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

(Dated: February 7, 2023)

One of the tenets of molecular biology is that dynamical transitions between three dimensional structures largely determine the function of individual proteins, as well as multi-protein assemblies. Therefore, it seems only natural that evolutionary analysis of proteins, presently based mainly on their primary sequence, needs to shift its focus towards their function as assessed by corresponding structural transitions. This can been facilitated by recent progress in cryogenic electron microscopy that provides atomic structures of multiple conformational states for multi-protein assemblies isolated from evolutionarily related species. In this work, we study evolutionary conservation of (multi)protein function by using mechanical strain as a quantitative footprint of structural transitions. We adopt the formalism of finite strain analysis, developed in condensed matter physics and engineering, and apply it, as a case study, to a classical multi-protein assembly, the ATP synthase. Our Protein Strain Analysis (PSA) 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 β chains responsible for ATP synthesis across different 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 are able to identify conserved strain patterns for ATP synthases isolated from different organisms. Our results promote strain as an important quantity close to protein function, which is independent of structural alignments and coordinate systems. The observed strain conservation across evolutionary distant species indicates that this quantity should be essential in future structure-based evolutionary studies of protein (and multi-protein assembly) function.

PACS numbers:

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 multi-protein 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 towards 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 to be 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 last 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 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 asses evolutionary conservation of protein function?

Protein Strain Analysis of ATP synthase. We adapted the complete formalism of finite strain analysis, which belongs to a long-established branch of physics, elasticity theory [15, 16], to proteins. Our Protein Strain Analysis (PSA, see *Supplementary Information* A for a detailed method summary) computes local quantities (defined for a neighborhood of each residue) analogous to those of finite strain theory. Examples of such quantities are the principal stretches, which measure the magni-

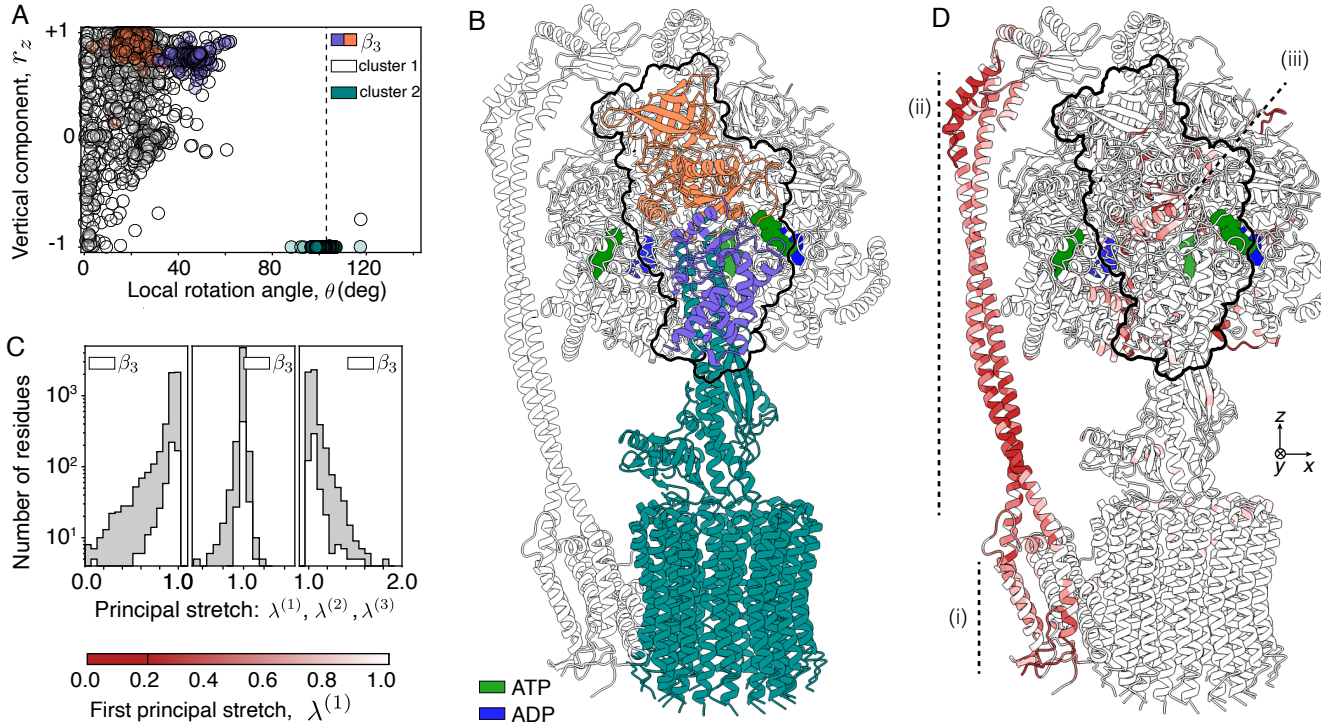


FIG. 1: *Rotation and strain analysis of ATP synthase structural transition.* **A.** Scatter plot of rotation angle θ against the vertical component of the rotation axis, r_z, for each residue of the structure. Each color labels a cluster obtained using the algorithm in [13] with parameter $\rho = 0.1$. The dark green cluster corresponds to $\theta \approx 100^o$ and agrees with the rotation of 103° reported in [14] for F_o. Amino acids of β_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_o and F_1. Focusing on β_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, $\lambda^{(i)}$. 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, the middle of the peripheral stalk bb', and the area where β and γ interact. ATP and ADP appear in green and blue, respectively.

tude 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 [17], and the recent availability of multiple structural conformations for several ATP synthases isolated from evolutionary distant organisms [18–26]. Furthermore, the well established connection between ATP synthase function and rotations is an ideal test-ground for PSA.

