Review Article

Emerging Methods and Applications to Decrypt Allostery in Proteins and Nucleic Acids

Pablo R. Arantes, Amun C. Patel and Giulia Palermo *

Department of Bioengineering, *University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States*
Department of Chemistry, *University of California Riverside, 900 University Avenue, Riverside, CA 52512, United States*

Correspondence to Giulia Palermo: *giulia.palermo@ucr.edu (G. Palermo), @pablitoarantes (P.R. Arantes), @palermo_lab (G. Palermo)*
https://doi.org/10.1016/j.jmb.2022.167518
Edited by Igor Berezovsky

Abstract

Many large protein-nucleic acid complexes exhibit allosteric regulation. In these systems, the propagation of the allosteric signaling is strongly coupled to conformational dynamics and catalytic function, challenging state-of-the-art analytical methods. Here, we review established and innovative approaches used to elucidate allosteric mechanisms in these complexes. Specifically, we report network models derived from graph theory and centrality analyses in combination with molecular dynamics (MD) simulations, introducing novel schemes that implement the synergistic use of graph theory with enhanced simulations methods and *ab-initio* MD. Accelerated MD simulations are used to construct "enhanced network models", describing the allosteric response over long timescales and capturing the relation between allostery and conformational changes. "Ab-initio network models" combine graph theory with *ab-initio* MD and quantum mechanics/molecular mechanics (QM/MM) simulations to describe the allosteric regulation of catalysis by following the step-by-step dynamics of biochemical reactions. This approach characterizes how the allosteric regulation changes from reactants to products and how it affects the transition state, revealing a tense-to-relaxed allosteric regulation along the chemical step. Allosteric models and applications are showcased for three paradigmatic examples of allostery in protein-nucleic acid complexes: (i) the nucleosome core particle, (ii) the CRISPR-Cas9 genome editing system and (iii) the spliceosome. These methods and applications create innovative protocols to determine allosteric mechanisms in protein-nucleic acid complexes that show tremendous promise for medicine and bioengineering.

Introduction

Allostery is a fundamental process by which biological macromolecules transmit the effect of a local perturbation (e.g. a binding event) at one site to a distal, functional site, allowing for regulation of activity.[1,2] This property plays a cardinal role in biological processes, as it enables signal transduction and drug-drug synergy.[3–8] Further, many large protein-nucleic acid complexes require facile communication between multi-domain structures for proper functionality.[9–13] Thus, structural remodeling and long-range allosteric communication are the main mechanistic determinants underlying their function in cells. Processes including DNA/RNA replication, chromatin packaging, gene editing and regulation involve these complexes, demonstrating the profound biological importance of allosteric regulation. Here we focus analysis of allostery on three systems for which this paradigm holds: 1. the nucleosome core particle,[14] which is the fundamental unit of chromatin, 2. the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas9 complex,[15] which recently emerged as a transformative gene editing technology, and 3. the spliceosome,[16] which performs RNA splicing in eukaryotes (Fig.1).

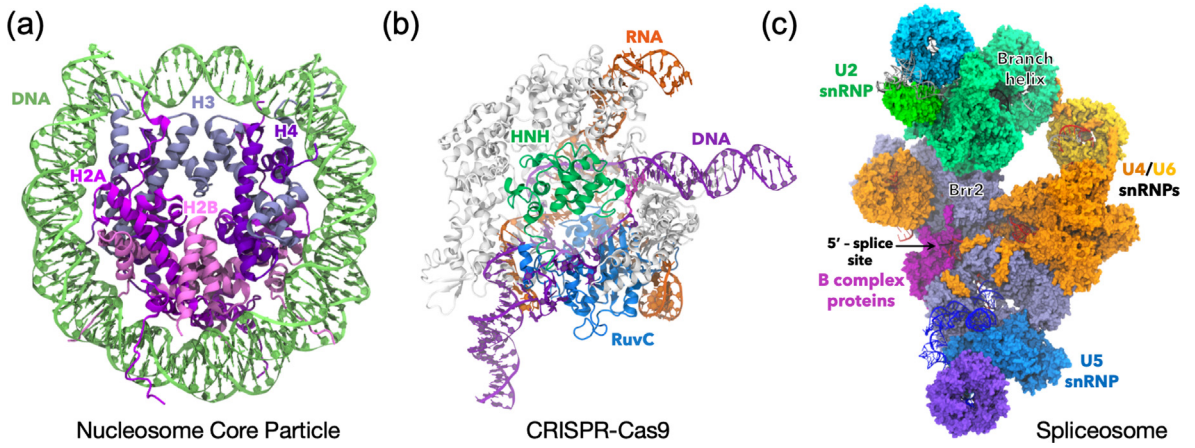

Figure 1. Overview of three allosteric protein and nucleic acid complexes. (a) The nucleosome core particle, composed of chromosomal DNA wrapped around an octamer of four core histone proteins (H3, H4, H2A and H2B, PDB: 1AOI).[14] **(b)** The CRISPR-Cas9 gene editing system (PDB: 5F9R),[46] composed of the Cas9 protein bound to RNA (orange) and DNA (purple). The HNH and RuvC catalytic domains are shown in green and blue, respectively. **(c)** The pre-catalytic spliceosome, composed of several proteins and five small nuclear ribonucleoprotein particles (snRNPs: U1, U2, U4, U5, and U6) (B complex, PDB ID: 5NRL).[47]

These systems hold tremendous potential for biomedicine, potential hampered by a poor fundamental understanding of the mechanisms that might be viable targets for their application.

Allosteric pathways in biomolecular systems have been widely investigated,[3–8,17–19] with a number of theoretical and experimental studies[20–23] revealing the significance of conformational dynamics. These studies led to the current idea that the propagation of the allosteric signaling arises from a concert of multiple timescales, over fast and slow motions, which transfer the information on multiple binaries.[24–26] This concept applies well to large macromolecular machines that process nucleic acids, where slow conformational changes and fast local perturbations could both impact the information transfer. Such phenomena can be examined in detail through the use of All-Atom MD simulations.[4–9,27,28] By capturing atomic fluctuations and conformations, MD can uniquely describe the subtle dynamical changes associated with allosteric signaling. This has greatly enriched the experimental determination of the allosteric phenomena attained through NMR, X-ray crystallography, cryo-EM and single-molecule experiments.

The data obtained from MD simulations, though beneficial on their own, can also be effectively organized into network models derived from graph theory and centrality analyses, a methodology particularly useful for examining allosteric mechanisms. Here we present synergistic schemes, such as “enhanced network models”, which enable access to long timescale dynamics and capture the relation between allostery and conformational changes. We introduce a novel approach combining graph theory with *ab-initio* MD and quantum mechanics/molecular mechanics (QM/MM) simulations to decrypt the allosteric regulation of catalysis. These “*ab-initio* network models” have been introduced to describe how long-range allosteric signals affect the DNA cleavage in CRISPR-Cas9, proposing a tense–*to*–relaxed allosteric regulation along the chemical step. Altogether, interfacing different simulation techniques with graph theory creates innovative protocols to access multiple timescales and capture the role of the signal transduction in the biophysical and chemical function of proteins and nucleic acids. The innovative methods and applications described here will inspire and impact future studies of allostery in large proteins and nucleic acid complexes.

Critical overview of the field

Nonstop development of more powerful supercomputers and smarter algorithms enables researchers to study biomolecules of increasing motion complexity, reaching timescales from nanoseconds (*ns*) to microseconds (μs). This is critical to identify transitions and short-lived conformations that escape experimental characterization.[29] Nevertheless, investigating allosteric mechanisms in large proteins through MD is often difficult. Indeed, the biological function of such large macromolecular machines relies on slow dynamical motions ranging from μs to milliseconds (*ms*), which are associated with the (re)organization of protein domains and long-range conformational effects.

This challenges the state-of-the-art computational approaches, requiring methods that enhance the sampling of the configurational space to access long timescale dynamics. Another intriguing aspect of large biomolecules is that allostery can often activate catalysis,[30] but the relation between the

two phenomena is yet to be fully clarified. This relation can rely on the transmission of signaling over short timescales.[17,24–26,30,31] This transfer of information "cliques" the catalysis, but a comprehensive characterization has not yet been given. It is also largely unknown how the allosteric regulation changes from reactants to (R) to products (P) and how it affects the transition state ($TS^{\ddagger}$). In this complex scenario, an in-depth investigation requires novel strategies that leverage a variety of computational methodologies, able to capture the different extent of the motions. Enhanced sampling simulations are required to extend the timescale limits of classical MD and capture long timescale conformational changes.[32] Furthermore, since classical MD simulations are based on force field parameterization, they are unable to describe bond breakage and formation, and cannot be employed to understand the relation between catalysis and allostery. To overcome this limitation, quantum mechanical simulations can be valuable to study the allosteric response along the catalysis. These simulations allow one to integrate the equation of motion from the first quantum mechanical principles and are commonly referred as *ab-initio* MD simulations.[33]

Other long-lasting limitations associated to classical MD refer to the accuracy of the empirical force field in reproducing the experimental (or quantum mechanical) properties. Though ceaseless refinement of the most popular force field models – OPLS,[34] AMBER,[35] GROMOS[36] and CHARMM[37] – is leading to good representations of proteins. For nucleic acids, the description of base-pairing, stacking and base/sugar interactions can be challenging. Improvements to the Cornell et al. model[38] led to the correction of several parameters, from van der Waals to electrostatic terms and dihedral parameters. These include the bsc0 corrections of the unbalanced α/γ transitions in DNA,[39] and the χOL3 corrections for RNA,[40,41] which balances the anti and the high-anti conformation related to the χ angle. Studies summarized here have indicated the reliability of these models for protein/nucleic acid complexes. Moreover, computational studies of protein allostery can be fundamentally assisted by solution NMR.[42] In this respect, recent force field models have improved the consistency of the backbone conformational ensemble with NMR experiments,[43] and were used in combined MD-NMR studies discussed here.[44,45]

Overall, considering these challenges, our contribution accounts methods and innovative approaches to efficiently decrypt dynamic allostery in proteins and nucleic acids. Methods and applications showcased here will help in creating novel protocols to determine the allosteric network of communication over multiple scales, as well as the relation between allostery and catalysis, which has remained unaddressed through classical approaches.

Allosteric proteins and nucleic acids

Here, we introduce the biomolecular function of the three proteins-nucleic acid complexes that will be used as paradigmatic examples of allosteric mechanisms.

