

Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings

Michael R. Mitchell^{a,b,1}, Tsvi Tlusty^{c,d,e,1}, and Stanislas Leibler^{a,b,c,1}

^aLaboratory of Living Matter, The Rockefeller University, New York, NY 10065; ^bCenter for Studies in Physics and Biology, The Rockefeller University, New York, NY 10065; ^cThe Simons Center for Systems Biology, School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540; ^dCenter for Soft and Living Matter, Institute for Basic Science, Ulsan 689-798, South Korea; and ^eDepartment of Physics, Ulsan National Institute of Science and Technology, Ulsan 689-798, South Korea

Contributed by Stanislas Leibler, August 2, 2016 (sent for review June 13, 2016; reviewed by Alexander Y. Grosberg and Henri Orland)

In many proteins, especially allosteric proteins that communicate regulatory states from allosteric to active sites, structural deformations are functionally important. To understand these deformations, dynamical experiments are ideal but challenging. Using static structural information, although more limited than dynamical analysis, is much more accessible. Underused for protein analysis, strain is the natural quantity for studying local deformations. We calculate strain tensor fields for proteins deformed by ligands or thermal fluctuations using crystal and NMR structure ensembles. Strains—primarily shears—show deformations around binding sites. These deformations can be induced solely by ligand binding at distant allosteric sites. Shears reveal quasi-2D paths of mechanical coupling between allosteric and active sites that may constitute a widespread mechanism of allostery. We argue that strain—particularly shear—is the most appropriate quantity for analysis of local protein deformations. This analysis can reveal mechanical and biological properties of many proteins.

strain | protein mechanics | protein allostery | elasticity

Although many proteins fold into well-defined, stable structures, their internal deformations around such average structures often play a vital role in their functions. In allosteric protein regulation, a protein's ability to catalyze reactions or to associate with binding partners is influenced by binding of an allosteric regulator to a spatially distinct site. For example, subunits of the tetrameric protein hemoglobin, which transports oxygen in the bloodstream, undergo allosteric structural shifts as they bind O₂ (1). These allosteric shifts alter the protein's O₂ binding affinity (2), enabling hemoglobin to deliver nearly twice as much oxygen to tissues as it could were it not allosteric (3). Elsewhere, allostery enables cells to modulate the activity of proteins more quickly than other regulatory mechanisms like control of protein synthesis and degradation would permit (e.g., ref. 4). This regulation enables cells to respond rapidly to changing conditions.

Given the significant role that structural shifts of proteins play in their function, there has been substantial effort toward better understanding them. Several models have attempted to explain the mechanism of allostery. The most prominent have been the well-known concerted model of Monod, Wyman, and Changeux (5) and the sequential model of Koshland, Nemethy, and Filmer (6). Recently, there has been increasing consideration of the thermodynamic nature of allostery; allosteric proteins' transitions between different functional states can correspond not to discrete switching between states, but rather to a statistical shift in a population distribution of structural states (reviewed in, e.g., refs. 7–11). In the present work, we remain largely agnostic among these models and simply attempt to exploit experimental data to understand the mechanical properties of allosteric proteins.

Dynamical experiments can provide the most direct information about the conduction of allosteric signals and about protein structural dynamics more generally. Studies using methods including room-temperature crystallography (e.g., ref. 12), time-resolved crystallography (e.g., ref. 13), FRET (e.g., ref. 14), and direct pulling measurements (15, 16) have been published. However,

these techniques are often labor-intensive and technically challenging. Some are applicable only to certain experimentally amenable proteins while some provide valuable but incomplete information. Thus, although such methods provide highly valuable information, many proteins are resistant to study using these methods.

In contrast, thousands of crystallography and NMR structures are publicly available for many proteins in multiple ligand-binding states. Furthermore, standard crystallography and NMR techniques are generally more accessible for the study of new proteins than are methods for direct dynamical measurements. Although these data, being static, cannot provide the richness of full dynamical experiments, they nevertheless contain valuable information concerning the net deformations that take place within proteins. It would therefore be highly desirable to extract useful insights regarding allosteric and other structural properties of proteins solely from datasets comprising several static crystal or NMR structures.

Comparison of Protein Structural States

The simplest and most widespread method used in detailed analysis of related crystal structures is direct, manual inspection of positions, distances, and angles of specific atoms, residues, and bonds in protein structures using molecular viewer programs such as PyMOL, VMD, or Chimera. Such inspection can provide detailed information about differences between a few related structures and may be appropriate for analysis of a small region within a protein such as an enzyme's active site. In such a situation, analysis is confined to a small region of the protein known

Significance

Regulation of biochemical activity is essential for proper cell growth and metabolism. Many proteins' activities are regulated by interactions with other molecules binding some distance away from the proteins' active sites. In such allosteric proteins, active sites should thus be mechanically coupled to spatially removed regulatory regions. We studied crystal and NMR structures of proteins in various regulatory and ligand-binding states. We calculated and analyzed distributions of strains throughout several proteins. Strains reveal allosteric and active sites and suggest that quasi-two-dimensional strained surfaces mediate mechanical couplings between them. Strain analysis of widely available structural data can illuminate protein function and guide future experimental investigation.

Author contributions: M.R.M., T.T., and S.L. designed research; M.R.M. performed research; M.R.M. contributed new reagents/analytic tools; M.R.M., T.T., and S.L. analyzed data; and M.R.M., T.T., and S.L. wrote the paper.

Reviewers: A.Y.G., New York University; and H.O., Alternative Energies and Atomic Energy Commission.

The authors declare no conflict of interest.

¹To whom correspondence may be addressed. Email: mich.r.mitchell@gmail.com, tsveltusty@gmail.com, or livingmatter@rockefeller.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1609462113/-DCSupplemental.

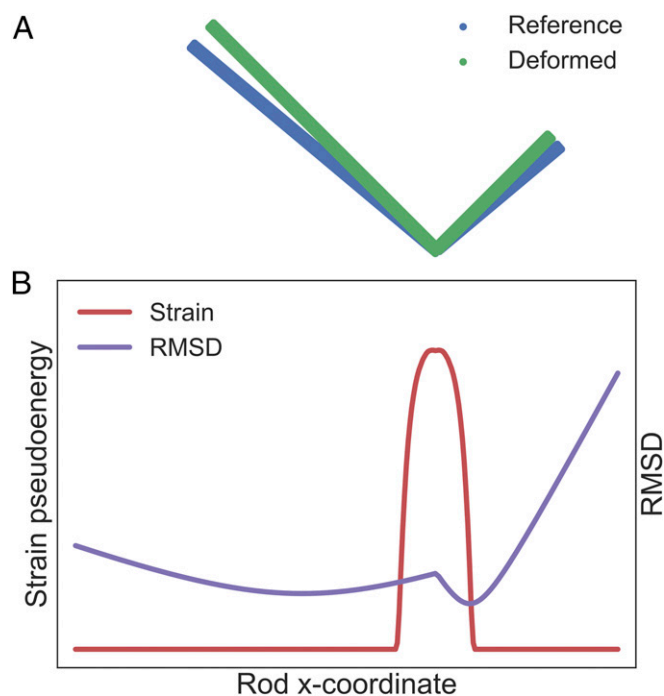


Fig. 1. rmsd vs. strain. (A) Two states of a rod with one flexible, hinge-like region. (B) rmsd of a global alignment of the rod's states primarily highlights regions with large displacement but no internal deformation. Strain highlights the region immediately around the deformation (here, strain tensors are reduced to scalars in the form of their apparent strain energies; see *Strain*).

to be functionally important and measurements can be informed by knowledge of the relevant reaction chemistry. However, this approach is labor-intensive, particularly for analysis of many or large structures. In addition, it does not provide an unbiased framework for analyzing large-scale structural changes or distinguishing isolated fluctuations from biologically important deformations.

Other methods have been applied to analyze large-scale protein deformations, usually based on some variant of global alignment of the structures to be compared (17–21). These global and semilocal methods for comparing protein structures are useful in some circumstances, such as estimating structural similarity of two different proteins. However, because displacements in a structure are cumulative, a global or even semilocal best-fit alignment may highlight apparent “differences” where no local deformation, but only net displacement due to deformation elsewhere, has occurred (Fig. 1). Consequently, alignment-based approaches are not appropriate for studying local structural changes in proteins.

