INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team abs

Algorithms, Biology, Structure

Sophia Antipolis - Méditerranée

THEME BIO

Table of contents

1. Team

Head of project
Frédéric Cazals [DR2 Inria, HdR]

Administrative Assistant
Agnès Bessière [TR Inria, assistant of GEOMETRICA and ABS]

Research scientists
Julie Bernauer [CR2 Inria; from December 2007 on]

Ph.D. Students
Sébastien Loriot [MESR monitor fellow]

Post-doctoral fellow
Benjamin Bouvier [ARC REFLEXP; from November 2006 to April 2008]

Student interns
Sushant Sachdeva [IIT Bombay - India, May-July 2007]

2. Overall Objectives

2.1. Introduction

Computational Biology and Computational Structural Biology. Understanding the lineage between species and the genetic drift of genes and genomes, apprehending the control and feed-back loops governing the behavior of a cell, a tissue, an organ or a body, and inferring the relationship between the structure of biological (macro)-molecules and their functions are amongst the major challenges of modern biology. The investigation of these challenges is supported by three types of data: genomic data, transcription and expression data, and structural data.

Genetic data feature sequences of nucleotides on DNA and RNA molecules, and are symbolic data whose processing falls in the realm of Theoretical Computer Science: dynamic programming, algorithms on texts and strings, graph theory dedicated to phylogenetic problems. Transcription and expression data feature evolving concentrations of molecules (RNAs, proteins, metabolites) over time, and fit in the formalism of discrete and continuous dynamical systems, and of graph theory. The exploration and the modeling of these data are covered by a rapidly expanding research field termed *systems biology*. Structural data encode informations about the $3d$ structures of molecules (nucleic acids, proteins, small molecules) and their interactions, and come from three main sources: X ray crystallography, NMR spectroscopy, cryo Electron Microscopy. Ultimately, structural data should expand our understanding of how the structure accounts for the function of macro-molecules —one of the central question in structural biology. This goal actually subsumes two equally difficult challenges, which are *folding* —the process through which a protein adopts its $3d$ structure, and *docking* —the process through which two or several molecules assemble. Folding and docking are driven by non covalent interactions, and for complex systems, are actually inter-twined [38]. Apart from the bio-physical interests raised by these processes, two different application domains are concerned: in fundamental biology, one is primarily interested in understanding the machinery of the cell; in medicine, applications to drug design are developed.

Modeling in Computational Structural Biology. Acquiring structural data is not always possible: NMR is restricted to relatively small molecules; membrane proteins do not crystallize, etc. As a matter of fact, as of October 2007, about 1,000 genomes have been fully sequenced or are about to be so, while the Protein Data Bank contains (a mere) 40,000 structures. For these reasons, *molecular modeling* is expected to play a key role in investigating structural issues.

Ideally, bio-physical models of macro-molecules should resort to quantum mechanics. While this is possible for small systems, say up to 50 atoms, large systems are investigated within the framework of the Born-Oppenheimer approximation which stipulates the nuclei and the electron cloud can be decoupled. Example force fields developed in this realm are AMBER, CHARMM, OPLS. Of particular importance are Van der Waals models, where each atom is modeled by a sphere whose radius depends on the atom chemical type. From an historical perspective, Richards [36], [24] and later Connolly [20], while defining molecular surfaces and developing algorithms to compute them, established the connexions between molecular modeling and geometric constructions. Remarkably, a number of difficult problems (e.g. additively weighted Voronoi diagrams) were touched upon in these early days.

The models developed in this vein are instrumental in investigating the interactions of molecules for which no structural data is available. But such models often fall short from providing complete answers, which we illustrate with the folding problem. On one hand, as the conformations of side-chains belong to discrete sets (the so-called rotamers or rotational isomers) [27], the number of distinct conformations of a poly-peptidic chain is exponential in the number of amino-acids. On the other hand, Nature folds proteins withing time scales ranging from milliseconds to hours, which is out of reach for simulations. The fact that Nature avoids the exponential trap is known as Levinthal's paradox. The intrinsic difficulty of problems calls for models exploiting several classes of informations. For small systems, *ab initio* models can be built from first principles. But for more complex systems, *homology* or template-based models integrating a variable amount of knowledge acquired on similar systems are resorted to.

