Journal of Molecular Graphics and Modelling 53 (2014) 72–81

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Skeleton-based shape analysis of protein models☆

Zhong Li[a,*], Shengwei Qin[a], Zeyun Yu[b], Yao Jin[a]

[a] College of Science, Zhejiang Sci-Tech University, Hangzhou 310018, China
[b] Department of Computer Science, University of Wisconsin, Milwaukee 53211, USA

ARTICLE INFO

Article history:
Accepted 30 June 2014
Available online 12 July 2014

Keywords:
Protein shapes
Shape descriptor
Similarity comparison
Skeleton
Local diameter

ABSTRACT

In order to compare the similarity between two protein models, a shape analysis algorithm based on skeleton extraction is presented in this paper. It firstly extracts the skeleton of a given protein surface by an improved Multi-resolution Reeb Graph (MRG) method. A number of points on the model surface are then collected to compute the local diameter (LD) according to the skeleton. Finally the LD frequency is calculated to build up the line chart, which is employed to analyze the shape similarity between protein models. Experimental results show that the similarity comparison using the proposed shape descriptor is more accurate especially for protein models with large deformations.

1. Introduction

Proteins are vital in cells and play a special role in many biological activities. The surface shape of a protein defines a geometric and biochemical domain where the protein interacts with other proteins or its environment, and thus is an important characteristic of the protein. The similarity comparison between protein shapes had recently become an important means of protein analysis, which can reveal the structural and functional relationship of the associated proteins. With the development of computer graphics, the shape similarity comparison between protein molecules has been applied in many fields, such as computer aided molecular design, drug discovery, protein structure retrieval and so on [1,2].

Many researchers have proposed different methods on shape similarity comparison for surface models and some of them are specifically for protein shape-based analysis. These methods can roughly be classified into three categories [3]. The first one is the similarity comparison based on the outline. For example, Osada et al. [4] provided a shape distribution method based on the statistical histogram to measure the whole model shape. Horn [5] applied the extended Gaussian image mapping for similarity comparison of convex polyhedron models. Michael et al. [6] proposed a functional analysis method according to the spherical harmonic expansion to compare the protein binding pocket and ligand. Kazhdan et al. [7] put forward an algorithm of analyzing geometric structures by using all kinds of frequent characteristics of three-dimensional shapes to retrieve these models.

The second category of methods on shape comparison is based on shape projections. Min et al. [8] proposed a method using a 2D sketch interface for a 3D model search engine. It mainly does a projection transformation in different directions on three-dimensional models, and gets a series of two-dimensional projection images for model retrieval. However, it has some limitations. For example, it can only describe the feature of brightness distributions and cannot reflect topology characteristics.

The third category of methods on shape comparison is based on geometric features of three dimensional models. Reuter et al. [9] proposed a spectral method using the Laplace–Beltrami operator, which possesses an isometry-invariant global geometric property. Other researchers use topological structures to compare the model shapes. For example, Hilaga et al. [10] provided a shape comparison method by using the multi-resolution Reeb graph (MRG). Though the MRG method describes the topological characteristics, it does not cover full geometric information (for example holes) and may get the mismatching for deformed models [10]. Therefore, for the protein models with holes or with the deformed shape, directly using this method may not obtain satisfactory similarity results. Some researchers used an efficient computation of a simplified medial axis for shape analysis [11,12]. The problem of the medial axis is that the computational complexity is large and sensitive to the holes on 3D models. Fang et al. [1] applied the local-diameter (LD) descriptor to compare the similarity of different flexible proteins. The LD method can improve the efficiency and quality of

☆ This research was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LY14A010032 and Scientific Research Foundation of Ministry of Education of China under Grant No. [2009]1590.

* Corresponding author. Tel.: +86 571 86843224; fax: +86 571 86843224.
E-mail addresses: lizhong@zstu.edu.cn, lizhongzju@hotmail.com (Z. Li).

http://dx.doi.org/10.1016/j.jmgm.2014.06.012
1093-3263/

similarity analysis, but there are still some problems. For example, it may not be precise for the deformed protein models with large topological changes.

In this paper we propose a new protein shape analysis method. It adopts the MRG algorithm based on collision detection and plane segmentation to obtain the skeleton for approximating the protein shape. The skeleton can describe the whole topological characteristics and optimize the details well. Also, it is insensitive to model noise and large topological changes caused by protein deformations. Based on the skeleton, we calculate the local-diameter of some sample points on the model surface to conduct similarity comparison. The experiment demonstrates that the shape analysis is invariant for rotations and translations of the protein models, and is also robust to shape deformations.

2. Materials and methods

2.1. Skeleton extraction for protein molecule models

There are many methods to extract the linear skeleton of a three-dimensional model [10,13]. Here we propose an improved MRG (Multi-resolution Reeb Graph) algorithm based on the collision detection and plane segmentation for protein molecule models.

