

# Revisiting the Voronoi Description of Protein-Protein Interfaces: Algorithms

Frédéric Cazals

► **To cite this version:**

Frédéric Cazals. Revisiting the Voronoi Description of Protein-Protein Interfaces: Algorithms. PRIB - 5th IAPR International Conference on Pattern Recognition in Bioinformatics - 2010, Sep 2010, Nijmegen, Netherlands. pp.419-430, 10.1007/978-3-642-16001-1\_36 . hal-00796062

**HAL Id: hal-00796062**

**<https://hal.inria.fr/hal-00796062>**

Submitted on 1 Mar 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Revisiting the Voronoi description of Protein-Protein interfaces: Algorithms

F. Cazals<sup>1</sup>

INRIA Sophia-Antipolis, France

Frederic.Cazals@sophia.inria.fr,

WWW home page: <http://http://www-sop.inria.fr/abs/Frederic.Cazals>

**Abstract.** Describing macro-molecular interfaces is key to improve our understanding of the specificity and of the stability of macro-molecular interactions, and also to predict complexes when little structural information is known. Ideally, an interface model should provide easy-to-compute geometric and topological parameters exhibiting a good correlation with important bio-physical quantities. It should also be parametric and amenable to comparisons. In this spirit, we recently developed an interface model based on Voronoi diagrams, which proved instrumental to refine state-of-the-art conclusions and provide new insights.

This paper formally presents this Voronoi interface model. First, we discuss its connexion to classical interface models based on distance cut-offs and solvent accessibility. Second, we develop the geometric and topological constructions underlying the Voronoi interface, and design efficient algorithms based on the Delaunay triangulation and the  $\alpha$ -complex.

We conclude with perspectives. In particular, we expect the Voronoi interface model to be particularly well suited for the problem of comparing interfaces in the context of large-scale structural studies.

**Keywords:** Protein interfaces, Computational Geometry, Voronoi diagrams, Geometric patterns

## 1 Introduction

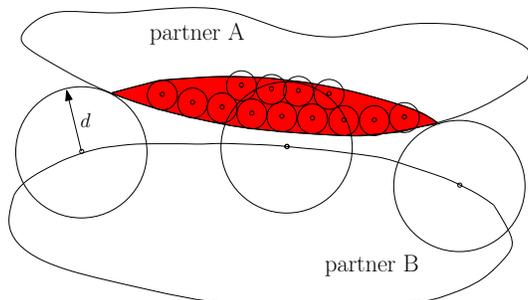
### 1.1 On classical protein - protein interface models

**Modeling interfaces.** Understanding the stability and the specificity of macro-molecular interactions is a key endeavour in computational structural biology. Such an endeavour requires on the one hand describing non-covalent interactions for the interfaces of complexes which have been solved experimentally, and on the other hand developing algorithms able to predict complexes when little or no structural information on the partners is known. On a per-complex basis, interface models allow one to investigate correlations between structural parameters and key bio-physical properties such as the conservation of residues, their polarity, the water dynamics at the interface, mutagenesis data, etc. For large scale experiments in the context of proteomics, the comparison of binding patches

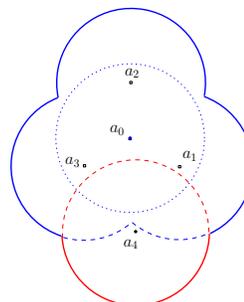
associated to interface models allow, in particular, the investigation of putative partners between orphan molecules.

**Classical interface models.** Classical interface models are twofold. May be the most widely used model is the so-called geometric footprint also called distance-based model, which consists of considering all pairs of atoms within a distance threshold  $d$ , typically in the range 5-8Å. But as illustrated on Fig. 1, considering all atoms of one partner which are within distance  $d$  from the second partner results in a bias towards convex regions [GLN04]. Another very popular interface model is that based on the Solvent Accessible Surface (SAS). Recall that the SAS is the boundary of balls with *expanded radii*, i.e. Van der Waals radii expanded by  $r = 1.4\text{Å}$  to account for a water probe. In this model, an interface atom is an atom contributing to the SAS of its own sub-unit, but losing part of this exposed surface in the complex. See Fig. 2. Interface atoms identified this way can further be classified as exposed or buried, depending on whether they retain accessibility in the complex, the former and latter making up the so-called rim and core of the interface. But as established in [CPBJ06] and explained in section 2.1, the SAS model actually omits privileged contacts.