The variety of approaches developed are illustrated by the two community wide experiments CASP (*Critical Assessment of Techniques for Protein Structure Prediction*; http://predictioncenter.org) and CAPRI (*Critical Assessment of Prediction of Interactions*; http://capri.ebi.ac.uk), which allow models and prediction algorithms to be compared to experimentally resolved structures.

As illustrated by the previous discussion, modeling macro-molecules touches upon biology, physics and chemistry, as well as mathematics and computer science. In the following, we present the topics investigated within ABS.

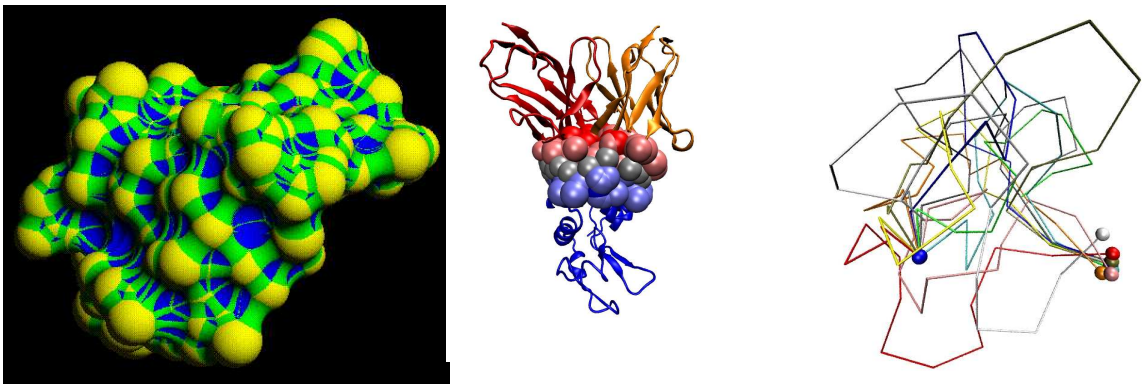

Figure 1. (a) Molecular surface (b) An antiboby-antigen complex, with interface atoms computed as described in [10] (c)Conformations of a backbone loop

2.2. Highlights of the year

The ABS team was created in July 2007. Julie Bernauer, postdoc with Michael Levitt in the Structural Biology Dpt at Stanford University, will take her position in December 2007 –Julie was recruited as CR2 in June 2007.

Following this post-doc position, a project funded by the France Stanford center for Inter-disciplinary studies got accepted, the principal investigators being F. Cazals, M. Levitt, and J. Bernauer.

3. Scientific Foundations

3.1. Introduction

The research conducted by ABS focuses on two main directions in Computational Structural Biology (CSB), each such direction calling for specific algorithmic developments. These direction are:
- Modeling interfaces and contacts,
- Modeling the flexibility of macro-molecules.

3.2. Modeling interfaces and contacts

Keywords: *Docking*, *Voronoi diagrams*, *arrangements of balls*, *interfaces*, *protein complexes*, *scoring functions*, *structural alphabets*.

Problems addressed. The Protein Data Bank, http://www.rcsb.org/pdb, contains the structural data which have been resolved experimentally. Most of the entries of the PDB feature isolated proteins [1], the remaining ones being protein - protein or protein - drug complexes. These structures feature what Nature does —up to the bias imposed by the experimental conditions inherent to structure elucidation, and are of special interest to investigate non-covalent contacts in biological complexes. More precisely, given two proteins defining a complex, interface atoms are defined as the atoms of one protein *interacting* with atoms of the second one. Understanding the structure of interfaces is central to understand biological complexes and thus the function of biological molecules [38]. Yet, in spite of almost three decades of investigations, the basic principles guiding the formation of interfaces and accounting for its stability are unknown [41]. Current investigations follows two routes. From the experimental perspective [23], directed mutagenesis allows one to quantify the energetic importance of residues, important residues being termed *hot* residues. Such studies recently evidenced the *modular* architecture of interfaces [35]. From the modeling perspective, the main issue consists of guessing the hot residues from sequence and/or structural informations [30].