The nucleosome core particle is the fundamental unit of chromatin, composed of chromosomal DNA of 145–147 base pairs, wrapped around an octamer of four core histone proteins (H3, H4, H2A and H2B, Figure 1(a)).[14] Nucleosomes have the essential role of compacting DNA in eukaryotic cells, where the majority of DNA assumes a packed conformation, rather than as free oligonucleotides or non-protein–DNA complexes. The allosteric regulation of nucleosomes have been well characterized,[11,48] revealing the effect of local alterations of the histones' dynamics on the protein and DNA binding properties at distal sites. Moreover, drug targeting of nucleosomes can exploit histones' allostery to achieve drug-drug synergistic effects.[49,50] This regulation is promising for developing new therapeutic strategies that could interfere with chromatin compaction.[51,52] In fact, allosteric drugs can hinder the binding of chromatin transcription factors that modulate gene expression in cancer cells, leading to new strategies for anticancer therapy.

CRISPR-Cas9 is the core of a transformative genome editing technology that is innovating biomedicine, pharmaceutics and agriculture.[15] Cas9 is an RNA-guided DNA endonuclease, which generates double-stranded breaks in DNA by first recognizing its protospacer-adjacent motif (PAM) sequence, and then cleaving the two DNA strands via the HNH and RuvC nuclease domains. The large multi-domain Cas9 protein comprises a recognition (REC) and a nuclease (NUC) lobe, the latter including the catalytic domains and the PAM-interacting region (Figure 1(b)).[46] At the molecular level, an intricate allosteric signaling regulates the CRISPR-Cas9 biochemical information transfer to activate double-stranded DNA cleavage.[53,54] This allosteric communication is critical for transmitting the DNA binding information, affecting the function and specificity of Cas9. Its knowledge is essential for the system's activation and for improving its genome editing applicability.

The spliceosome, a multi-mega Dalton assembly of proteins and small nuclear RNAs, is one of the most important non-coding RNA–protein complexes in humans (Figure 1(c)).[16] By performing RNA splicing, the spliceosome edits the premature messenger RNA (pre-mRNA), which is cleared of its non-coding sections (introns). The coding exons are then ligated, forming messenger RNA (mRNA) which is subsequently translated into proteins. The spliceosome is composed of hundreds of proteins and five small nuclear ribonucleoprotein particles (snRNPs: U1, U2, U4, U5, and U6), which undergo a continuous conformational and composi-

tional remodeling during the splicing cycles.[47] Understanding the inter-protein/RNA communication can pose the foundations to interfere with the spliceosome function, a clinically significant goal considering splicing deregulation is associated with more than 200 human diseases.

Allostery as a violin or a domino

The early models on protein cooperativity have been a stepping stone to the understanding of allostery as a fundamentally a dynamic phenomenon.[1,55,56] It is now established that allostery is characterized by a change in the dynamical properties. Indeed, the binding of an allosteric effector (e.g., a drug or a substrate) can induce major conformational rearrangements or subtle shifts in the conformational ensemble, resulting in the transmission of the binding information to distal sites.[5] This favorable free energy change, induced by effector binding, can be characterized by enthalpy, which implies an explicit change of conformation, or by entropy, as in the case of allostery without an explicit conformation change.[57]

A brilliant interpretation of the signal transmission suggested that this phenomenon can resemble a "violin" or a "domino".[58] In the violin model, the binding of an allosteric effector (i.e., a drug or substrate) triggers a pattern of vibrations, which results in the activation of the system at distal sites. This behavior is analogous to a violin, when the player pitches a string (ligand binding) and triggers a pattern of vibrations that transfers itself to the sounding board (the biomolecule) for function. Substrate binding thereby shifts the conformational ensemble by altering the system's motions in a non-specific way. On the other hand, in a domino model effector binding triggers a sequential set of local events propagating via a well-defined pathway from one allosteric active site to the other. Two examples that show the difference in mechanism between the "violin" and "domino" models are CRISPR-Cas9 and the nucleosome core particle (Figure 2).

In the CRISPR-Cas9 complex, the binding of the PAM recognition sequence triggers coupled motions within the protein framework, resulting in a shift in the conformational dynamics (Figure 2 (a)).[10] In this case, PAM is the allosteric effector "pitching" the right cord and triggering a pattern of vibrations that transfers across the system. In-depth analysis of the dynamics of CRISPR-Cas9 bound to PAM and of its analogue crystallized without the PAM sequence revealed that PAM binding substantially strengthens inter-domain correlations between the RuvC and HNH catalytic domains. This suggested that PAM binding activates the system for concerted cleavages of the two DNA strands. Analysis of the allosteric inhibition in CRISPR-Cas9 also suggested a violin model, where the inhibition shifts the conformational ensemble toward a less catalytically competent state.[59] In the nucleosome core particle, the binding of two unrelated

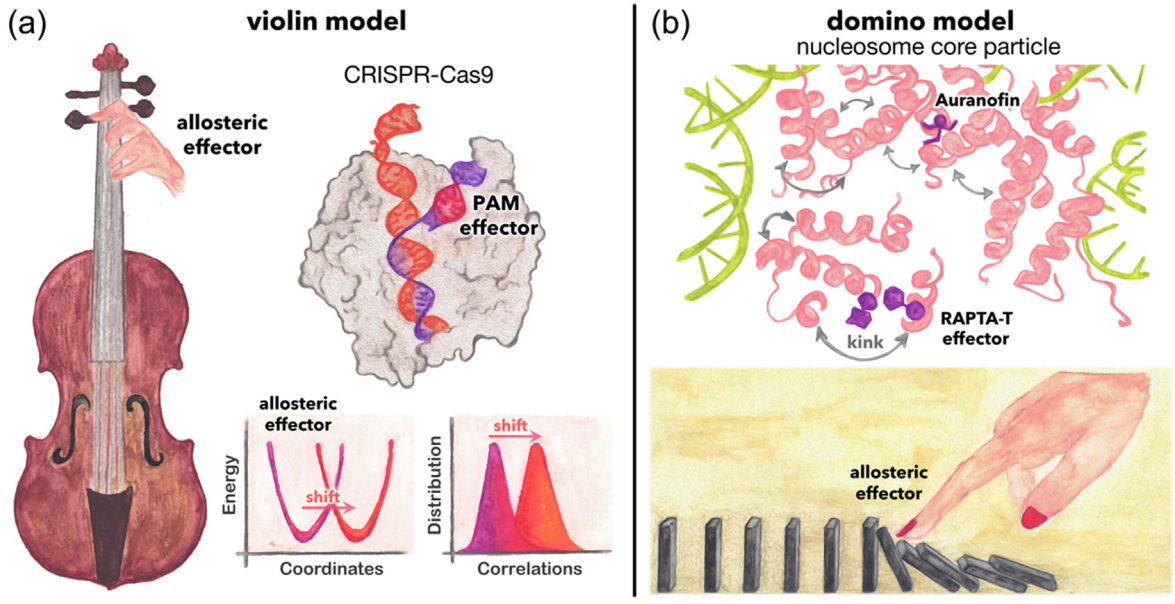

Figure 2. Violin and domino models for biomolecular allostery.[58] **(a)** In a violin model, the binding of an allosteric effector triggers a pattern of vibrations, similar to the player's pitch on a string, leading to the activation of the system at distal sites. In CRISPR-Cas9, the binding of the PAM recognition sequence (i.e., the allosteric effector) leads to a change in the conformational dynamics, as indicated by a shift in the free energy basin and in the correlated motions (bottom panel).[10] **(b)** In a domino model, the allosteric effector triggers a sequential set of local events propagating via a well-defined pathway, similar to the effect of a hand touch to a domino. In the nucleosome core particle, the biding of RAPTA-T induces a local kink, that is dynamically coupled to a path of adjacent α-helices (shown using arrows), allowing the transfer of information to the auranofin site.[49]

metal compounds – *viz.,* RAPTA-T and auranofin – that yield a synergistic activity in killing cancer cells, has shown to activate an allosteric response though a mechanism that resembles the "domino" model (Figure 2(b)).[49] Crystallographic studies have shown that RAPTA-T and auranofin bind the nucleosome core particle located at a $\sim$40 Å distance from each other. Hence, experimental characterization of the synergistic activity was limited. Molecular simulations revealed that the binding of RAPTA-T induces a local kink in the α-helix of the H2A histone, which was not observed in the absence of RAPTA-T. This kink was found to be dynamically coupled to a path of adjacent α-helices, allowing the transfer of information from RAPTA-T to the auranofin site. This suggested that the signal transmission could mainly occur through a domino of events, where a few pathways with particularly high local correlations result in a defined sequence of correlations among α-helices connecting the RAPTA-T and auranofin sites. These findings led to cross-linking the allosteric sites, increasing nucleosome stability and proposing a novel strategy for therapeutic applications.[50]

Interestingly, in this system, the signal transmission was also shown to arise from the reorganization in coupled motions, and thereby not limited the domino model. In this respect, it is important to note that the violin vs. domino representation is not absolute and there is no marked physical difference between the two models. As the effector binds, a dynamic perturbation is observed, which could manifest as a pattern of non-specific vibrations and can also hold more specific information. This indicates that allosteric communication lands on a continuum between the violin and domino model. Both models are an effect of the change in free energy induced by effector binding. An important aspect of this effect is causality, which links the perturbation to the allosteric communication. A recent theoretical model reveals the causal relation between effector binding and allosteric communication on a per-residue basis.[60–62] In this model, an "allosteric potential" is used to measure the signaling on a residue, as a result of the conformational changes of its neighborhood. Then, the entropic contribution to the allosteric free energy of the residue is computed by comparing the conformational ensembles in the effector-free and effector-bound systems. This effectively tracks the causal relation between effector binding and the per-residue allosteric communication.

Long-range communication and mutual dynamics

The examples above clearly pinpoint the importance of dynamic cooperativity for biomolecular allostery. Correlation analysis allows detecting the presence of possible dynamic correlations among spatially distant sites, and the molecular elements responsible for the "signal transmission" between them. Dynamic correlations can be detected through Cross-Correlation (CC) analysis, by computing the Pearson's correlations between the fluctuations of the Cα atoms relative to their average position. Cross-correlations measure the collinear coupling between two atoms, determining whether they tend to move in lockstep (positive CC) or show opposed motions (negative CC).

