

# Evolutionary Conservation of Mechanical Strain Distributions in Functional Transitions of Protein Structures


Pablo Sartori\*

*Instituto Gulbenkian de Ciência, Oeiras, Portugal*

Stanislas Leibler

*Simons Center for Systems Biology, School of Natural Sciences, Institute for Advanced Study,  
Princeton, New Jersey, USA*

*and Laboratory of Living Matter, The Rockefeller University, New York, New York, USA*

 (Received 18 July 2023; revised 10 January 2024; accepted 29 January 2024; published 8 March 2024)

One of the tenets of molecular biology is that dynamic transitions between three-dimensional structures determine the function of proteins. Therefore, it seems only natural that evolutionary analysis of proteins, presently based mainly on their primary sequence, needs to shift its focus toward their function as assessed by corresponding structural transitions. This can be facilitated by recent progress in cryogenic electron microscopy that provides atomic structures of multiple conformational states for proteins and protein assemblies isolated from evolutionarily related species. In this work, we study evolutionary conservation of multiprotein assembly function by using mechanical strain as a quantitative footprint of structural transitions. We adopt the formalism of finite strain theory, developed in condensed matter physics, and apply it, as a case study, to a classical multiprotein assembly, the ATP synthase. Protein strain analysis provides a precise characterization of rotation domains that agrees with the present biophysical knowledge. In addition, we obtain a strain distribution on the protein structure associated with functional transitions. By analyzing in detail the strain patterns of the chains responsible for ATP synthesis across distinct species, we show that they are evolutionarily conserved for the same functional transition. Such conservation is not revealed by displacement or rotation patterns. Furthermore, within each functional transition, we can identify conserved strain patterns for ATP synthases isolated from different organisms. The observed strain conservation across evolutionary distant species indicates that strain should be essential in future structure-based evolutionary studies of protein function.

DOI: [10.1103/PhysRevX.14.011042](https://doi.org/10.1103/PhysRevX.14.011042)

Subject Areas: Biological Physics

## I. INTRODUCTION

Evolution acts by natural selection of multifaceted features of living systems, phenotypic traits, that are encoded in their DNA sequence, genotype. At the sub-cellular scale, many phenotypic traits can be associated to the functioning of proteins and multiprotein assemblies [1]. Most approaches for evolutionary comparison of proteins focus on their DNA sequence [2]. While this sheds light into the genotypic space, taking phenotypic traits into account requires comparing proteins as they perform their multiple cellular functions. To study evolution of protein phenotypes, the focus must, therefore, shift toward the

evolutionary changes in the different structural transitions underlying protein functions.

Evolutionary comparison of protein structures encounters two fundamental hurdles [3]. First, despite recent efforts [4], spatial alignment remains an arbitrary element of structural comparison. Second, it is not a static structure that defines the function of a protein, but rather dynamical transitions between multiple structural states. Mechanical strain is an alignment-independent quantity that has recently been used to study structural transitions [5,6], and has therefore the potential to overcome both these barriers. The usage of strain as a probe for protein function is justified because on functional timescales proteins behave largely as elastic materials [7,8], which explains why elastic paradigms have been so successful in protein biophysics during the past decades [9]. However, despite recent applications to diverse protein structures [10–12], the relationship between strain patterns and rotation or displacement descriptions traditionally associated to protein function remains unaddressed. Here, we establish this

\*Corresponding author: [psartori@igc.gulbenkian.pt](mailto:psartori@igc.gulbenkian.pt)

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

relationship, and show that distinct strain patterns correspond to distinct functions performed by one same protein. This allows us to address the fundamental question: Are structural transitions of proteins corresponding to different functions evolutionary conserved? In other words, beyond their sequence similarity, can we assess evolutionary conservation of protein function using mechanical strain?

## II. PROTEIN STRAIN ANALYSIS (PSA) OF ATP SYNTHASE

We adapted the complete formalism of finite strain theory, which belongs to a long-established branch of physics, elasticity theory [13,14], to proteins. Protein strain analysis (PSA; see Appendix A for a detailed method summary) computes local quantities (defined for a neighborhood of each residue taking  $C_\alpha$  as reference) analogous to those of finite strain theory. Examples of such quantities are the principal stretches, which measure the magnitude of strain, or the local rotation angle, which measures structural changes that do not strain the structure. As a case study we chose the ATP synthase. This was motivated by its important functional conservation concomitant with evolutionary variability [15], abundant modeling of its dynamics [16–21], and the availability of multiple structures from evolutionary distant species resolved by cryogenic electron microscopy [22–30]. Furthermore, the well-established connection between ATP synthase function and rotations is an ideal test ground for PSA.

The  $F_0F_1$  assembly, as it is also called, synthesizes ATP driven by a membrane proton gradient [22,31,32] (see Appendix B for detailed schematics). Through its functional cycle the membrane-submerged  $F_0$  complex, composed of a  $c$  ring together with its central “stalk”  $\gamma$ , adopts three distinct rotational states (I, II, and III) relative to the rest of the assembly [33,34]. Proton driven rotation of  $F_0$  results in interactions with the  $F_1$  complex, with approximate threefold symmetry. The  $F_1$  complex includes three distinct  $\beta$  chains  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ , participating in ATP synthesis. In any of the three rotary states, each of the three distinct  $\beta$  chains occupies one of three distinct enzymatic (and conformational) states, which we call here E (for the “empty” APO state), D (for “ADP bound”), and T (for “ATP synthesis”). As the whole assembly performs its rotary cycle  $I \rightarrow II \rightarrow III \rightarrow I \rightarrow \dots$ , each of the three  $\beta$  chains undergoes the enzymatic cycle  $E \rightarrow D \rightarrow T \rightarrow E \rightarrow \dots$  [35,36]. Each of the three transitions of  $\beta$  correspond to a different function: ADP binding ( $E \rightarrow D$ ), ATP synthesis ( $D \rightarrow T$ ), and ATP release ( $T \rightarrow E$ ). Therefore, the ATP synthase can intuitively be described as a rotary engine, in which rotation of  $F_0$  is coupled to conformational transitions of  $\beta$  chains in  $F_1$  that perform specific functions.

To provide a quantitative account of this description, we determined the rotation angle and axis for the assembly transition  $I \rightarrow II$  [see Sec. B of Supplemental Material (SM)

for a table of supporting files with information on other transitions, and note that we aligned structures using the channel chain  $a$  [37]. Figure 1(a) shows a scatter plot of the local (per residue  $i$ ) rotation angle  $\theta(\text{res}_i)$  against the  $z$  component of the rotation vector  $r_z(\text{res}_i)$ . We identify a cluster (cluster 2, in green) that corresponds to a rotation angle  $\theta \approx 100^\circ$ , and maps onto the  $F_0$  region of the protein; see Fig. 1(b). This is in agreement with the rotation of  $103^\circ$  reported in Ref. [38]. Note that the fact that the rotation angle is not  $120^\circ$  emphasizes the precession and the lack of perfect global symmetry of the structure. In addition, within the chain  $\beta_3$  that undergoes  $E \rightarrow D$  to bind ADP, we also identify two clusters [in orange and purple in Fig. 1(a)], which map into two compact domains of the structure, Fig. 1(b). These domains are separated at the height of the site of ATP synthesis. Therefore, PSA robustly quantifies local rotations, which are distributed over large compact domains (at multi-chain levels and subchain levels). This agrees with the intuitive understanding of how the ATP synthase functions, yet our analysis requires no *a priori* knowledge of the detailed geometrical nature of the structure.

