



MATERIALS SCIENCE

Systematic modification of functionality in disordered elastic networks through free energy surface tailoring

Dan Mendels^{1*†}, Fabian Bylén¹, Timothy W. Sirk², Juan J. de Pablo^{1*}

A combined machine learning–physics–based approach is explored for molecular and materials engineering. Specifically, collective variables, akin to those used in enhanced sampled simulations, are constructed using a machine learning model trained on data gathered from a single system. Through the constructed collective variables, it becomes possible to identify critical molecular interactions in the considered system, the modulation of which enables a systematic tailoring of the system’s free energy landscape. To explore the efficacy of the proposed approach, we use it to engineer allosteric regulation and uniaxial strain fluctuations in a complex disordered elastic network. Its successful application in these two cases provides insights regarding how functionality is governed in systems characterized by extensive connectivity and points to its potential for design of complex molecular systems.

INTRODUCTION

Progress in the manufacturing and characterization of complex molecular and material systems is often hampered by the complexity of such systems and the enormity of the available design space. Engineering systems at molecular length scales remains a challenging, costly, and time-consuming endeavor. There is a need for new design and optimization methods that can, on the one hand, harness the underlying complexity and, on the other, identify the regions in design space that are most likely to provide fruitful solutions for a given problem.

Several promising strategies for devising such methods revolve around the use of machine learning and particularly its application to large libraries of data associated with sets of diverse systems corresponding to design problems being considered. Given the increasing amounts of data being generated through experiments and simulations, the use of machine learning for molecular and materials design in this way continues to grow (1–11). There are, however, limitations to such approaches’ effectiveness. These include (i) the volume of training data that is required, (ii) the inability to interpret certain outcomes given the black-box nature of many AI-based algorithms, and (iii) the limited applicability of the constructed models beyond the underlying training domain. While overcoming these issues is a matter of ongoing research within the field of machine learning at large, we have recently proposed a machine learning–based method designed to circumvent them. The method, referred to as collective variables for free energy surface tailoring (CV-FEST), does so by (i) using the considerable amount of data generated in simulations or experiments of a single system and by (ii) relying on the powerful ability of certain machine learning algorithms to generate insightful dimensionally reduced representations of complex high-dimensional data.

Specifically, CV-FEST relies on the notion that the functionality of many systems can often be characterized using a dimensionally reduced representation of their free energy surfaces (FES) within a space spanned by a set of collective variables (CVs). Such a depiction serves two interconnected objectives: (i) It provides insight into the mechanisms underlying the functionality of a considered system, and (ii) it allows to condense the most essential information about such a system into a small number of parameters that can be tuned with the aid of optimization algorithms [e.g., (12–17)] for design purposes.

In (18), we focused on analyzing and modifying the functionality of systems that consist of relatively small numbers of degrees of freedom, e.g., a small peptide. Here, we focus on a system of much higher complexity and focus on the question of whether its underlying FES can be manipulated at will. Specifically, we consider an elastic network consisting of roughly 1000 harmonic bonds. Elastic networks have become a subject of increasing interest in recent years given their functional similarities to proteins (19, 20) and their ability to exhibit metamaterial qualities (21, 22). In addition, elastic networks form a natural framework for studying network structure and behavior in the physical realm, potentially providing new and general perspectives on network behavior (23, 24).

For concreteness, we consider a two-dimensional disordered network consisting of identical beads connected by harmonic bonds of identical elastic modulus (see Fig. 1 for illustration). We focus on two properties of the network. The first, allosteric regulation, is analogous to that found in proteins, in which the conformation of a target site (referred to as the active site) in the network is regulated by the conformation of a different, distant site in the network (referred to as the allosteric site). The second functionality examined here is the network’s uniaxial mechanical behavior, as manifested by its uniaxial strain fluctuations. We find that using CV-FEST, we are able to modify these two network characteristics in a simple and tractable manner by systemically tailoring the FES associated with them.

¹Pritzker School of Molecular Engineering, University of Chicago, 5640 S. Ellis Avenue, Chicago, IL 60637 USA. ²Polymers Branch, U.S. CDC Army Research Laboratory, Aberdeen Proving Ground, MD 21005, USA.

*Corresponding author. Email: depablo@uchicago.edu (J.J.d.P.); danmendels@technion.ac.il (D.M.)

†Present address: Department of Chemical Engineering, Technion–Israel Institute of Technology, Haifa, 32000 Israel.

Copyright © 2023 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Downloaded from https://www.science.org at University of Chicago on September 20, 2023

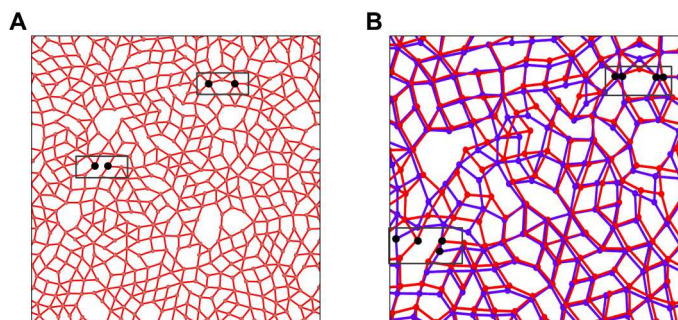


Fig. 1. Illustration of a simulated network from which 12 bonds were removed to embed allosteric response between the indicated sites. (A) The source (top) and target (bottom) beads are represented by the black circles. **(B)** Magnified view of the network in its activated state (in blue) overlaid with the network in its inactivated state (in red). The source beads are pinched toward one another in the activated state leading to the distancing of the target beads from one another.

