Computer-Aided Design 44 (2012) 835–845

Contents lists available at SciVerse ScienceDirect

Computer-Aided Design

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

QTF: Quasi-triangulation file format

Deok-Soo Kim [a,b,*], Youngsong Cho [b], Jae-Kwan Kim [a,b], Joonghyun Ryu [b]

[a] Department of Industrial Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, South Korea
[b] Voronoi Diagram Research Center, Hanyang University, Seoul, South Korea

ARTICLE INFO

Article history:
Received 19 October 2011
Accepted 7 April 2012

Keywords:
Voronoi diagram of spherical balls
Quasi-triangulation
Neutral file format
Molecular modeling
Structural/computational molecular biology
Protein Data Bank (PDB)

ABSTRACT

A quasi-triangulation is the dual structure of the Voronoi diagram of spherical balls and its properties and algorithms are well-studied in three-dimensional space. Quasi-triangulation has been used for efficiently solving various structure/shape related problems for biomolecules. The computation of the quasi-triangulation directly from an input file can take a significant amount of time. If the quasi-triangulation is computed a priori and stored in a file, an application software can directly load the file for solving application problems. In this paper, we propose a neutral file format, called the quasi-triangulation file format QTF, so that users can use the quasi-triangulation more effectively and efficiently by focusing more on his or her own application problems than the Voronoi diagram or the quasi-triangulation itself. The proposed QTF file format was thoroughly validated through an extensive experiment by computing the molecular volumes of one hundred molecular models in the Protein Data Bank. This approach has an important consequence: The QTF file format separates the computation of the Voronoi diagram from its applications.

1. Introduction

The Voronoi diagram of three-dimensional spheres and the corresponding quasi-triangulation have recently found important applications in various areas, particularly in computational chemistry/biology and structural molecular biology [1–4]. The Voronoi diagram of three-dimensional spherical balls, called *atoms*, represents the proximity among the atoms and its robust and efficient algorithms are now well-known [5,6]. The topological dual of the Voronoi diagram of atoms is called the *quasi-triangulation* and its useful properties have been reported with accompanying data structures [7–9]. Based on the quasi-triangulation, two useful geometric constructs called the *beta-complex* and the *beta-shape* are defined. The beta-complex is the subset of the quasi-triangulation where each of its simplexes represents the proximity among atoms within and on the boundary of a molecule. The boundary of a molecule is defined with respect to the radius of a probe representing solvent molecules. The beta-complex is very useful for applications due to its dual properties: it is a concise abstraction of an atom set while it has a correct and complete proximity information among the atoms. The beta-shape is the region of the three-dimensional space bounded by the boundary of the beta-complex and represents the proximity among the boundary atoms. Given the quasi-triangulation, the beta-complex and the beta-shape can be efficiently computed so that they can be used to solve various, otherwise difficult if not impossible, spatial reasoning problems among atoms. For details of the beta-complex, the readers may refer to the work of Kim et al. [8,10]. Hence, the quasi-triangulation becomes the core of the post-processing of the Voronoi diagram.

There can be two approaches to make the Voronoi diagram and the quasi-triangulation available for applications. As shown in Fig. 1(a), the first, obvious approach is to compute the Voronoi diagram of atoms (in $O(n^3)$ time) from scratch with an input set of n atoms and then convert it into the corresponding quasi-triangulation (in $O(m)$ time where m represents the number of simplexes in the quasi-triangulation). For molecular biology applications, the input set is usually a molecular structure stored in the Protein Data Bank (PDB) [11,12]. Fig. 1(b) shows the second approach that is to upload the quasi-triangulation (in $O(m)$ time) from an a priori computed and stored file in the quasi-triangulation database (QTDB). Then, the uploaded quasi-triangulation can be converted into the Voronoi diagram, if necessary (in $O(m)$ time).

Given the quasi-triangulation, the beta-complex can be computed in $O(\log n + k)$ time if a binary search is used for a sorted simplexes [8]. Being the boundary of the space taken by the beta-complex, the boundary of the beta-shape can be also easily computed from the beta-complex. A more general approach for the simplex query for the quasi-triangulation and the beta-complex was recently reported [13].

The time complexities above are all for the worst case. We recommend to adopt the second approach for most practical

* Corresponding author at: Department of Industrial Engineering, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul 133-791, South Korea. Tel.: +82 2 2220 0472; fax: +82 2 2292 0472.
E-mail address: dskim@hanyang.ac.kr (D.-S. Kim).

0010-4485/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cad.2012.04.003

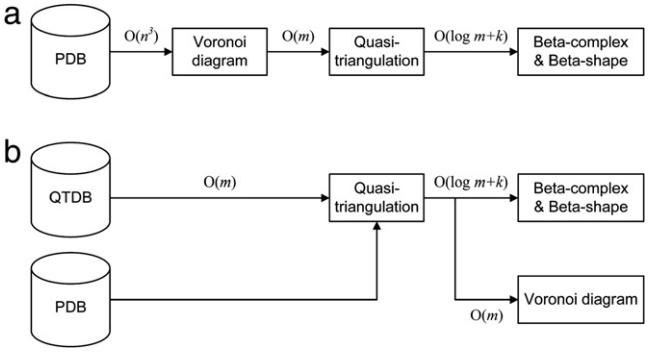

Fig. 1. Two approaches to make the quasi-triangulation available for applications. n: the number of atoms in a molecule; m: the number of simplexes in the quasi-triangulation; k: the number of simplexes in the beta-complex.

biology problems. For this purpose, it is necessary to define the file format of the quasi-triangulation without any ambiguity.