While the PSA framework provides a quantitative and local characterization of rotations, it is unlikely that rotation angle and axis encode information that is connected to protein function. The key reason is that rotations depend on the spatial alignment of structures, which is arbitrary, and so the value of the local rotation angle does not quantify an intrinsic property of the protein’s deformation. Therefore, we used PSA to quantify strain, which is alignment independent, and thus expected to be closer to a quantification of function itself. Figure 1(c) depicts a histogram of the principal stretches,  $(\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)})$ . Since values close to 1 correspond to small deformations, the indication of large strains is to be found in the tails of the histograms. We find that few residues belong to regions that are significantly strained. For instance, in  $\beta_3$  only  $\sim 5\%$  of residues are stretched above 20%, that is  $\lambda^{(3)} > 1.2$ .

Figure 1(d) shows how strains are distributed throughout the structure, with darker red tones denoting higher strain. There are three regions that show significant strain: (i) the proton channel at the interface of membrane protein  $a$  and the  $c$  components of  $F_0$  (see Fig. 5 for schematics); (ii) the central and upper region of the peripheral stalk  $bb'$ , which has been suggested to be elastically compliant [24,38,40,41]; (iii) the central part of  $\beta_3$ , near  $\gamma$ , which is where synthesis of ATP takes place. The geometry of high-strain regions is very different from that of rotation domains. Whereas rotation domains span large, compact regions, strain accumulates in smaller, disperse functional regions. While some such “strain pockets” are localized at the interface between rotation domains (see below), we emphasize that strain and rotation are not directly correlated, see also Fig. 9 in the SM [37], as strain characterizes deformations by explicitly excluding rotational information (see Appendix A).

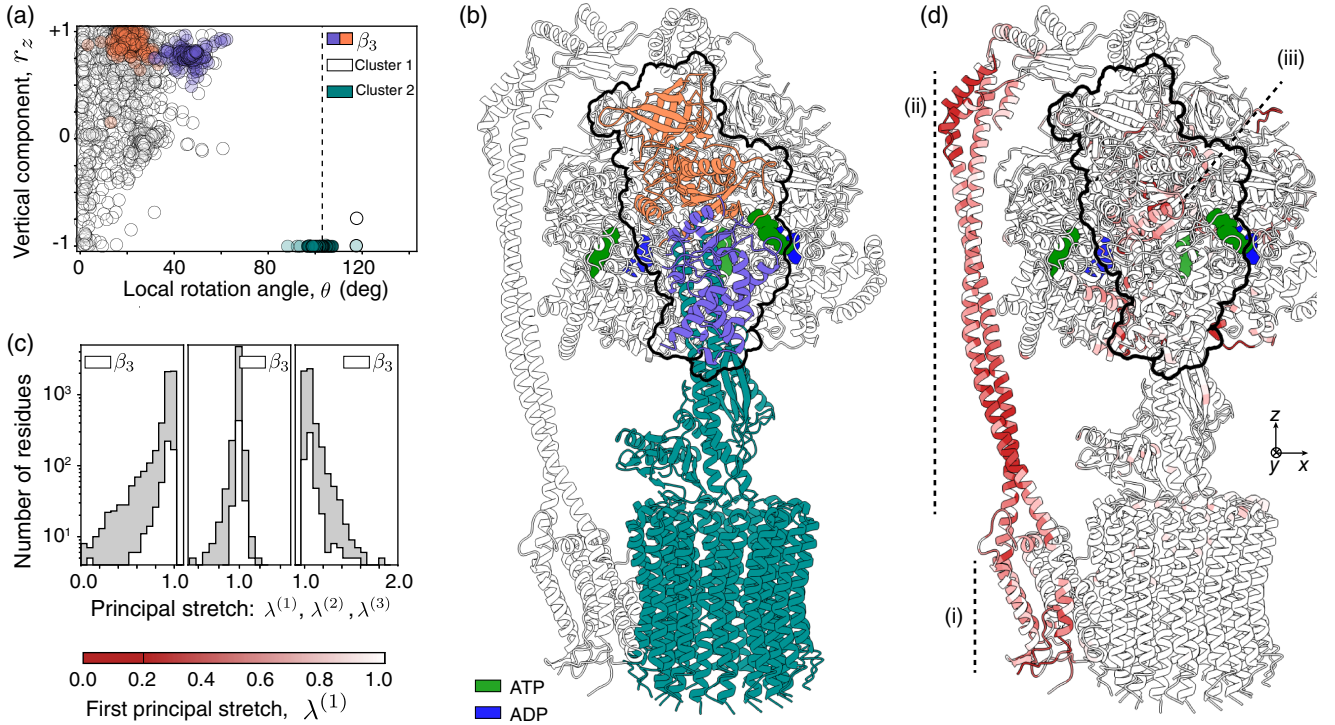


FIG. 1. Rotation and strain analysis of ATP synthase structural transition. (a) Scatter plot of rotation angle,  $\theta$ , against the vertical component of the rotation axis,  $r_z$ , for each residue of the structure. Each color labels a cluster obtained using the algorithm in Ref. [39] with parameter  $\rho = 0.1$ . The dark green cluster corresponds to  $\theta \approx 100^\circ$  and agrees with the rotation of  $103^\circ$  reported in Ref. [38] for  $F_0$ . Amino acids of  $\beta_3$  can be separated in two further clusters, one purple and one orange, with distinct rotation angles. (b) Mapping of the rotation-space clusters on the spatial structure of the synthase. At the whole assembly level, rotations occur in large compact regions corresponding to  $F_0$  and  $F_1$ . Focusing on  $\beta_3$ , outlined, the two rotation-space clusters map into compact regions of the chain. These regions are separated by an interface at the height where ATP is located. (c) Histogram of principal stretches  $\lambda^{(i)}$  (gray bars for the whole assembly, and white bars for  $\beta_3$ ). Large strain regions correspond with low values of  $\lambda^{(1)}$  and large values of  $\lambda^{(3)}$ . Note that  $\lambda^{(1)} \leq 1$ ,  $\lambda^{(2)} \approx 1$ , and  $\lambda^{(3)} \geq 1$ , which arises from near incompressibility; i.e.,  $\lambda^{(1)}\lambda^{(2)}\lambda^{(3)} \approx 1$ . (d) Mapping of stretch  $\lambda^{(1)}$  in ATP synthase structure using color scheme in (b). Residues of large strain reside in the proton channel (i), the middle of the peripheral stalk  $bb'$  (ii), and the area where  $\beta$  and  $\gamma$  interact (iii). ATP and ADP appear in green and blue, respectively.

### III. WITHIN-SPECIES CONSERVATION OF $\beta$ STRAIN FUNCTIONAL CYCLE

We focused our analysis on region (iii), i.e.,  $\beta_3$  and its interaction with  $\gamma$ . Figure 2(a) shows the distribution of strain found on  $\beta_3$  and  $\gamma$ , as the former undergoes the  $E \rightarrow D$  transition with the function of binding ADP. The tip of  $\gamma$  shows large strain, due to the large shearing that this small rotating region is subject to. In comparison, a large fraction of strained residues of  $\beta_3$  are localized at a 2D planar interface between the two rigid domains, identified in Figs. 1(a) and 1(b), that rotate relative to each other; see Figs. 2(b) and 2(c). This supports the long-standing hypothesis of strain accumulation in  $\beta$  during  $E \rightarrow D$  [16]. The involved strained amino acids are found in regions that are distant in space (from the ADP binding pocket G175 to the peripheral K191) and in sequence (such as V440 or A331). Furthermore, we found that the Walker-A motif, a set of eight residues that is evolutionary conserved and functionally critical [42], belongs to region (iii) and is clearly overstressed during transitions  $E \rightarrow D$

and  $T \rightarrow E$ ; see Table I of the SM [37]. We see, therefore, that PSA detects mechanically strained regions of proteins that participate in *functionally* important transitions (see Sec. A of the SM for characterization of strain in the  $bb'$  stalk).