Our goal is to learn about mechanical properties of proteins by examining structural variations between different protein states (free, substrate-bound, regulator-bound, etc.) using comparisons of multiple X-ray crystallography or solution NMR structures. For this purpose, it is most appropriate to consider local displacements, because most forces between atoms in a protein are short-range, contact-like forces (22). Some attempts have been made along these lines. Daily et al. (23) analyzed networks of contact rearrangements to find regions of the protein proposed to transmit structural rearrangements between allosteric and active sites. Some attempts to use normal mode analysis to study relationships between residue mobility and allostery have been made (24, 25), and Miyashita et al. (26) attempted to estimate deformation energies of the elastic network model. Yamato (27) calculated strain tensor fields induced by high pressure in lysozyme and myoglobin. Recently, Yamato and coworkers have applied a similar, molecular dynamics-based approach to supplement their analyses of protein dynamics (13, 28).

These and other studies have yielded insights into the mechanics of globular proteins; however, with the exception of Yamato's 1996 work (see ref. 27), they rely on models layered on top of data, rather than analyzing data directly. We argue that calculation of finite strain tensor fields in deformed protein structures, until now highly underused for this purpose, is the simplest and most natural method for measuring and analyzing protein structural properties. Strains can be calculated directly from a pair of structural coordinates and explicitly describe local deformations between these structures. We will demonstrate that strains reveal significant mechanical coupling between active and regulatory sites in allosteric proteins.

Strain

The standard method in mechanics to describe deformations in (approximately) continuous media is calculation and analysis of strain fields, applied to protein crystal structures by Yamato (27). Strain measures local deformations in an elastic medium. In one spatial dimension, strain is the spatial derivative of the displacement. That is, if a point in an elastic medium at coordinate x is deformed to coordinate $x' = x + u$, then the 1D strain is $\epsilon = \partial u(x)/\partial x$. Generalized to arbitrary dimension, the (Euler–Almansi) strain tensor is

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \sum_k \frac{\partial u_k}{\partial x_i} \frac{\partial u_k}{\partial x_j} \right), \quad [1]$$

where $u_i = x'_i - x_i$ is the i th component of the displacement vector (29). Note that this strain tensor is well-defined for finite deformations u , unlike the (more common, in the physics literature) infinitesimal strain tensor, which contains only the linear terms in Eq. 1.

The strain ϵ_{ij} is a second-rank tensor (similar to a 3×3 matrix). The six independent components of the symmetric strain tensor fully describe any local stretching/compression, twisting, and shear at a point in a structure (the tensor varies across the structure). The strain tensors for each point in an object characterize small deformations of that object. Moreover, strain is a local measure. Unlike rmsd and other global alignment-based quantities, strain measures only local deformations at each point in a structure, so regions of high strain correspond to the sites where deformations originate (and hence to the sites where deformation forces and energies are located; Fig. 1).

The strain tensor's trace (sum of diagonal elements) is a measure of bulk compression or expansion. Its traceless component is a measure of shear. Given strain tensors, shear tensors can be obtained as $\gamma = \epsilon - \frac{1}{3}(\text{Tr } \epsilon) \mathbb{I}$, where \mathbb{I} is the 3×3 identity matrix. Under the simplifying assumptions that both the shear and bulk moduli (strain analogs of spring constants) are isotropic and take the values 1 and 2, respectively, the linear elastic energy of a deformation at site i can be written in terms of shear and bulk components as

$$\begin{aligned} E_{\text{shear}} &= \sum_{m,n} (\gamma_{mn})^2 \\ E_{\text{bulk}} &= \sum_m (\epsilon_{mm})^2 \\ E_{\text{strain}} &= E_{\text{shear}} + E_{\text{bulk}}. \end{aligned} \quad [2]$$

Of course, protein deformation energies are clearly nonlinear, whereas the strain energy defined above assumes linearity. Nevertheless, linearization is a useful tool to aid our understanding of proteins' deformations. Not knowing the precise nature of the proteins' nonlinearities, it is better not to bias our analysis by making assumptions about them. Additionally, as we will demonstrate below, even highly deformed regions of proteins typically have strain magnitudes of only around 10^{-1} ; linearization should be reasonably accurate in this regime. Finally,

other linear methods, such as those of refs. 24 and 25, have been able to generate insights despite their use of this approximation.

In addition, calculation of energies requires knowledge of strain moduli that are not known for proteins and surely vary throughout each protein. An important source of anisotropy in protein strain moduli are forces produced by covalent bonds between backbone atoms, which are much stronger than non-covalent interactions between neighboring atoms and residues. Our approach is to neglect both the nonlinearities and the anisotropies present in the proteins. In this way, we can avoid biasing our analysis with (necessarily incorrect) assumptions regarding the actual distribution of strain moduli in the proteins we study. Instead, we interpret the strain tensors we estimate and the corresponding “pseudoenergies” as aggregate indications of the type and magnitude of local deformations occurring in the proteins under study, and of the local amino acid interactions present throughout those proteins. In doing so, we can use strains to analyze proteins’ deformations and identify residues essential for allostery and other protein functions.

Finally, the concept of strain is typically applied in the context of continuous media. Proteins are clearly not continuous, being composed of residues themselves made up of atoms. However, the theory of elasticity has been remarkably successful in describing the mechanical properties of real materials, all discretely composed of atoms. Indeed, much of the early development of the theory of elasticity—including the models of Fresnel, Navier, Cauchy, Poisson, Voigt, Born, and von Karman—occurred in the context of discrete, atomic media (30, 31), with continuum theories arising as a limit of molecular models. Moreover, the calculation of strain tensors can readily be extended to granular media (32), and strains remain highly useful in describing granular materials’ deformations. Ultimately, our calculation of strain tensors in protein structures is equivalent to calculation of the spatial derivatives of the displacement fields between pairs of structures; these spatial derivatives (i.e., strains) indicate the structure of local deformations independently of any interpretation under the theory of continuum elasticity.

Taken together, the simplifying assumptions of linearity and homogeneous strain moduli mean that, while the calculated pseudoenergy values have the mathematical form of an energy, they can alternatively be interpreted as merely the magnitude of local deformation at each site in the protein. This interpretation assumes neither linearity of strain energies nor homogeneity of strain moduli. In this sense, our analysis is closely analogous to the use of rmsd and related measures for comparing two or more structures, except that where rmsd identifies large global displacements but not their structural origin, our approach identifies local deformations that underlie proteins’ structural variations (Fig. 1).

Results

Protein Strains Are Primarily Shears and Reveal an Enzyme’s Active Site: Adenylate Kinase. We first illustrate the use of strain calculations to describe the equilibrium structure of adenylate kinase measured by solution NMR. We computed deformations between an ensemble of 20 solution NMR structures [Protein Data Bank (PDB) ID code 1P4S; ref. 33]. The NMR structures were obtained using free adenylate kinase, with no ligands present. Because there is no natural reference structure, we calculated the mean strain pseudoenergies across strains for all pairwise structure comparisons. Although the NMR structure ensemble corresponds to a collection of structures consistent with measured NOE constraints rather than a directly measured set of conformations, strains between the calculated structures should provide information about local deformations within the protein. Strain, shear strain, and bulk strain pseudoenergies for each residue in the protein are shown in Fig. 2 A–C. The strain distribution shows that relatively unconstrained loops at the top and

bottom undergo very high shears; in addition, a small pocket between the two loops undergoes relatively high shear deformations. Comparison with a crystal structure of the *Escherichia coli* adenylate kinase (PDB ID code 4JLA; ref. 34) shows that this sheared pocket corresponds to the ADP binding sites (ADP shown in light blue). Here we observe that strains corresponding to structural fluctuations captured by the NMR structure ensemble reveal functionally significant features in the protein, even without perturbation of those sites by ligand binding.

As seen in Fig. 2, distributions of total strain, shear strain, and bulk strain throughout adenylate kinase are quite similar; this observation holds true also for other proteins we studied. Because of this similarity and because globular proteins are generally not very compressible (i.e., bulk strain is a relatively small fraction of total strain), we will show only shear strains for subsequent proteins. Shear pseudoenergies were generally higher than bulk pseudoenergies, but they were typically within an order of magnitude of each other. The ratios of shear pseudoenergy to bulk pseudoenergy for adenylate kinase are shown in Fig. 2D; for all residues, shear pseudoenergy was larger than bulk pseudoenergy, but underconstrained surface sites tended to have lower shear/bulk ratios.