CVs for free energy tailoring

As described above, CV-FEST relies on the idea that the functionality of many systems in nature can be captured using a dimensionally reduced representation of their FES within a space spanned by a set of CVs. The CVs describe the relevant modes of behavior of a system, similar to the role they play in enhanced sampling methods such as umbrella sampling (25), metadynamics (26), and adaptive biasing force (27), which are used to accelerate the sampling of systems for which the characteristic time scales lie beyond the reach of ordinary simulations. Generally speaking, CVs are functions of the system's atomic coordinates $s(\mathbf{R})$ and can be defined through the following relations

$$P(s) = \int d\mathbf{R} \delta[s - s(\mathbf{R})] P(\mathbf{R}) \quad (1)$$

where $P(s)$ is the system probability to hold a set of CV values s , and $P(\mathbf{R})$ is the Boltzmann probability and is Dirac's delta function. A system's FES with respect to the used set of CVs then follows

$$F(s) = -\frac{1}{\beta} \log P(s) \quad (2)$$

where $\beta = 1/k_B T$ and k_B is Boltzmann's constant and T is the temperature.

While the construction of adequate CVs in the context of enhanced sampling can be challenging, once achieved, CVs encode the physical essence of the processes that determine a system's behavior at long time scales. This renders them potentially useful tools for engineering. While traditionally the construction of CVs was deemed to require a certain degree of expertise regarding the system being studied, new machine learning-based methods for this task have been introduced in recent years (28–34). CV-FEST uses harmonic linear discriminant analysis (HLDA) (29, 35–38), given its ease of use and straightforward interpretation. HLDA constructs CVs as linear weighted sums of descriptors and requires as input a limited amount of information, which can be collected via short simulations in states relevant to the processes being considered. Given their linear form, the interpretation of the HLDA CVs is straightforward; descriptors attaining larger weights in absolute value are deemed to be associated with the forces that encompass higher physicochemical importance with respect to the

relevant behavior or process. The HLDA CV descriptor hierarchy can thus be used to identify the set of forces and interactions in a system that can be tuned for the purposes of tailoring its FES and modifying its functionality in desirable ways.

While HLDA was originally designed to construct CVs that correspond to rare transitions occurring between metastable states (29, 35), its applicability has since been shown to extend also to cases in which the input data are collected in unstable states (39). In what follows, we take advantage of this aspect of the method. The application of HLDA requires that a list of system descriptors d_i be identified as an input, e.g., distances between beads, bond angles, or more complex variables such as the enthalpy or entropy of a system (40, 41). In the current context, however, we limit the type of descriptors to those which directly correspond to tunable force potentials of the system, namely, the system's bond potentials.

Once a descriptor set is assembled, HLDA requires as input the expectation value vectors Σ_I and covariance matrices μ_I with respect to the predefined descriptor space, corresponding to each of the states $I \in M$ associated with the relevant processes. The computation of the elements can be carried out using data collected in short unbiased simulations of each of the relevant states. To construct the CVs, HLDA estimates the directions \mathbf{W} in the N_d dimensional descriptor space on which the projections of the collected training distributions are best separated. This is done through the maximization of the ratio between the training data, the so-called between-class \mathbf{S}_b and within-class \mathbf{S}_w scatter matrices, and can be written as

$$\mathcal{J}(\mathbf{W}) = \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}} \quad (3)$$

with

$$\mathbf{S}_b = (\boldsymbol{\mu}_I - \bar{\boldsymbol{\mu}}) (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu}_I)^T \quad (4)$$

where $\boldsymbol{\mu}_I$ are the expectation value vectors of the I th metastable state, and $\bar{\boldsymbol{\mu}}$ is the overall mean of the distributions, i.e., $\bar{\boldsymbol{\mu}} = 1/M \sum_{I=1}^M \boldsymbol{\mu}_I$, and with

$$\mathbf{S}_w = \frac{1}{\frac{1}{\sum_1} + \frac{1}{\sum_2} + \dots + \frac{1}{\sum_M}} \quad (5)$$

Given the normalization $\mathbf{W}^T \mathbf{S}_w \mathbf{W} = 1$, Eq. 3 can be shown to be equivalent to solving the eigenvalue equation (35)

$$\mathbf{S}_w^{-1} \mathbf{S}_b \mathbf{W} = \lambda \mathbf{W} \quad (6)$$

The eigenvectors of Eq. 6 associated with the largest $M-1$ eigenvalues define the directions in the N_d space along which the distributions obtained from the M sampled states overlap the least and thus constitute the CVs that correspond to the transitions of interest. Using these CVs, we can now systematically tailor the system FES and modify its functionality in a purposeful manner. To do that, the leading descriptors of each of the constructed CVs are identified and their corresponding interaction potentials are changed. The modification of the force potentials is guided by inspection of their functional form relative to the system's overall FES (e.g., augmenting force potentials which have minima within a metastable state which is sought to be further stabilized) (18).

RESULTS

Allosteric response

The first network functionality we focus on is that of allosteric regulation. The term allosteric regulation originates from the study of proteins, referring to the alteration of the activity of a site in the protein (e.g., its ability to have a ligand bind to it) through the binding of an effector molecule to another distal site in it, i.e., the allosteric site. Allosteric regulation plays an important role in many biological processes, such as transcriptional regulation and metabolism, and is rooted in the fundamental physical properties of macromolecular systems. Its underlying mechanisms are still poorly understood and hence it constitutes a central theme of present research in the field of biology (42, 43).

