Approximating the Pathway Axis and the Persistence Diagram of a Collection of Balls in 3-Space*

[Extended Abstract]

Eitan Yaffe
School of Computer Science
Tel-Aviv University, Israel
eitanyaf@post.tau.ac.il

Dan Halperin
School of Computer Science
Tel-Aviv University, Israel
danha@post.tau.ac.il

ABSTRACT

Given a collection $\mathcal{B}$ of balls in three-dimensional space, each having a radius of at least 1, we present an approximation scheme that constructs a collection $\mathcal{K}_\varepsilon$ of unit balls that approximate $\mathcal{B}$, such that the Hausdorff distance between $\cup\mathcal{B}$ and $\cup\mathcal{K}_\varepsilon$ is at most ε. We define the *pathway axis* as the subset of the medial axis of the complement of $\cup\mathcal{B}$ for which the set of closest balls in $\mathcal{B}$ do not have a common intersection. It is the medial axis of the complement of $\cup\mathcal{B}$ without 'dead-ends' and therefore it is a good starting point for finding pathways that lie outside $\cup\mathcal{B}$. The recently introduced *persistence diagram* of the distance function from $\cup\mathcal{B}$ encodes topological characteristics of the function, giving a measure on the importance of topological features such as voids or tunnels during a uniform growth process of $\mathcal{B}$. In this paper we introduce the *pathway diagram* as a useful subset of the Voronoi diagram of the centers of the unit balls in $\mathcal{K}_\varepsilon$, which can be easily and efficiently computed. We show that the pathway diagram contains an approximation of the pathway axis of $\mathcal{B}$. We prove a bound on the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$, namely the ratio between the number of unit balls in $\mathcal{K}_\varepsilon$ and the number of balls in $\mathcal{B}$. We employ this bound to show how we efficiently approximate the persistence diagram of $\cup\mathcal{B}$. Finally, we show that our approach is superior to the standard point-sample approaches for the two problems that we address in this paper: Approximating the medial axis of the complement of $\cup\mathcal{B}$, and approximating the persistence diagram of $\cup\mathcal{B}$. In a companion paper we introduce MolAxis, a tool for the identification of channels in macromolecules, that demonstrates how the pathway diagram and the persistence diagram are used to identify pathways in the complement of molecules.

*This work has been supported in part by the IST Programme of the EU as Shared-cost RTD (FET Open) Project under Contract No IST-006413 (ACS - Algorithms for Complex Shapes), by the Israel Science Foundation (Grant no. 236/06), and by the Hermann Minkowski–Minerva Center for Geometry at Tel-Aviv University.

SCG'08, June 9–11, 2008, College Park, Maryland, USA.

Categories and Subject Descriptors

G.1.2 [**Approximation**]: Approximation of surfaces and contours; I.3.5 [**Computational Geometry and Object Modeling**]: Geometric algorithms, languages, and systems; I.3.5 [**Computational Geometry and Object Modeling**]: Curve, surface, solid, and object representations

General Terms

Algorithms, Theory, Experimentation

Keywords

Medial axis, alpha complex, molecular modeling, persistence diagram

1. INTRODUCTION

Let $\mathcal{B}$ be a finite collection of balls in $\mathbb{R}^3$ and let $\cup\mathcal{B}$ denote their union. We assume, without loss of generality, that every ball in $\mathcal{B}$ is not smaller than a unit ball. We wish to capture the shape of the complement of $\cup\mathcal{B}$. The *medial axis* of the complement of $\cup\mathcal{B}$ is the set of points in this complement that have more than one closest ball in $\mathcal{B}$. We define the *pathway axis* of $\mathcal{B}$ to be the set of points in the complement of $\cup\mathcal{B}$ for which the collection of (more than one) closest balls in $\mathcal{B}$ do not have a common intersection. It is the medial axis of the complement of $\cup\mathcal{B}$ without 'dead-ends' and therefore it is a good starting point for finding pathways that lie outside $\cup\mathcal{B}$. See Figure 1 for a two-dimensional illustration.

The exact medial axis of the complement of the union of balls is a subset of the Voronoi diagram [6, 8] of the balls and can be computed in an exact manner as shown by Boissonnat and Delage [7]. We opt for a piecewise linear approximation approach since the implementation is simpler. The underlying structure is the Voronoi diagram of points in space for which robust implementation is readily available. Moreover, since we further manipulate the diagram in our motivating application, we anticipated that it would be easier for us to work with polygons (the facets of the approximate structure) than to work with the curved facets of the Voronoi diagram of the spheres. In this paper we offer an approximation scheme that replaces $\mathcal{B}$ by a collection of unit balls $\mathcal{K}_\varepsilon$ such that the Hausdorff distance between $\cup\mathcal{B}$ and $\cup\mathcal{K}_\varepsilon$ is not larger than a prescribed ε. We introduce an algorithm that constructs a geometric entity we call the *pathway diagram* of the centers of $\mathcal{K}_\varepsilon$, and prove that the pathway diagram contains an approximation of the pathway axis of $\mathcal{B}$.

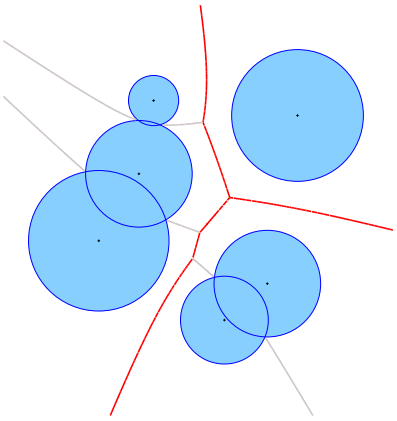


Figure 1: The pathway axis of a collection of discs (bold line). The discarded portions of the medial axis are colored gray.

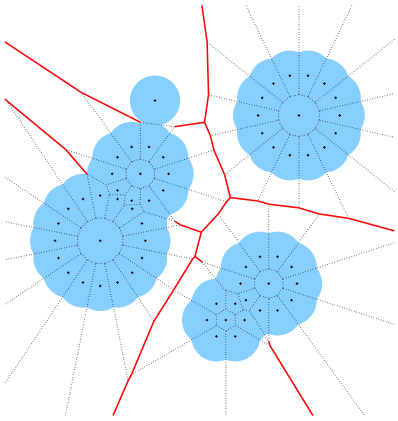


Figure 2: The pathway diagram (bold line) of a collection same-size discs that approximate the discs depicted in Figure 1. The discarded portions of the Voronoi diagram of their centers is depicted using dotted lines.

There are algorithms that approximate the medial axis of an object from a set of unorganized points sampled on the surface of the object [2, 15]. Oudot and Boissonnat [26] introduce an algorithm for computing the medial axis that has certified results for smooth shapes. However, the complement of the union of a collection of balls is not bounded by a smooth surface, making it difficult to directly apply the techniques (and hence have the topological guarantees) obtained in the papers cited above. In contrast to the aforementioned approaches we sample a volume with balls instead of sampling a surface with points.

There are algorithms that focus on a subset of the medial axis. Giesen *et al.* [22] approximate a useful subset of the medial axis of a shape with smooth boundary that captures the topology of the shape. The λ-medial axis [10, 11], introduced by Chazal and Lieutier, is a subset of the medial axis, that for some "regular" values of λ remains stable under Hausdorff distance perturbation. This leads to an algorithm [11] that constructs an approximation of the λ-medial axis of an object from a set of noisy unorganized points sampled on or close to the (not necessarily smooth) boundary surface of the object. We apply theoretical ideas introduced there to prove geometric convergence of our approximation.