In a complete synthesis cycle, each  $\beta$  chain performs three distinct functions, related to corresponding structural transition, as summarized by the table in Fig. 2(d). While the sequence of the three chains is the same, they are different in their spatial location relative to the peripheral stalk. The stalk in fact breaks the rotational symmetry of  $F_1$ , resulting in a precession of  $F_1$  relative to  $F_0$  as the latter rotates. Given these asymmetries, we investigated how well the strain patterns corresponding to each functional transition are conserved across the three  $\beta$  chains. In Fig. 2(d) we show cross sections of  $F_1$  at the height of the ADP binding site. Highly strained residues are colored in red, and the rest of them in beige. The strain patterns are characteristic of each transition and are found to be well conserved across the three  $\beta$ 's. For example, comparing



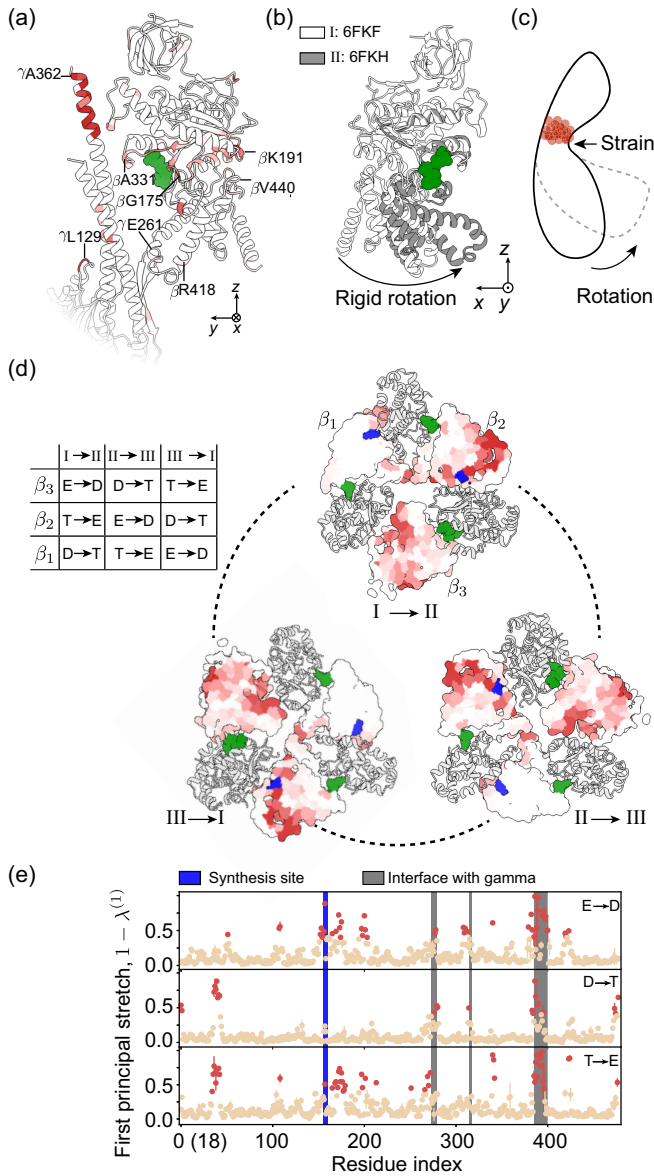


FIG. 2. Strain dynamics of  $\beta$  chains through synthesis cycle. (a) Distribution of strain in  $\beta_3$  and  $\gamma$  in the I  $\rightarrow$  II transition. Amino acids in  $\beta$  with large strain are located at the interface between the two rotation domains identified in Fig. 1(b), near the binding region of ADP. (b) Superposition of  $\beta$  in the reference (I, white) and deformed (II, gray) states. The lower and upper halves of the chain rotate relative to each other as rigid domains. (c) Schematic representation of the deformation undergone by  $\beta$  during E  $\rightarrow$  D. (d) Horizontal cross section of  $F_1$  showing strain distribution [color scheme as Fig. 1(d), with saturation at  $\lambda^{(1)} < 0.4$ ]. High strain accumulates in the plane of ATP synthesis in chains undergoing E  $\rightarrow$  D and T  $\rightarrow$  E, and in the basal bridge between chains undergoing T  $\rightarrow$  E and D  $\rightarrow$  T. The table contains the conformational changes undergone by the three  $\beta$  chains through a complete cycle of ATP synthesis. (e) Stretch profiles computed via  $\lambda^{(1)}$ . Points on each panel depict the mean for the three  $\beta$  when they undergo the same transition, with error bars mostly smaller than the point size. Note that profiles characterize each enzymatic transition. The (18) on the x axis denotes labeling of first residue in Ref. [38].

assembly states I  $\rightarrow$  II reveals that  $\beta_2$ , which performs the function of ATP release in the transition T  $\rightarrow$  E, has a strain pattern that forms a red loop. This red loop is replicated in  $\beta_1$  and  $\beta_3$  when they undergo T  $\rightarrow$  E. To quantify strain conservation we plotted the strain profiles, i.e., strain values along the sequence, for the three  $\beta$ 's when they are undergoing each transition, Fig. 2(e). The profiles are mostly flat, with peaks localized in small regions. The pattern of peaks is unique for each transition, acting as the mechanical fingerprint of the corresponding function. For example, there is a peak near the synthesis site when ADP is bound, E  $\rightarrow$  D, or ATP released, T  $\rightarrow$  E, but not in the synthesis step, D  $\rightarrow$  T (this may reflect large free-energy changes during nucleotide exchange, compared to moderate ones during bond formation). Furthermore, the strain profiles show a strong degree of conservation across chains for the same structural transition, as reflected by the small standard deviation across the three chains (depicted in the plot as error bars). We conclude that strain in  $\beta$  chains is localized in functionally relevant small regions distant from each other, with strain profiles that are characteristic of each enzymatic transition. Furthermore, despite overall precession of  $F_1$  and other structural asymmetries, these profiles show strong conservation across the three  $\beta$  chains when they perform the same function.

#### IV. STRAIN PATTERN ACROSS SPECIES

Having established strain as a functionally relevant property, robust to structural asymmetries, we now move to study its evolutionary conservation. To this end we compiled all published structures of ATP synthases, for which a full cycle has been resolved [23–28,38], and computed the distributions of strain (see Sec. B of the SM [37] for table of supporting files with comprehensive analysis of all structures). Figure 3(a) shows the spatial distribution of high strain for one third of a cycle on seven different ATP synthases. As one can see, these assemblies are structurally very different from one another, which in turn results in distinct distributions of strain. However, in all cases the region (iii), in which  $\gamma$  interacts with  $\beta$ , displays large strain, which parallels the functional conservation of ATP synthase function across species. We thus focused our attention on strain conservation in  $\beta$  chains.

