

Inria

Activity Report 2019

Project-Team ABS

Algorithms, Biology, Structure

RESEARCH CENTER
Sophia Antipolis - Méditerranée

THEME
Computational Biology

Table of contents

1. Team, Visitors, External Collaborators	1
2. Overall Objectives	2
3. Research Program	3
3.1. Introduction	3
3.2. Modeling interfaces and contacts	3
3.3. Modeling macro-molecular assemblies	4
3.3.1. Reconstruction by Data Integration	4
3.3.2. Modeling with Uncertainties and Model Assessment	5
3.4. Modeling the flexibility of macro-molecules	5
3.5. Algorithmic foundations	6
3.5.1. Modeling Interfaces and Contacts	6
3.5.2. Modeling Macro-molecular Assemblies	6
3.5.3. Modeling the Flexibility of Macro-molecules	6
4. New Software and Platforms	7
5. New Results	7
5.1. Modeling interfaces and contacts	7
5.2. Modeling the flexibility of macro-molecules	8
5.2.1. Wang-Landau Algorithm: an adapted random walk to boost convergence	8
5.2.2. Survey of the analysis of continuous conformational variability of biological macro-molecules by electron microscopy	8
5.3. Algorithmic foundations	8
5.3.1. Comparing two clusterings using matchings between clusters of clusters	9
5.3.2. Low-Complexity Nonparametric Bayesian Online Prediction with Universal Guarantees	9
5.3.3. How long does it take for all users in a social network to choose their communities?	9
6. Partnerships and Cooperations	10
6.1. Regional Initiatives	10
6.2. International Research Visitors	10
7. Dissemination	10
7.1. Promoting Scientific Activities	10
7.1.1. Scientific Events: Organisation	10
7.1.2. Scientific Events: Selection	10
7.1.2.1. Member of the Conference Program Committees	10
7.1.2.2. Reviewer	11
7.1.3. Invited Talks	11
7.1.4. Leadership within the Scientific Community	11
7.1.5. Scientific Expertise	11
7.1.6. Research Administration	11
7.2. Teaching - Supervision - Juries	12
7.2.1. Teaching	12
7.2.2. Supervision	12
7.2.3. Juries	12
7.3. Popularization	12
7.3.1. Internal or external Inria responsibilities	12
7.3.2. Articles and contents	12
7.3.3. Education	13
7.3.4. Interventions	13
7.3.5. Internal action	14
7.3.6. Creation of media or tools for science outreach	14
8. Bibliography	14

Project-Team ABS

Creation of the Project-Team: 2008 July 01

Keywords:

Computer Science and Digital Science:

A2.5. - Software engineering
A3.3.2. - Data mining
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A6.1.4. - Multiscale modeling
A6.2.4. - Statistical methods
A6.2.8. - Computational geometry and meshes
A8.1. - Discrete mathematics, combinatorics
A8.3. - Geometry, Topology
A8.7. - Graph theory
A9.2. - Machine learning

Other Research Topics and Application Domains:

B1.1.1. - Structural biology
B1.1.5. - Immunology
B1.1.7. - Bioinformatics

1. Team, Visitors, External Collaborators

Research Scientists

Frédéric Cazals [Team leader, Inria, Senior Researcher, HDR]
Mehmet Serkan Apaydin [Inria, Starting Research Position, until May 2019]
Dorian Mazauric [Inria, Researcher]

PhD Students

Thi Viet Ha Nguyen [Inria, PhD Student]
Timothée O'Donnell [Inria, PhD Student]
Méliné Simsir [Univ Côte d'Azur, PhD Student]

Interns and Apprentices

Gabriel Djebbar [Inria, from Jul 2019 until Aug 2019]
Maria Guramare [Inria, from May 2019 until Aug 2019]
Guilherme Sales Santa Cruz [Inria, from Feb 2019 until Jul 2019]

Administrative Assistant

Florence Barbara [Inria, Administrative Assistant]

Visiting Scientist

Thiziri Nait Saada [Eurecom, from Sep 2019]

External Collaborator

Charles Robert [CNRS, HDR]

2. Overall Objectives

2.1. Overall Objectives

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 (DNA, RNA), 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 questions 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 [44]. 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, the order of magnitude of the number of genomes sequenced is of the order of one thousand, which results in circa one million of genes recorded in the manually curated Swiss-Prot database. On the other hand, the Protein Data Bank contains circa 90,000 structures. Thus, the paucity of structures with respect to the known number of genes calls for modeling in structural biology, so as to foster our understanding of the structure-to-function relationship.

