



How round is a protein? Exploring protein structures for globularity using conformal mapping

Joel Hass^{1*} and Patrice Koehl²

¹ Department of Mathematics, University of California, Davis, Davis, CA, USA

² Department of Computer Science and Genome Center, University of California, Davis, Davis, CA, USA

Edited by:

Henri Orland, Commissariat à l'Energie Atomique, France

Reviewed by:

Ilpo Vattulainen, Tampere University of Technology, Finland

Michel Bauer, Commissariat à l'Energie Atomique, France

*Correspondence:

Joel Hass, Department of Mathematics, University of California, 1 Shields Avenue, Davis, CA 95616, USA
e-mail: hass@math.ucdavis.edu

We present a new algorithm that automatically computes a measure of the geometric difference between the surface of a protein and a round sphere. The algorithm takes as input two triangulated genus zero surfaces representing the protein and the round sphere, respectively, and constructs a discrete conformal map f between these surfaces. The conformal map is chosen to minimize a symmetric elastic energy $E_S(f)$ that measures the distance of f from an isometry. We illustrate our approach on a set of basic sample problems and then on a dataset of diverse protein structures. We show first that $E_S(f)$ is able to quantify the roundness of the Platonic solids and that for these surfaces it replicates well traditional measures of roundness such as the sphericity. We then demonstrate that the symmetric elastic energy $E_S(f)$ captures both global and local differences between two surfaces, showing that our method identifies the presence of protruding regions in protein structures and quantifies how these regions make the shape of a protein deviate from globularity. Based on these results, we show that $E_S(f)$ serves as a probe of the limits of the application of conformal mapping to parametrize protein shapes. We identify limitations of the method and discuss its extension to achieving automatic registration of protein structures based on their surface geometry.

Keywords: proteins, genus zero surfaces, conformal mapping, diffeomorphism, triangular mesh

1. INTRODUCTION

Proteins, the end products of the information encoded in the genome of any organism, play a central role in defining the life of this organism. They catalyze most biochemical reactions within cells and are responsible, among other functions, for the transport of nutrients and for signal transmission within and between cells. As a consequence, a major focus of bioinformatics is to study how the information contained in a gene is decoded to yield a functional protein (Pevsner, 2009). The overall principles behind this decoding are well understood. The sequence of nucleotides that forms a gene is first translated into an amino acid sequence, following the rules encoded in the genetic code. The corresponding linear chain of amino acids becomes functional only when it adopts a three-dimensional shape, the so-called tertiary, or native structure of the protein. This is by no means different from the macroscopic world: most proteins serve as tools in the cell and as such either have a defined or adaptive shape to function, much as the shapes of the tools we use are defined according to the functions they need to perform.

Protein structures come in a large range of sizes and shapes. They can be divided into four major groups, corresponding to *fibrous* proteins, *membrane* proteins, *globular* proteins, and *disordered* proteins. Fibrous proteins are elongated molecules in which the secondary structure forms the dominant structure (Fraser, 2012). They are insoluble, play a structural or supportive role in the cell, and are also involved in movement (such as in muscle and ciliary proteins). Membrane proteins are restricted to

the phospho-lipid bilayer membrane that surrounds the cell and many of its organelles (White and Wimley, 1999). These proteins cover a large range of shapes, from globular proteins anchored in the membrane by means of a tail, to proteins that are fully embedded in the membrane. Globular proteins, also referred to as *spheroproteins*, due to their compactness, have a unique structure derived from a non-repetitive sequence. They range in size from one to several hundred residues, and adopt a compact structure (Lim, 1974; Levitt and Chothia, 1976; Branden and Tooze, 1991). While proteins belonging to these three groups illustrate the shape-defines-function rule mentioned above, intrinsically disordered proteins form a significant group of exceptions, as they lack stable structures (Dyson and Wright, 1999, 2005; Dunker et al., 2008). Shape remains important for those proteins, although it is its flexibility and plasticity that is of essence, as shown for example in the case of P53 (Oldfield et al., 2009).

The overall importance of shapes for proteins underlines the importance of being able to study, measure and compare those shapes. The most relevant mathematical fields for this topic are Topology and Geometry. One of the first questions that arise in these fields is what distinguishes a space from the simplest and most symmetric shape, the sphere (Bryant and Sangwin, 2011). The 3-dimensional Poincaré conjecture for example, recently proved by Perelman (for review see Morgan, 2005), states that if a closed 3-manifold is simply connected then it is homeomorphic to the 3-sphere. In differential geometry, a main focus is how the local geometry of a space, as measured through its curvature,

differs from the local geometry of a sphere, and how that difference affects global properties of the space. The Sphere Theorem of differential geometry states that a simply-connected smooth manifold whose curvatures are sufficiently close to those of a sphere is itself a sphere (Brendle and Schoen, 2009).

The fundamental question that arises is how to describe the geometry of a shape such as a protein. The configuration of atoms that constitute a protein can be explicitly obtained by high-resolution experimental methods such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, or cryo-electron microscopy. As of September 2014, the geometries of over 100,000 proteins are available in the Protein Data Bank (PDB) (Bernstein et al., 1977; Berman et al., 2000). The PDB file corresponding to a protein contains the coordinates of all its atoms. This representation has its limitations. Indeed, it corresponds to a rigid representation of a protein, while proteins have dynamic structures, a key feature that explains their functions, over a large range of time scales, from the nanosecond to the minute time scales (Henzler-Widman and Kern, 2007; Henzler-Widman et al., 2007; Vendruscolo and Dobson, 2011). This means that modeling them with a single rigid representative in 3-space \mathbb{R}^3 can be problematic.

One approach to overcoming the challenges raised by flexibility is to work with the geometry of a 2-dimensional surface that encloses the protein, rather than with the 3-dimensional atomic coordinates. Following the space-filling models such as those of Corey-Pauling-Koltun (CPK; Corey and Pauling, 1953; Koltun, 1965), a protein is represented as the union of balls, whose centers match with the atomic centers and radii defined by van der Waals parameters. The structure of a protein is then fully defined by the coordinates of these centers, and the radii values. One option for generating a 2-dimensional surface that encloses a protein is to consider the geometric surface or boundary of its union of balls, the vdW surface of the protein. Note that other definitions are possible, such as the accessible surface (Lee and Richards, 1971), the molecular surface (Richards, 1977), and the skin surface (Edelsbrunner, 1999). While the dynamics of a protein can cause some distortion of its surface, the geometry of this surface is generally well preserved under motions, much more so than the occupied solid region in 3-space. Focusing attention on the surface of the boundary of a protein is also biologically reasonable, since the main biological functions of a protein take place at its surface.

Within this framework, the basic question about protein shape resemblance asks for a measure of the similarity of two protein surfaces. With this paper, we begin an investigation of this question. Our eventual aim is to get a meaningful measurement of the relative similarity of any pair of proteins. It seems useful however to first compare protein geometries to a single well understood benchmark. We could take some fixed protein as a benchmark, but the results we obtain would then be dependent on a rather arbitrary choice of a reference protein. To develop our method in a geometrically meaningful framework, we use the round sphere as a base shape to compare to a range of proteins. The sphere is the most symmetrical surface in 3-space, and the resemblance of a protein to a sphere reflects the symmetry, convexity, and globularity of the protein. With this in mind, we focus on the following question: How round

is a protein? A suitable answer would assign a nonnegative number to each protein that indicates how far away it is from being round. This number should be stable under small perturbations, and not change significantly for different poses of a single flexible protein. We also choose it to be independent of scale.

Ideally, shape comparison techniques aim at defining directly a map between any two shapes that is as close to an isometry as possible. This is however a difficult problem, as the space of possible near-isometric maps is extremely large and not straightforward to characterize mathematically. Despite these difficulties, there have been many methods developed to find such mappings, including one for mapping bio-molecular surfaces onto the sphere (Rahi and Sharp, 2007). These methods rest on the definition of a distance measure that evaluates how close the map is to an isometry, on choices of sets of points on the two shapes, and on an algorithm for finding the mapping between these sets of points that minimizes this distance measure. The harmonic or Dirichlet energy (Eck et al., 1995; Alliez et al., 2002), the Procrustes distance and its continuous variant (Lipman et al., 2013a), the Gromov-Hausdorff distance and variants (Bronstein et al., 2006; Mémoli, 2007), and the conformal Wassterstein distance (Boyer et al., 2011; Lipman and Daubechies, 2011; Lipman et al., 2013b) are popular distance measures used in this context. The closest to-isometric mapping is then found by exhaustive evaluation of the chosen distance measure over all permutations of the landmark points on the two surfaces (Mémoli and Sapiro, 2005), by direct optimization, such as the generalized multi-dimensional scaling algorithm proposed by Bronstein and colleagues in (Bronstein et al., 2006), or through conformal parametrization of the surfaces (Gu and Yau, 2003; Gu et al., 2004).