In the considered network, we simulate the binding of an effector molecule to an allosteric site as the pinching of two neighboring nonbonded beads toward each other. We refer to these two beads as the source beads (represented by blue circles in Fig. 1A) of the considered allosteric process. Correspondingly, a pair of target beads is defined at a different location in the network, representing the active site (also represented by blue circles in Fig. 1A). The allosteric activation of the target beads, defined as their distancing from one another in response to the pinching of the source beads, is embedded into the network using the “tuning by pruning” algorithm introduced in (20) (see Methods for more details). Thus, when the source beads are at their relaxed positions, the distance between the target beads assumes its initial value, and the active site is in its inactivated state. However, when the source beads are pinched beyond a threshold, the target beads respond by distancing from one another, and the active site is deemed to be activated. We consider an additional third state of the system, corresponding to the phenomenon of negative cooperativity (44). A scenario is considered in which the binding of a ligand to a site neighboring the allosteric site inhibits the binding of the effector molecule, for example, by steric repulsion. We simulate this scenario as a state of the network in which the source beads are stretched away from one another, thus inhibiting their ability to arrive to the pinched active state. The system in practice can thus be in one of three states: the inactivated, activated, or inhibited state.

To explore our ability to systematically tailor the system’s FES associated with its allosteric behavior, we start by defining a descriptor set for the problem at hand. In this case, we opt for a descriptor set composed of all the network bond distances. Next, we run short simulations in the three different states of the system at finite temperature, constraining the source beads to their respective positions at each state. Before calculating the HLDA CVs, to prevent the skewing of the obtained results (and following a common practice in the context of many machine learning algorithms), we compute the correlation matrices corresponding to each of the collected distributions and omit descriptors that exhibit correlations greater than 0.9 (in absolute value) to others in the set. Last, using the data collected in the simulations, we calculate the expectation values μ_I and covariance matrices Σ_I associated with each of the states and, using Eq. 6, compute the HLDA CVs corresponding to the three-state system. The weight distribution corresponding to the first eigenvector is presented in the inset of Fig. 2C.

While the HLDA eigenvectors corresponding to the top two eigenvalues provide good separation between the predefined states, to augment our ability to tailor the system’s FES, before selecting the

top weighted descriptors, we apply a two-dimensional rotation to the plane spanned by the two HLDA CVs. This is implemented in such a way that states A and B are best separated with respect to the direction corresponding to the rotated HLDA₁, while states B and C are best separated with respect to the direction corresponding to the rotated HLDA₂ (see fig. S1 in the Supplementary Materials for illustration). In analogy to the concept of normal modes, applying the described rotation allows us to substantially decouple the effects induced by the modification of bonds associated with HLDA₁ on the FES associated with those induced by the modification of the bonds involved in HLDA₂. In essence, modifying bonds associated with HLDA₁ predominantly affect the free energy difference between states A and B, whereas modifying bonds associated with HLDA₂ predominantly affects the free energy difference between states A and C, as can be seen in Fig. 2A. Given the linearity of HLDA, the resulting CVs tend to be dominated by the descriptors associated with the largest weights in the CV, rendering the weight distribution of the lower weighted descriptors potentially less accurate. To circumvent this issue, we apply HLDA iteratively, whereby after each iteration (and rotation), the top five weighted descriptors (in absolute value) of each of the two constructed CVs are selected and removed from the descriptor list for subsequent iterations. The top bonds selected in this way are highlighted in Fig. 3A. As can be seen, both bond sets corresponding to the two HLDA CVs form distinct patterns, reflecting their functional significance.

To test our ability to tailor the FES corresponding to the allosteric and cooperative behaviors of the system, we alter the bond coefficient of the selected bonds and then compute the FES. Given the large differences in free energy between the considered states, the computation of the system’s FES requires the use of an enhanced sampling approach. Here, we use well-tempered metadynamics (45), using the distance between the source beads d_s as the biasing CV for simulations. (See Methods for details and fig. S3 for an example of the time-dependent behavior of the source and target nodes in such simulations). Upon convergence of the well-tempered metadynamics runs, we compute the FES of the system using Eq. 9. Figure 2A presents the FES computed in this way for several realizations of the system. One can appreciate that altering the bond coefficient of the selected bonds gives rise to substantial changes in the system’s FES. In contrast, similar alterations of randomly selected bonds do not lead to noticeable changes of the FES. It can also be seen that altering the bonds associated with HLDA₁ mainly affects the FES branch corresponding to the $A \leftrightarrow B$ transition, while altering those associated with HLDA₂ predominantly affects the FES branch corresponding to the $B \leftrightarrow C$ transition.

An examination of the effects of altering the bond coefficient of the top 10 bonds corresponding to HLDA₁ reveals that one can systematically modify the extent to which the source nodes need to be pinched to initiate the “activation” of the target beads. Figure 2C illustrates this feature by showing the dependence of the distance d_s for which the allosteric response is initiated as a function of the bond coefficient of the top 10 HLDA₁ bonds. Our results show that as the selected bonds are weakened, the source beads need to be brought closer to one another for allosteric activation to occur and vice versa. One can envision inducing similar effects in real proteins; by systemically softening the environment of an allosteric site, one could alter the types of molecules (e.g., different dimensions or interactions with the allosteric site) that would lead to its activation. We find that we obtain such an effect, albeit to a slightly lesser