A result by Lieutier [24] states that under certain conditions the medial axis of an object and the object itself have the same 'shape' (they are *homotopy equivalent*). Yet in order to apply this result the sample quality depends on the *weak feature size* [10], which would require a dense sample and downgrade the performance dramatically. Instead, in order to state topological properties of our approximation scheme, we utilize the emerging concept of topological persistence. Edelsbrunner *et al.* [19] introduce the notion of topological persistence during a growth process of the union of balls. In that work an efficient algorithm is described that classifies topological changes during the growth process as topological features or topological noise depending on their lifetime during the process. The theoretical notion of persistence was extended independently by Carlsson *et al.* [9], by Chazal *et al.* [12], and by Cohen-Steiner *et al.* [13]. Cohen-Steiner *et al.* [13] introduce the *persistence diagram* of a real-valued function on a topological space, which encodes topological characteristics of the function, giving a measure on the importance of topological features. They show that the persistence diagram is stable under Hausdorff perturbation.

In this paper we define the *pathway diagram* as a subset of the Voronoi diagram of the centers of $\mathcal{K}_\varepsilon$, which is closely related to the celebrated α-complex introduced by Edelsbrunner *et al.* [20]; technically it is the collection of Voronoi facets dual to simplices that are *not* included in the $(\alpha = 1)$-complex of the centers of $\mathcal{K}_\varepsilon$. See Figure 2 for a two-dimensional illustration of a pathway diagram. We present an algorithm that computes it and prove several of its properties. As mentioned above, we prove that the pathway diagram contains an approximation of the pathway axis. We provide a bound on the number of balls in $\mathcal{K}_\varepsilon$ as a function of ε and the ratio between the largest and the smallest ball in $\mathcal{B}$. We employ this bound to show that the persistence diagram of $\cup\mathcal{B}$ can be approximated using $O(|\mathcal{B}|/\varepsilon^{4/3})$ unit balls. Finally, we compare our approach with a point-sample approach, proving a multiplicative-factor gain of $\Omega(1/\varepsilon^{2/3})$ in the worst case in the number of sample entities over the point samples for the approximation of both the pathway axis and the persistence diagram. In the last section we report on experimental results and show how the persistence diagram is used to identify the center of the largest chamber within a molecule, presented as a collection of balls.

In a companion paper [29] we present MolAxis, a new tool designed for the efficient identification of molecular channels. MolAxis was found to be a very efficient and accurate tool in identifying transmembrane (TM) channels and pathways leading from buried cavities within enzymes to outside of their convex hull. Being very fast, MolAxis has been successfully used to analyze channel dimensions and lining residues in hundreds of snapshots of Molecular Dynamic simulation of the human CYP3A4.

More details on the results described in this extended abstract, including proofs omitted for lack of space, can be found in the thesis [28]: `http://www.cs.tau.ac.il/~eitanyaf/thesis.pdf`. The MolAxis web server is available at: `http://bioinfo3d.cs.tau.ac.il/MolAxis`.

2. PRELIMINARIES

Our work builds on a large body of results concerning Voronoi diagrams and the medial axis. We assume some familiarity with alpha complexes [20]. We borrow notation mainly from the work of Attali *et al.* [4] and the work of Chazal and Lieutier [11]. For any set X we denote by $\bar{X}$, X^o, ∂X, X^c and $|X|$ the closure, the interior, the boundary, the complement and the cardinality of X respectively. $B(x,r)$, $B^o(x,r)$ and $S(x,r)$ denote the closed ball, open ball and sphere of center x and radius r in $\mathbb{R}^3$ respectively. We denote the Euclidean distance between two points $x, y \in \mathbb{R}^3$ by $d(x,y)$. The distance between two subsets A, B of $\mathbb{R}^3$ is defined to be $d(A,B) = \inf_{a\in A, b\in B} d(a,b)$.

The *one-sided Hausdorff distance* between two compact subsets A and B of $\mathbb{R}^3$ is:

$$d_H(A|B) = \sup_{x\in A} d(x,B) \ .$$

The *(symmetric) Hausdorff distance* between two compact subsets A and B of $\mathbb{R}^3$ is the maximum of the two one-sided distances, namely $d_H(A,B) = \max(d_H(A|B), d_H(B|A))$. We say that A is a *Hausdorff approximation* of B with an *approximation resolution* of ε if the Hausdorff distance between A and B is not larger than ε. In such a case, we will say for short that A is an ε-approximation of B.

Let $\mathcal{O}$ be a bounded open subset of $\mathbb{R}^3$. For any point $x \in \mathcal{O}$, we denote by $\Gamma_{\mathcal{O}}(x)$ the set of closest points to x in the complement $\mathcal{O}^c$, namely $\Gamma_{\mathcal{O}}(x) = \{y \in \mathcal{O}^c : d(x,y) = d(x,\mathcal{O}^c)\}$. The *medial axis* $M[\mathcal{O}]$ of the open set $\mathcal{O}$ is the set of points $x \in \mathcal{O}$ that have at least two closest boundary points:

$$M[\mathcal{O}] = \{x \in \mathcal{O} : |\Gamma_{\mathcal{O}}(x)| \geq 2\} \ .$$

Let E be a finite point set in $\mathbb{R}^3$. We denote by $\mathcal{V}[E]$ the collection of *Voronoi cells* of E of dimension 0,1,2 or 3, and by $\mathcal{D}[E]$ its dual structure, the *Delaunay triangulation* of E which contains simplices which are of dimension 0,1,2 or 3. Note that the Delaunay triangulation of E is a *simplicial complex* (see, e.g., [6, 17]).

The rest of the section is devoted to the formal definition of the persistence diagram. We repeat verbatim the definitions introduced by Cohen-Steiner *et al.* [13], which we need in order to state our results in Section 6. The reader is referred to [25] for an accessible introduction to Homology.

Given a topological space X and an integer k, we denote the k-th singular homology group of X by $H_k(X)$, and the k-th Betti number of X by $\beta_k(X) = \dim H_k(X)$. We work here with modulo 2 coefficients, so that homology groups are vector spaces over $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$.

DEFINITION 2.1. **[13]** *Let X be a topological space and f a real function on X. A homological critical value of f is a real number A for which there exists an integer k such that for all sufficiently small $\varepsilon > 0$ the map $H_k(f^{-1}(-\infty, A-\varepsilon]) \to H_k(f^{-1}(-\infty, A+\varepsilon])$ induced by inclusion is not an isomorphism.*

DEFINITION 2.2. **[13]** *A function $f : X \to \mathbb{R}$ is tame if it has a finite number of homological critical values and the homology groups $H_k(f^{-1}(-\infty, A])$ are finite-dimensional for all $k \in \mathbb{Z}$ and $A \in \mathbb{R}$.*

In other words, the homological critical values are the levels where the homology of the sub-level sets changes. Assuming a fixed integer k we write $F_x = H_k(f^{-1}(-\infty, x])$, and for $x < y$ we define $f_x^y : F_x \to F_y$ to be the map induced by inclusion of the sub-level set of x in that of y. We write $F_x^y = \text{im } f_x^y$ for the image of F_x in F_y. By convention we set $F_x^y = \{0\}$ whenever x or y is infinite. Let $\beta_x^y = \dim F_x^y$ denote the *persistent Betti number* for all $-\infty \leq x \leq y \leq +\infty$.

Let $f : X \to \mathbb{R}$ be a tame function, $(a_i)_{i=1\ldots n}$ its homological critical values, and $(b_i)_{i=0\ldots n}$ an interleaving sequence, namely $b_{i-1} < a_i < b_i$ for all i. We set $b_{-1} = a_0 = -\infty$ and $b_{n+1} = a_{n+1} = +\infty$. For two integers $0 \leq i < j \leq n+1$, we define the *multiplicity* of the pair (a_i, a_j) to be $\mu_i^j = \beta_{b_{i-1}}^{b_j} - \beta_{b_i}^{b_j} + \beta_{b_i}^{b_{j-1}} - \beta_{b_{i-1}}^{b_{j-1}}$. Denoting by $\bar{\mathbb{R}}$ the union $\mathbb{R} \cup \{-\infty, +\infty\}$ we are ready to define the persistence diagram.