As a first step, we confirmed that strain is conserved across  $\beta$ 's in the same species, spinach chloroplast, when they undergo the same transition, as suggested by Figs. 2(d) and 2(e). Figure 3(b), top panel, shows the fraction of strained residues that are conserved, as a function of the increasing number of residues considered to have high strain. Each curve represents a pair of chains and each color labels a structural transition or function [e.g., a single blue curve corresponds to  $(\beta_1, \beta_2)_{E \rightarrow D}$ , ADP binding]: in all cases the conserved fraction is very large. As a comparison to this strain conservation, we also calculated conservation of displacements among the same residues, depicted in the

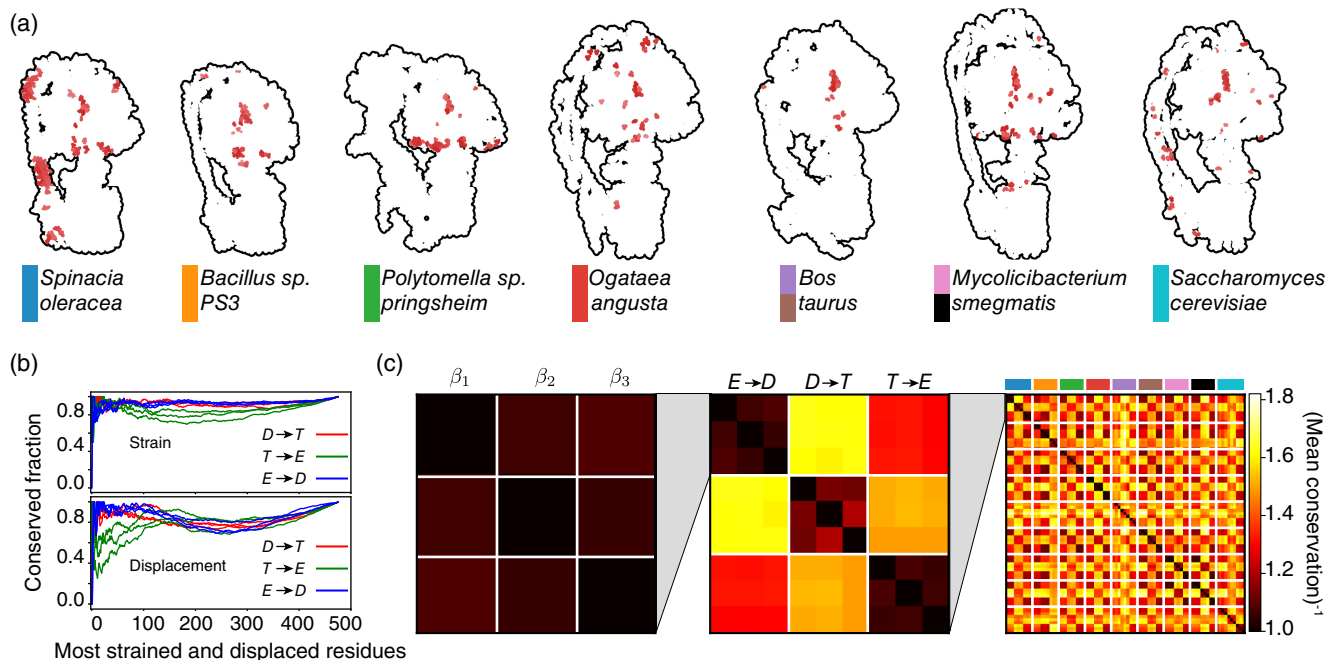


FIG. 3. Strain pattern across species. (a) Strain in ATP synthases of seven different species. Residues sustaining high strain ( $\lambda^{(1)} < 0.25$ ) are depicted in red on top of a silhouette of the corresponding synthase. In all cases high strain accumulates at the height of the  $\beta$ , where it interacts with  $\gamma$  [region (iii) lies approximately in the same location as in Fig. 1(d)]. Color bars are legends for subsequent panels (note that for *Bos taurus* and *Mycobacterium smegmatis* two sets of structures were considered; see Sec. B of the SM for identification of corresponding supporting file [37]). (b) Fraction of conserved residues between a pair of  $\beta$  (of the same species) plotted against the number of residues considered to be strained (obtained by decreasing the threshold separating low-strain from high-strain residues). For all enzymatic transitions, quantifying deformation via strain produces larger conservation than via displacements. (c) The inverse mean conservation [integral of curves in (b)] can be used to produce a distance matrix that compares different chains (first panel), undergoing different enzymatic transition (second panel), in different species (third panel). Note that darker colors correspond to higher conservation.

bottom panel. Although they are also significantly conserved, their level of conservation is lower.

Starting from this quantitative measure of strain conservation, we can take the mean conservation [area under each curve in the upper panel of Fig. 3(b)] as a single number that quantifies how conserved strain is between a pair of chains that undergo a particular enzymatic transition. Figure 3(c) shows matrices of mean conservation at three different levels: first panel, among the three different chains of the same species that undergo the same transition [this corresponds to Fig. 3(b)]; second panel, among the different chains when they undergo different transitions in the same species; and third panel, among different chains, undergoing different transitions, in different species. Therefore, strain conservation provides a systematic way of comparing different functional transitions of different proteins across different species.

## V. EVOLUTIONARY CONSERVATION OF STRAIN

Figure 3(c) quantitatively describes the evolutionary strain conservation in  $\beta$  chains participating in ATP synthesis, within seven evolutionarily distant multiprotein

assemblies. To elucidate further strain conservation from this multidimensional matrix, we used multidimensional scaling (MDS) to find its lower-dimensional representation. In brief, MDS finds an embedding in low dimensions (two, in this case) such that the distances among points in this embedding are as close as possible to those in the input metric [here the rightmost matrix in Fig. 3(c)].

Figure 4(a) shows that, when using strain conservation described above as the metric for MDS, enzymatic transitions of different species cluster together: strain patterns that correspond to the same functional transition are closer to each other than those that correspond to the same chain, or those of the same species. Importantly, such clustering is *not* observed when using as a metric conservation of residue displacement, Fig. 4(b), nor conservation of rotation angle; see Fig. 10 in the SM [37] (in both cases structures were aligned using chain  $a$  and observables computed using exclusively  $C_\alpha$ ). Therefore, strain, and not displacements or rotations, is an evolutionarily conserved characteristic of functional transitions in  $\beta$  chains. It is interesting to note that there is one outlier in the clustering of Fig. 4(a), corresponding to the ATP synthase from *Bacillus sp. PS3* (orange). Indeed, in Ref. [23] it was argued that in this ATP synthase  $\beta$  adopts “open-open-close” conformations (as opposed to

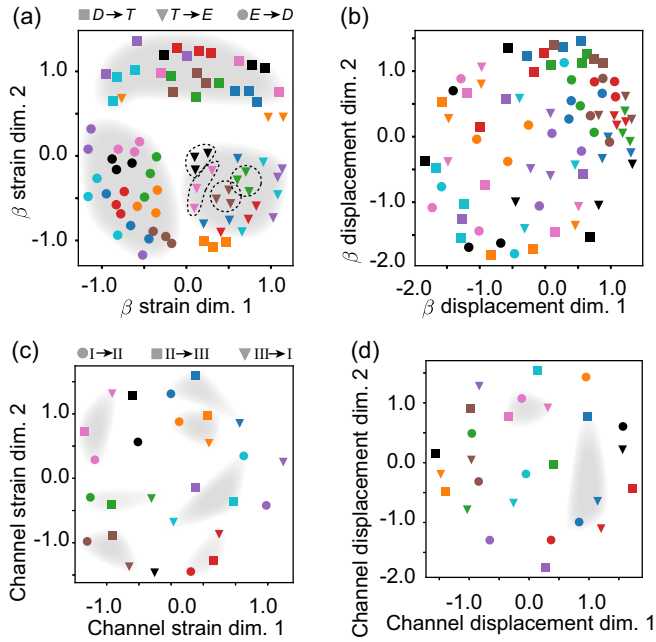


FIG. 4. Evolutionary conservation of strain patterns. (a),(b) Multidimensional scaling (MDS) of the dissimilarity matrix corresponding to conservation of strain (a) and displacement (b) in  $\beta$ . As is readily seen, with few outliers discussed in the text, strain conservation clusters the data according to the enzymatic transitions (gray shades are guides for the eye, and were hand drawn). This clustering is not observed when displacement conservation is used. Dashed contours highlight examples of within-species conservation, beyond the already mentioned functional conservation. (c),(d) Same as (a) and (b) but for the proton channel chain  $a$ . In this case, transitions of each species cluster together when using strain conservation, and they do not when using displacement conservation (as before, gray shades are guides for the eye). Colors correspond to legends in Fig. 3(a).

typical “close-open-close” ones) in order to perform the enzymatic transitions, which naturally results in misclassification of the strain patterns.