The description of interfaces is also of special interest to improve *scoring functions*. By scoring function, two things are meant: either a function which assign to a complex a quantity homogeneous to a free energy change [2], or a function stating that a complex is more stable than another one, in which case the value returned is a score and not an energy. Borrowing to statistical mechanics [15], the usual way to design scoring functions is to mimic the so-called potentials of mean force. To put it briefly, one reverts Boltzmann's law, that is, denoting $p_i(r)$ the probability of two atoms –defining type i– to be located at distance r, the (free) energy assigned to the pair is computed as $E_i(r) = -kT \log p_i(r)$. Estimating from the PDB one function $p_i(r)$ for each type of pair of atoms, the energy of a complex is computed as the sum of the energies of the pairs located within a distance threshold [39], [26]. To compare the energy thus obtained to a reference state, one may compute $E = \sum_i p_i \log p_i/q_i$, with p_i the observed frequencies, and q_i the frequencies stemming from an a priori model [31]. In doing so, the energy defined is nothing but the Kullback-Leibler divergence between the distributions $\{p_i\}$ and $\{q_i\}$.

Methodological developments. Describing interfaces poses problems in two settings: static and dynamic.

[1] For structures resolved by crystallography, the PDB contains the asymmetric unit of the crystal. Determining the biological unit from the asymmetric unit is a problem in itself.

[2] The Gibbs free energy of a system is defined by $G = H - TS$, with $H = U + PV$. G is minimum at an equilibrium, and differences in G drive chemical reactions.

In the static setting, one seeks the minimalist geometric model providing a relevant bio-physical signal. A first step in doing so consists of identifying interface atoms, so as relate the geometry and the bio-chemistry at the interface level [10]. To elaborate at the atomic level, one seeks a structural alphabet encoding the spatial structure of proteins. At the side-chain and backbone level, an example such alphabet is that of [16]. At the atomic level and in spite of recent observations on the local structure of the neighborhood of a given atom [40], no such alphabet is known. Specific important local conformations are known, though. One of them is the so-called dehydron structure, which is an under-desolvated hydrogen bond —a property that can be directly inferred from the spatial configuration of the C_α carbons surrounding a hydrogen bond [22].

A structural alphabet at the atomic level may be seen as an alphabet featuring for an atom of a given type all the conformations this atom may engage into, depending on its neighbors. One way to tackle this problem consists of extending the notions of molecular surfaces used so far, so as to encode multi-body relations between an atom and its neighbors [11]. In order to derive such alphabets, the following two strategies are obvious. On one hand, one may use an encoding of neighborhoods based on geometric constructions such as Voronoï diagrams (affine or curved) or arrangements of balls. On the other hand, one may resort to clustering strategies in higher dimensional spaces, as the p neighbors of a given atom are represented by $3p-6$ degrees of freedom —the neighborhood being invariant upon rigid motions.

In the dynamic setting, one wishes to understand whether selected (hot) residues exhibit specific dynamic properties, so as to serve as anchors in a binding process [34]. More generally, any significant observation raised in the static setting deserves investigations in the dynamic setting, so as to assess its stability. Such questions are also related to the problem of correlated motions, which we discuss next.

3.3. Modeling the flexibility of macro-molecules

Keywords: *Folding*, *Morse theory*, *docking*, *energy landscapes*, *induced fit*, *molecular dynamics; point clouds*, *reconstruction*, *shape learning*.

Problems addressed. Proteins in vivo vibrate at various frequencies: high frequencies correspond to small amplitude deformations of chemical bonds, while low frequencies characterize more global deformations. This flexibility contributes to the entropy thus the `free energy` of the system *protein - solvent*. From the experimental standpoint, NMR studies and Molecular Dynamics simulations generate ensembles of conformations, called `conformers`. Of particular interest while investigating flexibility is the notion of correlated motion. Intuitively, when a protein is folded, all atomic movements must be correlated, a constraint which gets alleviated when the protein unfolds since the steric constraints get relaxed [3]. Understanding correlations is of special interest to predict the folding pathway that leads a protein towards its native state. A similar discussion holds for the case of partners within a complex, for example in the third step of the *diffusion - conformer selection - induced fit* complex formation model.

Parameterizing these correlated motions, describing the corresponding energy landscapes, as well as handling collections of conformations pose challenging algorithmic problems.

Methodological developments. At the side-chain level, the question of improving rotamer libraries is still of interest [21]. This question is essentially a clustering problem in the parameter space describing the side-chains conformations.