DEFINITION 2.3. **[13]** *The persistence diagram $D(f) \subset \bar{\mathbb{R}}^2$ of f is the set of points (a_i, a_j), counted with multiplicity μ_i^j for $0 \leq i < j \leq n+1$, union all points on the diagonal, counted with infinite multiplicity.*

For points $p = (p_1, p_2)$ and $q = (q_1, q_2)$ in $\bar{\mathbb{R}}^2$, let $\|p-q\|_\infty$ be the maximum of $|p_1 - q_1|$ and $|p_2 - q_2|$. Similarly for functions f and g, let $\|f-g\|_\infty = \sup_x |f(x) - g(x)|$. Let X and Y be two multisets of points.

DEFINITION 2.4. **[13]** *The bottleneck distance between X and Y is*

$$d_B(X,Y) = \inf_\gamma \sup_x \|x - \gamma(x)\|_\infty \ ,$$

where $x \in X$ and $y \in Y$ range over all points and γ ranges over all bijections from X to Y. If $d_B(X,Y) < \varepsilon$ we say that Y is an *ε-approximation* of X. Cohen-Steiner *et al.* prove [13] that small changes in f imply small changes under the bottleneck metric in the persistence diagram. We state here a weaker version of their main theorem that is sufficient for our needs.

THEOREM 2.5. **[13]** *Let A, A' be two subsets of $\mathbb{R}^3$ such that $d_H(A, A') \leq \varepsilon$. Let $f_A, f_{A'}$ denote the distance from A, A' respectively. The persistence diagrams of $f_A, f_{A'}$ satisfy*

$$d_B(D(f_A), D(f_{A'})) \leq \varepsilon \ .$$

3. CONSTRUCTING THE PATHWAY DIAGRAM

In this section we define and explain what is the pathway diagram of a collection of points in $\mathbb{R}^3$ and give a formal description of our algorithm. The algorithm is fairly simple and it proceeds in two steps. First, we construct a collection $\mathcal{K}_\varepsilon$ of unit balls such that $\cup\mathcal{K}_\varepsilon$ constitutes an ε-approximation of $\cup\mathcal{B}$. In a second step we construct the pathway diagram of the centers of $\mathcal{K}_\varepsilon$, which we denote by $\mathcal{P}_\varepsilon$. We defer technical implementation details to Section 10. Properties of $\mathcal{K}_\varepsilon$ and the pathway diagram are presented and proved in the following sections.

3.1 Pathway Diagram

Let E be a finite point set in $\mathbb{R}^3$ and let σ be a Delaunay d-simplex of $\mathcal{D}[E]$, spanned by the $d+1$ point set T. We

say that a ball is *empty* in E if it does not contain any points of E in its interior. Let R_T denote the radius of the smallest empty ball that contains all points of T on its boundary surface. We say that the simplex σ is α-*exposed* if $\alpha > R_T$ [20]. The collection of α-exposed simplices is a simplicial complex, which is called the α-*complex* of E. We call the collection of the dual Voronoi faces of simplices that are *not* in the α-complex the α-*Voronoi graph* of E.

The $(\alpha = 1)$-Voronoi graph of E plays an important role in this paper, and we refer to it as the *pathway diagram* of E. Denoting by $\mathcal{K}(E)$ the collection of unit balls centered at points of E, we can define the pathway diagram of E in a more intuitive manner. It is the set of Voronoi faces in $\mathcal{V}[E]$ that do not intersect $\cup\mathcal{K}(E)$. It is a subset of the medial axis of the complement of $\cup\mathcal{K}(E)$ and it contains only flat facets, i.e., patches of planes bounded by simple polygons. Actually, the only difference between the pathway diagram of E and the whole medial axis of the complement of $\cup\mathcal{K}(E)$ is that the medial axis also contains planar patches bounded by arcs whenever the medial axis reaches the boundary surface of $\cup\mathcal{K}(E)$. The pathway diagram is thus defined such that it is completely piecewise linear and easy to compute, avoiding the need to construct facets that are bounded by arcs.

DEFINITION 3.1. **(ε-pathway diagram)** *Let X be a closed bounded subset of $\mathbb{R}^3$ and let E be a finite point set. If $\cup\mathcal{K}(E)$ is an ε-approximation of X we call the pathway diagram of E an* ε-pathway diagram *of X.*

3.2 Ball ε-Sample and the ε-Flower

Let X be as above, a closed bounded subset of $\mathbb{R}^3$. We say that a finite point set E is a *point sample* of X if E is contained in ∂X. The set E is a *point ε-sample* of X if it is a point sample of X and $d_H(\partial X|E) \leq \varepsilon$. We extend the ε-sample concept from points to balls. We call a ball $B(x,r)$ a *sample ball* of X if it is contained in X and the distance of its center to the boundary of X is equal to its radius, namely $d(x,\partial X) = r$. A set K of balls is a *ball sample* of X if all balls in K are sample balls.

DEFINITION 3.2. **(ball ε-sample)** *Given a set K of closed balls, a body $X \subset \mathbb{R}^3$ and a real parameter $\varepsilon > 0$ we say that K is a ball ε-sample of X if K is a ball sample of X and $d_H(\partial X|\bigcup K) \leq \varepsilon$.*

Note that if the balls in K have radius 0, then the definition of the ball sample coincides with the definition of the point set sample. Next we define an ε-flower.

DEFINITION 3.3. **(ε-flower)** *Let $B = B(x,r)$ be a ball and $\varepsilon \geq 0$ be real a parameter. A set K of closed* unit *balls are called an ε-flower of B if they constitute a ball ε-sample of B.*

Note that the sample balls which constitute an ε-flower of a ball $B = B(x,r)$ are tangent to B from the inside and are centered on the boundary surface of $B(x, r-1)$. See Figure 3 for examples of ε-flowers.

3.3 The Algorithm

The purpose of the algorithm is to construct an ε-pathway diagram of $\mathcal{B}$, and an ε-approximation of $\cup\mathcal{B}$ (note that we approximate all of $\cup\mathcal{B}$ and not only its boundary surface). The input is a collection $\mathcal{B}$ of three-dimensional balls such that every ball in $\mathcal{B}$ is not smaller than a unit ball, and a real parameter $\varepsilon < 1/2$. The algorithm constructs a collection E_ε of points such that the collection $\mathcal{K}_\varepsilon$ of unit balls centered at the points of E_ε constitute an ε-approximation of $\cup\mathcal{B}$. The output of the algorithm is both $\mathcal{K}_\varepsilon$ and the pathway diagram $\mathcal{P}_\varepsilon$ of E_ε, which is an ε-pathway diagram of $\cup\mathcal{B}$ by definition.

Input:
A collection $\mathcal{B}$ of balls each not smaller than a unit ball.
Output:
(1) A collection $\mathcal{K}_\varepsilon$ of unit balls such that $\cup\mathcal{K}_\varepsilon$ is an ε-approximation of $\cup\mathcal{B}$.
(2) An ε-pathway diagram of $\cup\mathcal{B}$.

$E \Leftarrow \emptyset$, $\mathcal{P}_\varepsilon \Leftarrow \emptyset$
for all $B = B(x,r) \in \mathcal{B}$ **do**
  $r' \Leftarrow r$
  while $r' > 0$ **do**
    $E \Leftarrow E \cup \text{FLOWER}(B(x,r'),\varepsilon)$
    $r' \Leftarrow r' - 1$
  end while
end for
$\mathcal{K}_\varepsilon \Leftarrow \mathcal{K}(E)$
$\mathcal{D}[E] \Leftarrow$ Delaunay triangulation of E
for all $\sigma \in \mathcal{D}[E]$ **do**
  if σ is not 1-exposed **then**
    $\mathcal{P}_\varepsilon \Leftarrow \mathcal{P}_\varepsilon \cup \text{DUAL}(\sigma)$
  end if
end for
return $\mathcal{K}_\varepsilon$, $\mathcal{P}_\varepsilon$

Figure 4: Pseudocode of `PATHWAY_DIAGRAM`, the algorithm that constructs the pathway diagram.

