Vaccine xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Isometrically invariant and allosterically aware description of deformable macromolecular surfaces: Application to the viral neuraminidase

Q1 Eric Paquet[a,b,*], Herna L. Viktor[b]

[a] Vaccines Program, National Research Council, 1200 Montreal Road, Ottawa, Canada
[b] School of Electrical Engineering and Computer Science, University of Ottawa, 800 King Edward Avenue, Ottawa, Canada

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Macromolecular surface
Description
Neuraminidase
Allosteric
Deformable
Heat kernel
Fractional

ABSTRACT

Motivation: The macromolecular surfaces associated with proteins and macromolecules play a key role in determining their functionality and interactions, and are also of importance in structural analysis and classification. As a result of their interaction with their environment, the macromolecular surfaces experience random conformational deformations. Consequently, a realistic description of the molecular surface must be invariant under these deformations. Further, the motion associated with disconnected regions on the molecular surface may be correlated. This property is known as the allosteric effect. In this paper, we address these two requirements. To this end, we propose an approach based on discrete differential geometry and the fractional Fokker–Planck equation which provides an isometrically invariant and allosteric aware description of macromolecular surfaces. Our method is applied to the influenza neuraminidase.

1. Introduction

Q3 The shape of macromolecular structures, such as proteins, play a fundamental and multifaceted role in their classification, evolution [1], functionality [2], description [3] and in defining their mutual interactions [4]. For instance, the potential docking of two macromolecular structures is, in a large part, determined by the geometrical compatibility in between the receptor and the ligand [5]. The shape of a protein is more resilient than the underlying amino acid sequences, as the former must be preserved over time in order to maintain the functionality, while the latter is prone to mutations. As a result, methods based on macromolecular shape are more suited than phylogenetic approaches, in order to study evolution over very long periods of time [1].

The creation of an informative and compact description of macromolecular shapes is of paramount importance in structural proteomics, macromolecular docking, computational vaccinology and in the study of protein evolution, amongst others [5,6]. Yet, this is not the end of the story. Macromolecules undergo random fluctuations of their shape (conformational changes). This characteristic is due to their interaction with their environment [7] as well as their interactions with one another. For instance, conformational changes are caused by the induced fit between a receptor and a ligand during macromolecular docking [4]. As a result, an accurate and robust shape description should be invariant, under the most common deformations, in order to be of practical relevance.

Many approaches have been proposed to allow for the invariant description of macromolecular surfaces. The reader is referred to [8] for a complete review. For instance, some of these methods are based on the Bayesian analysis of backbone deformations [9]. Other approaches employ the spherical harmonics description of the molecular surface [10] or concern the projection of the molecular surface on Zernike functions [11]. Further, there are techniques that utilise the convolution kernel [12] or consider the heat propagation on the molecular surface [13] and [14]. Yet, all of these methods have the same drawback in that they fail to put disconnected regions into relation. Recall that this ability is a required characteristic in order to describe allosteric effects. Only recently [15] has proposed such an approach (although not invariant). However, their method provides a description which is restricted to the residues, while our work as detailed in this paper considers the entire macromolecular surface.

In this paper, we propose an approach based on discrete differential geometry [16,17] and the Fokker–Planck equation [18], from which an invariant, compact and informative shape signature

Q2 * Corresponding author at: Vaccines Program, National Research Council, 1200 Montreal Road, Ottawa, Canada. Tel.: +1 6139935242.
E-mail addresses: eric.paquet@nrc-cnrc.gc.ca (E. Paquet), hviktor@uottawa.ca (H.L. Viktor).

http://dx.doi.org/10.1016/j.vaccine.2015.08.098
0264-410X/

(or descriptor) is extracted. Our method provides invariance under isometry (non-elastic deformations). Further, it may also be made invariant under local dilatation, by following an approach based on Gaussian curvatures [19]. From a macromolecular point of view, an isometric invariance corresponds to a macromolecular surface formed by a very large, potentially infinite, number of rotamers or hinges [6] which account for non-elastic deformations. Consequently, this invariance is a generalisation of the backbone and side chains rotation invariance which is at the basis of macromolecular docking algorithms such as the FireDock [6] method.

Our shape signature thus allows for the description of disconnected regions, such as those associated with allosteric effects [20]. This is important because, despite of the fact that the conformational deformations are random, they may be correlated in between themselves over regions which are apparently disconnected. This implies that a purely local shape analysis may not be entirely satisfactory in that particular case.

Our paper is organised as follows. In Section 2, we apply discrete differential geometry to macromolecular surfaces in order to obtain an intrinsic description. In Section 3, we review some of our earlier results for the description of macromolecular surfaces in terms of the heat equation. This approach is reformulated in terms of a Gaussian random walk in Section 4. This random walk is generalised, in Section 5, in order to obtain a non-local, isometrically invariant description of macromolecular surfaces based on the fractional Fokker–Planck equation. Experimental results for the influenza neuramidase are presented in Section 6, while a conclusion follows in Section 7.

2. Macromolecular surface geometry and discrete differential geometry

In this section, we present the mathematical framework from which an isometrically invariant description of a macromolecular surface may be obtained. We assume that the macromolecular surface is represented in terms of a triangular tessellation (triangular mesh) which may be assimilated to a discrete graph. This is the most common representation adopted amongst molecular modelling software.

