Published in final edited form as:

Bull Math Biol. ; 82(8): 108. doi:10.1007/s11538-020-00783-2.

The de Rham–Hodge Analysis and Modeling of Biomolecules

Rundong Zhao[1], **Menglun Wang**[2], **Jiahui Chen**[2], **Yiying Tong**[1], **Guo-Wei Wei**[2,3,4]

[1]Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

[2]Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA

[3]Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI 48824, USA

[4]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

Abstract

Biological macromolecules have intricate structures that underpin their biological functions. Understanding their structure–function relationships remains a challenge due to their structural complexity and functional variability. Although de Rham–Hodge theory, a landmark of twentieth-century mathematics, has had a tremendous impact on mathematics and physics, it has not been devised for macromolecular modeling and analysis. In this work, we introduce de Rham–Hodge theory as a unified paradigm for analyzing the geometry, topology, flexibility, and Hodge mode analysis of biological macromolecules. Geometric characteristics and topological invariants are obtained either from the Helmholtz–Hodge decomposition of the scalar, vector, and/or tensor fields of a macromolecule or from the spectral analysis of various Laplace–de Rham operators defined on the molecular manifolds. We propose Laplace–de Rham spectral-based models for predicting macromolecular flexibility. We further construct a Laplace–de Rham–Helfrich operator for revealing cryo-EM natural frequencies. Extensive experiments are carried out to demonstrate that the proposed de Rham–Hodge paradigm is one of the most versatile tools for the multiscale modeling and analysis of biological macromolecules and subcellular organelles. Accurate, reliable, and topological structure-preserving algorithms for implementing discrete exterior calculus (DEC) have been developed to facilitate the aforementioned modeling and analysis of biological macromolecules. The proposed de Rham–Hodge paradigm has potential applications to subcellular organelles and the structure construction from medium- or low-resolution cryo-EM maps, and functional predictions from massive biomolecular datasets.

Keywords

Algebraic topology; Differential geometry; De Rham–Hodge theory; Macromolecular flexibility; Macromolecular Hodge mode analysis; Cryo-EM analysis

Yiying Tong, ytong@msu.edu; Guo-Wei Wei, wei@math.msu.edu.

1 Introduction

One of the most amazing aspects of biological science is the intrinsic structural complexity of biological macromolecules and its associated functions. The understanding of how changes in macromolecular structural complexity alter their function remains one of the most challenging issues in biophysics, biochemistry, structural biology, and molecular biology. This understanding depends crucially on our ability to model three-dimensional (3D) macromolecular shapes from original experimental data and to extract geometric and topological information from the architecture of molecular structures. Very often, macromolecular functions depend not only on native structures but also on nascent, denatured, or unfolded states. As a result, understanding the structural instability, flexibility, and collective motion of macromolecules is of vital importance. Structural bioinformatics searches for patterns among diverse geometric, topological, instability, and dynamic features to deduce macromolecular function. Therefore, the development of efficient and versatile computational tools for extracting macromolecular geometric characteristics, topological invariants, instability spots, flexibility traits, and mode analysis is a key to infer their functions, such as binding affinity, folding, folding stability change upon mutation, reactivity, catalyst efficiency, and allosteric effects.

Geometric modeling and characterization of macromolecular 3D shapes have been an active research topic for many decades. Surface models not only provide a visual basis for understanding macromolecular 3D shapes, but also bridge the gap between experimental data and theoretical modelings, such as generalized Born and Poisson–Boltzmann models for biomolecular electrostatics (Natarajan et al. 2008; Yu et al. 2008). A space-filling model with van der Waals spheres was introduced by Corey, Pauling, and Koltun (Corey and Pauling 1953). Solvent-accessible surface (SAS) and solvent-excluded surface were proposed (Lee and Richards 1971; Richards 1977) to provide a more elaborate 3D description of biomolecular structures. However, these surface definitions admit geometric singularities, which lead to computational instability. Smooth surfaces, including Gaussian surfaces (Blinn 1982; Duncan and Olson 1993; Zheng et al. 2012; Chen et al. 2012; Li et al. 2013), skinning surfaces (Cheng and Shi 2009), minimal molecular surface (Bates et al. 2008), and flexibility–rigidity index (FRI) surfaces (Xia et al. 2013; Nguyen et al. 2016), were constructed to mitigate the computational difficulty.

Another important property of macromolecules is their structural instability or flexibility. Such property measures macromolecular intrinsic ability to respond to external stimuli. Flexibility is known to be crucial for biomolecular binding, reactivity, allosteric signaling, and order–disorder transition (Ma 2005). It is typically studied by standard techniques, such as normal mode analysis (NMA) (Go et al. 1983; Tasumi et al. 1982; Brooks et al. 1983; Levitt et al. 1985,) Gaussian network model (GNM) (Bahar et al. 1997), and anisotropic network model (ANM) (Atilgan et al. 2001). These methods have the computational complexity of $\mathcal{O}\left(N^3\right)$, with N being the number of unknowns. As a geometric graph-based method, FRI was introduced to reduce the computational complexity and improve the accuracy of GNM (Xia et al. 2013; Opron et al. 2014). NMA and ANM offer the collective

motions which are manifested in normal modes and may facilitate the functionally important conformational variations of macromolecules.

The aforementioned Gaussian surface or FRI surface defines a manifold structure embedded in 3D, which makes the analysis of geometry and topology accessible by differential geometry and algebraic topology. Recently, differential geometry has been introduced to understand macromolecular structure and function (Feng et al. 2012; Xia et al. 2014). In general, the protein surface has many atomic scale concave and convex regions which can be easily characterized by Gaussian curvature and/or mean curvature. In particular, the concave regions of a protein surface at the scale of a few residues are potential ligand-binding pockets. Differential geometry-based algorithms in both Lagrangian and Cartesian formulations have been developed to generate multiscale representations of biomolecules. Recently, a geometric flow-based algorithm has been proposed to detect protein-binding pockets by Zhao et al. (2018). Morse functions and Reeb graphs are employed to characterize the hierarchical pocket and sub-pocket structure (Zhao et al. 2018; Dey et al. 2013).

More recently, persistent homology (Carlsson et al. 2005; Edelsbrunner and Harer 2010), a new branch of algebraic topology, has become a popular approach for the topological simplification of macromolecular structural complexity (Yao et al. 2009; Xia and Wei 2014; Xia et al. 2015). Topological invariants are macromolecular-connected components, rings, and cavities. Topological analysis is able to unveil the topology–function relationship, such as ion channel open/close, ligand binding/disassociation, and protein folding–unfolding. However, persistent homology neglects chemical and biological information during its geometric abstraction. Element-specific persistent homology has been introduced to retain crucial chemical and biological information during the topological simplification (Cang and Wei 2018). It has been integrated with deep learning to predict various biomolecular properties, including protein–ligand-binding affinities and protein folding stability changes upon mutation by Cang and Wei (2017).

It is interesting to note that most current theoretical models for macromolecules are built from classical mechanics, namely computational electromagnetics, fluid mechanics, elasticity theory, and molecular mechanics based on Newton’s law. These approaches lead to multivalued scalar, vector, and tensor fields, such as macromolecular electrostatic potential, ion channel flow, protein anisotropic motion, and molecular dynamics trajectories. Biomolecular cryogenic electron microscopy (cryo-EM) maps are also scalar fields. Mathematically, macromolecular multivalued scalar, vector, and tensor fields contain rich geometric, topological, stability, flexibility, and Hodge mode information that can be analyzed to reveal molecular function. Unfortunately, unified geometric and topological analysis of macromolecular multivalued fields remains scarce. It is more challenging to establish a unified mathematical framework to further analyze macromolecular flexibility and Hodge modes. There is a pressing need to develop a unified theory for analyzing the geometry, topology, flexibility, and collective motion of macromolecules so that many existing methods can be calibrated to better uncover macromolecular function, dynamics, and transport.

The objective of the present work is to construct a unified theoretical paradigm for analyzing the geometry, topology, flexibility, and Hodge mode of macromolecules in order to reveal their function, dynamics, and transport. To this end, we introduce de Rham–Hodge theory for the modeling and analysis of macromolecules. De Rham–Hodge theory is a cornerstone of contemporary differential geometry, algebraic topology, geometric algebra, and spectral geometry (Hodge 1989; Bott and Tu 2013; Mitchell 1998). It provides not only the Helmholtz–Hodge decomposition to uncover the interplay between geometry and topology and the conservation of certain physical observables, but also the spectral representation of the underlying multivalued fields, which further unveils the geometry and topology. Specifically, as a ubiquitous computational tool, the Helmholtz–Hodge decomposition of various vector fields, such as electromagnetic fields by Hekstra et al. (2016), velocity fields by De La Torre and Bloomfield (1977), and deformation fields by Atilgan et al. (2001), can reveal their underlying geometric and topological features (see a survey by Bhatia et al. (2013)). Additionally, de Rham–Hodge theory interconnects classic differential geometry, algebraic topology, and partial differential equation (PDE) and provides a high-level representation of vector calculus and the conservation law in physics. Finally, the spectra of Laplace–de Rham operators in various differential forms also contain the underlying geometric and topological information and provide a starting point for the theoretical modeling of macromolecular flexibility and Hodge modes. The corresponding computational tool is discrete exterior calculus (DEC) (Hirani 2003; Desbrun et al. 2005; Arnold et al. 2006; Zhao et al. 2019). Lim discussed discrete Hodge Laplacians on graphs, which might not recover all the properties of the Laplace–de Rham operator (Lim 2015). De Rham–Hodge theory has had great success in theoretical physics, such as electrodynamics, gauge theory, quantum field theory, and quantum gravity. However, this versatile mathematical tool has not been applied to biological macromolecules, to the best of our knowledge. The proposed de Rham–Hodge framework seamlessly unifies previously developed differential geometry, algebraic topology, spectral graph theory, and PDE-based approaches for biological macromolecules (Xia and Wei 2016). Our specific contributions are summarized as follows:

- We provide a spectral analysis tool based on the de Rham–Hodge theory to extract geometric and topological features of macromolecules. In addition to the traditional spectra of scalar Hodge Laplacians, we enrich the spectra by using vector Hodge Laplacians with various boundary conditions.
- We construct a de Rham–Hodge theory-based analysis tool for the orthogonal decomposition of various vector fields, such as electric field, magnetic field, velocity field from molecular dynamics and displacement field, associated with macromolecular modeling, analysis, and computation.
- We propose a novel multiscale flexibility model based on the spectra of various Laplace–de Rham operators. This new method is applied to the Debye–Waller factor prediction of a set of 364 proteins (Opron et al. 2014). By comparison with experimental data, we show that our new model outperforms GNM, the standard bearer in the field (Bahar et al. 1997; Opron et al. 2014).

- We introduce a multiscale Hodge mode model by constraining a vector Laplace–de Rham operator with a Helfrich curvature potential. The resulting Laplace–de Rham–Helfrich operator is applied to analysing the Hodge modes of cryo-EM data. Unlike previous normal mode analysis which assumes harmonic potential around the equilibrium, our approach allows unharmonic motions far from the equilibrium. The multi-resolution nature of the present method makes it a desirable tool for the multiscale analysis of macromolecules, protein complexes, subcellular structures, and cellular motions.
- We demonstrate electrostatic field analysis based on Hodge decomposition and eigenfield analysis. The eigenfield analysis is applied to the reaction potential calculated by solving the Poisson–Boltzmann equation. We show that local dominant Hodge eigenfields exist for electrostatic analysis.

2 Results

Our results are twofold: We first describe our contribution to computational tools for Laplace–de Rham operators based on the simplicial tessellation of volumes bounded by biomolecular surfaces and then we present the modeling and analysis of de Rham–Hodge theory for biological macromolecules.

2.1 Theoretical Modeling and Analysis

This section introduces de Rham–Hodge theory for the analysis of biomolecules. To establish notation, we provide a brief review of de Rham–Hodge theory. Then, we introduce topological structure-preserving analysis tools, such as discrete exterior calculus (DEC) (Desbrun et al. 2005), discretized differential forms, and Hodge–Laplacians, on the compact manifolds enclosing biomolecular boundaries. We use simple finite-dimensional linear algebra to computationally realize our structure-preserving analysis on various differential forms. We construct appropriate physically relevant boundary conditions on biomolecular manifolds to facilitate various scalar and vector Laplace–de Rham operators such that the resulting spectral bases are consistent with three basic singular value decompositions of the gradient, curl and divergence operators through dualities.

2.1.1 De Rham–Hodge Theory for Macromolecules—While the spectral analysis can be carried out using scalar, vector, and tensor calculus, differential forms and exterior calculus are required in de Rham–Hodge theory to reveal the intrinsic relations between differential geometry and algebraic topology on biomolecular manifolds. Since biomolecular shapes can be described as 3-manifolds with a 2-manifold boundary in the 3D Euclidean space, we represent scalar and vector fields on molecular manifolds as well as their interconversion through differential forms. As a generalization of line integral and flux calculation of vector fields, a differential k-form $w^k \in \Omega^k(M)$ is a field that can be integrated on a k-dimensional submanifold of M, which can be mathematically defined through a rank-k antisymmetric tensor defined on a manifold M. By treating it as a multi-linear map from k vectors spanning the tangent space to a scalar, it turns an infinitesimal k-dimensional cell into a scalar, whose sum over all cells in a tessellation of a k-dimensional submanifold produces the integral in the limit of infinitesimal cell size. In $\mathbb{R}^3$, 0-forms and 3-forms have

one degree of freedom at each point and can be regarded as scalar fields, while 1-forms and 2-forms have three degrees of freedom and can be interpreted as vector fields.

The differential operator (also called exterior derivative) d can be seen as a unified operator that corresponds to gradient (∇), curl ($\nabla\times$), and divergence (∇) when applied to 0-, 1-, and 2-forms, mapping them to 1-, 2-, and 3-forms, respectively. On a boundaryless manifold, a codifferential operator δ is the adjoint operator under L_2-inner product of the fields (integral of pointwise inner product over the whole manifold), which corresponds to $-\nabla\cdot$, $\nabla\times$, and $-\nabla$, for 1-, 2-, and 3-forms, mapping them to 0-, 1-, and 2-forms, respectively.

One key property of $d:\Omega^k(M)\to\Omega^{k+1}(M)$ is that $dd=0$, which allows the space of differential forms Ω^k to form a chain complex, which is called the de Rham complex

$$0\to\Omega^0(M)\xrightarrow[(\nabla)]{d}\Omega^1(M\xrightarrow[(\nabla\times)]{d}\Omega^2(M)\xrightarrow[(\nabla\cdot)]{d}\Omega^3(M)\xrightarrow{d}0. \tag{1}$$

It also matches the identities of second derivatives for vector calculus in $\mathbb{R}^3$, i.e., $(\nabla\times)\nabla=0$ and $(\nabla\cdot)\nabla\times=0$. The topological property associated with differential forms is given by the de Rham cohomology,

$$H^k_{dR}(M)=\frac{\ker d^k}{\mathrm{im}\,d^{k-1}}. \tag{2}$$

The de Rham theorem states that the de Rham cohomology is isomorphic to the singular cohomology, which is derived purely from the topology of the biomolecular manifold.