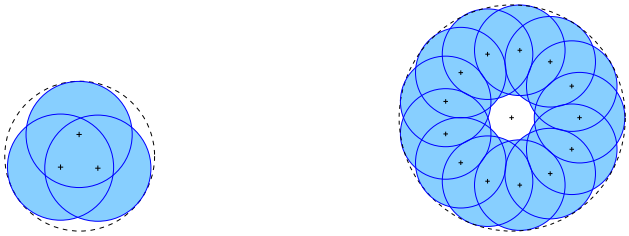
In the pseudocode of `PATHWAY_DIAGRAM` (see Figure 4) the procedure DUAL(σ) returns the dual Voronoi face of a simplex σ, and FLOWER(B,ε) is a procedure that returns the centers of the unit balls in an ε-flower of a ball B. In Section 10 we describe our implementation of the procedure FLOWER(B,ε), and explain how we construct, with little effort, the pathway diagram of a point set using the CGAL library [1]. Note that although a single ε-flower imposes a *one-sided* Hausdorff distance the combination of multiple ε-flowers per each ball of $\mathcal{B}$ ensures that the *symmetric* Hausdorff distance between $\cup\mathcal{B}$ and $\cup\mathcal{K}_\varepsilon$ is bounded by ε.

4. THE RELATION BETWEEN THE PATHWAY DIAGRAM AND THE PATHWAY AXIS

In this section we establish the relation between the pathway diagram and the medial axis. Namely, we show that the pathway diagram contains an approximation of a subset of the medial axis called the pathway axis.

We limit ourselves from now on to a large open ball $Q = B(c_q, r_q)$ that contains $\cup\mathcal{B}$. For any closed bounded subset $X \subseteq Q$ we define $\mathcal{O}(X)$ to be $Q \setminus X$, which is an open subset of Q. For each point $x \in M[\mathcal{O}(\cup\mathcal{B})]$ we define $I_\mathcal{B}(x)$ to be the set of balls in $\mathcal{B}$ closest to x:

$$I_\mathcal{B}(x) = \{B \in \mathcal{B} : \Gamma_{\mathcal{O}(\cup\mathcal{B})}(x) \cap B \neq \emptyset\} .$$



(a) (b)

Figure 3: (a) An ε-flower (in the plane) without a void inside. (b) An ε-flower with a void inside. In both cases, the dashed line bounds the disc that is being approximated.

We define the *pathway axis* of $\mathcal{B}$, denoted by $PA[\mathcal{B}]$, to be the subset of $M[\mathcal{O}(\cup\mathcal{B})]$ for which the balls of $I_{\mathcal{B}}(x)$ do not share a common point:

$$PA[\mathcal{B}] = \{x \in M[\mathcal{O}(\cup\mathcal{B})] : \bigcap_{B \in I_{\mathcal{B}}(x)} B = \emptyset\} .$$

We consider the pathway axis sufficient for finding pathways since it is a subset of the medial axis of $\mathcal{O}(\cup\mathcal{B})$ without 'dead ends'. See Figure 1 for a two-dimensional illustration.

The following theorem states that the pathway diagram that our algorithm constructs contains an approximation of the pathway axis of $\mathcal{B}$. A collection of balls is said to be in *general position* if no degeneracies occur, namely the common intersection of the boundary spheres of any two, three or four balls in the collection is not a single point.

THEOREM 4.1. *Let $\mathcal{B}$ be a finite collection of balls in general position that are each not smaller than a unit ball. For any $\delta > 0$ there exists an $\varepsilon > 0$ such that the pathway diagram $\mathcal{P}_\varepsilon$, which the algorithm* `PATHWAY_DIAGRAM` *constructs, satisfies $d_H(PA[\mathcal{B}] \mid \mathcal{P}_\varepsilon) < \delta$.*

The proof of the theorem proceeds by using the λ-medial axis [11]. Specifically, we use the λ-medial axis as a mediator, i.e., we show that the pathway diagram contains an approximation of the $(\lambda = 1)$-medial axis of the union of balls in $\mathcal{B}$ shrunk by one, and that the $(\lambda = 1)$-medial axis contains the pathway axis. The (rather involved) proof is omitted for lack of space; see [28].

5. BOUND ON THE NUMBER OF UNIT BALLS

In this section we address the complexity of our approximation scheme, focusing on the number of unit balls in $\mathcal{K}_\varepsilon$. We state an upper bound on the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$. In Section 7 we employ this bound to show what is gained from using ball samples over using point samples, the latter being the standard practice.

Based on the definitions of a ball sample and a point sample, presented in Section 3.2, we define the *approximation quality* of a sample. Let X be a closed bounded subset of $\mathbb{R}^3$, let E be a finite point sample of X, and let K be ball sample of X. We call the one-sided Hausdorff distance $d_H(\partial X|E)$ the *approximation quality* of E. In a similar fashion, we call the one-sided Hausdorff distance $d_H(\partial X| \cup K)$ the *approximation quality* of K. Given an integer $\kappa > 0$, and denoting by μ the approximation quality of E, we say that E is *κ-light* if the number of sample points in any ball of radius μ is not greater than κ, namely $\forall x' \in \mathbb{R}^3, |B(x', \mu) \cap E| \leq \kappa$.

The following theorem gives an upper bound on the number of unit balls needed to construct an ε-flower of a single ball $B \in \mathcal{B}$.

THEOREM 5.1. *Let $B(x, r)$ be a ball with $r \geq 1$, let E be a finite point sample of $B(x, r-1)$. Let $K = K(E)$ be the collection of unit balls centered at E, which are a ball sample of $B(x, r)$. Denoting by ε the approximation quality of K, if $\varepsilon < 1/2$ and E is κ-light then:*

$$|E| \leq \kappa \frac{16r^2}{3\varepsilon^{4/3}} .$$

Next we state two corollaries of the theorem, which give bounds on the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$.

COROLLARY 5.2. *Let $\rho > 0$. If for every ball $B(x, r) \in \mathcal{B}$ it holds that $r \leq \rho$, and the sampling procedure $FLOWER(B, \varepsilon)$ produces a κ-light sampling for an integer κ then:*

$$|\mathcal{K}_\varepsilon|/|\mathcal{B}| \leq \kappa \frac{16\lceil\rho\rceil^3}{3\varepsilon^{4/3}} .$$

COROLLARY 5.3. *Assuming that the parameters ρ, κ in Corollary 5.2 are constants, if for every ball $B(x, r) \in \mathcal{B}$ it holds that $1 \leq r \leq \rho$, then the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$ is $O(1/\varepsilon^{4/3})$.*

If we were interested only in asymptotic bounds, we could have required (without loss of generality) that all radii of balls in $\mathcal{B}$ are at least 2 (while approximating with unit balls). In such a setting the asymptotic value of the ratio $|\mathcal{K}_\varepsilon|/|\mathcal{B}|$ is actually $O(1/\varepsilon)$. Yet for practical reasons it is useful to allow the balls in $\mathcal{B}$ to have a radius 1 or very close to 1, which results in much fewer sample unit balls. This requires a more refined analysis which results in the looser bound above.

In order to prove Theorem 5.1 we first prove several auxiliary claims. The following lemma gives upper and lower bounds on the number of points in a μ-sample which is κ-light.