This analysis neglects correlated motions occurring out of phase, prompting the introduction of a Generalized Correlation (GC) method.[63] This approach measures the degree of correlation between residues based on their mutual information, capturing also non-linear correlations. Here, the correlations of fluctuations in the positions of Cα atoms is based on the GC coefficients, namely $r_{MI}[x_i, x_j] = [1 - \exp(-\frac{2}{3})I[x_i, x_j]]^{\frac{1}{2}}$ and is computed in terms of mutual information $I[x_i, x_j] = H[x_i], +H[x_j] - H[x_i, x_j]$. Here, $H[x_i]$, $H[x_j]$ are the marginal Shannon entropies, while $H[x_i, x_j]$ is the joint Shannon entropy for atomic vector displacements (x_i, x_j). These are computed as ensemble averages over trajectories from multiple replicates. This provides a method for detecting any type of dependence in the atomic motions, regardless of the direction of motion. Notably, the GC method uses the mutual information to discretize positional displacements. This prompted alternative approaches to also consider the dihedral space as a valuable descriptor of coupled dynamics.[64] Overall, the GC method is a normalized measure of how much information on one atom's position is dependent on that of another atom. It does not inform the sign, nor the direction of the correlations.

The synergistic application of CC and GC analyses is useful to characterize any type of correlation and how it contributes the signal transmission. An example is given by the study of the conformational dynamics of CRISPR-Cas12a, a recent CRISPR system that expands applications to nucleic acid detection (Figure 3(a)).[65] CC analysis revealed that the REC lobe of Cas12a, which mediates nucleic acid binding, moves in the opposite direction with respect to the NUC lobe, suggesting an "open-to-close" conformational transition for nucleic acid binding. CC analysis did not detect relevant differences in the RNA- and DNA-bound states. On the other hand, GC analysis revealed an overall increase in the system's correlations upon DNA binding, suggesting a DNA-induced allosteric dynamic switch to favor the conformational activation of Cas12a toward DNA cleavage. Taken together, the CC and GC methods inform how coupled motions contribute to the allosteric phenomenon. Since both methods provide information on the per-residue correlations, it is difficult for very large biomolecules to localize

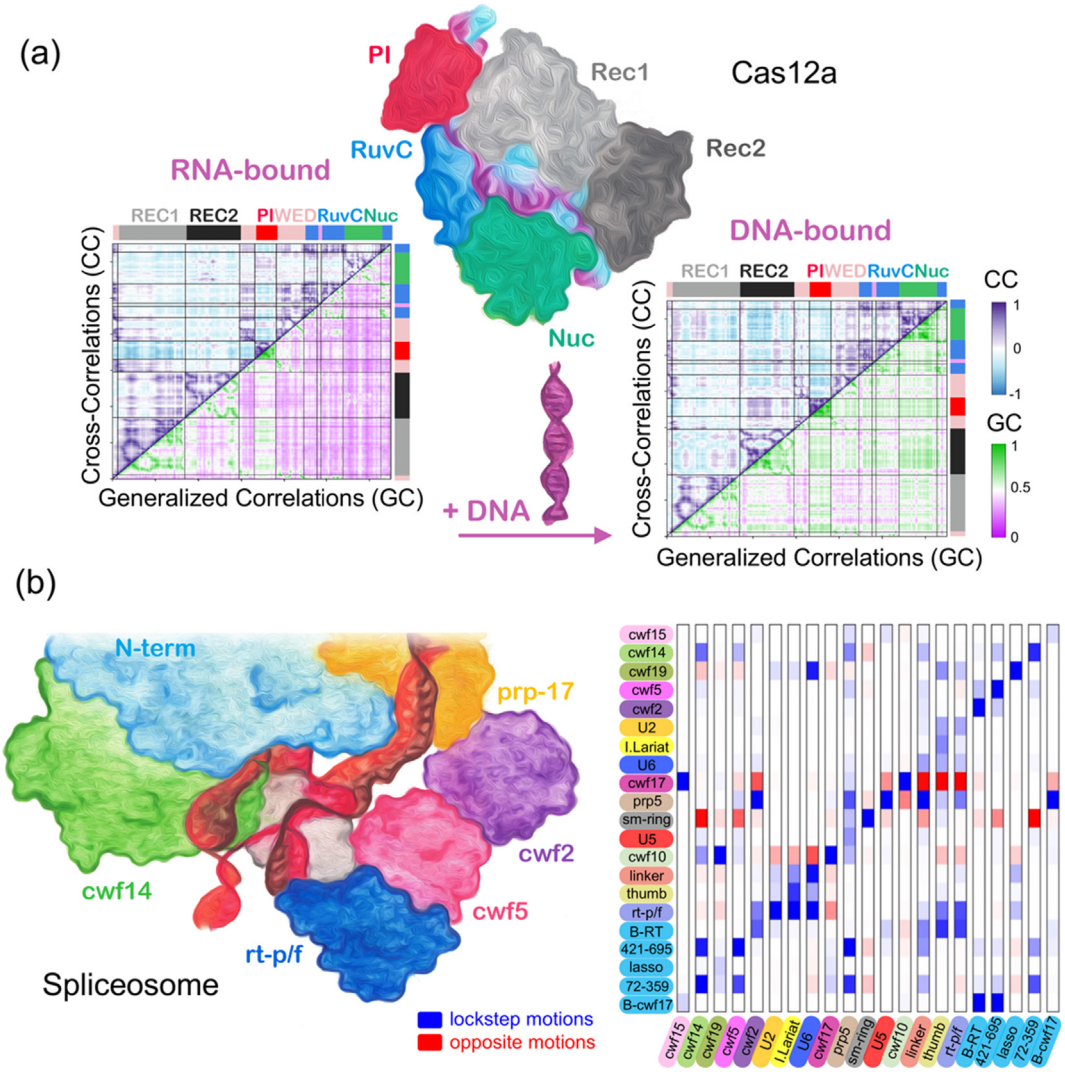

Figure 3. Correlation analysis. (a) Coupled motions in the CRISPR-Cas12a system, suggesting an "open-to-close" conformational transition for nucleic acid binding.[45] Cross-Correlation (CC, upper triangles) and Generalized Correlations (GC, lower triangles) matrices, computed for Cas12a in the RNA-bound (left) and DNA-bound (right) states (color-coded according to the scales on the right). DNA binding induces a sensible increase in GCs. Adapted with permission from Saha et al. (2020).[65] Copyright 2020 American Chemical Society. https://pubs.acs.org/doi/full/10.1021/acs.jcim.0c00929. Further permissions related to the material excerpted should be directed to the American Chemical Society. **(b)** Per-domain CC histogram of the spliceosome dynamics.[46] The inter-domain cross-correlations reveal domains moving in lockstep (blue) and through opposite motions (red). Adapted with permission from Casalino et al. (2018).[65] Copyright 2018 National Academy of Sciences.

the correlations effectively responsible for the signaling, due to a high noise. To overcome this limitation, a coarse representation of the correlation matrix, which accumulates the inter-domain correlations, has been introduced to provide a measure of the per-domain correlations.[10] This approach has been useful in detecting allosteric responses in large systems, where >1500 amino acids generate a large noise. An example is the human spliceosome, where a per-domain correlation matrix has been instrumental in describing the functional dynamics (Figure 3(b)).[66]

Using graphs to describe communication networks

Graph theory is a sub-discipline of mathematics and computer science pioneered by Leonard Euler in the eighteenth century. Euler was a mathematician in Königsberg (now Kaliningrad, Russia), where the Pregel river dissected the city in a peculiar fashion and was crossed by seven bridges. Euler answered the question of whether it is possible to find a path crossing each bridge exactly once, and in doing so pioneered graph

theory. This field experienced an enormous development, leading to the widespread use of networks in communication science (e.g., for social media), economics, geology, physics, and systems biology.[67]

Network models for protein allostery build on correlation analysis to construct a network of long range interactions, which efficiently describes the allosteric transmission.[9] In a typical dynamical network model, the biomolecular system is described as a graph of nodes and edges (Figure 4(a–b)), where nodes represent the amino acids (Cα atoms) and the nucleotides (P atoms, N1 in purines and N9 in pyrimidines), while edges denote the connection between them. An edge's length is weighted as a function of the strength of the correlations, placing strongly correlated nodes close to each other (i.e., displaying shorter edge-length). This network model has been developed building on Pearson's cross-correlations.[9] Recent applications have shown that a GC-based dynamical network analysis sensibly improves the description of the allosteric network.[10,68,69] Indeed, the GC method does not only provide a more complete estimation of the overall coupled motions but, building on Shannon's entropy, it allows for a direct evaluation of the system's entropy redistribution, induced by effector binding.

As noted above, the edge lengths are obtained by "weighting" the system's correlations, with the weight (w_{ij}) of the edge connecting nodes *i* and *j* computed as $w^0_{(ij)} = -\log[r_{MI}(x_i, x_j)]$. The resulting "*weighted graph*" defines the system as a dynamical network, with information on the critical nodes that are important for the communication within the complex. The weighted network is then structured in a set of "*communities*", groups of nodes in which the network connections are dense but between which they are sparse (Figure 4(c)). These local substructures of highly correlated residues can be obtained through the Girvan-Newman algorithm,[70] a divisive algorithm that uses the "*edge betweenness*" (EB) partitioning criterion. The EB is the number of shortest pathways that cross the edge, and is computed using the Floyd-Warshall algorithm,[71] which sums the lengths (w_{ij}) of all edges in different paths of nodes, and identifies the pathway displaying the shortest total length. In a typical community network plot (Figure 4(c)), the communities are linked by bonds whose thickness is proportional to the total EB, indicating the strength of the communication between communities. Recently, a community network model of the transcription preinitiation complex was constructed.[72] This work delivered critical insights into the biomolecular function of an important complex involved in the expression of protein-encoding genes, demonstrating the tremendous value of MD simulations in combination with graph theory.

