



# Geometric, topological and contact analysis of interfaces in macro-molecular complexes: from the atomic to the complex scale using Intervor

Frédéric Cazals

## ► To cite this version:

Frédéric Cazals. Geometric, topological and contact analysis of interfaces in macro-molecular complexes: from the atomic to the complex scale using Intervor. [Research Report] RR-5864, INRIA. 2006, pp.36. inria-00070162

**HAL Id: inria-00070162**

**<https://inria.hal.science/inria-00070162>**

Submitted on 19 May 2006

**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.

***Geometric, topological and contact analysis of  
interfaces in macro-molecular complexes: from the  
atomic scale to the complex scale using Intervor***

Frédéric Cazals

**N° 5864**

Mars 2006

Thème SYM



***rapport  
de recherche***





# Geometric, topological and contact analysis of interfaces in macro-molecular complexes: from the atomic scale to the complex scale using Intervor

Frédéric Cazals

Thème SYM — Systèmes symboliques  
Projet Geometrica

Rapport de recherche n° 5864 — Mars 2006 — 36 pages

**Abstract:** Understanding the *sociology* of interactions between the proteins encoded in a genome is a central question of structural biology, and interface models between molecules forming a complex are instrumental in this perspective. Qualifying interface atoms as atoms losing solvent accessibility in the complex, or pairs of atoms within a distance threshold, several interface models have been proposed. Yet, until recently, no interface model existed to answer coherently (if at all) the following questions: can one bridge the gap from atoms losing solvent accessibility to interface pairs? is the interface flat or curvy? is it connected or not (does it have a multi-patch structure)? is a connected component of the interface simply connected or not (does it have a hole)? what is precisely the role played by interface structural water?

Using the  $\alpha$ -complex of the Van der Waals balls, a construction derived from the Voronoi diagram, we designed such an interface model, and validated it on the usual database of co-crystallized protein-protein complexes. This paper is a methodological paper aiming at easing the access of the interface model to structural biologists. As such, the paper overviews: (i) the geometric principles underlying the interface model (ii) the definitions of the interface and its extension to accommodate structural water (iii) the statistics one can compute from the interface model (iv) the Software **Intervor** and the associated web site. These presentations are accompanied by illustrations and insights on protein - protein complexes.

**Key-words:** Structural Biology, Molecular Modeling, Molecular Interfaces, Voronoi diagrams, Geometry and Topology of interfaces.

# Analyse géométrique, topologique et des contacts à l'interface de complexes macro-moléculaires: de l'échelle atomique à celle du complexe avec Intervor

**Résumé :** Comprendre la sociologie des interactions entre toutes les protéines codées par un génome est une question centrale en biologie, et les modèles d'interface entre deux ou plusieurs molécules sont d'importance capitale dans cette perspective. Jusqu'à récemment, les modèles classiques d'interface définissaient les *atomes à l'interface* comme ceux soit perdant de l'accessibilité au solvant dans le complexe, soit ayant un atome proche sur une autre molécule participant au complexe. Cependant, aucun de ces modèles ne permettait de répondre précisément et de façon cohérente aux questions suivantes: y a-t-il un lien entre les atomes perdant de l'accessibilité au solvant et les paires d'atomes à une distance donnée? une interface est-elle plate ou courbée? est-elle connexe ou pas? une composante connexe est-elle simplement connexe? quel est le rôle précis joué par les molécules d'eau à l'interface?

En utilisant l' $\alpha$ -complexe des boules de Van der Waals balls, une construction dérivée du diagramme Voronoï diagramme, nous avons développé un modèle d'interface répondant à ces questions. De plus, ce modèle a été validé sur les complexes protéine - protéine classiques.

Ce papier, de nature méthodologique, est rédigé à l'intention des biologistes intéressés par l'étude de complexes. Sont ainsi couverts: (i) les pre-requis géométriques pour aborder le modèle; (ii) la définition de l'interface d'un complexe, et son extension pour caractériser le rôle de l'eau cristallographique; (iii) les statistiques que l'on peut calculer à partir du modèle; (iv) le logiciel **Intervor** et le site Web associé.

Le propos est accompagné d'illustrations et de propriétés mise en lumière par notre modèle d'interface sur certains complexes protéine - protéine.

**Mots-clés :** Biologie Structurale, Modélisation Moléculaire, Interfaces, Protéines, Diagrammes de Voronoï.

# 1 Complexes and interfaces

## 1.1 Previous work

Van der Waals models, where an atom is represented by a ball whose radius depends on the chemical type of the element as well as its environment, are instrumental in investigating the structure of macro-molecules and also the interaction between such molecules. Because co-crystallized complexes features what Nature does, an important amount of investigations focused on the design of statistics able to discriminate biological versus non biological complexes. Over the years, a wealth of such statistics have been proposed, using a variety of methods and algorithms.

Based on atoms loosing accessibility in a complex, parameters of interfaces at the atomic and the interface scales are studied in [CJ99, ME02, CJ02, BCRJ04]. At the atomic level, the classification of interface atoms as buried or exposed, the description of neighbors, the packing properties using Voronoi volumes, the dissection of chemical properties of interface atoms are investigated. In order to discriminate homodimeric and monomeric proteins from the crystalline state, statistical potentials mixing distances and surface areas are developed in [PHT00]. Statistical potentials charging surface areas to pairs of residues are defined in [AT97]. Aside these applied contributions, more conceptual interface models have been developed. In [GW96], an interface is defined as the points in-between two molecules where two potentials emitted by the molecules take the same value. Of particular interest is [BER04], which defines an interface from a *retraction* process based on the  $\alpha$ -complex [Ede92, EM94].

## 1.2 Contributions and paper overview

In all the applied contributions afore-mentioned, interface atoms are either described as atoms loosing solvent accessibility upon formation of a complex, or atoms having a neighbor within a distance cut-off. Interestingly, these contributions raise a number of interesting questions:

can one bridge the gap from atoms loosing solvent accessibility to interface pairs? is the interface flat or curvy? does it have a multi-patch structure (mathematically, is it *connected* or not)? does a patch contain holes (mathematically, is it *simply connected* or not)? is a connected component of the interface simply connected or not (does it have a hole)? what is precisely the role played by interface structural water?

We recently developed an interface model answering these questions using a unique geometric constructions derived from the Voronoi diagram of the Van der Waals balls. At the atomic level, our construction lists the interface atoms and the neighbors of a given atom, and quantifies the contacts between neighbors. At the interface level, the construction provides informations on the geometry of the interface (its flatness, its surface area), on its topology (the number of patches, the presence of holes within a patch). This interface model is precisely described from the mathematical and algorithmic viewpoints in [CP05], while a

detailed study on the usual data-base of protein-protein complexes can be found in [CP<sup>+</sup>]. Aside these two publications, this paper aims at easing the access of our interface model to structural biologists, and presenting the software **Intervor**. The paper is organized as follows.

The geometric principles underlying our interface model are over-viewed in section 2. The definitions of the interface and its extension to accommodate structural water are presented in section 4. The statistics one can compute from the interface model are enumerated in section 5. An overview of the **Intervor** software is provided in section 6. Finally, appendices provide supplementary material. Selected illustrations and insights gained from the interface model are listed in section 8. The classes of atomic annotations used by **Intervor**, the files generated for a given complex, and the chemical annotations of atoms and pairs, are respectively developed in sections 9, 10 and 11.

## 2 Voronoi diagram, $\alpha$ -complex, $\alpha$ -shape

In this section, we review the geometric pre-requisites of our interface model. Notice these are of major interest for the definition and the calculation of molecular surfaces and also for the characterization of pockets in molecules.

### 2.1 Delaunay - Voronoi

We aim at investigating Van der Walls models, where each atom is represented by a sphere, and for such a sphere, the associated ball refers to the solid region bounded by the sphere.

For Van der Walls models, the pioneering work of Richards [Ric74] launched the systematic use of Voronoi diagrams and relatives for structural studies. Richards originally used the Euclidean Voronoi diagram, where the Voronoi cell of an atom is the convex polyhedron defined from the bisecting planes associated to the center of the atom and the centers of its neighbors. For balls of different radii, the natural generalization of the Euclidean Voronoi diagram is the so-called *power diagram* [Aur87]. To define it, recall that the *power* of a point  $p$  wrt a sphere  $S_i(a_i, r_i)$  centered at  $a_i$  and of square radius  $w_i = r_i^2$  is defined by  $\pi(p, S_i) = a_i p^2 - w_i$ . Given two spheres, the sets of points having equal power wrt two spheres is called the radical flat of the spheres.

Given a collection of spheres, replacing the bisecting planes by the radical flats yields the power diagram. For balls of equal radii, the power diagram reduces to the Euclidean Voronoi diagram. The two being equivalent (they are affine diagrams), we shall call the power diagram a Voronoi diagram and associate it with its dual, the Delaunay triangulation. The Delaunay triangulation is a simplicial complex containing simplices of dimension zero (vertices), one (edges), two (triangles), and three (tetrahedra), built as follows from the Voronoi diagram: whenever two or more Voronoi cells have a common intersection (Voronoi facet, Voronoi edge or Voronoi vertex), one finds the simplex spanning the sphere centers in the Delaunay triangulation. As an example, the Delaunay triangulation of Fig. 1 features four vertices, six edges and three triangles, defined from the atoms' centers  $a_1, \dots, a_4$ .

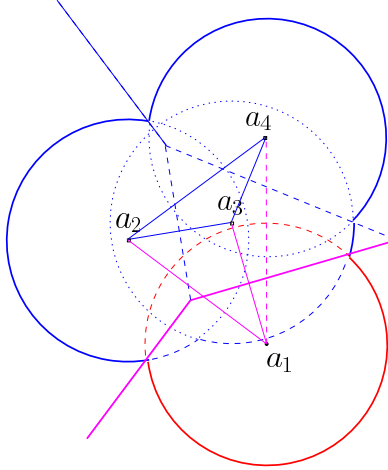


Figure 1: Bicolor interface defined from the  $\alpha$ -complex of the atoms. Interface edges are  $a_1a_2$  and  $a_1a_3$ .