In this paper we introduce a new method for measuring the similarity between a protein and the sphere that is based entirely on intrinsic geometry. It compares the two shapes by measuring the distortion of an optimal conformal mapping of the surface of one to the surface of the other. A preliminary report of this method was published in Koehl and Hass (2014). We assume that the surface of the protein is a surface of genus zero in \mathbb{R}^3 . This allows us to look for optimal diffeomorphisms (differentiable maps with differentiable inverses) between the surface and the sphere. The restriction to genus zero is appropriate for a wide variety of natural surface comparison problems, including facial recognition (Wang et al., 2005), alignment and comparison of brain cortical surfaces (see for example Gu et al., 2004; Hurdal and Stephenson, 2009), and geometric identification and comparison of bones (for example Boyer et al., 2011), in addition to protein surfaces (Rahi and Sharp, 2007). Compared to the other techniques for comparing genus zero surfaces mentioned above, the method we describe here has the advantage of being both computationally efficient and dependent only on the intrinsic surface geometry of the protein. Computational efficiency allows for comparisons with large collections of shapes, such as those found in the Protein Data Bank. Dependence on the intrinsic surface geometry makes our method well suited for modeling geometric similarities of flexible shapes, shapes that can bend over time to realize varying configurations in space. A substantial number of proteins demonstrate substantial flexibility, and thus our method seems well suited to their study.

As mentioned above, this paper is an extension of a previous study (Koehl and Hass, 2014). It differs mainly in that we have modified the elastic energy used to measure the difference between the optimal conformal mapping designed to map a surface onto another and an isometry, and we justify why. We also introduce a new quantitative measure of the similarity between a protein surface and the round sphere, and describe how this measure allows us to set the limits of the applications of conformal mapping to analyzing protein shapes. The paper is organized as follows. Section 2 provides the mathematical background for our algorithm: conformal geometry and measures of similarity between surfaces of genus zero. In Section 3, we provide the details of its implementation on discrete surfaces, as well as a description of the test cases used in the Results section. Section 4 presents and discusses the results obtained by our algorithm first on simple test cases to show the validity and power of the approach, then on a large dataset of proteins that are compared to the round sphere. We conclude the paper with a brief discussion on future developments.

2. MATHEMATICAL BACKGROUND

2.1. BASIC IDEA: FINDING AN OPTIMAL CONFORMAL MAPPING BETWEEN TWO SURFACES OF GENUS ZERO

Let F_1 and F_2 be two surfaces of genus zero. Our goal is to define a map $f : F_1 \rightarrow F_2$ that is as close as possible to an isometry, i.e., that minimizes the distortion of pairwise geodesic distances between points. When $F_2 = S^2$, i.e., the unit 2-sphere in \mathbb{R}^3 and F_1 and F_2 are scaled to have the same area, then f gives a measure of the roundness of F_1 . We always in this paper scale two surfaces to have the same area, which we can take to be 4π , the area of the unit sphere. We then say that F_1 is round if f is an isometry. For a surface that is not round, some metric distortion is found in any map to or from the sphere.

We now fix $F_2 = S^2$ to be isometric to the unit sphere. A deep result, the *Uniformization Theorem*, states that given any smooth genus zero surface F , there is always a conformal diffeomorphism from F_1 to S^2 (see Bers, 1972). Such conformal maps are not unique. Each conformal diffeomorphism $f : F_1 \rightarrow S^2$ is part of a family of conformal diffeomorphisms. The space of conformal diffeomorphisms from S^2 to itself forms the group $\text{PSL}(2, \mathbb{C})$, called the *Möbius* or *Linear-Fractional* transformations. Any conformal map $C : F_1 \rightarrow S^2$ can be composed with a conformal Möbius transformation $\phi : S^2 \rightarrow S^2$ to give a new conformal map $\phi \circ C : F_1 \rightarrow S^2$, and this construction gives all orientation-preserving conformal maps from F_1 to S^2 .

Given two surfaces F_1 and F_2 and a conformal mapping f between them, f can be understood as the composition of three conformal mapping functions, C_1 , m and C_2^{-1} (see Figure 1). In this composition, m is a Möbius transformation that may arise through composition with transformations ϕ_1 and ϕ_2 as described above. We can choose m among the six-dimensional space of Möbius transformations to yield minimal distortion.

2.2. DISTORTION FROM AN ISOMETRY

At a point $p \in F_1$, a conformal map $f : F_1 \rightarrow F_2$ stretches the metric of F_1 uniformly in all directions by a positive factor $\lambda(p)$. A conformal diffeomorphism then defines a real valued function

$\lambda : F_1 \rightarrow \mathbb{R}^+$ that measures this point-wise stretching. The function $\lambda > 0$ is called the *dilation* and is defined by the formula

$$f^*(g_2) = \lambda^2 g_1 \quad (1)$$

where g_1, g_2 are the metrics on F_1, F_2 respectively. Since $\lambda > 0$, it can be represented in the form $\lambda = e^u$, where $u : F_1 \rightarrow \mathbb{R}$ is a real-valued function.

We use the following energy function to measure the distortion of a conformal map $f : F_1 \rightarrow F_2$ from an isometry. Recall that we have scaled all surfaces to have area equal to one.

Definition. The *symmetric elastic energy* of a conformal diffeomorphism $f : F_1 \rightarrow F_2$ with dilation function $\lambda = e^u$ is given by

$$E_S(f) = E_n(f) + E_n(f^{-1}) = \int_{F_1} u(x)^2 dA + \int_{F_2} u(y)^2 dA. \quad (2)$$

In (Koehl and Hass, 2014), we considered a different distortion energy function:

$$E(f) = \int_{F_1} (\lambda(x) - 1)^2 dA. \quad (3)$$

Equations 2 and 3 differ at two levels. First, the distortion over a whole surface is computed using either the logarithm u of the dilation function λ , or λ directly. The latter varies between 0 and $+\infty$, with values smaller than 1 corresponding to compression and values larger than 1 corresponding to expansion. As such, large compressions can contribute less to the total distortion than large dilations. In contrast, the function $u = \ln(\lambda)$ varies between $-\infty$ and 0 for compression, and between 0 and $+\infty$ for expansion, leading to a more balanced contribution for the two types of distortion. Second, $E_S(f)$ is symmetric and treats equally the distortions induced by f and those induced by f^{-1} . In contrast, $E(f)$ only accounts for the distortions induced by f . For these two reasons, we believe that $E_S(f)$ may be a better measure of distortion from an isometry.

The symmetric elastic energy defined in Equation 2 has the following properties (Hass and Koehl, in preparation):