Overall, a general drawback of these two models is that they do not provide a rich framework to compute pieces of information such as volume and surface areas, curvature information, dissection of the interface into patches. Instead, the computation of such quantities requires running dedicated algorithms.



**Fig. 1.** Defining an interface based on a distance threshold creates a bias towards convex regions.



**Fig. 2.** Defining an interface from atoms losing solvent accessibility. The dashed regions are exposed in the red and blue sub-units, but get covered in the complex.

## 1.2 The Voronoi interface

**Intuitive presentation.** Recall that the Euclidean Voronoi diagram of a collection of *sample* points is the partition of the ambient space into convex cells, such that all points in a cell have the sample associated to the cell as nearest neighbor. In bio-chemistry, since atoms' radii depend on their chemical type,

one replaces the Euclidean distance by the so-called power distance<sup>1</sup>. Abusing terminology, we still call the resulting diagram a Voronoi diagram. In the sequel, we shall consider the Voronoi diagram of atoms with expanded radii. See Fig. 3 for a 2D illustration.

Neighbors in a Voronoi diagram are actually *privileged neighbors*. That is, given two neighboring spheres  $S_i$  and  $S_j$ , and for any point  $p$  found on the dual Voronoi face, one has  $\pi(p, S_i) = \pi(p, S_j) < \pi(p, S_k), \forall k \neq i, j$ . This property is the main incentive for using pairs of neighboring regions to report interface neighbors. However, two atoms may share a Voronoi face, yet their relative distance might be arbitrarily large. To get around this difficulty, let a *restricted ball* or *restriction* be the intersection of this ball with its Voronoi region, see e.g. the red ball on Fig. 3(a). Focusing on pairs of neighboring restrictions instead of pairs of neighboring Voronoi regions allows one to report pairs of privileged neighbors without resorting to a distance cutoff. We illustrate this construction to define our Voronoi interface model.

Consider a complex involving two partners  $A$  and  $B$ , and denote  $W$  the water molecules, if any. These types are also referred to as *colors*. Atoms of type  $A/B/W$  are denoted  $a_i/b_i/w_i$ , respectively.

Let an *interface water molecule* be a water molecule such that its restriction has neighboring restrictions of type  $A$  and  $B$ . Water molecules which are not at the interface are called *bulk* water molecules. As illustrated on Fig. 3(a), our interface features pairs of restrictions of type  $[A; B]$  or  $[A; W]$  or  $[B; W]$ , with  $W$  standing for interface water molecules. Each pair actually conveys two pieces of information, namely the atoms associated to the restrictions, and the Voronoi facet, also called *tile*, dual of this edge. As illustrated on Fig. 3(b), an interface atom is an atom involved in at least one pair. Focusing on two types allows one to define three *bicolor* interfaces. That is, tiles of type  $AB$  ( $AW$  and  $BW$ ) define the interface  $AB$  ( $AW$  and  $BW$  respectively). Tiles of type  $AB$  define direct contacts between the partners, while tiles of type  $AW$  and  $BW$  define contacts between these partners which are mediated by interface water molecules. The union of tiles  $AW$  and  $BW$  defines the  $AW$ – $BW$  interface. The union of the  $AW$ – $BW$  and  $AB$  interfaces defines the  $ABW$  interface, which separates the partners and gives a global overview of the interaction area, regardless of the role played by water molecules. See Fig. 3(c,d).

Interestingly, the interface  $ABW$  can be *shelled into concentric shells*—prosaically speaking the process is similar to peeling an onion from the outside to the inside. The process yields a integer called the *Voronoi Shelling Order* or VSO to tiles and atoms. This VSO qualifies the depth of an atom at the interface—from one for rim atoms to an integer in the range 7-10 for most complexes. See Fig. 3(c,d).