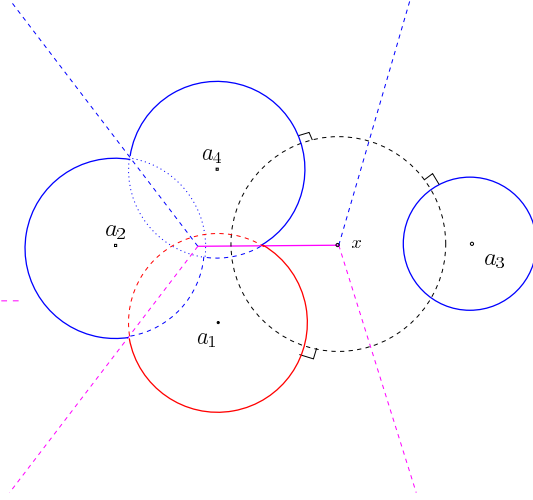


Figure 2: The sphere centered at  $x$  is orthogonal to spheres centered at  $a_1, a_3, a_4$ , and is the largest sphere orthogonal to the spheres centered at  $a_1$  and  $a_4$ . Its radius can be used to accept/reject the interface Voronoi facet associated to the Delaunay edge  $a_1a_4$ .

## 2.2 The $\alpha$ -complex and the $\alpha$ -shape — union of balls

Noting that the radical flat of two spheres is invariant if the same quantity  $\alpha$  is added to the square radii of both spheres, Edelsbrunner [Ede92] introduced the  $\alpha$ -complex. For a given  $\alpha$ , the  $\alpha$ -complex is a subset of the Delaunay triangulation, built as the Delaunay triangulation, excepted that one restricts each Voronoi cell to its associated ball, or vice-versa, and seeks intersections between these restricted regions. The restriction of a ball to its Voronoi region is called a *restricted ball*. More precisely, whenever two or more restricted balls have a common intersection, one finds the simplex spanning the sphere centers in the  $\alpha$ -complex. The  $\alpha$ -shape is defined as the domain of the  $\alpha$ -complex, i.e. the union of all points found on simplices of the  $\alpha$ -complex. Getting back to Fig. 1, it is easily checked that the 0-complex features four vertices, five edges and two triangles. Notice in particular that edge  $a_1a_4$  does not belong to the 0-complex since the restrictions of the balls centered in  $a_1$  and  $a_4$  to their respective Voronoi regions do not intersect. If we increase  $\alpha$ , more simplices are found in the  $\alpha$ -complex, which coincides with the Delaunay triangulation for large enough  $\alpha$ .



To proceed with the  $\alpha$ -complex, we shall need some terminology. The faces<sup>1</sup> of a simplex  $s$  are the simplices whose vertices are a subset of those of  $s$ . Dually, the cofaces of a simplex  $s$  are the simplices having  $s$  as a face.

This said, for a fixed value of  $\alpha$ , simplices of the  $\alpha$ -complex are classified into three categories: a simplex is called *singular* if it is on its own i.e. does not have any coface; it is *interior* if it has the same cofaces as in the Delaunay triangulation and is not located on the convex hull; it is *regular* otherwise. Consider the following two examples. On Fig. 1, the Delaunay edge  $a_2a_3$  is interior since its two Delaunay triangles are present as cofaces, while all other edges are regular. On Fig. 2, the vertex  $a_3$  is singular.

Notice that practically, computing the  $\alpha$ -complex of a union of balls is a non trivial task. Fortunately, robust and efficient implementation exists: see in particular the implementation available in the `Alpha_shape_3` package of the `Computational Geometry Algorithms Library`, [www.cgal.org](http://www.cgal.org).

## 2.3 Applications to molecular modeling

Several important properties of Van der Waals models can be inferred from the  $\alpha$ -complex.

**Molecular surfaces.** Given a collection of (possibly intersecting) balls, the *boundary of the union* of these balls refers to the spherical caps of all spheres which are not contained in any other ball. The  $\alpha$ -complex and more precisely the 0-complex can be used to retrieve the boundary of a union of balls [Ede92, EFL98]. For example, the Van der Waals surface just reads from the  $\alpha$ -complex of the balls  $S_i(a_i, r_i)$ , with  $\alpha = 0$ . Similarly, the SAS surface is described by the boundary of the union of balls  $S_i(a_i, r_i + r_w)$ , i.e. the Van der Waals balls expanded by the water probe, and read from the 0-complex of these expanded balls. More precisely, simplices contributing to the boundary of a union of balls are exactly those which are singular or regular.

**Atoms: buried or exposed?** Consider an atom in a union of balls. This atom is called *exposed* if it contributes to the boundary of the union, and *buried* otherwise. Following the property of molecular surfaces just recalled, an atom is exposed (buried) iff the corresponding vertex is singular or regular (interior) in the 0-complex of the balls. Notice this property is valid for the Van der Waals and SAS surfaces.

**Filtering (long) Delaunay edges.** Consider a Delaunay edge such that the restricted balls associated to its vertices do not intersect. Such an edge does not belong to the  $\alpha$ -complex, which can thus be used to filter long Delaunay edges. In particular, this strategy avoids using (fake) explicit water molecules to remove long Delaunay edges.

**Filtering (large) Voronoi facets.** Recall that two intersecting spheres / balls are called orthogonal if they meet at a right angle—see Fig. 2. For any point  $p$  on a (relatively open) Voronoi face  $f$ , there exists a unique ball centered at  $p$  and orthogonal to the balls found

<sup>1</sup>For example a tetrahedron has respectively 4/6/4 faces of dimension 2/1/0, usually called triangles, edges, vertices.

at the vertices of the simplex dual of the Voronoi face  $f$ . In particular, the radius  $r_x$  of the largest such ball reads from the  $\alpha$ -complex of the balls. See the ball centered at  $x$  on Fig. 2.

Consider a Delaunay edge  $e$  and its dual Voronoi facet  $f$ , and let  $r_e$  stand for the smallest radius of the balls found at the endpoints of  $e$ . Given a threshold  $M$ , one may discard the Voronoi facet  $f$  and its dual Delaunay edge  $e$  based upon the following criterion:

$$r_x/r_e > M. \quad (1)$$

Such a filtering mechanism can be applied to Delaunay simplices of all dimensions. As an example, consider again Fig. 2: comparing the radius of the ball centered at  $x$  and that of atom  $a_1$  (or  $a_4$ ), one can decide the Voronoi facet dual of  $a_1a_4$  is too large.

### 3 Crystals and structural water: the $AB$ and $ABW$ models

Consider a PDB file featuring a complex. For example, for a protein-protein complex, each partner is specified as a collection of polypeptidic chains. All atoms found in the PDB file are either tagged as  $A$  or  $B$  for the two partners,  $W$  for structural water,  $X$  for the remaining atoms. These four tags are called the types of the atoms, and a pair (triple) of atoms is called *bicolor* (*tricolor*) if it features two (three) different types. Since we define our interface model from the Voronoi diagram of the atomic balls, we introduce two models:

- $ABW$  model: all atoms found in the PDB file, excepted hydrogens and possibly selected water molecules, are inserted into the Voronoi diagram.
- $AB$  model: all atoms found in the PDB file excepted hydrogens and water molecules are inserted into the Voronoi diagram.

In particular, notice the  $ABW$  model is highly relevant for high resolution crystals — resolution better than  $2\text{\AA}$ , where water molecules are spotted reliably. Notice also that one may wish to retain all such molecules, or only those having a relatively low temperature factor.

These two models call for a comment regarding buried atoms. As recalled above, an atom is buried in a collection of balls if it does not contribute to the boundary of the union. For complexes, by *boundary of the union* we refer to the Solvent Accessible surface (SAS). In particular, an atom contributing to the SAS is in contact with the water probe used to define the SAS. Thus, in the  $AB$  model, an atom is buried if it does not contribute to the SAS of the complex in the collection of balls of types  $A, B, X$ . Similarly, in the  $ABW$  model, this atom is buried if it does not contribute to the SAS of the complex in the union of all Van der Waals balls. But then, notice that a buried atom, which does not contribute to the SAS of the complex, may actually be in contact with structural water found in the crystal.

## 4 Bicolor and tricolor interfaces

### 4.1 The $AB$ , $AW$ – $BW$ , $ABW$ interfaces

**Interface atoms and edges.** Given a Voronoi diagram, an appealing way to report neighbours is to report all pairs of atoms of different types corresponding to Voronoi cells sharing a Voronoi facet—or phrased differently bicolor Delaunay edges. Unfortunately, this definition is not satisfactory since atoms far apart may have neighboring Voronoi cells. To circumvent this difficulty, and following the construction of the  $\alpha$ -complex, we restrict a Voronoi region to its Van der Waals ball (or vice-versa), and focus on intersections between these restricted regions. Thus, a bicolor Delaunay edge is selected as interface edge if:

- The Delaunay edge is such that its two restricted balls intersect;
- If one wishes to filter out large Voronoi facets, the Delaunay edge is not discarded by Equation (1).

Any atom involved in at least one interface edge is termed an *interface atom*. Pairs selected this way do not account for all pairs within a distance threshold, but it is proved in [CP05] that such interface atoms form a superset of atoms losing solvent accessibility. Discussing more precisely the types of edges selected, we define 3 interface types:

- $AB$  interface: the interface specified by edges of type  $AB$ . This interface describes the contacts between both partners.
- $AW$ – $BW$  interface: the interface specified by edges of type  $AW$  or  $BW$ . This interface describes the contacts between the partners and structural water.
- $ABW$  interface: the union of the  $AB$  and  $AW$ – $BW$  interfaces. This interface positions relatively to one another the previous two interfaces.

Finally, since a Delaunay edge is dual of a Voronoi facet, notice that an interface is either specified from bicolor Delaunay edges found in the 0-complex, or from the Voronoi facets dual of these edges.