We also calculated strains in a closely related protein, guanylate kinase, based on X-ray structures (PDB ID codes 1ZNV, 1ZNX, and 1ZNY; ref. 35), with similar results (Fig. S1).

Strain Analysis of an Allosteric Protein Reveals Mechanical Coupling Between Allosteric and Active Sites Not Demonstrated by rmsd: Glucokinase. The previous example has shown that analyzing strain distributions in proteins can highlight enzymes’ active sites, but adenylate kinase is not known to be allosteric. As a result the strained region in that protein seems to be confined to the active site. In the case of allosteric proteins, it can be possible to observe strained regions connecting active and allosteric sites, demonstrating a direct mechanical coupling between these sites.

We now consider glucokinase, a monomeric allosteric enzyme that phosphorylates glucose to glucose-6-phosphate. Among its regulators is glucokinase regulatory protein (GRP), which acts as an inhibitor. There are also numerous synthetic activators and inhibitors. We computed shears across 26 different structures of glucokinase including free enzyme, glucose-bound enzyme, GRP-bound enzyme, and multiple synthetic inhibitor- and activator-bound states (PDB ID codes 1V4S, 1V4T, 3A0I, 3F9M, 3FGU, 3FR0, 3GOI, 3H1V, 3ID8, 3IDH, 3IMX, 3S41, 3VEV, 3VEY, 3VF6, 4DCH, 4DHY, 4ISE, 4ISF, 4ISG, 4L3Q, 4LC9, 4MLE, 4MLH, 4NO7, and 4RCH; refs. 36–48), again studying the mean pseudoenergy across all pairwise structure comparisons.

Shear pseudoenergies are shown in Fig. 3A. There is a quasiplanar region of high shear connecting the GRP binding site and the glucose binding site. This finding suggests that in glucokinase information about the GRP binding state is transmitted along the observed manifold of shear strain. This observation is consistent with previous reports that the two domains of glucokinase undergo relative displacement during regulation (43), but by examining strains we can see that this displacement corresponds to a deformation propagated across the protein to the active site by the strained residues shown in Fig. 3. Thus, for this allosteric protein, strain-based analysis reveals not only binding sites (which, in this case, were apparent from the structures) but also how functionally coupled sites are mechanically connected.

We compared the distribution of shears across the glucokinase structure to the distribution of rmsds across the structure. The rmsd distribution is shown in Fig. 3B; it is distinctly different from the distribution of shears. It is diffusely spread across the structure, and the regions of largest rmsd are the top and bottom lobes of the protein, which are largely free of shear. The sheared region extending between the GRP binding site and the glucose binding site, in fact, is near the region of lowest

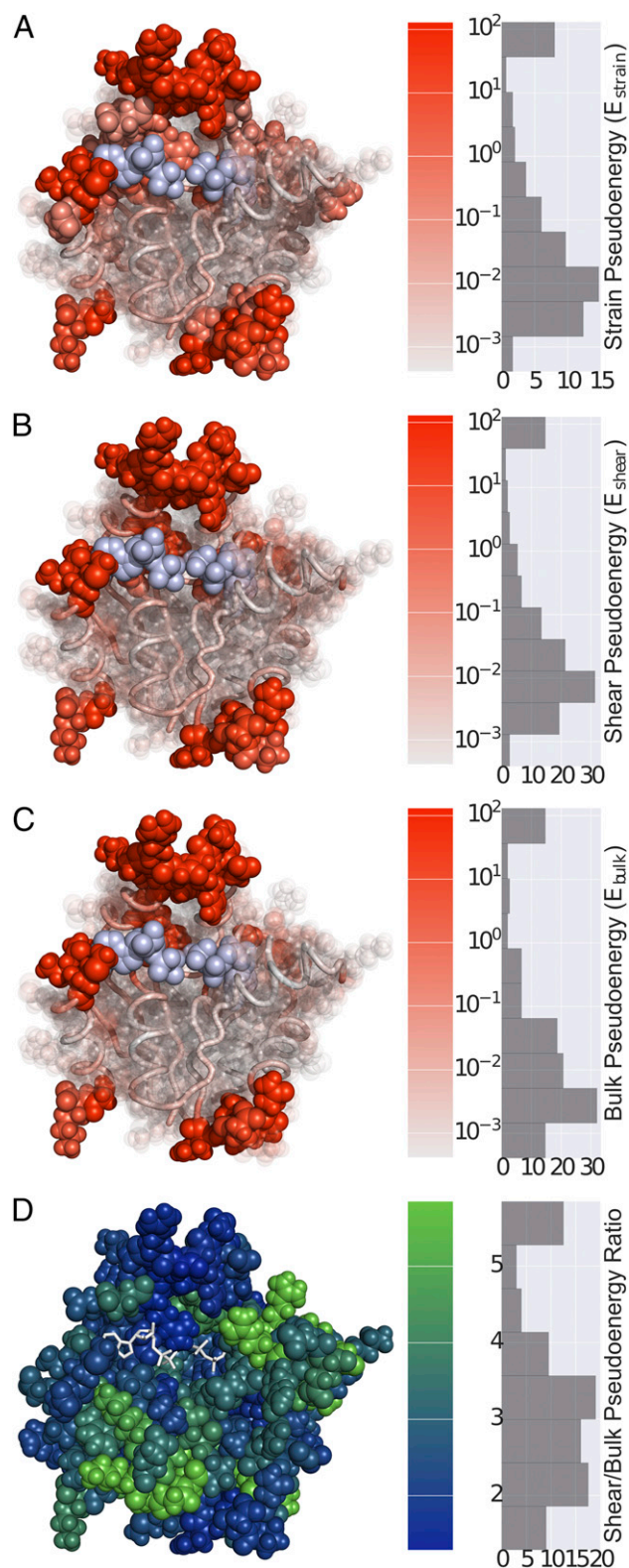


Fig. 2. *Mycobacterium tuberculosis* adenylate kinase strains (PDB ID code 1P4S). (A) Strain pseudoenergies (E_{strain} ; Eq. 2). (B) Shear strain pseudoenergies (E_{shear}). (C) Bulk strain pseudoenergies (E_{bulk}). (D) Ratio of shear to bulk pseudoenergies ($E_{\text{shear}}/E_{\text{bulk}}$). Coloring in A–C is log-scaled (same scale for all subfigures except D). For clarity, only the most strained residues are opaque. Histograms at the right show the distribution of pseudoenergies across all residues. ADP binding sites are indicated by blue ADP spheres (white sticks in D). ADP was not bound in the NMR structure; the binding site was obtained by superimposing an ADP-bound *E. coli* adenylate kinase

rmsd. Although rmsd does show displaced regions of the proteins it does not identify the site of the local deformations causing those displacements. In fact, the shear and rmsd distributions across the glucokinase structure look remarkably like those in the idealized example shown in Fig. 1.

Ligand Binding Reveals Multimeric Allosteric Coupling: Aspartate Carbamoyltransferase. Many allosteric enzymes are multimeric, unlike glucokinase, but it remains possible to detect strained regions that mediate allosteric regulation extending across monomers. Aspartate carbamoyltransferase (ATCase) is a multimeric allosteric enzyme in the pyrimidine biosynthesis pathway whose active form consists of a hexamer of catalytic and regulatory dimers (configuration illustrated in Fig. 4A). **The enzymatic activity of ATCase is allosterically regulated by the binding of one or two nucleotides to each regulatory subunit, with purines (adenosine and guanosine bases) enhancing activity and pyrimidines (cytidine and thymidine bases) inhibiting activity (49).**

We calculated shears between T-state ATCase alone (PDB ID code 6AT1; ref. 50), in the presence of CTP (PDB ID code 5AT1; ref. 50), and in the presence of ATP (PDB ID code 4AT1; ref. 50) and R-state ATCase in the presence of a bisubstrate analog (PDB ID code 8ATC; ref. 51), in the presence of substrate analogs and CTP (PDB ID code 8AT1; ref. 52), and in the presence of substrate analogs and ATP (PDB ID code 7AT1; ref. 52), again studying the mean pseudoenergy across all pairwise structure comparisons. Shear strain pseudoenergies for a single catalytic (C) and regulatory (R) subunit are shown in Fig. 4B; this dimer forms the asymmetric unit in the crystal structures analyzed. Substantially sheared residues are visible around both the active site at bottom and the allosteric site at top, but the middle region (constituting the lower part of the regulatory subunit and the upper part of the catalytic subunit) seems very static. It is unclear from this view how the shears at the allosteric site are able to modulate activity at the active site.