At the atomic level, flexibility is essentially investigated resorting to methods based on a classical potential energy (molecular dynamics), and (inverse) kinematics. A molecular dynamics simulation provides a point cloud sampling the conformational landscape of the molecular system investigated, as each step in the simulation corresponds to one point in the parameter space describing the system (the conformational space) [37]. The standard methodology to analyze such a point cloud consists of resorting to normal modes. Recently, though, more elaborate methods resorting to more local analysis [33], to Morse theory [28] and to analysis of meta-stable states of time series [29] have been proposed.

[3] Assuming local forces are prominent, which in turn subsumes electrostatic interactions are not prominent.

Given a sampling on an energy landscape, a number of fundamental issues actually arise: how does the point cloud describes the topography of the energy landscape (a question reminiscent from Morse theory)? can one infer the effective number of degrees of freedom of the system over the simulation, and is this number varying? Answers to these questions would be of major interest to refine our understanding of folding and docking, with applications to the prediction of structural properties. It should be noticed in passing such questions are probably related to modeling phase transitions in statistical physics where geometric and topological methods are being used [32].

From an algorithmic standpoint, such questions are reminiscent from *shape learning*. Given a collection of samples on an (unknown) *model*, *learning* consists of guessing the model from the samples —the result of this process may be called the *reconstruction*. In doing so, two types of guarantees are sought: topologically speaking, the reconstruction and the model should (ideally!) be isotopic; geometrically speaking, their Hausdorff distance should be small. Motivated by applications in CAGD, surface reconstruction triggered a major activity in the Computational Geometry community over the past five years [7]. Aside from applications, reconstruction raise a number of deep issues: the study of distance functions to the model and to the samples, and their comparison [17]; the study of Morse-like constructions stemming from distance functions to points [25]; the analysis of topological invariants of the model and the samples, and their comparison [18], [19].

Last but not the least, gaining insight on such questions would also help to effectively select a reduced set of conformations best representing a larger number of conformations. This selection problem is indeed faced by flexible docking algorithms that need to maintain and/or update collections of conformers for the second stage of the *diffusion - conformer selection - induced fit* complex formation model.

4. Software

4.1. Web services

4.1.1. Modeling macro-molecular interfaces

Participant: Frédéric Cazals.

We recently proposed an interface model of (macro-)molecular interfaces based upon power diagrams [10]. The corresponding software, *Intervor*, has been made available to the community from the web site http://cgal.inria.fr/Intervor. Our publication appeared in the September 2006 issue of *Protein Science*, and the server has been used about 600 times since then. To the best of our knowledge, this code is the only publicly available one for analyzing interfaces in complexes.

4.2. CGAL and Ipe

Participant: Sébastien Loriot.

In collaboration with L. Rineau and S. Pion, GEOMETRICA. *Work started by Nicolas Carrez, summer intern, 2005. http://www.cgal.org*

CGAL is a C++ library of geometric algorithms initially developed within two European projects (project ESPRIT IV LTR CGAL December 97 - June 98, project ESPRIT IV LTR GALIA november 99 - august 00) by a consortium of eight research teams from the following institutes: Universiteit Utrecht, Max-Planck Institut Saarbrücken, INRIA Sophia Antipolis, ETH Zürich, Tel Aviv University, Freie Universität Berlin, Universität Halle, RISC Linz. The goal of CGAL is to make the solutions offered by the computational geometry community available to the industrial world and applied domains.

The IPE editor, see http://tclab.kaist.ac.kr/ipe, is a graphical editor which combines XFIG like facilities together with standard Computational Geometry algorithms. It is intensively used by the computational geometry community for making presentations as well as illustrating papers.

Based on the $2D$ algorithms present in the CGAL library, we developed in C++ a set of plugins, so as to make the following algorithms available from IPE: triangulations (Delaunay, constrained Delaunay, regular) as well as their duals, a convex hull algorithm, polygon partitioning algorithms, polygon offset, arrangements of linear and degree two primitives. The first version was released on 08/13/2007. These plugins are available under the Open Source LGPL license, and are subject to the constraints of the underlying CGAL packages. They can be downloaded from http://cgal-ipelets.gforge.inria.fr.

5. New Results

5.1. Modeling interfaces and contacts

Keywords: *Docking*, *arrangements of balls*, *interfaces*, *protein complexes*, *scoring*, *scoring functions; Voronoi diagrams*, *structural alphabets*.