LEMMA 5.4. *Let E be a μ-sample of a ball $B = B(x, r)$. If E is κ-light and $\mu < \min(r, 1/2)$, then $16r^2/3\mu^2 \leq |E| \leq \kappa(16r^2/3\mu^2)$.*

The proof is omitted here (see [28]). In the next lemma we establish a relation between the approximation quality of a point sample E of $B(x, r-1)$ and the approximation quality of the ball sample $K(E)$ of $B(x, r)$:

LEMMA 5.5. *Let $B = B(x,r)$ be a ball such that $r \geq 1$, and let E be a point sample of $B(x,r-1)$. Let ε denote $d_H(\partial B(x,r)|\cup K(E))$, and μ denote $d_H(\partial B(x,r-1)|E)$. If $\varepsilon < \min(r-1, 1/2)$ then:*

$$\mu^2 \geq 2\varepsilon\frac{r-1}{r} .$$

PROOF. Let $p \in \partial B$ be a point at distance $\varepsilon = d_H(\partial B, \cup K)$ from $\cup K$. We know p exists since B and $\cup K$ are compact and ε is the approximation quality. Let $k = B(c,1)$ be a ball in K that is ε distant from p. We denote by f the radial projection of p onto the ball $B(x,r-1)$; see Figure 5 for an illustration. We note that $c \in E$ is the closest point of E to f, since k is the closest ball to p. Therefore $\mu \geq d(f,c)$ giving an implicit bound on μ which we will work out below.

We use the following notation. The intersection of the segment $\overline{pc}$ with the boundary of k is marked with d, the closest point to c on the segment $\overline{px}$ is marked with g, and the midpoint of $\overline{cf}$ is marked with m. We denote the length of $\overline{cf}$ by $\bar{\mu}$. Since $\mu \geq d(f,c) = \bar{\mu}$ it suffice to prove the bound for $\bar{\mu}$. First we note that $|\overline{cx}| = r-1$, $|\overline{px}| = r$, $|\overline{pd}| = \varepsilon$, $|\overline{xf}| = r-1$, all from the definition. From triangle similarity we know that $|\overline{fg}|/|\overline{cf}| = |\overline{mf}|/|\overline{xf}|$ or in other words $|\overline{fg}| = \frac{(\bar{\mu})^2}{2(r-1)}$. Considering the triangle $\triangle cgp$ and the equality $|\overline{pg}| = |\overline{fg}| + 1$ we can express ε as a function of $\bar{\mu}$:

$$\begin{aligned}
\varepsilon &= (|\overline{pg}|^2 + |\overline{cg}|^2)^{1/2} - 1 \\
&= ((|\overline{fg}| + 1)^2 + |\overline{cg}|^2)^{1/2} - 1 \\
&= (|\overline{cf}|^2 + 2|\overline{fg}| + 1)^{1/2} - 1 \\
&= (\bar{\mu}^2 + 1 + \frac{(\bar{\mu})^2}{r-1})^{1/2} - 1 .
\end{aligned}$$

We can now work towards expressing $\bar{\mu}$ as a function of ε:

$$\begin{aligned}
(\varepsilon+1)^2 &= \bar{\mu}^2 + 1 + \frac{(\bar{\mu})^2}{r-1} \\
\varepsilon^2 + 2\varepsilon &= (\bar{\mu})^2\frac{r}{r-1} .
\end{aligned}$$

Reorganizing the terms we obtain

$$\bar{\mu}^2 = (\varepsilon^2 + 2\varepsilon)\frac{r-1}{r} \geq 2\varepsilon\frac{r-1}{r} .$$

Recall that $\mu \geq d(f,c)$ or in other words $\mu \geq \bar{\mu}$. The bound asserted in the lemma follows.

□

We can finally prove Theorem 5.1, harnessing the lemmas above.

Proof (Theorem 5.1):

Let E be a sample as defined in the theorem and let μ be its approximation quality. We handle three cases, according to the approximation quality ε of $K(E)$. First, the trivial case of a small ball. If $(r-1) \leq \varepsilon$ then E obviously contains a constant number of points. In the two other cases we can use Lemma 5.5, which states that $\mu^2 \geq 2\varepsilon\frac{r-1}{r}$. We also use Lemma 5.4 which states that $|E| \leq \kappa\frac{16(r-1)^2}{3\mu^2}$. In case $\varepsilon < (r-1) \leq \varepsilon^{\frac{1}{3}}$ we get:

$$\mu^2 \geq 2\varepsilon\frac{r-1}{r} \geq 2\varepsilon - \frac{2\varepsilon}{\varepsilon+1} = \frac{2\varepsilon^2}{\varepsilon+1} \geq \varepsilon^2 .$$

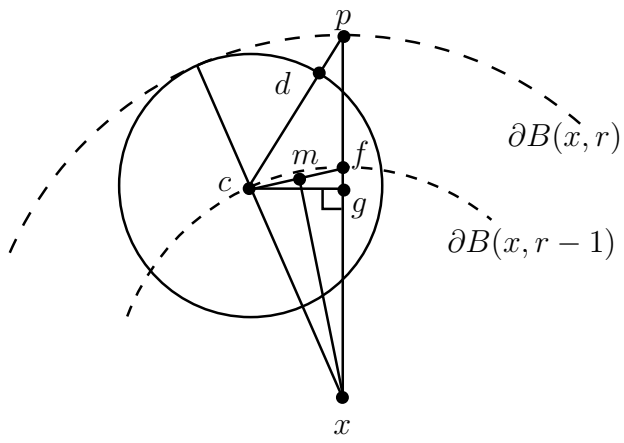


Figure 5: Geometric relation between μ and ε. The length of $\overline{dp}$ is ε and the length of $\overline{cf}$ bounds μ.

We use this bound in conjunction with Lemma 5.4 and get a bound on $|E|$:

$$|E| \leq \kappa\frac{16(r-1)^2}{3\mu^2} \leq \kappa\frac{16\varepsilon^{\frac{2}{3}}}{3\varepsilon^2} \leq \kappa\frac{16}{3\varepsilon^{\frac{4}{3}}} \leq \kappa\frac{16r^2}{3\varepsilon^{\frac{4}{3}}} .$$

In the remaining case, where $\varepsilon^{\frac{1}{3}} < (r-1)$ we get:

$$\begin{aligned}
\mu^2 &\geq 2\varepsilon\tfrac{r-1}{r} \geq 2\varepsilon - \tfrac{2\varepsilon}{r} \geq 2\varepsilon - \tfrac{2\varepsilon}{\varepsilon^{\frac{1}{3}}+1} = \varepsilon(2 - \tfrac{2}{\varepsilon^{\frac{1}{3}}+1}) \\
\mu^2 &\geq 2\tfrac{\varepsilon^{\frac{4}{3}}}{\varepsilon^{\frac{1}{3}}+1} \geq \varepsilon^{\frac{4}{3}} .
\end{aligned}$$

By using the above inequality together with the bound of Lemma 5.4 we get:

$$|E| \leq \kappa\frac{16(r-1)^2}{3\mu^2} \leq \kappa\frac{16r^2}{3\varepsilon^{\frac{4}{3}}} .$$

□

6. APPROXIMATING THE PERSISTENCE DIAGRAM

The collection $\mathcal{K}_\varepsilon$ of unit balls can be used to construct an efficient approximation of the persistence diagram of the distance function from $\cup\mathcal{B}$, which we call the persistence diagram of $\cup\mathcal{B}$ for short. The set of unit balls $\mathcal{K}_\varepsilon$, which our algorithm constructs for a given set $\mathcal{B}$ of balls, satisfies $d_H(\cup\mathcal{B}, \cup\mathcal{K}_\varepsilon) \leq \varepsilon$. According to Theorem 2.5, the persistence diagram of $\cup\mathcal{K}_\varepsilon$ is therefore an ε-approximation of the persistence diagram of $\cup\mathcal{B}$. This implies that 'major' topological features, i.e., large voids or tunnels are the same for both unions. The following theorem asserts the efficiency of our approximation, and its proof follows easily from Corollary 5.3.