This protein assembly synthesizes ATP driven by a membrane proton gradient [18, 27, 28](see *SI* Fig. 5 for detailed schematics). Through its functional cycle the membrane-submerged F_o complex, composed of a c-ring together with its central "stalk" γ, adopts three disctinct rotational states (I, II and III) relative to the rest of the assembly [29, 30]. Proton driven rotation of F_o results in interactions with the F_1 complex, with approximate three-fold symmetry. The F_1 complex includes three distinct β chains β_1, β_2 and β_3, participating in ATP synthesis. In any of the three rotary states, each of the three distinct β 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→II→III→..., each of the three β chains undergoes the enzymatic cycle E→D→T→... [31, 32]. Each of the three transitions of β correspond to a different function: ADP binding (E→D), ATP synthesis (D→T), and ATP release (T→E). Therefore, the ATP synthase can intuitively be described as a rotary engine, in which rotation of F_o is coupled to conformational transitions of βs that perform specific functions.

To provide a quantitative account of this functional description, we determined the rotation angle and axis for the assembly transition I→II (see *SI* D for other transitions). Fig 1A 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 $\theta \approx 100^o$, and maps onto the F_o region of the protein, see Fig 1B. This is in agreement with the

rotation of 103° reported in [14]. Note that the fact that the rotation angle is not 120° emphasizes the precession and the lack of perfect global symmetry of the structure. In addition, within the chain β_3 that undergoes E→D to bind ADP, we also identify two clusters (in orange and purple, Fig 1A), which map into two compact domains of the structure, Fig 1B. 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 sub-chain 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.

Because rotations depend on the arbitrary spatial alignment of structures, they enclose information that is not connected to protein function. Therefore, we used PSA to quantify strain, which is alignment independent, and thus expected to be closer to the function itself. Fig. 1C 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 β_3 only $\sim 5\%$ of residues are stretched or compressed above 20%, that is $\lambda^{(1)} < 0.1$ or $\lambda^{(3)} > 1.2$. Figure 1D 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 *a* membrane protein and *c* components of F_o (see *SI* Fig. 5); (*ii*) the central and upper region of the peripheral stalk bb', which has been suggested to be elastically compliant [14, 20, 33, 34]; (*iii*) the central part of β_3, near γ, 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 that can be localized at the interface between rotation domains (see below).

Within species conservation of β strain functional cycle. We focused our analysis on region (*iii*), i.e. β_3 and its interaction with γ. Figure 2A shows the distribution of strain found on β_3 and γ, as the former undergoes the E→D transition with the function of binding ADP. The tip of γ shows large strain, due to the large rotation that this small region is subject to. In comparison, a large fraction of strained residues of β_3 are localized at a 2D planar interface between the two rigid domains, identified in Fig. 1A and B, that rotate relative to each other, see Fig. 2B and C. This supports the long-standing hypothesis of strain accumulation in β during E→D [35]. 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 also in sequence (such as V440 or A331). We see, therefore, that PSA detects mechanically strained regions of proteins that participate in *functionally* important tran-