5.1.1. Computing the exact arrangement of circles on a sphere, with applications in structural biology

Participants: Frédéric Cazals, Sébastien Loriot.

Given a collection of circles on a sphere, we adapt the Bentley-Ottmann algorithm to the spherical setting to compute the *exact* arrangement of the circles [13], [11]. The algorithm consists of sweeping the sphere with a meridian, which is non trivial because of the degenerate cases and the algebraic specification of event points.

From an algorithmic perspective, and with respect to general sweep-line algorithms, we investigate a strategy maintaining a linear size event queue. (The algebraic aspects involved in the development of the predicates involved in our algorithm are reported in [12].)

From an implementation perspective, we present the first effective arrangement calculation dealing with general circles on a sphere in an exact fashion, as exactness incurs a mere factor of two with respect to calculations performed using *double* floating point numbers on generic examples. In particular, we stress the importance of maintaining a linear size queue, in conjunction with arithmetic filter failures.

From an application perspective, we present an application in structural biology. Given a collection of atomic balls, we adapt the sweep-line algorithm to report all balls covering a given face of the spherical arrangement on a given atom. This calculation is used to define molecular surface related quantities going beyond the classical exposed and buried solvent accessible surface areas. Spectacular differences w.r.t. traditional observations on protein - protein and protein - drug complexes are also reported.

5.2. Algorithmic foundations

Keywords: *Computational geometry*, *differential geometry*, *optimization*, *robustness*.

5.2.1. Spheres, circles and circular arcs in 3D

Participants: Frédéric Cazals, Sébastien Loriot.

In collaboration with P. Machado Manhães de Castro and M. Teillaud, GEOMETRICA. *The first part of this work was partially done while Pedro Machado Manhães de Castro was visiting* GEOMETRICA *in 2006, in the framework of the INRIA Intership Programme.*

A 3D kernel for the manipulations of spheres, circles and circular arcs in 3D was submitted to the CGAL Editorial board. The package follows the same overall design as the 2D circular kernel. It proposes functionalities involving these objects. It also defines the concept of an algebraic kernel dedicated to the special case of spheres, lines and circles in 3D.

We proposed more recently to expand this package by adding objects and functionalities dedicated to the case where all the objects handled are located on a reference sphere. We showed how the two frameworks can be combined [12].

5.2.2. Estimating the Differential Properties of a Sampled Smooth Surface

Participant: Frédéric Cazals.

In collaboration with M. Pouget (VEGAS project, LORIA, NANCY).

Surfaces of $\mathbb{R}^3$ are ubiquitous in science and engineering, and estimating the local differential properties of a surface discretized as a point cloud or a triangle mesh is a central building block in all situations where surfaces are dealt with. One strategy to perform such an estimation consists of resorting to polynomial fitting, either interpolation or approximation, but this route is difficult for several reasons: choice of the coordinate system, numerical handling of the fitting problem, extraction of the differential properties.

This work [14] presents a generic C++ software package solving these problems. On the theoretical side and as established in a companion paper, the interpolation and approximation methods provided achieve the best asymptotic error bounds known to date. On the implementation side and following state-of-the-art coding rules in Computational Geometry, genericity of the package is achieved thanks to three template classes accounting for (a) the type of the input points (b) the internal geometric computations and (c) the linear algebra operations. An instantiation using CLAPACK is also provided within the Computational Geometry Algorithms Library (CGAL, from version 3.3 on).

6. Other Grants and Activities

6.1. National initiatives

6.1.1. INRIA Cooperative Research Initiative (ARC) REFLEXP

Participants: Benjamin Bouvier, Frédéric Cazals, David Cohen-Steiner, Sébastien Loriot.

The goal of this ARC is to develop new methodologies to handle the flexibility of proteins. The ARC, which is coordinated by F. Cazals, features a collaboration with Institut Pasteur Paris (M. Nilges), and MPI Saarebrucken (J. Giesen). The following topics have been investigated: characterization of dynamic properties of interface atoms during Molecular Dynamics simulations of a complex; optimization algorithms to perform conformer selection amongst a molecular ensemble; characterization of induced fit mechanisms in protein-drug binding. Additional details can be found on the following web site http://www-sop.inria.fr/geometrica/team/Frederic.Cazals/arc-reflexp.