**Connexions to classical interface parameters.** We now discuss the finding made with our interface model, and note in passing that the corresponding soft-

<sup>1</sup> Recall that the power distance of a point  $p$  to a sphere  $S(c_i, r_i)$  is defined by  $\pi(p, S_i) = \|c_i p\|^2 - r_i^2$ .

ware, **Intervor** [LC10], can be run and retrieved from <http://cgal.inria.fr/abs/Intervor>, together with plugins for VMD and Pymol.

We established in [CPBJ06] that our model identifies a superset of interface atoms losing solvent accessibility [BCR<sup>+</sup>04], which actually draws the attention to interactions between main chain atoms upon association. (Algorithms in section 2.1.) Interface tiles are naturally gathered into patches, which have been shown [CPBJ06] to be coherent with those obtained with classical clustering algorithms [CJ02]. (Algorithms in sections 2.2, 2.3 and 3.2.) Quantifying the planarity of interfaces and patches is important, e.g. to estimate (de-)solvation energies and also to identify putative binding regions for docking. While previous studies have used strategies based on plane fitting [JT96], the Voronoi interface comes with a notion of discrete (mean) curvature [CPBJ06], which allows to assess the curvature properties at any scale (from two tiles to the whole interface). (Algorithms in section 2.4.) Finally, the VSO provides a discrete interface depth parameterization which refines the dissection into a core and a rim [CCJ99,BCR<sup>+</sup>04]. In [BGNC09], this parameterization allowed us to sharpen the investigation of correlations between (i) the interface geometry, (ii) the location of polar residues [CJ02], (iii) the location of conserved residues [GC05], (iv) the dynamics of interfacial water [MRL07]. (Algorithms in section 3.3.)

We note in passing that another Voronoi interface definition has been proposed in [BER04]. This interface model uses the Voronoi diagram of the Van der Waals atoms (rather than the expanded radii), and closes small gaps at the interface using a growth process of the atoms which consists of expanding their squared radii by a value  $\alpha$ . As a consequence, the atoms reported are not qualified with respect to solvent accessibility.

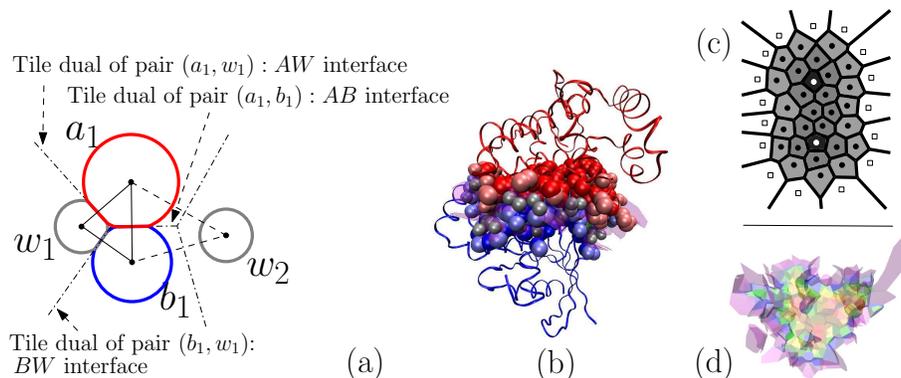
## 2 Bicolor Voronoi interfaces

In this section, we formally define bicolor Voronoi interfaces. We use the terminology of bicolor  $AB$  interface—the interface between the two proteins  $A$  and  $B$ , although the presentation is identical for any bicolor interface. We assume that the reader is familiar with the  $\alpha$ -complex of a union of balls [Ede92].

### 2.1 Bicolor interface and interface neighbors

To account for privileged contacts in the Voronoi diagram of atomic balls, we seek pairwise intersections of restrictions of different colors. To balls whose restrictions intersect actually define an edge in the  $\alpha$ -complex of the balls for  $\alpha = 0$ , whence the following:

**Definition 1.** *An  $AB$  interface edge is an edge of type  $AB$  in the  $\alpha$ -complex of the balls  $B_i$ , with  $\alpha = 0$ ; its endpoints are called interface atoms. The interface neighbors of a sphere  $S_i$  are the atoms of the second molecule sphere  $S_i$  is connected to through an interface edge.*



**Fig. 3.** (a) A fictitious complex with two atoms (red and blue) and two water molecules (in grey). The Voronoi diagram consists of the dashed-dotted line-segments. The interface comprises three pairs namely  $[a_1; b_1]$ ,  $[a_1; w_1]$ , and  $[b_1; w_1]$ ; water  $w_2$  is not at the interface. (b) Signal transduction complex (1tx4.pdb) : chains and interface atoms displayed with radii expanded by  $1.4\text{\AA}$ , with interface water molecules in grey. (c) Shelling a fictitious 2D interface into three shells (d) Shelling the  $ABW$  interface of complex 1txa into concentric shells: transparent view of the shells.