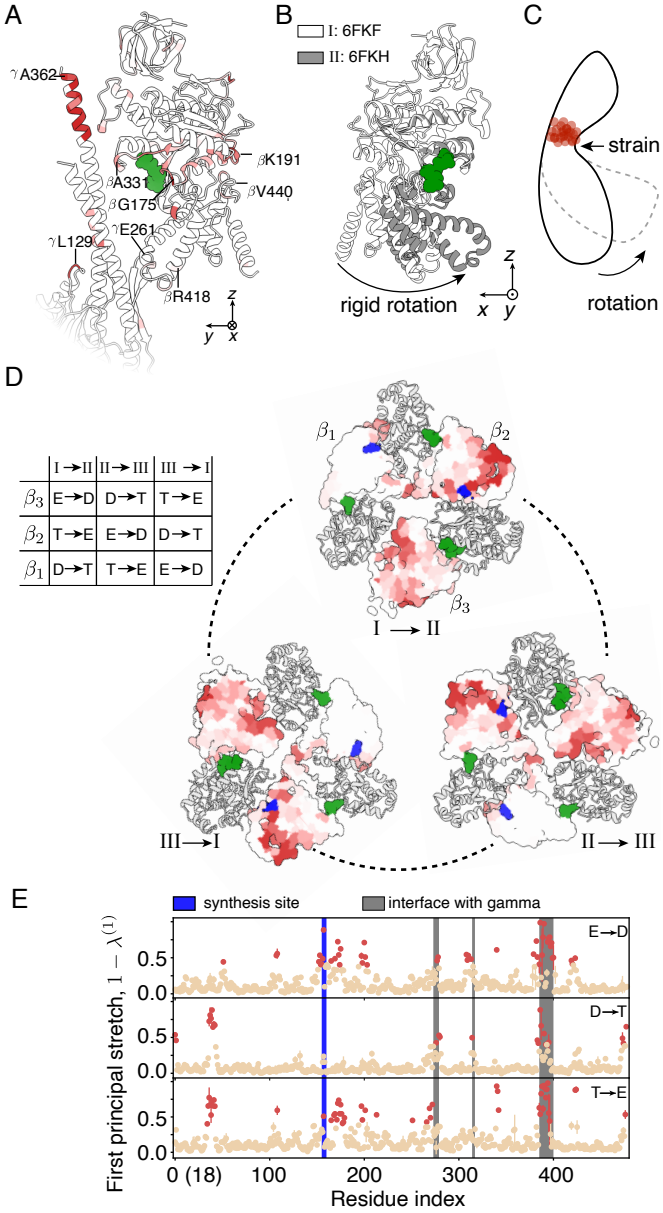


FIG. 2: *Strain dynamics of β chains through synthesis cycle.* **A.** Distribution of strain in β_3 and γ. Amino acids in β with large strain are located at the interface between the two rotation domains identified in Fig. 1B, near the binding region of ADP. **B.** Superposition of β in the reference (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 β during E→D. **D.** Horizontal cross-section of F_1 showing strain distribution (color scheme as Fig. 1D, with saturation at $\lambda^{(1)} < 0.4$). High strain accumulates in the plane of ATP synthesis in chains undergoing E→D and T→E, and in the basal bridge between chains undergoing T→E and D→T. The table contains the conformational changes undergone by the three β 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 β 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 labelling of first residue in [14].

sitions (see *SI* C for characterization of strain in the bb' stalk).

In a complete synthesis cycle, each β chain performs three distinct functions, related to corresponding structural transition, as summarized by the table in Fig. 2D. We investigated how well the strain patterns corresponding to each functional transition are conserved across the three β chains. In Fig. 2D 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 βs. For example, comparing assembly states I→II reveals that β_2, which performs the function of ATP release in the transition T→E, has a strain pattern that forms a red loop. This red loop is replicated in β_1 and β_3 when they undergo T→E. To quantify strain conservation we plotted the strain profiles, i.e. strain values along the sequence, for the three βs when they are undergoing each transition, Fig. 2E. The profiles are mostly flat, with peaks localized in small regions. We remark that the pattern of peaks is unique for each transitions, acting as the mechanical fingerprint of the corresponding function. 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 β 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 β chains when they perform the same function.

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 [14, 19–24], and computed the distributions of strain (see *SI* D for comprehensive analysis of all structures). Figure 3A 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 γ interacts with β, displays large strain, which parallels the functional conservation of ATP synthase function across species. We thus focused our attention on strain conservation in β chains.

As a first step, we confirmed that strain is conserved across βs in the same species, spinach chloroplast, when they undergo the same transition, as suggested by Fig. 2D and E. Figure 3B, 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 $(\beta_1, \beta_2)_{\mathrm{E}\to\mathrm{D}}$, 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 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 conserved fraction (area under each curve in the upper panel of Fig. 3B) as a single number that quantifies how conserved strain is between a pair of chains that undergo a particular enzymatic transition. Figure 3C shows matrices of mean conserved fraction at three different levels: first panel, among the three different chains of the same species that undergo the same transition (this corresponds to Fig. 3B); 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.

Evolutionary conservation of strain. Figure 3C quantitatively describes the evolutionary strain conservation in β chains participating in ATP synthesis, within seven evolutionarily distant multi-protein assemblies. To elucidate further strain conservation from this multi-dimensional matrix, we used multi-dimensional 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 right-most matrix in Fig. 3C).

Figure 4A 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 displacement conservation as a metric, Fig. 4B. Therefore, *strain, and not displacements, is an evolutionarily conserved characteristic of functional transitions in* β chains. It is interesting to note that there is one outlier in the clustering of 4A, corresponding to the ATP synthase from *Bacillus sp. PS3* (orange). Indeed, in [19] it was argued that in this ATP synthase β adopts "open-open-close" conformations (as opposed to 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 4A. 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 [22]. This second level of clustering indicates