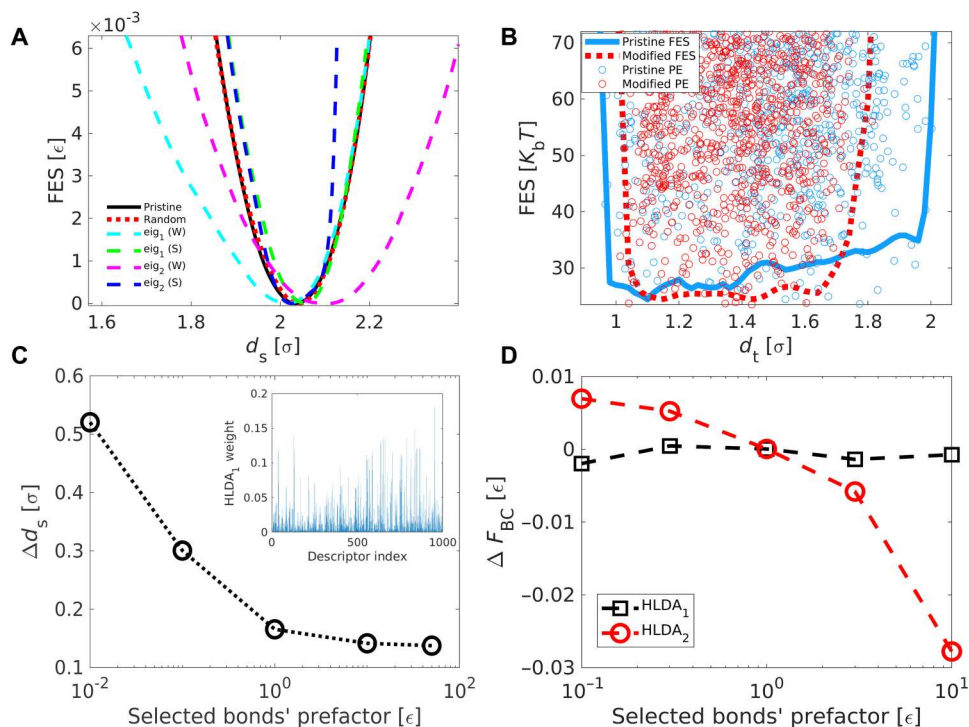


Fig. 2. Systematic modifications of the system's allosteric response. (A) FES with different realizations of the allosteric network as a function of the distance between the source nodes. FES of the pristine network (solid line) and FES obtained after 30 random bonds are strengthened to 50ϵ (dotted line). Also shown: FES obtained of networks in which the 10 highest weighted bonds of either HLDA₁ or HLDA₂ are either strengthened (S) to 50ϵ or weakened (W) to 0.1ϵ (dashed lines). (B) Potential energy distributions and reweighted FES as a function of d_t calculated using the methodology of (60). The pristine network can reach the fully activated state at $d_t > 17\sigma$, while this state is not accessible to the network in which the bonds ranked 12 to 27 in the HLDA₁ weight hierarchy are weakened. (C) Minimal pinching length for which activation of the target beads is initiated as a function of the bond coefficient of the 10 highest weighted HLDA₁ bonds. Inset: Absolute value of the weight distribution of HLDA₁, generated in the first iteration of the calculation. (D) ΔF_{BC} (calculated using equation S10), the free energy difference between state C (the stretched state), $d_s = 2.19\sigma$, and the pinched state (defined as the point at which activation of the target site is initiated) as a function of the bond coefficient of the targeted bonds.

extent, also by modifying only the top ranked HLDA₁ bond as shown in fig. S5.

While altering the top bonds associated with HLDA₁ has a significant effect on the distance d_s for which the allosteric response is initiated, we find that it induces a more modest effect on the free energy difference between states A and B and consequently also on that between states B and C. In contrast, however, modifying the bonds associated with HLDA₂ leads to large changes in the free energy difference between states B and C, thereby allowing us to easily alter this free energy difference between the activated and inhibited states, as shown in Fig. 2D.

By repeating the bond selection procedure for several iterations, we find that we can reveal segments that compose the primary channel of mechanical communication between the allosteric and active sites. The corresponding bonds, placed in positions 12 to 27 in the hierarchy of HLDA₁, are highlighted in Fig. 3B. Specifically, we find that weakening this set of bonds limits the communication between the sites, precluding the activation of the active site, as illustrated in fig. S4. Figure 2B illustrates this point further by presenting the FES corresponding to the distance between the target beads, plotted for the pristine system and the modified system. It can be seen that the fully activated region corresponding to $d_t > 1.7\sigma$ is inaccessible in the case of the modified system. Considering again the analogy to proteins, it would be intriguing to explore if the proposed methodology would be able to help shed light on the

prominent question of how communication occurs between allosteric and active sites.

Uniaxial strain fluctuations

To complement the analysis of the network's allosteric behavior, we apply CV-FEST also to a comparably more global attribute of the network, namely, its uniaxial strain fluctuations (46). The network is constructed in the same manner as before. As in the previous case, to systematically modify the behavior of the network, we start by collecting training data. We do this by running a slow uniaxial compression simulation at constant temperature and constant lateral dimension. The network is deformed in the x direction to a strain of $\delta L_x/L_x = 0.0075$. To construct the HLDA CV that corresponds to the network's strain fluctuations in the x direction, we use data collected in two short segments of the compression simulation. The first, taken from the beginning of the simulation when the network is nearly relaxed, and the second from the end of the simulation when the network is nearly fully deformed.