In this paper, we propose a very simple neutral file format, the quasi-triangulation file format QTF (using the file extension `qtf`), for storing the topology of the Voronoi diagram in its dual structure, the quasi-triangulation. The rationale behind this proposal is explained in Section 2. We have actually built and made available to the public the QTDB containing the quasi-triangulation files for all molecular structures of the PDB: See Voronoi Diagram Research Center, abbreviated as the VDRC (http://voronoi.hanyang.ac.kr/). Hence, researchers in relevant application domains (e.g. biologists, chemists, or physicists) who are not necessarily experts in geometric computations and computer programming, can download the quasi-triangulation files of interested molecules from the QTDB and easily write application programs to solve any structure- or geometry-related problems of biomolecules. This approach frees application experts from tedious algorithm/programming work because most computations can be done in a rather straightforward fashion using the simplexes in the quasi-triangulation which are vertices, edges, triangular faces, and tetrahedral cells. Hence, we claim that the QTF is a significant advance in the application of the Voronoi diagram of atoms.

On our web site, we have made a few softwares developed at the VDRC freely available to the public. `BetaMol` is a powerful molecular modeling, analysis, and visualization software which is based on the theory of the beta-complex and the Voronoi diagram [14]. Loading a molecular structure stored in a file format such as PDB, `BetaMol` computes the Voronoi diagram and transforms to the quasi-triangulation if a corresponding `qtf` file is not available. If the `qtf` file for a particular molecule is available, `BetaMol` reads in both the `qtf` file and the corresponding PDB file and synchronizes both. `QTFier` is a software that reads in a PDB structure file for a molecule and computes its Voronoi diagram to transform into the corresponding quasi-triangulation to store in the `qtf` file. Hence, readers may use either software to get the quasi-triangulation of a molecule of interest.

Once the definition and properties of the quasi-triangulation and the QTF are correctly understood, it is easy to develop algorithms and software using the quasi-triangulation. Therefore, it is important to understand the basic theory of the Voronoi diagram and the quasi-triangulation: For the Voronoi diagram, the readers may refer to the work of Kim et al. [5,15]; for quasi-triangulation, the readers may refer to the work of Kim et al. [7–9].

2. Why the quasi-triangulation file format?

There are three important strategic reasons about our proposal of the quasi-triangulation file format as follows:

(1) *Separation of the topology computation and applications*: By devising a neutral file format for the quasi-triangulation, it is possible to separate the effort to study the algorithms (and their implementations) for computing the quasi-triangulation (or the Voronoi diagram of atoms) from the effort to use the topology and geometry of the quasi-triangulation for solving biomolecular problems in application softwares. As the research of the Voronoi diagram continues, new and more powerful algorithms with better performances in both a computational speed, a solution quality, and a robustness will be developed. Having defined a neutral file format for the quasi-triangulation, different softwares/algorithm may be used to compute the Voronoi diagram and store it in the quasi-triangulation file without influencing the valuable application softwares developed by biologists, chemists, and physicist who are not necessarily experts in geometric computation and computer programming.

(2) *A priori fixed model*: In most biomolecular applications, the structure of a target protein is fixed a priori and various kinds of analysis are performed on the structure. The data size of the target protein is usually large and repeatedly referred to during the computation for an analysis. Therefore, it is more convenient to read in the topology of the quasi-triangulation directly from a file, rather than to compute it on the fly, where the quasi-triangulation is a priori computed and stored via preprocessing. For example, in a protein-ligand docking and a structure-based virtual screening which are examples with a particular importance in drug design, a fixed target protein is usually determined beforehand and many small compounds in a database are attempted to measure their fit to the target such as a shape complementarity.

(3) *Direct application of the quasi-triangulation*: For many biomolecular applications, the quasi-triangulation itself can be used for solving many structure- or geometry-related problems without converting it to the corresponding Voronoi diagram. The computation of the beta-complex and the beta-shape are good examples for such cases. According to our experience, most application problems in molecular biology only need knowledge of the topology among atoms; there are very few problems additionally requiring the geometry of the Voronoi diagram beyond the topology. Computation of the volume of *atomic occupation* is an example of the latter case where the atomic occupation is the region of the space that an atom occupies in competition with others.

In addition, there are other significant operational reasons as follows:

(4) The computation of the Voronoi diagram of spherical balls in $\mathbb{R}^3$ is relatively time consuming because the algorithm requires $O(n^3)$ time in the worst case where n is the number of input balls. Even though the computation takes $O(n)$ time on average for molecules [16–18], it still takes a significant amount of computation time for moderate-sized molecules. According to our experience (with the regression analysis), the computation time T for computing the Voronoi diagram of a molecule roughly follows $T = 0.001n$ (s) using the computation environment used in this experiment (for details, see Section 7). Hence, on this machine, a little more than 1000 atoms take one second for the Voronoi diagram computation. For details of the algorithm for computing the Voronoi diagram, see [5].

(5) From the quasi-triangulation, the topology of the Voronoi diagram can be easily reconstructed in $O(m)$ time in