This question is clarified by considering the biological multimer, shown in Fig. 4C. The biological assembly consists of two trimers of catalytic subunits at the top (C1–3) and bottom (C4–6) in a threefold symmetry (seen edge-on) bound to three dimers of regulatory subunits arranged around the catalytic subunits in a threefold symmetry (also seen edge-on; here, regulatory subunits R3 and R4 are hidden behind the structure). The allosteric sites are exposed at the outer surface of each regulatory dimer, while the active sites are at the interface between the two catalytic trimers in the center of the image.

In this view, it appears that shears produced by purine or pyrimidine binding at the allosteric site are transmitted through the unstrained regions of the regulatory and catalytic subunits to produce displacements and corresponding shears at the active sites. In this way, shear strains are transmitted through the quaternary structure of a multimeric allosteric enzyme to provide the mechanical coupling required for regulation.

Strained Residues Lie in Quasi-2D Manifolds. The distribution of strained residues we observed in the studied proteins seems to be structured. To characterize the geometry of strained regions within proteins, we computed the correlation dimension of the structure composed of the most strained residues (53, 54). The correlation dimension is a form of the fractal dimension appropriate for use on datasets with relatively few points. Using the correlation

structure (PDB ID code 4JLA) on the *M. tuberculosis* NMR structure. A region of high strain surrounds the ADP binding sites. High-strain regions visible at the top and bottom may be due to underconstraint of exposed residues there (see Fig. S5 for discussion). Patterns of total strain, shear strain, and bulk strain are similar, but bulk strain pseudoenergies are consistently lower than shear and total strain pseudoenergies.

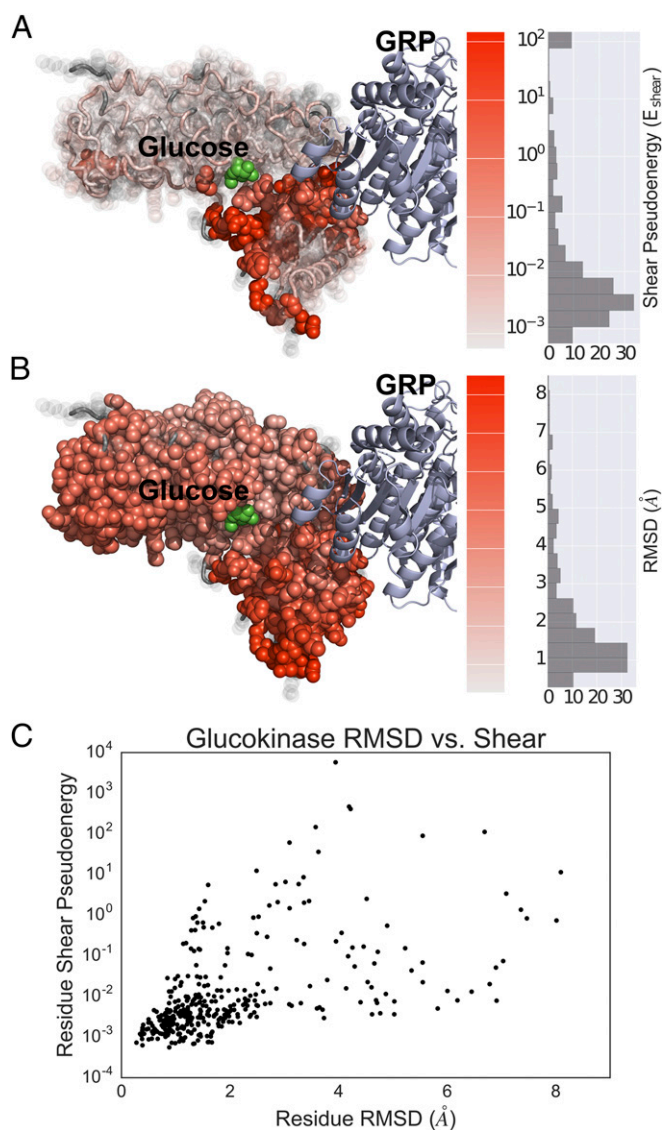


Fig. 3. *Homo sapiens* glucokinase shear vs. rmsd. (A) Glucokinase shear pseudoenergies (E_{shear}) (PDB ID code 4LC9). Gray corresponds to low shear energy, and red corresponds to high energy. Coloring is log-scaled. For clarity, only the most sheared residues are opaque. The histogram at the right shows the distribution of pseudoenergies across all residues. Dark gray residues at the top left were unresolved in some of the structures analyzed, so no strain values were computed for them. The glucose substrate is shown in green. GRP, a regulator of glucokinase, is shown as a light blue ribbon. A path of high shear is visible connecting the GRP binding site and the active site. (B) Distribution of rmsds across the glucokinase structure; compare with the shear distribution in A. Unlike the compact, quasiplanar region of shear extending from the GRP binding site to the glucose binding site, rmsds are diffusely spread across the structure, primarily localizing in the two, largely unshaped regions at top and bottom. (C) Comparison of per-residue rmsd values and log shear pseudoenergies. rmsd and shear are correlated, but there are distinct discrepancies (Pearson correlation coefficient: 0.46). These discrepancies indicate that rmsd and similar global alignment-based approaches do not necessarily highlight the sites corresponding to local deformations.

dimension, we are able to estimate the approximate dimensionality of the structure formed by strained residues. Diffuse, unstructured distribution throughout a protein would produce a correlation dimension close to 3. A correlation dimension significantly less than 3 suggests that residues are positioned along some lower-dimensional manifold extending between distinct regions of the protein.

We analyzed the correlation dimension of strained regions in four allosteric proteins: glucokinase and ATCase (as previously described) in addition to human serum albumin and hemoglobin. Human serum albumin is a monomeric allosteric protein; it is known that binding of warfarin allosterically decreases the protein's heme affinity by an order of magnitude (55). We calculated strains (Fig. S2) between apo-human serum albumin and human serum albumin bound to heme, warfarin, oxyphenbutazone, or phenylbutazone (two other drugs that bind in the same site as warfarin) (PDB ID codes 1E78, 2BXB, 2BXC, 2BXD, and 1N5U; refs. 56–58). Hemoglobin is a tetrameric allosteric protein. We calculated hemoglobin strains (Fig. S3) based on an ensemble of solution NMR structures (PDB ID code 2M6Z; ref. 59).

For all four allosteric proteins, we computed the correlation dimension of the N most sheared residues, varying N from only a few residues to the size of the protein. Results were similar for all proteins we studied. For very small clusters, anomalously high dimensions were observed, presumably because those clusters were not connected. As cluster size increased, the cluster dimension decreased to approximately 2 and remained constant across a size span of several tens of residues, before eventually increasing to a dimension around 3, consistent with the dimension of the protein as a whole (Fig. 5). This apparent two-dimensionality suggests that the most strained residues of these proteins are distributed nonrandomly in some way. In particular, they seem to lie on quasi-2D manifolds, consistent with the occurrence of shears along a shear surface between relatively rigid 3D regions.

To verify that this apparent two-dimensionality was not an artifact of our calculation or the proteins' structures, we repeated the calculation using randomized shear pseudoenergies and found that the dimensionality of randomly ranked residue coordinates never decreased significantly below 3. To check that the apparent two-dimensionality was not simply a consequence of shears on the proteins' surfaces, we also performed this randomization while maintaining the distribution of surface and interior sites across the ranking, with similar results (Fig. S4).

It is significant that this low-dimensionality result holds for both the two multimeric proteins studied (ATCase and hemoglobin) and the two monomeric proteins (glucokinase and human serum albumin). Although it is easy to see that a multimeric protein might be susceptible to shear along intersubunit boundaries, similar quasi-2D surfaces are present in monomeric proteins with no such interfaces.