2.1.1. MRG algorithm

The basic idea of the MRG algorithm [10] is to find a good function μ, for example, the height function, to extract the skeleton by the Reeb graph. The function μ can be defined for each vertex v as follows

$$\mu = \sum g(v, b_i) \cdot area(b_i) \quad (1)$$

where $\{b_i\}$ is the fundamental point set sampled on the surface of the model, $g(v, b_i)$ is the geodesic distance between a vertex v and the fundamental points b_i, and $area(b_i)$ is the local surface area related on each point b_i.

Then we can simplify the function μ as

$$\mu = \sum g(v, b_i) \cdot C \quad (2)$$

where C is the average area defined as $C = S/N$, where S is the surface area of the model and N is the number of fundamental points. In practice, C can be chosen as a constant. In our experiments we set it as 1.

After the function μ of each vertex is calculated from (2), we then normalize the value of the function μ and divide the values of the function μ into k ranges. The point set in each range is then constructed. We finally replace each point set by a joint point according to the barycentric coordinates and automatically connect these joint points based on the connecting relationship to extract the skeleton. Fig. 1 is the skeleton results for two protein molecular models by using the above process.

We find the traditional MRG algorithm is easy to implement, but for some protein molecular models especially with holes, the skeleton results sometimes cannot effectively preserve the topology and shape feature. For example, the skeleton may not be located in the center of the model and the skeleton may beyond the body of the model (see Fig. 1). In the next subsection we aim to improve the MRG algorithm to solve these problems.

2.1.2. Our improved MRG algorithm for the skeleton extraction

In order to extract more accurate skeletons for protein molecular models, we provide an improved MRG algorithm based on the collision detection and plane segmentation. The main algorithm steps are as follows.

Step (1) We extract the origin skeleton R from model M based on the traditional MRG algorithm, save the joint points and record the numbers of joint points as N.
Step (2) We check whether each skeleton segment composed of two neighboring joint points (g_i, g_{i+1}) (where $i < N$) is beyond the surface of the model. If it stays in the model, go to step (4); otherwise, go to step (3).
Step (3) If the skeleton segment is beyond the surface of the model, we calculate the intersection points, record them and count the number S of these intersections. Then we do the following operations according to the number of intersections.
(i) If $S = 1$, there are three cases. Case 1: one endpoint of segment (g_i, g_{i+1}) is out of the object and another one is in the object. We compute the midpoint between the intersection point and

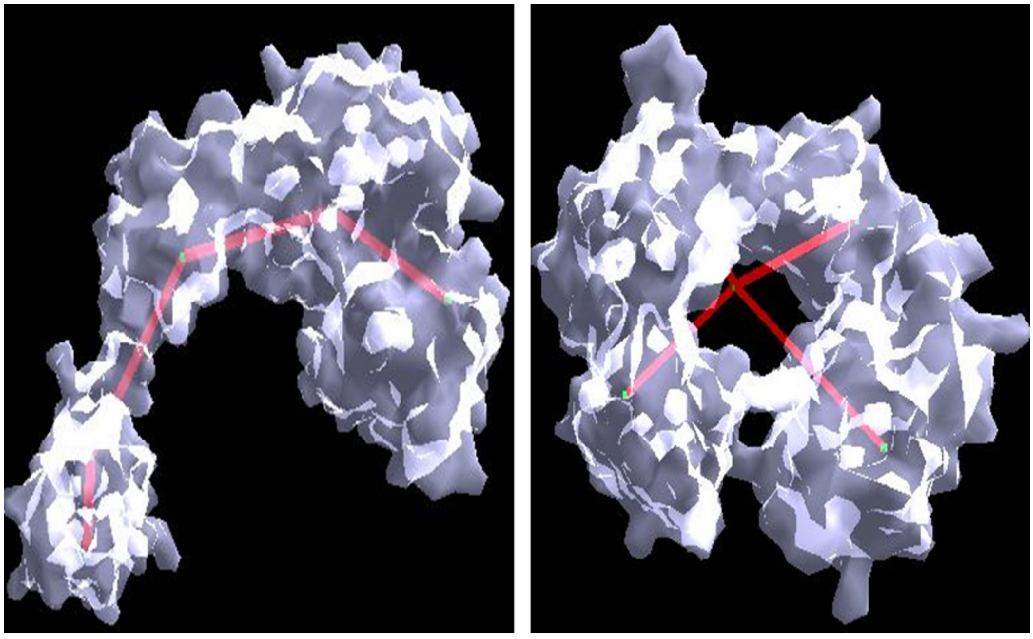

Fig. 1. Skeleton extraction based on the traditional MRG algorithm.