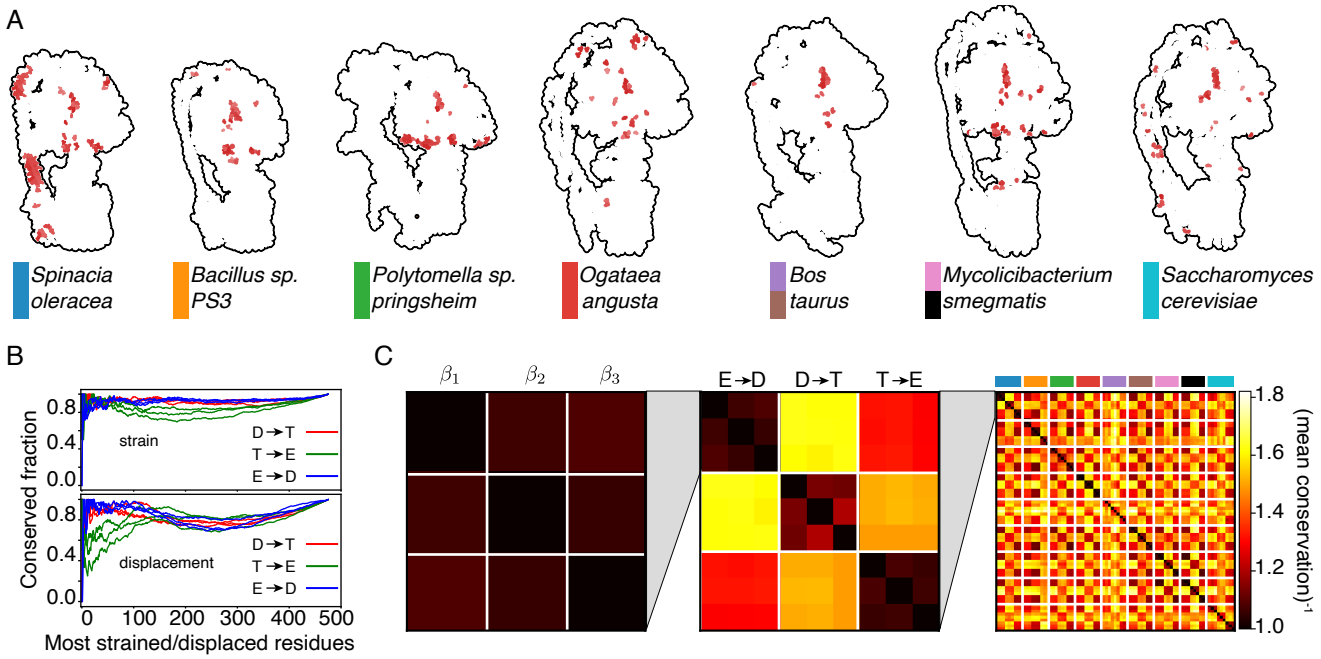


FIG. 3: *Strain pattern across species.* **A**. Strain in ATP synthases of seven different species. Residues sustaining high strain ($\lambda^{(1)} < 0.25$) are depicted in red on top of a silouhette of the corresponding synthase. In all cases high strain accumulates at the height of the β, where it interacts with γ. Color bars are legends for subsequent panels (note that for *Bos taurus* and *Mycolicibacterium smegmatis* two sets of structures were considered, see *SI* D). **B**. Fraction of conserved residues between a pair of β (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 panel 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.

that species-specific adaptations are indeed present for each component of the cycle, but become only apparent once the dominant functional/structural effect has been first taken into account.

Our 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. 4C, than by using displacement conservation as a metric, Fig. 4D. Of the two outliers seen in Fig. 4C, one corresponds again to low resolution *Bos taurus* structures. The other outlier, depicted in black, corresponds to the structures of *Mycolicibacterium smegmatis*, obtained in presence of the inhibitory drug bedaquiline (BDQ) [25]. This is to be expected as BDQ binds to the c-ring, and is known to affect the interaction between the c-ring and the channel protein a. Once again, as in the case of β chains in *Bacillus sp. PS3*, our 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.

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 changes that a protein can undergo. The set of the allowed functional changes 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 [36, 37] and reconstructing protein atomic maps [38–40]. In contrast, quantitative methods for the comparative analysis of such maps are still underdeveloped. The need for such methods is further amplified by the advent of accurate structures predicted computationally [41, 42]. In response to this need, PSA provides a quantitative description of structural transi-

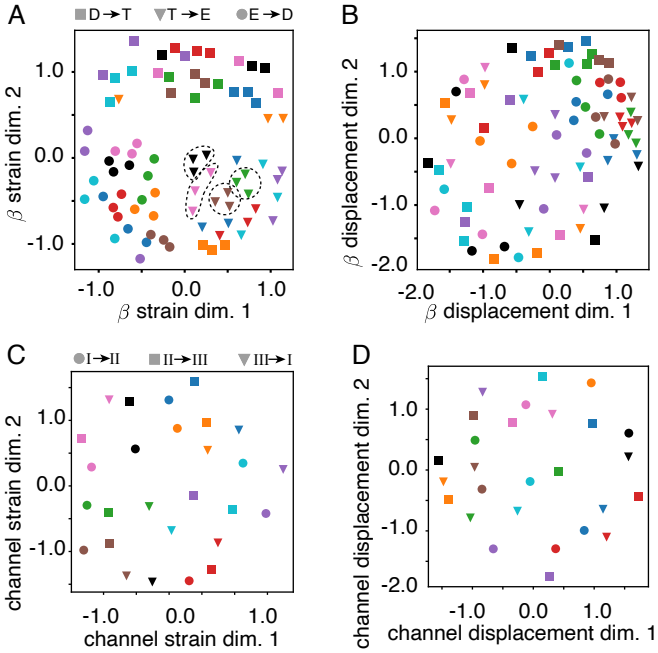


FIG. 4: *Evolutionary conservation of strain patterns.* **A** and **B**. Multidimensional scaling (MDS) of the dissimilarity matrix corresponding to conservation of strain (A) and displacement (B) in β. As is readily seen, with few outliers discussed in the text, strain conservation clusters the data 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** and **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. 3A.

tions that is grounded in fundamental physical principles. In particular, mechanical strain, a local quantity that is by design independent of alignment, is closely related to functional characteristics of multi-protein assemblies, as we showed for the ATP synthase. Furthermore, unlike nowadays prevalent machine-learning methods, PSA requires no training. Overall, mechanical strain is a robust, alignment-independent method that quantifies functional transitions of proteins at the structural level.