Defining the 1007 distances corresponding to all of the network's bonds as our descriptor set and calculating the relaxed and compressed states' expectation vectors and covariance matrices, we apply Eq. 6 to obtain the HLDA CV. As previously done, to circumvent the limitation imposed by the linearity of HLDA, we apply it iteratively, whereby the descriptors attaining the three largest weights in absolute value at each iteration are selected and

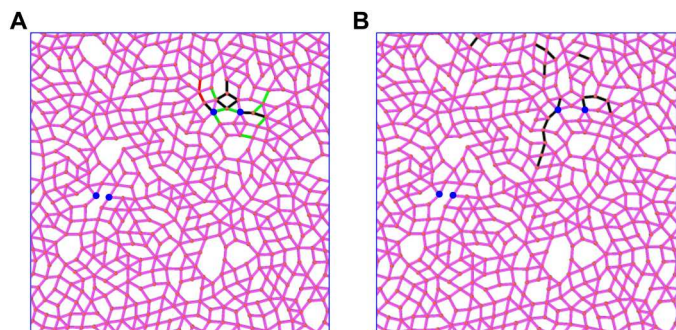


Fig. 3. Depiction of the highest weighted HLDA bonds. (A) Depiction of the 10 highest HLDA CVs weighted bonds in absolute value. The top bonds corresponding to HLDA₁ are shown in black, and the top bonds corresponding to HLDA₂ are shown in green. The top bonds corresponding to both HLDA₁ and HLDA₂ are red. (B) Depiction of the bonds ranked 12 to 27 in the weight hierarchy of HLDA₁. Decreasing the bond coefficient of these bonds inhibits the activation of the active site.

removed from the descriptor set used in the iterations to follow. Figures 4 (A and B) highlight, respectively, the top ~1% and top ~6% of bonds selected in this manner. The selected bonds are distributed fairly homogeneously across the network, in contrast to the allosteric case. Such bonds appear to be organized in small clusters, consisting of two to six bonds each.

To modify the FES corresponding to the network's uniaxial strain fluctuations, we systematically alter the bond coefficient of the selected bonds. To quantify the resulting behavior, we simulate the network under constant temperature and constant pressure, applying the Parrinello-Rahman barostat (47, 48) in the x direction of the simulation cell, keeping the simulation cell edge in the perpendicular direction, L_y , constant. From these simulations, we compute the probability density $P(L_x)$ and, using Eq. 1, $F(L_x)$, the corresponding FES of the network (49).

Figure 4B exhibits the FES of the pristine and modified networks. Altering the bonds selected by CV-FEST gives rise to substantially greater changes in the system's FES compared to the case in which randomly selected bonds are altered. This is

particularly apparent when less than 1% (i.e., nine bonds) of the network's bonds are altered. In that case, CV-FEST is able to select strategically important bonds of the network, whereas a random selection yields no apparent change of the system's FES. This stark difference illustrates the importance of the critical bonds' positions in the network, given that all the bonds in the pristine network have the same elastic modulus. Examining the FES of the different networks, we find that altering the bond coefficient of the selected bonds leads to three different effects. The first is a modification of the steepness of the FES. The second is a change of the functional form of the FES, namely, the extent to which it deviates from a parabola (as would be expected for entropic contributions to the system's elasticity). The third is a shift of the minimum. These three differences are further illustrated in figs. S6 and S7 in which the FES of the different networks are aligned and in fig. S8, which illustrates how the stress-strain curves and Young's modulus of the system are changed given the underlying modifications. The ability to modify the FES of the network in these ways points to the promise of using CV-FEST to engineer systems that exhibit targeted, desirable mechanical properties.

DISCUSSION

We have studied the ability to systematically tailor the FES of a disordered elastic network with respect to allosteric regulation and uniaxial strain fluctuations at finite temperature and pressure conditions. We find that CV-FEST is capable of (i) identifying the important bonds in the network with respect to each of these functionalities and (ii) by altering these bonds' stiffness, of tailoring in a tractable way the system's FES and the corresponding functional behavior.

Given the complex interconnected nature of the networks, CV-FEST's demonstrated capabilities offer potential as a tool for design and analysis of complex systems in general, including systems such as proteins, macromolecules, and materials. In this context, it is also worthwhile mentioning the interesting works of (50, 51) which also use contrastive machine learning methodologies for network design. While these works focus on continuous, simultaneous and

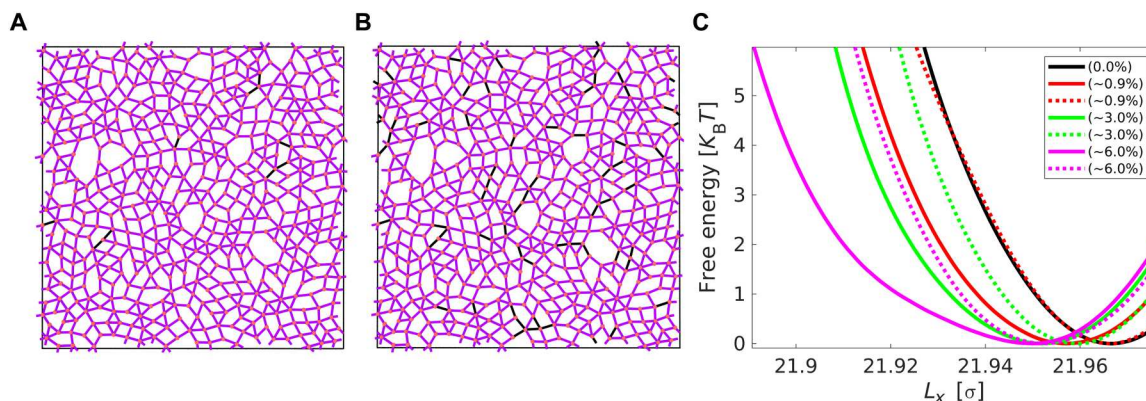


Fig. 4. Systematic modifications of the system's uniaxial strain fluctuations. (A) Illustration of the simulated network with the top 9 (~1%) HLDA bonds highlighted in dark blue and (B) with the top 60 (~6%) HLDA bonds highlighted in dark blue. (C) The computed FES of the simulated pristine network as function of L_x (solid black) along with the FES of modified realizations of the network in which the bond coefficient of the highest weighted HLDA bonds was reduced to $0.01\epsilon/\sigma^2$ (dotted lines) or in which the bond coefficient of randomly selected bonds was reduced to $0.01\epsilon/\sigma^2$ (solid lines). In parenthesis, the percentage of bonds that was modified in the network in each case.

incremental modifications of all the forces in a considered network and hence are less suitable for more realistic systems such as proteins and materials, investigating the potential for combining their proposed methods with the approach introduced here could prove to be an interesting direction for future research. Last, considering that CV-FEST relies solely on kinematic information for its input, it would be interesting to explore its direct applicability to macroscopic mechanical experimental systems (21, 52) for which such information is relatively easily obtained. In the case of microscopic systems, we envision the use of CV-FEST to come to fruition in the context of combined experimental-simulation studies of macromolecular systems such as biomolecules and complex materials such as polymer networks.