**Filtering large Voronoi facet.** Equation (1) provides a way to filter large Voronoi facet. In particular, all infinite Voronoi facets, which are dual of Delaunay edges found on the convex hull, may be filtered this way. An example such facet is the facet dual of  $a_1a_2$  of Fig. 2. More generally, as also illustrated on the same figure, finite facets may be filtered using the radius  $r_x$  of the largest orthogonal ball.

Practically, such a filtering mechanism is of interest for visualization purposes, or if one wishes to interpret the surface area of the Voronoi facets as in [CP<sup>+</sup>].

**The topology of interfaces.** Prosaically, topology is the geometry of rubber made objects: topological properties on an object are indeed invariant upon geometric deformations that neither tear the object apart nor (self-)glue it. We aim at investigating two topological questions about interfaces: is an interface connected, i.e. is made of one or several patches? given a connected component, does it contain a hole—in which case we say it is *non simply connected*, or not—in which case the patch is said to be a *topological disk*? As we shall see, these questions are of central interest to assess the role of structural water molecules at interfaces.

Because the topology of an object consisting of pieces is about the way these pieces patch together, we need to examine how Voronoi facets glue together. Refer to Fig. 3 for an illustration. Two Voronoi facets sharing a Voronoi edge are called *edge-connected*, or connected for short. A Voronoi edge shared by two edge-connected facets is termed *interior*, and *boundary* otherwise. Consider a collection of Voronoi facets of an interface of any of the three types. If there exists a transitive path between any two of them using edge-connectivity, this collection of facets is called edge-connected. Moreover, if this collection is maximal, i.e. cannot be added any edge-connected facet, it is called a *connected component* or cc for short. A cc corresponds to *transitive* pairwise intersections between atoms. For example, a cc of the *AB* interface corresponds to a collection of *A* and *B* atoms such that a restricted atom of type *A* intersects a restricted atom of type *B* which intersects a restricted atom of type *A* and so on.

Having defined a cc, describing the topology of an interface amounts to describing the way the interface facets patch together. In particular, in mathematics, a *manifold* is a space of *homogeneous* dimension. For example, a curve is a one-manifold, an ellipsoid and a cube are two dimensional manifolds. But the two line-segments defining the X letter do not define a one-manifold because the neighborhood of the intersection consists of two *branches* and not one. Are our interfaces manifold surfaces?

To investigate this question, consider a Delaunay tetrahedron *T*. The six Voronoi facets dual of the six edges of *T* meet at the Voronoi vertex dual of *T*. For all 3 interface types, configurations where two of the six Voronoi facets meet in a *beak-to-beak* fashion at the Voronoi vertex occur, as illustrated on Fig. 4. Such a Voronoi vertex is called a *pinched* Voronoi vertex, and actually corresponds to two pairs of *independent* atoms in *T*, in the sense that the atoms in each pair define a bicolor edge. Consequently, this connectivity property is not used to reconnect edge connected components. In any case, interfaces are not manifold at pinched Voronoi vertices. On the other hand, it is shown in [CP05] that Voronoi edges of the *AB* and *AW–BW* interfaces are either boundary edges or manifold edges, while edges of the *ABW* interface may not be manifold because three Voronoi facets can share such an edge. See Fig. 5 for an illustration, and [CP05] for the details.

Summarizing this discussion and abusing terminology by omitting the pinched vertices and focusing on Voronoi edges, we shall say the *AB* and *AW–BW* interface are manifold interfaces. The curve bounding a cc of the *AB* interface is called a *boundary loop*, while the collection of boundary edges bounding a cc of the is called a *boundary network*.

Finally, a comment is in order wrt cc of the  $AB$  and  $AW-BW$  interfaces. The  $AB$  interface describes the contacts between both partners, while the  $AW-BW$  interface describes the contacts between the partners and structural water. When one cc of the  $AB$  and one cc of the  $AW-BW$  interfaces have a common Voronoi edge, we merge them into a single cc of the  $ABW$  interface. The  $ABW$  interface is thus defined as the union of the  $AB$  and  $AW-BW$  interfaces, and quantifies the relative positions of cc of the  $AB$  and  $AW-BW$  interfaces. In particular, one expects the  $ABW$  interface to provide informations about packing defects filled by structural water in-between the two partners  $A$  and  $B$ .

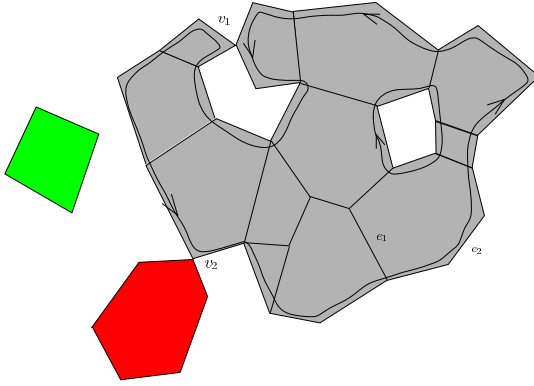


Figure 3: An interface with 3 edge connected components. The red and grey connected components, connected by a pinched Voronoi vertex, are not edge connected. The boundary of the grey component consists of two loops. Edge  $e_1$  is an interior Voronoi edge, edge  $e_2$  is a boundary Voronoi edge.

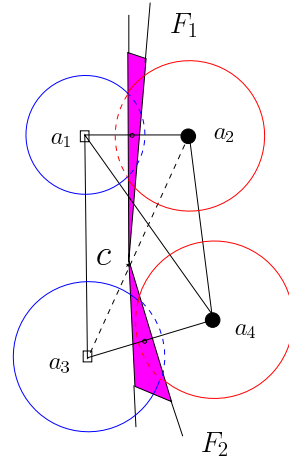


Figure 4: Voronoi facets  $F_1$  and  $F_2$  meet at a pinched Voronoi vertex  $c$ . Out of the 4 bicolor edges of the tetrahedron, only  $a_1a_2$  and  $a_3a_4$  are found in the  $\alpha$ -complex

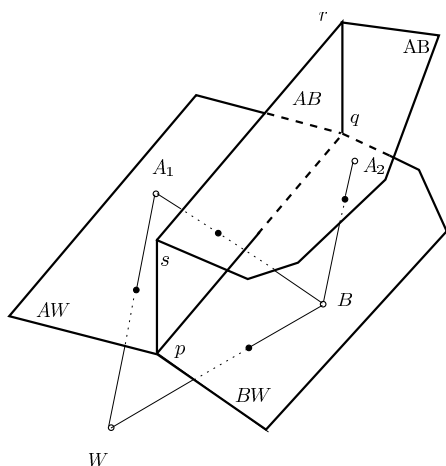


Figure 5: An  $ABW$  interface with four atoms  $A_1, A_2, B, W$ : edge  $pq$  is shared by three Voronoi facets of types  $AB, AW, BW$  and is non manifold; edge  $rs$  is shared by two Voronoi facets and is manifold. The boundary net consists of 15 edges (all edges but edge  $pq$ ).

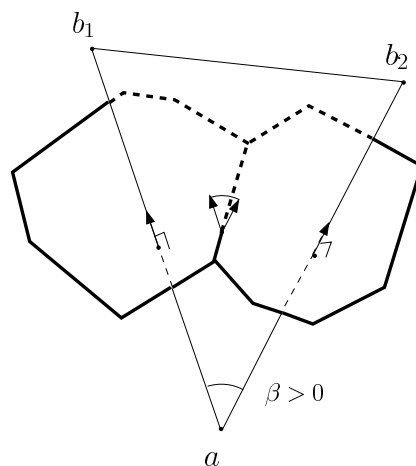


Figure 6: A triangle with one atom of type  $A$  and two atoms of type  $B$ : the dihedral angle between the Voronoi facets dual of edges  $ab_1$  and  $ab_2$  is also the angle between the two Delaunay edges. For a triangle with two (one) atoms of type  $A$  ( $B$ ), the angle is counted negatively.

## 4.2 Curvature of the $AB$ interface

In mathematics, the curvature of a smooth surface is characterized by its so-called Gauss and mean curvatures. While the Gauss curvature is related to global topological properties via the so-called Gauss-Bonnet theorem, the mean curvature encodes what the eyes see, i.e. the way the surface is *embedded* into  $\mathbb{R}^3$ .

As just outlined, the  $AB$  interface is a polyhedral surface, which, apart from pinched Voronoi vertices, is a manifold surface. Interestingly, mathematicians worked out curvature notions for polyhedral surfaces, and we resort to the so-called *discrete mean curvature* to characterize the curvature of the interface. Consider a Voronoi edge  $e$  shared by two facets of the  $AB$  interface, and denote  $l(e)$  its length, and  $\beta(e)$  the *signed* dihedral angle. The mean curvature carried by edge  $e$  is defined by  $h(e) = \beta(e)l(e)$ . See Fig. 6.

Notice that we measure the curvature of the  $AB$  interface, rather than computing a curvature notion associated to the atoms themselves. However, it should be kept in mind that each Voronoi facet witnesses the intersection of two atomic balls whose centers are located apart from the plane containing the Voronoi facet, so that the curvature of the interface is a fingerprint of the *curvature* of the union of interface atoms.

## 5 Statistics

In this section, we review the statistics one can compute from the interface model, and mention their relevance to investigate complexes. All informations required to compute such statistics are available in the files generated by **Intervor**.

### 5.1 Notations

Denote  $\#A / \#B / \#W / \#X$  the number of interface atoms of each of the four species. Also denote  $\#XY$  the number of interface edges between atoms of types  $X$  and  $Y$ , and  $\#X_{XY}$  the number of interface atoms of type  $X$  involved in the  $XY$  interface. In the  $AB$  model, we always have  $\#A_{AB} = \#A$ . But in the  $ABW$  model, we may have  $\#A_{AB} < \#A$  since some  $A$  interface atoms may be so due to contacts with interface water molecules. In the  $AB$  or  $ABW$  model, the fraction of atoms buried is denoted  $bur$ . To compare a statistic  $S$  between the two models, we denote  $R(S)$  the ratio ( $S$  in the  $ABW$  model) / ( $S$  in the  $AB$  model).