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), it is unrelated to sequence-alignment based methods [43]. Furthermore, despite the patent similarity among structural states via visual inspection, simple analysis of alignment dependent quantities, such as residues displacements, does not result in a comparable clustering, i.e. comparable evolutionary conservation. The reason is that rotations and displacements depend on spatial alignment, which is arbitrary and thus contains function-unrelated information. Therefore, 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 the elasticity theory, should be used to better understand their functioning and their evolution.

Acknowledgments

Acknowledgements. 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 supplementary material. This work was partly funded by a laCaixa grant to PS. Early stages of this research have 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).

[1] B. Alberts, "The cell as a collection of protein machines: preparing the next generation of molecular biologists," *cell*, vol. 92, no. 3, pp. 291–294, 1998.

[2] C. Pál, B. Papp, and M. J. Lercher, "An integrated view of protein evolution," *Nature reviews genetics*, vol. 7, no. 5, pp. 337–348, 2006.

[3] L. Holm and C. Sander, "Mapping the protein universe," *Science*, vol. 273, no. 5275, pp. 595–602, 1996.

[4] C. Zhang, M. Shine, A. M. Pyle, and Y. Zhang, "Us-align: Universal structure alignments of proteins, nucleic acids, and macromolecular complexes," *bioRxiv*, 2022.

[5] T. Yamato, "Strain tensor field in proteins," *Journal of molecular graphics*, vol. 14, no. 2, pp. 105–107, 1996.

[6] M. R. Mitchell, T. Tlusty, and S. Leibler, "Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings," *Proceedings of the National Academy of Sciences*, vol. 113, no. 40, pp. E5847–E5855, 2016.

[7] Y. Wang and G. Zocchi, "The folded protein as a viscoelastic solid," *EPL (Europhysics Letters)*, vol. 96, no. 1, p. 18003, 2011.

[8] H. Qu and G. Zocchi, "How enzymes work: A look through the perspective of molecular viscoelastic properties," *Physical Review X*, vol. 3, no. 1, p. 011009, 2013.

[9] J. Howard, *Mechanics of Motor Proteins and the Cytoskeleton*. Sinauer Associates, Publishers, 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," *Proceedings of the National Academy of Sciences*, vol. 119, no. 10, p. 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," *Nature communications*, vol. 11, no. 1, pp. 1–13, 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*, vol. 611, no. 7935, pp. 380–386, 2022.

[13] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *science*, vol. 344, no. 6191, pp. 1492–1496, 2014.

[14] A. Hahn, J. Vonck, D. J. Mills, T. Meier, and W. Kühlbrandt, "Structure, mechanism, and regulation of the chloroplast atp synthase," *Science*, vol. 360, no. 6389, p. eaat4318, 2018.

[15] A. J. M. Spencer, *Continuum mechanics*. Courier Corporation, 2004.

[16] P. G. Ciarlet, *Mathematical elasticity: Three-dimensional elasticity*. SIAM, 2021.

[17] D. Stock, A. G. Leslie, and J. E. Walker, "Molecular architecture of the rotary motor in atp synthase," *Science*, vol. 286, no. 5445, pp. 1700–1705, 1999.

[18] W. Kühlbrandt, "Structure and mechanisms of f-type atp synthases," *Annual review of biochemistry*, vol. 88, pp. 515–549, 2019.

[19] H. Guo, T. Suzuki, and J. L. Rubinstein, "Structure of a bacterial atp synthase," *Elife*, vol. 8, p. e43128, 2019.

[20] 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 f1-fo coupling," *Science*, vol. 364, no. 6446, p. eaaw9128, 2019.

[21] 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," *Proceedings of the National Academy of Sciences*, vol. 113, no. 45, pp. 12709–12714, 2016.

[22] 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*, vol. 4, p. e10180, 2015.

[23] T. E. Spikes, M. G. Montgomery, and J. E. Walker, "Structure of the dimeric atp synthase from bovine mitochondria," *Proceedings of the National Academy of Sciences*, vol. 117, no. 38, pp. 23519–23526, 2020.

[24] 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 f1fo atp synthase accommodates symmetry mismatch," *Nature communications*, vol. 11, no. 1, pp. 1–10, 2020.

[25] 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*, vol. 589, no. 7840, pp. 143–147, 2021.

[26] H. Guo and J. L. Rubinstein, "Structure of atp synthase under strain during catalysis," *Nature communications*, vol. 13, no. 1, pp. 1–9, 2022.

[27] P. Mitchell, "Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism," *Nature*, vol. 191, no. 4784, pp. 144–148, 1961.

[28] P. D. Boyer, "The atp synthasea splendid molecular machine," *Annual review of biochemistry*, vol. 66, no. 1, pp. 717–749, 1997.