Remarkably, inside each functional cluster, we also observe a large degree of within-species clustering; see contours shapes in Fig. 4(a). As for the functional clustering, for this within-species clustering there seem to be some notable exceptions. Most notably, the *Bos taurus* dataset (in violet) is very spread out. It is likely that this may arise from a low resolution of the corresponding cryo-EM structure [26]. This second level of clustering indicates that species-specific adaptations are indeed present for each component of the cycle but become apparent only once the dominant functional or structural effect has been first taken into account.

The analysis of strain conservation can be extended to other relevant regions of the ATP synthase. For instance, an analogous MDS embedding performed on the proton channel protein  $a$  shows that all transitions of the same species can be better clustered together while using strain conservation, Fig. 4(c), than by using displacement

conservation as a metric, Fig. 4(d). Of the two outliers seen in Fig. 4(c), one corresponds again to low resolution *Bos taurus* structures. The other outlier, depicted in black, corresponds to the structures of *Mycobacterium smegmatis*, obtained in the presence of the inhibitory drug bedaquiline (BDQ) [29]. BDQ binds to the  $c$  ring and is known to affect the interaction between the  $c$  ring and the channel protein  $a$ . In agreement with this, we found that pockets of strain characterizing the deformations of the  $a$  protein in the absence of BDQ were suppressed in the presence of BDQ; see Fig. 11 in the SM [37]. Once again, as in the case of  $\beta$  chains in *Bacillus sp. PS3*, strain-based quantitative analysis captures subtle differences in structural transitions in an unbiased manner. Overall, the analysis of the proton channel further supports that species-specific adaptations are revealed when using a function-related metric, as we have shown strain is.

In conclusion, strain profiles can be used for functional evolutionary analysis as they quantify conservation properties (as well as point to subtle differences) better than measures based on residues displacements.

## VI. DISCUSSION

Evolution of proteins is associated with changes in their encoded sequence, translating directly into changes of their three-dimensional structure. In turn, these changes in structure, together with the changes taking place for all cellular partners, modify the structural transitions that a protein can undergo. The set of the allowed functional transitions in structure can be viewed as the protein’s “molecular phenotype,” which encompasses, for instance, its allosteric or enzymatic qualities. Therefore, quantification of functional transitions in protein structures is crucial to provide a quantitative assessment of protein genotype-phenotype maps.

The recent explosion of cryo-EM data has resulted in a surge of methods aimed at interpreting ensembles of electron microscopy images [43,44] and reconstructing protein atomic maps [45–47]. In contrast, quantitative methods for the comparative analysis of such maps are still underdeveloped. Most existing approaches come from the molecular dynamics literature [48,49], and require large sets of simulation-generated atomic maps to perform low-dimensional embeddings. This requirement renders them inadequate for the handful of atomic maps that are experimentally accessible. Instead, the closest approach to protein strain analysis (PSA) is normal mode analysis (NMA) [50,51], which characterizes a structure by the normal modes of its elastic network model. However, while NMA characterizes fluctuations of a single structure by displacement fields, PSA characterizes potentially large structural transitions between a pair of structures by strain fields, which are alignment independent. Therefore, PSA presents a new paradigm to compare pairs of atomic maps grounded on fundamental physical principles.



We have shown that strain shows a large extent of evolutionary conservation for the same function. This implies that mechanical strain is not only an important functional, but also *evolutionary* quantity. Because our analysis corresponds to structural changes of the same genotype (for each species), the concept of conservation we have discussed so far is different from that of sequence conservation. Nevertheless, given the relevance of sequence conservation in determining the structure of proteins [52], an important question is how strain relates to sequence conservation. To address this we performed multiple sequence alignment on the set of all  $\beta$  chains studied in this work, and quantified conservation by the Shannon entropy on the position-specific score matrix; see Fig. 12 in the SM [37]. Remarkably, we found no correlation between the Shannon entropy and the magnitude of strain. This suggests that strain patterns are not a direct consequence of sequence conservation, and instead reflect a convergent phenotypic trait in the light of genotypic variations.

In conclusion, strain emerges as an evolutionarily conserved quantity that characterizes well the nature of functional transitions in the ATP synthase. This should not be surprising: after all, proteins are (visco)elastic evolved materials, so it is natural that strain, introduced by elasticity theory, should be used to better understand their functioning and their evolution.

In order to facilitate the usage of PSA by the research community, we have created a PYTHON 3 library that contains a detailed implementation of PSA [53].

## ACKNOWLEDGMENTS

The authors would like to thank W. Kühlbrandt, H. Guo, J. L. Rubinstein, J. Howard, and T. Tlusty for helpful discussions, as well as E. Kussell and V. H. Mello for a critical reading of the manuscript. The authors would also like to thank V. H. Mello for support in the design of Box 2 in the Supplemental Material. This work was partially funded by a laCaixa (LCF/BQ/PI21/11830032) grant to P. S. Early stages of this research were partially supported by grants from the Simons Foundation to S. L. through the Rockefeller University (Grant No. 345430) and the Institute for Advanced Study (Grant No. 345801).

## APPENDIX A: PROTEIN STRAIN ANALYSIS IN A NUTSHELL

We now summarize the key aspects of finite strain theory (FST) [9,13], and then elaborate on how to adapt FST for the analysis of proteins, which we refer to as protein strain analysis (PSA). The main aim of the FST formalism, summarized in Appendix C, is constructing the (Lagrangian) strain tensor  $\gamma_{ij}$  at every point of a material. This mathematical object quantifies the difference between

measuring distances in a reference and a deformed structure. In other words,  $\gamma_{ij} = \frac{1}{2}(g_{ij}^{\text{ref}} - g_{ij}^{\text{def}})$ , where the metric  $g_{ij}$  quantifies how distances in a material measured in the direction of coordinate  $i$  change along coordinate  $j$  [54]. As distances are coordinate-independent properties, so is the strain tensor. In addition, FST also provides information on the local rotation angle  $\theta$  and rotation axis  $\mathbf{r}$ , which are, however, alignment dependent. We remark that the strain tensor  $\boldsymbol{\gamma}$  does not depend on an elasticity model. In contrast, to discuss stress a strain-stress relation needs to be defined. For hyperelastic solids this can be done from a particular elastic energy density (e.g., Saint-Venant, neo-Hookean, etc.).

In this work, we characterize the strain tensor by the principal axes of deformation  $\mathbf{v}^{(i)}$  (with  $i = 1, 2, 3$ ). The corresponding principal stretches along these three axes,  $\lambda^{(i)}$ , quantify the amount of deformation (note that  $\lambda^{(i)} = 1$  corresponds to lack of deformation). Importantly, the principal stretches are invariant under coordinate transformations. A local deformation can be thought of as transforming a small sphere of material into a small ellipsoid:  $\mathbf{v}^{(i)}$  correspond then to the direction of the ellipsoid axes and  $\lambda^{(i)}$  to their relative stretch (Fig. 6). We emphasize that strain is a rate of change in distances, and not rotations. While rotations in a material can indirectly and nonlocally affect strain, there is no *a priori* direct correlation between the two (see Fig. 9 in the SM [37]).