The Hodge k-star $\star^k$ (also called Hodge dual) is a linear map from a k-from to its dual form, $\star^k:\Omega^k(M)\to\Omega^{n-k}(M)$. Given two k-forms $\alpha,\beta\in\Omega^k(M)$, the ($L_2$-)inner product between them can be defined along with star operator as

$$\left\langle\alpha,\beta\right\rangle=\int_M\alpha\wedge\star\beta=\int_M\beta\wedge\star\alpha. \tag{3}$$

Under the inner products, the adjoint operators of d are the codifferential operators $\delta^k:\Omega^k(M)\to\Omega^{k-1}(M)$, $\delta^k=(-1)^k\star d\star$ satisfies $\delta\delta=0$. Hodge further established the isomorphism

$$H^k_{dR}(M)\cong H^k_\Delta(M), \tag{4}$$

where $H^k_\Delta(M)=\left\{\omega\mid\Delta\omega=0\right\}$ is the kernel of the Laplace–de Rham operator $\Delta\equiv d\delta+\delta d=(d+\delta)^2$, also known as the space of harmonic forms. A corollary of the result is the Hodge decomposition,

$$\omega=d\alpha+\delta\beta+h, \tag{5}$$

which is an L_2-orthogonal decomposition of any form ω into d and δ of two potential fields $\alpha \in \Omega^{k-1}(M)$ and $\beta \in \Omega^{k+1}(M)$ respectively, and a harmonic form $h \in H^k_\Delta(M)$. This means that harmonic forms are the non-integrable parts of differential forms, which form a finite-dimensional space determined by the topology of the biomolecular domain due to de Rham's and Hodge's theorems.

2.1.2 Macromolecular Spectral Analysis—The Laplace–de Rham operator $\Delta = d\delta + \delta d$, when restricted to a 3D object embedded in the 3D Euclidean space, is simply $-\nabla^2$. As it is a self-adjoint operator with a finite-dimensional kernel, it can be used to build spectral bases for differential forms. For irregularly shaped objects, these bases can be very complicated. However, for simple geometry, these bases are well-known functions. For example, 0-forms on a unit circle can be expressed as the linear combination of sine and cosine functions, which are eigenfunctions of the Laplacian for 0-forms Δ_0. Similarly, spherical harmonics are eigenfunctions of Δ_0 on a sphere and it has also been extended to manifold harmonics on Riemannian 2-manifolds.

We further extend the analysis to any rank k and to 3D shapes such as macromolecular shapes where analysis can be carried out in two types of cases. In the first type, one may treat the surface of the molecular shape as a boundaryless compact manifold and analyzes any field defined on such a 2D surface. In fact, this approach is relevant to protein surface electrostatic potentials or the behavior of cell membrane or mitochondrial ultrastructure. In this work, we shall restrain from any further exploration in this direction. In the second type, we consider the volumetric data enclosed by a macromolecular surface. As a result, the molecular shape has a boundary. In this setting, the harmonic space becomes infinite-dimensional unless certain boundary conditions are enforced. In particular, tangential or normal boundary conditions (also called Dirichlet or Neumann boundary conditions, respectively) are enforced to turn the harmonic space into a finite-dimensional space corresponding to algebraic topology constructions that lead to absolute and relative homologies.

We first discuss the natural separation of the eigenbasis functions into curl-free and div-free fields in the continuous theory, assuming that the boundary condition is implicitly enforced, before providing details on the discrete exterior calculus with the boundary taken into consideration.

Given any eigenfield ω of the Laplacian,

$$\Delta\omega = \lambda\omega, \tag{6}$$

we can decompose it into $\omega = d\alpha + \delta\beta + h$. For $\lambda \neq 0$, based on $dd = 0$ and $\delta\delta = 0$, it is easy to see that both $d\alpha$ and $\delta\beta$ are eigenfunctions of Δ with eigenvalue λ due to the uniqueness of the decomposition, unless one of them is 0. It is typically the case that ω is either a curl field or a gradient field; otherwise, λ has a multiplicity of at least 2, in which case both eigenfields associated with λ are the linear combinations of the same pair of the gradient field and the curl field.

2.1.3 Discrete Spectral Analysis of Differential Forms—In a simplicial tessellation of a manifold mesh, d^k is implemented as a matrix D_k, which is a signed incidence matrix between (k+1)-simplices and k-simplices. We provide the details in Sect. 3, but the defining property in de Rham–Hodge theory is preserved through such a discretization: $D_{k+1}D_k = 0$. The adjoint operator δ^k is implemented as $S_{k-1}^{-1}D_{k-1}^T S_k$, where S_k is a mapping from a discrete k-form to a discrete $(n-k)$-form on the dual mesh, which can be treated as a discretization of the L_2-inner product of k-forms. As S_k is always a symmetric positive matrix, the L_2-inner product between two discrete k-forms can be expressed as $\left(\omega_1^k\right)^T S_k \omega_2^k$. The discrete Hodge Laplacian maps a discrete k-form to a discrete $n-k$-form which is defined as

$$L_k = D_k^T S_{k+1} D_k + S_k D_{k-1} S_{k-1}^{-1} D_{k-1}^T S_k, \tag{7}$$

which is a symmetric matrix and $S_k^{-1}L_k$ corresponds to Δ_k. The eigenbasis functions are found through a generalized eigenvalue problem,

$$L_k \omega^k = \lambda^k S_k \omega^k. \tag{8}$$

Depending on whether the tangential or normal boundary condition is enforced, D_k includes or excludes the boundary elements, respectively. Thus, the boundary condition is built into discrete linear operators. When we need to distinguish these two cases, we use $L_{k,t}$ and $L_{k,n}$ to denote the tangential and normal boundary conditions, respectively.

In general, it is not necessarily efficient to take the square root of the discrete Hodge star operator, $S_k^{\frac{1}{2}}$ or to compute its inverse,S_k^{-1}. However, for analysis, we can always convert a generalized eigenvalue problem in Eq. (8) into a regular eigenvalue problem,

$$\bar{L}_k \bar{\omega}^k \equiv S_k^{-\frac{1}{2}} L_k S_k^{-\frac{1}{2}} \bar{\omega}^k = \lambda^k \bar{\omega}^k, \tag{9}$$

where $\bar{\omega} \equiv S_k^{\frac{1}{2}}\omega$. We can further rewrite the symmetrically modified Hodge Laplacian as

$$\bar{L}_k = \bar{D}_k^T \bar{D}_k + \bar{D}_{k-1} \bar{D}_{k-1}^T, \tag{10}$$

where $\bar{D}_k \equiv S_{k+1}^{\frac{1}{2}} D_k S_k^{-\frac{1}{2}}$ must satisfy $\bar{D}_{k+1}\bar{D}_k = 0$. Now the L_2-inner product between two discrete differential forms in the modified space is simply $\left(\bar{\omega}_1^k\right)^T \bar{\omega}_2^k$, and the adjoint operator of $\bar{D}_k$ is simply $\bar{D}_k^T$.

Now the partitioning of the eigenbasis functions into harmonic fields, gradient fields, and curl fields for 1-forms and 2-forms and their relationship can be understood from the singular value decomposition of the differential operator

$$\overline{D}_k = U_{k+1}\Sigma_k V_k^T, \tag{11}$$

where U_{k+1} and V_k are orthogonal matrices and Σ_k is a rectangular matrix that only has nonzero entries on the diagonal, which can be sorted in ascending order as $\sqrt{\lambda_i^k}$ with trailing zeros. As the Hodge decomposition is an orthogonal decomposition, each column of V_k that corresponds to a nonzero singular value $\sqrt{\lambda_i^k}$ is orthogonal to any column of U_k that corresponds to a nonzero $\sqrt{\lambda_j^{k-1}}$. Here, V_k and U_k, together with the finite-dimensional set of harmonic forms h_k (which satisfy both $D_k h_k = 0$ and $D_{k-1}^T h_k = 0$), span the entire space of k-forms. Moreover, the spectrum (i.e., set of eigenvalues) of the symmetric modified Hodge Laplacian in Eq. (10) consists of 0s, the set of λ_i^k's, and the set of λ_j^{k-1}'s. Note that in the spectral basis, taking derivatives $\overline{D}$ (or $\overline{D}^T$) is simply performed through multiplying the corresponding singular values, and integration is done through division by the corresponding singular values, mimicking the situation in the traditional Fourier analysis for scalar fields.

2.1.4 Boundary Conditions and Dualities in 3D Molecular Manifolds—Overall, appropriate boundary conditions are prescribed to preserve the orthogonal property of the Hodge decomposition. In 3D molecular manifolds, 0- and 3-forms can be seen as scalar fields and 1- and 2-forms as vector fields. For the spectral analysis of scalar fields (0-forms or 3-forms), two types of typical boundary conditions are used: Dirichlet boundary condition $f|_{\partial M} = f_0$ and Neumann boundary condition $\mathbf{n}\cdot\nabla f|_{\partial M} = g_0$, where f_0 and g_0 are functions on the boundary $\partial \boldsymbol{M}$ and $\mathbf{n}$ is the unit normal on the boundary. For spectral analysis, harmonic fields satisfying the arbitrary boundary conditions can be dealt with through spectral analysis of f_0 or g_0 on the boundary, and the following boundary conditions are used for the volumetric function f. The normal 0-forms (tangential 3-forms) satisfy

$$f|_{\partial M} = 0, \tag{12}$$

and the tangential 0-forms (normal 3-forms) satisfy

$$\mathbf{n}\cdot\nabla f|_{\partial M} = 0. \tag{13}$$

For the spectral analysis of vector fields, boundary conditions are for the three components of the field. Based on the de Rham–Hodge theory, it is more convenient to also use two types of boundary conditions. For tangential vector field (representing tangential 1-forms or normal 2-forms) $\mathbf{v}$, we use the Dirichlet boundary condition for the normal component and the Neumann condition for the tangential components:

$$\mathbf{v}\cdot\mathbf{n} = 0,\ \mathbf{n}\cdot\nabla(\mathbf{v}\cdot\mathbf{t}_1) = 0,\ \mathbf{n}\cdot\nabla(\mathbf{v}\cdot\mathbf{t}_2) = 0, \tag{14}$$

where $\mathbf{t}_1$ and $\mathbf{t}_2$ are two local tangent directions forming a coordinate frame with the unit normal $\mathbf{n}$. The corresponding spectral fields are shown in Fig. 1. For normal vector field (representing normal 1-forms or tangential 2-forms) $\mathbf{v}$, we use the Neumann boundary condition on the normal component and the Dirichlet boundary condition on the tangential components:

$$\mathbf{v} \cdot \mathbf{t}_1 = 0, \ \mathbf{n} \cdot \mathbf{t}_2 = 0, \ \mathbf{n} \cdot \nabla(\mathbf{v} \cdot \mathbf{n}) = 0 . \tag{15}$$

The corresponding spectral fields are shown in Fig. 2. Aside from the harmonic spectral fields, there are two types of fields involved for the spectral fields of both boundary conditions—the set of divergence-free fields (also called curl fields) and the set of curl-free fields (also called gradient fields). In summary, the above four boundary conditions account for both types of boundary conditions of all four differential forms, since the tangential boundary conditions of k-forms are equivalent to the normal boundary conditions of $n-k$-forms.

2.1.5 Reduction and Analysis—For the four types of k-forms ($k \in \{0, 1, 2, 3\}$ in $\mathbb{R}^3$) in combinations with the two types of boundary conditions (tangential and normal), there are eight different Laplace–de Rham operators ($L_{k,t}$ and $L_{k,n}$) in total. However, based on Eq. (10), the nonzero parts of the spectrum L_k can be assembled from the singular values of $\overline{D}_k$ and $\overline{D}_{k-1}$. Thus, for each type of boundary conditions, there are only three spectra associated with $\overline{D}_0$, $\overline{D}_1$, and $\overline{D}_2$, since $\overline{D}_3 = 0$ for 3D space. (One still has eight Laplace–de Rham operators.) Moreover, according to the Hodge duality discussed in the above paragraph, there is a one-to-one map between tangential k-forms and normal $(3-k)$-forms, which further identifies $\overline{D}_{0,t}$ with $\overline{D}^T_{2,n}$, $\overline{D}_{0,n}$ with $\overline{D}^T_{2,t}$, and $\overline{D}^T_{1,n}$ with $\overline{D}^T_{1,n}$. As a result, one has four independent Laplace–de Rham operators. Finally, due to the self-adjointness, there are only three intrinsically different spectra: (1) The first contains singular values of the gradient operator $D_{0,t}$ on tangential scalar potential fields (or equivalently, the singular values of the divergence operator $D_{2,n}$ on tangential gradient fields) as shown in Fig. 3b; (2) the second contains singular values of the gradient operator $D_{0,n}$ on normal scalar potential fields (or equivalently, the singular values of the divergence operator $D_{2,t}$ on normal gradient fields) as shown in Fig. 3c; and (3) the third contains singular values of the curl operator $D_{1,t}$ applied to tangential curl fields (or equivalently, the singular values of the curl operator $D_{1,n}$ applied to normal curl fields) as shown in Fig. 3d.

As discussed above, each of the eight Hodge Laplacians defined for smooth fields on a smooth shape has a spectrum that is simply the combination of one or two of the three sets of singular values along with possibly a 0. However, the numerical evaluation of the singular values of the differential operators for tangential k-forms $\overline{D}_{k,t}$ can differ from those of the discrete operators for normal $3-k$-forms $\overline{D}^T_{2-k,n}$, as shown in Fig. 3d. One immediate reason is that the degrees of freedom (DoFs) associated with tangential/normal scalar/vector fields represented as tangential forms are not the same as those represented by normal forms on a given tessellation, leading to different sampling accuracies. For example, the

tessellation of the shape in Fig. 3 consists of approximately 1000 vertices, 7000 edges, 10,000 triangles, and 5000 tetrahedra. Thus, each tangential 0-form only has 1000 DoFs, and each normal 3-form has 5000. Hence, $L_{3,n}$ is capable of handling higher-frequency signals in any given smooth scalar field than $L_{0,t}$ when we approach the Nyquist frequencies of the sampling. The convergence of both discretizations for the same continuous operator can be observed with increasing DoFs for both differential forms under refinement of the tet meshes (Fig. 3e, left). For low frequencies (smallest eigenvalues), there is a good agreement to begin with (Fig. 3e, middle), while for any given high frequency, the convergence with increased resolutions can be clearly observed (Fig. 3e, right).

On the other hand, $\bar{D}_k \bar{D}_k^T$ and $\bar{D}_k^T \bar{D}_k$ will have strictly the same set of nonzero eigenvalues. For instance, the spectrum of $L_{0,t}$ and the partial spectrum of $L_{1,t}$ that correspond to gradient fields are identical, since $\bar{D}_{0,t} \bar{D}_{0,t}^T$ and $\bar{D}_{0,t}^T \bar{D}_{0,t}$ have the same nonzero eigenvalues.