### 5.2 Connectivity properties of interface atoms

We address the following questions: *What are the neighbors of a given atom? are the connectivity properties of atoms of the two partners equivalent? How does structural water influence these properties?*

In the  $AB$  model, one is interested in the number of interface atoms of types  $A$  and  $B$ . In the  $ABW$  model, one wishes to report the number of interface atoms of types  $A, B, W$ , together with the ratios  $\#A_{AB}/\#A$  and  $\#A_{AW}/\#A$  —and similarly for atoms of type  $B$ . These ratios quantify the contributions of the  $AB$  and  $AW$  interfaces to the qualification of an atom of type  $A$  or  $B$  as interface atom. For both models, we report the fraction of buried atoms. To compare the two models and assess the influence of water molecules, one reports the ratios  $R(\#A + B)$  and  $R(bur)$ .

The number of atoms and interface edges can be used to compare the average number of neighbors of a given atom. In the  $AB$  model, one compute the average number of neighbors for  $A$  and  $B$  —given by the formula  $n_X = \#XY/\#X$ . Combining these values for  $A$  and  $B$  yields the interface average number of neighbors:

$$n_g = \frac{\#A}{\#A + B} n_A + \frac{\#B}{\#A + B} n_B = \frac{2\#AB}{\#A + B}. \quad (2)$$

In the  $ABW$  model, one computes the average number of neighbors for  $A$  from the formula  $n_A = (\#AB + \#AW)/\#A$ , and similarly for  $B$ . Again, combining these values with the frequencies of  $A$  and  $B$  yields the  $ABW$  interface average number of neighbors:

$$n_g = \frac{\#AW + \#BW + 2\#AB}{\#A + B}. \quad (3)$$

To study the potential asymmetric situations, a statistic of interest is  $r_{Mm} = \text{Max}(n_A, n_B) / \text{min}(n_A, n_B)$ . Again, this statistic can be used to compare both models from the ratio  $R(n_g)$ .

### 5.3 The geometry of interfaces

We address the following questions: *How large is an interface? Is an interface flat or curvy?*

**Voronoi Interface Surface Area.** Interface surface areas are usually computed from the Surface Area Buried (BSA) upon formation of the complex [CJ99, CMJW03]. Denoting SASA the SAS Area, one has  $BSA(\text{complex}) = SASA(A) + SASA(B) - SASA(A \cup B)$ . Notice this BSA is usually computed without structural water, and assumes no conformational change in the complex since SAS of the partners are computed from their geometries in the complex.

Given a Voronoi interface, the Voronoi Interface Surface Area (VISA) is naturally defined as the sum of the surface areas of the facets defining the interface. Denote  $\text{Area}(XY)$  the surface area of the bicolor interface  $XY$ . For both models, we define:

- $A_{AB}$ ;  $AB$  model. Defined as the sum of the surface areas of the  $AB$  facets, that is  $A_{AB} = \text{Area}(AB)$ .
- $A_{ABW}$ ;  $ABW$  model. In the  $ABW$  model, we have  $AB$  but also  $AW$  and  $BW$  facets. Since a water molecule is always *sandwiched* between atoms of type  $A$  and  $B$ , it is natural to define the VISA by  $\text{Area}(ABW) = \text{Area}(AB) + (\text{Area}(AW) + \text{Area}(BW))/2$ .

**Flatness and curvature properties.** To report on curvature properties, we focus on the discrete mean curvature of the  $AB$  interface introduced in section 4.2. Let  $IE$  stand for all bicolor edges of the  $AB$  interface. To get a global view of the interface absolute mean curvature, we compute

$$\overline{s_H} = \sum_{e \in IE} |h(e)|. \quad (4)$$

To capture cooperative effects at the interface scale, we compute

$$s_H = \sum_{e \in IE} h(e). \quad (5)$$

Large values of  $s_H$  indicate that at the interface scale, Voronoi facets bend in a coherent fashion. But a low value of  $s_H$  does not necessarily mean the interface is flat. Like the faces of a cube, the interface may feature several *flat* components —each having a small  $s_H$  value but located in different planes. Thus  $s_H$  aims at characterizing *curvy* rather than *non planar* interfaces. Finally, to get a local view of curvature properties, we report the expectation and the standard deviation of the absolute value of the dihedral angle —the weight of the angle at edge  $e$  being  $l(e)/L$ , with  $L$  the sum of the lengths of all internal edges.



## 5.4 The topology of interfaces

We address the following questions: *Is an interface connected or not (does it have a multi-patch structure)? Is a connected component of the interface simply connected or not (does it have a hole)? What is precisely the role played by interface structural water?*

**Connected components.** The interface model defines connected components (cc). Given a threshold  $t_s \in [0, 1]$ , define a significant connected component (scc) as a cc whose surface area is at least a fraction  $t_s$  of the VISA of the  $AB$  interface. In the  $AB$  model, we report the number of cc and scc, denoted  $\#cc$  and  $\#scc$ .

Consider now the  $ABW$  model. As outlined above, we merge the edge connected components of the  $AB$  and  $AW-BW$  interfaces. We therefore report the number of cc and scc before and after the merge process, denoted  $\#cc_{bm}$ ,  $\#scc_{bm}$ ,  $\#cc_{am}$ ,  $\#scc_{am}$ . By definition, the number of cc and scc in the  $ABW$  model before the merge process corresponds to  $AB$  connected components only —i.e. components of the  $AW-BW$  interface are ignored.

**On the relative numbers of scc.** To fully understand the relationship between the number of connected components and structural water, it is instructive to compare the numbers of scc in the  $AB$  and  $ABW$  models.

To begin with, notice that  $\#cc_{bm} \geq \#cc$  since water either create new components or split components already present. One also expect  $\#cc \geq \#cc_{am}$  —if the number of cc of the  $AB$  interface split and/or reconnected by the merge process is larger than the number of cc of the  $AW-BW$  interface created. In general, these relationships do not translate onto their counterparts on significant cc. For example, one cannot expect  $\#scc_{bm} \geq \#scc$  since water molecules may have split significant cc into non significant cc. However, for all complexes of resolution better than  $2\text{\AA}$  investigated in  $[\text{CP}^+]$ , one has:

$$\#scc_{bm} \geq \#scc \geq \#scc_{am}. \quad (6)$$

Moreover, enumerating the type ( $=, >$ ) of the two binary operators, one gets the following discussion:

- $\#scc_{bm} = \#scc = \#scc_{am}$ . Such complexes have several well separated patches, which are not merged upon consideration of structural water. For high resolution structures (resolution better than  $2\text{\AA}$ ), we did not observe this configuration, because the following one takes over.
- $\#scc_{bm} = \#scc > \#scc_{am}$ . The number of scc in the  $AB$  model equals that in the  $ABW$  model before the merge. In other words, there are significant gaps which are not filled if water molecules are not there, but which are reconnected by water molecules. Example such complexes are 1dan, 1wej, 1osp, 2pcc, 1gg2, 1l0y, 1ycs.
- $\#scc_{bm} > \#scc = \#scc_{am}$ . The number of cc in the  $AB$  equals that in the  $ABW$  model after the merge. In other words, water molecules fill small gaps which are already filled in the  $AB$  model. Example such complexes are 1cse, 1tx4, 1hwg.

- $\#scc_{bm} > \#scc > \#scc_{am}$ . The number of  $scc$  in the  $AB$  is in-between the numbers before and after the merge process. This situation is a mix of the two previous ones. An example such complex is 1tco.

**Hydration patterns.** A puzzling question about structural water is the way molecules patch together at interfaces [RBCJ05]. But the  $cc$  of the  $AW-BW$  interface provide clusters of water molecules. To investigate the way water molecules are distributed at the interface, we therefore report  $\#ccw$  ( $\#sccw$ ), the number of (significant)  $cc$  of the  $AW-BW$  interface. These numbers quantify the propensity of water molecules to form networks. But the number  $\#sccw$  can be misleading since isolated water molecules yield as many significant patches. Given a threshold  $t_w$ , e.g. set to two, we therefore report the fraction  $p_w$  of water molecules participating to a  $cc$  of the  $AW-BW$  interface containing at most  $t_w$  water molecules. Low values of  $p_w$  indicate that water molecules tend to form (a) large network(s).

## 6 Intervor: overview

The constructions described in this paper have been implemented in C++ using the Computational Geometry Algorithms Library, [www.cgal.org](http://www.cgal.org), and more precisely the **Alpha.shape.3** package which provides the only public/robust/efficient implementation for  $\alpha$ -shapes and  $\alpha$ -complexes. The corresponding software, called **Intervor**, can be run from the command line, as well as from the following web site [bombyx.inria.fr/Intervor/intervor.html](http://bombyx.inria.fr/Intervor/intervor.html). Results can also be visualized under VMD —Visual Molecular Dynamics. The input consists of: a PDB file featuring a complex; a description of the partners in the complex; a boolean stating whether water molecules are kept or discarded; an optional threshold on the temperature factor to filter out unstable water molecules; an optional threshold if one wishes to filter out large Voronoi facets by Eq. (1). A typical run on a PDB file featuring between 1,000 and 10,000 atoms requires less than 5 seconds on a Pentium processor at 2 GHz.

The classes of atomic annotations used by **Intervor**, the files generated by **Intervor** and the chemical annotations of atoms and pairs, are respectively developed in sections 9, 10 and 11.

## 7 Conclusion

This paper presents an interface model which unifies most (if not all) the statistics used so far to describe interfaces in macro-molecular complexes, together with a software tool, **Intervor**, constructing interfaces from a PDB file. A key property of this interface model is to use a unique construction to evaluate all the parameters reported, thus providing a coherent view of a complex from the atomic to the interface scale, in the geometric, topological, and bio-chemical senses.

We hope this interface model and the associated tool will help structural biologists to address questions such as the role played by interface atoms and in particular structural

water, the flexibility issues involved in complex formation, or the specificity of recognition mechanisms.