THEOREM 6.1. *The persistence diagram of the distance function from $\cup\mathcal{B}$ can be ε-approximated using $O(|\mathcal{B}|/\varepsilon^{4/3})$ unit balls.*

We compute and make use of the ($k = 2$) persistence diagram of $\cup\mathcal{K}_\varepsilon$ to identify the center of the largest chamber within a molecule; see Section 8.

7. COMPARISON WITH POINT SAMPLING TECHNIQUES

In this section we compare our ball-sample approach with a standard point-sample approach (see, e.g., [22, 26]) for the two problems that we address in this paper: approximating the medial axis of the complement of $\cup\mathcal{B}$ and approximating the persistence diagram of $\cup\mathcal{B}$. To solve both problems with point samples, the boundary surface of each of the balls in $\mathcal{B}$ is sampled with a collection of points S_ε, such that S_ε contains an ε-sample of the boundary surface of $\cup\mathcal{B}$. As described in [4], the λ-Voronoi graph of S_ε for $\lambda > \varepsilon$ serves as an approximation of the medial axis, which tends to the medial axis as λ tends to 0. The persistence diagram of the distance function from S_ε can serve as an approximation of the persistence diagram of $\cup\mathcal{B}$.

Let us focus on a single ball $B \in \mathcal{B}$. The advantage of our approach is the relatively small number of unit balls needed to construct an ε-flower of B, compared to the number of points needed to construct a point ε-sample of B. The number N_{ball} of unit balls needed by `PATHWAY_DIAGRAM` to constitute an ε-flower B is not more than $\kappa(16r^2/3\varepsilon^{4/3})$, according to Theorem 5.1. On the other hand, the number N_{point} of points needed to constitute a point ε-sample of B is at least $16r^2/3\varepsilon^2$, as we show in Lemma 5.3 in [28]. Therefore, by regarding κ as a constant, we get that the ratio $N_{\text{point}}/N_{\text{ball}}$ is $\Omega(1/\varepsilon^{2/3})$.

8. EXPERIMENTAL RESULTS

In this section we present results that demonstrate the performance of the algorithm. (We describe ample biological results in the MolAxis paper [29].) All tests were carried out on a Pentium IV 3.0 GHz machine with 1GB of RAM running a LINUX native operating system. We report on a set of tests that were performed on a collection of 3251 balls, which model a P450 Enzyme isozyme (H atoms discarded). The ratio between the largest and smallest input balls is $\rho = 1.21$. In Table 1 we report on a set of tests performed on these input balls with varying resolution. In Figure 6 we provide (a) a graph of the size of $\mathcal{K}_\varepsilon$ as a function of ε, comparing the experimental results to the theoretical bound proved in Theorem 5.1 and (b) a graph of the number of Voronoi vertices in the pathway diagram as a function of the number of balls in $\mathcal{K}_\varepsilon$. Note that as ε becomes smaller and the number of unit balls increases, the ratio between the number of Voronoi vertices and number of unit balls tends to one. We believe that the explanation of this phenomenon is that our sampled points behave like points sampled on a smooth surface [5] or points sampled on a polyhedral surface [3]. In these two cases the number of faces in the Voronoi diagrams is less than the worst case $\Theta(n^2)$ faces, and is $O(n \log n)$ and $O(n)$ respectively. In Table 2 we provide the runtime breakdown of three runs, for $\varepsilon = 0.4$, $\varepsilon = 0.1$, $\varepsilon = 0.01$.

Pocket Persistence Diagram

We call the persistence diagram of $\cup\mathcal{B}$ for $k = 2$ the *pocket diagram* of $\mathcal{B}$. To give some intuition as to what the pocket diagram is imagine a continuous uniform growth process of the balls in $\mathcal{B}$. During the growth process the complement of the expanding $\cup\mathcal{B}$ is divided into connected components, with one unbounded component and zero or more *voids*. Voids are transient in the sense that they may appear and disappear during the growth process. Each void has a *birth time* and a *death time*. We call these temporal voids, which existed during the growth process, the *pockets* (or *chambers*) of $\mathcal{B}$, similar to the definition by Edelsbrunner *et al.* [18], yet with a more natural geometric meaning.[1] The *lifetime* of a pocket is defined to be its death time minus its birth time. We extract the lifetime of pockets from the pocket diagram that we have computed. The *pocket center* is the location of the Voronoi vertex dual to the last tetrahedron associated with the void during the growth process. See Figure 7 for an example of the pocket diagram of a collection of balls that model a large molecule (the P450 Enzyme) that has one large inner chamber. An interesting direction for further investigation can be to compare the persistence homology under the Euclidean distance function which we compute, with the persistent homology under the power distance function which is presented in [19].

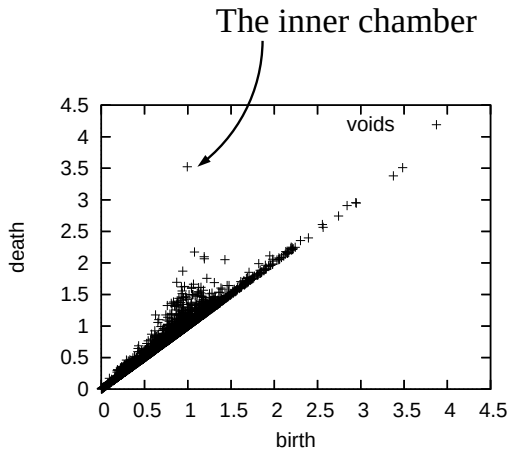


Figure 7: The pocket diagram of a collection of balls that model the P450 Enzyme with one large inner chamber. Notice that all pockets except the inner chamber are close to the $x = y$ diagonal and can be seen as topological noise.

Constructing the geometric location of persistent homology classes that represent wide tunnels from the $k = 1$ persistence diagram is a non-trivial matter. We believe this problem can be addressed using techniques as the one introduced by Freedman *et al.* (see, e.g., [21]).

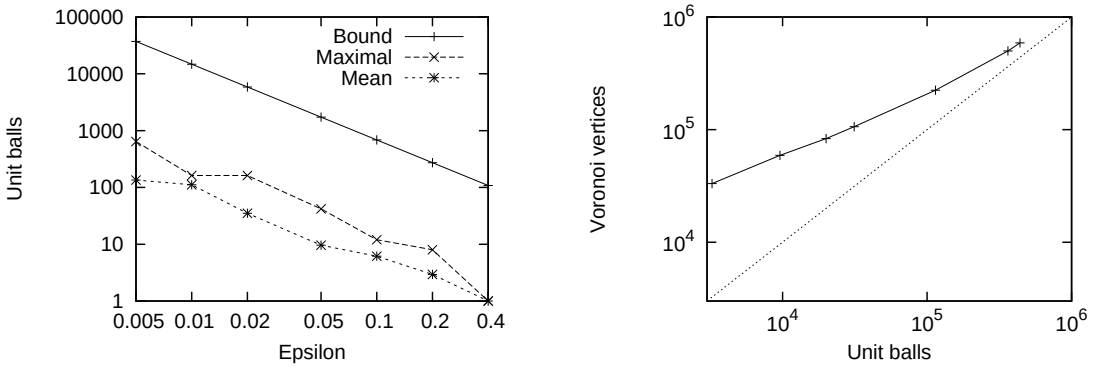
9. MOLAXIS

MolAxis is a tool designed to assist the biologist or the biochemist in the identification of molecular channels. In a companion paper [29] we introduce MolAxis, and present analysis of channels in molecules achieved with MolAxis. A major advantage of our approach is that since the medial axis is composed of two-dimensional surface patches it reduces the dimension of the problem, i.e., it transforms a three-dimensional problem to a two-dimensional problem. This dimension reduction, combined with the novel sampling technique introduced in this paper, leads to a highly efficient algorithm. We compare MolAxis to state-of-the art tools for channel identification, showing that MolAxis is ex-

[1] Edelsbrunner *et al.* [18] define pockets using the *power distance*, which means that in their growth process the balls are expanded in a non-uniform manner, i.e., larger balls grow faster.