[29] H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita, "Direct observation of the rotation of f1-atpase," *Nature*, vol. 386, no. 6622, pp. 299–302, 1997.

[30] R. Yasuda, H. Noji, K. Kinosita Jr, and M. Yoshida, "F1-atpase is a highly efficient molecular motor that rotates with discrete 120 steps," *Cell*, vol. 93, no. 7, pp. 1117–1124, 1998.

[31] 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," *Proceedings of the National Academy of Sciences*, vol. 70, no. 10, pp. 2837–2839, 1973.

[32] J. P. Abrahams, A. G. Leslie, R. Lutter, and J. E. Walker, "Structure at 2.8 a resolution of f1-atpase from bovine heart mitochondria," *Nature*, vol. 370, no. 6491, pp. 621–628, 1994.

[33] 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*, vol. 5, 2016.

[34] 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*, vol. 8, p. e43864, 2019.

[35] G. Oster and H. Wang, "Reverse engineering a protein: the mechanochemistry of atp synthase," *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1458, no. 2-3, pp. 482–510, 2000.

[36] A. Dashti, P. Schwander, R. Langlois, R. Fung, W. Li, A. Hosseinizadeh, 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*, pp. 463–475, World Scientific, 2018.

[37] E. D. Zhong, T. Bepler, B. Berger, and J. H. Davis, "Cryodrgn: reconstruction of heterogeneous cryo-em structures using neural networks," *Nature methods*, vol. 18, no. 2, pp. 176–185, 2021.

[38] 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*, vol. 7, p. e42166, 2018.

[39] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, "cryosparc: algorithms for rapid unsupervised cryo-em structure determination," *Nature methods*, vol. 14, no. 3, pp. 290–296, 2017.

[40] T. Grant, A. Rohou, and N. Grigorieff, "cistem, user-friendly software for single-particle image processing," *elife*, vol. 7, p. e35383, 2018.

[41] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, *et al.*, "Highly accurate protein structure prediction with alphafold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[42] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N.

Kinch, R. D. Schaeffer, *et al.*, "Accurate prediction of protein structures and interactions using a three-track neural network," *Science*, vol. 373, no. 6557, pp. 871–876, 2021.

[43] 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 co-evolution captures native contacts across many protein families," *Proceedings of the National Academy of Sciences*, vol. 108, no. 49, pp. E1293–E1301, 2011.

[44] P. Gullett, M. Horstemeyer, M. Baskes, and H. Fang, "A deformation gradient tensor and strain tensors for atomistic simulations," *Modelling and Simulation in Materials Science and Engineering*, vol. 16, no. 1, p. 015001, 2007.

[45] J. A. Zimmerman, D. J. Bammann, and H. Gao, "Deformation gradients for continuum mechanical analysis of atomistic simulations," *International Journal of Solids and Structures*, vol. 46, no. 2, pp. 238–253, 2009.

Supplementary Material

A. Protein Strain Analysis in a nutshell

A quantitative description of protein structural changes should capture local (residue scale) deformations in a manner that is independent of the alignment of the structures being compared. The problem of finding appropriate invariant quantities for material deformations is well known to engineers and geologists who study materials in configurations far from their reference state [9, 15]. The solution there is found within the mathematical formalism of finite strain theory (FST), which forms the basis for non-linear elasticity. Indeed, FST compares a pair of different states of one same material (e.g. a metal beam with and without a load or a protein with and without a ligand), and provides different types of quantitative local information. We now summarize the key aspects of FST, and then elaborate on how to apply to proteins the formalism of strain analysis (PSA).

Box 1 — Summary of Finite Strain Theory

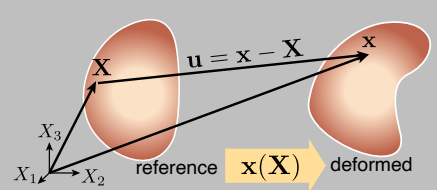
A body characterized by material cartesian coordinates X_i, with $i = 1, 2, 3$, is deformed to $\mathbf{x}(\mathbf{X})$. The *deformation gradient* $F_{ij}(\mathbf{X})$ is then given by

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \mathbf{R}\mathbf{U}$$

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

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

where we used the *displacement field* $u_i = x_i - X_i$, see schematic below.



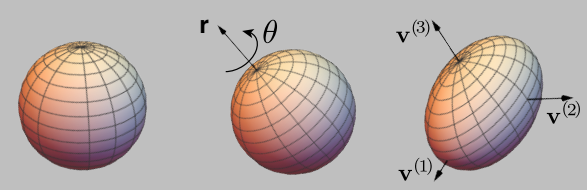
The local *rotation angle* $\theta(\mathbf{X})$ and *axis* $r_i(\mathbf{X})$ are calculated from the rotation matrix via

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

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

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

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



Invariance. $\mathbf{F}$, $\mathbf{R}$ 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.

Decomposition. Using this formalism a deformation $\mathbf{x}(\mathbf{X})$ can be decomposed in a global translation, local rotations (characterized by $\mathbf{r}$ and θ) and local strain (characterized by $\mathbf{v}^{(i)}$ and $\lambda^{(i)}$).