In PSA, we apply FST to protein structural changes. To this end, we first identify the positions of all the carbon atoms  $C_\alpha$  in the reference state  $\mathbf{X}$  and deformed state  $\mathbf{x}$ . This information is obtained from atomic maps, readily available in the form of pdb or cif files. The next step is to calculate the deformation gradient  $\mathbf{F}$ , which involves derivatives of  $\mathbf{x}$  with respect to  $\mathbf{X}$ . Because we are attempting to compute  $\mathbf{F}$ , which is a continuum concept, in a discrete material, an approximation method is needed. Different such approaches have been developed in the literature of discrete materials [55,56] (see Appendix D for an overview of three such methods). In this paper, we use the method of Ref. [55] that estimates  $\mathbf{F}$  at a given atomic position by considering all neighboring atoms within a sphere of radius  $r$ . Once the deformation gradient field is calculated, deducing the rotation and strain features can be done as described in Appendix C.

## APPENDIX B: SCHEMATICS OF THE ATP SYNTHASE STRUCTURE

The ATP synthase protein assembly consists of two motor complexes,  $F_0$  and  $F_1$ , that oppose each other; see Fig. 5. The  $F_0$  complex is constituted by the  $c$  ring (immersed in the membrane) and the central “stalk”  $\gamma$  and  $\epsilon$ .  $F_0$  rotates driven by a controlled flux of protons

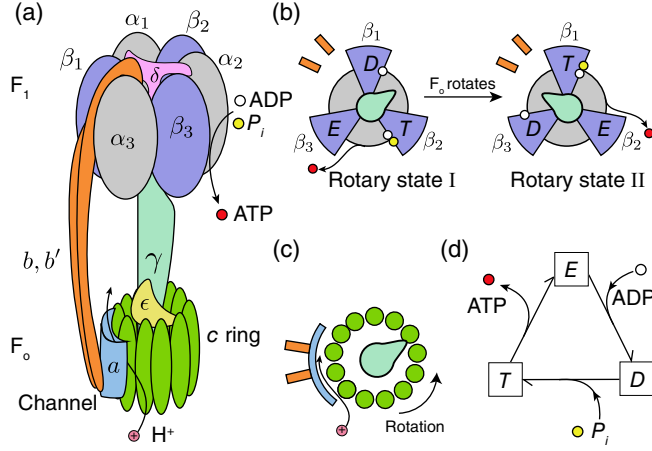


FIG. 5. Schematics of the structure and functioning of ATP synthase. (a) Overview of the ATP synthase protein assembly highlighting main components. (b) Schematic view from the top of the  $F_1$  complex as  $F_0$  undergoes a  $120^\circ$  rotation. Each of the three  $\beta$ , blue circular sections, adopt a distinct enzymatic state: empty (E), ADP bound (D), and ready for the synthesis of ATP (T). Orange bars represent  $bb'$  and gray circular sections  $\alpha$ . (c) Schematic view from the top of the  $F_0$  complex, indicating rotation direction. (d) Enzymatic diagram of the synthesis cycle at an individual  $\beta$ . Colors of parts are consistent across panels.

through the channel between the  $c$  ring and  $a$ , which acts as a stator. The  $F_1$  complex is prevented from rotating by the stalk  $bb'$  with which it interacts via  $\delta$ . As a consequence, rotation of  $\gamma$  triggers sequential conformational changes of the three  $\beta$ , which participate in the synthesis of ATP from ADP and inorganic phosphate  $P_i$ .

One important structural feature of the ATP synthase assembly is that the number of protein monomers in the  $c$  ring is not always a multiple of 3, and so there is a stoichiometric mismatch between  $F_0$  and  $F_1$ . In addition, the rotation of  $\gamma$  is not homogeneous, but occurs through three discrete jumps. These three jumps do not span exactly  $120^\circ$  each, likely due to the aforementioned mismatch in stoichiometry.

### APPENDIX C: SUMMARY OF FINITE STRAIN THEORY

A body characterized by material Cartesian coordinates  $X_i$ , with  $i = 1, 2, 3$ , is deformed to  $\mathbf{x}(\mathbf{X})$ ; see Fig. 6(a). The deformation gradient  $F_{ij}(\mathbf{X})$  is then given by

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \mathbf{R}\mathbf{U}, \quad (\text{C1})$$

where in the second equation we used the polar decomposition to obtain the rotation matrix  $R_{ij}(\mathbf{X})$  and the right stretch tensor  $U_{ij}(\mathbf{X})$ . The definition of nonlinear

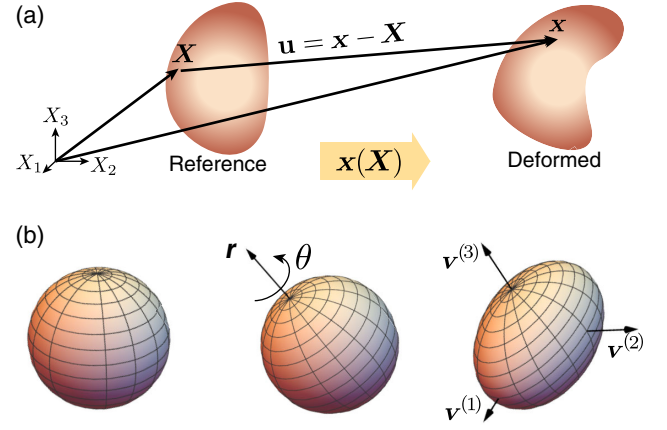


FIG. 6. Finite strain theory decomposition. (a) A continuous body with material coordinates  $\mathbf{X}$  in its reference state is described by coordinates  $\mathbf{x}(\mathbf{X}) = \mathbf{X} + \mathbf{u}$  in a deformed state, with  $\mathbf{u}(\mathbf{X})$  the displacement field. (b) The deformation of an elementary sphere inside a material is described by the local rotation angle  $\theta$  around the axis  $\mathbf{r}$  and the principal axes of deformation  $\mathbf{v}^{(i)}$  with magnitudes given by the principal stretches  $\lambda^{(i)}$  and  $i = 1, 2, 3$ .

Lagrangian strain tensor  $\gamma_{ij}(\mathbf{X})$  follows directly:

$$\boldsymbol{\gamma} = \frac{1}{2}(\mathbf{U}^2 - \mathbf{I}) = \underbrace{\frac{1}{2}\left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} + \frac{\partial \mathbf{u}^\top}{\partial \mathbf{X}}\right)}_{\text{linear strain } \boldsymbol{\gamma}_{\text{lin}}} + \underbrace{\frac{1}{2}\frac{\partial \mathbf{u}^\top}{\partial \mathbf{X}}\frac{\partial \mathbf{u}}{\partial \mathbf{X}}}_{\text{nonlinear correction}}, \quad (\text{C2})$$

where we used the displacement field  $u_i = x_i - X_i$ ; see schematic in Fig. 6(a).

The local rotation angle  $\theta(\mathbf{X})$  and axis  $r_i(\mathbf{X})$  are calculated from the rotation matrix via

$$\cos(\theta) = \frac{1}{2}[\text{Tr}(\mathbf{R}) - 1] \quad \text{and} \quad \mathbf{R} \cdot \mathbf{r} = \mathbf{r}. \quad (\text{C3})$$

The principal stretches  $\lambda^{(i)}(\mathbf{X})$  and principal axes  $\mathbf{v}^{(i)}(\mathbf{X})$  of deformation are obtained from the eigenproblem:

$$\mathbf{U} \cdot \mathbf{v}^{(i)} = \lambda^{(i)} \mathbf{v}^{(i)}. \quad (\text{C4})$$

The  $\lambda^{(i)}$ , customarily ordered by increasing magnitude, are related to the eigenvalues of the Lagrangian strain tensor through  $\gamma^{(i)} = [(\lambda^{(i)})^2 - 1]/2$ .