Resolution	Max BPS	Unit Balls	Runtime(sec's)	File size(Mb)
0.4	1	3251	2.456	2.6
0.2	8	9585	5.120	4.5
0.1	12	19996	8.737	6.3
0.05	42	31287	12.229	7.9
0.02	162	114222	35.522	16.9
0.01	162	362262	101.830	37.7
0.005	642	439182	123.444	44.4

Table 1: Data for a set of runs with different resolutions on a single instance of the P450 enzyme. Resolution is the Hausdorff approximation quality ε. Max BPS is the maximal number of unit balls used to approximate a single ball. Unit Balls is the total number of unit balls used in the approximation (the number of balls in $\mathcal{K}_\varepsilon$). Runtime is the total runtime in seconds, including the construction of the pathway diagram.



(a) (b)

Figure 6: Graphs generated for the set of runs that are analyzed in Table 1. All scales are logarithmic. (a) Experimental and theoretical values of the ratio $|\mathcal{B}|/|\mathcal{K}_\varepsilon|$. Maximal is the maximal BPS, i.e. the maximal number of unit balls used to approximate a single ball. Mean is the mean BPS over all input balls. Bound is the theoretical upper bound as expressed in Theorem 5.1, with $\kappa = 3$. Note that for $\varepsilon > 0.2$ we need exactly one unit ball to approximate each input ball. (b) The number of Voronoi vertices in the pathway diagram as a function of the number of unit balls in $\mathcal{K}_\varepsilon$. We draw the function $y = x$ as a reference using a dotted line. Note that as the number of unit balls increases the ratio between the number of Voronoi vertices and the number of unit balls becomes closer to one.

tremely efficient and accurate. To the best of our knowledge it is the first attempt to approximate and analyze the medial axis of the complement of a molecule in order to construct channels.

We follow a standard practice in biology of modeling a molecule by a collection of three-dimensional balls, one ball per atom. The term *channel* is often used in molecular biology to refer to a probable route taken by a small molecule passing through a hole in the molecule. A *pathway* is a curve in the space that lies outside the molecule. If $\mathcal{B}$ is the set of atom balls of a molecule then the boundary surface of $\cup\mathcal{B}$ is called the *van der Waals surface* (VDW for short) of the molecule. The *clearance* $c(p)$ of a point p outside the molecule is the distance between the point and the van der Waals surface of the molecule. Given a pathway π, the *profile* of π is the clearance of the points on π as a function of the distance along the pathway. The *pathway ball* of $p \in \pi$ is the ball with radius $c(p)$ that is centered at p. The *pathway surface* of π is the boundary (envelope) surface of the union of all pathway balls of π. The *bottleneck radius* of π is the minimal clearance along the pathway, and the *bottleneck point* of π is the point in π where the bottleneck radius is achieved. The *bottleneck atoms* (resp. *bottleneck residues*) are the lining atoms (resp. lining residues) of the bottleneck point.

MolAxis is adapted for two scenarios. In the first scenario MolAxis identifies transmembrane channels while in the second scenario it identifies channels that connect an inner chamber in a molecule to the bulk solvent. We focus here on the latter scenario. MolAxis identifies the center of the chamber using the pocket diagram (defined in Section 8). MolAxis constructs *corridors* which are pathways that are contained in the pathway diagram and that balance between length and clearance (defined formally in [29]). For each corridor, MolAxis reports its profile, lining atoms/residues, bottleneck radius and bottleneck atoms/residues. The pathway and pathway surface can be visualized using VMD [23].

The underlying algorithm used by MolAxis is `PATHWAY_DIAGRAM` (Figure 4), which constructs the pathway diagram of the molecule. Yet the pathway diagram is composed of two-dimensional patches whereas corridors are 1-dimensional entities. Therefore on top of the pathway diagram we construct a graph called the *pathway graph*. By applying Dijkstra's algorithm [16] on the pathway graph we compute a tree which we call the *corridor tree* that contains the desired paths. See [29] for full details.

Cytochrome P450 proteins constitute a large family of mono-oxygenases heme containing enzymes that oxidize a variety of chemical compounds in microorganisms. The ox-

Phase	$\varepsilon = 0.4$	$\varepsilon = 0.1$	$\varepsilon = 0.01$
Construct ball sample	0.016	0.088	1.720
Compute triangulation	0.088	0.568	10.645
Compute alpha shape family	0.456	2.848	51.471

Table 2: Runtime breakdown of MolAxis on a single instance of the P450 enzyme (see the relevant rows in Table 1 for more details about these runs). The time is given in seconds. Note that the most time-consuming phase is the Alpha shape computation, and as ε decreases it becomes more dominant.

idation of a substrate occurs at the hydrophobic core of the protein. It is of great mechanistic and biochemical interest to identify and characterize all channels that link the active site to bulk solvent both statically and dynamically by means of Molecular Dynamics (MD, for short) simulations. In Figure 8 we show the channels of the human CYP3A4, a P450 isozyme with 3251 atoms, as computed using MolAxis. Being very fast, MolAxis has been successfully used to analyze channel dimensions and lining residues in hundreds of snapshots of an MD simulation of the human CYP3A4.

10. IMPLEMENTATION DETAILS

In Section 10.1 we describe how we implemented the procedure FLOWER(B, ε) which constructs an ε-flower for a single ball $B \in \mathcal{B}$. In Section 10.2 we explain how we construct the pathway diagram $\mathcal{P}_\varepsilon$ using ready-made tools. (See Section 3 for the definitions of these entities.)

10.1 ε-Flower Construction

For each ball $B = B(c, r) \in \mathcal{B}$ we construct a set of unit balls K_B that are an ε-flower of B. Constructing an ε-flower with a minimal number of unit balls is an optimization problem closely related to the following problem: "how can n points be distributed on a unit sphere such that they maximize the minimum distance between any pair of points?". Such a configuration of points is called a *spherical code* and its construction has been intensively studied [27]. We employ two heuristic sampling techniques for producing an ε-flower.

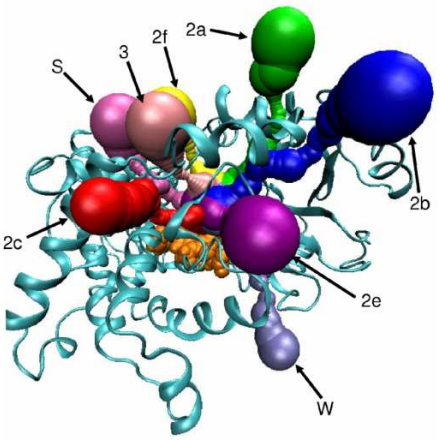


Figure 8: CYP3A4 channels as detected by MolAxis. CYP3A4 is represented by cartoons and the heme prosthetic group is represented by its VDW surface. Each channel surface has different shades of grey for the sake of clarity.

The first heuristic is **Icosahedron refinement.** An *icosahedron* is the Platonic solid P_3 having 12 vertices, 30 edges, and 20 congruent equilateral triangular faces. We denote by $I_0(c, r)$ an icosahedron that has its vertices on the sphere $S = S(c, r - 1)$. $I_0(c, r)$ can be refined by adding a vertex in the midpoint of each polyhedron edge and centrally projecting it onto the sphere S. Three new polyhedron edges are added within each triangle, that connect the three new vertices on its boundary edges. Each triangle of $I_0(c, r)$ is split into four new smaller triangles. In this way we get a *refined icosahedron* of degree one, denoted by $I_1(c, r)$, which comprises 42 vertices, 80 triangles and 120 edges. We iterate the refinement and define a sequence of refined icosahedra: $I_0(c, r), I_1(c, r), I_2(c, r)$, etcetera. For any natural number η we define $\varepsilon_{\text{ico}}(\eta)$ to be the one-sided Hausdorff distance between the boundary surface of B and the union of the collection of unit balls centered at the vertices of $I_\eta(c, r)$. We compute $\varepsilon_{\text{ico}}(\eta)$ by scanning all triangles of $I_\eta(c, r)$. Given the user-specified parameter $\varepsilon > 0$ we find the smallest η such that $\varepsilon_{\text{ico}}(\eta) \leq \varepsilon$, and return the vertices of $I_\eta(c, r)$.