*The  $AB$  interface is defined as the collection Voronoi facets dual of the  $AB$  interface edges. A Voronoi edge bounding an interface Voronoi facet is called an interface Voronoi edge.*

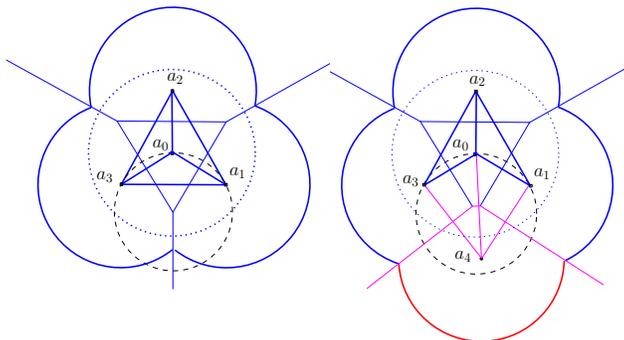
With respect to interface atoms defined with the BSA criterion, one can prove the following:

**Observation. 1** *Any atom  $S_i$  such that losing solvent accessibility during complex formation is an interface atom by Def. 1.*

However, the converse is false, as already mentioned in section 1.1, since an interface atom by Def. 1 may not lose solvent accessibility. A sufficient—but not necessary—condition for that is met when the atom is buried in its own sub-unit, and a 2D illustration is provided on Fig. 4. On that figure, the maintenance of the so-called *empty ball property* which characterizes Delaunay triangulations results in the creation an edge between the buried atom centered at  $a_0$  and the red atom centered at  $a_4$ .

## 2.2 Topology of bicolor Voronoi interfaces

Consider a bicolor interface, say the  $AB$  interface. Since the interface is a subset of the Voronoi diagram, it is a cell complex. To further qualify its topology, we need to examine how Voronoi facets patch together. Since a facet is the dual of a bicolor edge, we examine the tetrahedra containing this edge. We define:



**Fig. 4.** In the Voronoi model, interface atoms can be buried (a) Atom centered at  $a_0$  is buried in its subunit (b) Yet, it makes an interface edge with atom centered at  $a_4$  in the complex.

**Definition 2.** *The type of a tetrahedron featuring atoms of types  $A$  and  $B$  is denoted by a pair  $(i, j)$  where  $i, j$  respectively count the number of atoms of each type, and  $i + j = 4$ . Similarly, the type of a tetrahedron featuring an additional atomic type out of  $W, X$  is denoted by a triple  $(i, j, k)$ , with  $k$  the number of atoms of the third type, and  $i + j + k = 4$ .*

A case analysis of the types of tetrahedra yields:

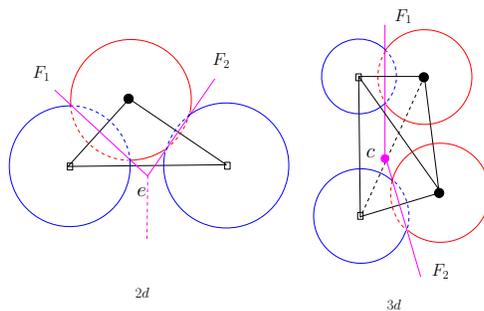
**Observation. 2** *The bicolor interface has the following properties:*

- *A Voronoi edge bounding an  $AB$  Voronoi facet has one or two incident  $AB$  Voronoi facets. Such a Voronoi edge is on the interface boundary iff only one edge of its dual triangle is an interface edge of type  $AB$ .*
- *The neighborhood of every Voronoi vertex is either a topological disk, a half-topological disk, or two half-topological disks pinched together at the Voronoi vertex. This later case correspond to a tetrahedron of type  $(2, 2)$ , where two bicolor edges are interface edges, and these two edges span the four vertices of the tetrahedron.*