METHODS

Network construction

Construction of the networks followed the protocol put forward in (20). Briefly, two-dimensional configurations of soft disks placed in a simulation cell with periodic boundary conditions and allowed to relax to a local energy minimum using a standard jamming algorithm (53). The network is then constructed by placing nodes at the center of each disk and by linking nodes corresponding to disks which overlap in the resulting configuration. To implement the elastic networks, beads of identical mass are placed at every node position in the network, and links are replaced by harmonic bonds with an elastic energy of the form

$$V(\mathbf{r}_{ij}) = \frac{K_{ij}}{(\mathbf{r}_{ij}^0)^2} (\mathbf{r}_{ij} - \mathbf{r}_{ij}^0)^2 \quad (7)$$

where K_{ij} is the bond coefficient corresponding to the bond between the i and j beads and is set initially to 1 for all bonds in the network; \mathbf{r}_{ij}^0 is the rest length of the bond, and \mathbf{r}_{ij} is the distance between the beads.

Embedding allosteric response into the network

Allosteric response is embedded into the network by randomly selecting two pairs of neighboring nonbonded beads, referred to as the source beads and target beads (see Fig. 1). The target beads are chosen to be spatially distant from the source beads to achieve the long-range effect that characterizes allostery.

The allosteric effect is defined as an imposed change of the distance between the target beads given a change in the distance between the source beads. To optimize the network such that this effect will emerge, a tuning-by-pruning of bonds strategy is used with the objective to minimize the fitness function of Eq. 8, which measures the difference between the desired target beads' response and the actual response (20, 54). Namely, the ratio of the target strain to the source strain, $\eta = \Delta d_t / \Delta d_s$, is measured and compared to the desired ratio η^* , set to 5, rendering the fitness function to be

$$\Delta^2 = (\eta / \eta^* - 1)^2 \quad (8)$$

To compute Eq. 8, the source beads are "pinched" to 50% of their initial distance and frozen at their new positions, after which a second minimization of network energy is carried out and the ratio η is calculated. The optimization procedure was applied iteratively, whereby at each iteration, Δ^2 resulting from a trial removal of

each bond in the network was computed. A greedy algorithm was followed in which the bond the removal of which lead to the largest decrease in Δ^2 with respect to its previous value was permanently deleted. To keep local stability, however, the bond is deleted only if all the beads it was connected to were connected to at least three remaining bonds (55). Otherwise, the bond that created the next-largest decrease in Δ^2 is permanently deleted, given that it satisfied this constraint and so on. This iterative process is continued until the desired strain ratio (Eq. 8) is attained. The energy minimization of the network was performed using LAMMPS (56).

Dynamic simulations of the allosteric network

All simulations were run with LAMMPS (56) patched with PLUMED2.6 (57). The network consisted of 499 beads and 999 bonds. For simulations run at $T = 8.6 \cdot 10^{-6}$, all beads had a mass of $M = m$ with the exception of the source beads which had a mass of $M_s = 1000 m$. For the sake of preventing the destabilization of the network, for well-tempered metadynamics, simulations run at $T = 4.3 \cdot 10^{-5}$ all beads had a mass of $M = 100 m$ with the exception of the source beads which had a mass of $M_s = 1000 m$. All simulations were initially energetically relaxed at zero pressure and subsequently run at a constant temperature using a Langevin thermostat (58) with a damping parameter of 1 and a time step of 0001. In the unbiased training simulations, the source bead positions were constrained in the allosteric activated state B, $d_s = 1.87\sigma$, and the trap state C, $d_s = 2.1\sigma$, to keep the network from relaxing back to the inactivated state A. The FES of the system was computed using Eq. 9

$$F(s) = \frac{\gamma}{1 - \gamma} V(s) \quad (9)$$

where $V(s)$ is the bias potential deposited in the well-tempered metadynamics simulations, and γ is the so-called bias factor. Well-tempered metadynamics simulations were run with a bias factor of $\gamma = 180$, a hill height of 0.0001ϵ , and a hill width of 0.01σ .

Probing uniaxial fluctuations

Simulations were run with LAMMPS (56) patched with PLUMED2.6 (57). The network consisted of 499 beads and 1007 bonds. All beads had a mass m . Training simulations, in which the networks were slightly compressed in the x direction, were run at constant temperature $T = 8.6 \cdot 10^{-6}$ (with $k_B = 1$) using a Langevin thermostat (58) with a damping parameter of 1 and a time step of 0.01τ . Compression was executed using a constant deformation rate of $10^{-7}\tau^{-1}$. The FESs of the networks were measured in constant pressure simulations run at zero pressure, using a Parrinello-Rahman barostat (48, 59) applied in the x direction, corresponding to the direction of compression in the training simulations. The simulation box side in the y direction was kept constant in all simulations.