**Acknowledgments.** J. Janin is acknowledged for discussions and for providing feedback of the the files generated. M. Yvinec is acknowledged for polishing the CGAL `Alpha_shape_3` package. F. Proust is acknowledged for helping with VMD.

## References

- [AT97] R. Abagyan and M. Totrov. Contact area difference (cad): A robust measure to evaluate accuracy of protein models. *J. Mol. Biol.*, 268, 1997.
- [Aur87] F. Aurenhammer. Power diagrams: properties, algorithms and applications. *SIAM J. Comput.*, 16:78–96, 1987.
- [BCRJ04] R.P. Bahadur, P. Chakrabarti, F. Rodier, and J. Janin. A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Bio.*, 336, 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.
- [CJ75] C. Chotia and J. Janin. Principles of protein-protein recognition. *Nature*, 256:705–708, 1975.
- [CJ99] L. Lo Conte and J. Janin. The atomic structure of protein-protein recognition sites. *J. Mol. Bio.*, 185, 1999.
- [CJ02] P. Chakrabarti and J. Janin. Dissecting protein-protein recognition sites. *Proteins*, 47, 2002.
- [CMJW03] R. Chen, J. Mintseris, J. Janin, and Zhiping Weng. A protein-protein docking benchmark. *Proteins*, 52:88–91, 2003.
- [CP<sup>+</sup>] F. Cazals, F. Proust, , R. Bahadur, and J. Janin. Revisiting the voronoi description of protein-protein interfaces. Submitted to *J. of Molecular Biology*.
- [CP05] F. Cazals and F. Proust. Revisiting the description of protein-protein interfaces. part i: algorithms. Technical Report 5346, INRIA, 2005. Submitted.
- [Ede92] H. Edelsbrunner. Weighted alpha shapes. Technical Report UIUCDCS-R-92-1760, Dept. Comput. Sci., Univ. Illinois, Urbana, IL, 1992.
- [EFL98] H. Edelsbrunner, M. Facello, and J. Liang. On the definition and the construction of pockets in macromolecules. *Discrete Appl. Math.*, 88:83–102, 1998.
- [EM94] H. Edelsbrunner and E. P. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graph.*, 13(1):43–72, January 1994.

- 
- [GW96] R.R. Gabdouliline and R.C. Wade. Analytically defined surfaces to analyze molecular interactions proeprties. *J. Molecular Graphics*, 14, 1996.
- [ME02] B.J. McConkey and V. Sobolev M. Edelman. Quantification of protein surfaces, volumes and atom-atom contacts using a constrained voronoi procedure. *Bioinformatics*, 18:1365–1373, 2002.
- [PHT00] H. Ponstingl, K. Henrick, and J.M. Thornton. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins Struct. Funct. Genet.*, 41, 2000.
- [RBCJ05] F. Rodier, R.P. Bahadur, P. Chakrabarti, and J. Janin. Hydration of protein - protein interfaces. *Proteins*, To Appear, 2005.
- [Ric74] F. M. Richards. The interpretation of protein structures: Total volume, group volume distributions and packing density. *Journal of Molecular Biology*, 82:1–14, 1974.

## 8 Illustrations

In this section, we present selected illustrations of interfaces, and selected insights developed from this interface model. The colors of the Voronoi interface facets depend on the chemical properties of the two atoms involved, and are not discussed here.

### 8.1 The 3 interfaces: $AB$ , $AW-BW$ , $ABW$

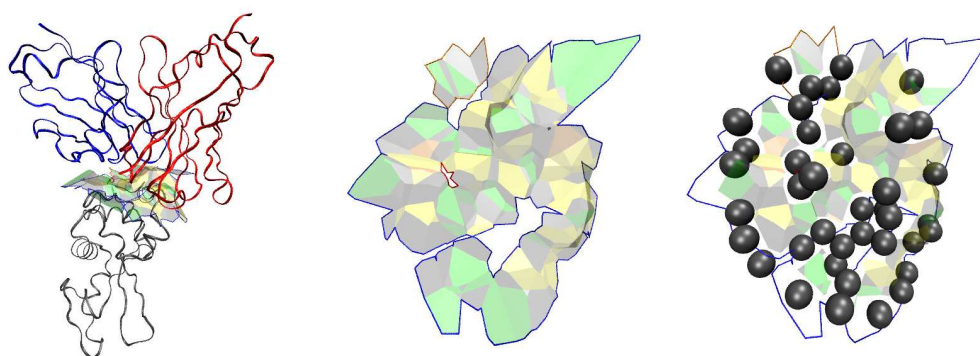


Figure 7: (a)Side view of the 1vf complex, with chains drawn as ribbons: grey: Lysozyme; blue-red: Fv fragments of mouse monoclonal antibody D1.3 (b)Top view of the AB interface in the AB model (c)The boundary of the  $AB$  interface in the AB model, the  $AB$  interface (without boundary) and the structural water molecules in the  $ABW$  model.

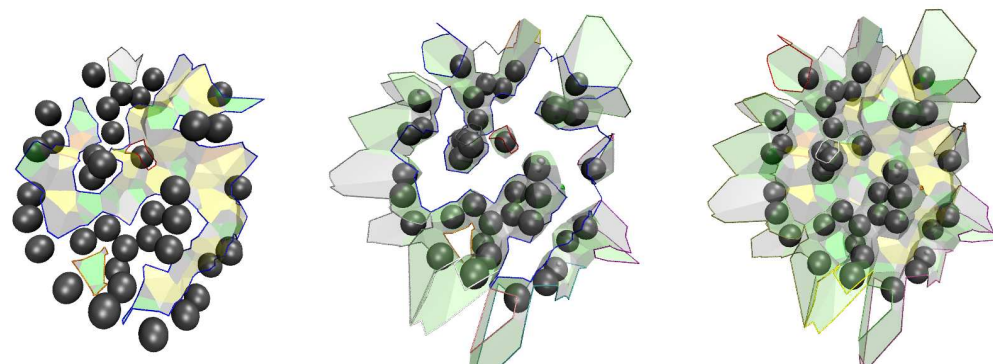


Figure 8: Complex 1vfb,  $ABW$  model (a)The  $AB$  interface and the water molecules (b)The  $AW-BW$  interface (c)The  $ABW$  interface. Notice common boundaries have been merged.

## 8.2 Geometry and topology of interfaces

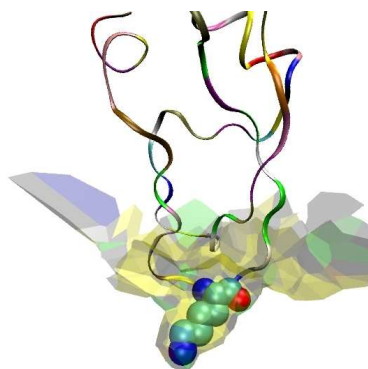


Figure 9: Interface of complex Kallikrein - pancreatic trypsin inhibitor (2kai) in the *AB* model. The inhibitor is drawn as a ribbon, and its Lys 15 drawn in VdW mode. The concave active site of the protease accommodates Lys 15 of the inhibitor. The cylindrical geometry of the active site is reflected by a high *discrete mean curvature*. Equivalently, an atom of the inhibitor has more than twice more neighbors than an atom of the protease.

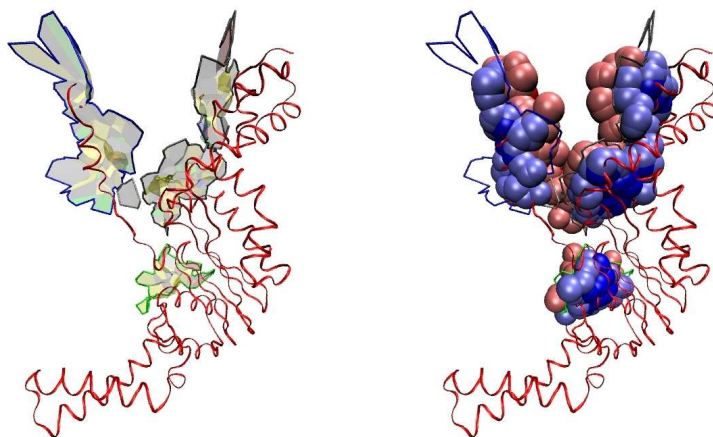


Figure 10: Interface patches of complex E. Coli EF-Tu/Ts (1efu) in the *AB* model. EF-Ts is shown as a ribbon (a)The interface splits into three main connected components whose boundaries are depicted in green, blue, brown (b)Interface atoms for each connected component. The blue/red atoms respectively correspond to EF-Tu and EF-Ts.

### 8.3 Structural water

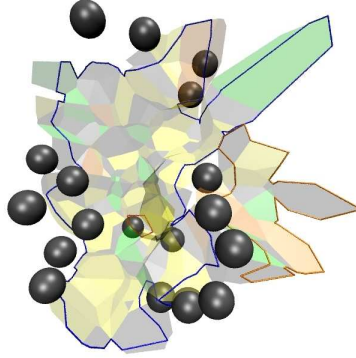


Figure 11: Complex Chymotrypsin-eglin (1acb): interfaces in the  $AB$  and  $ABW$  models. The grey balls represent crystallographic water in the  $ABW$  model. (a) In the  $AB$  model, the interface, whose boundary is not depicted, is a topological disk (b) In the  $ABW$  model, this disk is split into two parts: the topological disk with brown boundary; the (non simply connected) region with blue boundary, with one a hole punched by a water molecule on the left —small brown boundary.