Discussion

Protein mechanics underlie many important protein functions, including ligand binding and allostery, making better understanding of how the mechanical properties of proteins contribute to these functions quite valuable. Although direct dynamical experiments are the most powerful approach for detecting structural transitions within proteins, the required techniques are challenging, labor-intensive, and not yet widely accessible. Moreover, some of these techniques such as room-temperature crystallography produce datasets whose meaning is not clear without powerful analytic methods. A great deal of effort has been invested in generating static structural information, which is available for over 67,000 distinct protein sequences (60), with multiple biochemical states present for many of these proteins. Methods to infer mechanical properties of proteins from these already available data can supplement (although not replace) direct experimental approaches.

We observed shear around allosteric and active sites as a consequence of ligand binding at those sites, of ligand binding at mechanically coupled sites, and of mobility at those sites in solution NMR ensembles. A limitation of our method is that, to avoid imposing strain moduli that we are not capable of measuring, we interpreted apparent strain energies computed using

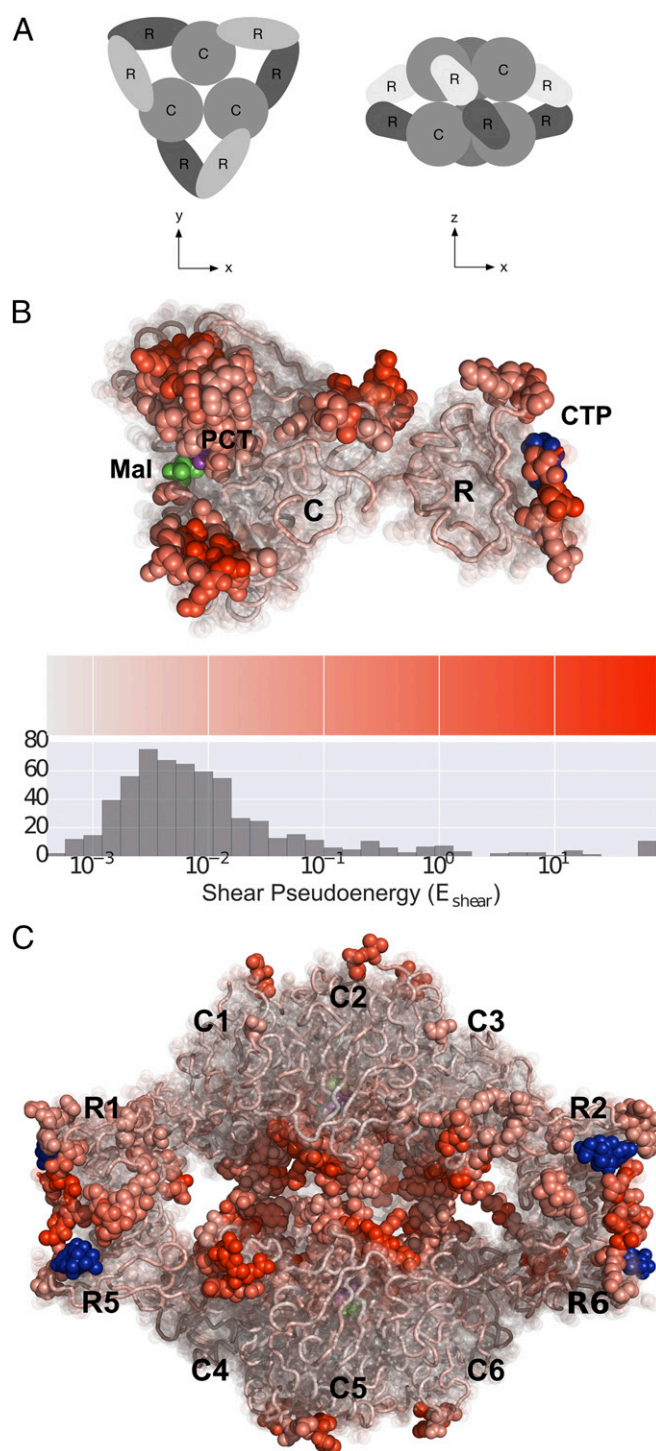


Fig. 4. *E. coli* ATCase shear pseudoenergy distribution. (A) ATCase quaternary structure. The complex consists of two triangular catalytic trimers stacked next to each other, with three regulatory dimers arranged around them in a threefold symmetry. (B) ATCase shear pseudoenergies (E_{shear}) in one regulatory (R) and one catalytic (C) subunit (the crystallographic asymmetric unit) (PDB ID code 8AT1). Gray corresponds to low shear energy, and red corresponds to high energy. Coloring is log-scaled. For clarity, only the most sheared residues are opaque. The histogram at the bottom shows the distribution of pseudoenergies across all residues. Substrates maltose (Mal) and phosphonoacetamide (PCT) are shown in purple and green, respectively. CTP (blue) is bound at the allosteric site. Shear is visible around both the allosteric and active sites, but no propagation of shear is visible from one to the other. (C) In the biological assembly, it seems that shear is transmitted from the allosteric sites to the active sites (the red regions in the center) via

nominal strain moduli as indications of deformation magnitude rather than as true energies. This approximation means we cannot differentiate between flexible but weakly coupled low-strain regions and strongly coupled but rigid low-strain regions.

Heuristically, we can filter out strained residues with relatively low coordination number (and perhaps correspondingly low local rigidity); this filtering tends to distinguish isolated surface residues and loops with high shear pseudoenergy from strained regions around active and allosteric sites (Fig. S5). Alternatively, one can divide strain pseudoenergies by the β -factors at each site; in this way, the influence of thermal fluctuations can be minimized relative to the contribution of concerted deformations (Fig. S6). Ishikura et al. (61) proposed using short molecular dynamics simulations to thermalize protein crystal structures, then using the molecular dynamics force field to calculate the stress field describing the protein's state. Such an approach would sidestep our limitation but necessitate recourse to a less data-driven and more computationally intensive approach, namely molecular dynamics. A combination of stress and strain analysis could permit more complete analysis of the mechanical properties of individual proteins.

We expect that the allosteric nature of proteins exhibiting the shear phenomena we describe should depend strongly on the mechanical properties of the sheared residues. Our results, therefore, can readily be subjected to experimental verification by determining the robustness of proteins' allosteric regulation to mutation of these highly sheared residues. For example, mutations in the sheared region of glucokinase (the simplest example here because it is monomeric) that limit mobility in that region should reduce the protein's responsiveness to regulation by GRP.

Strain (and shear strain in particular) is a natural quantity to use in describing local deformations in proteins because it emphasizes the short-range forces that dominate protein structural dynamics and deformations. We have shown that active and allosteric sites are highly strained regions in proteins. In at least some allosteric proteins there seem to be shear surfaces transmitting deformations between active and allosteric sites. These sheared regions appear to be roughly 2D, indicating that strains are distributed nonrandomly through the proteins studied and may form surfaces along which deformations can be transmitted through proteins.

For some time conventional wisdom indicated that allostery was limited to multimeric proteins owing to the need for the subunits to communicate their states to each other to "cooperate." There are now numerous monomeric allosteric proteins known (we have included analysis of two of them here), but it seems possible that multimeric allosteric proteins are common not just because they enable cooperation between multiple active sites but also because they provide a convenient shearable manifold between the subunits, which can act to transmit deformations between subunits. Even the monomeric allosteric enzymes we analyzed, though, seem to use quasi-2D surfaces to transmit deformations between binding sites, even transverse to the peptide chain (Fig. S7).

The apparent two-dimensionality of shear surfaces in the proteins described here raises the possibility that 2D shear "faults" are a common mechanism for propagation of protein deformations. A quasi-2D shear manifold may minimize the number of amino acids involved in a protein deformation. By extension, a quasi-2D shear manifold could minimize both size of the sequence space to be explored in evolving a particular function and the deformation energy involved in an allosteric transition, although

the unsheared regions of the catalytic domains. Shears appear localized along the planar interface between the two halves of the complex.