Overall, finite strain theory allows us to decompose a deformation  $\mathbf{x}(\mathbf{X})$  into a global translation, local rotations (characterized by  $\mathbf{r}$  and  $\theta$ ), and local strain (characterized by  $\mathbf{v}^{(i)}$  and  $\lambda^{(i)}$ ); see Fig. 6(b). We remark that  $\mathbf{F}$ ,  $\mathbf{R}$ , and  $\mathbf{U}$  are invariant under rigid body translations of either reference or deformed structures.  $\mathbf{U}$  and  $\boldsymbol{\gamma}$  are additionally invariant under finite angle rotations, and  $\lambda^{(i)}$  are, furthermore, coordinate invariant.

### APPENDIX D: DEFORMATION GRADIENTS IN DISCRETE MATERIALS

Consider a protein that transitions from a reference state (with atomic positions  $\mathbf{X}$ ) to a deformed state (with atomic



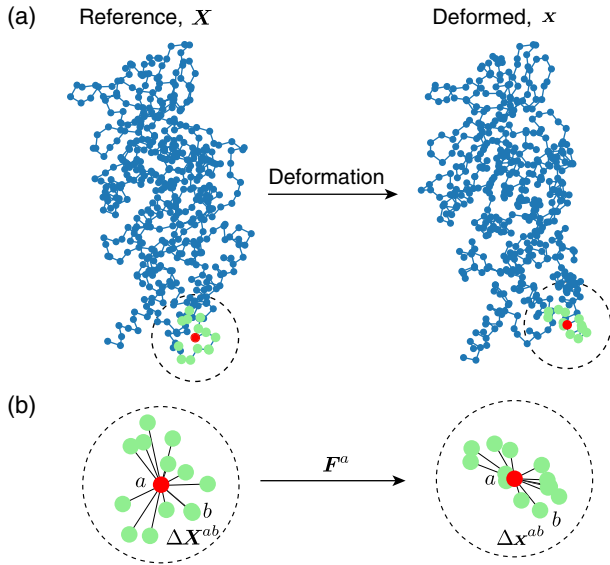


FIG. 7. Example of protein deformation. (a) Two-dimensional projection of the  $C_\alpha$  from chain  $\beta$  in two conformations, one taken as reference and the other as deformed. A particular atom is highlighted in red, and the surrounding ones within a sphere of radius 8 Å in light green. (b) The atom highlighted in (a), labeled  $a$ , is characterized by vectors  $\Delta \mathbf{X}^{ab}$  to the neighboring  $b$  atoms in the reference state, and  $\Delta \mathbf{x}^{ab}$  in the deformed state. The matrix  $\mathbf{F}^a$  estimates the map of one set of vectors to the other; see Eq. (D1).

positions  $\mathbf{x}$ ), for which we want to perform FST. To estimate at a given atom  $a$  the deformation gradient  $\mathbf{F}^a$ , we consider all atoms  $b$  within a sphere of radius  $r$  centered at  $a$ ; see Fig. 7.

For the reference conformation we define vectors connecting the target atom  $a$  with its neighbors  $b$ ,  $\Delta \mathbf{X}_{ab}$ , and analogously for the deformed state,  $\Delta \mathbf{x}^{ab}$ . A second order approximation of the deformation gradient satisfies

$$\Delta x_i^{ab} \approx \sum_j F_{ij}^a \Delta X_j^{ab} + \frac{1}{2} \sum_{jk} H_{ijk}^a \Delta X_j^{ab} \Delta X_k^{ab} \quad (\text{D1})$$

for each  $(ab)$  pair.

Different algorithms, which we now summarize, attempt at estimating  $\mathbf{F}_a$  from the expression above.

- (i) Minimal.  $\mathbf{H}^a$  is set to zero, which keeps the approximation linear; see Ref. [55]. Furthermore, only the three  $b$  neighbors closest to  $a$  are considered. Since on the previous equation each neighbor adds three conditions on the components of  $\mathbf{F}^a$ , which has nine components, this is sufficient to determine a unique  $\mathbf{F}^a$ .
- (ii) First order. Still  $\mathbf{H}^a$  is set to zero, but all  $b$  neighbors are kept, which leads to an overdetermined system for  $\mathbf{F}^a$  when there are more than 3 neighbors. An error function is defined  $\phi(\mathbf{F}^a) = \sum_b \|\Delta \mathbf{x}^{ab} - \mathbf{F}^a \cdot \Delta \mathbf{X}^{ab}\|^2$ . The minimum of this function is taken as an estimator for  $\mathbf{F}^a$ ; see Ref. [55].

- (iii) Second order. Analogous to the previous, but the error function takes into account the second order corrections. The system is overdetermined for more than 12 neighbors, and the computational cost is substantially larger. The gain is a discrete strain tensor that satisfies compatibility conditions [56].

- 
- [1] B. Alberts, *The cell as a collection of protein machines: Preparing the next generation of molecular biologists*, *Cell* **92**, 291 (1998).
  - [2] C. Pál, B. Papp, and M. J. Lercher, *An integrated view of protein evolution*, *Nat. Rev. Genet.* **7**, 337 (2006).
  - [3] L. Holm and C. Sander, *Mapping the protein universe*, *Science* **273**, 595 (1996).
  - [4] C. Zhang, M. Shine, A. M. Pyle, and Y. Zhang, *US-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes*, *Nat. Methods* **19**, 1109 (2022).
  - [5] T. Yamato, *Strain tensor field in proteins*, *J. Mol. Graphics* **14**, 105 (1996).
  - [6] M. R. Mitchell, T. Tlustý, and S. Leibler, *Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5847 (2016).
  - [7] Y. Wang and G. Zocchi, *The folded protein as a viscoelastic solid*, *Europhys. Lett.* **96**, 18003 (2011).
  - [8] H. Qu and G. Zocchi, *How enzymes work: A look through the perspective of molecular viscoelastic properties*, *Phys. Rev. X* **3**, 011009 (2013).
  - [9] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton* (Sinauer Press, Sunderland, 2001).
  - [10] A. Yu, E. M. Lee, J. A. Briggs, B. K. Ganser-Pornillos, O. Pornillos, and G. A. Voth, *Strain and rupture of HIV-1 capsids during uncoating*, *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2117781119 (2022).
  - [11] J. Lara, L. Diacovich, F. Trajtenberg, N. Larrieux, E. L. Malchiodi, M. M. Fernández, G. Gago, H. Gramajo, and A. Buschiazzo, *Mycobacterium tuberculosis FasR senses long fatty acyl-CoA through a tunnel and a hydrophobic transmission spine*, *Nat. Commun.* **11**, 3703 (2020).
  - [12] M. J. Reynolds, C. Hachicho, A. G. Carl, R. Gong, and G. M. Alushin, *Bending forces and nucleotide state jointly regulate F-actin structure*, *Nature (London)* **611**, 380 (2022).
  - [13] A. J. M. Spencer, *Continuum Mechanics* (Courier Corporation, London, 2004).
  - [14] P. G. Ciarlet, *Mathematical Elasticity: Three-Dimensional Elasticity* (SIAM, Philadelphia, 2021).
  - [15] D. Stock, A. G. Leslie, and J. E. Walker, *Molecular architecture of the rotary motor in ATP synthase*, *Science* **286**, 1700 (1999).
  - [16] G. Oster and H. Wang, *Reverse engineering a protein: The mechanochemistry of ATP synthase*, *Biochim. Biophys. Acta Bioenerg.* **1458**, 482 (2000).
  - [17] E. Lathouwers, J. N. Lucero, and D. A. Sivak, *Nonequilibrium energy transduction in stochastic strongly coupled rotary motors*, *J. Phys. Chem. Lett.* **11**, 5273 (2020).