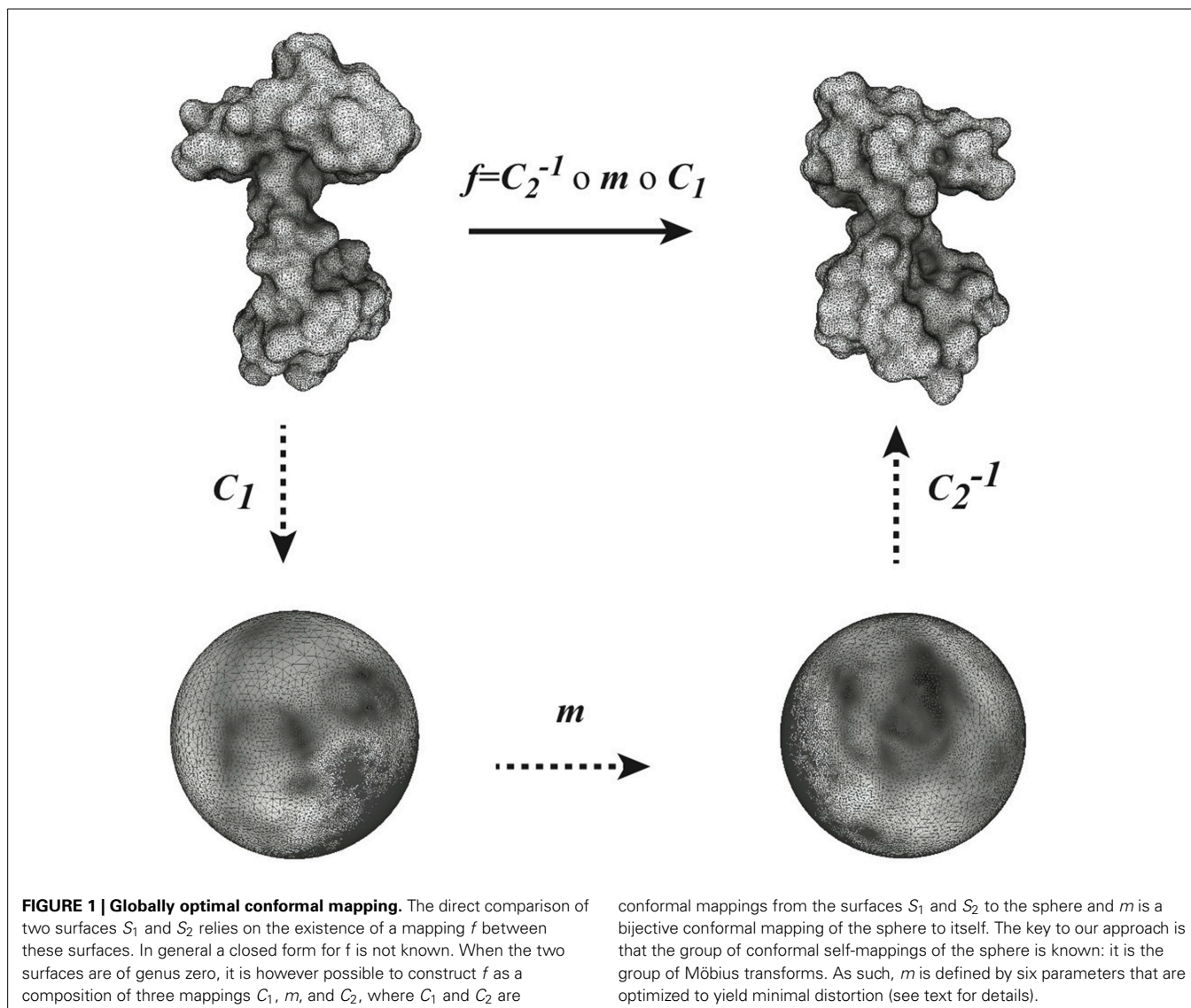
1. For any pair of genus zero surfaces there is a smooth conformal homeomorphism between them that minimizes the symmetric elastic energy.
2. The symmetric elastic energy of a map is zero if and only if the map is an isometry. (Recall that we are assuming that all surface areas are equal to 4π .)

3. MATERIALS AND METHODS

3.1. A GENERAL ALGORITHM FOR MAPPING TWO SURFACES OF GENUS ZERO

The algorithm described below is derived from our initial study of conformal mapping of genus zero surfaces described in Koehl and Hass (2014), which gives a comprehensive description. We focus here on the general concepts and on the differences with the original algorithm.

Let F_1 and F_2 be two surfaces of genus zero, represented by the meshes \mathcal{M}_1 and \mathcal{M}_2 , respectively. Both meshes are taken to



be triangular, with $\mathcal{M}_i = (V_i, E_i, T_i)$, $i = 1, 2$, where $\{V_i, E_i, T_i\}$ denote the vertices, edges and triangles, respectively. We note that these two meshes are completely independent of each other, and are likely to have different combinatorics.

As illustrated in **Figure 1**, we rely on the idea that a conformal mapping f between two surfaces F_1 and F_2 of genus zero can be written as the composition of two discrete conformal mappings C_1 and C_2 that parametrize S_1 and S_2 onto the sphere, and a Möbius transformation m . In optimizing the map produced from this composition, C_1 and C_2 are fixed, while m is variable and depends on six degrees of freedom, summarized in a parameter vector \vec{h} . The key to our approach is to choose the transformation m to minimize the sum of the distortions between the mesh \mathcal{M}_1 representing F_1 and its image $W_m(\mathcal{M}_1)$ warped by f onto F_2 , and between the mesh \mathcal{M}_2 representing F_2 and its image $W_m^{-1}(\mathcal{M}_2)$ warped by f^{-1} onto F_1 . The total distortion is a discrete version of the symmetric elastic energy given by Equation 2 and is computed as a sum over all edges of the two surface meshes:

$$E_S(f) = \sum_{e_{ij} \in E_1} \left(\ln \frac{l'_{ij}}{l_{ij}} \right)^2 \frac{A_{ijk} + A_{ijm}}{3} + \sum_{e_{kn} \in E_2} \left(\ln \frac{l'_{kn}}{l_{kn}} \right)^2 \frac{A_{knp} + A_{knq}}{3} \quad (4)$$

Here E_1, E_2 denote the set of edges in the meshes on F_1 and F_2 respectively, l_{ij} denotes the length of the edge $e_{ij} \in E_1$ that connects vertices v_i, v_j and l'_{ij} the distance from $f(v_i)$ to $f(v_j)$. Similarly l_{kn} denotes the length of the edge $e_{kn} \in E_2$ that connects vertices v_k, v_n and l'_{ij} the distance from $f^{-1}(v_k)$ to $f^{-1}(v_n)$. The areas of the two triangles adjacent to the edge e_{ij} are given by A_{ijk} and A_{ijm} . When f maps a pair of vertices v_i, v_j of F_1 to arbitrary points in F_2 , the distance between these points is computed by extending the metric on the edges of F_2 to a flat Euclidean metric on each 2-simplex of the triangulation.

We have developed all the tools we need to search for a conformal map between two surfaces of genus zero that has minimal distortion, as defined by Equation 4.

- (i) An algorithm for computing the discrete conformal mappings C_1 and C_2 :

While Riemann's Uniformization Theorem guarantees that any smooth genus zero surface F can be mapped conformally to the unit sphere, the theoretical underpinnings of the theory of discrete conformal maps are still being developed. Many methods have been developed to compute them in practice. We follow the approach proposed by Springborn and colleagues, which introduces a notion of discrete conformal equivalence (Springborn et al., 2008). In this method, the mesh \mathcal{M} representing a genus zero surface F is first made topologically equivalent to a disk by removing a vertex v_0 and its star. The transformed mesh is projected conformally on a plane through an optimization procedure (Springborn et al., 2008). The planar mesh is then warped onto the sphere by stereographic projection. Vertex v_0 is reinstated on the North pole of the sphere and connected back to the mesh. Finally, we apply a Möbius normalization to ensure that the center of mass of all vertices is at the origin of the sphere. Full details on the implementation of this algorithm are provided in Koehl and Hass (2014).

- (ii) An algorithm for generating the warping of a discrete mesh onto a surface for a given Möbius transformation $m : S^2 \rightarrow S^2$:

This algorithm works as follows. A vertex v_i in \mathcal{M}_1 has image $v'_i = C_1(v_i)$ in the spherical mesh $C_1(\mathcal{M}_1)$. We locate the image $v''_i = m(v'_i)$ on the spherical mesh $C_2(\mathcal{M}_2)$, namely we identify the triangle t of $C_2(\mathcal{M}_2)$ that contains v''_i and compute barycentric coordinates (α, β, γ) of v''_i in t . Finally, we compute the position of $v'''_i = f(v_i)$ on the surface F_2 by propagating the barycentric coordinates (α, β, γ) onto the triangle t' in \mathcal{M}_2 that corresponds to t . Full details on the implementation of this method are provided in Koehl and Hass (2014).

To simplify the notation, we write $E_S(f) = E_S(m(\vec{h})) = E_S(\vec{h})$ as the map f is determined by m which in turn is determined by the six parameters of \vec{h} . Simple calculations provide the analytical expressions for the symmetric elastic energy function $E_S(\vec{h})$ and its gradient with respect to \vec{h} . This allows us to apply a steepest descent algorithm to search for an optimum for the Möbius transformation m . Our general algorithm for comparing the two surfaces F_1 and F_2 represented with the discrete meshes \mathcal{M}_1 and \mathcal{M}_2 respectively, is then:

The scaling of the surface meshes in step (1) makes our comparison method insensitive to global changes of scale. While not necessary, this step is appropriate to measure scale invariant properties such as roundness. It is also appropriate when the global scale used to describe the vertex positions of the input surfaces is unknown. The damping parameter α_n in step (6) is obtained by solving the equation $E_S(h_n + \alpha_n \nabla E_S(h_n)) \leq E_S(h_n)$ using a line search method. The value of TOL is set to a small constant related to machine error.

Algorithm 1 | Conformal mapping with minimal distortion between discrete surfaces of genus zero.