The second heuristic that we employ is **Random points.** The icosahedron refinement technique has a major drawback: The number of vertices in each icosahedron in the sequence jumps in large steps, i.e., $12, 42, \ldots$ Since the first icosahedron has 12 vertices, we have a 'gap' between 2 and 11 that we wish to bridge. We use a naive random sampling technique to generate preprocessed samples as described next.

We repeat the following procedure for each $i = 2, \ldots 11$. Using a typically large integer constant N_{rnd}, we generate $j = 1, \ldots, N_{\text{rnd}}$ random point sets E_{ij}, each containing exactly i points, such that the points of E_{ij} are located on the sphere $S(c, r - 1)$. Recall that for any finite point collection $E \subset \mathbb{R}^3$ we denote by $K(E)$ the collection of unit balls centered at E. We choose the set $E_{i\tilde{j}}$, $1 < \tilde{j} \leq N_{\text{rnd}}$, such that for any $1 < j \leq N_{\text{rnd}}$ it holds that $d_H(\partial B| \cup K(E_{i\tilde{j}})) \leq d_H(\partial B| \cup K(E_{ij}))$.

We denote $E_{i\tilde{j}}$ by E_i and denote the one-sided Hausdorff distance $d_H(\partial B| \cup K(E_{i\tilde{j}}))$ by $\varepsilon_{\text{rnd}}(i)$. After completing the procedure for all $2 \leq i \leq 11$ we have 10 computed point sets $\{E_i\}_{i=2,\ldots 11}$ with their respective one-sided Hausdorff distances $\{\varepsilon_{\text{rnd}}(i)\}_{i=2,\ldots 11}$. Now, given the user-specified parameter ε, we choose the minimal $2 \leq i \leq 11$ such that $\varepsilon_{\text{rnd}}(i) \leq \varepsilon$. If i exists we use E_i as the centers of the ε-flower. If no i satisfies $\varepsilon_{\text{rnd}}(i) \leq \varepsilon$ we use icosahedron refinement.

10.2 Computing the Pathway Diagram

We compute the $(\alpha = 1)$-Voronoi graph of E_ε using the 3D Alpha Shapes package [14] of the CGAL library. We use exact arithmetic to ensure that the algorithm is robust. The Alpha Shapes data structure allows to retrieve the α-complex for any α value. We set the pathway diagram $\mathcal{P}_\varepsilon$ to be the collection of dual Voronoi faces of the simplices that are *not* in the $(\alpha = 1)$-complex of E_ε.

Acknowledgement

The authors thank David Cohen-Steiner for helpful comments on an earlier draft.

11. REFERENCES

[1] The CGAL Manaul, Release 3.2.1, 2006, http://www.cgal.org.

[2] N. Amenta, S. Choi, and R. Kolluri. The power crust, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2–3):127–153, 2001.

[3] D. Attali and J.-D. Boissonnat. A linear bound on the complexity of the Delaunay triangulation of points on polyhedral surfaces. *Discrete & Computational Geometry*, 31(3):369–384, 2004.

[4] D. Attali, J.-D. Boissonnat, and H. Edelsbrunner. Stability and computation of medial axes: A state of the art report. In B. H. T. Möller and B. Russell, editors, *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*. Springer-Verlag, Mathematics and Visualization, 2007.

[5] D. Attali, J.-D. Boissonnat, and A. Lieutier. Complexity of the Delaunay triangulation of points on surfaces: The smooth case. In *Proceedings of the Symposium on Computational Geometry*, pages 201–210, 2003.

[6] F. Aurenhammer and R. Klein. Voronoi diagrams. In J.-R. Sack and J. Urrutia, editors, *Handbook of Computational Geometry*, pages 201–290. Elsevier Science Publishers B.V. North-Holland, Amsterdam, 2000.

[7] J.-D. Boissonnat and C. Delage. Convex hull and Voronoi diagram of additively weighted points. In *Proceedings of the European Symposium on Algorithms*, pages 367–378, 2005.

[8] J.-D. Boissonnat and M. Yvinec. *Algorithmic Geometry*. Cambridge University Press, UK, 1998. Translated from the French version by H. Brönnimann.

[9] G. Carlsson and A. Zomorodian. The theory of multidimensional persistence. In *Proceedings of the Symposium on Computational Geometry*, pages 184–193, 2007.

[10] F. Chazal and A. Lieutier. Stability and homotopy of a subset of the medial axis. In *Proceedings of the Symposium on Solid Modeling and Applications*, pages 243–248, 2004.

[11] F. Chazal and A. Lieutier. The "Lambda-medial axis". *Graphical Models*, 67(4):304–331, 2005.

[12] F. Chazal and A. Lieutier. Weak feature size and persistent homology: Computing homology of solids in $\mathbb{R}^n$ from noisy data samples. In *Proceedings of the Symposium on Computational Geometry*, pages 255–262, 2005.

[13] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. In *Proceedings of the Symposium on Computational Geometry*, pages 263–271, 2005.

[14] T. K. F. Da and M. Yvinec. 3D Alpha Shapes. In CGAL Editorial Board, editor, CGAL- *3.2 User and Reference Manual*. 2006. http://www.cgal.org/Manual/3.2/doc_html/cgal_manual/Alpha_shapes_3/Chapter_main.html.

[15] T. K. Dey and W. Zhao. Approximate medial axis as a Voronoi subcomplex. In *Proceedings of the Symposium on Solid Modeling and Applications*, pages 356 – 366, 2002.

[16] E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959.

[17] H. Edelsbrunner. *Geometry and topology for mesh generation*. Cambridge University Press, New York, NY, USA, 2001.

[18] H. Edelsbrunner, M. A. Facello, and J. Liang. On the definition and the construction of pockets in macromolecules. *Discrete Applied Mathematics*, 88(1-3):83–102, 1998.

[19] H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete & Computational Geometry*, 28(4):511–533, 2002.

[20] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, 1994.

[21] D. Freedman and C. Chen. Measuring and localing homology classes. *The Computing Research Repository (CoRR)*, abs/0705.3061, 2007. http://arxiv.org/abs/0705.3061.

[22] J. Giesen, E. A. Ramos, and B. Sadri. Medial axis approximation and unstable flow complex. In *Proceedings of the Symposium on Computational Geometry*, pages 327–336, 2006.

[23] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual Molecular Dynamics. *J. Mol Graphics*, 14.1:33–38, 1996.

[24] A. Lieutier. Any open bounded subset of $\mathbb{R}^n$ has the same homotopy type than its medial axis. In *Proceedings of the Symposium on Solid Modeling and Applications*, pages 65–75, 2003.

[25] J. Munkres. *Elements of Algebraic Topology*. Addison-Wesley, 1984.

[26] S. Oudot and J.-D. Boissonnat. Provably good surface sampling and approximation. In *Proceedings of the Symposium on Geometry Processing*, pages 9–19, 2003.

[27] E. W. Weisstein. Spherical code, from mathworld — a wolfram web resource, http://mathworld.wolfram.com/sphericalcode.html, 2000.

[28] E. Yaffe. Efficient construction of pathways in the complement of the union of balls in $\mathbb{R}^3$. Master's thesis, Tel-Aviv University, September 2007. http://www.cs.tau.ac.il/~eitanyaf/thesis.pdf.

[29] E. Yaffe, D. Fishelovitch, H. J. Wolfson, R. Nussinov, and D. Halperin. MolAxis: Efficient and accurate identification of channels in macromolecules. *To appear in PROTEINS : Structure, Function, and Bioinformatics*.