- [18] K.-i. Okazaki and G. Hummer, *Elasticity, friction, and pathway of  $\gamma$ -subunit rotation in  $F_0F_1$ -ATP synthase*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10720 (2015).
- [19] S. Kubo, T. Niina, and S. Takada, *Molecular dynamics simulation of proton-transfer coupled rotations in ATP synthase  $F_0$  motor*, *Sci. Rep.* **10**, 8225 (2020).
- [20] J. Czub, M. Wieczor, B. Prokopowicz, and H. Grubmüller, *Mechanochemical energy transduction during the main rotary step in the synthesis cycle of  $F_1$ -ATPase*, *J. Am. Chem. Soc.* **139**, 4025 (2017).
- [21] B. I. Costescu and F. Gräter, *Time-resolved force distribution analysis*, *BMC Biophys.* **6**, 5 (2013).
- [22] W. Kühlbrandt, *Structure and mechanisms of F-type ATP synthases*, *Annu. Rev. Biochem.* **88**, 519 (2019).
- [23] H. Guo, T. Suzuki, and J. L. Rubinstein, *Structure of a bacterial ATP synthase*, *eLife* **8**, e43128 (2019).
- [24] B. J. Murphy, N. Klusch, J. Langer, D. J. Mills, Ö. Yildiz, and W. Kühlbrandt, *Rotary substates of mitochondrial ATP synthase reveal the basis of flexible  $F_1$ - $F_0$  coupling*, *Science* **364**, eaaw9128 (2019).
- [25] K. R. Vinothkumar, M. G. Montgomery, S. Liu, and J. E. Walker, *Structure of the mitochondrial ATP synthase from *Pichia angusta* determined by electron cryo-microscopy*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12709 (2016).
- [26] A. Zhou, A. Rohou, D. G. Schep, J. V. Bason, M. G. Montgomery, J. E. Walker, N. Grigorieff, and J. L. Rubinstein, *Structure and conformational states of the bovine mitochondrial ATP synthase by cryo-EM*, *eLife* **4**, e10180 (2015).
- [27] T. E. Spikes, M. G. Montgomery, and J. E. Walker, *Structure of the dimeric ATP synthase from bovine mitochondria*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 23519 (2020).
- [28] M. Sobti, J. L. Walshe, D. Wu, R. Ishmukhametov, Y. C. Zeng, C. V. Robinson, R. M. Berry, and A. G. Stewart, *Cryo-EM structures provide insight into how *E. coli*  $F_1F_0$  ATP synthase accommodates symmetry mismatch*, *Nat. Commun.* **11**, 2615 (2020).
- [29] H. Guo, G. M. Courbon, S. A. Bueler, J. Mai, J. Liu, and J. L. Rubinstein, *Structure of mycobacterial ATP synthase bound to the tuberculosis drug bedaquiline*, *Nature (London)* **589**, 143 (2021).
- [30] H. Guo and J. L. Rubinstein, *Structure of ATP synthase under strain during catalysis*, *Nat. Commun.* **13**, 2232 (2022).
- [31] P. Mitchell, *Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism*, *Nature (London)* **191**, 144 (1961).
- [32] P. D. Boyer, *The ATP synthase—A splendid molecular machine*, *Annu. Rev. Biochem.* **66**, 717 (1997).
- [33] H. Noji, R. Yasuda, M. Yoshida, and K. Kinoshita, *Direct observation of the rotation of  $F_1$ -ATPase*, *Nature (London)* **386**, 299 (1997).
- [34] R. Yasuda, H. Noji, K. Kinoshita, Jr, and M. Yoshida,  *$F_1$ -ATPase is a highly efficient molecular motor that rotates with discrete 120 steps*, *Cell* **93**, 1117 (1998).
- [35] P. D. Boyer, R. L. Cross, and W. Momsen, *A new concept for energy coupling in oxidative phosphorylation based on a molecular explanation of the oxygen exchange reactions*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2837 (1973).
- [36] J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker, *Structure at 2.8 Å resolution of  $F_1$ -ATPase from bovine heart mitochondria*, *Nature (London)* **370**, 621 (1994).
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevX.14.011042> for figures containing extended analysis on protein deformation, as well as analysis on the relationship between sequence conservation and strain. Also includes PYTHON notebooks with detailed analysis of all the structures that appear in this paper.
- [38] A. Hahn, J. Vonck, D. J. Mills, T. Meier, and W. Kühlbrandt, *Structure, mechanism, and regulation of the chloroplast ATP synthase*, *Science* **360**, eaat4318 (2018).
- [39] A. Rodriguez and A. Laio, *Clustering by fast search and find of density peaks*, *Science* **344**, 1492 (2014).
- [40] M. Sobti, C. Smits, A. S. Wong, R. Ishmukhametov, D. Stock, S. Sandin, and A. G. Stewart, *Cryo-EM structures of the autoinhibited *E. coli* ATP synthase in three rotational states*, *eLife* **5**, e21598 (2016).
- [41] M. Sobti, R. Ishmukhametov, J. C. Bouwer, A. Ayer, C. Suarna, N. J. Smith, M. Christie, R. Stocker, T. M. Duncan, and A. G. Stewart, *Cryo-EM reveals distinct conformations of *E. coli* ATP synthase on exposure to ATP*, *eLife* **8**, e43864 (2019).
- [42] J. E. Walker, M. Saraste, M. J. Runswick, and N. J. Gay, *Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold*, *EMBO J.* **1**, 945 (1982).
- [43] A. Dashti, P. Schwander, R. Langlois, R. Fung, W. Li, A. Hosseinzadeh, H. Y. Liao, J. Pallesen, G. Sharma, V. A. Stupina et al., *Trajectories of the ribosome as a Brownian nanomachine*, in *Single-Particle Cryo-Electron Microscopy: The Path Toward Atomic Resolution: Selected Papers of Joachim Frank with Commentaries* (World Scientific, Singapore, 2018), pp. 463–475.
- [44] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis, *CryoDRGN: Reconstruction of heterogeneous cryo-EM structures using neural networks*, *Nat. Methods* **18**, 176 (2021).
- [45] J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl, and S. H. Scheres, *New tools for automated high-resolution cryo-EM structure determination in RELION-3*, *eLife* **7**, e42166 (2018).
- [46] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, *cryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination*, *Nat. Methods* **14**, 290 (2017).
- [47] T. Grant, A. Rohou, and N. Grigorieff, *cisTEM, User-friendly software for single-particle image processing*, *eLife* **7**, e35383 (2018).
- [48] M. Ernst, F. Sittel, and G. Stock, *Contact- and distance-based principal component analysis of protein dynamics*, *J. Chem. Phys.* **143**, 244114 (2015).
- [49] F. Sittel and G. Stock, *Perspective: Identification of collective variables and metastable states of protein dynamics*, *J. Chem. Phys.* **149**, 150901 (2018).
- [50] D. A. Case, *Normal mode analysis of protein dynamics*, *Curr. Opin. Struct. Biol.* **4**, 285 (1994).
- [51] I. Bahar, A. R. Atilgan, and B. Erman, *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential*, *Folding Des.* **2**, 173 (1997).

- [52] F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J.N. Onuchic, T. Hwa, and M. Weigt, *Direct-coupling analysis of residue coevolution captures native contacts across many protein families*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1293 (2011).
- [53] <https://github.com/Sartori-Lab/PSA>.
- [54] E. Efrati, E. Sharon, and R. Kupferman, *The metric description of elasticity in residually stressed soft materials*, *Soft Matter* **9**, 8187 (2013).
- [55] P. Gullett, M. Horstemeyer, M. Baskes, and H. Fang, *A deformation gradient tensor and strain tensors for atomistic simulations*, *Model. Simul. Mater. Sci. Eng.* **16**, 015001 (2007).
- [56] J.A. Zimmerman, D.J. Bammann, and H. Gao, *Deformation gradients for continuum mechanical analysis of atomistic simulations*, *Int. J. Solids Struct.* **46**, 238 (2009).