Initialization. (1) Scale $\mathcal{M}_1, \mathcal{M}_2$ to have total area one.
 (2) Find C_1 and C_2 that conformally map \mathcal{M}_1 and \mathcal{M}_2 onto the sphere, using the method described above.
 (3) Initialise Möbius transformation $m_0 = m(\vec{h}_0)$.
for $n = 0, \dots$ until convergence
 (4) Generate $f(\mathcal{M}_1)$ and $f^{-1}(\mathcal{M}_2)$ using the warping method described above, where $m = m(\vec{h}_n)$.
 (5) Compute $E_S(\vec{h}_n)$ and its gradient $\nabla E_S(\vec{h}_n)$ with respect to \vec{h}_n .
 (6) Update $\vec{h}_{n+1} = \vec{h}_n - \alpha_n \nabla E_S(\vec{h}_n)$.
 (7) Check for convergence: if $E_S(\vec{h}_{n+1}) < \text{TOL}$, stop.
end for

We have implemented the whole procedure outlined in **Algorithm 1** into a Fortran program, RoundProtein. The results of a run of this program include a warping of the mesh \mathcal{M}_1 onto the surface F_2 , $\mathcal{W}_2(\mathcal{M}_1)$ and its corresponding inverse, a warping of the mesh \mathcal{M}_2 onto the surface F_1 , $\mathcal{W}_1(\mathcal{M}_2)$, that minimizes distortion from an isometry among nearby conformal maps, as measured by the symmetric elastic energy. In addition, it gives a numeric measure of the geometric difference between \mathcal{M}_1 and \mathcal{M}_2 based on Equation 4. When the surfaces F_1 and F_2 are isometric, any energy minimizer is an isometry.

When F_2 is set to be the round sphere, $d(F_1, S^2)$ is a measure of the roundness of the surface F_1 .

3.2. TRIANGULAR MESHES FOR REGULAR SHAPES

To compare surfaces of genus zero to the round 2-sphere S^2 , we need a triangular mesh $\mathcal{M}(S^2)$. We generate $\mathcal{M}(S^2)$ by positioning N points uniformly on the sphere and forming a triangulation from these N points.

Distributing points uniformly on the 2-sphere is one of eighteen unsolved mathematics problems proposed by the mathematician (Smale, 1998). We adopt the Thompson formulation of this problem and define it as the problem of determining the minimum electrostatic potential energy configuration of N electrons on the surface of a unit sphere, that repel each other with a force given by Coulomb's law, (Thomson, 1904). The total electrostatic potential energy of a N -electron configuration is expressed as the sum of all its pair-wise interactions,

$$U(N) = \frac{1}{4\pi\epsilon_0} \sum_{i < j} \frac{1}{\|r_i - r_j\|} \quad (5)$$

where ϵ_0 is the vacuum permittivity and r_i is the coordinate vector of electron i . A minimum value of $U(N)$ over the configurations of N distinct points is found by numerical minimization. We used for this the Matlab package "Uniform sampling of the sphere" available from Semeshko (2012). Once a minimum configuration is obtained, a triangular mesh is generated using QHull (Barber et al., 1996). We note that the optimization of $U(N)$ is computationally intensive. To generate a mesh that is dense enough on the sphere, we have used the method described here for $N = 1000$ and subdivided the corresponding mesh recursively using triangular quadrisection (in this process, a triangle is subdivided into

4 triangles by adding the three edges that join the midpoint of its three sides).

In parallel, we have generated dense triangular meshes of the surfaces of the Platonic solids using a similar procedure. Starting from the vertices of a platonic solid, we generate a triangular mesh using QHull. This mesh is then subdivided recursively using triangular quadrisection.

Table 1 summarizes the characteristics of the triangular meshes generated for the sphere and the five Platonic solids.

3.3. DATA SET OF PROTEIN STRUCTURES

The set of structures considered in this study is extracted from the database of 2930 sequence-diverse CATH (Orengo et al., 1997) v2.4 domains used in a previous study (Kolodny et al., 2005). As we focus on three-dimensional structures, we consider the first three levels of CATH, Class, Architecture and Topology, to give a CAT classification. We refer to a set of structures with the same CAT classification as a *fold*. We selected five of the most populated folds in the database of 2930 structures as the test set for all computational experiments run in the studies presented in this paper, including at least one fold from each CATH class: CATH fold 1.10.10, a fully α fold (arc repressor, 55 representatives), CATH fold 2.60.40, a fully β fold (immunoglobulin-like, 156 representatives), and three mixed $\alpha - \beta$ folds: 3.20.20, (TIM-like, 52 representatives), 3.30.70, (two layer sandwich, 85 representatives) and 3.40.50 (Rossmann fold, 185 representatives). These five folds include a total of 533 proteins.

We represent the surface of each protein by its skin surface (Edelsbrunner, 1999), given as a triangulated mesh that surround the atoms of the protein. We use the standard model in chemistry of representing a protein structure as a union of balls, with each ball corresponding to an atom. The skin surface of a protein is then computed from the boundary of the union of these balls, where the center of a ball is given by the coordinates of the corresponding atom, and its radius is set to $2^{1/6}\sigma + R_{H_2O}$, where σ is the vdW parameter for the atom in the AMBER94 force field (Cornell et al., 1995) and R_{H_2O} is the radius of the solvent probe, set to 1.4 Å.

We generated high quality meshes for the skin surfaces of all 533 proteins using the program smesh, described in detail in Cheng and Shi (2004, 2009). Briefly, the algorithm implemented in smesh uses a Delaunay-based method to generate quality mesh for the skin surface incrementally. In particular, points are sampled one by one on the skin surface using a front advancing method. The Delaunay triangulation of the sample

points is maintained using an incremental flipping algorithm developed by Lawson (1972). A subset of the Delaunay triangulation is extracted that defines candidate surface triangles. These candidate surface triangles form a partial mesh and guides the subsequent point samplings. The procedure is applied iteratively until an ϵ -sampling of the whole surface is obtained. The corresponding surface triangles define the skin surface mesh. The corresponding triangular meshes have similar sizes for all proteins, with approximately 25,000 vertices and 50,000 triangles on average. We checked that all the meshes have genus zero.

4. RESULTS AND DISCUSSION

4.1. HOW ROUND ARE THE PLATONIC SOLIDS?

We first consider the surfaces formed by the boundaries of the five Platonic solids: the tetrahedron (4 faces), the hexahedron, or cube (6 faces), the octahedron (8 faces), the dodecahedron (12 faces), and the icosahedron (20 faces). These highly symmetric surfaces serve as a collection of coarse to fine discrete representations of the sphere, with known measures of the quality of the approximation. As such, they provide natural test cases for the effectiveness of our approach to measure surface roundness.

Figure 2 illustrates the quality of the optimal mapping obtained with RoundProtein between the sphere and the icosahedron, both represented with fine discrete triangular meshes whose characteristics are given in **Table 1**. The resulting warping of the icosahedron mesh onto the surface of the sphere shows 12 dense spots, corresponding to the 12 vertices of the icosahedron (left panel). In contrast, the warping of the discrete mesh representing the sphere onto the surface of the icosahedron shows smaller distortion. It represents the icosahedron surface well, with relatively large dilation at the vertices (red spots on the right panel of **Figure 2**). These dilations are expected as the mesh of the sphere needs to adapt to the angle defect at these vertices. Similar results were observed for the four other Platonic solids (results not shown).

Two common measures of the roundness of a surface $F \subset \mathbb{R}^3$ can also be computed analytically for the Platonic solids:

- (i) The *sphericity* of a surface measures how efficiently the surface encloses volume. It is given as the ratio of the surface area of a sphere (with the same volume enclosed by the surface F) to the surface area of F :

$$Sph = \pi^{1/3}(6V)^{2/3}/A,$$

where V is the volume enclosed and A is the surface area. The sphericity is at most one, and equals one only for the round sphere.

- (ii) The ratio R_{IC} of the radii of inscribed and circumscribed spheres. This is often used as a measure of roundness for convex surfaces, but is less useful for general shapes.

We will compare these roundness measures with E_S . Note however that these measures are *extrinsic*, depending on the particular embedding of a surface into \mathbb{R}^3 . They will not be preserved under flexing and bending, unlike E_S .

Table 1 | Characteristics of the discrete meshes of regular shapes.