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 [42], [31] and later Connolly [27], 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) [33], 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 within time scales ranging from milliseconds to hours, while time-steps used in molecular dynamics simulations are of the order of the femto-second, so that biologically relevant time-scales are out 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.

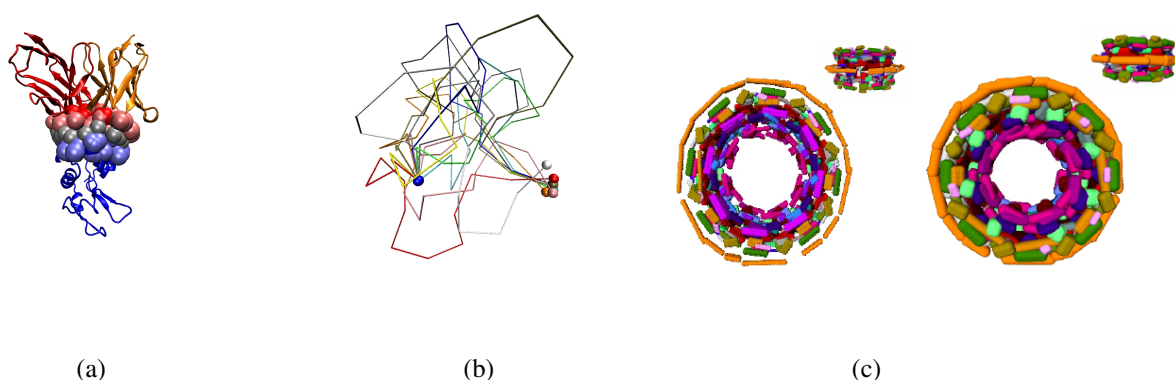


Figure 1. Geometric constructions in computational structural biology. (a) An antibody-antigen complex, with interface atoms identified by our Voronoi based interface model. This model is instrumental in mining correlations between structural and biological as well as biophysical properties of protein complexes [12]. (b) A diverse set of conformations of a backbone loop, selected thanks to a geometric optimization algorithm [8]. Such conformations are used by mean field theory based docking algorithms. (c) A tolerated model (TOM) of the nuclear pore complex, visualized at two different scales [9]. The parameterized family of shapes coded by a TOM is instrumental to identify stable properties of the underlying macro-molecular system.

3. Research Program

3.1. Introduction

The research conducted by ABS focuses on three main directions in Computational Structural Biology (CSB), together with the associated methodological developments:

- Modeling interfaces and contacts,
- Modeling macro-molecular assemblies,
- Modeling the flexibility of macro-molecules,
- Algorithmic foundations.

3.2. Modeling interfaces and contacts

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

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 [44]. Yet, in spite of almost three decades of investigations, the basic principles guiding the formation of interfaces and accounting for its stability are unknown [47]. Current investigations follow two routes. From the experimental perspective [30], directed mutagenesis enables one to quantify the energetic importance of residues, important residues being termed *hot* residues. Such studies recently evidenced the *modular* architecture of interfaces [41]. From the modeling perspective, the main issue consists of guessing the hot residues from sequence and/or structural informations [36].

The description of interfaces is also of special interest to improve *scoring functions*. By scoring function, two things are meant: either a function which assigns 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 [25], 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 [45], [32]. 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 [37]. In doing so, the energy defined is nothing but the Kullback-Leibler divergence between the distributions $\{p_i\}$ and $\{q_i\}$.

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

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 to relate the geometry and the bio-chemistry at the interface level [12]. 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 of such alphabet is that of [26]. At the atomic level and in spite of recent observations on the local structure of the neighborhood of a given atom [46], 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 [29].

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 [40]. 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 macro-molecular assemblies

Keywords: Macro-molecular assembly, reconstruction by data integration, proteomics, modeling with uncertainties, curved Voronoi diagrams, topological persistence.

3.3.1. Reconstruction by Data Integration

Large protein assemblies such as the Nuclear Pore Complex (NPC), chaperonin cavities, the proteasome or ATP synthases, to name a few, are key to numerous biological functions. To improve our understanding of

¹For 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.