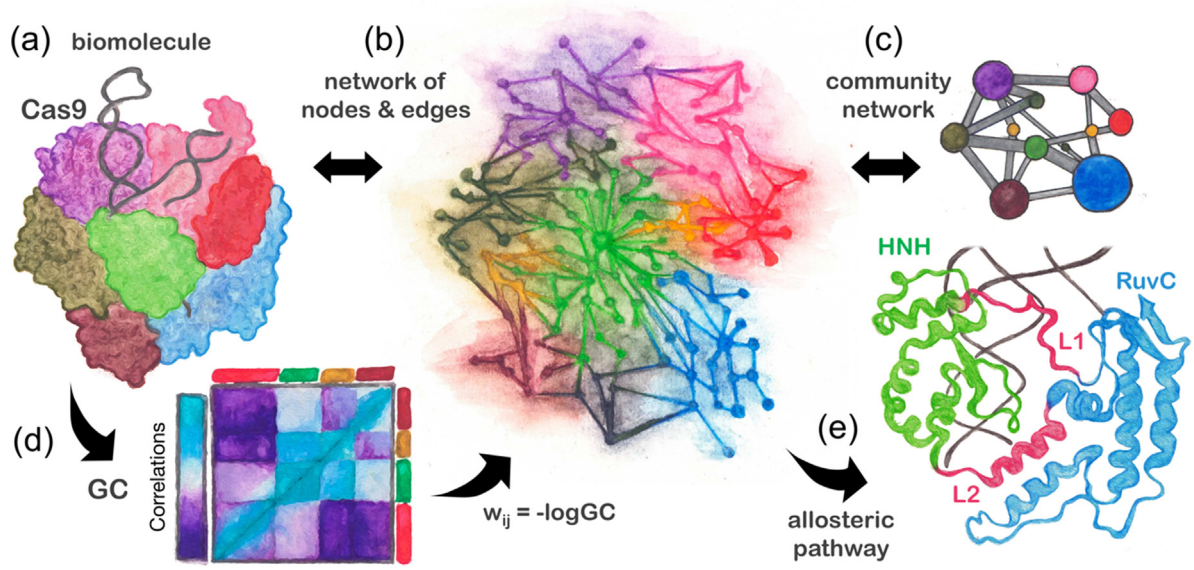

Figure 4. Network models for biomolecular allostery, shown for the CRISPR-Cas9 system. The biomolecule **(a)** can be described as a network of residue nodes and edges whose length is weighted by the strength of the residues' correlations **(b)**, and as a network of communities connected by bonds measuring their intercommunication strength **(c)**.[9] The network model builds on correlation analysis **(d)**, whereby the Generalized Correlations ($GC = r_{MI}[x_i, x_j]$) are used to weight the edges connecting nodes ($w_{ij} = -\log GC$), such placing strongly correlated nodes close to each other. **(e)** From the network model, the shortest pathways crossing the edges between distal sites can be computed as efficient communication routes among allosteric sites. This is shown for the L1/L2 loops in CRISPR-Cas9, connecting the HNH and RuvC domains as shortest pathways.[10]

Circular networks to depict allosteric gain and loss

The EB is an important measure of the "traffic" passing through edges. It accounts for the number of times an edge acts as a bridge in the communication flow between nodes of the network. Hence, the total EB between couples of communities (i.e., the sum of the EB of all edges connecting two communities) is an important measure of their communication strength. In a recent study, this measure has been used to construct circular networks of the mutation-induced allosteric gain/loss (Figure 5).[44]

In detail, in a study of improved specificity by lysine-to-alanine mutations in CRISPR-Cas9, the mutation-induced EB change (ΔEB) was computed as a difference between the EB of the mutant and the WT system. The normalized ΔEB were plotted using circular networks, where the communities are displayed in a circle and connected using links with thickness proportional to the ΔEB. Negative values of ΔEB ($-1 < 0$, red) represent loss of communication, as opposed to positive values ($0 < 1$, blue), which indicate a communication gain upon mutation. As a result, a dramatic loss of communication was observed between the allosteric communities that connect the functional sites (i.e., the A1–A3 communities). On the contrary, the non-allosteric sites (NA1–NA4) gained in communication, overall indicating that mutations increasing Cas9 specificity also perturb its allosteric signaling. In summary, circular networks are useful to spot the allosteric gain or loss upon mutation in biomolecular systems.

Shortest paths linking allosteric sites

A useful applications of network analysis is the identification of "*shortest pathways*" between distal sites through the Floyd-Warshall algorithm. These pathways are likely to be efficient communication routes among allosteric sites, as shown for the allosteric pathways communicating the HNH and RuvC catalytic sites in CRISPR-Cas9.[10] Calculation of the shortest routes for information transfer revealed that the crosstalk between HNH and RuvC flows through the L1/L2 loops (Figure 4(e)), which have been indicated as "signal transducers". Structural studies supporting this notion[46,73] revealed that the L1/L2 loops allow the conformational activation of the HNH domain toward DNA cleavage. The Floyd-Warshall algorithm was also employed for the identification of the shortest pathways in the spliceosome[12] and its components.[74] The principal routes of communication identified the splicing cofactors as signal conveyors for pre-mRNA maturation. This is a paradigmatic example of how the signaling transfer may not happen exclusively through a single optimal path, granting in-depth investigation of the alternative or sub-optimal paths.

For increasingly complex biological systems, however, the number of paths between distal nodes increases with the total number of interconnected nodes and poses a combinatorial problem. Hence, the identification of the near

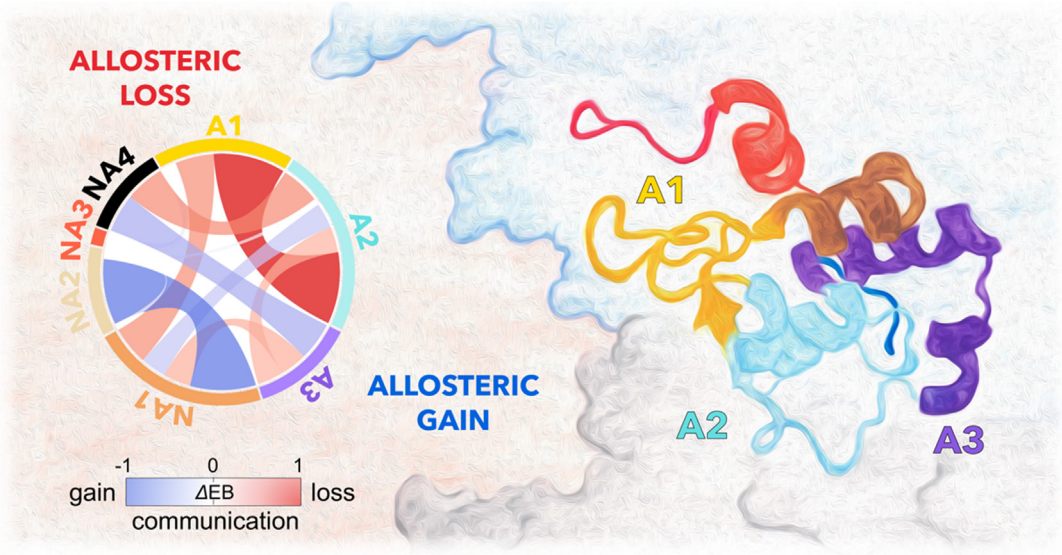

Figure 5. Circular network of the allosteric communication, reporting the mutation-induced Edge Betweenness change (ΔEB), a measure of communication gain or loss between couples of communities upon mutation.[44] The communities are displayed in a circle and are connected by links with thickness proportional to ΔEB. Communities connecting allosteric sites (A1–A3) display a loss of communication (negative ΔEB, red), while the non-allosteric sites (NA1–NA4) gain in communication (positive ΔEB, blue). Circular network adapted from Nierzwicki et al. (2021)[44], published in eLife under a Creative Commons Attribution license. https://elifesciences.org/articles/73601.

optimal pathways for the signaling transfer might quickly become computationally intractable as the number of nodes increases.[75] Recent contributions have shown that alternative path searches are valuable to overcome this possible limitation. Among them, distance fluctuation analysis[76,77] and machine learning-based analyses,[78,79] which can provide excellent agreement with experimental testing, holding a tremendous potential in the drug discovery field. An efficient algorithm for path search in large biomolecular systems is the scheme proposed by Dijkstra,[80] which is widely used in cartography to find the shortest roads leading to the desired destination. The Dijkstra's algorithm defines a starting and destination point and optimizes iteratively a pathway from the former to the latter (Figure 6(a)).

For protein allostery, the algorithm uses the correlation coefficients as a metric to define the iterative optimization problem. It finds the roads, composed by the $w^0_{(ij)}$ inter-node connections, which minimize the total distance (and so maximize the correlation) between nodes. In this way, the Dijkstra's algorithm finds the roads that optimize the momentum transport between distal sites (nodes) and therefore are efficient signaling pathways. The Dijkstra's algorithm was applied to identify the allosteric signaling within the HNH nuclease of CRISPR-Cas9 (Figure 6(b)).[23,45] Dijkstra revealed a potential route of signal transduction connecting the DNA recognition region to the catalytic sites. This signaling route, comprising the top ten most likely pathways, displayed a remarkable overlap with the slow residues of HNH identified through NMR Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments. Altogether this suggested a mechanism of activation, where the transfer of the DNA binding information is critical to activate DNA cleavage.

Central residues in the allosteric network

One of the cornerstones of the network theory is the concept of centrality,[81] i.e. the relative influence of a node or cluster of amino acid nodes in the network.[82–84] The application of graph theory to social media emphasizes the importance of centrality in the information transfer. In a social network, some friends hold more connections, becoming the hubs of the communication, where the information centralizes and transfers more efficiently. Analogously, central residues are the hubs governing the biomolecular dynamics. Three main measures can be harnessed to define centrality: (i) degree centrality (DC), (ii) betweenness centrality (BC) and (iii) eigenvector centrality (EC). DC is simply the number of edges hold by a node and can be interpreted as a local centrality measure. BC is the number of shortest pathways passing through a node and quantifies the number of times a node acts as a bridge along the shortest path between two other nodes. BC is the most popular centrality measure and has been applied in studies of allostery in CRISPR-Cas9,[10] the nucleosome core particle[48] and the spliceosome.[12,74] The EC is the third centrality measure, which defines influence of a node in the network, weighting the nodes based on their contribution to the system's dynamics. This approach relies on assigning the functional dynamics to the major collective mode of the system, i.e., the first eigenvector of the adjacency matrix A (based on the generalized correlations r_{MI}). The EC of a node, c_i, is defined as the sum of the centralities of all nodes that are connected to it by an edge, $c_i = \frac{1}{\lambda}\sum_{j=1}^{n} A_{ij}c_j$, where the edges A_{ij} are elements of the adjacency matrix A and λ is the eigenvalue associated to the eigenvector composed by c_i elements. The EC estimation quantifies the degree of connectivity of each amino acid (or nucleobase) within the system and quantifies how well nodes are interconnected. Hence, this measure holds significant promise to identify the hubs of the signal transduction in protein/nucleic acid complexes.

Application of the centrality analysis to the HNH–RuvC cross-talk in CRISPR-Cas9 revealed that the E584, Q771, K775, and R905 residues of the L1/L2 loops act as central hubs of the dynamics.[10] Mutations of the central node residues, through which the majority of allosteric pathways pass, reported an increased specificity, observed for the K775A and R905A residues in the eCas9[85] and HypaCas9[86] variants. This indicates that altering the central hubs of the dynamics affects the system function, also suggesting the targeting of the allosteric regulation as a strategy for the specificity enhancement.

Enhanced network models and conformational control

In large biomolecular complexes, such as those of multidomain proteins with nucleic acid elements, the conformational dynamics are characterized "*per se*" by slow dynamical motions, affecting the transmission of the allosteric response that occurs over longer timescales. Understanding this conformational control of the allosteric response is difficult though conventional MD simulations, which are notoriously limited to short *ns*–to–μ*s* motions. Another hurdle of MD simulations of large ribonucleoproteins is that the protein components are tightly entangled to nucleic acids, requiring extensive sampling for converged results.