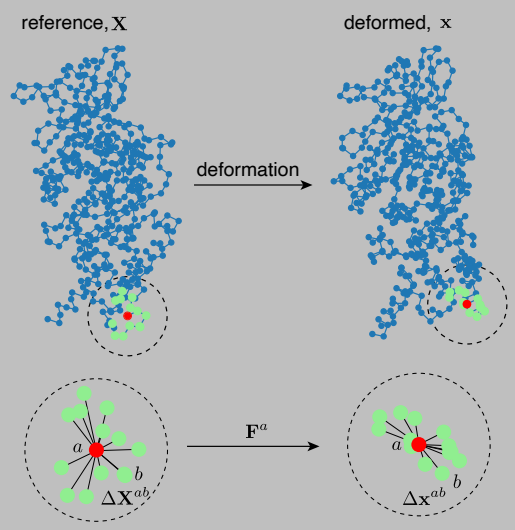
The mathematical formalism of FST, summarized in Box 1, derives information on the local rotation angle θ and rotation axis $\mathbf{r}$ implied by the structural change from the reference to the deformed state. In addition, FST also provides the alignment-independent strain tensor, which we characterize by the principal axes of deformation $\mathbf{v}^{(i)}$ (with $i = 1, 2, 3$). The corresponding principal stretches along the three axes, $\lambda^{(i)}$, which are, importantly, invariant under coordinate transformations, quantify the amount of deformation (note that $\lambda^{(i)} = 1$ corresponds to lack of deformation). Intuitively, 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 the ellipsoid axes and $\lambda^{(i)}$ to their relative stretch (second figure in Box 1). This set of quantities are local, i.e. defined in the neighborhood of each point, and they fully characterize any deformation undergone by a structure from a mechanical point of view.

In Protein Strain Analysis, PSA, we apply FST to protein structural changes. To this end, we first identify the positions of all the carbon atoms C_α 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 [44, 45] (Box 2 for an overview of three such methods). In this paper, we use the method of Ref. [44] 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 Box 1.

Box 2 — 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 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 figure.



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

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

for each (ab) pair.

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

- **Minimal.** $\mathbf{H}^a$ is set to zero, which keeps the approximation linear, see [44]. 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 $\mathbf{F}^a$.
- **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 [44].
- **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 [45].

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, `github.com/???/PySA`. The library takes as input a pair of reference names from the Protein Data Bank, downloads the corresponding `.pdb` files, and applies FST using the discrete method described in Box 2. The results can be plotted, as well as saved in modified `.pdb` files, where FST information is encoded on the $\beta-$factors of the corresponding residues. This allows rendering structures using standard structural biology visualization software, such as `ChimeraX` or `PyMol`, applying color schemes that reflect FST quantities, such as rotation angle or principal stretch.

B. Schematics of ATP synthase

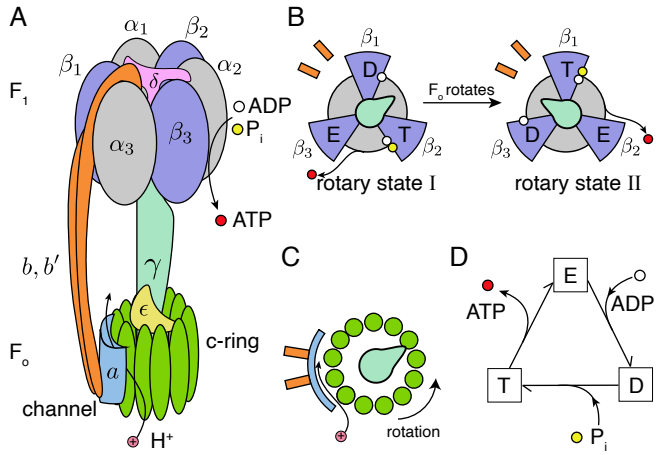


FIG. 5: *Schematic overview of the structure and functioning of ATP synthase.* **A.** ATP synthase complex structure highlighting main components. The F_o complex is constituted by the $c-$ring (immersed in the membrane) and the central "stalk" γ and ϵ. F_o rotates driven by a controlled flux of protons 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' that with which it interacts via δ. As a consequence, rotation of γ triggers sequential conformational changes of the three β, which participate in the synthesis of ATP from ADP and inorganic phosphate P_i. **B.** Schematic view from the top of the F_1 complex as F_o undergoes a 120° rotations. Each of the three β, 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 α. **C.** Schematic view from the top of the F_o complex, indicating rotation direction. **D.** Enzymatic diagram of the synthesis cycle at an individual β. Colors of parts are consistent across panels.

C. Extension-compression cycle in peripheral stalk

So far we have discussed strain on the catalytic β. We now focus instead on the peripheral stalk, which has been suggested to be responsible for energy storage via elastic deformation. Figure **??** revealed a large accumulation of strain on bb' during assembly transition I→II, see also Fig. 6A. This corresponds to compressive strain, i.e. to a region with low values of $\lambda^{(1)}$. In contrast, Fig. 6B shows that the same I→II transition does not contain significant extensile strain, large values of $\lambda^{(3)}$. The converse is true for the transition III→I, which corresponds to significant extensile strain, but very little compressive strain. Finally, during the II→III transition both extensile and compressive strain are low. Therefore, throughout a full cycle, the peripheral stalk undergoes compression (I→II), bending (II→III), and extension (III→I) transitions.