For eigenfields vector Laplacians represented as 1-forms or 2-forms, i.e., the eigenfields of L_1 or L_2, we can observe some typical traits in the distributions of eigenvalues under normal and tangential boundary conditions. The normal boundary condition tends to allow more gradient eigenfields associated with eigenvalues below a given threshold than those under the tangential boundary condition for eigenvalues below the same threshold. We conjecture that it is due to the more stringent Dirichlet boundary condition on the potential scalar fields than the Neumann boundary condition. The relation between the tangential boundary condition gradient-type eigenfields and curl-type eigenfields for low-frequency range seems to be highly dependent on the shape (Fig. 3b, d). Figure 1 shows different vector eigenfields for tangential boundary condition with EMD 7972 surface. The first row shows different harmonic fields corresponding to the number of handles of the shape, the second row shows different gradient fields, and the third row shows different curl fields. Figure 2 shows different vector eigenfields for normal boundary condition with the protein and DNA complex crystal structure 6D6V. Since there are no cavities for this shape, there are no harmonic fields. The first row shows different gradient fields, and the second row shows different curl fields. Note that the scalar potentials for gradient fields and the vector potentials for curl fields are also themselves eigenfields associated with the same eigenvalues, although for different Laplacians.

Summarizing the above discussion on the properties of Laplacian spectra for 3D shape, we propose the following suggestions for practical spectral analysis:

- Only three independent spectra (e.g., singular values of $D_{0,t}$, $D_{1,t}$, and $D_{2,t}$) are necessary to avoid redundancy.
- Laplace–de Rham operators with higher DoFs can be used for more accurate calculation (at a higher computational cost) given the same tessellation.
- When computing eigenvalues given the same high-frequency truncation threshold, the differences in the numbers of eigenvalues in the three spectra vary with the shape.

2.2 Macromolecular Modeling and Analysis

Biological macromolecules and their complexes offer a rich variety of geometric and topological features, which often exhibit close relations with their functionalities. For instance, protein pockets can often be identified as a geometrically concave region on the protein surface, or as a topological cavity of an offset surface. Ion channels that regulate important biological functions can be usually associated with a topological tunnel. Mitochondrial ultrastructures admit various geometric and topological complexity which is related to their functions (Wollenman et al. 2017). Hence, a unified approach for quantitatively analyzing such geometric and topological features is in great need. Our de Rham–Hodge analysis and Laplace–de Rham operator modeling provide such a unified approach for capturing both geometric and topological features simultaneously.

Our de Rham–Hodge analysis offers a powerful new tool for characterizing macromolecular geometry, identifying macromolecular topology, and modeling macromolecular structural flexibility and collective motion. We have carried out extensive computational experiments using protein structural datasets and cryo-EM maps to demonstrate the utility and usefulness of the proposed de Rham–Hodge tools and models.

2.2.1 Molecular Shape Generation—The geometric modelling of macromolecular 3D shapes bridges the gap between experimental data and theoretical models for macromolecular function, dynamics, and transport. To carry out our de Rham–Hodge analysis on a macromolecule or a protein complex, we need a given domain containing the 3D macromolecular shape. Theoretically, such a domain for a macromolecule can be generated by taking an isosurface of a cryo-EM map or constructed from the atomic coordinates of the macromolecule. For a given set of atomic coordinates $\mathbf{r}_i$, $i = 1, 2, \ldots, N$, van der Waals surface, solvent-accessible surface, and the solvent-excluded surface can be constructed. However, these surfaces are typically singular, leading to computational instability for de Rham–Hodge analysis. Alternatively, minimal molecular surface (MMS) generated by differential geometry, Gaussian surface (Li et al. 2013), and flexibility rigidity index (FRI) surface (Xia et al. 2013; Opron et al. 2014) are computationally preferred and used widely in many studies. In fact, FRI surface is simpler than MMS and more stable than Gaussian surface (Nguyen et al. 2016). To generate an FRI surface, we use a discrete-to-continuum mapping to define an unnormalized molecular density (Xia et al. 2013; Nguyen et al. 2016)

$$\rho(\mathbf{r}, \eta) = \sum_{j=1}^{N} \Phi\big(\|\mathbf{r} - \mathbf{r}_j\|; \eta\big) \tag{16}$$

where η is a scale parameter and in this paper, it is set to twice of the atomic van der Waals radius r_j. Φ is density estimator that satisfies the following admissibility conditions

$$\Phi\big(\|\mathbf{r} - \mathbf{r}_j\|; \eta\|\big) = 1, \text{ as } \|\mathbf{r} - \mathbf{r}_j\| \to 0, \tag{17}$$

$$\Phi\big(\|\mathbf{r} - \mathbf{r}_j\|; \eta\|\big) = 0, \text{ as } \|\mathbf{r} - \mathbf{r}_j\| \to \infty\,. \tag{18}$$

Monotonically decaying radial basis functions are all admissible. Commonly used correlation kernels include generalized exponential functions

$$\Phi\left(\|\mathbf{r}-\mathbf{r}_j\|;\eta\|\right)=e^{-\left(\|\mathbf{r}-\mathbf{r}_j\|/\eta\right)^{\kappa}},\ \kappa>0, \tag{19}$$

and generalized Lorentz functions

$$\Phi\left(\|\mathbf{r}-\mathbf{r}_j\|;\eta\right)=\frac{1}{1+\left(\|\mathbf{r}-\mathbf{r}_j\|/\eta\right)^{\upsilon}},\ \upsilon>0. \tag{20}$$

The Gaussian kernel ($\kappa=2$) is employed in this work.

A family of biomolecular domains can be defined by varying level set parameters $c>0$

$$M=\{\mathbf{r}\mid\rho(\mathbf{r},\eta)\geq c\}. \tag{21}$$

2.2.2 Topological Analysis—In this work, we discuss topology in the mathematical sense. Therefore, topological features are those stable structural characteristics that do not change with deformation, such as the number of connected components, the number of holes on each connected component, and the number of cavities. They are captured in the null spaces of the corresponding Laplace–de Rham operators. In other words, the invariant spaces associated with the eigenvalue of 0, i.e., the lowest ends of the spectra. Specifically, the dimension of the null space of $L_{1,t}$ and $L_{2,n}$ is the same as the number of tunnels as shown in Fig. 4a. The dimension of the null space of $L_{1,n}$ and $L_{2,t}$ provides the number of cavities as shown in Fig. 4b. The dimension of the $L_{0,t}$ is equal to the number of connected components. In persistent homology, the geometric measurement for characterizing the persistence of a topological feature has been proven crucial to the practical use of these otherwise overly stable features. The eigenfields associated with the eigenvalue 0 in our spectral analysis can also provide such information. For instance, the strength of the eigenfield associated with the eigenvalue 0 for $L_{1,t}$ can indicate how narrow the handle/tunnel is in the region. In the tangential harmonic fields of Fig. 1, the colors show the strength of eigenfields such that red colors stand for high strengths and indigo colors stand for low strengths. One can see that strengths are higher in the middle narrow tunnels than the top and bottom parts.

2.2.3 Geometric Analysis—Although the spectra of the Laplace–de Rham operators do not uniquely determine the geometry (sometimes referred to as "you cannot hear the shape of the drum"), they do provide key information when comparing shapes, which, sometimes, is referred to as shape "DNA". Thus, the traits of the nonzero parts of the spectra can be regarded as geometrical features. These geometrical features are rigid transformation invariant. The scalar Hodge Laplacian spectrum has already been used in computer graphics and computer vision to distinguish various structures in shape analysis and shape retrieval. It has also been extended to 1-form Hodge Laplacian on surfaces for shape analysis. However, on surfaces, L_1 spectrum is identical to L_0 spectrum, except that the multiplicity is doubled for nonzero eigenvalues. Note that the multiplicity for the zero eigenvalues is determined by

the number of genus instead of the number of connected components for scalar Hodge Laplacian. In our 3D extension, we have three unique spectra for each molecule. Figure 5 shows nonzero spectrum traits for three simple proteins (PDB IDs: 2Z5H (Murakami et al. 2008), 6HU5 (Lanza et al. 2019), and 5HY9 (Kuglstatter et al. 2017), where the clear distinction among the spectra can be observed. We have tested on various biomolecules and observed the same discriminating ability of the spectra on these shapes.

Geometric analysis and topological analysis based on the de Rham–Hodge theory can be readily applied to characterizing biomolecules in machine learning and to biomolecular modeling. To further demonstrate the capability of de Rham–Hodge spectral analysis for macromolecular analysis, we propose a set of de Rham–Hodge models for protein flexibility analysis and a vector de Rham model for biomolecular Hodge mode analysis.

2.2.4 Flexibility Analysis—Biomolecular flexibility analysis and B-factor prediction have been commonly performed by normal mode analysis (Go et al. 1983; Tasumi et al. 1982; Brooks et al. 1983; Levitt et al.1985; Ma2005) and Gaussian network model (GNM) by Bahar etal. (1997). The flexibility is strongly correlated with protein functions, such as structural support, catalyzing chemical reactions, and allosteric regulation (Frauenfelder et al. 1991). Recently, graph theory-based FRI has been shown to outperform other methods (Opron et al. 2014). However, all of the aforementioned methods are based on the discrete coordinate representation of biomolecules. As such, it is not very convenient to use these methods for flexibility analysis at different scales. For example, for some large macromolecules, such as an HIV viral capsid which involves millions of atoms, one may wish to analyze their flexibility at atomic, residue, protein domain, protein, and protein complex scales by using a unified approach so that the results from cross-scales can be compared on an equal footing. However, current approaches cannot provide such a unified cross-scale flexibility analysis. In this work, we introduce ade Rham–Hodge theory-based model to quantitatively analyze macromolecular flexibility across many scales.

We assume that the de Rham–Hodge B-factor at the ith atom estimated by L_k is given by

$$B_{k,i}^{\mathrm{dRH}} = a \sum_j \frac{1}{\lambda_j^k} \left[\omega_j^k(\mathbf{r}) \left(\omega_j^k(\mathbf{r}') \right)^T \right]_{\mathbf{r}=\mathbf{r}_i, \mathbf{r}'=\mathbf{r}_i}, \forall \lambda_j^k > 0, \tag{22}$$

where a is a parameter to be determined by the least squares regression. Its value depends on structural resolution, diffraction intensity, experimental method (i.e., x-ray scattering, electron microscopy, etc.), number of diffraction angles, experimental temperature, sample quality, and structure reconstruction method. In the computation, the value of $\omega_j^k(\mathbf{r})$ is given on a set of mesh points. The linear regression over a cutoff radius d is used to obtain the required values in atomic centers $\mathbf{r}_i$ where the B-factor values are reported. $L_{0,t}$ is applied in test cases.

We perform numerical experiments to confirm that our flexibility analysis on C-alpha atoms is robust and reliable. In fact, our method can analyze the flexibility of all atoms or a subset of atoms. The cutoff radius is set to 7 Å. Our method involves several parameters including level set value c and grid spacing r and cutoff radius d (Fig. 6). Figure 7 shows statistics of

the average Pearson correlation coefficient with various parameters on the test set of 364 proteins.

Level Set: The level set parameter c in Eq. (21) controls the general distance from the surface to C-alpha atoms (Fig. 6a). A larger level set value will result in a smaller domain with richer topology structures, including many tunnels and cavities. A smaller level set value will make the surface fatter so that it will lead to a ball-like shape.

Grid Spacing: The grid spacing r controls the density of tetrahedrons of the mesh. A finer mesh will lead to a better prediction but is computationally more expensive (Fig. 6b).

Cutoff Radius: The parameter cutoff radius d controls the linear regression region around the specific C-alpha atom (tets within the radius d to the specific C-alpha atom which is colored purple in Fig. 6d). Our approach will potentially introduce a denser mesh, which will lead to small local vibrations (high frequencies introduced due to the increasing number of matrix elements) that should be filtered out. This treatment is the same as throwing away higher frequencies.

We consider a benchmark test set of 364 proteins studied in earlier work by Opron et al. (2014) to systematically validate our method. Our test indicates that the best parameters are $c = 0.4$, $r = 1.6$ Å, $d = 4.0$ Å. Figure 8 shows several examples with the best parameters and comparisons with GNM. Table 1 shows the average Pearson correlation coefficient of predicting the benchmark set of 364 proteins Opron et al. (2014) at a cutoff radius 4.0 Å, which includes the overall best average Pearson correlation coefficient at grid spacing 1.6 Å and level set value 0.4. The contour level value should not be too large such that only those C-alpha atoms that are close enough to each other will have interactions, as well as not be too small such that enough geometric and topological features are preserved. The cutoff radius should be a proper value such that higher frequencies are mitigated, while lower frequencies are well kept. There is not much influence of resolution if the previous two parameters are well set (see statistics at cutoff radius 5 Å). This provides the foundation for analyzing large protein complexes with coarse resolution.

The proposed flexibility analysis can be easily extended to analyze the flexibility of cryo-EM data at given level sets. The computed (relative) B-factors are located at vertices but can be interpolated to any desirable location if necessary. Due to the multi-resolution nature of our approach, the computational cost is determined by the number of unknowns, i.e., the mesh size. For a given computational domain, the mesh size depends on the grid spacing. Therefore, for large macromolecules with millions of atoms, which is intractable for coordinate-based methods, the proposed de Rham–Hodge approach can still be very efficient.

The commonly used method that produces the B-factors that wind up in the PDB files is the least squares fit. This method connects diffraction intensity profiles and structural model predicted densities in the PDB with B-factors. In our model, we connect experimental structures (the coordinates of structural model predicted densities) and B-factors in the PDB files with our Hodge eigenvalue- and eigenvector-based model.

2.2.5 Hodge Mode Analysis—Normal mode analysis is an important approach for understanding biomolecular collective behavior, residue coupling, protein domain motion, and protein–protein interaction, reaction pathway, allosteric signaling, and enzyme catalysis (Go et al. 1983; Tasumi et al. 1982; Brooks et al. 1983; Levitt et al. 1985; Ma 2005). However, normal model analysis becomes very expensive for large biomolecules. In particular, it is difficult to carry out the anisotropic network model (ANM) analysis (Atilgan et al. 2001) for cryo-EM maps which do not have atomic coordinates. Virtual particle-based ANM methods were proposed to tackle this problem (Tama et al. 2002; Ming et al. 2002). Being based on the harmonic potential assumption, these methods are restricted to relatively small elastic motions. In this work, we propose an entirely different strategy for biological macromolecular anisotropic motion analysis based on de Rham–Hodge theory.

Laplace–de Rham Operator: It is noted that a mass–spring system is underlying many earlier successful elastic network models. This system describes the interconversion between the kinetic energy and potential energy during the dynamic motion. In our construction, we take advantage of de Rham–Hodge theory. In fact, de Rham–Hodge theory provides a general framework to model the dynamic behavior of macromolecules. In the present work, we just illustrate this approach with special construction.

In order for de Rham–Hodge theory to be able to describe anisotropic motions, we utilize the 1-form Laplace–de Rham operator

$$\Delta_1 = d_0 \star_0^{-1} d_0 \star_1 + \star_1^{-1} d_1 \star_2 d_1, \tag{23}$$

where d_k denote exterior derivatives on $\Omega^k(M)$ and $\star_k$ denote Hodge star operators. Note that the 2-form Laplace–de Rham operator works similarly well, but we will limit our discussion with 1-form. The first term on the right hand side of Eq. (23) is the quadratic energy form measuring the total divergence energy, while the second term measures the total curl energy. Both terms are kinetic energy physically or Dirichlet energy mathematically.