Considering the above, to investigate the allosteric response in large biomolecular systems, it is essential to capture their long timescale dynamics, and to combine enhanced simulations methods with graph theory. Through "enhanced network models", the conformational landscape obtained through enhanced sampling methods

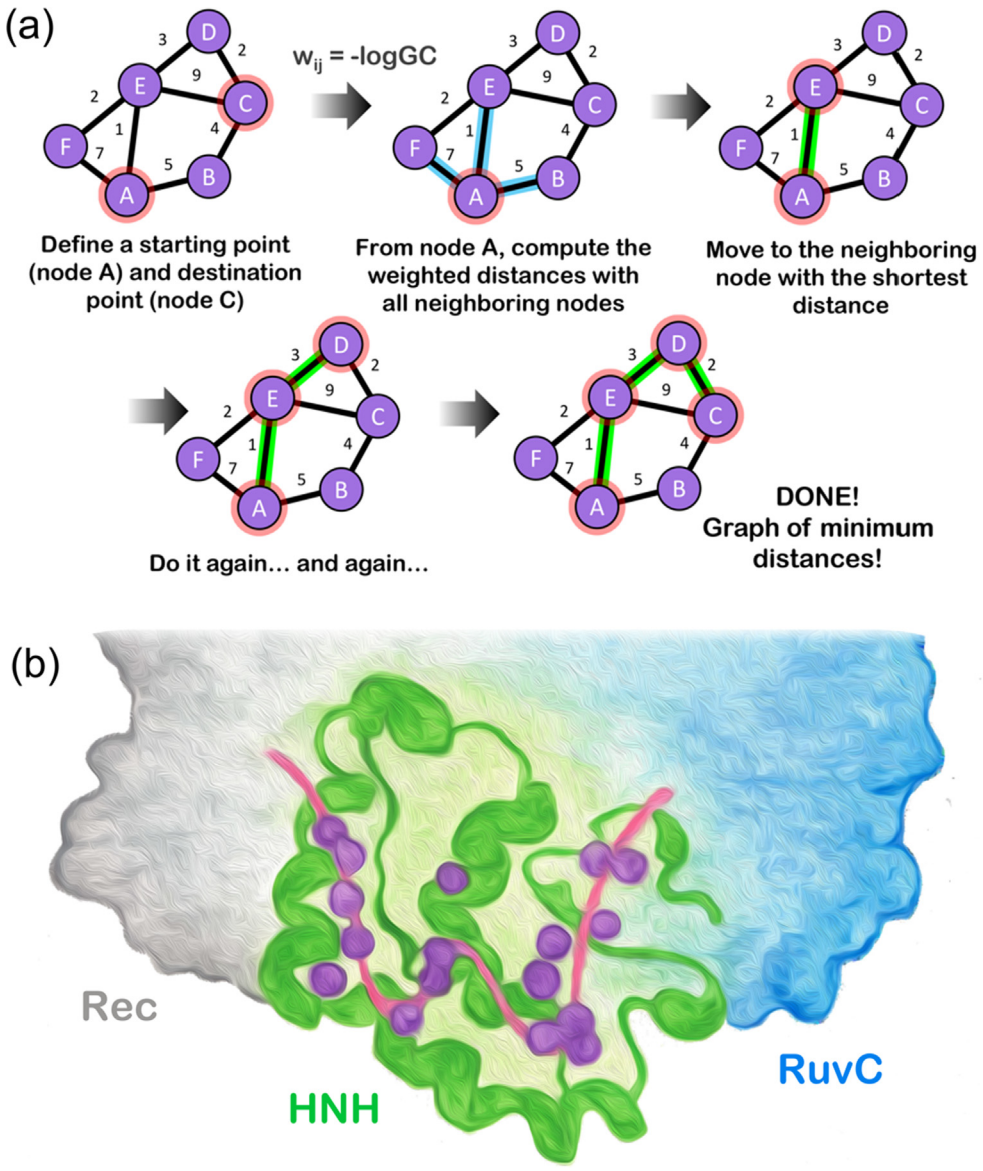

Figure 6. Shortest paths connecting allosteric sites. (a) Dijkstra algorithm for shortest path calculation. The algorithm defines a starting and a destination point (i.e., nodes A and C) and optimizes iteratively a path from the former to the latter. In each iteration, the closest unvisited node is designated as the current node, updating the remaining unvisited nodes until the destination is reached. For biomolecular allostery, the algorithm uses the correlation coefficients as a metrics to identify the closest nodes (i.e., $w_{ij} = -\log GC$), maximizing the correlation between starting and destination nodes. **(b)** Allosteric pathway within the HNH domain of CRISPR-Cas9 connecting the DNA recognition region (REC2) to the RuvC cleavage site. The signaling route identified through the Dijkstra algorithm (pink line) overlaps with slow dynamical residues found through solution NMR (purple spheres).[23]

can be used as a basis to construct the allosteric network (Figure 7(a)). In this way, the allosteric network preserves the memory of the long timescale conformational change. Enhanced network models can also be constructed through Markov State Models of biomolecular allostery, to elucidate the kinetic pathways connected to the allosteric transmission.[87–89] Recent studies combined a Gaussian accelerated MD (GaMD) method[90] with network models to characterize the allosteric response over long timescales. GaMD adds a harmonic boost potential to smoothen the potential energy surface of the simulation system, accelerating the transitions between low-energy states. In GaMD, when the threshold energy is set to the maximum potential ($E = V_{max}$), the system's potential energy surface is smoothened by adding a harmonic boost potential that follows a Gaussian distribution (Figure 7(b)). Briefly, when the system potential $V(\vec{r})$ is lower than a threshold energy E

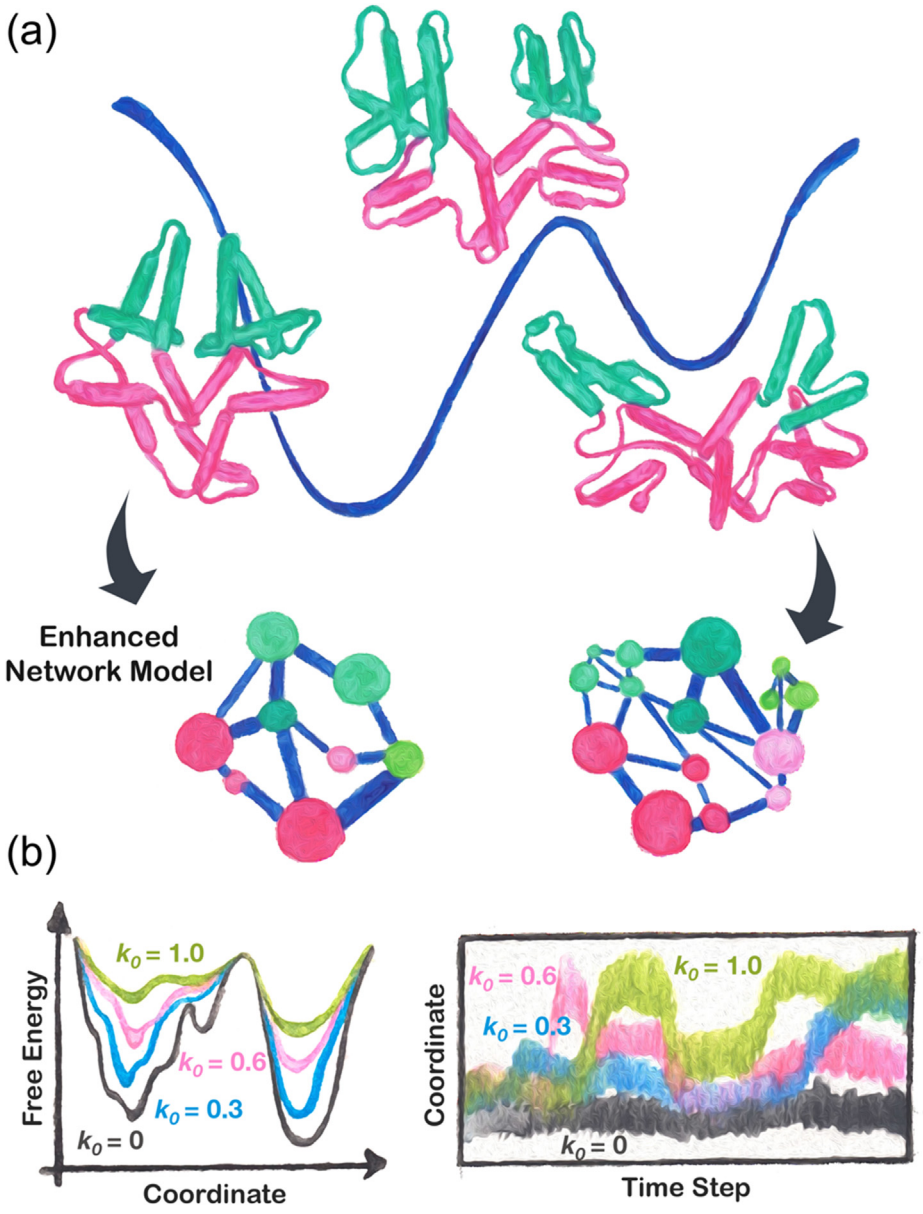

Figure 7. Enhanced network models. (a) The conformational landscape obtained through enhanced sampling molecular dynamics (top) is used as a basis to construct the allosteric network (bottom). Enhanced network models are constructed by applying graph theory on the reweighted conformational landscape, which reports the canonical ensemble. Inspired by Wereszczynski & McCammon (2012) *Proc. Natl. Acad. Sci. USA* 109, 7759–7764.[94] **(b)** Gaussian accelerated MD (GaMD) method.[86] Quadratic functions are used to modify the original potential energy of the system to overcome energetic barriers. The extent of acceleration is controlled by the harmonic constant k_0, varying from 0 to 1. The greater the value of k_0, the greater the acceleration and the easier the system overcomes the barrier between states, enhancing the conformational ensemble.