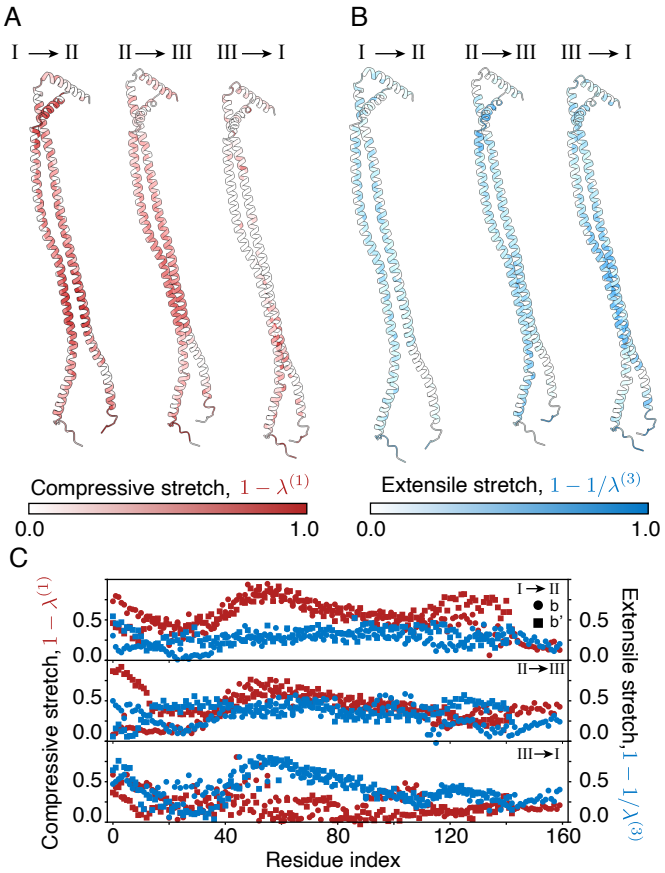


FIG. 6: *Strain cycle of the peripheral stalk.* **A.** Peripheral stalk bb' colored by compressive strain, where small values of $\lambda^{(1)}$ correspond to strong compressions. As one can see, transition I→II implies a strong compression, transition II→III intermediate, and the transition III→I shows almost no compression. **B.** Same as A but with the extensile strain. The results mirror those of A, with extensile strain being contained in III→I, with the other two transitions containing small extensile strain. **C.** Compressive and extensile strain profiles of the two chains in the peripheral stalk for all three transitions. Whereas in II→III extension and compression domains are similar, in I→II and III→I either one or the other dominates. Both chains, b and b', exhibit similar behavior.

Figure. 6C allows to quantify this cycle. Indeed, we observe that compressive strain dominates in I→II, whereas extensile strain dominates in III→I. It is worth noting that both chains, b and b', behave similarly, which speaks to the fact that strain is a local, yet coarse-grained, characterization of deformations. Taken together, our analysis shows that the peripheral stalk undergoes a deformation cycle that may allow for storage and release of energy during ATP synthesis. Such elastic energy accumulation has been suggested to mediate the loose coupling between F_o and F_1 [14, 33, 34], which allows to accommodate their stoichiometric mismatch: $3\alpha\beta$ compared to14 chains in c-ring.

D. Description of additional materials

We append all the `jupyter` notebooks necessary to reproduce the results in the main text, see table below. The notebooks make use of `PyPSA`. Except `Strain_evolution.ipynb`, all other notebooks correspond to a detailed analysis of a pair of `pdb` files that were contributed in the corresponding paper. Since all papers contain, at least, three structures, at the beginning of the notebook there is an option to choose among the corresponding `pdb` files. There are also options to highlight diverse chains (chain β_3 is the one highlighted by default). Therefore, while in the main text only a detailed analysis of the spinach chloroplast structures in [14] is presented, the notebooks contain analogous analysis for the structures of all the other species.

The notebook `strain_evolution.ipynb` contains data structures that standardize the nomenclature of chains, assembly states, and enzymatic transitions. This is necessary in order to perform comparative analysis across species. Finally, the notebook `strain_evolution.ipynb` corresponds to the evolutionary analysis behind Figures 3 and fig:straincons. We note that this second notebook loads the output of `strain_evolution.ipynb`.

File name	Description
`Hahn_2018.ipynb`	Detailed analysis of structures in [14]
`Guo_2019.ipynb`	Detailed analysis of structures in [19]
`Guo_2021.ipynb`	Detailed analysis of structures in [25]
`Guo_2022.ipynb`	Detailed analysis of structures in [26]
`Murphy_2019.ipynb`	Detailed analysis of structures in [20]
`Sobti_2020.ipynb`	Detailed analysis of structures in [24]
`Spikes_2020.ipynb`	Detailed analysis of structures in [23]
`Vinothkumar_2016.ipynb`	Detailed analysis of structures in [21]
`Zhou_2016.ipynb`	Detailed analysis of structures in [22]
`reference_tables.ipynb`	Standardizes the labels for chains and assembly states
`strain_evolution.ipynb`	Evolutionary analysis of data presented in all papers above