Shape	Vertices	Faces	Vertices in fine mesh	Faces in fine mesh
Tetrahedron	4	4	8194	16,384
Cube	8	6	6146	12,288
Octahedron	6	8	16,386	32,768
Dodecahedron	20	12	18,434	36,864
Icosahedron	12	20	10,242	20,480
Sphere			15,970	31,936

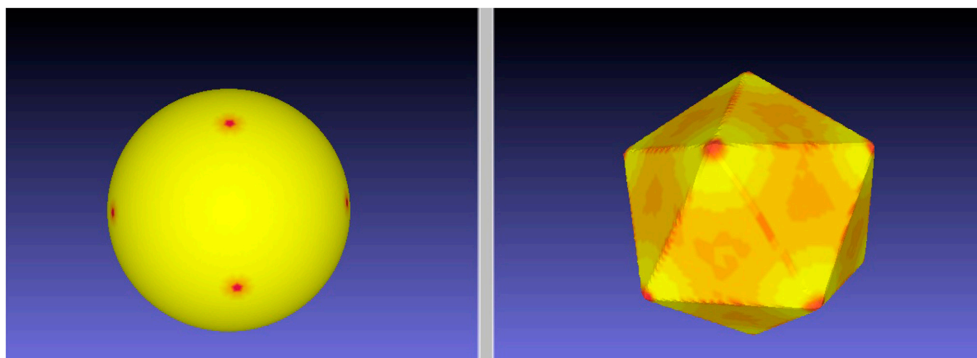


FIGURE 2 | An E_S minimizing map between the sphere and the icosahedron. We computed the minimal distortion conformal map between the discrete mesh \mathcal{M}_1 representing the icosahedron and the discrete mesh \mathcal{M}_2 representing the sphere, where distortion is defined by the symmetric

elastic energy given by Equation 4. The left panel shows the warping of the mesh \mathcal{M}_1 onto the surface of the sphere, while the right panel shows the warping of the mesh \mathcal{M}_2 onto the surface of the icosahedron. Red on the target indicates large dilation in the source.

In addition, we can measure local deformations between a Platonic solid and the sphere by computing the solid angle Ω at each vertex. The solid angle Ω is given by

$$\Omega = q\theta - (q - 2)\pi, \quad (6)$$

where

$$\sin \frac{\theta}{2} = \frac{\cos(\pi/q)}{\sin(\pi/p)}. \quad (7)$$

θ is the interior angle between any two face planes of the solid, p is the number of edges of each face, and q is the number of faces meeting at each vertex.

In **Table 2** we report the values of these measures of roundness for all five Platonic solids as well as the minimal symmetric elastic energies obtained when computing the conformal mapping between the solids and the sphere using RoundProtein. As expected, the sphericity, R_{IC} , and the solid angles Ω increase as the number of faces of the solid increases, i.e., as the solid becomes a better approximation of the sphere. In parallel, E_S decreases, i.e., the differences between the conformal mapping constructed between the solid and the sphere and the isometry become smaller as the number of faces increases. The decrease in E_S is highly correlated with the increases in sphericity, R_{IC} , and solid angles, with Pearson's correlation coefficients of -0.92 , -0.92 , and -0.84 , respectively.

We note that the order of the different measures of roundness does not precisely coincide. Sph and R_{IC} increase monotonically as the number of faces increases. These two measures capture the global shape of the solid. In contrast, the solid angle Ω shows a non-monotonic behavior, illustrated in **Figure 3**. Ω is a measure of local differences with the sphere, as it measures how the local shape around a vertex of the solid differs from a round sphere. While the octahedron has more faces than the cube, its vertices have a smaller solid angle, i.e., they have less local resemblance to the sphere. The same difference in ordering is observed between the dodecahedron and the icosahedron. Interestingly, the symmetric elastic energy E_S captures these local differences between

Table 2 | Roundness of the Platonic solids.

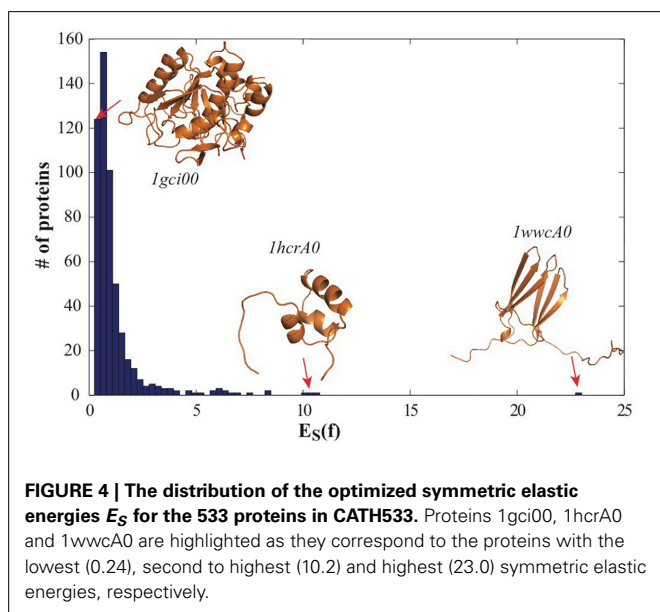
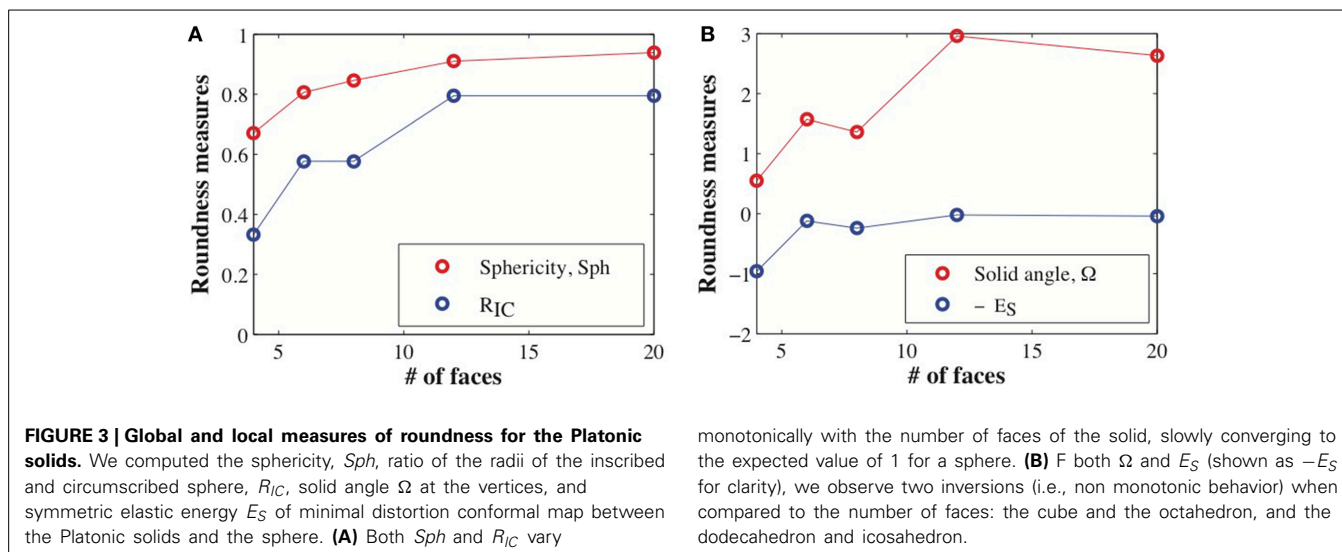
Surface	# of faces	Sphericity, Sph	R_{IC}	Solid angle, Ω	E_S
Tetrahedron	4	0.671	$\frac{1}{3} \approx 0.333$	0.551	0.96
Cube	6	0.806	$\frac{1}{\sqrt{3}} \approx 0.577$	1.571	0.12
Octahedron	8	0.846	$\frac{1}{\sqrt{3}} \approx 0.577$	1.360	0.24
Dodecahedron	12	0.910	0.795	2.962	0.02
Icosahedron	20	0.939	0.795	2.635	0.04
Sphere	Not defined	1.000	1.000	Not defined	0.00

the shapes, while still decreasing as a shape gets closer globally to the sphere. As such, E_S is able to capture both local and non-local differences between a surface and a sphere.

4.2. HOW ROUND IS A PROTEIN?

Proteins come in a wide variety of sizes and shapes. Fibrous proteins, such as collagens that are important for structuring cellular tissues, have elongated shapes while globular spheroproteins that are responsible for catalyzing chemical reactions within cells adopt a compact structure. Understanding the relationship between a protein sequence, its shape, and its function is one of the fundamental problems in biology. Here we address a very specialized question within this problem, namely the characterization of the globularity of a protein, or a quantification of its roundness. A protein structure can be depicted in many different ways, each emphasizing different features of the protein. We focus on the geometry of a 2-dimensional surface that encloses the protein, as defined by the skin surface (Edelsbrunner, 1999).

We use CATH533 as our data set of proteins to assess our approach to measuring the roundness of a surface. CATH533 is a database of 533 protein structures that covers the three main classes of CATH: one fully α fold, one fully β fold, and three $\alpha - \beta$ folds (the TIM fold, an α/β plait, and the Rossmann fold) (see Materials and Methods section above for details). We generated a mesh for each protein in CATH533 using the program smesh (Cheng and Shi, 2004, 2009) and computed the optimal



conformal mapping between this corresponding mesh and the discrete mesh representing the 2-sphere using RoundProtein. In **Figure 4**, we show the distribution of corresponding optimized symmetric elastic energies E_S .