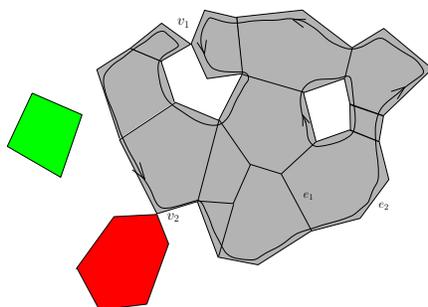
An interface can be connected through a pinched Voronoi vertex of a type  $(2, 2)$  tetrahedron. For such a tetrahedron, out of the four bicolor edges, only two actually occur in the  $\alpha$ -complex. Phrased differently, we have two independent bicolor pairs. See Figs. 5 and 6. Since the intersection of the corresponding Voronoi facets does not encode a joint property of these pairs, we define:

**Definition 3.** *Two Voronoi facets are called edge-connected if they share a Voronoi edge. An edge-connected component of the interface is a collection of edge-connected Voronoi facets. The closed curves bounding an edge connected component are called the loops.*

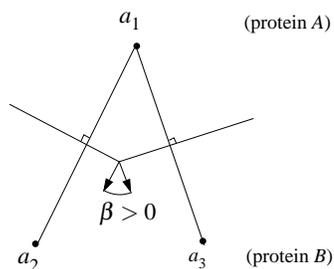
We now present the algorithms used to retrieve connected components and loops.



**Fig. 5.** (a) Top view of a Delaunay triangle in  $2d$ . Voronoi facets  $F_1$  and  $F_2$  meet along the Voronoi edge dual of the triangle, and are edge-connected (b) Side view of Delaunay tetrahedron contributing 2 bicolor interface edges. The dual Voronoi facets  $F_1$  and  $F_2$  meet at a pinched Voronoi vertex dual of a type  $(2, 2)$  tetrahedron, and are not edge-connected.



**Fig. 6.** An interface with three connected components and four boundary loops.



**Fig. 7.** Dihedral angle between Voronoi facets.

### 2.3 Computing bicolor interfaces and their boundaries

*Retrieving connected components.* Given an initial Voronoi facet  $f_0$  dual of an edge  $e_0$ , the exploration of the corresponding connected component requires running a breadth-first-search (or depth-first-search) like algorithm anchored at  $f_0$ . To run such an algorithm, the only information required is the list of Voronoi facets edge connected to a given facet  $f$ . But two edge connected Voronoi facets are dual of edges of the same triangle. Therefore, to report the facets edge connected to a facet  $f$  dual of an edge  $e$ , we just need to report the Delaunay triangles (i) incident to  $e$  (ii) featuring a second bicolor edge whose type is that of  $e$ . This is easily done by rotating around edge  $e$  in the Delaunay triangulation.

Upon completion of a connected component, we have a collection of Delaunay edges. We record them without any additional topological information as any

such information is encoded in the Delaunay triangulation. As an example, we present an algorithm to retrieve the cycles bounding a connected component. This algorithm operates on the Delaunay triangulation, but takes constant time per boundary edge.

*Retrieving the cycles bounding a connected component.* Starting from an initial boundary Voronoi edge  $e$ , a given cycle (also called loop) can be computed by iteratively following the successor of  $e$  on the boundary of the connected component. Assume  $e$  is oriented and denote  $s(e), t(e)$  the corresponding source and target Voronoi vertices, and let  $T(v)$  be the tetrahedron associated with Voronoi vertex  $v$ . The successor of edge  $e$  is one of the four Voronoi edges dual of the facets of  $T(t(e))$ . Following observation 2, if the neighborhood of  $t(e)$  is a half-topological disk, tetrahedron  $T(t(e))$  has only two facets whose dual Voronoi edges are boundary edges. Since  $e$  is one of them, extending the loop requires retrieving and following the second one. On the other hand, if  $t(e)$  is a pinched Voronoi vertex and if the four edges belong to the boundary of the connected component processed, there are three potential outgoing edges, but only one bounding the Voronoi facet which has  $e$  on its boundary. Again, extending the loop requires finding and following this edge.