Laplace–de Rham–Helfrich Operator: Physically, a potential energy term is required to constrain the elastic motion of biological macromolecules. There are many options, such as Willmore energy, which minimize the difference between two principle curvatures. Additionally, Helfrich introduced a curvature energy for modeling cell membrane or closed lipid vesicles (Helfrich 1973; Du et al. 2004). In our case, we assume the curvature energy of the form

$$V = \mu \int_{\partial M} (H - H_0)^2 \, \mathrm{d}A, \tag{24}$$

where μ is the molecular bending rigidity, H is the mean curvature on the molecular surface, and H_0 is the spontaneous curvature of the molecule. The potential energy in Eq. (24) is defined on the compact manifold enclosing a smooth molecular surface.

Conceptually, our curvature model deals with a dynamical system with a thin shell having a thickness much smaller than other dimensions. Computationally, the 2D curvature model serves as a boundary condition to complete the Laplace–de Rham operator on a

macromolecule. The curvature energy increases as the mean curvature H deforms away from its rest state. Therefore, H is a function of surface displacement. The quadratic energy generated from surface deformation is given by (see Tamstorf and Grinspun (2013) for discretization details)

$$Q = \partial^2 V / \partial X^2, \tag{25}$$

where X is a displacement vector field on the surface. Due to the isomorphism between vector fields and 1-forms, we can evaluate the volumetric 1-form ω as a displacement vector field and restrict it to the boundary surface. We denote the restriction as a linear operator G,

$$X = G\omega\,. \tag{26}$$

Then, the quadratic form for the curvature energy in terms of the 1-form is $G^T QG$. Finally, the total 1-form quadratic energy is given by the following one-parameter Laplace–de Rham–Helfrich operator

$$E_\mu = d_0 \star_0^{-1} d_0^T \star_1 + \star_1^{-1} d_1^T \star_2 d_1 + G^T QG\,. \tag{27}$$

We can solve the eigenvalue problem for the Laplace–de Rham–Helfrich operator E_μ to extract the natural vibration modes of biomolecules. It is a standard procedure to assemble required matrix G and Q together with our Laplace–de Rham matrix.

In fact, an advantage of the proposed anisotropic motion theory is that it allows to treat the divergence energy and curl energy differently. For example, we can introduce a bulk modulus type of parameter λ to the divergence energy term, which leads to a weightedLaplace–deRhamoperator.As a result, we have a two-parameter Laplace–de Rham–Helfrich operator

$$E_{\lambda\mu} = \lambda \cdot d_0 \star_0^{-1} d_0^T \star_1 + \star_1^{-1} d_1^T \star_2 d_1 + G^T QG\,. \tag{28}$$

We need to choose appropriate weight parameters λ and μ. Generally, the two-parameter Laplace–de Rham–Helfrich operator and boundary condition matrix can be tuned separately. What we would like to achieve is letting the curvature energy drive the motion and let our system penalize the compressibility (i.e., the divergence energy). Therefore, we select an appropriate λ at a different scale and choose $\mu > \lambda > 1$.

Modal analysis, compared to fluctuation analysis, provides more information. In addition to the description of flexibility, modal analysis also provides the collective motion of a molecule and its potential function. The dynamics of a macromolecule can be described by the linear combination of its natural modes. Figure 9 shows several Hodge modes for core spliceosomal components, EMD 1258 (Sander et al. 2006), which indicates the success of our Laplace–de Rham–Helfrich operator.

It is noted that the original Laplace–de Rham operator with appropriate boundary conditions admits the orthogonal Hodge decomposition in terms of divergence-free, curl-free, and harmonic eigenmodes. In contrast, the Laplace–de Rham–Helfrich operator does not preserve these properties. Nonetheless, the eigenmodes generated by the Laplace–de Rham–Helfrich operator are mutually orthogonal and subject to different physical interpretations. For example, the first three eigenmodes are associated with 3D translational motions. Therefore, the operator is translational invariant. The modes in Figure 9 have little to do with the topological singularity of EMD 1258.

Additionally, the eigenmodes in Fig. 1 have a fixed boundary. In contrast, boundaries of eigenmodes generated with the Laplace–de Rham–Helfrich operator as shown in Fig. 9 are allowed to change. The Laplace–de Rham–Helfrich operator can predict significant macromolecular deformations, which are controllable with two weight parameters, λ and μ. In contrast, existing normal mode analysis methods can only admit small deformations due to the use of the harmonic potential.

Moreover, due to its continuous nature, the proposed Laplace–de Rham–Helfrich operator can be easily employed for the Hodge mode analysis at any given scale. It can be directly applied to the analysis of cryo-EM maps and other volumetric data at an arbitrary scale. One specific example of potential applications is the analysis of subcellular organelles, such as mitochondrial ultrastructure and endoplasmic reticulum.

Finally, the proposed Laplace–de Rham–Helfrich model is phenomenological in nature but can describe physical observations. Like the Navier–Stokes equation for fluid mechanics and the Ginzburg–Landau equation for superconductivity, Laplace–de Rham–Helfrich model is not rigorously derived from the fundamental laws of physics or first principles.

2.2.6 Field Decomposition and Analysis—Our Laplace–de Rham operators constructed from different boundary conditions can also perform vector field decomposition tasks. Following the discussion of boundary conditions in Sect. 2.1, a Hodge decomposition for a k-form bounded manifolds in 3D is constructed as $\omega^k = d\alpha_n^{k-1} + \delta\beta_t^{k+1} + h^k$, where α_n^{k-1} is in the space of normal (k–1)-forms Ω_n^{k-1}, β_t^{k+1} is in the space of tangential (k+1)-forms, and h^k is in H_Δ^k. Moreover, H_Δ^k is further decomposed based on boundary conditions and a five-component orthogonal decomposition (Cantarella et al. 2002) is given as

$$\omega^k = d\alpha_n^{k-1} + \delta\beta_t^{k+1} + h_t^k + h_n^k + \eta^k, \tag{29}$$

where h_t^k is a tangential harmonic form, h_n^k is a normal harmonic form, and η^k is central harmonic form which is both exact and coexact. There are naturally various vector fields existing in biomolecules, such as electric fields, magnetic fields, and elastic displacement fields. De Rham–Hodge theory can help provide a mutually orthogonal decomposition to investigate source, sink, and vortex features presented in those fields. An example of this analysis is given in Fig. 10 for a synthetic vector field on a vacuolar ATPase motor, EMD 1590 (Muench et al. 2009). We expect this decomposition becomes more interesting for biomolecular electric fields, dipolar fields, and magnetic fields. Various components from

the decomposition can be naturally used as the components of machine learning feature vectors. Moreover, each orthogonal component can be represented in the basis formed by eigenfields of Laplace–de Rham operators, and the low-frequency coefficients can be used as machine learning features as well. The following session illustrates an example of an eigenfield representation, for the gradient of the reaction potential for molecular electrostatics.

Electrostatics Analysis: Electrostatic interactions are of paramount importance in biomolecular simulations due to their ubiquitous existence and vital contribution to force fields. Two major types of electrostatic analyses are the qualitative analysis for general electrostatic characteristics and the quantitative analysis for statistical, thermodynamic, and kinetic observables. An important two-scale implicit solvent model for electrostatic analysis is the Poisson–Boltzmann (PB) model (Sharp and Honig 1990; Fogolari et al. 2002), in which the explicit water molecules are treated as a dielectric continuum and the dissolved electrolytes are modeled with the Boltzmann distribution. The PB model has been widely applied in biomolecular simulations such as protein structures (Cherezov et al. 2007), protein–protein interactions (Dong et al. 2003), pKa (Alexov et al. 2011; Antosiewicz et al. 1996; Nielsen and McCammon 2003), membranes (Zhou et al. 2010), binding energies (Nguyen et al. 2017), and solvation-free energies (Wagoner and Baker 2006).

The Poisson–Boltzmann Model for a Solvated Molecule: The PB model is illustrated in Fig. 11, in which the molecular surface Γ separates the solute domain Ω_1 and the solvent domain Ω_2. The molecule domain Ω_1 consists of a set of atomic charges q_k located at atomic centers $\mathbf{x}_k$ for $k = 1, \ldots, N_c$. In domain Ω_2, a Boltzmann distribution describes the free ions. For computational purposes, the Boltzmann term is often linearized.

Thus, the electrostatic potential $\phi(\mathbf{x})$ here satisfies the linearized PB equation,

$$-\nabla \cdot \epsilon(\mathbf{x}) \nabla \phi(\mathbf{x}) + \bar{\kappa}^2(\mathbf{x}) \phi(\mathbf{x}) = \sum_{k=1}^{N_c} q_k \delta(\mathbf{x} - \mathbf{x}_k), \tag{30}$$

where $\epsilon(\mathbf{x})$ is the piecewise-constant dielectric function

$$\epsilon(\mathbf{x}) = \begin{cases} \epsilon_1, & \mathbf{x} \in \Omega_1, \\ \epsilon_2, & \mathbf{x} \in \Omega_2, \end{cases} \tag{31}$$

and $\bar{\kappa}$ is the screening parameter with the relation $\bar{\kappa}^2 = \epsilon_2 \kappa^2$, where κ is the inverse Debye length measuring the ionic length. The interface conditions on the molecular surface are

$$\phi_1(\mathbf{x}) = \phi_2(\mathbf{x}), \ \epsilon_1 \frac{\partial \phi_1(\mathbf{x})}{\partial n} = \epsilon_2 \frac{\partial \phi_2(\mathbf{x})}{\partial n}, \ \mathbf{x} \in \Gamma, \tag{32}$$

where ϕ_1 and ϕ_2 are the limit values when approaching the interface from the inside and the outside domains, n is the outward unit normal vector on Γ, and the normal derivatives are $\frac{\partial \phi_i}{\partial n} = n \cdot \nabla \phi_i$. The PB model assumes the far-field boundary condition of $\lim_{|\mathbf{x}| \to \infty} \phi(\mathbf{x}) =$

0. Taking interface Γ as the solvent-excluded surface, the PB model is usually solved numerically. Two types of methods have been developed: Grid-based finite-difference and finite-element methods discretize the entire domain (Im et al. 1998; Honig and Nicholls 1995; Baker et al. 2001), such as MIBPB (Yu et al. 2007; Chen et al. 2011) and boundary-element methods discretize only the molecular surface (Juffer et al. 1991; Liang and Subranmaniam 1997; Vorobjev and Scheraga 1997; Lu et al. 2007; Geng and Krasny 2013). We use boundary-element methods according to the same surface mesh used as the molecular surface and the boundary for our volumetric manifold, for the simplicity of calculating the reaction potential.

Solving PB Model and Reaction Potential: A well-conditioned boundary integral form of PB implicit solvent model is derived by applying Green's second identity and properties of fundamental solutions to Eq. (30), which yields the electrostatic potential,

$$\phi(\mathbf{x}) = \int_{\Gamma}\left[G_0(\mathbf{x},\mathbf{y})\frac{\partial\phi(\mathbf{y})}{\partial n} - \frac{\partial G_0(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{y}}}\phi(\mathbf{y})\right]\mathrm{d}S_{\mathbf{y}} + \sum_{k=1}^{N_c} q_k G_0(\mathbf{x},\mathbf{y}_k),\ \mathbf{x}\in\Omega_1, \tag{33a}$$

$$\phi(\mathbf{x}) = \int_{\Gamma}\left[-G_\kappa(\mathbf{x},\mathbf{y})\frac{\partial\phi(\mathbf{y})}{\partial n} + \frac{\partial G_\kappa(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{y}}}\phi(\mathbf{y})\right]\mathrm{d}S_{\mathbf{y}},\ \mathbf{x}\in\Omega_2, \tag{33b}$$

where the Green's function for Coulomb interaction is $G_0(\mathbf{x},\mathbf{y}) = \frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}$ and the Green's function for the screened Coulomb interaction $G_k(\mathbf{x},\mathbf{y}) = \frac{e^{-\kappa|\mathbf{x}-\mathbf{y}|}}{4\pi|\mathbf{x}-\mathbf{y}|}$. Then, applying the interface condition in Eq. (32) with the differentiation of electrostatic Potential in each domain yields a set of boundary integral equations relating the surface potential ϕ_1 and its normal derivative, $\partial\phi_1/\partial n$ on Γ,

$$\frac{1}{2}(1+\epsilon)\phi_1(\mathbf{x}) = \int_{\Gamma}\left[K_1(\mathbf{x},\mathbf{y})\frac{\partial\phi_1(\mathbf{y})}{\partial n} + K_2(\mathbf{x},\mathbf{y})\phi_1(\mathbf{y})\right]\mathrm{d}S_{\mathbf{y}} + S_1(\mathbf{x}),\ \mathbf{x}\in\Gamma, \tag{34a}$$

$$\frac{1}{2}\left(1+\frac{1}{\epsilon}\right)\frac{\partial\phi_1(\mathbf{x})}{\partial n} = \int_{\Gamma}\left[K_3(\mathbf{x},\mathbf{y})\frac{\partial\phi_1(\mathbf{y})}{\partial n} + K_4(\mathbf{x},\mathbf{y})\phi_1(\mathbf{y})\right]\mathrm{d}S_{\mathbf{y}} + S_2(\mathbf{x}),\ \mathbf{x}\in\Gamma, \tag{34b}$$

where $\epsilon = \epsilon_2/\epsilon_1$. As given in Eqs. (35a–35b) and (36), the kernels $K_{1,2,3,4}$ and source terms $S_{1,2}$ arelinearcombinationsoftheCoulombandscreenedCoulombinteractions, and their first- and second-order normal derivatives,

$$K_1(\mathbf{x},\mathbf{y}) = G_0(\mathbf{x},\mathbf{y}) - G_K(\mathbf{x},\mathbf{y}),\ K_2(\mathbf{x},\mathbf{y}) = \epsilon\frac{\partial G_\kappa(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{y}}} - \frac{\partial G_0(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{y}}}, \tag{35a}$$

$$K_3(\mathbf{x},\mathbf{y}) = \frac{\partial G_0(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{x}}} - \frac{1}{\epsilon}\frac{\partial G_\kappa(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{x}}},\ K_4(\mathbf{x},\mathbf{y}) = \frac{\partial^2 G_\kappa(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{x}}\partial n_{\mathbf{y}}} - \frac{\partial^2 G_0(\mathbf{x},\mathbf{y})}{\partial n_{\mathbf{x}}\partial n_{\mathbf{y}}}, \tag{35b}$$

and the source terms $S_{1,2}$ are

$$S_1(\mathbf{x}) = \frac{1}{\epsilon_1}\sum_{k=1}^{N_c} q_k G_0(\mathbf{x}, \mathbf{y}_k),\ S_2(\mathbf{x}) = \frac{1}{\epsilon_1}\sum_{k=1}^{N_c} q_k \frac{\partial G_0(\mathbf{x}, \mathbf{y}_k)}{\partial n_{\mathbf{x}}}. \tag{36}$$

Once the potential and normal derivaives of the potential on the boundary of Eqs. (33a) and (33b) are solved, the reaction potential $\phi_{\text{reac}}(\mathbf{x}) = \phi(\mathbf{x}) - S_1(\mathbf{x})$ and for $\mathbf{x} \in \Omega_1$ it is given as

$$\phi_{\text{reac}}(\mathbf{x}) = \int_\Gamma \left[G_0(\mathbf{x}, \mathbf{y}) \frac{\partial \phi(\mathbf{y})}{\partial n} - \frac{\partial G_0(\mathbf{x}, \mathbf{y})}{\partial n_{\mathbf{y}}} \phi(\mathbf{y}) \right] \mathrm{d}S_{\mathbf{y}}. \tag{37}$$

Numerically solving boundary integral forms of the PB model requires speedup techniques, for which we directly apply the software package presented in Chen and Geng (2018). The reaction potential describes the potential caused by the solvent and solute near their interface. It is important to calculate the electrostatic solvation energy, given as $\Delta G_{\text{sol}} = \frac{1}{2}\sum_{k=1}^{N_c} q_k \phi_{\text{reac}}(\mathbf{x}_k)$, where N_c is the number of charges and q_k are charges.