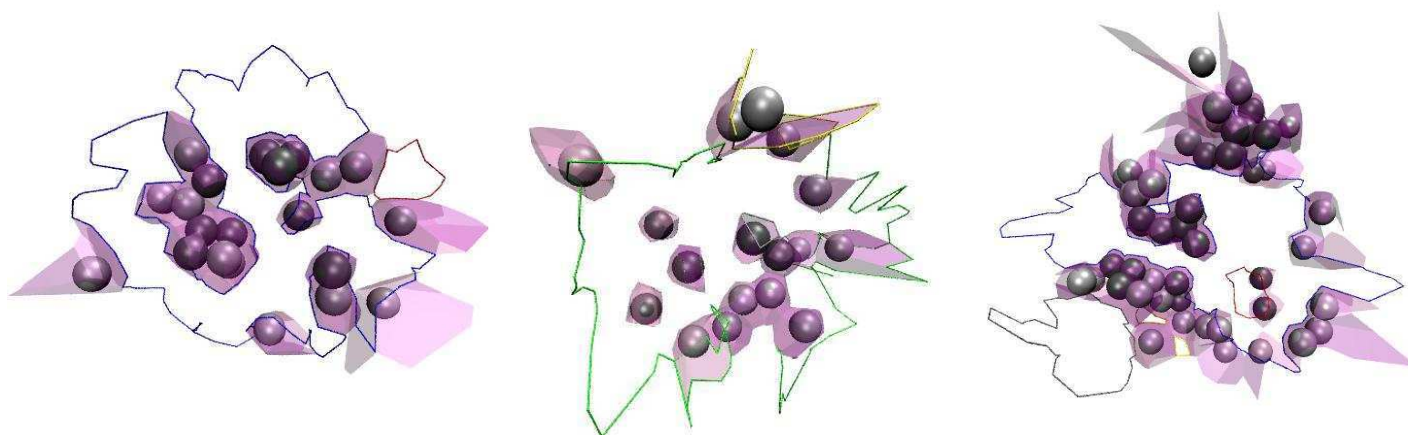


Figure 12: Different hydration patterns, where the facets represent the  $AW-BW$  interface, grey balls represent crystallographic water, solid lines represent the boundaries of the connected components of the  $AB$  interface in the  $ABW$  model. (a)Complex 1vfb. Antibody Antigen complexes (usually) feature creeks filled by water molecules near the CDR regions (b)Complex 1ppe. Protease-Inhibitor complexes (usually) feature water molecules isolated in the *middle* of the interface (c)Complex 1tx4. Signal Transduction complexes (usually) feature large clusters of water molecules re-connecting disconnected patches of the interface.

## 8.4 Atoms and neighbors

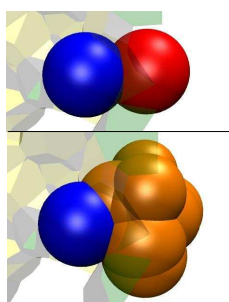


Figure 13: Buried atom making a contact: the red atom, although buried by atoms of its subunit (the orange atoms), makes a contact with the blue atom.



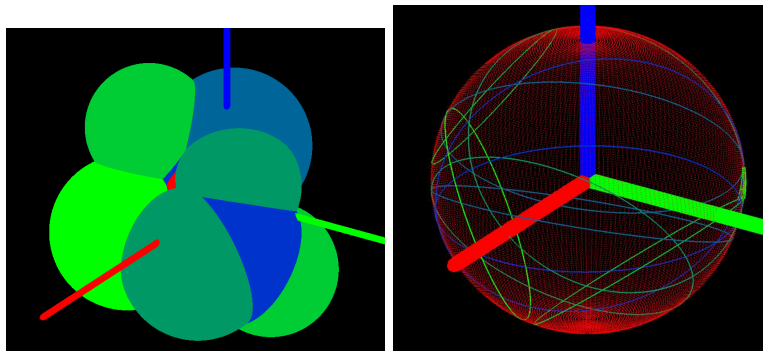


Figure 14: A  $C_\alpha$  atom and its 9 neighbors: the arrangement of the 9 intersection circles subdivides the Solvent Accessible Surface of the atom into 54 regions, each covered by a number  $n$  of neighbors with  $n \in 0..5$

## 8.5 Highlights

In this section, we briefly highlight some facts raised by the experimental study of 96 protein-protein interface carried out in [CP<sup>+</sup>]. The data set splits into five families of complexes, *Protein - Inhibitor*, *Enzyme - Substrate/Inhibitor*, *Antigen - Antibody*, *Signal Transduction*, *Misc*. Out of these 96 complexes, 30 are at resolution equal or better than 2Å, and constitute the 2Å set used to investigate the *ABW* model.

**Interface atoms.** In the *AB* model, the number of interface atoms specified by our definition increases in average of 13% wrt atoms loosing solvent accessibility —that is atoms with Buried Surface Area > 0. Main chain atoms, which account for 19% of the BSA represent 39% of our interface atoms, and represent about 2/3 of interface atoms that do not loose solvent accessibility. This suggests main chain atoms play a role more important than anticipated so far in stabilizing protein-protein complexes.

**Connected components.** In the *AB* model, the number of connected components varies in the range 1..6 with an average of 1.9 by complex. Discarding all cc contributing less than 7.5% of the Voronoi Interface Surface Area yields an average of 1.21 scc. Out of the 96 complexes, 81 have a unique scc, 9 have two, and 6 have three. Out of these 15 complexes, 7 are found in the Signal Transduction family. This later family features larger and more fragmented interfaces.

**Curvature in the *AB* model.** The distribution of dihedral angles observed for interior edges of the *AB* interface follows a bimodal distribution. In particular, the mode in between 12-24 degrees corresponds to covalently bonded atoms. At the interface scale, Protease-

Inhibitor complexes have a curvature larger than other families of complexes, a signature of their active site which has a cylindrical like geometry.

A property correlating with the curvature is the following: an atom in the convex region (e.g. an inhibitor) has up to twice more neighbors than an atom in the concave region (e.g. a protease).

Equivalently, an atom of the inhibitor has more than twice more neighbors than an atom of the protease.

**Structural water.** The following remarks are concerned with the  $2\text{\AA}$  set. For such complexes, consideration of interface water yields on average an increase of 45% of interface atoms, and the ratio of buried atoms shifts from 38% to 68%.

With an average of 6,6 cc, the *AW-BW* interface is more fragmented than the *AB* interface. The average number of water molecules by connected component of the *AW-BW* interface for Antibody-Antigen complexes is 2.82 times that of Protease-Inhibitor complexes, a feature corresponding to the creeks of water molecules found near the Complementarity Determining Regions of the antibody. Finally, in the *ABW* model, all interfaces of the  $2\text{\AA}$  set have a unique cc, which shows the cc of the *AW-BW* actually reconnect the patches of the *AB* interface. Phrased differently, these statistics also quantify the role of water to fill packing defects.

**Pairwise contacts.** In both *AB* and *ABW* models, tagging atoms as Hydrophobic (H) or Polar (P), the ratio of HH, HP, and PP interface edges is roughly the random distribution *given* the ratio of H and P atoms. Thus, pairwise contact fail to provide a detailed view of chemical interactions, which calls for the development of more elaborate models.

## 9 Classes of atomic annotations used by Intervor

In this section, we present the several annotations schemes used by **Intervor**. All of them are recorded in the *annotations file*, presented in section 11. **Intervor** is developed in C++, and requires a .C and .h files generated by a perl script from this annotations file. Thus, a version of **Intervor** meeting special needs is easily recompiled for any annotations file.

### 9.1 Specifying partners in a complex

We aim at studying *binary* complexes, either Protein-Protein complexes, or Protein-Ligand (typically protein-drug) complexes. We assume each partner in the complex is identified by a *label* which distinguishes its atoms. Labels in both cases are specified as follows:

- Label for a Protein-Protein complex. The label of a protein consists of the collection of its chain labels.
- Label for a Protein-Ligand complex. The label of the ligand consists of its residue id. (One cannot use the residue name since some ligands may be amino-acids.)

### 9.2 Atoms and their annotations

**Annotation as  $A, B, W, X$ .** Any atom or hetatm belonging to a chain or residue specified as a partner of the complex is tagged as  $A$  or  $B$ . Water oxygen are tagged as  $W$ . Other atoms or hetatoms are tagged as  $X$ .

**Chemical elements.** In a PDB file, atoms are distinguished from hetatoms (hetero-atoms) from the record name, namely ATOM vs HETATM. For each atom or hetatom, we need the *element symbol*, specified by the PDB format as 15th entry on any ATOM or HETATM record. Unfortunately, some files do not feature it, and others (old files) spoil this field with non relevant information. To get around these difficulties, we retrieve the element symbol from the first two characters of the atom name —see [http://www.rcsb.org/pdb/docs/format/pdbguide2.2/part\\_76.html](http://www.rcsb.org/pdb/docs/format/pdbguide2.2/part_76.html). If the first character is a digit, the element is an hydrogen. If not, these two characters specify the element symbol in a right-justified manner. Finally, oxygen atoms are identified from the residue name HOH.

**Chemical annotations for atoms.** As specified by the entry ATOMS ANNOTATIONS in section 11, we annotate atoms with one of the following 12 entries: Cali, Caro, Cpep, Nhbd, Naro, NchP, Ohbd, Opep, OchM, Owat, Sh, Unk.

To assign such an annotation, we proceed as follows. If the atom belongs to a standard amino-acid, the corresponding type is retrieved. If not, if the atom is a water oxygen, we label it as Ohdb. If not, the type is retrieved from the element symbol, and the type attributed is the generic one of that element —see the unknown or UNK residue in section

11. Each chemical annotation comes with a radius, which is a so-called *group radius* since hydrogens atoms are discarded. These radii, from [CJ75], can be found in section 11.

**Chemical annotations for edges.** Interactions between pairs of annotated atoms are classified into the following nine categories listed in the section PAIRS ANNOTATIONS in section 11, while the correspondence between the pairs of atoms and the annotation can be found in ANNOTATED PAIRS MATRIX section. The nine categories are (in parenthesis, the color used for the corresponding Voronoi facet in VMD):

- 1. ali (yellow): aliphatic
- 2. aro (orange): aromatic
- 3. elf (red): electrostatic favourable i.e. charges have opposite signs
- 4. elu (blue): electrostatic un-favourable
- 5. ssb (purple): sulfur-sulfur
- 6. hbd (green): H-bond
- 7. hbw (pale green): H-bond involving an oxygen of a crystallographic water molecule
- 8. unk (dark grey): unknown
- 9. unkw (light grey): unknown involving an oxygen of a crystallographic water molecule.

