of hidden allosteric states in the ABL1 protein kinase. *J. Chem. Theory Comput.* **20**, 5317–5336 (2024).
47. C. A. Rohl, C. E. Strauss, K. M. Misura, D. Baker, Protein structure prediction using rosetta. *Methods Enzymol.* **383**, 66–93 (2004).
48. Y. Yuan, Q. Zhu, R. Song, J. Ma, H. Dong, A two-ended data-driven accelerated sampling method for exploring the transition pathways between two known states of protein. *J. Chem. Theory Comput.* **16**, 4631–4640 (2020).
49. M. Gur, J. D. Madura, I. Bahar, Global transitions of proteins explored by a multiscale hybrid methodology: Application to adenylate kinase. *Biophys. J.* **105**, 1643–1652 (2013).
50. J. A. McCammon, B. R. Gelin, M. Karplus, P. G. Wolynes, The hinge-bending mode in lysozyme. *Nature* **262**, 325–326 (1976).
51. D. A. Potoyan, C. Bueno, W. Zheng, E. A. Komives, P. G. Wolynes, Resolving the NFκB heterodimer binding paradox: Strain and frustration guide the binding of dimeric transcription factors. *J. Am. Chem. Soc.* **139**, 18558–18566 (2017).
52. Z. Qiao, W. Nie, A. Vahdat, T. F. Miller III, A. Anandkumar, State-specific protein-ligand complex structure prediction with a multiscale deep generative model. *Nat. Mach. Intell.* **6**, 195–208 (2024).
53. M. Mirdita *et al*., ColabFold: Making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
54. G. A. Papoian, J. Ulander, P. G. Wolynes, Role of water mediated interactions in protein–protein recognition landscapes. *J. Am. Chem. Soc.* **125**, 9170–9178 (2003).
55. W. L. DeLano *et al*., An open-source molecular graphics tool. *CCP4 Newsl. Pro. Crystallogr.* **40**, 82–92 (2002).
56. X. Guan *et al*., Example code. GitHub. https://github.com/Gxy-with-luv/AFF2024. Deposited 1 August 2024.