In order to apply our approach, some notions of discrete differential geometry are required which are briefly reviewed here. We chose differential geometry as our mathematical framework for two reasons. Firstly, the differential geometry allows for an intrinsic description (i.e. no external reference frames) of the macromolecular surface which is required in order to obtain invariance under isometric deformations. Secondly, because the description is intrinsic, the geometry of the macromolecular surface becomes Riemannian (i.e. a curved space as the molecule is not embedded in Euclidean space anymore) which means that the mathematical description should be consistent with such geometry. One of the most important notions in differential geometry is de Rham operator (also known as Laplace–Beltrami operator which is a generalisation of the Laplacian) which characterises both the geometry (the metric aspect, e.g. distances, scalar products) and the topology (holes, handles, etc.) of the underlying manifold. The de Rham operator is defined as:

$$\Delta f = (dd^* + d^*d)f \tag{1}$$

where $df = (\partial f_{\mu_1\dots\mu_k}/\partial x^\mu)dx^\mu \wedge dx^{\mu_1} \wedge \dots dx^{\mu_k}$ is the exterior derivative, f a differential k-form and $\wedge$ is the exterior product, while d^* (also written as δ) is the codifferential which is defined as

$$d^* = -1^{kn+n+1} * \quad d \quad * \tag{2}$$

where $*\omega_{\mu_1\dots\mu_{n-k}} = 1/k! \quad \omega^{\nu_1\dots\nu_k}\sqrt{|\det g|}\ \varepsilon_{\nu_1\dots\nu_k\mu_1\dots\mu_{n-k}}$ is the dual operator, g is the metric associated with the macromolecular surface, ε is the completely antisymmetric tensor and n is the dimensionality of the manifold which is two for a macromolecular surface. For a scalar function or 0-form, the de Rham operator reduces to:

$$\Delta = d^*d \tag{3}$$

The discrete counterpart of differential geometry, discrete differential geometry [16,17], which is required as a discrete representation is used for the macromolecular surface, is based on the concept of the incidence matrix. Let σ_j^p be a p-simplex or cell. For instance, a 0-simplex is a vertex or node, a 1-simplex is an edge and a 2-simplex is a triangle. The incidence matrix $\mathbf{N}_p^T$ encodes the relationships in between the p-cells and the $(p-1)$-cells and is defined as

$$(\mathbf{N}_p)_{ij} = \begin{cases} 0 & \Leftrightarrow \quad \sigma_j^{p-1} \notin \partial\sigma_i^p \\ 1 & \Leftrightarrow \quad \mathcal{O}(\sigma_j^{p-1}) = \mathcal{O}(\sigma_i^p) \\ -1 & \Leftrightarrow \quad \mathcal{O}(\sigma_j^{p-1}) = -\mathcal{O}(\sigma_i^p) \end{cases} \tag{4}$$

where $\mathcal{O}$ is the orientation operator which is equal to +1 when the orientation of the p-simplex is positive and −1 otherwise. Consequently, $\mathbf{N}_p$ is equal to zero if the $(p-1)$-simplex j is not in the neighbourhood of the p-simplex i. Secondly, it is equal to +1 if the orientation of the $(p-1)$-simplex j is compatible with the one of the p-simplex i. Finally, $\mathbf{N}_p$ is equal to −1 if their respective orientations are not compatible. In the particular case of an edge and a vertex, the incidence matrix is equal to 1 if the edge is entering the vertex and −1 otherwise.

From the incident matrix, it is possible to define a discrete exterior derivative and a discrete codifferential [16,17] by introducing a discrete dual and by associating a metric to each p-simplex:

$$\begin{aligned} d_p &\sim \mathbf{N}_p^T \\ d_p^* &\sim (\mathbf{N}_{p+1}^T)^* = *\mathbf{N}_p^T* = \mathbf{G}_{n-p+1}^*\mathbf{N}_p^T\mathbf{G}_p^{-1} \end{aligned} \tag{5}$$

where $\mathbf{G}_p$ is the metric (a symmetric matrix) associated with the p-simplex. From these equivalences, we may define the discrete de Rham operator as [16,17]:

$$\mathbf{L}_p = \mathbf{N}_p\mathbf{N}_p^* + \mathbf{N}_{p+1}^*\mathbf{N}_{p+1} \sim \Delta \equiv dd^* + d^*d \tag{6}$$

In our work, we are more interested in the Laplacian associated with the nodes, since our description is based on the Fokker–Planck equation which only required the node Laplacian [21]. In this case, the de Rham operator takes a particularly simple form:

$$\mathbf{L}_0 = \mathbf{N}_0\mathbf{N}_0^* + \mathbf{N}_1^*\mathbf{N}_1 = \mathbf{N}_0\mathbf{A}^T\mathbf{G}_1^{-1}\mathbf{A} \tag{7}$$

The metrics may be defined in various ways [16,17]. In our case, the metric associated with the edges $\mathbf{G}_1$, is defined as the affinity in between connected vertices:

$$(\mathbf{G}_1)_{ij} = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right] \tag{8}$$

where $\sigma = \text{med}\{\|x_i - x_j\|\}$ is the median of the length of the edges [22] while $\mathbf{G}_0$, the metric associated with the vertices, is a diagonal matrix of the areas of the neighbourhood associated with each vertex [16,17]. The neighbourhood may be defined either as the 1-ring neighbourhood (the total area of all the triangles connected to vertex i) or from the dual cell associated with a particular vertex. It may be constructed, for instance, by joining the barycentres of all the triangles connected to vertex i through their respective common edges.

Please cite this article in press as: Paquet E, Viktor HL. Isometrically invariant and allosterically aware description of deformable macromolecular surfaces: Application to the viral neuraminidase. Vaccine (2015), http://dx.doi.org/10.1016/j.vaccine.2015.08.098