All proteins included in CATH533 are enzymes and therefore they are expected to be globular. Indeed, we observe that computing the optimal mappings f between these proteins and the sphere leads to mappings that are close to isometries, as measured by $E_S(f)$, the symmetric elastic energy of the optimal mapping given in Equation 4. Of the 533 proteins, 352 have an optimized $E_S(f)$ below 1, and 106 of those have an optimized $E_S(f)$ below 0.5. The “best” mapping, i.e. the one closest to an isometry, is observed for the protein with CATH code 1gci00. The latter corresponds to PDB code 1gci which contains the ultra-high resolution (0.78 Å) of *B. Lenti* subtilisin, a serine protease that is known to form

a very compact beta barrel at its core (Kuhn et al., 1998). The corresponding optimized symmetric elastic energy of 0.24 would make this serine protease similar to an octahedron when compared to the sphere (see **Table 2**). The “worst” mapping, i.e., the least similar to an isometry, with an optimized symmetric elastic energy of 23.0, is observed for the protein with CATH code 1wwcA0. This is chain A from the PDB file 1wwc that contains the crystal structures of the neurotrophin-binding domains TrkA, TrkB, and TrkC, with chain A corresponding to TrkA. The TrkA domain is known to fold into an immunoglobulin-like structure, with a core of β -sheet and two long loops at the N and C termini (Ultsch et al., 1999). It is the presence of these two long loops that makes the structure deviate significantly from the sphere (see insert in **Figure 4**). Interestingly, the next to worst comparison of a protein surface with the 2-sphere is observed for the protein with CATH code 1hcrA0. This is chain A from the PDB file 1hcr, corresponding to the complex of a prokaryotic Hin recombinase bound to DNA. The recombinase adopts a 3 helix-bundle conformation, with two long flanking extended polypeptide regions that contact bases in the minor groove of the DNA (Feng et al., 1994). As we only consider the structure of the recombinase, these two regions stand aside from the core helix bundle, leading to a less compact structure (see insert in **Figure 4**).

In **Figure 5**, we compare the optimized symmetric energy $E_S(f)$ of the mapping f between a protein surface and the sphere with the sphericity of the protein surface, computed using Equation 6, for all proteins in CATH533. Just as for the Platonic solids, $E_S(f)$ and the sphericity Sph are correlated: as the sphericity increases, the mapping between the protein surface and the sphere improves, and $E_S(f)$ decreases. Interestingly, the correlation coefficient between $E_S(f)$ and Sph for protein surfaces, -0.64 , is significantly lower than the corresponding correlation coefficient for the Platonic solids, -0.92 . We assign this difference to the fact that the latter are convex while the geometry of even globular proteins is more diverse, with more significant local differences to a round surface that are not captured by sphericity.

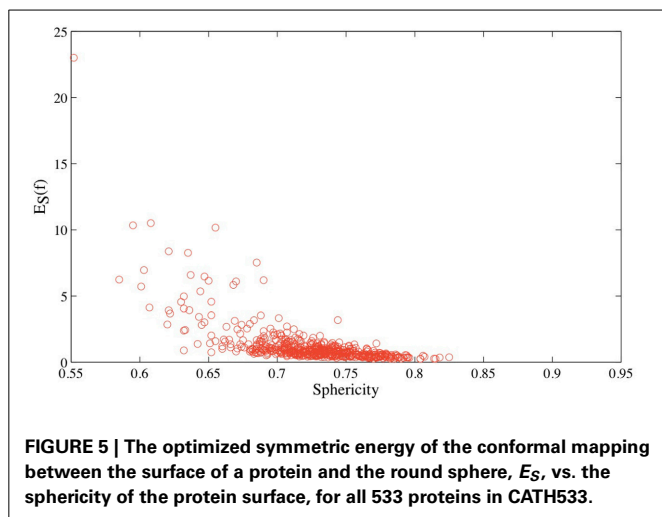


Figure 4 illustrates that the optimal conformal mapping between a protein surface that has long protruding regions and the sphere deviates significantly from an isometry. To help understand why this is the case, we compare in **Figure 6** the surfaces of the three representative proteins identified in **Figure 4** with the surfaces generated from the corresponding warping $f^{-1}(\mathcal{M}(S^2))$ of the mesh represented the sphere onto the surfaces of the three proteins, where the warping is generated with RoundProtein.

If the conformal mapping between a protein surface and the sphere is close to an isometry, it is expected that $f^{-1}(\mathcal{M}(S^2))$ closely follows the surface of the protein. This is indeed observed for the very compact protein 1gci00. The main distortions observed in the warped mesh occur at bumps in the surface (which correspond to the spherical representations of the atoms at the surface of the protein). In the case of the less compact proteins 1hcrA0 and 1wwcA0 however, the warped surfaces generated from $f^{-1}(\mathcal{M}(S^2))$ deviate significantly from the actual surfaces of the proteins. Most of the distortions occur at the protruding regions that are not present in the images of the spheres on the protein surfaces. The discrete conformal mappings of these protruding regions to the sphere introduce very large negative conformal factors on their vertices, which in turn lead to infinitesimally small edge lengths in the projected meshes and consequently large numerical errors. We have observed similar behaviors when computing conformal mappings between generic genus zero surfaces (Koehl and Hass, 2014). This problem is not specific to our method, as it appears in many conformal mapping procedures. In some cases approximating by a conformal map appears to be too restrictive. One solution is to introduce cone singularities in the regions with the worst distortions (see for example Springborn et al., 2008).

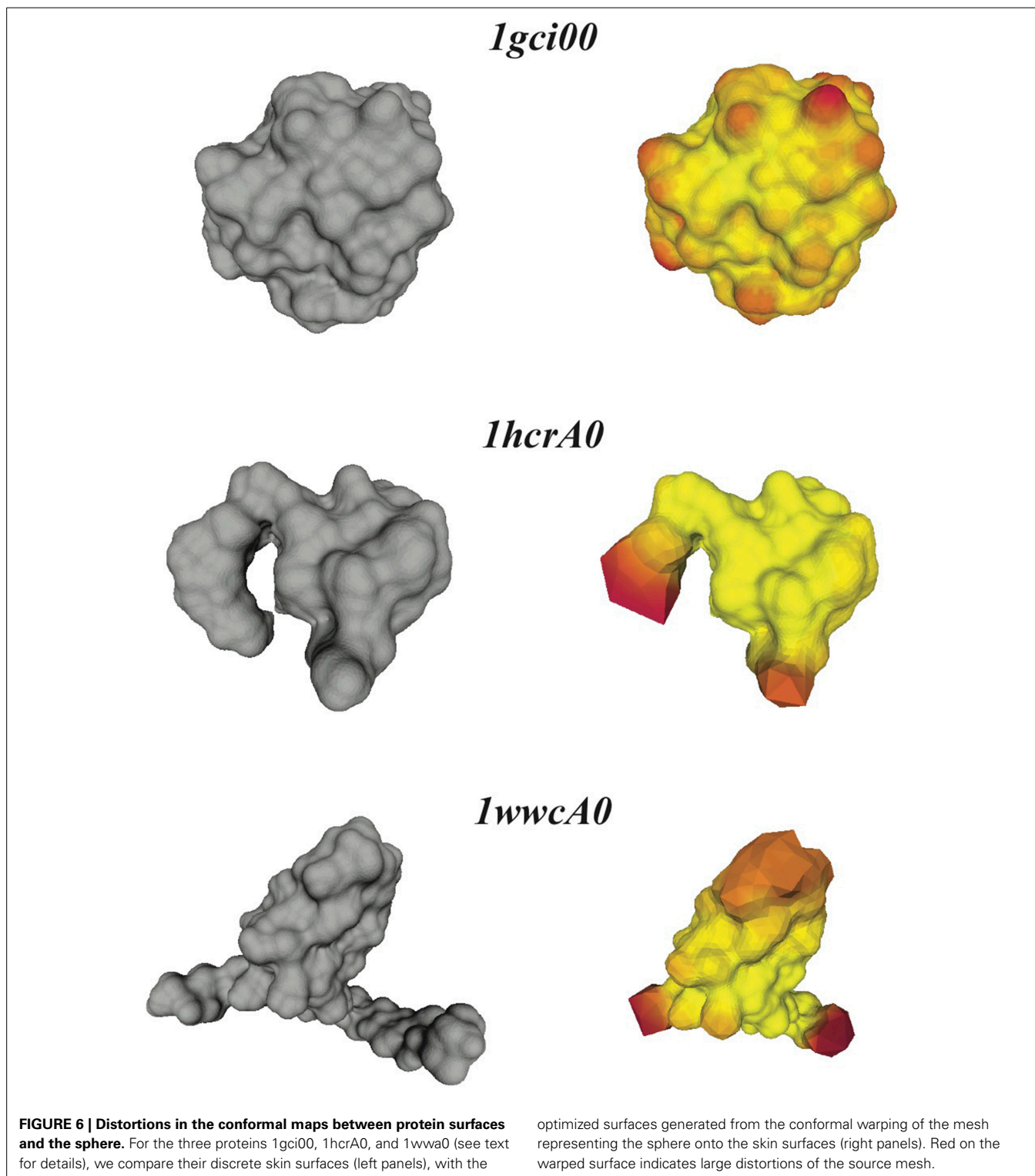
Figure 6 illustrates that the distortions introduced by the restrictive condition that the mapping between the protein surface and the sphere be conformal lead to an image $f^{-1}(\mathcal{M}(S^2))$ of the mesh of the sphere onto the surface of the protein that does not capture well the geometry of this surface. One approach to measuring these distortions is to compute the ratio of the surface