²The 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.

these functions, one would ideally like to build and animate atomic models of these molecular machines. However, this task is especially tough, due to their size and their plasticity, but also due to the flexibility of the proteins involved. In a sense, the modeling challenges arising in this context are different from those faced for binary docking, and also from those encountered for intermediate size complexes which are often amenable to a processing mixing (cryo-EM) image analysis and classical docking. To face these new challenges, an emerging paradigm is that of reconstruction by data integration [24]. In a nutshell, the strategy is reminiscent from NMR and consists of mixing experimental data from a variety of sources, so as to find out the model(s) best complying with the data. This strategy has been in particular used to propose plausible models of the Nuclear Pore Complex [23], the largest assembly known to date in the eukaryotic cell, and consisting of 456 protein *instances* of 30 *types*.

3.3.2. Modeling with Uncertainties and Model Assessment

Reconstruction by data integration requires three ingredients. First, a parametrized model must be adopted, typically a collection of balls to model a protein with pseudo-atoms. Second, as in NMR, a functional measuring the agreement between a model and the data must be chosen. In [22], this functional is based upon *restraints*, namely penalties associated to the experimental data. Third, an optimization scheme must be selected. The design of restraints is notoriously challenging, due to the ambiguous nature and/or the noise level of the data. For example, Tandem Affinity Purification (TAP) gives access to a *pullout* i.e. a list of protein types which are known to interact with one tagged protein type, but no information on the number of complexes or on the stoichiometry of proteins types within a complex is provided. In cryo-EM, the envelope enclosing an assembly is often imprecisely defined, in particular in regions of low density. For immuno-EM labelling experiments, positional uncertainties arise from the microscope resolution.

These uncertainties coupled with the complexity of the functional being optimized, which in general is non convex, have two consequences. First, it is impossible to single out a unique reconstruction, and a set of plausible reconstructions must be considered. As an example, 1000 plausible models of the NPC were reported in [22]. Interestingly, averaging the positions of all balls of a particular protein type across these models resulted in 30 so-called *probability density maps*, each such map encoding the probability of presence of a particular protein type at a particular location in the NPC. Second, the assessment of all models (individual and averaged) is non trivial. In particular, the lack of straightforward statistical analysis of the individual models and the absence of assessment for the averaged models are detrimental to the mechanistic exploitation of the reconstruction results. At this stage, such models therefore remain qualitative.

3.4. Modeling the flexibility of macro-molecules

Keywords: Folding, docking, energy landscapes, induced fit, molecular dynamics, conformers, conformer ensembles, point clouds, reconstruction, shape learning, Morse theory.

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 generate ensembles of conformations, called *conformers*, and so do molecular dynamics (MD) simulations. 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 relaxed³. 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.

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

At the side-chain level, the question of improving rotamer libraries is still of interest [28]. 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) [43]. 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 [39], to Morse theory [34] and to analysis of meta-stable states of time series [35] have been proposed.

3.5. Algorithmic foundations

Keywords: Computational geometry, computational topology, optimization, data analysis.

Making a stride towards a better understanding of the biophysical questions discussed in the previous sections requires various methodological developments, which we briefly discuss now.

3.5.1. Modeling Interfaces and Contacts

In modeling interfaces and contacts, one may favor geometric or topological information.

On the geometric side, the problem of modeling contacts at the atomic level is tantamount to encoding multi-body relations between an atom and its neighbors. On the one hand, one may use an encoding of neighborhoods based on geometric constructions such as Voronoi 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. The information gathered while modeling contacts can further be integrated into interface models.

On the topological side, one may favor constructions which remain stable if each atom in a structure *retains* the same neighbors, even though the 3D positions of these neighbors change to some extent. This process is observed in flexible docking cases, and call for the development of methods to encode and compare shapes undergoing tame geometric deformations.

3.5.2. Modeling Macro-molecular Assemblies

In dealing with large assemblies, a number of methodological developments are called for.

On the experimental side, of particular interest is the disambiguation of proteomics signals. For example, TAP and mass spectrometry data call for the development of combinatorial algorithms aiming at unraveling pairwise contacts between proteins within an assembly. Likewise, density maps coming from electron microscopy, which are often of intermediate resolution (5-10Å) call the development of noise resilient segmentation and interpretation algorithms. The results produced by such algorithms can further be used to guide the docking of high resolutions crystal structures into maps.