Eigenfield Decomposition: The 1-form electrostatic reaction field ω is generated from the gradient of the reaction potential $\nabla\phi_{\text{reac}}$ by taking line integral on each edge. Our goal is to project ω onto the eigenvectors of Hodge Laplacian by L_2-inner products of Eq. (3). The molecular surface Γ created by the solute and the solvent is considered as the boundary of the volumetric manifold M. The space of k-forms $\Omega_k(M)$ is a Hilbert space equipped with the aforementioned L_2-inner products. Therefore, the corresponding 1-form of the electrostatic reaction field inside the molecule surface is in the space $\Omega_1(M)$. Moreover, as shown in Eq. (29), aside from a harmonic component, the gradient of the reaction potential is in the spaced of normal gradient fields, which is spanned by the eigenvectors corresponding to the normal gradient fields. Represented in the basis formed by these eigenvectors, the electrostatic reaction field (without the harmonic component) is a linear combination of these eigenvectors. However, the coefficients are with only large absolute values for certain modes, since dominant eigenmodes often exist due to the geometry characteristics of the molecular domain. We illustrate the Hodge mode decomposition for two examples. Table 2 shows the square of coefficients of ith eigenvector projected on the electrostatic reaction field ω as $\langle \omega, e_i \rangle^2$, and their sums. The dominant eigenvectors for p–p and n–p are the first and second eigenvectors, respectively, as shown in Fig. 12, in which the eigenvectors are sorted in ascending order of their corresponding eigenvalues. As the number of eigenvectors increases, the difference between the electrostatic reaction field and the approximated electrostatic reaction field decreases. Table 3 shows another example with four changes arranged in five ways as shown in Fig. 13. The first case has four positive charges. The first Hodge eigenvector is the dominant mode among all the eigenvectors as shown in Fig. 13. In the second and third cases, where two same type charges located in either the top–bottom or right–left manner, the second and third Hodge eigenvectors dominate their electrostatic reaction fields. The dominant Hodge eigenvector for the third case is the forth Hodge mode. The last case illustrates a molecule that has three positive charges and one negative charge, for which the first Hodge eigenvector is the dominant

mode. In all cases, the accumulated contributions of the first 11 Hodge modes have a similar magnitude. This method is readily applicable to the electrostatic reaction field analysis of complex biomolecular systems and the general Hodge mode analysis of any biomolecular vector fields.

3 Method Preliminaries

We provide the details for our design of computational tools, data structures, and parameters in our implementation of the present de Rham–Hodge spectral analysis. Through efficient implementation, our method is highly scalable and capable of handling molecular data ranging from protein crystal structures to cryo-EM maps.

3.1 Simplicial Complex Generation

The domain of our Laplace–de Rham operators is first tessellated into a simplicial complex, which is a tetrahedral mesh in our 3D case. There are quite a few well-developed software packages for tetrahedral mesh generation given a boundary with a surface triangle mesh as input. We chose CGAL (computational geometry algorithm library) over others for its superior control on element quality.

In theory, we can generate tetrahedral meshes with any highly accurate closed surface. However, macromolecule complexes with atom-level resolution often make the output mesh intractable with typical computing platforms. Moreover, a dense mesh is unnecessary for the calculation of the low-frequency range of the spectrum. Thus, we produce a coarse resolution with a spatial sampling density higher than twice the spatial frequencies (wavenumbers, i.e., square root of eigenvalues of the Laplacians) of the geometrical and topological features to be computed in the given biomolecule complexes.

For protein crystal structures, we tested the construction of the surface using only the C_a positions. First, a Gaussian kernel is assigned to each atomic position to approximate the electron density. Then, a level set surface is generated to construct the contour of the protein closely enclosing the high electron density regions.

For cryo-EM data, to produce a smooth contour surface, Gaussian kernels are associated with data points. Other approaches, such as mean curvature flow (Bates et al. 2008; Zhao et al. 2018), can be used as well. When dealing with noisy and densely sampled data, we can carefully choose the level set that corresponds to a fairly smooth contour surface that encloses the original cryo-EM data.

Given a volumetric data, we can either directly use CGAL to produce a tetrahedral mesh or first convert it to a triangular surface mesh through the marching cubes algorithm, and use that to generate a tetrahedral mesh. Different sampling densities are tested to meet typical quality requirements while balancing computational cost and mesh quality.

3.2 Discrete Exterior Calculus

As a topological structure-preserving discretization of the exterior calculus on differential forms, discrete exterior calculus (DEC) has been widely applied in recent years for various

successful applications on geometrical problems and finite-element analysis, including meshing and computational electromagnetics (Hekstra et al. 2016). It is an appropriate tool for our de Rham–Hodge analysis of biomolecules, as all the related operations, including exterior derivatives and the Hodge stars, are represented as matrices that preserve the defining properties in the continuous setting. More precisely, the discrete exterior derivative operators strictly satisfy $D_{k+1}D_k = 0$, mimicking $d_{k+1}d_k = 0$, and the discrete Hodge star operators are realized by symmetric positive definite matrices. Hence, the discrete Laplace–de Rham operators can be assembled using finite-dimensional linear algebra with the aforementioned three distinct spectra.

To allow replication of our results, we recap our implementation of DEC (Zhao et al. 2019). We start by a tetrahedral tessellation of the volumetric domain, i.e., a tetrahedral mesh, which is the collection of a vertex set $\mathscr{V}$, an edge set $\mathscr{E}$, a triangle set $\mathscr{F}$, and a tetrahedron set $\mathscr{T}$. The vertices are points in 3D Euclidean space; the edges/triangles/tetrahedra are represented as 1-/2-/3-simplices, i.e., pairs/triples/quadruples of vertex indices, respectively, and regarded as the convex hull of these vertices. We further choose an arbitrary orientation for each k-simplex, which is an order set of k+1 vertices, up to an even permutation. We denote an oriented k-simplex as

$$\sigma = [v_0, v_1, \ldots, v_k]. \tag{38}$$

The boundary operator is defined as

$$\partial\sigma = \sum_{i=0}^{k} (-1)^i [v_0, v_1, \ldots, \hat{v}_i, \ldots, v_k], \tag{39}$$

where $\hat{v}_i$ means that the ith vertex is omitted. Thus, the boundary operator will take all the 1-degree lower faces of σ with an induced orientation. We will take the following strategy to handle orientation in the implementation. We usually assign each tet an orientation such that, when applying the boundary operator, each facet has an outward pointing orientation. The total boundary of the tet mesh conforms naturally with the surface with outward pointing orientation. But for each edge and facet, we pre-assignan orientation by increasing indices of incident vertices. In this case, we need to take care of the boundary operator when there is a conflict between the pre-assigned orientation and the induced orientation. The algorithm for calculating the cohomology basis of boundary operators is similar to the algorithm in simplicial homology (Edelsbrunner et al. 2000). However, DEC needs further constructions.

Scalar fields are naturally encoded as 0-forms and 3-forms. A 0-form is the same with the finite-element method such that the coefficients are sampled on vertices equipped with basis functions. A 3-form is, different from a 0-form, stored per tet as volume integration of the scalar field. Vector fields are naturally encoded as 1-form and 2-form. A 1-form is sampled by the line integral on each oriented edge. A 2-form is sampled by surface flux on each oriented facet. Whitney forms (Bossavit 1988) can help convert forms back to piecewise linear vector fields on each tet, which can be used in, e.g., the construction of the operator G.

We will store discrete k-forms as column vectors. Then, as mentioned before, all the discrete operators can be formed as matrices applying on the column vectors. Then, we start to construct discrete exterior derivative and discrete Hodge star matrices. Suppose we are dealing with the discrete differential form $d\omega$ on simplices σ, according to Stokes' theorem

$$\int_{\partial\sigma}\omega=\int_{\sigma}\mathrm{d}\omega, \tag{40}$$

$d\omega$ is just an oriented summation of ω on facets of σ. So the discrete exterior derivative operator D_k is just a matrix filled with −1, 0,1 (Fig. 14), depending on whether the pre-assigned orientation is conforming with the induced orientation. The preservation of Stokes' theorem is what guarantees the preservation of the de Rham cohomology, as the discrete de Rham k-cohomology is isomorphic to the simplicial $n-k$-homology due to the boundary operator, which is in turn isomorphic to singular k-cohomology and thus to the continuous de Rham k-cohomology.

One can easily observe that the discrete exterior derivative operators for dual forms are merely D_k^T. The discrete Hodge star operator S_k is just converting primal form and dual form back a forth by the following equation:

$$\frac{1}{|\sigma_k|}\int_{\sigma_k}\omega=\frac{1}{|{}^*\sigma_k|}\int_{{}^*\sigma_k}\star\,\omega\,. \tag{41}$$

Each primal element in the tet mesh has one corresponding dual element (Fig. 15). So the discrete Hodge star operator is merely a diagonal matrix. Note that here we use a diagonal matrix to approximate the Hodge star operator, where non-diagonal Hodge star with higher accuracy can be applied as well. But a diagonal Hodge star is enough for our current application. The diagonal Hodge star matrix just has diagonal entries as dual-element volume over primal-element volume. For example, given a 1-form on each edge, applying the Hodge star is turning the primal 1-form into dual 2-form stored on each dual facet. This can be interpreted as we sample the vector field at the center of the edge. One way is to compute the 1-form as the sampled vector integrated the primal edge as the line integral; the other way is to compute the 2-form as the sampled vector integrated on the dual facet as vector flux. So the transition can be encoded as a number of dual-element volume over primal-element volume. See Fig. 16 for relations between differential forms and operators.

Once we have these related matrices for discrete operators, we are ready to construct the Laplacian matrix L_k for $k=0,\ldots,3$ as

$$\begin{aligned} L_0&=D_0^TS_1D_0, & L_1&=D_1^TS_2D_1+S_1D_0S_0^{-1}D_0^TS_1,\\ L_2&=D_2^TS_3D_2+S_2D_1S_1^{-1}D_1^TS_2, & L_3&=S_3D_2S_2^{-1}D_2^TS_3, \end{aligned} \tag{42}$$

where D_k are pre-assembled discrete exterior derivatives, S_k are discrete Hodge star matrices and L_k correspond to $\star\Delta_k$. The assembly of Laplace–de Rham operators L_k are just starting from primal k-forms, multiplying matrices along the circular direction as shown in Fig. 16.

Note that the usual Hodge Laplacian matrix is not symmetric generally. In practice, we usually left multiply by Hodge star to turn it into a symmetric one. After this, we need to take care of the boundary conditions. Boundary condition treatment can be incorporated when assembling d matrices. Recall that the d matrices are merely for creating an oriented summation of discrete differential forms stored on simplices. We can just delete corresponding columns and rows for boundary elements. We use $L_{k,t}$ to denote Laplace–de Rham operator with boundary elements and $L_{k,n}$ to denote those without boundary elements (Demlow and Hirani 2014).

Finally, the spectral analysis can be done with a generalized eigenvalue problem in Eq. (8). The smallest eigenvalues and their corresponding eigenvectors are associated with useful low frequencies. In principle, large eigenvalues also contain useful information but are often impaired by large computational errors. We use an eigensolver with parameter starting from small magnitude eigenvalues.

4 Conclusion

The de Rham–Hodge theory is a landmark of twentieth-century mathematics that interconnects differential geometry, algebraic topology, and partial differential equation. It provides a solid mathematical foundation to electromagnetic theory, quantum field theory, and many other important physics. However, this important mathematical tool has never been applied to macromolecular modeling and analysis, to the best of our knowledge. This work introduces the de Rham–Hodge theory as a unified paradigm to analyze biomolecular geometry, topology, flexibility, and Hodge modes based on three-dimensional (3D) coordinates or cryo-EM maps. Specifically, de Rham–Hodge spectral analysis has been carried out to reveal macromolecular geometric characteristic and topological invariants with normal and tangential boundary conditions. The Helmholtz–Hodge decomposition is employed to analyze the divergence-free, curl-free, and harmonic components of macromolecular vector fields. Based on the 0-form scalar Hodge–Laplacian, an accurate multiscale model is constructed to predict protein fluctuations. By equipping a vector Laplace–de Rham operator with a boundary constraint based on Helfrich-type curvature energy, a 1-form Laplace–de Rham–Helfrich operator is proposed to predict the Hodge modes of biomolecules, particularly cryo-EM maps. In addition to its versatile nature for a wide variety of modelling and analysis, the proposed de Rham–Hodge paradigm also provides a unified approach to handle biomolecular problems at various spatial scales and with different data formats. A state-of-the-art 3D discrete exterior calculus algorithm is developed to facilitate accurate, reliable, and topological structure preserving spectral analysis and modeling of biomolecules. Extensive numerical experiments indicate that the proposed de Rham–Hodge paradigm offers one of the most powerful tools for the modeling and analysis of biological macromolecules.

The proposed de Rham–Hodge paradigm provides a solid foundation for a wide variety of other biological and biophysical applications. For example, the present de Rham–Hodge flexibility and Hodge mode analysis can be directly applied to subcellular organelles, such as vesicle, endoplasmic reticulum, golgi apparatus, cytoskeleton, mitochondrion, vacuole, cytosol, lysosome, and centrosome, for which the existing atomistic biophysical approaches

have very limited accessibility. Additionally, features extracted from de Rham–Hodge flexibility and Hodge mode analysis can be incorporated into deep neural networks for the structure reconstruction from medium- and low-resolution cryo-EM maps (Haslam et al. 2018). Finally, due to its ability to characterize geometric traits and describe topological invariants, the proposed de Rham–Hodge paradigm opens an entirely new direction for the quantitative structure–function analysis of molecular and macromolecular datasets. The integration of de Rham–Hodge features and machine learning algorithms for the predictions of protein–ligand-binding affinity, protein–protein-binding affinity, protein-folding stability change upon mutation, drug toxicity, solubility, partition coefficient, permeability, and plasma protein binding is under our consideration.

Acknowledgements

This work was supported in part by NSF Grants DMS-1721024, DMS-1761320, and IIS1900473 and NIH Grant GM126189. GWW was also funded by Bristol-Myers Squibb and Pfizer.