6.2. International initiatives

6.2.1. Project France-Stanford Center for Interdisciplinary Studies

Participants: Frédéric Cazals, Julie Bernauer, Sébastien Loriot.

The France-Stanford Center for Interdisciplinary Studies is funding a two-year project (2007-08) entitled *Developments of Geometric Methods and Algorithms for the study of macro molecular assemblies*. The PIs are F. Cazals (INRIA) and M. Levitt (chair of the Structural Biology Dpt, Stanford University). The goal of the project is to make a stride towards improved multi-scale modeling of large protein complexes.

7. Dissemination

7.1. Animation of the scientific community

7.1.1. Conference program committees

- Frédéric Cazals was member of the paper committee of the Eurographics Symposium on Geometry Processing 07, of the ACM Symposium on Solid and Physical Modeling 07, and of New Advances in Shape Analysis and Geometric Modeling 07.

7.1.2. INRIA committees

- Frédéric Cazals is member of the INRIA Scientific Steering Committee (COST).

7.1.3. WWW server

http://www-sop.inria.fr/abs/
The ABS web server is under construction –release scheduled for December 2007.

7.2. Teaching

7.2.1. Teaching at universities

- Ecole Centrale Paris, 3rd year (master); Introduction to Computational Structural Biology; F. Cazals, 21h.
- Master Bioinformatique et Biostatistiques (BIBS), Orsay University; Algorithmic Problems in Computational Structural Biology; F. Cazals (12h), J. Janin (6h), C. Robert (6h).
- Cursus Ecole Normale Supérieure de Lyon à Sophia-Antipolis; Topics in Structural Biology; F. Cazals, 24h.

7.2.2. Internships

Internship proposals can be seen on the web at http://www-sop.inria.fr/geometrica/positions/
- Sushant Sachdeva, *Optimization problems for conformer selection*, IIT Bombay.

7.2.3. Ongoing Ph.D. theses

- Sébastien Loriot, *Modélisation mathématique, calcul et classification de poches d'arrimage de médicaments sur les protéines*, université de Bourgogne.

7.3. Participation to conferences, seminars, invitations

7.3.1. Invited talks

Members of the project have presented their published articles at conferences. The reader can refer to the bibliography to obtain the corresponding list. We list below all other talks given in seminars or summer schools.

- *Modèles et algorithmes pour la description des interactions macro-moléculaires: le triptyque biophysique - géométrie - statistiques*; French Academy of Sciences, Colloquium *Avancées en Sciences de l'Information présentées par leurs auteurs*; F. Cazals.
- *Geometric and topological inference in the non linear realm: on the importance of singularity theory*; Non-Linear Computational Geometry, Institute for Mathematics and its Applications, Univ. of Minnesota; F. Cazals.
- *Modèles pour la description de la structure des protéines: de la géométrie à la bio-physique en passant par les statistiques*; Bioinformatique, modélisation des systèmes biologiques Journées 2007, ACI-IMPBIO & GDR BIM, Institut Henri Poincaré (Paris); F. Cazals.
- *Describing protein-protein and atomic environments: a geometric perspective*; Dpts of Biopharmaceutical Sciences and Pharmaceutical Chemistry (SALILAB), UCSF; Dpt of Structural Biology, Stanford Univ., F. Cazals.

7.3.2. The ABS seminar

http://www-sop.inria.fr/geometrica/seminars/
Joint seminar with GEOMETRICA.

7.3.3. Scientific visits

- F. Cazals visited Stanford University for two weeks (November).

ABS has hosted the following scientists:
- Raik Gruenberg, Univ. of Barcelona, one week, July 2007.

8. Bibliography

Major publications by the team in recent years

[1] J.-D. BOISSONNAT, F. CAZALS. *Smooth Surface Reconstruction via Natural Neighbour Interpolation of Distance Functions*, in "Comp. Geometry Theory and Applications", 2002, p. 185–203.

[2] F. CAZALS. *Computational Geometry and Topology: Concepts, Algorithms, Applications*, Habilitation à diriger des recherches, Université de Nice-Sophia Antipolis, France, 2006.

[3] F. CAZALS. *La morphologie des contacts entre bio-molécules*, in "Pour la Science", vol. Juillet - Août, 2006.