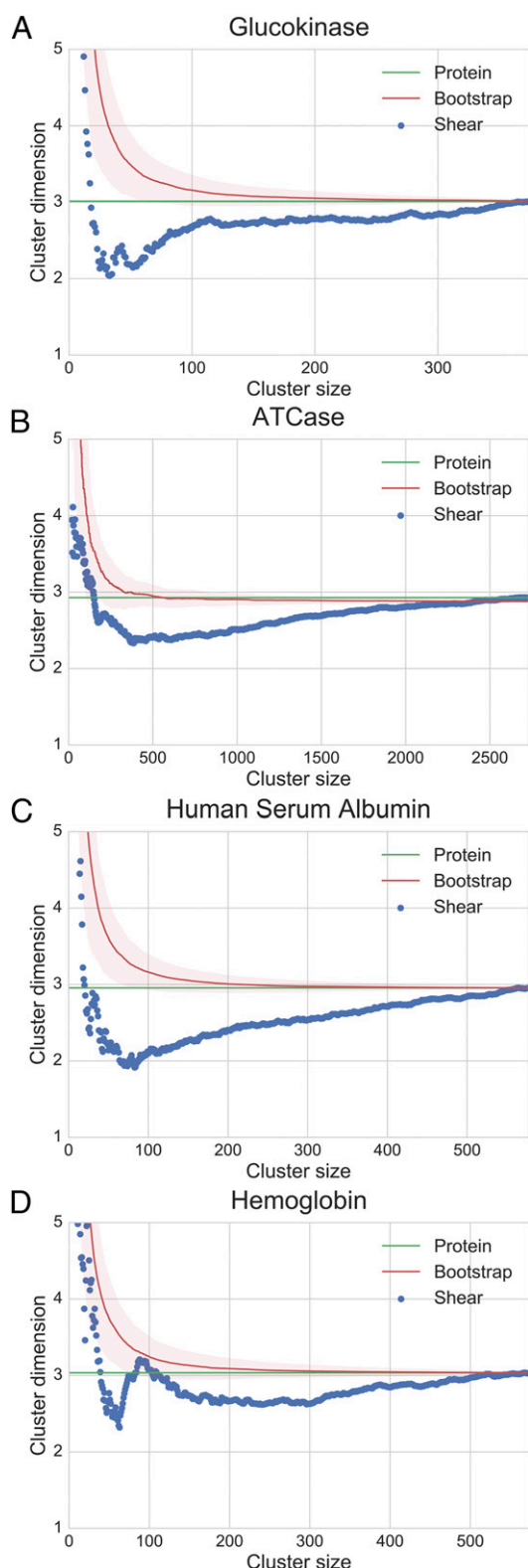


Fig. 5. Strain region dimension for several proteins. Correlation dimension of the N most strained residues was computed for N ranging from 10 up to the protein size, as described in *Materials and Methods*. Correlation dimension of the protein is shown as a green horizontal line for reference. A background distribution of the dimensions of random residue rankings is shown (mean \pm SD) for comparison. (A) Glucokinase. (B) ATCase. (C) Human serum albumin. (D) Hemoglobin. In all cases, there is a period of quasi-two-dimensionality, typically for N around 10% of the protein's size. Glucokinase and human serum albumin are monomeric, whereas ATCase and hemoglobin are multimeric.

further study is needed to better understand the origin of this phenomenon.

The examples we have shown indicate that analysis of strains and shears in protein structures can reveal mechanically important sites in proteins. This method can highlight regions with high deformations and is more appropriate for this purpose than rmsd-based and similar analyses (Fig. 3 A–C). Because the presence of shears is expected to be a mechanism for propagation of deformations between functionally important sites in proteins, analysis of shears caused by ligand binding at one site in a protein can reveal mechanically and functionally coupled sites not yet experimentally investigated. Shears seem to be propagated along quasi-2D manifolds in all cases studied here, suggesting that this may be a widespread mechanism for transmission of deformations through proteins. Finally, analysis of strains even in thoroughly studied proteins can illuminate the mechanics and mechanisms of their regulation.

Materials and Methods

Protein Strain Calculation. There are at least two straightforward methods for computing strain between two structures. One approach is to calculate the Delaunay tessellation of each structure and compute the continuum strain of each resulting cell produced by the deformation between the two structures (27). The method used in this work employs reference and displaced coordinates to estimate at each site (atom or amino acid, for example) the matrix describing the local deformation. From this matrix, the stress tensor can be readily computed (62).

Consider a protein structure described by a set of points x_α in some reference state (perhaps with no binding partners) and a corresponding structure described by the points $x'_\alpha = \chi(x_\alpha)$ in a deformed state (perhaps with an allosteric effector bound). Locally, the deformation of the structure at a point is given by the derivative $F_\alpha = \partial\chi/\partial x_\alpha = \partial x'_\alpha/\partial x_\alpha$. To first order, this derivative can be written as $dx'_\alpha \approx F_\alpha dx_\alpha$. Given the matrix F_α at a point α , the Eulerian strain tensor can be calculated as $\epsilon^\alpha = \frac{1}{2}[\mathbb{I} - (F_\alpha^T F_\alpha)^{-1}]$, which is equivalent to Eq. 1. Hence, calculation of strain tensors throughout a protein structure is reduced to estimating the local deformation matrix at each atom or amino acid.

Note that the approximation $dx'_\alpha \approx F_\alpha dx_\alpha$ assumes that the material being studied is continuous rather than granular, like a protein consisting of collections of atoms. Our calculated strains, therefore, are discrete approximations that nevertheless represent local deformations in the proteins. Results shown here are based on strains calculated between amino acids; calculating strains between individual atoms produces similar results (Fig. S8).

Determination of the deformation gradients F_α essentially corresponds to estimation of spatial derivatives in the structures. For each site α , consider all other sites β within some neighborhood N_α of point α . Define the relative positions of points α and β in the reference and deformed structures as $\Delta x_{\alpha\beta} = x_\alpha - x_\beta$ and $\Delta x'_{\alpha\beta} = x'_\alpha - x'_\beta$, respectively. To first order,

$$\Delta x'_{\alpha\beta} = F_\alpha \Delta x_{\alpha\beta}. \quad [3]$$

Provided N_α , the neighborhood of α , contains at least three points, then the system Eq. 3 for all $\beta \in N_\alpha$ is overdetermined and can be solved in the least squares sense to give an estimate of F_α for each point α in the structure. In practice, this least-squares solution was weighted according to simple distance-based weight functions. Amino acid-resolution strain calculations used the piecewise linear radial weight function

$$w_{\alpha\beta} = \begin{cases} 1 & : r_{\alpha\beta} \leq 6\text{\AA} \\ 1 - \frac{1}{2}(r_{\alpha\beta} - 6) & : 6\text{\AA} < r_{\alpha\beta} \leq 8\text{\AA} \\ 0 & : r_{\alpha\beta} > 8\text{\AA} \end{cases} \quad [4]$$

although results did not depend strongly on the exact choice of weight function, provided the resultant neighborhoods were large enough to fully determine Eq. 3 (typically at least 6Å) and not so large as to include too much of the protein.

It is important to note that the quantity we study—shear pseudoenergy—is invariant under rotations of the coordinate system used to represent the protein structures. If not for this fact, then the outcome of our analysis would depend arbitrarily on the axes chosen when assigning atomic coordinates in structures. One should expect a spring energy to be so invariant, and indeed

the shear strain energy is rotationally invariant. The three invariants of any symmetric second-order tensor A are

$$\begin{aligned} I_A &= \text{Tr}(A) \\ II_A &= \frac{1}{2} (\text{Tr}(A)^2 - \text{Tr}(A^2)) \\ III_A &= \det(A). \end{aligned}$$

It should therefore be possible to write the shear strain energy as a function of these invariants. Shear strain energy can be written as $\text{Tr}(\epsilon^2) - \frac{1}{3} \text{Tr}(\epsilon)^2$, which is a function of I_A and II_A , by observing that the energy is the squared Frobenius norm of the shear tensor, which can be written $\text{Tr}(\gamma^2)$. Substituting the definition of the shear tensor γ yields the previous expression in terms of the first two symmetric tensor invariants. Although we could analyze other invariants of the shear tensor such as the trace or determinant, we are interested in a quadratic spring energy and so consider a function involving the second tensor invariant.