References

Alexov E, Mehler EL, Baker N, Baptista AM, Huang Y, Milletti F, Erik Nielsen J, Farrell D, Carstensen T, Olsson MH et al. (2011) Progress in the prediction of pka values in proteins. Proteins Struct Funct Bioinf 79(12):3260–3275

Antosiewicz J, McCammon JA, Gilson MK (1996) The determinants of p*Ka*s in proteins. Biochemistry 35(24):7819–7833 [PubMed: 8672483]

Arnold DN, Falk RS, Winther R (2006) Finite element exterior calculus, homological techniques, and applications. Acta Numer 15:1–155

Atilgan AR, Durell S, Jernigan RL, Demirel M, Keskin O, Bahar I (2001) Anisotropy of fluctuation dynamics of proteins with an elastic network model. Biophys J 80(1):505–515 [PubMed: 11159421]

Bahar I, Atilgan AR, Erman B (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. Fold Des 2:173–181 [PubMed: 9218955]

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Nat Acad Sci USA 98(18):10037–10041 [PubMed: 11517324]

Baradaran R, Wang C, Siliciano AF, Long SB (2018) Cryo-em structures of fungal and metazoan mitochondrial calcium uniporters. Nature 559(7715):580–584 [PubMed: 29995857]

Bates PW, Wei GW, Zhao S (2008) Minimal molecular surfaces and their applications. J Comput Chem 29(3):380–91 [PubMed: 17591718]

Bhatia H, Norgard G, Pascucci V, Bremer P-T (2013) The Helmholtz–Hodge decomposition-a survey. IEEE Trans Vis Comput Graphics 19(8):1386–1404

Blinn JF (1982) A generalization of algebraic surface drawing. ACM Trans Graph 1:235–256

Bossavit A (1988) Whitney forms: a class of finite elements for three-dimensional computations in electromagnetism. IEE Proc A (Phys Sci Meas Instrum Manag Educ Rev) 135(8):493–500

Bott R, Tu LW (2013) Differential forms in algebraic topology, vol 82. Springer, Berlin

Brooks BR, Bruccoleri RE, Olafson BD, States D, Swaminathan S, Karplus M (1983) Charmm: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 4:187–217

Cang ZX, Wei GW (2017) TopologyNet: topology based deep convolutional and multi-task neural networks for biomolecular property predictions. PLoS Comput Biol 13(7):e1005690. 10.1371/journal.pcbi.1005690 [PubMed: 28749969]

Cang ZX, Wei GW (2018) Integration of element specific persistent homology and machine learning for protein-ligand binding affinity prediction. Int J Numer Methods Biomed Eng. 10.1002/cnm.2914

Cantarella J, DeTurck D, Gluck H (2002) Vector calculus and the topology of domains in 3-space. Am Math Mon 109(5):409–442

Carlsson G, Zomorodian A, Collins A, Guibas LJ (2005) Persistence barcodes for shapes. Int J Shape Model 11(2):149–187

Chen D, Chen Z, Chen C, Geng WH, Wei GW (2011) MIBPB: a software package for electrostatic analysis. J Comput Chem 32:657–670

Chen J, Geng W (2018) On preconditioning the treecode-accelerated boundary integral (TABI) Poisson–Boltzmann solver. J Comput Phys 373:750–762

Chen M, Tu B, Lu B (2012) Triangulated manifold meshing method preserving molecular surface topology. J Mole Graph Model 38:411–418

Cheng H-L, Shi X (2009) Quality mesh generation for molecular skin surfaces using restricted union of balls. Comput Geom 42(3):196–206

Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SG, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK et al. (2007) High-resolution crystal structure of an engineered human β2-adrenergic g protein–coupled receptor. Science 318(5854):1258–1265 [PubMed: 17962520]

Corey RB, Pauling L (1953) Molecular models of amino acids, peptides, and proteins. Rev Sci Instrum 24(8):621–627

De La Torre JG, Bloomfield VA (1977) Hydrodynamic properties of macromolecular complexes. i. translation. Biopolym Original Res Biomol 16(8):1747–1763

Demlow A, Hirani AN (2014) A posteriori error estimates for finite element exterior calculus: the de Rham complex. Found Comput Math 14(6):1337–1371

Desbrun M,Hirani AN,Leok M,Marsden JE(2005)Discreteexteriorcalculus.arXivpreprintmath/0508341

Dey TK, Fan F, Wang Y (2013) An efficient computation of handle and tunnel loops via Reeb graphs. ACM Trans Graph 32(4):32

Dong F, Vijaykumar M, Zhou HX (2003) Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. Biophys J 85(1):49–60 [PubMed: 12829463]

Du Q, Liu C, Wang X (2004) A phase field approach in the numerical study of the elastic bending energy for vesicle membranes. J Comput Phys 198(2):450–468

Duncan BS, Olson AD (1993) Shape analysis of molecular surfaces. Biopolymers 33(2):231–8 [PubMed: 8485297]

Edelsbrunner H, Harer J (2010) Computational topology: an introduction. American Mathematical Soc, Providence

Edelsbrunner H, Letscher D, Zomorodian A (2000) Topological persistence and simplification. In: 41st annual symposium on foundations of computer science, 2000. Proceedings. IEEE, pp 454–463

Feng X, Xia K, Tong Y, Wei G-W (2012) Geometric modeling of subcellular structures, organelles and large multiprotein complexes. Int J Numer Methods Biomed Eng 28:1198–1223

Fogolari F, Brigo A, Molinari H (2002) The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. J Mol Recognit 15(6):377–92 [PubMed: 12501158]

Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. Science 254(5038):1598–1603 [PubMed: 1749933]

Geng W, Krasny R (2013) A treecode-accelerated boundary integral Poisson-Boltzmann solver for electrostatics of solvated biomolecules. J Comput Phys 247:62–78

Go N, Noguti T, Nishikawa T (1983) Dynamics of a small globular protein in terms of low-frequency vibrational modes. Proc Natl Acad Sci 80:3696–3700 [PubMed: 6574507]

Hanawa-Suetsugu K, Sekine S-I, Sakai H, Hori-Takemoto C, Terada T, Unzai S, Tame JR, Kuramitsu S, Shirouzu M, Yokoyama S (2004) Crystal structure of elongation factor p from Thermus thermophilus hb8. Proc Nat Acad Sci 101(26):9595–9600 [PubMed: 15210970]

Haslam D, Zeng T, Li R, He J (2018) Exploratory studies detecting secondary structures in medium resolution 3d cryo-em images using deep convolutional neural networks. In: Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics. ACM, pp 628–632

Hekstra DR,White KI,Socolich MA,Henning RW,Šrajer V,Ranganathan R(2016)Electric-field-stimulated protein mechanics. Nature 540(7633):400 [PubMed: 27926732]

Helfrich W (1973) Elastic properties of lipid bilayers: theory and possible experiments. Zeitschrift für Naturforschung Teil C 28:693–703

Hirani AN (2003) Discrete exterior calculus. PhD thesis, California Institute of Technology

Hodge WVD (1989) The theory and applications of harmonic integrals. CUP Archive

Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268(5214):1144–9 [PubMed: 7761829]

Im W, Beglov D, Roux B (1998) Continuum solvation model: electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. Comput Phys Commun 111(1–3):59–75

Jiang J, Wang Y, Sušac L, Chan H, Basu R, Zhou ZH, Feigon J (2018) Structure of telomerase with telomeric dna. Cell 173(5):1179–1190 [PubMed: 29775593]

Juffer A, van Keulen BE, van der Ploeg A, Berendsen H (1991) The electric potential of a macromolecule in a solvent: a fundamental approach. J Comput Phys 97:144–171

Kuglstatter A, Stihle M, Neumann C, Müller C, Schaefer W, Klein C, Benz J, Research RP, Development E (2017) Structural differences between glycosylated, disulfide-linked heterodimeric knob-into-hole fc fragment and its homodimeric knob-knob and hole-hole side products. Protein Eng Des Sel 30(9):649–656 [PubMed: 28985438]

Lanza A, Margheritis E, Mugnaioli E, Cappello V, Garau G, Gemmi M (2019) Nanobeam precession-assisted 3d electron diffraction reveals a new polymorph of hen egg-white lysozyme. IUCrJ 6(2):178–188

Lee B, Richards FM (1971) The interpretation of protein structures: estimation of static accessibility. J Mol Biol 55(3):379–400 [PubMed: 5551392]

Levitt M, Sander C, Stern PS (1985) Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J Mol Biol 181(3):423–447 [PubMed: 2580101]

Li L, Li C, Zhang Z, Alexov E (2013) On the dielectric constant of proteins: smooth dielectric function for macromolecular modeling and its implementation in delphi. J Chem Theory Comput 9(4):2126–2136 [PubMed: 23585741]

Liang J, Subranmaniam S (1997) Computation of molecular electrostatics with boundary element methods. Biophys J 73:1830–1841 [PubMed: 9336178]

Lim L-H (2015) Hodge Laplacians on graphs. arXiv preprint arXiv:1507.05379

Lu B, Cheng X, McCammon JA (2007) new-version-fast-multipole-method accelerated electrostatic calculations in biomolecular systems. J Comput Phys 226(2):1348–1366 [PubMed: 18379638]

Ma JP (2005) Usefulness and limitations of normal mode analysis in modeling dynamics of biomolecular complexes. Structure 13:373–380 [PubMed: 15766538]

Ming D, Kong Y, Lambert MA, Huang Z, Ma J (2002) How to describe protein motion without amino acid sequence and atomic coordinates. Proc Nat Acad Sci 99(13):8620–8625 [PubMed: 12084922]

Mitchell JC (1998) Hodge decomposition and expanding maps on the flat tori. PhD thesis, University of California, Berkeley

Muench SP, Huss M, Song CF, Phillips C, Wieczorek H, Trinick J, Harrison MA (2009) Cryo-electron microscopy of the vacuolar ATPase motor reveals its mechanical and regulatory complexity. J Mol Biol 386(4):989–999 [PubMed: 19244615]

Murakami K, Stewart M, Nozawa K, Tomii K, Kudou N, Igarashi N, Shirakihara Y, Wakatsuki S, Yasunaga T, Wakabayashi T (2008) Structural basis for tropomyosin overlap in thin (actin) filaments and the generation of a molecular swivel by troponin-t. Proc Nat Acad Sci 105(20):7200–7205 [PubMed: 18483193]

Natarajan V, Koehl P, Wang Y, Hamann B (2008) Visual analysis of biomolecular surfaces. In: Linsen L, Hagen H, Hamann B (eds) Mathematical methods for visualization in medicine and life science. Springer, Berlin, pp 237–256

Nguyen DD, Wang B, Wei GW (2017) Accurate, robust and reliable calculations of Poisson–Boltzmann binding energies. J Comput Chem 38:941–948 [PubMed: 28211071]

Nguyen DD, Xia KL, Wei GW (2016) Generalized flexibility-rigidity index. J Chem Phys 144:234106 [PubMed: 27334153]

Nielsen JE, McCammon JA (2003) Calculating pka values in enzyme active sites. Protein Sci 12(9):1894–1901 [PubMed: 12930989]

Nishino T, Rago F, Hori T, Tomii K, Cheeseman IM, Fukagawa T (2013) Cenp-t provides a structural platform for outer kinetochore assembly. EMBO J 32(3):424–436 [PubMed: 23334297]

Opron K, Xia KL, Wei GW (2014) Fast and anisotropic flexibility-rigidity index for protein flexibility and fluctuation analysis. J Chem Phys 140:234105 [PubMed: 24952521]

Richards FM (1977) Areas, volumes, packing, and protein structure. Ann Rev Biophys Bioeng 6(1): 151–176 [PubMed: 326146]

Sander B, Golas MM, Makarov EM, Brahms H, Kastner B, Lührmann R, Stark H (2006) Organization of core spliceosomal components u5 snrna loop i and u4/u6 di-snrnp within u4/u6. u5 tri-snrnp as revealed by electron cryomicroscopy. Mol Cell 24(2):267–278 [PubMed: 17052460]

Sharp KA, Honig B (1990) Electrostatic interactions in macromolecules—theory and applications. Ann Rev Biophys Biophys Chem 19:301–332 [PubMed: 2194479]

Singh AK, McGoldrick LL, Twomey EC, Sobolevsky AI (2018) Mechanism of calmodulin inactivation of the calcium-selective TRP channel trpv6. Sci Adv 4(8):eaau6088 [PubMed: 30116787]

Tama F, Wriggers W, Brooks CL III (2002) Exploring global distortions of biological macromolecules and assemblies from low-resolution structural information and elastic network theory. J Mol Biol 321(2):297–305 [PubMed: 12144786]

Tamstorf R, Grinspun E (2013) Discrete bending forces and their Jacobians. Graph Models 75(6):362–370

Tasumi M, Takenchi H, Ataka S, Dwidedi AM, Krimm S (1982) Normal vibrations of proteins: glucagon. Biopolymers 21:711–714 [PubMed: 7066480]

Vorobjev YN, Scheraga HA (1997) A fast adaptive multigrid boundary element method for macromolecular electrostatic computations in a solvent. J Comput Chem 18(4):569–583

Wagoner JA, Baker NA (2006) Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms. Proc Nat Acad Sci USA 103(22):8331–6 [PubMed: 16709675]

Wollenman LC, Vander Ploeg MR, Miller ML, Zhang Y, Bazil JN (2017) The effect of respiration buffer composition on mitochondrial metabolism and function. PLoS ONE 12(11):e0187523 [PubMed: 29091971]

Xia K, Wei G-W (2016) A review of geometric, topological and graph theory apparatuses for the modeling and analysis of biomolecular data. arXiv preprint arXiv:1612.01735

Xia KL, Feng X, Tong YY, Wei GW (2014) Multiscale geometric modeling of macromolecules i: Cartesian representation. J Comput Phys 275:912–936

Xia KL, Feng X, Tong YY, Wei GW (2015) Persistent homology for the quantitative prediction of fullerene stability. J Comput Chem 36:408–422 [PubMed: 25523342]

Xia KL, Opron K, Wei GW (2013) Multiscale multiphysics and multidomain models—flexibility and rigidity. J Chem Phys 139:194109 [PubMed: 24320318]

Xia KL, Wei GW (2014) Persistent homology analysis of protein structure, flexibility and folding. Int J Numer Methods Biomed Eng 30:814–844

Yao Y, Sun J, Huang X, Bowman GR, Singh G, Lesnick M, Guibas LJ, Pande VS, Carlsson G (2009) Topological methods for exploring low-density states in biomolecular folding pathways. J Chem Phys 130(14):04B614

Yu SN, Geng WH, Wei GW (2007) Treatment of geometric singularities in implicit solvent models. J Chem Phys 126:244108 [PubMed: 17614538]

Yu ZY, Holst M, Cheng Y, McCammon JA (2008) Feature-preserving adaptive mesh generation for molecular shape modeling and simulation. J Mol Graphics Model 26:1370–1380

Zhao R, Cang Z, Tong Y, Wei G-W (2018) Protein pocket detection via convex hull surface evolution and associated Reeb graph. Bioinformatics 34(17):i830–i837 [PubMed: 30423105]

Zhao R, Desbrun M, Wei G-W, Tong YY (2019) 3D Hodge decompositions of edge-and face-based vector fields

Zheng Q, Yang S, Wei G-W (2012) Biomolecular surface construction by PDE transform. Int J Numer Methods Biomed Eng 28(3):291–316

Zhou Y, Lu B, Gorfe AA (2010) Continuum electromechanical modeling of protein-membrane interactions. Phys Rev E 82(4):041923

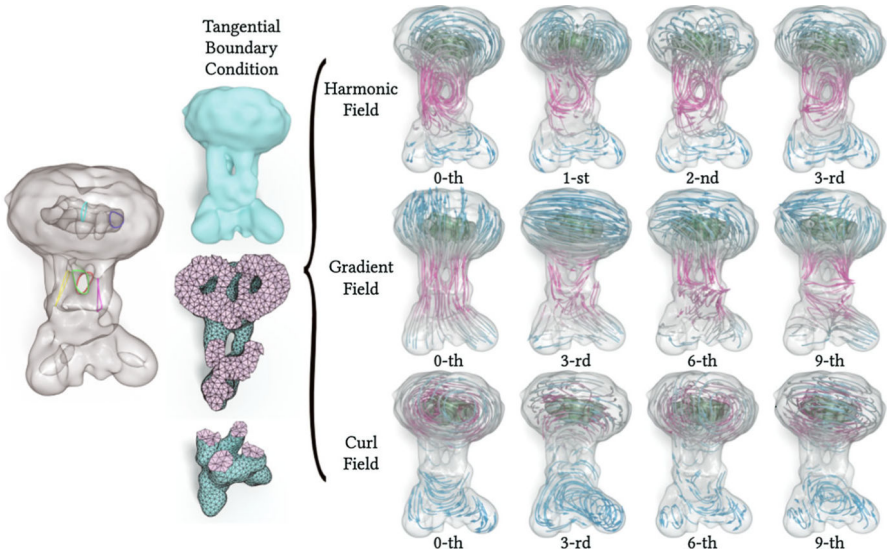

Fig. 1.