## 10 Files generated by Intervor

### 10.1 Conventions

**Intervor** generates files containing informations about interface atoms and their contacts. When dumping informations involving an atom, we use the plain PDB format i.e. the 16 standard fields, or the reduced PDB format:

**Definition. 1** *The reduced PDB record of an atom is the record featuring the serial number, the PDB atom name, the residue name and chain id.*

### 10.2 Overview of files generated

When called on a file `fprefix.pdb`, **Intervor** generates several text files. These files are self-contained, whence some redundant informations, and decompose into three groups.

#### 10.2.1 Annotations

The following files are meant to make sure all atoms are annotated correctly:

`fprefix.Chem.annotations-IV.txt` The atoms which are not standard atoms from the standard amino-acids. In particular, one finds water oxygens and hetatms. Each atom is dumped in PDB format, together with its chemical annotation.

`fprefix.ABW.annotations-IV.txt` All atoms tagged as *X*, as well as all hetatoms tagged as *A*, *B* or *W*. Atoms are dumped in the PDB format.

`fprefix.atoms.skipped-IV.txt` All atoms which are not inserted into the Delaunay triangulation. Hydrogens atoms are always discarded. In the *AB* model, all water molecules (i.e. oxygens) are discarded. In the *ABW* model, water molecules whose temperature factor is beyond a threshold are discarded —default threshold value is  $80\text{\AA}^2$ . Atoms discarded are dumped in PDB format.

#### 10.2.2 Results

`fprefix.log-IV.txt` This is the only file containing high level informations on the complex and its interfaces, and it should be consulted first. These informations are complemented by the other files. (When **Intervor** is called from the command line, the informations found in `pdbid.log.txt` are dumped in the shell from which the program is called.) See details below.

`fprefix.iar.pdb` The interface atoms. See details below.

`fprefix.iar-IV.txt` The interface atoms in PDB format, for visualisation softwares. The atom serial number reported is that of the original file. (Notice though some programs require atoms to be listed in sequential order.)

**fprefix.contacts-IV.txt** The pairwise contacts. See details below.

**fprefix.interface-IV.txt** Informations on the Voronoi interfaces: its geometry and topology, as well dissection in terms of interacting pairs. See details below.

**fprefix.patches-IV.txt** The distribution of interface atoms in the connected components of the interface. See details below.

### 10.2.3 Visualization

From the result files, a perl script generates a VMD file **fprefix.vmd** meant for visualization. See details below.

## 10.3 Global informations: file **fprefix.log-IV.txt**

This file contains high level informations, and should be read first. The informations provided split into three groups.

### 10.3.1 For any complex

**Header.** The header lists general informations about the complex:

- number of (finite) simplices of each dimension in the Delaunay triangulation underlying the construction,
- number of water molecules in the PDB file, and number of *stable* such molecules. A water molecule is termed stable if its temperature factor is below a threshold, set to  $80\text{\AA}^2$  by default.
- the number of stable water molecules in the complex and at the interface,
- the number of atoms tagged  $X$  in the complex and at the interface,
- the number of atoms and interface atoms of the four types  $A, B, W, X$ ,
- the number of interface edges ( $AB, AW, BW$ ) respectively on the convex hull, not on the convex hull but beyond the threshold  $M$  used in Eq. (1).

**$AB$  interface.** Informations on the  $AB$  interface decompose into three groups. First, each connected components (cc) is described by:

- the type of edge the cc was started with, the number of {facets, boundary loops, internal edges, boundary edges},
- the cc surface area,
- the number of atoms of each type — $A, B, W, X$ ,
- the number of edges of each boundary loop.

Second, we get four lines summarizing interface properties:

- the number of connected components of the interface,
- a summary of global statistics for all connected components of the interface: number of {cc, faces, boundary edges, loops}.
- the surface area information.

Third, we get the curvature information  $sAbsH$  and  $sH$  defined by Eqs (4) and (5), together with the Voronoi Interface Surface Area, denoted  $sA$ , and the length of the boundary loops, denoted  $sL$ . The last two lines provide the average and standard deviation of the dihedral angle  $\beta(e)$  and its absolute value, in radians.

### 10.3.2 Complexes with interface water

**$AW-BW$  interface.** If water molecules have been found at the interface, we first report for the  $AW-BW$  interface all the informations just outlined for the  $AB$  interface.

**$ABW$  interface and the Union-Find process.** The  $ABW$  interface is defined from the merge of the  $AB$  and  $AW-BW$  interfaces, using the so-called *Union-Find algorithm*. The elementary merge operation consists of merging two connected components along a common Voronoi edge. This edge may belong to two boundary loops of two connected components of the  $AB$  and  $AW-BW$  interfaces, or may involve one boundary and one interior Voronoi edges.

Statistics are reported for each connected component of the merge. First, one finds informations on the Union-Find process itself. These informations provide the number of interior and boundary Voronoi edges involved in the process, together with the number of components and loops after the merge. Second, each component is detailed. We indicate for each boundary net the number of loops involved. Finally, we provide the surface area information for the merged connected.

## 10.4 Interface atoms: file `fprefix_iar.pdb`

First, one finds informations for interface atoms. Each line features the atom reduced PDB format, together with the  $A, B, W, X$  label; a boolean stating whether the atoms is buried (true) or exposed (false), the chemical annotation; and the number of neighbours of each type.

Second, for the bicolor interface  $XY \in \{AB, AW, BW, WW\}$ , one encounters a summary of connectivity informations: the average number of neighbors of the party type, the numbers of atoms of both types, the number of edges of the bicolor interface.

## 10.5 Interface pairs: file `fprefix_contacts-IV.txt`

For each  $XY$  interface, with  $X$  and  $Y \in \{A, B, W, X\}$ , assuming  $X < Y$  according to the lexicographic order, we dump for each interface pair the two atoms in reduced PDB format, together with the annotation of the interaction. Since the file `fprefix_contacts-IV.txt` contains all pairs of contacts, notice one can easily infer all the neighbours of a given atom.

## 10.6 Interface : file `fprefix.interface-IV.txt`

An interface consists of Voronoi facets —each being the dual of an interface edge, and is bounded by loops and nets respectively before and after the merge of the *AB* and *AW-BW* interfaces.

### 10.6.1 Informations before the merge process

For the manifold interfaces *AB* and *AW-BW*, the following informations are dumped.

**Voronoi facet.** For an interface Voronoi facet, one first encounters the two atoms involved —in reduced PDB format, together with the annotation of the interaction. Next, one finds the *xyz* coordinates of the atoms. Finally, the list of coordinates of Voronoi centers bounding the facet is provided.

**Boundary loops of a connected component.** Once all facets of the interface have been dumped, we get for each connected component the description of the boundary loops. Each loop is described as follows:

- the loop id (an integer between 1 and the total number of loops created for all interfaces), and the bicolor edge associated to the creation of the connected component this loop belongs to;
- a list of Voronoi vertices bounding the loop;
- two positive real numbers giving the sum of the total length of the loop, and the sum of the lengths of boundary edges. Before the merge process, all edges of a loop are boundary edges, so that the two values agree.

**VIA.** Two dissections of the Voronoi Interface Area are provided: in terms of bicolor interfaces *AB*, *AW*, *BW*; and in terms of the annotations of pairs of atoms.

### 10.6.2 Informations after the merge process

After the merge process, we reconsider for each connected component the structure of loops and nets, and the dissection of the surface area of the merged components.

**Boundary nets after Union-Find.** During the merge process, all edges common to two components get identified, so that the length of the boundary edges of the net is less than its total length. These two values are reported. (Notice a merge occurs in two situations: the two edges are boundary edges; or a boundary edge merges with an edge in the interior of a component.)

**VIA.** The two dissections mentioned before the merge process are reported.

## 10.7 Atoms and connected components; File `fprefix.patches-IV.txt`

In the *AB* model, for each cc of the *AB* interface, all atoms involved in that cc are listed. Atoms involved in several cc are also reported. The same informations are reported after the merge step in the *ABW* model.



## 10.8 Visualization file: `pdbid_log.vmd`

A perl script, `vor2vmd.pl`, can be invoked on the text files described above to generate a VMD file containing the following informations:

- the interface atoms. Atoms of type A (B)(W) are colored in light/deep blue (pink/red) (light/dark grey) depending on whether they are accessible or buried in the complex. These atoms are represented in two venues, namely with their Van der Waals radii, or expanded by the water probe.
- the bicolor interface  $AB$  and the interface  $AW-BW$ . Facets of the bicolor interface follow the color conventions listed in section 9.
- the boundary loops of each interface ( $AB$ ,  $AW-BW$ ) before the merge step, and the boundary nets of the  $ABW$  interface. Different loops and nets are identified by different colors, but these color do not have any (chemical) meaning.

To ease the visual inspection of interfaces, all the objects mentioned above are independent VMD object and can be selected independently from the VMD main.