Equipped with this extension operation, computing the loop of a given edge  $e_0$  requires picking an arbitrary orientation for  $e_0$ , and following the boundary until the Voronoi vertex  $T(s(e_0))$  is encountered again. To retrieve all the loops, we just have to iterate over the remaining boundary edges which are not already part of a loop.

The previous description actually eludes a difficulty, namely the way infinite tetrahedra are handled. (Recall these are the tetrahedra featuring a triangle of the convex hull of the Delaunay triangulation). When such a tetrahedron is encountered during the extension of a loop, we do not report its Voronoi center but the weighted circumcenter of the finite facet —that is the center of the smallest sphere orthogonal to the three spheres associated with the vertices of the facet.

## 2.4 Geometry of connected components

The most straightforward geometric statistics for an interface are its surface area and its boundary length. Apart from these, another interesting quantity is the interface curvature. Since a bicolor interface is a piecewise linear orientable surface, the natural way to characterize its *extrinsic* curvature consists of using the mean curvature. Notice that since we aim at studying the way a surface is embedded in  $\mathbb{R}^3$ , the extrinsic curvature is more suited than the intrinsic Gauss curvature, which is related to topological invariants and in particular boundary properties —cf the Gauss-Bonnet theorem. Recall that the mean curvature of a polyhedral surface is carried out along edges [San79], the amount of mean curvature attached to an edge being defined by  $h(e) = \beta(e)l(e)$ , with  $l(e)$  the edge length and  $\beta(e)$  the angle between the normals to the facets incident to  $e$ ,

the angle being counted positively (negatively) if  $e$  is convex (concave). We thus define  $s_H = \sum_{e \in \text{IVE}} h(e)$ , with IVE the collection of interior Voronoi edges of the interface.

In a bio-chemical setting, one expects dihedral angles to alternate. Therefore, large values of  $s_H$  indicate that the interface facets bend in a coherent fashion at the interface scale. Notice that in case of interfaces with several connected components, the components must be oriented coherently for the sum to make sense. But since the angle between the Voronoi facets matches the angle between the corresponding Delaunay edges, implementing this constraint means initializing the orientation of all the components in the same way —e.g. from protein  $A$  to protein  $B$ . See Fig. 7.

## 2.5 On the geometry of interface facets

As seen in section 1.2, bicolor edges selected from the 0-complex allow one to report interface neighbors without resorting to a distance cutoff. Phrased differently, long edges belong to the Delaunay triangulation but not the  $\alpha$ -complex. This filtering mechanism avoids using explicit solvent molecules, a strategy often resorted to in applications deriving statistical potentials from the Delaunay triangulation. However, this filtering mechanism does not provide any control on the geometry of the interface Voronoi facets, and in particular, large facets are expected near the convex hull of the atoms centers. In other words, interface edges encode the topology of the interface but not its geometry.

To retrieve a relevant geometric information, we build upon the observation that boundary atoms do not play a major role from an energetic standpoint [BT98], so that one may discard selected boundary edges these atoms are involved in. One way to discard large Voronoi facets is the following. Recall that any simplex in the  $\alpha$ -complex comes with a value  $\bar{\mu}$  which gives the weight of its largest orthogonal ball [Ede92]. For an interface edge  $e$ , denoting  $w_e$  the weight of its smallest ball, one can therefore discard the edge if  $\bar{\mu}/w_e \geq M$ , with  $M$  a positive number. Since weights of balls are equal to their square radii, the condition amounts to saying that the radii of the balls are within a factor  $\sqrt{M}$ .

## 3 Tricolor interfaces and water molecules

### 3.1 The $AW$ – $BW$ interface

When considering an interface, an interesting question is the role played by structural water molecules <sup>2</sup>. As these water molecules are described from the crystal as protein atoms are, we also expand their radius by the quantity  $r_w$ . Notice again that this expansion aims at mimicking an implicit continuous layer of solvent molecules on the atoms found in the crystal —be they protein atoms or water molecules.