(i.e., for $V(\vec{r}) < E$), the energy surface is modified by adding a boost potential as $V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r})$, where $\Delta V(\vec{r}) = \frac{1}{2}k(E - V(\vec{r}))^2$. The harmonic constant k_0 determines the magnitude of the applied boost potential, accelerating the conformational sampling by orders of magnitude and reducing energy barriers. k_0 is an adjustable parameter (ranging from 0 to 1), which can be determined from conventional MD runs, considering the max/min, average and standard deviation of the system's potential energy (V). We refer to the original article for a theoretical description,[90] and to a recent review for the explanation of its application.[91] This method allows enhancing the exploration of the configurational space and "routine access" to the *ms* timescale.[92] The canonical ensemble average system is obtained by reweighting each point in the configura-

tion space on the modified potential by the strength of the Boltzmann factor of the bias energy, $exp[\beta\Delta V(r_{t(i)})]$ at that particular point. Since the boost potential follows a near-Gaussian distribution, this also allows for an improved reweighting procedure (through cumulant expansion to the 2nd order).[93] Enhanced network models are then constructed by applying graph theory on the reweighted conformational landscape, which reports the canonical ensemble.

GaMD simulations have been employed to generate a μs–to–ms timescale conformational ensemble of the CRISPR-Cas9 HNH nuclease, which was used for the application of the Dijkstra's algorithm for shortest path search.[23] As a result, the enhanced network model identified a route of residue–to–residue amino acids that maximizes the correlations from long timescale motions. The residues composing this pathway also displayed slow dynamics through CPMG relaxation dispersion experiments (*vide supra*, Figure 6(b)). Hence, residues that are critical for the long timescale dynamics, as computed through GaMD and the enhanced network model, overlap with residues displaying slow dynamics experimentally. This consistency indicates that the experimental dynamics is captured well by the enhanced network model. On the other hand, classical MD simulations described faster timescales (*ns* range), as arising from the comparison of the computed and measured S^2 order parameters. Hence, the use of enhanced sampling could properly describe the long timescale dynamics responsible for the transmission of the allosteric signaling. This application thereby demonstrates the value of "enhanced network models" in the characterization of allosteric mechanisms in large ribonucleoproteins, and suggests the combination of graph theory with other enhanced sampling methods.[32] This will allow the description of the allosteric network as arising from extended timescales (μ*s* to *ms*), characterizing how slow dynamical motions affect the communication network.

Ab-initio network models and allosteric regulation of the catalysis

The allosteric regulation is critical in enzymes to activate their catalytic function.[30] This modulation can occur over long timescale conformational changes, and through an instantaneous sub-*ns* transfer of motions.[17,24–26,30,31] Allosteric modifications can activate catalysis relying on short timescale motions with instantaneous transfer of information. However, the mechanism of sub-*ns* transfer and the relation between allostery and catalysis has not yet been fully clarified. It is also not clear how the allosteric regulation changes from reactants to (R) to products (P) and how it affects the transition state ($TS^{\ddagger}$). This is mainly because graph theory-based techniques have been widely applied on classical and enhanced MD simulations, but little is known on the network dynamics along chemical reactions. To comprehend the allosteric regulation of catalysis, graph theory met *ab-initio* quantum mechanics/molecular mechanics (QM/MM) simulations. In this approach, local events within the biomolecular system – such as catalytic reactions or short timescale fluctuations – can be sampled though MD at a high QM level, while the rest of the system is treated using the classical force field.[33] *Ab-initio* QM/MM MD can provide the step-by-step dynamics along the chemical reaction, and allow sub-*ns* sampling at critical steps. Building on these dynamics, graph theory can be used to inform how the allosteric pathway of communication changes from the R to P states, passing through the $TS^{\ddagger}$. This approach has been introduced to study the allosteric regulation of the RuvC catalysis in Cas9 (Figure 8(a)).[95,96] Analysis of the generalized correlations (GC) revealed highly coupled motions in the R state, which are progressively reduced in the $TS^{\ddagger}$ and in the P states (Figure 8(a), bottom panel). The R is therefore in a "tense" state, with highly entangled dynamics. The $TS^{\ddagger}$ reduces its tension toward a more "relaxed" state, while the signal fades away in the P state, which demonstrates quenched GC coefficients. Building on this observation, the allosteric regulation of the catalysis follows a tense–*to*–relaxed model, where "tense" refers to highly correlated and "relaxed" to poorly coupled. This suggests that the R is in a tense state, allosterically "prone" due to highly coupled dynamical motions and enthalpically poised for catalysis.

As the reaction proceeds, the dynamics of the $TS^{\ddagger}$ starts losing correlated motions as an effect of the starting of bond breaking. As the system fully relaxes, it reaches the P state. To further understand how the allosteric signaling changes from the R to P, the pathways communicating the DNA recognition region (REC) to the RuvC catalytic core have been computed using the Dijkstra's algorithm. In the R and $TS^{\ddagger}$ states, the information transfers through a pathway of charged and polar residues (Figure 8(b), shown for the R state). This suggests that electrostatic and conformational effects could influence the chemistry from a long-range though allostery, in addition to the electrostatics being critical at the active site level.[97] It is also notable that alanine mutation of central nodes in the REC lobe reduces the catalytic efficiency of Cas9 toward off-target DNAs,[86] reinforcing the notion that distal point mutations could allosterically modulate the catalysis.[98,99] Analysis of the P state revealed that the allosteric pathway is largely disrupted, in line with alteration of the correlation system. Overall, this was the first attempt to characterize the allosteric regulation of catalysis by following the step-by-step dynamics of biochemical reactions. Future studies by our lab will delve into further understanding the role of allostery along the catalysis.

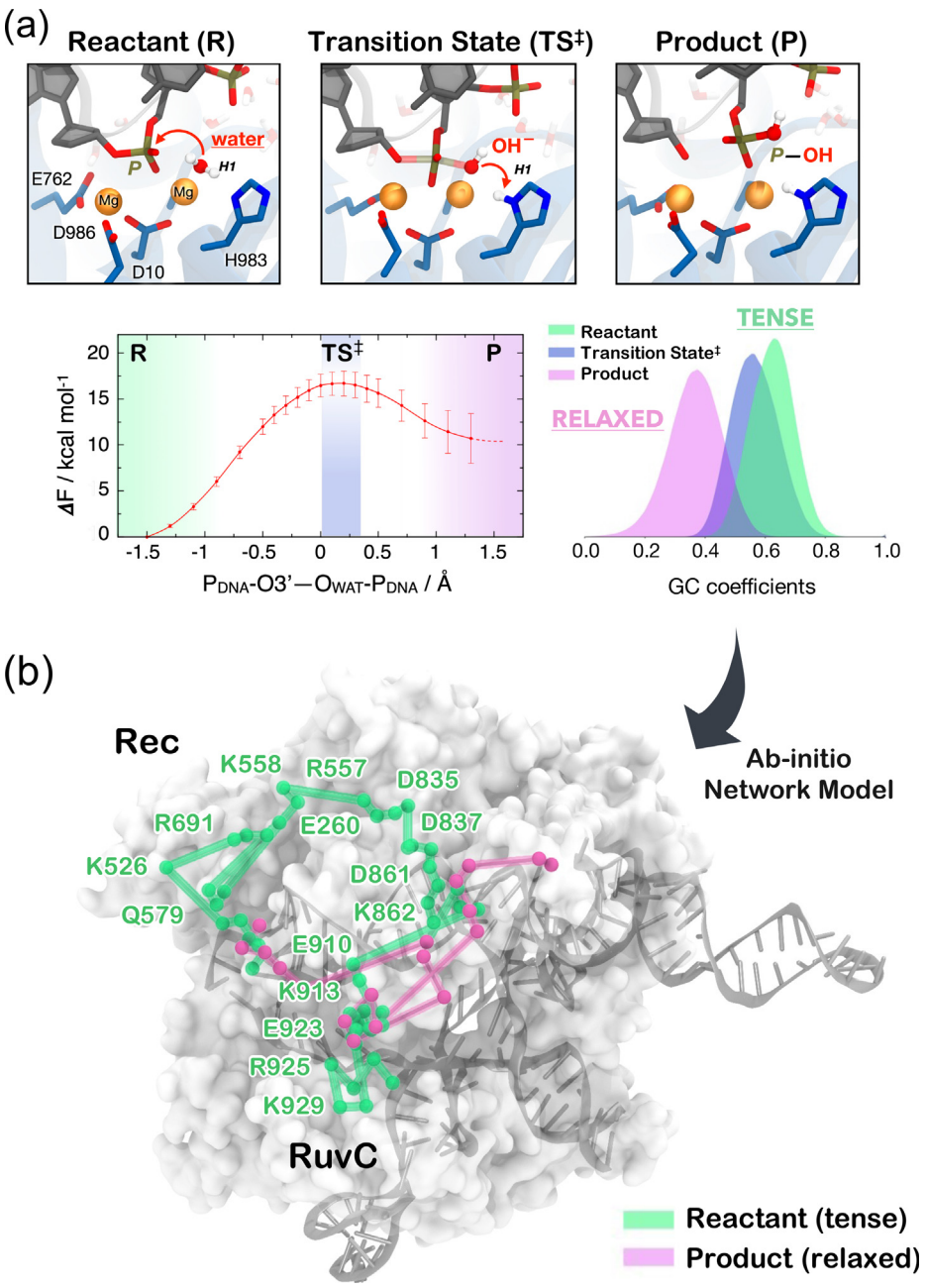

Figure 8. Ab-initio network models. (a) Catalytic mechanism of DNA cleavage in the RuvC active site of CRISPR-Cas9, investigated through QM/MM *ab-initio* simulations.[95,96] The reaction evolves from the Reactants (R) to the Transition State (TS‡) and Product though an associative S_N2 mechanism activated by H983, and with an activation barrier of ~16.5 kcal mol^{-1} (free energy profile at the bottom left). Analysis of the generalized correlations (GC, bottom right) reveals highly coupled motions in the R state, which are progressively reduced in the TS‡ and in the P states. This reveals a tense-to-relaxed model for the allosteric regulation of the chemical step (details in the main text). **(b)** Allosteric pathways connecting the DNA recognition region (REC) to the RuvC catalytic core in the R (green) and P (magenta) states. Adapted with permission from Casalino et al. (2020). Copyright 2020 American Chemical Society.

Perspectives

Here, we review established and innovative approaches to decrypt allostery in proteins and nucleic acids. We discuss the use of classical and enhanced molecular dynamics (MD) simulations in combination with network models and centrality analyses. We report emerging schemes, such as the synergistic use of enhanced simulations methods and network models, which define the

allosteric response over long timescales, and "*ab initio* network models" to describe the allosteric regulation of catalysis. These approaches revealed the allosteric transfer in three paradigmatic examples: (i) the nucleosome core particle, (ii) the CRISPR-Cas9 genome editing system and (iii) the spliceosome. Examples of applications also highlight the current challenges and the prospects of the field, attempting to capture the allosteric response over multiple timescales, relating allostery to conformational changes and catalysis. Indeed, investigating allosteric mechanisms in large biomolecules is often difficult, due to slow dynamical motions virtually inaccessible through classical simulation methods. Along the same lines, little is known about the evolution of the allosteric response along chemical reactions. Taken together, methods and applications showcased here will help overcome these challenges, creating novel protocols to determine the allosteric network of communication over multiple time scales, as well as the relation between allostery and catalysis, which has remained unaddressed through classical approaches.