[4] F. CAZALS. *Effective nearest neighbors searching on the hyper-cube, with a pplications to molecular clustering*, in "Proc. 14th Annu. ACM Sympos. Comput. Geom.", 1998, p. 222–230.

[5] F. CAZALS, F. CHAZAL, T. LEWINER. *Molecular shape analysis based upon the Morse-Smale complex and the Connolly function*, in "ACM Symposium on Computational Geometry", 2003.

[6] F. CAZALS, J.-C. FAUGÈRE, M. POUGET, F. ROUILLIER. *The implicit structure of ridges of a smooth parametric surface*, in "Computer Aided Geometric Design", INRIA Tech Report 5608, vol. 23, n^{o} 7, 2006, p. 582-598, http://hal.inria.fr/inria-00071237/fr/.

[7] F. CAZALS, J. GIESEN. *Delaunay Triangulation Based Surface Reconstruction*, in "Effective Computational Geometry for curves and surfaces", J.-D. BOISSONNAT, M. TEILLAUD (editors), Springer-Verlag, Mathematics and Visualization, 2006.

[8] F. CAZALS, C. KARANDE. *An algorithm for reporting maximal c-cliques*, in "Theoretical Computer Science", vol. 349, n^{o} 3, 2005, p. 484–490.

[9] F. CAZALS, M. POUGET. *Estimating Differential Quantities using Polynomial fitting of Osculating Jets*, in "Computer Aided Geometric Design", Conf. version: Symp. on Geometry Processing 2003, vol. 22, n^{o} 2, 2005, p. 121–146.

[10] F. CAZALS, F. PROUST, R. BAHADUR, J. JANIN. *Revisiting the Voronoi description of Protein-Protein interfaces*, in "Protein Science", vol. 15, n^{o} 9, 2006, p. 2082–2092.

Year Publications

Publications in Conferences and Workshops

[11] F. CAZALS, S. LORIOT. *Computing the exact arrangement of circles on a sphere, with applications in structural biology*, in "Proc. 23th Annu. ACM Sympos. Comput. Geom. —Video/Multimedia track", 2007.

Internal Reports

[12] P. M. M. D. CASTRO, F. CAZALS, S. LORIOT, M. TEILLAUD. *Design of the CGAL Spherical Kernel and application to arrangements of circles on a sphere*, Submitted, Research Report, n^{o} 6298, INRIA, 09 2007, https://hal.inria.fr/inria-00173124.

[13] F. CAZALS, S. LORIOT. *Computing the exact arrangement of circles on a sphere, with applications in structural biology*, Revised version. Submitted., Research Report, n° 6049, INRIA, 12 2007, https://hal.inria.fr/inria-00118781.

[14] F. CAZALS, M. POUGET. *Jet_fitting_3: A Generic C++ Package for Estimating the Differential Properties on Sampled Surfaces via Polynomial Fitting*, Submitted, Research Report, n° 6093, INRIA, 01 2007, https://hal.inria.fr/inria-00123501.

References in notes

[15] O. BECKER, A. D. MACKERELL, B. ROUX, M. WATANABE. *Computational Biochemistry and Biophysics*, M. Dekker, 2001.

[16] A.-C. CAMPROUX, R. GAUTIER, P. TUFFERY. *A Hidden Markov Model derived structural alphabet for proteins*, in "J. Mol. Biol.", 2004, p. 591-605.

[17] F. CHAZAL, D. COHEN-STEINER, A. LIEUTIER. *A Sampling Theory for Compacts in Euclidean Spaces*, in "ACM Symp. Comp. Geometry", 2006.

[18] F. CHAZAL, A. LIEUTIER. *Weak Feature Size and persistent homology : computing homology of solids in $\mathbb{R}^n$ from noisy data samples*, 2005.

[19] D. COHEN-STEINER, H. EDELSBRUNNER, J. HARER. *Stability of Persistence Diagrams*, in "ACM Symp. Comp. Geometry", 2005.

[20] M. L. CONNOLLY. *Analytical molecular surface calculation*, in "J. Appl. Crystallogr.", vol. 16, 1983.

[21] R. DUNBRACK. *Rotamer libraries in the 21st century*, in "Curr Opin Struct Biol", vol. 12, n° 4, 2002, p. 431-440.