Supplementary Materials

This PDF file includes:

Figs. S1 to S9

REFERENCES AND NOTES

1. A. Agrawal, A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science. *APL Mater.* **4**, 053208 (2016).

2. B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).
3. Z. Yang, X. Li, L. Catherine Brinson, A. N. Choudhary, W. Chen, A. Agrawal, Microstructural materials design via deep adversarial learning methodology. *J. Mech. Des.* **140**, 111416 (2018).
4. M. A. Webb, N. E. Jackson, P. S. Gil, J. J. de Pablo, Targeted sequence design within the coarse-grained polymer genome. *Sci. Adv.* **6**, eabc6216 (2020).
5. Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, W. Cai, Generative model for the inverse design of metasurfaces. *Nano Lett.* **18**, 6570–6576 (2018).
6. Y. Mao, Q. He, X. Zhao, Designing complex architected materials with generative adversarial networks. *Sci. Adv.* **6**, eaaz4169 (2020).
7. B. Kim, S. Lee, J. Kim, Inverse design of porous materials using artificial neural networks. *Sci. Adv.* **6**, eaax9324 (2020).
8. K. Swanson, S. Trivedi, J. Lequieu, K. Swanson, R. Kondor, Deep learning for automated classification and characterization of amorphous materials. *SoftMatter* **16**, 435–446 (2020).
9. J. M. Stokes, K. Yang, K. Swanson, W. Jin, A. Cubillos-Ruiz, N. M. Donghia, C. R. MacNair, S. French, L. A. Carrae, Z. Bloom-Ackermann, V. M. Tran, A. Chiappino-Pepe, A. H. Badran, I. W. Andrews, E. J. Chory, G. M. Church, E. D. Brown, T. S. Jaakkola, R. Barzilay, J. J. Collins, A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702.e13 (2020).
10. C. Shen, M. Krenn, S. Eppel, A. Aspuru-Guzik, Deep molecular dreaming: Inverse machine learning for *de-novo* molecular design and interpretability with surjective representations. *Mach. Learn. Sci. Technol.* **2**, 03LT02 (2021).
11. R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Sci.* **4**, 268–276 (2018).
12. A. W. Long, A. L. Ferguson, Rational design of patchy colloids via landscape engineering. *Mol. Syst. Des. Eng.* **3**, 49–65 (2018).
13. A. D. White, G. A. Voth, Efficient and minimal method to bias molecular simulations with experimental data. *J. Chem. Theory Comput.* **10**, 3023–3030 (2014).
14. D. B. Amirkulova, A. D. White, Recent advances in maximum entropy biasing techniques for molecular dynamics. *Mol. Simul.* **45**, 1285–1294 (2019).
15. A. Gil-Ley, S. Bottaro, G. Bussi, Empirical corrections to the amber RNA force field with target metadynamics. *J. Chem. Theory Comput.* **12**, 2790–2798 (2016).
16. F. Marinelli, J. D. Faraldo-Gómez, Ensemble-biased metadynamics: A molecular simulation method to sample experimental distributions. *Biophys. J.* **108**, 2779–2782 (2015).
17. A. Cesari, S. Reißer, G. Bussi, Using the maximum entropy principle to combine simulations and solution experiments. *Comput. Secur.* **6**, 15 (2018).
18. D. Mendels, J. J. de Pablo, Collective variables for free energy surface tailoring: Understanding and modifying functionality in systems dominated by rare events. *J. Phys. Chem. Lett.* **13**, 2830–2837 (2022).
19. I. Bahar, T. R. Lezon, L.-W. Yang, E. Eyal, Global dynamics of proteins: Bridging between structure and function. *Annu. Rev. Biophys.* **39**, 23–42 (2010).
20. J. W. Rocks, N. Pashine, I. Bischofberger, C. P. Goodrich, A. J. Liu, S. R. Nagel, Designing allostery-inspired response in mechanical networks. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2520–2525 (2017).
21. D. R. Reid, N. Pashine, J. M. Wozniak, H. M. Jaeger, A. J. Liu, S. R. Nagel, J. J. de Pablo, Auxetic metamaterials from disordered networks. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1384–E1390 (2018).
22. D. Hexner, A. J. Liu, S. R. Nagel, Periodic training of creeping solids. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 31690–31695 (2020).
23. R. Albert, A.-L. Barabási, Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
24. Y.-Y. Liu, A.-L. Barabási, Control principles of complex systems. *Rev. Mod. Phys.* **88**, 035006 (2016).
25. G. M. Torrie, J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **23**, 187–199 (1977).
26. A. Laio, M. Parrinello, Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562–12566 (2002).
27. E. Darve, A. Pohorille, Calculating free energies using average force. *J. Chem. Phys.* **115**, 9169–9183 (2001).
28. J. M. L. Ribeiro, P. Bravo, Y. Wang, P. Tiwary, Reweighted autoencoded variational bayes for enhanced sampling (rave). *J. Chem. Phys.* **149**, 072301 (2018).
29. D. Mendels, G. Piccini, M. Parrinello, Collective variables from local fluctuations. *J. Phys. Chem. Lett.* **9**, 2776–2781 (2018).
30. L. Bonati, V. Rizzi, M. Parrinello, Data-driven collective variables for enhanced sampling. *J. Phys. Chem. Lett.* **11**, 2998–3004 (2020).
31. W. Chen, A. R. Tan, A. L. Ferguson, Collective variable discovery and enhanced sampling using autoencoders: Innovations in network architecture and error function design. *J. Chem. Phys.* **149**, 072312 (2018).
32. C. Wehmeyer, F. Noé, Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics. *J. Chem. Phys.* **148**, 241703 (2018).
33. M. M. Sultan, V. S. Pande, Automated design of collective variables using supervised machine learning. *J. Chem. Phys.* **149**, 094106 (2018).