<sup>2</sup> A water molecule is termed *structural* if it is as stable as the surrounding atoms. In a crystal structure, this can be assessed thanks to B-factors.

If one has three molecular types  $A, B, W$ , one can define three types of bicolor interfaces. But since we primarily care for the  $AB$  interface, contact of type  $AW$  and  $BW$  are of interest only when located near the  $AB$  interface, see Fig. 8. We therefore define:

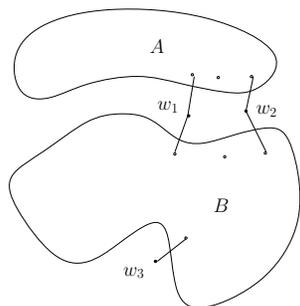
**Definition 4.** *An interface water molecule is a ball of type  $W$  which is the vertex of at least one edge of type  $AW$  and at least one edge of type  $BW$ , both edges belonging to the  $\alpha$ -complex of the balls  $B_i$ , with  $\alpha = 0$ . An  $AW$  (or  $BW$ ) interface edge is an edge of type  $AW$  (or  $BW$ ), with  $W$  an interface water molecule. The  $AW$  ( $BW$ ) interface is defined as the collection Voronoi facets dual of the  $AW$  ( $BW$ ) interface edges.*

A further refinement consists of aggregating Voronoi facets of type  $AW$  and  $BW$ :

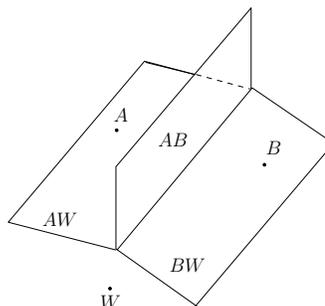
**Definition 5.** *The  $AW$ – $BW$  interface is the collection of Voronoi facets dual of edges of type  $AW$  or  $BW$ . A connected component of  $AW$ – $BW$  interface is a collection of edge-connected Voronoi facets dual of interface edges of type  $AW$  or  $BW$ .*

To study the  $AW$  or the  $BW$  or the  $AW$ – $BW$  interface, observe that edge connected Voronoi facets of types  $AW$  and  $BW$  are defined from bicolor, tricolor or quadricolor tetrahedra. Let us analyze the last two cases. In such a tetrahedron, we identify the labels  $A$  and  $B$  —since we do not report facets dual of such edges, so that the configurations found are those of bicolor tetrahedra. More precisely, a tetrahedron of type  $AABW$  where  $AB$  edges are omitted yields the same topological configurations as a bicolor  $(3, 1)$  tetrahedron for any bicolor interface. A tricolor tetrahedron of type  $ABWW$  is similar to a bicolor  $(2, 2)$  tetrahedron for any bicolor interface. Finally, a  $ABWX$  tetrahedron is equivalent to a  $(2, 1, 1)$  tetrahedron for any bicolor interface.

Therefore, the  $AW$ , the  $BW$ , and the  $AW$ – $BW$  interfaces have the same topological properties as the  $AB$  interface i.e. are surfaces with possibly pinched vertices.



**Fig. 8.** Water molecules centered at  $w_1$  and  $w_2$  are interface water molecules; that centered at  $w_3$  is not.



**Fig. 9.** The boundary of the union of the  $AB$  and  $AW$ – $BW$  interfaces may not be a one-manifold

### 3.2 The *ABW* interface

To assess the role of water molecules, and position relatively to one another the connected components of the *AB* and *AW–BW* interfaces, we define:

**Definition 6.** *The *ABW* interface is defined as the union of the *AB* and *AW–BW* interfaces.*

But the topology of the union is more involved than that of the singletons. First, non-manifold Voronoi edges may appear—if the three facets dual of the corresponding triangle are present in the union of the two interfaces. Second, the boundary of the union may not be a one-manifold, and we call it a *curve network* or net for short, see Fig. 9. To deal with these difficulties, it is actually sufficient to compute the *AB* and *AW–BW* interfaces separately, run a Union-Find algorithm to maintain the connected components of the edge-connected components, and another union-find algorithm to maintain the connected components of the boundary loops of the connected components of the union. Finding the Voronoi edges along which connexions occur can be done while computing the interfaces, while running  $m$  Union-Find operations on a  $n$ -element set takes  $O(m\alpha(m, n))$  with  $\alpha(m, n)$  the inverse of Ackerman’s function [Tar83].