area A_W of $f^{-1}(\mathcal{M}(S^2))$ to the surface area A_P of the source mesh representing this protein. We plot this ratio against the symmetric elastic energy of the refined mapping f , $E_S(f)$, in **Figure 7** for all 533 proteins in CATH533. If the mapping f is close to an isometry, there should be minimal distortion and $f^{-1}(\mathcal{M}(S^2))$ should be a good representation of the surface of the protein (as illustrated in **Figure 6** for 1gci00). The ratio A_W/A_P should then be close to 1. This is indeed observed for the majority of the proteins in CATH533. We find that A_W/A_P is greater than 0.99 for 226 proteins, greater than 0.98 for 471 proteins, and greater than 0.95 for 512 proteins. This ratio decreases significantly as f deviates more and more from an isometry, with a minimal value of 0.79 for protein 1wwcA0. Interestingly, A_W/A_P and $E_S(f)$ are strongly correlated with a Pearson's coefficient of correlation of 0.95. This indicates that $E_S(f)$ has value as a tool to test whether a conformal map is accurately representing a given surface.

5. SUMMARY AND CONCLUSIONS

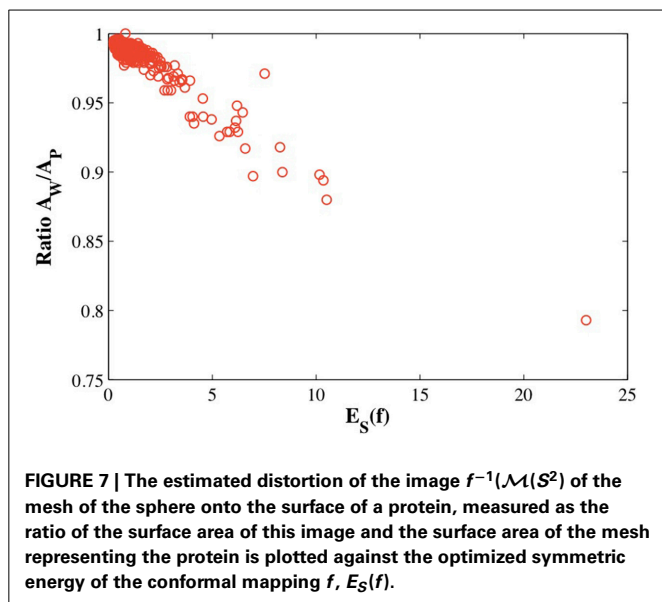
We have developed a new method for quantifying the compactness of a protein structure. In this new approach we compute the conformal map f between the surface of the protein (required to be of genus zero) and the 2-sphere that has minimal distortion, where distortion is defined as a symmetric elastic energy $E_S(f)$ that measures the distance between f and an isometry. It leads to flexible registration of the two surfaces and accurate measurements of their geometric dissimilarities. Its implementation within the program RoundProteins is based on fast and robust numerical methods, making surface comparisons feasible for large data sets of proteins. We have illustrated its use for quantifying the roundness of the Platonic solids and of 533 diverse protein structures. We have demonstrated that the elastic energy $E_S(f)$ captures both global and local differences between two surfaces. We have shown that our method identifies and measures the presence of protruding regions in protein structures that make them deviate from a compact shape.

This paper is a first step toward achieving automatic registration of protein structures based on their surfaces. The method described here is an extension of the approach described in Koehl and Hass (2014) and suffers from similar limitations. We note that it only applies to surfaces of genus zero and that it works best for surfaces that have uniform geometry, without long protrusions (Koehl and Hass, 2014). In this paper, we have shown that this limitation can be used to generate valuable information. The difficulty that RoundProtein encounters in finding a conformal mapping f between a highly non-spherical protein surface and the 2-sphere translates into a high value for the symmetric elastic energy E_S of f . Such a high value measures the extent of the deviation of the protein from being approximately round. It also indicates the limits of the application of conformal mapping to parametrize protein shapes, as high values for E_S correspond to significant deviations between the representations of a surface given by its source mesh and the representation given by the parametrization formed by the target mesh (see **Figure 6**). For the limitation to genus zero surfaces, we note that the concept of discrete conformal structures can be extended to surfaces with arbitrary topology, either through the introduction of cone singularities (Springborn et al., 2008), or through the definition of a



discrete conformal equivalence between a Euclidean triangulation on the surface and a flat or hyperbolic triangulation (Bobenko et al., 2010; Tsui et al., 2013). Finding closest-to-isometric mappings for surfaces with genus greater than zero remains a topic for future studies.

Finally, we note that while the symmetric elastic energy of a conformal mapping between two surfaces F_1 and F_2 defined in Equation 2 is useful for measuring the differences between these two surfaces, it is not clear that it establishes a distance on the space of genus zero shapes. A number of important applications



would benefit from an actual metric on the space of genus zero surfaces.

AUTHOR CONTRIBUTIONS

The two authors contributed equally to the work, as well as to the draft and following revisions of the manuscript.

ACKNOWLEDGMENTS

Patrice Koehl acknowledges support from the Ministry of Education of Singapore through Grant Number: MOE2012-T3-1-008. Joel Hass acknowledges support from the National Science Foundation through Grant Number IIS-1117663.