34. J. McCarty, M. Parrinello, A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* **147**, 204109 (2017).
35. G. Piccini, D. Mendels, M. Parrinello, Metadynamics with discriminants: A tool for understanding chemistry. *J. Chem. Theory Comput.* **14**, 5040–5044 (2018).
36. D. Mendels, G. Piccini, Z. F. Brotzakis, Y. I. Yang, M. Parrinello, Folding a small protein using harmonic linear discriminant analysis. *J. Chem. Phys.* **149**, 194113 (2018).
37. V. Rizzi, D. Mendels, E. Sicilia, M. Parrinello, Blind search for complex chemical pathways using harmonic linear discriminant analysis. *J. Chem. Theory Comput.* **15**, 4507–4515 (2019).
38. Y.-Y. Zhang, H. Niu, G. Piccini, D. Mendels, M. Parrinello, Improving collective variables: The case of crystallization. *J. Chem. Phys.* **150**, 094509 (2019).
39. Z. F. Brotzakis, D. Mendels, M. Parrinello, Augmented harmonic linear discriminant analysis. arXiv 1902.08854 (2019).
40. P. M. Piaggi, O. Valsson, M. Parrinello, Enhancing entropy and enthalpy fluctuations to drive crystallization in atomistic simulations. *Phys. Rev. Lett.* **119**, 015701 (2017).
41. D. Mendels, J. McCarty, P. M. Piaggi, M. Parrinello, Searching for entropically stabilized phases: The case of silver iodide. *J. Phys. Chem. C* **122**, 1786–1790 (2018).
42. S. J. Wodak, E. Paci, N. V. Dokholyan, I. N. Berezovsky, A. Horovitz, J. Li, V. J. Hilser, I. Bahar, J. Karanicolas, G. Stock, P. Hamm, R. H. Stote, J. Eberhardt, Y. Chebaro, A. Dejaegere, M. Cecchini, J. P. Changeux, P. G. Bolhuis, J. Vreebe, P. Faccioli, S. Orioli, R. Ravasio, L. Yan, C. Brito, M. Wyart, P. Gkeka, I. Rivalta, G. Palermo, J. A. McCammon, J. Panecka-Hofman, R. C. Wade, A. di Pizio, M. Y. Niv, R. Nussinov, C. J. Tsai, H. Jang, D. Padhorny, D. Kozakov, T. McLeish, Allostery in its many disguises: From theory to applications. *Structure* **27**, 566–578 (2019).
43. A. J. Faure, J. Domingo, J. M. Schmiedel, C. Hidalgo-Carcedo, G. Diss, B. Lehner, Mapping the energetic and allosteric landscapes of protein binding domains. *Nature* **604**, 175–183 (2022).
44. C. A. Hunter, H. L. Anderson, What is cooperativity? *Angew. Chem. Int. Ed.* **48**, 7488–7499 (2009).
45. A. Barducci, G. Bussi, M. Parrinello, Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100**, 020603 (2008).
46. M. Parrinello, A. Rahman, Strain fluctuations and elastic constants. *J. Chem. Phys.* **76**, 2662–2666 (1982).
47. M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).
48. G. J. Martyna, D. J. Tobias, M. L. Klein, Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
49. R. Martoňák, A. Laio, M. Parrinello, Predicting crystal structures: The parrinello-rahman method revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).
50. B. Scellier, Y. Bengio, Equilibrium propagation: Bridging the gap between energy-based models and backpropagation. *Front. Comput. Neurosci.* **11**, 24 (2017).
51. M. Stern, D. Hexner, J. W. Rocks, A. J. Liu, Supervised learning in physical networks: From machine learning to learning machines. *Phys. Rev. X* **11**, 021045 (2021).
52. N. Pashine, D. Hexner, A. J. Liu, S. R. Nagel, Directed aging, memory, and nature's greed. *Sci. Adv.* **5**, eaax4215 (2019).
53. A. J. Liu, S. R. Nagel, The jamming transition and the marginally jammed solid. *Annu. Rev. Condens. Matter Phys.* **1**, 347–369 (2010).
54. J. W. Rocks, H. Ronellenfitsch, A. J. Liu, S. R. Nagel, E. Katifori, Limits of multifunctionality in tunable networks. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2506–2511 (2019).
55. C. P. Goodrich, A. J. Liu, S. R. Nagel, The principle of independent bond-level response: Tuning by pruning to exploit disorder for global behavior. *Phys. Rev. Lett.* **114**, 225501 (2015).
56. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
57. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
58. T. Schneider, E. Stoll, Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **17**, 1302–1322 (1978).
59. M. Parrinello, A. Rahman, Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* **45**, 1196–1199 (1980).

60. P. Tiwary, M. Parrinello, A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **119**, 736–742 (2015).

Acknowledgments

Funding: This work is supported by the Department of Energy, Basic Energy Sciences, through the Midwest Center for Computational Materials (MiCCoM). Additional support for a collaboration on materials design for impact mitigation between the Army Research Laboratory and the University of Chicago was provided by the Center for Hierarchical Materials Design (CHiMaD), supported by NIST. **Author contributions:** D.M., J.J.d.P., and T.W.S. designed the

research. D.M. and F.B. performed the research, and D.M. and J.J.D.P. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 14 November 2022

Accepted 1 May 2023

Published 7 June 2023

10.1126/sciadv.adf7541

Systematic modification of functionality in disordered elastic networks through free energy surface tailoring

Dan Mendels, Fabian Byléhn, Timothy W. Sirk, and Juan J. de Pablo

Sci. Adv., **9** (23), eadf7541.
DOI: 10.1126/sciadv.adf7541

View the article online

<https://www.science.org/doi/10.1126/sciadv.adf7541>

Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)