## 11 Annotations of atoms and pairs

```
//types and radii
BEGIN ATOMS ANNOTATIONS
Cali 1.87
Caro 1.76
Cpep 1.76
Nhbd 1.65
Naro 1.65
NchP 1.50
Ohbd 1.4
Opep 1.4
OchM 1.4
Owat 1.4
Sh 1.85
Unk 2.0
END ATOMS ANNOTATIONS

BEGIN PAIRS ANNOTATIONS
ali
aro
hbd
elf
elu
ssb
unk
hbw
unkw
END PAIRS ANNOTATIONS

BEGIN ATOM ANNOTATIONS
RESIDUE UNK
ATOM " C " Cali
ATOM " O " Ohbd
ATOM " N " Nhbd
ATOM " S " Sh
ATOM " P " Unk
ATOM " F " Unk
ATOM " X " Unk
END

RESIDUE GLY
ATOM " C " Cpep
ATOM " CA " Cali
ATOM " O " Opep

ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE ALA
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE LEU
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " CD1" Cali
ATOM " CD2" Cali
ATOM " OXT" Ohbd
END

RESIDUE ILE
ATOM " C " Cpep
ATOM " N " Nhbd
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " O " Opep
ATOM " CD " Cali
ATOM " CD1" Cali
ATOM " CG1" Cali
ATOM " CG2" Cali
ATOM " OXT" Ohbd
END

RESIDUE VAL
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " O " Opep

ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

ATOM " N " Nhbd
ATOM " CG1" Cali
ATOM " CG2" Cali
ATOM " OXT" Ohbd
END

RESIDUE ASP
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Caro
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " OD1" OchM
ATOM " OD2" OchM
ATOM " OXT" Ohbd
END

RESIDUE GLU
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " O " Opep
ATOM " CD " Caro
ATOM " N " Nhbd
ATOM " OE2" OchM
ATOM " OE1" OchM
ATOM " OXT" Ohbd
END

RESIDUE ASN
ATOM " C " Cpep
ATOM " ND2" Nhbd
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Caro
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " OD1" Ohbd
ATOM " OXT" Ohbd
END
```

```

RESIDUE GLN
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " O " Opep
ATOM " CD " Caro
ATOM " N " Nhbd
ATOM " NE2" Nhbd
ATOM " OE1" Ohbd
ATOM " OXT" Ohbd
END

RESIDUE HIS
ATOM " C " Cpep
ATOM " CD2" Caro
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Caro
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " CE1" Caro
ATOM " ND1" Naro
ATOM " NE2" Naro
ATOM " OXT" Ohbd
END

RESIDUE HSE
ATOM " C " Cpep
ATOM " CD2" Caro
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Caro
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " CE1" Caro
ATOM " ND1" Naro
ATOM " NE2" Naro
ATOM " OXT" Ohbd
END

RESIDUE SER
ATOM " C " Cpep
ATOM " OG " Ohbd
ATOM " CB " Cali
ATOM " CA " Cali

ATOM " O " Opep
ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE THR
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " OG1" Ohbd
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " CG2" Cali
ATOM " OXT" Ohbd
END

RESIDUE CYS
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " SG " Sh
ATOM " OXT" Ohbd
END

RESIDUE MET
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " CE " Cali
ATOM " N " Nhbd
ATOM " SD " Sh
ATOM " O " Opep
ATOM " OXT" Ohbd
END

RESIDUE MSE
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " CE " Cali
ATOM " N " Nhbd
ATOM " SD " Sh

ATOM " O " Opep
ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE PRO
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " O " Opep
ATOM " CD " Cali
ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE ARG
ATOM " NE " NchP
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " O " Opep
ATOM " CD " Cali
ATOM " CZ " Caro
ATOM " NH1" NchP
ATOM " NH2" NchP
ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE LYS
ATOM " NZ " NchP
ATOM " C " Cpep
ATOM " CB " Cali
ATOM " CA " Cali
ATOM " CG " Cali
ATOM " CE " Cali
ATOM " CD " Cali
ATOM " O " Opep
ATOM " N " Nhbd
ATOM " OXT" Ohbd
END

RESIDUE PHE
ATOM " O " Opep
ATOM " OXT" Ohbd
ATOM "SE " Sh
END

```

ATOM " C " Cpep	ATOM " OXT" Ohbd	Nhbd NchP hbd
ATOM " CD2" Caro	END	Nhbd Ohbd hbd
ATOM " CB " Cali		Nhbd Opep hbd
ATOM " CA " Cali	RESIDUE HOH	Nhbd OchM hbd
ATOM " CG " Caro	ATOM " O " Owat	Nhbd Owat hbw
ATOM " O " Opep	END	Nhbd Sh hbd
ATOM " CZ " Caro		Nhbd Unk unk
ATOM " N " Nhbd	END ATOM ANNOTATIONS	Naro Naro hbd
ATOM " CD1" Caro		Naro NchP hbd
ATOM " CE1" Caro	BEGIN ANNOTATED PAIRS MATRIX	Naro Ohbd hbd
ATOM " CE2" Caro	Cali Cali ali	Naro Opep hbd
ATOM " OXT" Ohbd	Cali Caro ali	Naro OchM hbd
END	Cali Cpep ali	Naro Owat hbw
	Cali Nhbd unk	Naro Sh hbd
RESIDUE TYR	Cali Naro unk	Naro Unk unk
ATOM " C " Cpep	Cali NchP unk	NchP NchP elu
ATOM " CD2" Caro	Cali Ohbd unk	NchP Ohbd hbd
ATOM " OH " Ohbd	Cali Opep unk	NchP Opep hbd
ATOM " N " Nhbd	Cali OchM unk	NchP OchM elf
ATOM " CB " Cali	Cali Owat unkw	NchP Owat hbw
ATOM " CA " Cali	Cali Sh unk	NchP Sh hbd
ATOM " CG " Caro	Cali Unk unk	NchP Unk unk
ATOM " O " Opep	Caro Caro aro	Ohbd Ohbd hbd
ATOM " CZ " Caro	Caro Cpep ali	Ohbd Opep hbd
ATOM " CD1" Caro	Caro Nhbd unk	Ohbd OchM hbd
ATOM " CE1" Caro	Caro Naro aro	Ohbd Owat hbw
ATOM " CE2" Caro	Caro NchP unk	Ohbd Sh hbd
ATOM " OXT" Ohbd	Caro Ohbd unk	Ohbd Unk unk
END	Caro Opep unk	Opep Opep unk
	Caro OchM unk	Opep OchM unk
RESIDUE TRP	Caro Owat unkw	Opep Owat hbw
ATOM " CZ2" Caro	Caro Sh unk	Opep Sh hbd
ATOM " CZ3" Caro	Caro Unk unk	Opep Unk unk
ATOM " CD1" Caro	Cpep Cpep ali	OchM OchM elu
ATOM " CD2" Caro	Cpep Nhbd unk	OchM Owat hbw
ATOM " CH2" Caro	Cpep Naro unk	OchM Sh hbd
ATOM " C " Cpep	Cpep NchP unk	OchM Unk unk
ATOM " CB " Cali	Cpep Ohbd unk	Owat Owat hbw
ATOM " CA " Cali	Cpep Opep unk	Owat Sh hbw
ATOM " CG " Caro	Cpep OchM unk	Owat Unk unk
ATOM " O " Opep	Cpep Owat unkw	Sh Sh ssb
ATOM " N " Nhbd	Cpep Sh unk	Sh Unk unk
ATOM " CE3" Caro	Cpep Unk unk	Unk Unk unk
ATOM " CE2" Caro	Nhbd Nhbd hbd	END ANNOTATED PAIRS MATRIX
ATOM " NE1" Naro	Nhbd Naro hbd	

## Contents

<b>1</b>	<b>Complexes and interfaces</b>	<b>3</b>
1.1	Previous work . . . . .	3
1.2	Contributions and paper overview . . . . .	3
<b>2</b>	<b>Voronoi diagram, <math>\alpha</math>-complex, <math>\alpha</math>-shape</b>	<b>4</b>
2.1	Delaunay - Voronoi . . . . .	4
2.2	The $\alpha$ -complex and the $\alpha$ -shape — union of balls . . . . .	5
2.3	Applications to molecular modeling . . . . .	6
<b>3</b>	<b>Crystals and structural water: the <math>AB</math> and <math>ABW</math> models</b>	<b>7</b>
<b>4</b>	<b>Bicolor and tricolor interfaces</b>	<b>8</b>
4.1	The $AB$ , $AW$ – $BW$ , $ABW$ interfaces . . . . .	8
4.2	Curvature of the $AB$ interface . . . . .	11
<b>5</b>	<b>Statistics</b>	<b>12</b>
5.1	Notations . . . . .	12
5.2	Connectivity properties of interface atoms . . . . .	12
5.3	The geometry of interfaces . . . . .	13
5.4	The topology of interfaces . . . . .	14
<b>6</b>	<b>Intervor: overview</b>	<b>15</b>
<b>7</b>	<b>Conclusion</b>	<b>15</b>
<b>8</b>	<b>Illustrations</b>	<b>18</b>
8.1	The 3 interfaces: $AB$ , $AW$ – $BW$ , $ABW$ . . . . .	18
8.2	Geometry and topology of interfaces . . . . .	19
8.3	Structural water . . . . .	20
8.4	Atoms and neighbors . . . . .	21
8.5	Highlights . . . . .	22
<b>9</b>	<b>Classes of atomic annotations used by Intervor</b>	<b>24</b>
9.1	Specifying partners in a complex . . . . .	24
9.2	Atoms and their annotations . . . . .	24
<b>10</b>	<b>Files generated by Intervor</b>	<b>26</b>
10.1	Conventions . . . . .	26
10.2	Overview of files generated . . . . .	26
10.2.1	Annotations . . . . .	26
10.2.2	Results . . . . .	26
10.2.3	Visualization . . . . .	27

---

10.3	Global informations: file <code>fprefix_log-IV.txt</code> . . . . .	27
10.3.1	For any complex . . . . .	27
10.3.2	Complexes with interface water . . . . .	28
10.4	Interface atoms: file <code>fprefix_iar.pdb</code> . . . . .	28
10.5	Interface pairs: file <code>fprefix_contacts-IV.txt</code> . . . . .	28
10.6	Interface : file <code>fprefix_interface-IV.txt</code> . . . . .	29
10.6.1	Informations before the merge process . . . . .	29
10.6.2	Informations after the merge process . . . . .	29
10.7	Atoms and connected components; File <code>fprefix_patches-IV.txt</code> . . . . .	29
10.8	Visualization file: <code>pdbid_log.vmd</code> . . . . .	30
11	Annotations of atoms and pairs . . . . .	31

## Contents



---

Unité de recherche INRIA Sophia Antipolis  
2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Futurs : Parc Club Orsay Université - ZAC des Vignes  
4, rue Jacques Monod - 91893 ORSAY Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique  
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

---

Éditeur  
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)  
<http://www.inria.fr>  
ISSN 0249-6399