Illustration of tangential spectra of a cryo-EM map EMD 7972. Topologically, EMD 7972 (Baradaran et al. 2018) has six handles and two cavities. The left column is the original shape and its anatomy showing the topological complexity. On the right-hand side of the parenthesis, the first row shows tangential harmonic eigenfields, the second row shows tangential gradient eigenfields, and the third row shows tangential curl eigenfields. The credit for the leftmost picture belongs to Hayam Mohamed Abdelrahman

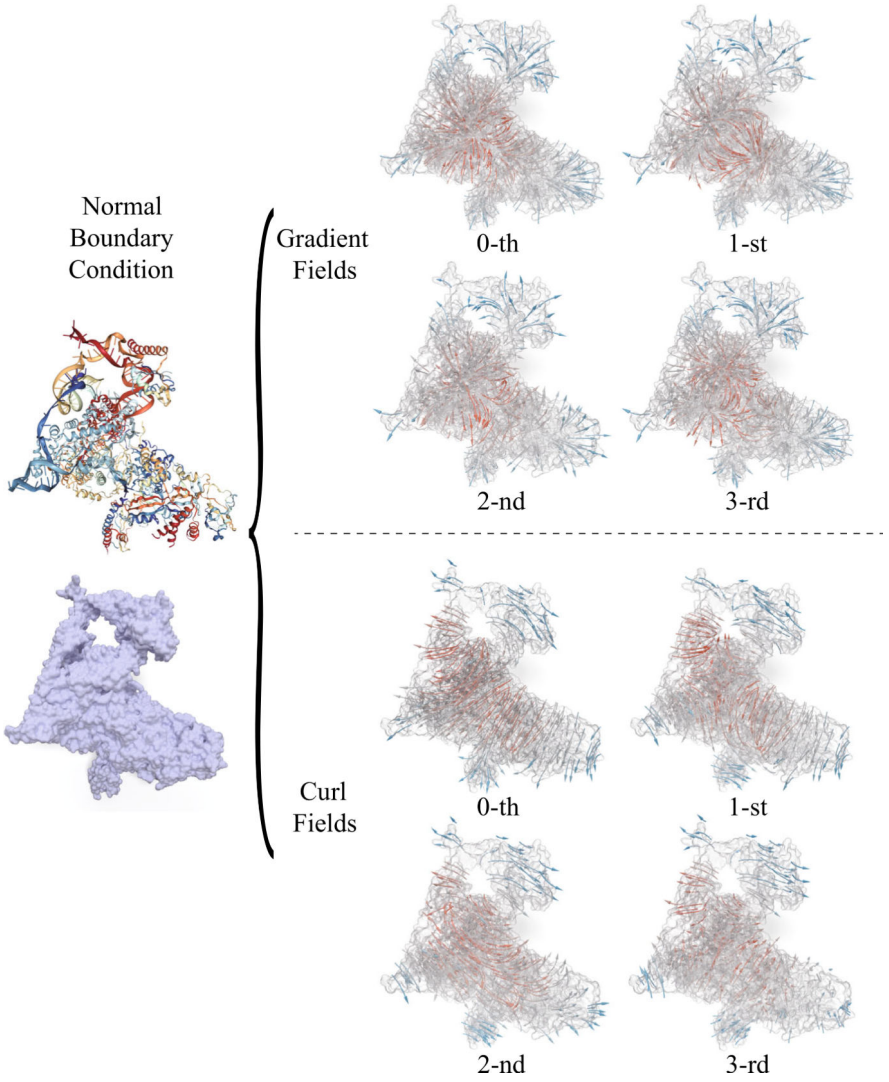

Fig. 2.

Illustration of the normal spectra of protein and DNA complex 6D6V. Topologically, the crystal structure of 6D6V (Jiang et al. 2018) has 1 handle. The left column shows the secondary structure and the solvent-excluded surface (SES). On the right-hand side, the first two rows show normal gradient eigenfields, and the last two rows show normal curl eigenfields

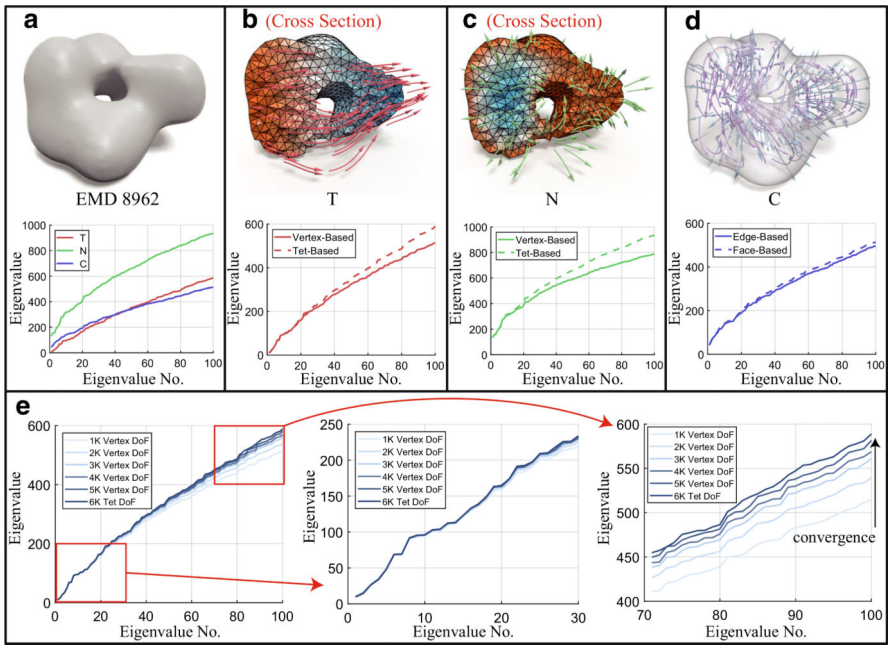

Fig. 3.

Illustration of Hodge Laplacian spectra. This figure shows the properties of three spectral groups, namely tangential gradient eigenfields (*T*), normal gradient eigenfields (*N*), and curl eigenfields (*C*), for EMD 8962 (Singh et al. 2018). **a** The original input surface and three distinct spectral groups. **b** The cross-section of a typical tangential gradient eigenfield and the distribution of eigenvalues for group *T*. **c** The cross-section of a typical normal gradient eigenfield and the distribution of eigenvalues for group *N*. **d** A typical curl eigenfield and the distribution of eigenvalues for group *C*. **e** The left chart shows the convergence of spectra in the same spectral group due to the increase in the mesh size, i.e., the DoFs from 1000 (1K) to 6000 (6K). Obviously, low-order eigenvalues converge fast (middle chart) and high-order eigenvalues converge slowly (right chart)

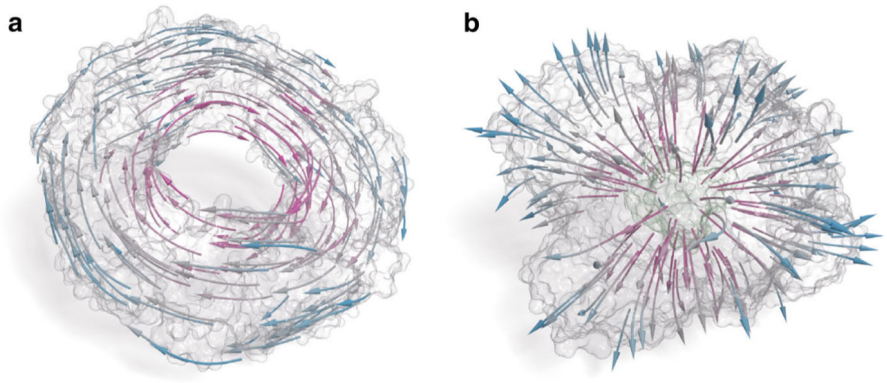

Fig. 4.

Illustration of topological analysis. **a** Eigenfields by null space of tangential Laplace–de Rham operators correspond to handles. **b** Eigenfields by null space of normal Laplace–de Rham operators correspond to cavities

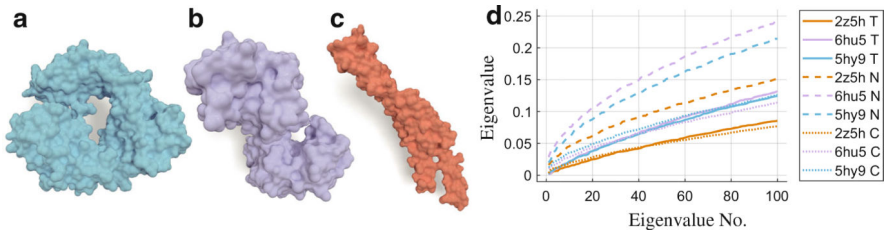

Fig. 5.

Illustration of geometric analysis. The geometry of different molecules (PDB IDs: 2Z5H (**a**), 6HU5 (**b**), and 5HY9 (**c**)) can be captured by three groups of different Hodge Laplacian spectra with clear separations shown in **d**. Note that the color of the line plot corresponds to the color of the molecules. The solid lines show the tangential gradient (T) spectrum, the dashed lines show the normal gradient (N) spectrum, and the dot lines show the curl spectrum (C). While there is a possibility that certain spectral sets may be close to each other (see group T of proteins 6HU5 and 5HY9), the other two groups of spectra (see groups N and C of proteins 6HU5 and 5HY9) will show a clear difference. In addition, our topological features will also provide a definite difference. For example, protein 6HU5 has trivial topology (ball), but protein 5HY9 has a handle

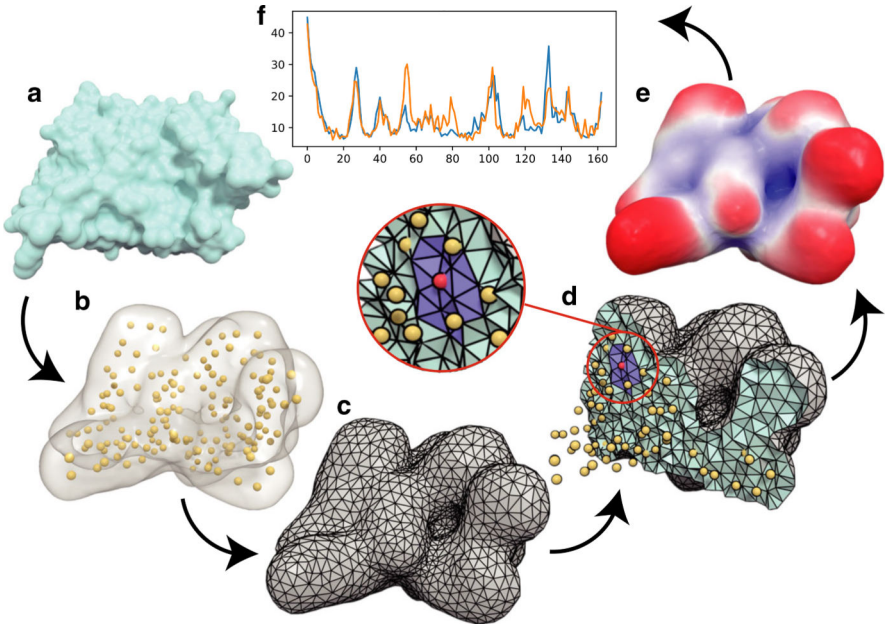

Fig. 6.

Illustration of the procedure for flexibility analysis. We use protein 3VZ9 (Nishino et al. 2013) as an example to demonstrate our procedure from **a** to **f**. **a** The input protein crystal structure. **b** That only C-alpha atoms (yellow spheres) are considered in this case. We assign a Gaussian kernel to each C-alpha atom and extract the level set surface (transparent surface) as our computation domain. **c** That standard tetrahedral mesh is generated with the domain. (Boundary faces are gray; inner faces are indigo.) We use a standard matrix diagonalization procedure to obtain eigenvalues and eigenvectors. B-factor at each mesh vertex is computed as shown in Eq. (22). **d** B-factor at the position of a C-alpha atom is obtained by the linear regression using within the nearby region. (For the red C-alpha atom, the linear regression region is colored as purple, which is within the cutoff radius.) **e** The predicted B-factors on the surface. **f** The predicted B-factors at C-alpha atoms (orange), compared with the experimental B-factors in the PDB file (blue). Our prediction for 3VA9 has the Pearson correlation coefficient of 0.8081

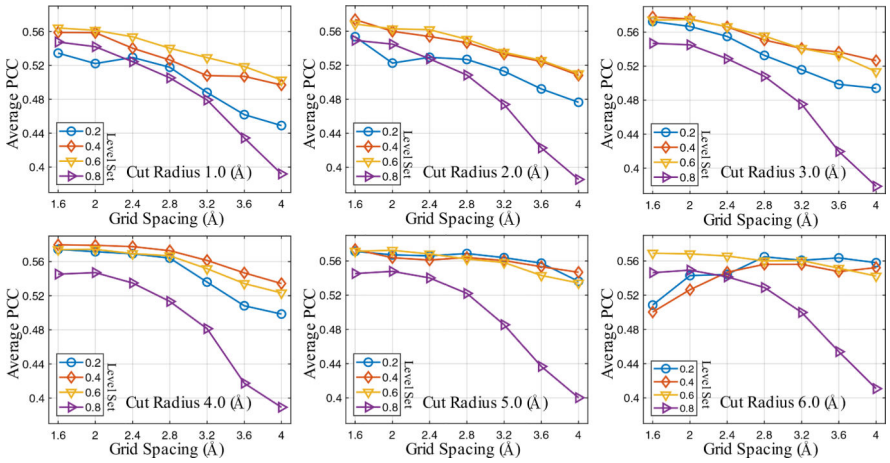

Fig. 7.