REFERENCES

- Alliez, P., Meyer, M., and Desbrun, M. (2002). "Interactive geometry remeshing," in *Computer Graphics (Proceedings SIGGRAPH 02)* (New York, NY: ACM), 347–354.
- Barber, C. B., Dobkin, D., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* 22, 469–483. doi: 10.1145/235815.235821
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucl. Acids Res.* 28, 235–242. doi: 10.1093/nar/28.1.235
- Bernstein, F. C., Koetzle, T. F., Williams, G., Meyer, D. J., Brice, M. D., Rodgers, J. R., et al. (1977). The protein databank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112, 535–542. doi: 10.1016/S0022-2836(77)80200-3
- Bers, L. (1972). Uniformization, moduli, and kleinian groups. *Bull. London Math. Soc.* 4, 257–300. doi: 10.1112/blms/4.3.257
- Bobenko, A., Pinkall, U., and Springborn, B. (2010). Discrete conformal maps and ideal hyperbolic polyhedra. arXiv: 1005.2698
- Boyer, D., Lipman, Y., StClair, E., Puente, J., Patel, B., Funkhouser, T., et al. (2011). Algorithms to automatically quantify the geometric similarity of anatomical surface. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18221–18226. doi: 10.1073/pnas.1112822108
- Branden, C., and Tooze, J. (1991). *Introduction to Protein Structure, Vol. 2*. New York, NY: Garland Science.
- Brendle, S. and Schoen, R. (2009). Sphere theorems in geometry. *Surv. Differ. Geometry* 13, 49–84. doi: 10.4310/SDG.2008.v13.n1.a2
- Bronstein, A., Bronstein, M., and Kimmel, R. (2006). Generalized multi-dimensional scaling: a framework for isometry-invariant partial surface matching. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1168–1172. doi: 10.1073/pnas.0508601103
- Bryant, J., and Sangwin, C. (2011). *How Round is Your Circle?* Princeton, NJ: Princeton University Press.
- Cheng, H., and Shi, X. (2004). "Guaranteed quality triangulation of molecular skin surfaces," in *Proceedings IEEE Visualization* (Washington, DC: IEEE Computer Society), 481–488.
- Cheng, H., and Shi, X. (2009). Quality mesh generation for molecular skin surfaces using restricted union of balls. *Comput. Geom.* 42, 196–206. doi: 10.1016/j.comgeo.2008.10.001
- Corey, R., and Pauling, L. (1953). Molecular models of amino acids, peptides and proteins. *Rev. Sci. Instr.* 24, 621–627. doi: 10.1063/1.1770803
- Cornell, W., Cieplak, P., Bayly, C., Gould, I. R., Merz, K. M. Jr., Ferguson, D., et al. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117, 5179–5197. doi: 10.1021/ja00124a002
- Dunker, A., Silman, I., Uversky, V., and Sussman, J. (2008). Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18, 756–764. doi: 10.1016/j.sbi.2008.10.002
- Dyson, H., and Wright, P. (1999). Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *J. Mol. Biol.* 293, 321–331. doi: 10.1006/jmbi.1999.3110
- Dyson, H., and Wright, P. (2005). Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6, 197–208. doi: 10.1038/nrm1589
- Eck, M., DeRose, T., Duchamp, T., Hoppe, H., Lounsbery, M., and Stuetzle, W. (1995). "Multiresolution analysis of arbitrary meshes," in *Computer Graphics (Proceedings SIGGRAPH 95)* (New York, NY: ACM), 175–182.
- Edelsbrunner, H. (1999). Deformable smooth surface design. *Discrete Comput. Geom.* 21, 87–115. doi: 10.1007/PL00009412
- Feng, J., Johnson, R., and Dickerson, R. (1994). Hin recombinase bound to DNA: the origin of specificity in major and minor groove interactions. *Science* 263, 348–355. doi: 10.1126/science.8278807
- Fraser, R. (2012). *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*. Waltham, MA: Academic Press.
- Gu, X., Wang, Y., Chan, T., Thompson, P., and Yau, S.-T. (2004). Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Trans. Med. Imaging* 23, 949–958. doi: 10.1109/TMI.2004.831226
- Gu, X., and Yau, S.-T. (2003). "Global conformal surface parametrization," in *Eurographics Symposium on Geometry Processing* (Aire-la-Ville: Eurographics Association), 127–137.
- Henzler-Widman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964–972. doi: 10.1038/nature06522
- Henzler-Widman, K., Lei, M., Thai, V., Kerns, S., Karplus, M., and Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450, 913–916. doi: 10.1038/nature06407
- Hurdal, M., and Stephenson, K. (2009). Discrete conformal methods for cortical brain flattening. *Neuroimage* 45, 586–598. doi: 10.1016/j.neuroimage.2008.10.045
- Koehl, P., and Hass, J. (2014). Automatic alignment of genus-zero surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 466–478. doi: 10.1109/TPAMI.2013.139
- Kolodny, R., Koehl, P., and Levitt, M. (2005). Comprehensive evaluation of protein structure alignment methods: scoring by geometric measures. *J. Mol. Biol.* 346, 1173–1188. doi: 10.1016/j.jmb.2004.12.032
- Koltun, W. (1965). Precision space-filling atomic models. *Biopolymers* 3, 665–679. doi: 10.1002/bip.360030606
- Kuhn, P., Knapp, M., Soltis, S., Ganshaw, G., Thoene, M., and Bott, R. (1998). The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin. *Biochemistry* 37, 13446–13452. doi: 10.1021/bi9813983
- Lawson, C. L. (1972). *Generation of a Triangular Mesh with Applications to Contour Plotting*. Technical Report MEMO-299, Jet Propulsion Laboratory, Pasadena, CA.
- Lee, B., and Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55, 379–400. doi: 10.1016/0022-2836(71)90324-X
- Levitt, M., and Chothia, C. (1976). Structural patterns in globular proteins. *Nature* 261, 552–558. doi: 10.1038/261552a0
- Lim, V. (1974). Structural principles of the globular organization of protein chains. a stereochemical theory of globular protein secondary structure. *J. Mol. Biol.* 88, 857–872. doi: 10.1016/0022-2836(74)90404-5

- Lipman, Y., Al-Aifari, R., and Daubechies, I. (2013a). The continuous procrustes distance between two surfaces. *Commun. Pure Appl. Math.* 66, 934–964. doi: 10.1002/cpa.21444
- Lipman, Y., and Daubechies, I. (2011). Conformal Wasserstein distances: comparing surfaces in polynomial time. *Adv. Math.* 227, 1047–1077. doi: 10.1016/j.aim.2011.01.020
- Lipman, Y., Puente, J., and Daubechies, I. (2013b). Conformal wasserstein distance: II. computational aspects and extensions. *Math. Comp.* 82, 331–381. doi: 10.1090/S0025-5718-2012-02569-5
- Mémoli, F. (2007). “On the use of Gromov–Hausdorff distances for shape comparison,” in *Proceedings Point Based Graphics* (Boston, MA: AK Peters), 81–90.
- Mémoli, F. and Sapiro, G. (2005). A theoretical and computational framework for isometry invariant recognition of point cloud data. *Found. Comput. Math.* 5, 313–347. doi: 10.1007/s10208-004-0145-y
- Morgan, J. W. (2005). Recent progress on the poincaré conjecture and the classification of 3–manifolds. *Bull. Am. Math. Soc.* 42, 57–78. doi: 10.1090/S0273-0979-04-01045-6
- Oldfield, C., Meng, J., Yang, J., Yang, M., Uversky, V., and Dunker, A. (2009). Flexible nets: disorder and induced fit in the associations of p53 and 14–3–3 with their partners. *BMC Genomics* 9:S1. doi: 10.1186/1471-2164-9-S1-S1
- Orengo, C., Michie, A., Jones, S., Jones, D., Swindells, M., and Thornton, J. (1997). CATH: a hierarchic classification of protein domain structures. *Structure* 5, 1093–1108. doi: 10.1016/S0969-2126(97)00260-8
- Pevsner, J. (2009). *Bioinformatics and Functional Genomics, 2nd Edn.* Hoboken, NJ: Wiley & Sons.
- Rahi, S., and Sharp, K. (2007). Mapping complicated surfaces on the sphere. *Int. J. Comput. Geom. Appl.* 17, 305–329. doi: 10.1142/S0218195907002355
- Richards, F. M. (1977). Areas, volumes, packing, and protein-structure. *Annu. Rev. Biophys. Bioeng.* 6, 151–176. doi: 10.1146/annurev.bb.06.060177.001055
- Semeshko, A. (2012). *Uniform Sampling of the Sphere - a MATLAB Package.* Available online at: <http://www.mathworks.com/matlabcentral/fileexchange/37004-uniform-sampling-of-a-sphere>
- Smale, S. (1998). Mathematical problems for the next century. *Math. Intell.* 20, 7–15. doi: 10.1007/BF03025291
- Springborn, B., Schröder, P., and Pinkall, U. (2008). “Conformal equivalence of triangle meshes,” in *Proceedings SIGGRAPH Asia* (New York, NY: ACM), 79–89.
- Thomson, J. J. (1904). On the structure of the atom: an investigation of the stability and periods of oscillation of a number of corpuscles arranged at equal intervals around the circumference of a circle; with application of the results to the theory of atomic structure. *Philos. Mag. Ser. 7*, 237–265. doi: 10.1080/14786440409463107
- Tsui, A., Fenton, D., Vuong, P., Hass, J., Koehl, P., Amenta, A., et al. (2013). Globally optimal cortical surface matching with exact landmark correspondence. *Inf. Process. Med. Imaging* 23, 487–498. doi: 10.1007/978-3-642-38868-2/41
- Ultsch, M., Wiesmann, C., Simmons, L., Henrich, J., Yang, M., Reilly, D., et al. (1999). Crystal structures of the neurotrophin-binding domain of TrkA, TrkB and TrkC. *J. Mol. Biol.* 290, 149–159. doi: 10.1006/jmbi.1999.2816
- Vendruscolo, M., and Dobson, C. (2011). Protein dynamics: Moores law in molecular biology. *Curr. Biol.* 21, R68–R70. doi: 10.1016/j.cub.2010.11.062
- Wang, Y., Chiang, M.-C., and Thompson, P. (2005). “Mutual information-based 3d surface matching with applications to face recognition and brain mapping,” in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, Vol. 1.* (Washington, DC: IEEE Computer Society), 527–534.
- White, S., and Wimley, W. (1999). Membrane protein folding and stability: physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28, 319–365. doi: 10.1146/annurev.biophys.28.1.319

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 03 October 2014; paper pending published: 10 November 2014; accepted: 21 November 2014; published online: 09 December 2014.

Citation: Hass J and Koehl P (2014) How round is a protein? Exploring protein structures for globularity using conformal mapping. *Front. Mol. Biosci.* 1:26. doi: 10.3389/fmolb.2014.00026

This article was submitted to *Mathematics of Biomolecules*, a section of the journal *Frontiers in Molecular Biosciences*.

Copyright © 2014 Hass and Koehl. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.