In analysis of experimental data, the lower limit for meaningful measurement of strain pseudoenergies is an important consideration. That is, very small apparent strain pseudoenergies might derive from the resolution limit of experimental data. We calculated shear pseudoenergies across an ensemble of five crystal structures of similarly prepared *Bos taurus* trypsin, which had been produced for the purpose of analyzing reproducibility of crystallography data (PDB ID codes 4I8G, 4I8H, 4I8J, 4I8K, and 4I8L; ref. 63). Typical shear pseudoenergies in that dataset were $10^{-5} \sim 10^{-4}$, suggesting that shear pseudoenergies larger than this are most likely due to meaningful differences between structures (Fig. S9).

Protein structure files were parsed using the Biopython PDB module (64, 65). Where PDB residue numbering differed between structure files for the same protein, residues were aligned using the EMBOSS water Smith–Waterman alignment program (66).

Tables of raw strain, shear strain, and bulk strain pseudoenergies in the proteins described here will be provided on request.

Protein rmsd Calculation. To compare strain calculations to rmsd-based structural comparison, we calculated per-residue rmsds for some proteins. We aligned ensembles of structures to globally minimize total rmsd using the iterative method of Wang and Snoeyink (67). We then computed the rmsd for each residue relative to the average structure.

Protein Structure Selection. Because our analysis made use of preexisting structural data, comparability of the structures involved is an important consideration; experimental conditions such as pH, salt content, and crystal symmetry could affect structures and produce spurious results. For our analysis of adenylate kinase, we made use of a solution NMR structure ensemble; experimental variability of this type is not a concern for NMR data, because all structures were obtained in the same solution. For our analysis of guanylate kinase, we used three structures crystallized by the same group for the same publication; all structures had the same symmetry group and very similar unit cell parameters, and chemical conditions were quite similar for all three. In the case of glucokinase, more variability between crystal properties was unavoidable, because some structures were of free glucokinase whereas others were of glucokinase complexed with other proteins. However, in this case, we were able to average over 325 pairwise comparisons between 26 distinct structures; any idiosyncratic variations due to experimental conditions should contribute negligibly to the final result in this case. Our analysis of ATCase made use of six crystal structures, all obtained under as similar

chemical conditions as could be expected, and all exhibiting the same symmetry group and very similar unit cell parameters. Our analysis of hemoglobin was based on analysis of solution NMR structures, for which these concerns are not relevant. The human serum albumin structures were based on crystals produced under similar chemical conditions, and four of the five structures we analyzed were based on crystals with the same space group. Finally, the consistency of our results across all of the proteins studied suggests that experimental variations contributed negligibly to the phenomena we describe. The care that must be exercised in this matter does emphasize the importance of additional, carefully consistent experimentation.

Strain Region Dimension. The fractal dimensions of point clouds corresponding to the most strained protein residues were estimated using the correlation dimension (53, 54). For a set of points $\{\vec{X}_i\}$, the correlation dimension is determined from the correlation integral

$$\begin{aligned} C(r) &= \lim_{M \rightarrow \infty} \frac{1}{M^2} \sum_{i,j=1}^M \theta(r - |\vec{X}_i - \vec{X}_j|) \\ &= \int_0^r d^d r' c(r'), \end{aligned} \quad [5]$$

which can be approximated for a finite set of size N (68) by

$$C(N, r) \approx \frac{1}{N(N-1)} \sum_{i \neq j} \theta(r - |\vec{X}_i - \vec{X}_j|). \quad [6]$$

The correlation dimension ν , then, is

$$\nu = \lim_{r \rightarrow 0} \frac{\log C(N, r)}{\log r}. \quad [7]$$

The correlation integral was computed across a range of radii spanning the smallest and largest pairwise distances in the protein structure. The gradient $d \log C(N, r) / d \log r$ was estimated using the first-order central difference. The dimension was estimated as the maximum of the smoothed gradient for $r > 8 \text{ \AA}$.

To explore the relationship between strain cluster size and dimension, correlation dimension of the N most strained residues (by strain, shear strain, or bulk strain) was estimated for N from 10 residues up to the protein size.

To help interpret the results, the correlation dimension of the entire protein was also computed. Additionally, a background dimension distribution was bootstrapped by repeatedly computing dimension as a function of cluster size using random residue rankings rather than strain pseudoenergies. To ensure the apparent dimensionality of strain clusters was not a surface artifact, this bootstrapping was also performed by randomly permuting strain labels among surface and interior residues separately; results were very similar (Fig. S4). Surface residues were defined using the Biopython PDB module in conjunction with the MSMS tool (64, 65, 69).

ACKNOWLEDGMENTS. We thank Jean-Pierre Eckmann, Doeke Hekstra, David Huse, and Olivier Rivoire for discussions, comments, and suggestions. M.R.M. thanks the Institute for Advanced Study for its hospitality. This research has been partly supported by grants from the Simons Foundation to S.L. through The Rockefeller University (Grant 345430) and the Institute for Advanced Study (Grant 345801). This material is based upon work supported by National Science Foundation Graduate Research Fellowship Grant DGE-1325261. T.T. has been partly supported by the Institute for Basic Science Grant IBS-R020-D1 and The Simons Center for Systems Biology at the Institute for Advanced Study.

- Perutz MF (1970) Stereochemistry of cooperative effects in haemoglobin. *Nature* 228(5273):726–739.
- Hill AV (1910) The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J Physiol* 40(suppl):iv–vii.
- Stryer L (1995) *Biochemistry* (Freeman, New York), 4th Ed.
- Van Schaftingen E (1989) A protein from rat liver confers to glucokinase the property of being antagonistically regulated by fructose 6-phosphate and fructose 1-phosphate. *Eur J Biochem* 179(1):179–184.
- Monod J, Wyman J, Changeux JP (1965) On the nature of allosteric transitions: A plausible model. *J Mol Biol* 12:88–118.
- Koshland DE, Jr, Némethy G, Filmer D (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry* 5(1):365–385.
- Kern D, Zuiderweg ER (2003) The role of dynamics in allosteric regulation. *Curr Opin Struct Biol* 13(6):748–757.
- Cui Q, Karplus M (2008) Allostery and cooperativity revisited. *Protein Sci* 17(8):1295–1307.
- Tsai CJ, Del Sol A, Nussinov R (2009) Protein allostery, signal transmission and dynamics: A classification scheme of allosteric mechanisms. *Mol Biosyst* 5(3):207–216.
- Changeux JP (2012) Allostery and the Monod-Wyman-Changeux model after 50 years. *Annu Rev Biophys* 41:103–133.
- Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. *Nature* 508(7496):331–339.
- Fraser JS, et al. (2009) Hidden alternative structures of proline isomerase essential for catalysis. *Nature* 462(7273):669–673.
- Tomita A, et al. (2009) Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. *Proc Natl Acad Sci USA* 106(8):2612–2616.
- Rice S, et al. (1999) A structural change in the kinesin motor protein that drives motility. *Nature* 402(6763):778–784.
- Choi B, et al. (2005) Artificial allosteric control of maltose binding protein. *Phys Rev Lett* 94(3):038103.
- Wang Y, Zocchi G (2010) Elasticity of globular proteins measured from the ac susceptibility. *Phys Rev Lett* 105(23):238104.