### 3.3 Shelling the *ABW* interface

Considering the edge-connectivity of interface tiles, define the *depth* or the *Voronoi Shelling Order* of a tile as the number of tiles visited to reach it from the interface boundary—any tile which has a boundary edge is at depth one. This VSO provides an integer-valued parameterization of the *ABW* interface, which refines the binary core-rim model discussed in section 1.1.

## 4 Conclusion and Outlook

The interface model presented in this paper proved instrumental to refine our understanding of correlations between structural properties of protein interfaces, and important bio-physical quantities. However, the topic of modeling interfaces remains largely open for several reasons.

First, as evidenced by the scoring round of the community-wise docking experiment CAPRI, the design of scoring functions is an active area of research, and structural parameters defined from interface models should prove instrumental in this context.

Second, the question of precisely aligning and comparing interfaces has barely been touched upon. The Voronoi interface model proved instrumental for the description of bio-chemical properties, and we believe that the precise topological and geometric information it encodes should ease the comparison of interfaces in exhaustive structural classification studies.

Finally, while modeling single molecules and complexes is done routinely using methods from potential theory (molecular dynamics simulations, normal

modes), we are not aware of any significant work for the problem of modeling dynamic interfaces so as to possibly incorporate entropy-related terms into scoring functions.

## References

- [BCR<sup>+</sup>04] R. Bahadur, P. Chakrabarti, F. Rodier, , and J. Janin. A dissection of specific and non-specific protein–protein interfaces. *JMB*, 336(4):943–955, 2004.
- [BER04] Y.-E. A. Ban, , H. Edelsbrunner, and J. Rudolph. Interface surfaces for protein-protein complexes. In *RECOMB*, pages 205–212, San Diego, 2004.
- [BGNC09] B. Bouvier, R. Grünberg, M. Nilges, and F. Cazals. Shelling the voronoi interface of protein-protein complexes reveals patterns of residue conservation, dynamics and composition. *Proteins*, 76(3):677–692, 2009.
- [BT98] A.A. Bogan and K.S. Thorn. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.*, 280, 1998.
- [CCJ99] L. Lo Conte, C. Chothia, and J. Janin. The atomic structure of protein-protein recognition sites. *JMB*, 285(5):2177–2198, 1999.
- [CJ02] P. Chakrabarti and J. Janin. Dissecting protein-protein recognition sites. *Proteins*, 47(3):334–43, 2002.
- [CPBJ06] F. Cazals, F. Proust, R. Bahadur, and J. Janin. Revisiting the voronoi description of protein-protein interfaces. *Protein Science*, 15(9):2082–2092, 2006.
- [Ede92] H. Edelsbrunner. Weighted alpha shapes. Technical Report UIUCDCS-R-92-1760, Dept. Comput. Sci., Univ. Illinois, Urbana, IL, 1992.
- [GC05] M. Guharoy and P. Chakrabarti. Conservation and relative importance of residues across protein-protein interfaces. *PNAS*, 102(43):15447–15452, Oct 2005.
- [GLN04] Raik Grünberg, Johan Leckner, and Michael Nilges. Complementarity of structure ensembles in protein-protein binding. *Structure*, 12(12):2125–2136, Dec 2004.
- [JT96] S. Jones and JM Thornton. Principles of protein-protein interactions. *PNAS*, 93(1):13, 1996.
- [LC10] S. Lorient and F. Cazals. Modeling macro-molecular interfaces with intervor. *Bioinformatics*, 26(7):964–965, 2010.
- [MRL07] I. Mihalek, I. Reš, and O. Lichtarge. On itinerant water molecules and detectability of protein–protein interfaces through comparative analysis of homologues. *JMB*, 369(2):584–595, 2007.
- [San79] L. Santaló. *Integral Probability and Geometric Probability*, volume 1 of *Encyclopedia of Mathematics and its Applications*. Addison-Wesley, 1979.
- [Tar83] R. E. Tarjan. *Data Structures and Network Algorithms*, volume 44 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1983.