Acknowledgments

This material is based upon work supported by the National Institute of Health (Grant No. R01GM141329) and by the National Science Foundation (Grant No. CHE-1905374). Computer time for MD simulations has been awarded by XSEDE (Grant No. TG-MCB160059) and by NERSC (Grant No. M3807).

Author contribution

PRA and GP wrote the manuscript. ACP provided critical comments and editing of the manuscript. PRA and GP created the graphics. GP created the illustrations in Figs. 2–7 though handmade painting. PRA performed digital manipulation. GP conceived this research.

Competing Interests

The authors declare no competing financial interest.

Additional Information

Computational codes and script files for the analysis of allosteric mechanism can be downloaded from Github: https://github.com/palermolab.

Received 24 December 2021;
Accepted 23 February 2022;
Available online 28 February 2022

Keywords:
molecular dynamics;
graph theory;
CRISPR-Cas9;
nucleosome core particle;
spliceosome

References

1. Monod, J., Wyman, J., Changeux, J.-P., (1965). On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **12**, 88–118.
2. Liu, J., Nussinov, R., (2016). Allostery: An Overview of Its History, Concepts, Methods, and Applications. *PLOS Comput. Biol.* **12**, e1004966
3. Nussinov, R., Tsai, C.J., (2013). Allostery in disease and in drug discovery. *Cell* **153**, 293–305.
4. Wagner, J.R., Lee, C.T., Durrant, J.D., Malmstrom, R.D., Feher, V.A., Amaro, R.E., (2016). Emerging Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem. Rev.* **116**, 6370–6390.
5. Guo, J., Zhou, H.X., (2016). Protein Allostery and Conformational Dynamics. *Chem. Rev.* **116**, 6503–6515.
6. Dokholyan, N.V., (2016). Controlling Allosteric Networks in Proteins. *Chem. Rev.* **116**, 6463–6487.
7. Nussinov, R., (2016). Introduction to Protein Ensembles and Allostery. *Chem. Rev.* **116**, 6263–6266.
8. Papaleo, E., Saladino, G., Lambrughi, M., Lindorff-Larsen, K., Gervasio, F.L., Nussinov, R., (2016). The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem. Rev.* **116**, 6391–6423.
9. Sethi, A., Eargle, J., Black, A.A., Luthey-Schulten, Z., (2009). Dynamical networks in tRNA: protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6620–6625.
10. Palermo, G., Ricci, C.G., Fernando, A., Basak, R., Jinek, M., Rivalta, I., Batista, V.S., McCammon, J.A., (2017). Protospacer Adjacent Motif-Induced Allostery Activates CRISPR-Cas9. *J. Am. Chem. Soc.* **139**, 16028–16031.
11. Bowerman, S., Wereszczynski, J., (2016). Effects of MacroH2A and H2A.Z on Nucleosome Dynamics as Elucidated by Molecular Dynamics Simulations. *Biophys. J.* **110**, 327–337.
12. Saltalamacchia, A., Casalino, L., Borišek, J., Batista, V.S., Rivalta, I., Magistrato, A., (2020). Decrypting the Information Exchange Pathways across the Spliceosome Machinery. *J. Am. Chem. Soc.* **142**, 8403–8411.
13. Palermo, G., Casalino, L., Magistrato, A., McCammon, J. A., (2019). Understanding the mechanistic basis of non-coding RNA through molecular dynamics simulations. *J. Struct. Biol.* **206**, 267–279.
14. Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J., (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260.
15. Doudna, J.A., Charpentier, E., (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096.
16. Wilkinson, M.E., Charenton, C., Nagai, K., (2020). RNA Splicing by the Spliceosome. *Annu. Rev. Biochem.* **89**, 359–388.

17. Wodak, S.J., Paci, E., Dokholyan, N.V., Berezovsky, I.N., Horovitz, A., Li, J., Hilser, V.J., Bahar, I., et al., (2019). Allostery in Its Many Disguises: From Theory to Applications. *Structure* **27**, 566–578.
18. Bowerman, S., Wereszczynski, J., (2016). Detecting Allosteric Networks Using Molecular Dynamics Simulation. *Methods Enzymol.* **578**, 429–447.
19. Vendruscolo, M., (2011). The statistical theory of allostery. *Nature Chem. Biol.* **7**, 411–412.
20. Kern, D., Zuiderweg, E.R., (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.* **13**, 748–757.
21. Popovych, N., Sun, S., Ebright, R.H., Kalodimos, C.G., (2006). Dynamically driven protein allostery. *Nature Struct. Mol. Biol.* **13**, 831–838.
22. Lisi, G.P., Loria, J.P., (2016). Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem. Rev.* **116**, 6323–6369.
23. East, K.W., Newton, J.C., Morzan, U.N., Narkhede, Y.B., Acharya, A., Skeens, E., Jogl, G., Batista, V.S., et al., (2020). Allosteric Motions of the CRISPR–Cas9 HNH Nuclease Probed by NMR and Molecular Dynamics. *J. Am. Chem. Soc.* **142**, 1348–1358.
24. Buchenberg, S., Sittel, F., Stock, G., (2017). Time-resolved observation of protein allosteric communication. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E6804–E6811.
25. Buchli, B., Waldauer, S.A., Walser, R., Donten, M.L., Pfister, R., Blöchliger, N., Steiner, S., Caflisch, A., et al., (2013). Kinetic response of a photoperturbed allosteric protein. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11725–11730.
26. Hawkins, R.J., McLeish, T.C.B., (2006). Coupling of global and local vibrational modes in dynamic allostery of proteins. *Biophys. J.* **91**, 2055–2062.
27. Hertig, S., Latorraca, N.R., Dror, R.O., (2016). Revealing Atomic-Level Mechanisms of Protein Allostery with Molecular Dynamics Simulations. *PLOS Comput. Biol.* **12**, e1004746
28. Holliday, M.J., Camilloni, C., Armstrong, G.S., Vendruscolo, M., Eisenmesser, E.Z., (2017). Networks of Dynamic Allostery Regulate Enzyme Function. *Structure* **25**, 276–286.
29. Nierzwicki, Ł., Palermo, G., (2021). Molecular Dynamics to Predict Cryo-EM: Capturing Transitions and Short-Lived Conformational States of Biomolecules. *Front. Mol. Biosci.* **8**, 120.
30. Goodey, N.M., Benkovic, S.J., (2008). Allosteric regulation and catalysis emerge via a common route. *Nature Chem. Biol.* **4**, 474–482.
31. Henzler-Wildman, K.A., Lei, M., Thai, V., Kerns, S.J., Karplus, M., Kern, D., (2007). Hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* **450**, 913–916.
32. Bernardi, R.C., Melo, M.C.R., Schulten, K., (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* **5**, 872–877.
33. Brunk, E., Ashari, N.A., Athri, P., Campomanes, P., de Carvalho, F.F., Curchod, B.F.E., Diamantis, P., Doemer, M., et al., (2011). Pushing the Frontiers of First-Principles Based Computer Simulations of Chemical and Biological Systems. *Chimia (Aarau)* **65**, 667–671.
34. Jorgensen, W.L., Maxwell, D.S., Tirado-Rives, J., (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **118**, 11225–11236.
35. Duan, Y., Wu, C., Chowdhury, S., Lee, M.C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., et al., (2003). A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012.
36. Christen, M., Hünenberger, P.H., Bakowies, D., Baron, R., Bürgi, R., Geerke, D.P., Heinz, T.N., Kastenholz, M.A., et al., (2005). The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **26**, 1719–1751.
37. MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., et al., (1998). *J. Phys. Chem. B* **102**, 3586–3616.
38. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K. M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., et al., (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **117**, 5179–5197.
39. Perez, A., Marchan, I., Svozil, D., Sponer, J., Cheatham 3rd, T.E., Laughton, C.A., Orozco, M., (2007). Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **92**, 3817–3829.
40. Zgarbova, M., Otyepka, M., Sponer, J., Mladek, A., Banas, P., Cheatham, T.E., Jurecka, P., (2011). Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **7**, 2886–2902.
41. Banas, P., Hollas, D., Zgarbova, M., Jurecka, P., Orozco, M., Cheatham 3rd, T.E., Sponer, J., Otyepka, M., (2010). Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J. Chem. Theor. Comput.* **6**, 3836–3849.
42. East, K.W., Skeens, E., Cui, J.Y., Belato, H.B., Mitchell, B., Hsu, R.V., Batista, V.S., Palermo, G., et al., (2020). NMR and computational methods for molecular resolution of allosteric pathways in enzyme complexes. *Biophys. Rev.* **12**, 155–174.
43. Yu, L., Li, D.W., Brüschweiler, R., (2020). Balanced Amino-Acid-Specific Molecular Dynamics Force Field for the Realistic Simulation of Both Folded and Disordered Proteins. *J. Chem. Theory Comput.* **16**, 1311–1318.
44. Nierzwicki, Ł., East, K.W., Morzan, U.N., Arantes, P.R., Batista, V.S., Lisi, G.P., Palermo, G., (2021). Enhanced Specificity Mutations Perturb Allosteric Signaling in the CRISPR-Cas9 HNH Endonuclease. *eLife* **10**, e73601.
45. Belato, H.B., D'Ordine, A.M., Nierzwicki, Ł., Arantes, P.R., Jogla, G., Palermo, G., Lisi, G.P., (2022). Structural and dynamic insights into the HNH nuclease of divergent Cas9 species. *J. Struct. Biol.* **214**, 107814
46. Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., Doudna, J.A., (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867–871.
47. Plaschka, C., Lin, P.-C., Nagai, K., (2017). Structure of a pre-catalytic spliceosome. *Nature* **546**, 617–621.
48. Bowerman, S., Hickok, R.J., Wereszczynski, J., (2019). Unique Dynamics in Asymmetric macroH2A–H2A Hybrid Nucleosomes Result in Increased Complex Stability. *J. Phys. Chem. B* **123**, 419–427.
49. Adhireksan, Z., Palermo, G., Riedel, T., Ma, Z., Muhammad, R., Rothlisberger, U., Dyson, P.J., Davey, C. A., (2017). Allosteric cross-talk in chromatin can mediate drug-drug synergy. *Nature Commun.* **8**, 14860.