17. Nichols WL, Rose GD, Ten Eyck LF, Zimm BH (1995) Rigid domains in proteins: an algorithmic approach to their identification. *Proteins* 23(1):38–48.
18. Kelley LA, Gardner SP, Sutcliffe MJ (1997) An automated approach for defining core atoms and domains in an ensemble of NMR-derived protein structures. *Protein Eng* 10(6):737–741.
19. Wriggers W, Schulten K (1997) Protein domain movements: Detection of rigid domains and visualization of hinges in comparisons of atomic coordinates. *Proteins* 29(1):1–14.
20. Schneider TR (2002) A genetic algorithm for the identification of conformationally invariant regions in protein molecules. *Acta Crystallogr D Biol Crystallogr* 58(Pt 2): 195–208.
21. Damm KL, Carlson HA (2006) Gaussian-weighted RMSD superposition of proteins: A structural comparison for flexible proteins and predicted protein structures. *Biophys J* 90(12):4558–4573.
22. Finkelstein A, Pitsyn O (2002) *Protein Physics: A Course of Lectures, Soft Condensed Matter, Complex Fluids and Biomaterials* (Elsevier, Amsterdam).
23. Daily MD, Upadhyaya TJ, Gray JJ (2008) Contact rearrangements form coupled networks from local motions in allosteric proteins. *Proteins* 71(1):455–466.
24. Hinsen K, Thomas A, Field MJ (1999) Analysis of domain motions in large proteins. *Proteins* 34(3):369–382.
25. Panjkovich A, Daura X (2012) Exploiting protein flexibility to predict the location of allosteric sites. *BMC Bioinformatics* 13(1):273.
26. Miyashita O, Onuchic JN, Wolynes PG (2003) Nonlinear elasticity, proteinquakes, and the energy landscapes of functional transitions in proteins. *Proc Natl Acad Sci USA* 100(22):12570–12575.
27. Yamato T (1996) Strain tensor field in proteins. *J Mol Graph* 14(2):105–107, 98–99.
28. Koike K, Kawaguchi K, Yamato T (2008) Stress tensor analysis of the protein quake of photoactive yellow protein. *Phys Chem Chem Phys* 10(10):1400–1405.
29. Landau L, Lifshitz E, Kosevich A, Pitaevskii L (1986) *Theory of Elasticity, Course of Theoretical Physics* (Butterworth-Heinemann, Amsterdam).
30. Born M, Huang K (1954) *Dynamical Theory of Crystal Lattices* (Clarendon, London).
31. Capecchi D, Ruta G, Trovati P (2010) From classical to Voigt's molecular models in elasticity. *Arch Hist Exact Sci* 64(5):525–559.
32. Bagi K (1996) Stress and strain in granular assemblies. *Mech Mater* 22(3):165–177.
33. Miron S, Munier-Lehmann H, Craescu CT (2004) Structural and dynamic studies on ligand-free adenylate kinase from *Mycobacterium tuberculosis* revealed a closed conformation that can be related to the reduced catalytic activity. *Biochemistry* 43(1): 67–77.
34. Kerns SJ, et al. (2015) The energy landscape of adenylate kinase during catalysis. *Nat Struct Mol Biol* 22(2):124–131.
35. Hible G, et al. (2006) Unique GMP-binding site in *Mycobacterium tuberculosis* guanosine monophosphate kinase. *Proteins* 62(2):489–500.
36. Kamata K, Mitsuya M, Nishimura T, Eiki JI, Nagata Y (2004) Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. *Structure* 12(3):429–438.
37. Mitsuya M, et al. (2009) Discovery of novel 3,6-disubstituted 2-pyridinecarboxamide derivatives as GK activators. *Bioorg Med Chem Lett* 19(10):2718–2721.
38. Nishimura T, et al. (2009) Identification of novel and potent 2-amino benzamide derivatives as allosteric glucokinase activators. *Bioorg Med Chem Lett* 19(5): 1357–1360.
39. Petit P, et al. (2011) The active conformation of human glucokinase is not altered by allosteric activators. *Acta Crystallogr D Biol Crystallogr* 67(Pt 11):929–935.
40. Beberitz GR, et al. (2009) Investigation of functionally liver selective glucokinase activators for the treatment of type 2 diabetes. *J Med Chem* 52(19):6142–6152.
41. Takahashi K, et al. (2009) The design and optimization of a series of 2-(pyridin-2-yl)-1H-benzimidazole compounds as allosteric glucokinase activators. *Bioorg Med Chem* 17(19):7042–7051.
42. Pfefferkorn JA, et al. (2011) Designing glucokinase activators with reduced hypoglycemia risk: Discovery of N,N-dimethyl-5-(2-methyl-6-(5-methylpyrazin-2-yl)-carbamoyl)benzofuran-4-yloxy)pyrimidine-2-carboxamide as a clinical candidate for the treatment of type 2 diabetes mellitus. *Med Chem Commun* 2(9):828–839.
43. Liu S, et al. (2012) Insights into mechanism of glucokinase activation: Observation of multiple distinct protein conformations. *J Biol Chem* 287(17):13598–13610.
44. Cheruvallath ZS, et al. (2013) Design, synthesis and SAR of novel glucokinase activators. *Bioorg Med Chem Lett* 23(7):2166–2171.
45. Filipinski KJ, et al. (2013) Pyrimidine-based series of glucokinase activators with alternative donor-acceptor motif. *Bioorg Med Chem Lett* 23(16):4571–4578.
46. Beck T, Miller BG (2013) Structural basis for regulation of human glucokinase by glucokinase regulatory protein. *Biochemistry* 52(36):6232–6239.
47. Hinklin RJ, et al. (2013) Identification of a new class of glucokinase activators through structure-based design. *J Med Chem* 56(19):7669–7678.
48. Hinklin RJ, et al. (2014) Discovery of 2-pyridylureas as glucokinase activators. *J Med Chem* 57(19):8180–8186.
49. Cockrell GM, et al. (2013) New paradigm for allosteric regulation of *Escherichia coli* aspartate transcarbamoylase. *Biochemistry* 52(45):8036–8047.
50. Stevens RC, Gouaux JE, Lipscomb WN (1990) Structural consequences of effector binding to the T state of aspartate carbamoyltransferase: Crystal structures of the unligated and ATP- and CTP-complexed enzymes at 2.6-Å resolution. *Biochemistry* 29(33):7691–7701.
51. Ke HM, Lipscomb WN, Cho YJ, Honzatko RB (1988) Complex of N-phosphonacetyl-L-aspartate with aspartate carbamoyltransferase. X-ray refinement, analysis of conformational changes and catalytic and allosteric mechanisms. *J Mol Biol* 204(3): 725–747.
52. Gouaux JE, Stevens RC, Lipscomb WN (1990) Crystal structures of aspartate carbamoyltransferase ligated with phosphonoacetamide, malonate, and CTP or ATP at 2.8-Å resolution and neutral pH. *Biochemistry* 29(33):7702–7715.
53. Grassberger P, Procaccia I (1983) Characterization of strange attractors. *Phys Rev Lett* 50(5):346–349.
54. Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. *Physica D* 9(1–2):189–208.
55. Baroni S, et al. (2001) Effect of ibuprofen and warfarin on the allosteric properties of haem-human serum albumin. A spectroscopic study. *Eur J Biochem* 268(23): 6214–6220.
56. Bhattacharya AA, Curry S, Franks NP (2000) Binding of the general anesthetics propofol and halothane to human serum albumin. High resolution crystal structures. *J Biol Chem* 275(49):38731–38738.
57. Ghuman J, et al. (2005) Structural basis of the drug-binding specificity of human serum albumin. *J Mol Biol* 353(1):38–52.
58. Wardell M, et al. (2002) The atomic structure of human methemalbumin at 1.9 Å. *Biochem Biophys Res Commun* 291(4):813–819.
59. Fan JS, et al. (2013) Solution structure and dynamics of human hemoglobin in the carbonmonoxy form. *Biochemistry* 52(34):5809–5820.
60. Berman HM, et al. (2000) The Protein Data Bank. *Nucleic Acids Res* 28(1):235–242.
61. Ishikura T, Hatano T, Yamato T (2012) Atomic stress tensor analysis of proteins. *Chem Phys Lett* 539–540:144–150.
62. Gullett PM, Horstemeyer MF, Baskes MJ, Fang H (2007) A deformation gradient tensor and strain tensors for atomistic simulations. *Model Simul Mater Sci Eng* 16(1):015001.
63. Liebschner D, Dauter M, Brzuszkiewicz A, Dauter Z (2013) On the reproducibility of protein crystal structures: five atomic resolution structures of trypsin. *Acta Crystallogr D Biol Crystallogr* 69(Pt 8):1447–1462.
64. Cock PJA, et al. (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11):1422–1423.
65. Hamelryck T, Manderick B (2003) PDB file parser and structure class implemented in Python. *Bioinformatics* 19(17):2308–2310.
66. Rice P, Longden I, Bleasby A (2000) EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 16(6):276–277.
67. Wang X, Snoeyink J (2006) Multiple structure alignment by optimal RMSD implies that the average structure is a consensus. *Comput Syst Bioinformatics Conf* 2006: 79–87.
68. Theiler J (1990) Estimating fractal dimension. *J Opt Soc Am A Opt Image Sci Vis* 7(6): 1055–1073.
69. Sanner MF, Olson AJ, Spehner JC (1996) Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* 38(3):305–320.