Statistics of the average Pearson correlation coefficient (PCC) with various parameters on the test set of 364 proteins. Each plot has the same cutoff radius varying from 1.0 Å to 6.0 Å with interval 1.0 Å. In each plot, the level set value varies from 0.2 to 0.8 with interval 0.2 shown by different lines; the grid spacing varies from 1.6 Å to 4.0 Å with interval 0.4 Å shown in the horizontal axis

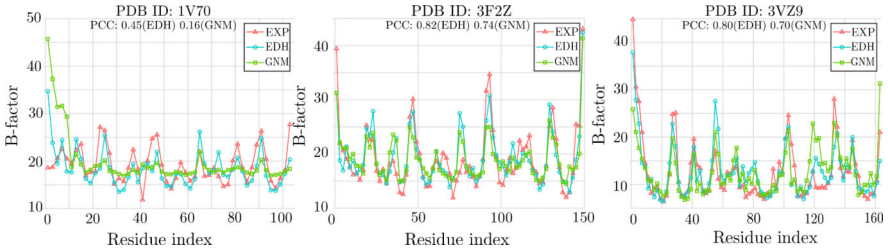

Fig. 8.

Illustration of B-factor prediction. We use proteins 1V70 (Hanawa-Suetsugu et al. 2004), 3F2Z, and 3VZ9 as examples to show our predictions compared with the experiments. The red lines with triangles are the ground truth from experimental data. The blue lines with circles are predictions with our method (EDH). The green lines with cubes are predictions from Gaussian network method (GNM)

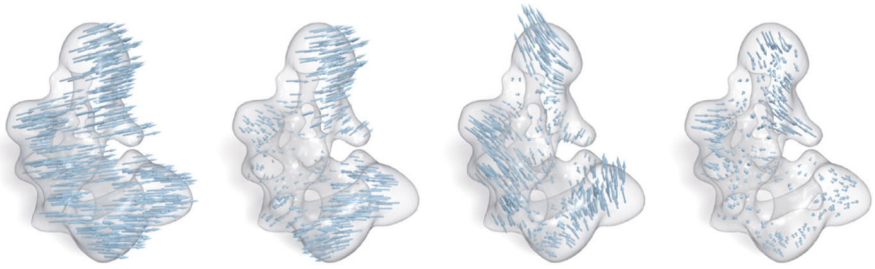

Fig. 9.
Hodge modes of EMD 1258. The 0th, 4th, 8th, and 12th Hodge modes are shown

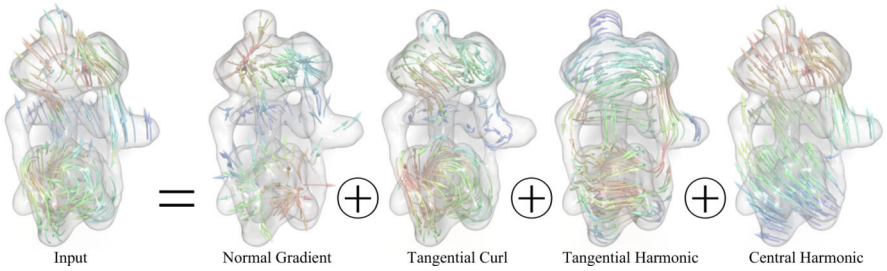

Fig. 10.

Biological flow decomposition. Illustration of a synthetic vector field in EMD 1590 that is decomposed into several mutually orthogonal components based on different boundary conditions

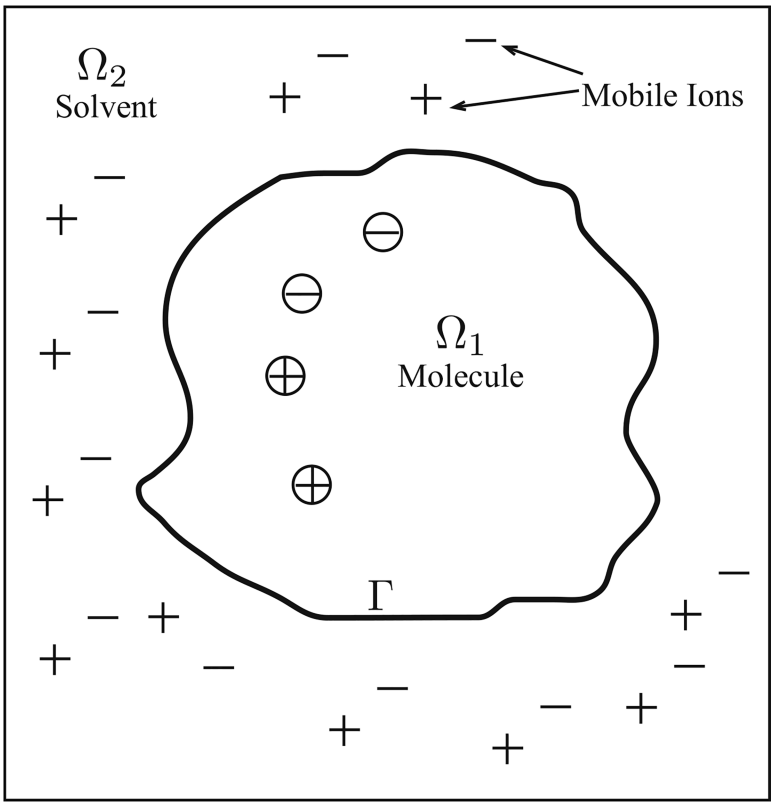

Fig. 11.

The PB implicit solvent model. Γ is the molecular surface separating space into the solute region Ω_1 and the solvent region Ω_2

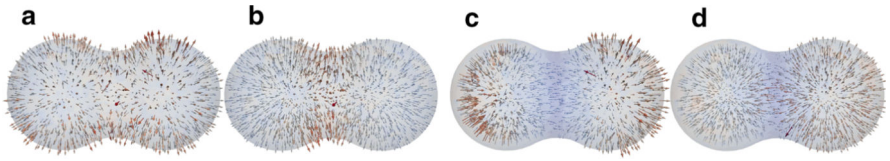

Fig. 12.

a The force field of two positive charges; **b** the first eigenvector; **c** the force field of one negative and one positive charge; **c** the second eigenvector

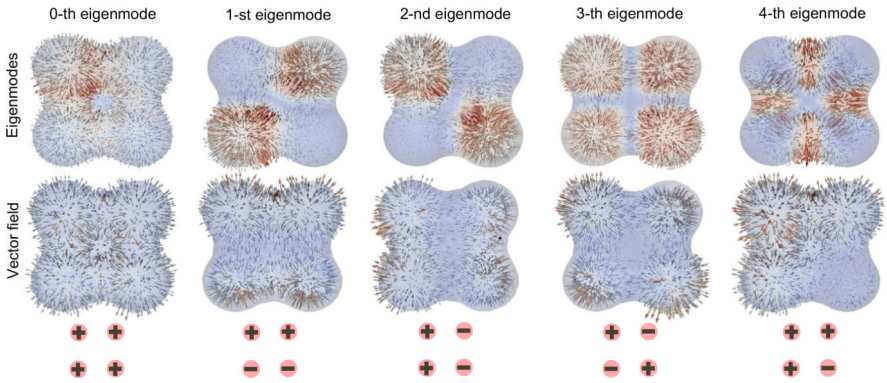

Fig. 13.

The first row shows the first five eigenmodes. The second row shows vector fields under corresponding charge combinations

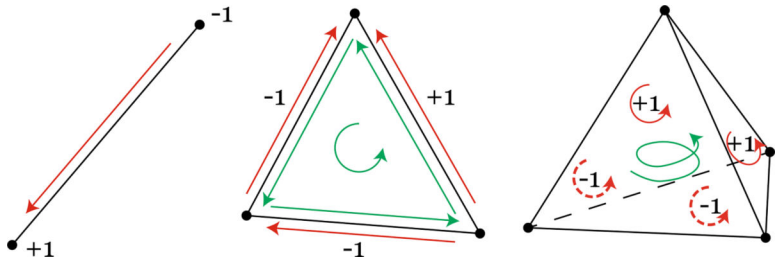

Fig. 14.

Illustration of orientation. The pre-assigned orientation is colored in red. Induced orientation by ∂ is colored in green. The vertices are assumed to have a positive pre-assigned orientation. Therefore, the induced orientation from edge orientation is +1 at the head and −1 at the tail. For a triangle facet, +1 is assigned whenever the pre-assigned orientation conforms with the induced orientation, and −1 vice versa. A similar rule applies to tets which obey a right-hand orientation with the normal pointing outward. Non-adjacent vertices give 0

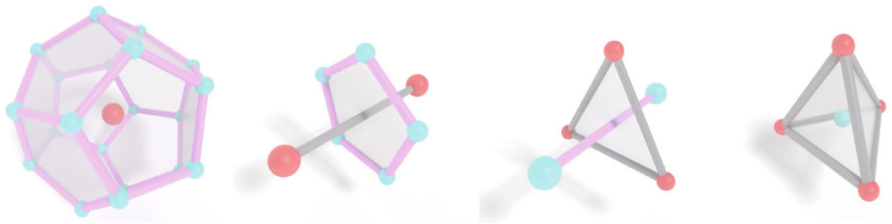

Fig. 15.

Illustration of the primal and dual elements of the tetrahedral mesh. All the red vertices are mesh primal vertices. All the indigo vertices are dual vertices at the circumcenter of each tet. All the gray edges are primal edges. All the pink edges are dual edges connecting adjacent dual vertices. The first chart shows the dual cell of a primal vertex. The second chart shows the dual facet of the primal edge. The third chart shows the dual edge of the primal facet. The last chart shows the dual vertex of the primal cell (tet)

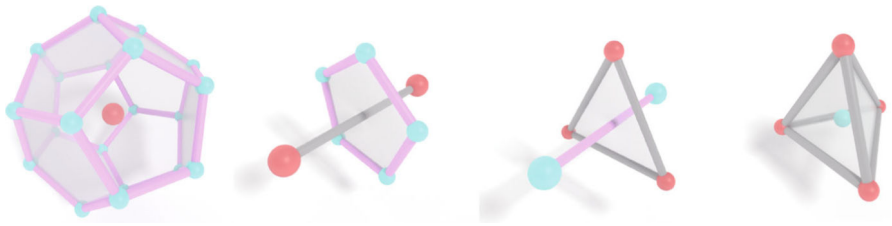

Fig. 16.

Illustration of cohomology. This figure illustrates the relation by exterior derivative and Hodge star operators. The assembly of Laplacian operator Lk is just starting from primal k-forms, multiplying matrices along the circular direction

Table 1

The average Pearson correlation coefficient for predicting 364 proteins at cutoff radius 4.0 Å. The overall best average Pearson correlation coefficient is 0.580 (in bold), compared to that of 0.565 for GNM on the same dataset (Opron et al. 2014)

		Grid spacing (Å)						
		1.6	**2.0**	**2.4**	**2.8**	**3.2**	**3.6**	**4.0**
Level set	0.2	0.574	0.572	0.569	0.564	0.536	0.508	0.498
	0.4	**0.580**	0.579	0.578	0.573	0.561	0.547	0.534
	0.6	0.574	0.574	0.569	0.567	0.552	0.534	0.523
	0.8	0.545	0.547	0.535	0.513	0.481	0.417	0.389

Table 2

Example 1 considers two cases: p–p for two positive charges and n–p for a negative charge on the left and a positive charge on the right. Here, $\langle \omega, e_i \rangle^2$ is the squared inner product of the normalized electrostatic reaction field ω with ith eigenvector, which is normalized too. The second row of each case is the squared sum of inner products. The sum recovers the normalized electrostatic reaction field if summation is carried out over the inner products with all the eigenfields according to Parseval's theorem

Eigenvector		e_0	e_1	e_2	e_3	e_4	⋯	e_{10}	⋯	e_{100}	⋯	e_{200}
p–p	$(\omega, e_i)^2$	0.538	0.006	0.025	0.000	0.000	⋯	0.001	⋯	0.000	⋯	0.000
	$\sum_{j=1}^{i} (\omega, e_j)^2$	0.538	0.544	0.569	0.569	0.569	⋯	0.576	⋯	0.928	⋯	0.964
n–p	$(\omega, e_i)^2$	0.002	0.479	0.000	0.000	0.01	⋯	0.000	⋯	0.000	⋯	0.000
	$\sum_{j=1}^{i} (\omega, e_j)^2$	0.002	0.481	0.481	0.481	0.482	⋯	0.556	⋯	0.906	⋯	0.941

Table 3

Example 2 considers four charges arranged in five cases, namely p–p–p–p, p–p–n–n, p–n–p–n, p–n–n–p, and p–p–p–n, where "p" stands for positive and "n" stands for negative, specified in the order of top left, top right, bottom left, and bottom right. Here, $\langle \omega, e_i \rangle^2$ is the inner product square of the normalized electrostatic reaction field ω with ith eigenvector, which is normalized too. The second row of each case is the squared sum of inner products. The sum recovers the normalized electrostatic reaction field if summation is carried out over the inner products with all the eigenfields according to Parseval's theorem

Eigenvector		e_0	e_1	e_2	e_3	e_4	⋯	e_{10}	⋯	e_{100}	⋯	e_{200}
p–p–p–p	$(\omega, e_i)^2$	0.547	0.017	0.000	0.001	0.000	⋯	0.000	⋯	0.001	⋯	0.000
	$\sum_{j=1}^{i}(\omega, e_j)^2$	0.547	0.564	0.564	0.565	0.565	⋯	0.566	⋯	0.853	⋯	0.911
p–p–n–n	$(\omega, e_i)^2$	0.008	0.268	0.211	0.001	0.000	⋯	0.001	⋯	0.000	⋯	0.000
	$\sum_{j=1}^{i}(\omega, e_j)^2$	0.008	0.276	0.487	0.488	0.488	⋯	0.546	⋯	0.839	⋯	0.895
p–n–p–n	$(\omega, e_i)^2$	0.005	0.198	0.272	0.005	0.000	⋯	0.000	⋯	0.000	⋯	0.000
	$\sum_{j=1}^{i}(\omega, e_j)^2$	0.005	0.203	0.475	0.480	0.480	⋯	0.533	⋯	0.840	⋯	0.892
p–n–n–p	$(\omega, e_i)^2$	0.002	0.002	0.002	0.434	0.000	⋯	0.000	⋯	0.000	⋯	0.000
	$\sum_{j=1}^{i}(\omega, e_j)^2$	0.002	0.004	0.006	0.440	0.440	⋯	0.459	⋯	0.839	⋯	0.888
p–p–p–n	$(\omega, e_i)^2$	0.434	0.055	0.000	0.047	0.000	⋯	0.000	⋯	0.001	⋯	0.000
	$\sum_{j=1}^{i}(\omega, e_j)^2$	0.434	0.489	0.489	0.536	0.536	⋯	0.557	⋯	0.848	⋯	0.908