50. Batchelor, L.K., De Falco, L., Erlach, T., Sharma, D., Adhireksan, Z., Roethlisberger, U., Davey, C.A., Dyson, P. J., (2019). Crosslinking Allosteric Sites on the Nucleosome. *Angew. Chem. Int. Ed.* **131**, 15807–15811.
51. Palermo, G., Magistrato, A., Riedel, T., von Erlach, T., Davey, C.A., Dyson, P.J., Rothlisberger, U., (2016). Fighting Cancer with Transition Metal Complexes: From Naked DNA to Protein and Chromatin Targeting Strategies. *ChemMedChem* **11**, 1199–1210.
52. Ma, Z., Palermo, G., Adhireksan, Z., Murray, B.S., von Erlach, T., Dyson, P.J., Rothlisberger, U., Davey, C.A., (2016). An Organometallic Compound Displays a Unique One-Stranded Intercalation Mode that is DNA Topology-Dependent. *Angew. Chem. Int. Ed.* **128**, 7441–7444.
53. Zuo, Z., Liu, J., (2020). Allosteric regulation of CRISPR-Cas9 for DNA-targeting and cleavage. *Curr. Opin. Struct. Biol.* **62**, 166–174.
54. Nierzwicki, Ł., Arantes, P.R., Saha, A., Palermo, G., (2021). Establishing the Allosteric Mecanism in CRISPR-Cas9. *WIREs Comput. Mol. Sci.*, e1503.
55. Koshland, D.E., Némethy, G., Filmer, D., (1966). Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5**, 365–385.
56. Cuendet, M.A., Weinstein, H., LeVine, M.V., (2016). The Allostery Landscape: Quantifying Thermodynamic Couplings in Biomolecular Systems. *J. Chem. Theory Comput.* **12**, 5758–5767.
57. Cooper, A., Dryden, D.T.F., (1984). Allostery without conformational change. *Eur. Biophys. J.* **11**, 103–109.
58. Kornev, A.P., Taylor, S.S., (2015). Dynamics-Driven Allostery in Protein Kinases. *Trends Biochem. Sci.* **40**, 628–647.
59. Li, X., Wang, C., Peng, T., Chai, Z., Ni, D., Liu, Y., Zhang, J., Chen, T., Lu, S., (2021). Atomic-scale insights into allosteric inhibition and evolutional rescue mechanism of Streptococcus thermophilus Cas9 by the anti-CRISPR protein AcrIIA6. *Comput. Struct. Biotechnol. J.* **19**, 6108–6124.
60. Guarnera, E., Berezovsky, I.N., (2016). Structure-Based Statistical Mechanical Model Accounts for the Causality and Energetics of Allosteric Communication. *PLOS Comput. Biol.* **12**, e1004678
61. Guarnera, E., Berezovsky, I.N., (2019). Toward Comprehensive Allosteric Control over Protein Activity. *Structure* **27**, 866–878.
62. Guarnera, E., Berezovsky, I.N., (2019). On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr. Opin. Struct. Biol.* **56**, 18–27.
63. Lange, O.F., Grubmuller, H., (2006). Generalized correlation for biomolecular dynamics. *Proteins-Struct. Funct. Bioinforma.* **62**, 1053–1061.
64. Pandini, A., Fornili, A., Fraternali, F., Kleinjung, J., (2012). Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *FASEB J.* **26**, 868–881.
65. Saha, A., Arantes, P.R., Hsu, R.V., Narkhede, Y.B., Jinek, M., Palermo, G., (2020). Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *J. Chem. Inf. Model.* **60**, 6427–6437.
66. Casalino, L., Palermo, G., Spinello, A., Rothlisberger, U., Magistrato, A., (2018). All-atom simulations disentangle the functional dynamics underlying gene maturation in the intron lariat spliceosome. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6584–6589.
67. Koutrouli, M., Karatzas, E., Paez-Espino, D., Pavlopoulos, G.A., (2020). A Guide to Conquer the Biological Network Era Using Graph Theory. *Front. Bioeng. Biotechnol.* **8**, 34.
68. Rivalta, I., Sultan, M.M., Lee, N.S., Manley, G.A., Loria, J. P., Batista, V.S., (2012). Allosteric pathways in imidazole glycerol phosphate synthase. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 1428–1436.
69. Melo, M.C.R., Bernardi, R.C., de la Fuente-Nunez, C., Luthey-Schulten, Z., (2020). Generalized correlation-based dynamical network analysis: a new high-performance approach for identifying allosteric communications in molecular dynamics trajectories. *J. Chem. Phys.* **153**, 134104
70. Girvan, M., Newman, M.E.J., (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 7821–7826.
71. Floyd, R.W., (1962). Algorithm 97: Shortest path. *Commun. ACM* **5**, 345.
72. Yan, C., Dodd, T., He, Y., Tainer, J.A., Tsutakawa, S.E., Ivanov, I., (2019). Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat. Struct. Mol. Biol.* **26**, 397–406.
73. Bravo, J.P.K., Liu, M.S., McCool, R.S., Jung, K., Johnson, K.A., W, T.D., (2021). Structural basis for mismatch surveillance by CRISPR/Cas9. *bioRxiv*. https://doi.org/10.1101/2021.11.18.469088.
74. Ghaemi, Z., Guzman, I., Gnutt, D., Luthey-Schulten, Z., Gruebele, M., (2017). Role of Electrostatics in Protein–RNA Binding: The Global vs the Local Energy Landscape. *J. Phys. Chem. B* **121**, 8437–8446.
75. Feher, V.A., Durrant, J.D., Van Wart, A.T., Amaro, R.E., (2014). Computational approaches to mapping allosteric pathways. *Curr. Opin. Struct. Biol.* **25**, 98–103.
76. Moroni, E., Agard, D.A., Colombo, G., (2018). The Structural Asymmetry of Mitochondrial Hsp90 (Trap1) Determines Fine Tuning of Functional Dynamics. *J. Chem. Theory Comput.* **14**, 1033–1044.
77. Sanchez-Martin, C., Moroni, E., Ferraro, M., Laquatra, C., Cannino, G., Masgras, I., Negro, A., Quadrelli, P., et al., (2020). Rational Design of Allosteric and Selective Inhibitors of the Molecular Chaperone TRAP1. *Cell Rep.*, 107531.
78. Verkhivker, G.M., Agajanian, S., Hu, G., Tao, P., (2020). Allosteric Regulation at the Crossroads of New Technologies: Multiscale Modeling, Networks, and Machine Learning. *Front. Mol. Biosci.* **7**, 136.
79. Zhou, H., Dong, Z., Verkhivker, G., Zoltowski, B.D., Tao, P., (2019). Allosteric mechanism of the circadian protein Vivid resolved through Markov state model and machine learning analysis. *PLOS Comput. Biol.* **15**, e1006801
80. Dijkstra, E.W., (1959). A Note on Two Problems in Connection with Graphs. *Numer. Math.* **1**, 269–271.
81. Borgatti, S.P., (2005). Centrality and network flow. *Soc. Netw.* **27**, 55–71.
82. Doshi, U., Holliday, M.J., Eisenmesser, E.Z., Hamelberg, D., (2016). Dynamical network of residue-residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4735–4740.
83. Alvarez-Socorro, A.J., Herrera-Almarza, G.C., González-Díaz, L.A., (2015). Eigencentrality based on dissimilarity

measures reveals central nodes in complex networks. *Sci. Rep.* **5**, 17095.
84. Negre, C.F.A., Morzan, U.N., Hendrickson, H.P., Pal, R., Lisi, G.P., Loria, J.P., Rivalta, I., Batista, V.S., (2018). Eigenvector Centrality Distribution for Characterization of Protein Allosteric Pathways. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12201–12208.
85. Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W. X., Zhang, F., (2016). Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88.
86. Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., et al., (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. *Nature* **550**, 407–410.
87. Lu, S., Ni, D., Wang, C., He, X., Lin, H., Wang, Z., Zhang, J., (2019). Deactivation Pathway of Ras GTPase Underlies Conformational Substates as Targets for Drug Design. *ACS Catal.* **9**, 7188–7196.
88. Ni, D., Wei, J., He, X., Rehman, A.U., Li, X., Qiu, Y., Pu, J., Lu, S., Zhang, J., (2021). Discovery of cryptic allosteric sites using reversed allosteric communication by a combined computational and experimental strategy. *Chem. Sci.* **12**, 464–476.
89. Lu, S., He, X., Yang, Z., Chai, Z., Zhou, S., Wang, J., Rehman, A.U., Ni, D., Pu, J., Sun, J., et al., (2021). Activation pathway of a G protein-coupled receptor uncovers conformational intermediates as targets for allosteric drug design. *Nature Commun.* **12**, 4721.
90. Miao, Y., Feher, V.A., McCammon, J.A., (2015). Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theor. Comput.* **11**, 3584–3595.
91. Wang, J., Arantes, P.R., Bhattarai, A., Hsu, R.V., Pawnikar, S., Huang, Y.M., Palermo, G., Miao, Y., (2021). Gaussian accelerated molecular dynamics: Principles and applications. *WIREs Comput. Mol. Sci.*, e1521.
92. Pierce, L.C.T., Salomon-Ferrer, R., Oliveira, C.A.F., McCammon, J.A., Walker, R.C., (2012). Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theor. Comput.* **8**, 2997–3002.
93. Miao, Y., Sinko, W., Pierce, L., Bucher, D., Walker, R.C., McCammon, J.A., (2014). Improved Reweighting of Accelerated Molecular Dynamics Simulations for Free Energy Calculation. *J. Chem. Theory Comput.* **10**, 2677–2689.
94. Wereszczynski, J., Mccammon, J.A., (2012). Nucleotide-dependent mechanism of Get3 as elucidated from free energy calculations. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7759–7764.
95. Casalino, L., Nierzwicki, Ł., Jinek, M., Palermo, G., (2020). Catalytic Mechanism of Non-Target DNA Cleavage in CRISPR-Cas9 Revealed by Ab Initio Molecular Dynamics. *ACS Catal.* **10**, 13596–13605.
96. Palermo, G., (2019). Structure and Dynamics of the CRISPR–Cas9 Catalytic Complex. *J. Chem. Inf. Model.* **59**, 2394–2406.
97. Warshel, A., Sharma, P.K., Kato, M., Xiang, Y., Liu, H., Olsson, M.H.M., (2006). Electrostatic Basis for Enzyme Catalysis. *Chem. Rev.* **106**, 3210–3235.
98. Liu, M.S., Gong, S., Yu, H.H., Jung, K., Johnson, K.A., Taylor, D.W., (2020). Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA. *Nature Commun.* **11**, 1–13.
99. Gong, S., Yu, H.H., Johnson, K.A., Taylor, D.W., (2018). DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. *Cell Rep.* **22**, 359–371.