[22] A. FERNÁNDEZ, R. BERRY. *Extent of Hydrogen-Bond Protection in Folded Proteins: A Constraint on Packing Architectures*, in "Biophysical Journal", vol. 83, 2002, p. 2475-2481.

[23] A. FERSHT. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, 1999.

[24] M. GERSTEIN, F. RICHARDS. *Protein geometry: volumes, areas, and distances*, in "The international tables for crystallography (Vol F, Chap. 22)", vol. F (Chapter 22.1.1), 2001.

[25] J. GIESEN, M. JOHN. *The Flow Complex: A Data Structure for Geometric Modeling*, in "ACM SODA", 2003.

[26] H. GOHLKE, G. KLEBE. *Statistical potentials and scoring functions applied to protein-ligand binding*, in "Curr. Op. Struct. Biol.", vol. 11, 2001, p. 231-235.

[27] J. JANIN, S. WODAK, M. LEVITT, B. MAIGRET. *Conformations of amino acid side chains in proteins*, in "J. Mol. Biol.", vol. 125, 1978, p. 357–386.

[28] V. K. KRIVOV, M. KARPLUS. *Hidden complexity of free energy surfaces for peptide (protein) folding*, in "PNAS", vol. 12, 2004.

[29] E. MEERBACH, C. SCHUTTE, I. HORENKO, B. SCHMIDT. *Metastable Conformational Structure and Dynamics: Peptides between Gas Phase and Aqueous Solution*, in "Analysis and Control of Ultrafast Photoinduced Reactions. Series in Chemical Physics 87", O. KUHN, L. WUDSTE (editors), Springer, 2007.

[30] I. MIHALEK, O. LICHTARGE. *On Itinerant Water Molecules and Detectability of Protein-Protein Int erfaces through Comparative Analysis of Homologues*, in "JMB", vol. 369, n^{o} 2, 2007.

[31] J. MINTSERIS, B. PIERCE, K. WIEHE, R. ANDERSON, R. CHEN, Z. WENG. *Integrating statistical pair potentials into protein complex prediction*, in "Proteins", vol. 69, 2007, p. 511–520.

[32] M. PETTINI. *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, 2007.

[33] E. PLAKU, H. STAMATI, C. CLEMENTI, L. KAVRAKI. *Fast and Reliable Analysis of Molecular Motion Using Proximity Relations and Dimensionality Reduction*, in "Proteins: Structure, Function, and Bioinformatics", vol. 67, n^{o} 4, 2007, p. 897–907.

[34] D. RAJAMANI, S. THIEL, S. VAJDA, C. CAMACHO. *Anchor residues in protein-protein interactions*, in "PNAS", vol. 101, 2004, p. 11287-11292.

[35] D. REICHMANN, O. RAHAT, S. ALBECK, R. MEGED, O. DYM, G. SCHREIBER. *From The Cover: The modular architecture of protein-protein binding interfaces*, in "PNAS", vol. 102, n^{o} 1, 2005, p. 57-62, http://www.pnas.org/cgi/content/abstract/102/1/57.

[36] F. M. RICHARDS. *Areas, volumes, packing and protein struture*, in "Ann. Rev. Biophys. Bioeng.", vol. 6, 1977, p. 151-176.

[37] G. RYLANCE, R. JOHNSTON, Y. MATSUNAGA, C.-B. LI, A. BABA, T. KOMATSUZAKI. *Topographical complexity of multidimensional energy landscapes*, in "PNAS", vol. 103, n^{o} 49, 2006, p. 18551-18555.

[38] G. SCHREIBER, L. SERRANO. *Folding and binding: an extended family business*, in "Current Opinion in Structural Biology", vol. 15, n^{o} 1, 2005.

[39] M. SIPPL. *Calculation of Conformational Ensembles from Potential of Mean Force: An Approach to the Knowledge-based prediction of Local Structures in Gobular Proteins*, in "J. Mol. Biol.", vol. 213, 1990, p. 859-883.

[40] C. SUMMA, M. LEVITT, W. DEGRADO. *An atomic environment potential for use in protein structure prediction*, in "JMB", vol. 352, 2005.

[41] S. WODAK, J. JANIN. *Structural basis of macromolecular recognition*, in "Adv. in protein chemistry", vol. 61, 2003.