As for modeling, two classes of developments are particularly stimulating. The first one is concerned with the design of algorithms performing reconstruction by data integration, a process reminiscent from non convex optimization. The second one encompasses assessment methods, in order to single out the reconstructions which best comply with the experimental data. For that endeavor, the development of geometric and topological models accommodating uncertainties is particularly important.

3.5.3. Modeling the Flexibility of Macro-molecules

Given a sampling on an energy landscape, a number of fundamental issues actually arise: how does the point cloud describe 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 noted in passing that such questions are probably related to modeling phase transitions in statistical physics where geometric and topological methods are being used [38].

From an algorithmic standpoint, such questions are reminiscent of *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 Computer Aided Geometric Design, surface reconstruction triggered a major activity in the Computational Geometry community over the past ten years. Aside from applications, reconstruction raises a number of deep issues: the study of distance functions to the model and to the samples, and their comparison; the study of Morse-like constructions stemming from distance functions to points; the analysis of topological invariants of the model and the samples, and their comparison.

4. New Software and Platforms

4.1. SBL

Structural Bioinformatics Library

KEYWORDS: Structural Biology - Biophysics - Software architecture

FUNCTIONAL DESCRIPTION: The SBL is a generic C++/python cross-platform software library targeting complex problems in structural bioinformatics. Its tenet is based on a modular design offering a rich and versatile framework allowing the development of novel applications requiring well specified complex operations, without compromising robustness and performances.

More specifically, the SBL involves four software components (1-4 thereafter). For end-users, the SBL provides ready to use, state-of-the-art (1) applications to handle molecular models defined by unions of balls, to deal with molecular flexibility, to model macro-molecular assemblies. These applications can also be combined to tackle integrated analysis problems. For developers, the SBL provides a broad C++ toolbox with modular design, involving core (2) algorithms, (3) biophysical models, and (4) modules, the latter being especially suited to develop novel applications. The SBL comes with a thorough documentation consisting of user and reference manuals, and a bugzilla platform to handle community feedback.

RELEASE FUNCTIONAL DESCRIPTION: In 2018, major efforts targeted two points. First, the simplification of installation procedures – now possible with conda/python. Second, the development of packages revolving on molecular flexibility at large: representations in internal and Cartesian coordinates, generic representation of molecular mechanics force fields (and computation of gradients), exploration algorithms for conformational spaces.

- Contact: Frédéric Cazals
- Publication: [The Structural Bioinformatics Library: modeling in biomolecular science and beyond](#)
- URL: <https://sbl.inria.fr/>

5. New Results

5.1. Modeling interfaces and contacts

Keywords: docking, scoring, interfaces, protein complexes, Voronoi diagrams, arrangements of balls.

5.1.1. Characterizing molecular flexibility by combining IRMSD measures

Participants: F. Cazals, R. Tetley.

The root mean square deviation (RMSD) and the least RMSD are two widely used similarity measures in structural bioinformatics. Yet, they stem from global comparisons, possibly obliterating locally conserved motifs. In this work [16], we correct these limitations with the so-called *combined RMSD*, which mixes independent IRMSD measures, each computed with its own rigid motion. The combined RMSD is relevant in two main scenarios, namely to compare (quaternary) structures based on motifs defined from the sequence (domains, SSE), and to compare structures based on structural motifs yielded by local structural alignment methods.

We illustrate the benefits of combined RMSD over the usual IRMSD on three problems, namely (i) the assignment of quaternary structures for hemoglobin (scenario #1), (ii) the calculation of structural phylogenies (case study: class II fusion proteins; scenario #1), and (iii) the analysis of conformational changes based on combined RMSD of rigid structural motifs (case study: one class II fusion protein; scenario #2). Using these, we argue that the combined RMSD is a tool of choice to perform positive and negative discrimination of degree of freedom, with applications to the design of move sets and collective coordinates.

Executables to compute combined RMSD are available within the Structural Bioinformatics Library (<http://sbl.inria.fr>).

5.2. Modeling the flexibility of macro-molecules

Keywords: protein, flexibility, collective coordinate, conformational sampling dimensionality reduction.

5.2.1. Wang-Landau Algorithm: an adapted random walk to boost convergence

Participants: F. Cazals, A. Chevallier.

The Wang-Landau (WL) algorithm is a recently developed stochastic algorithm computing densities of states of a physical system, and also performing numerical integration in high dimensional spaces. Since its inception, it has been used on a variety of (bio-)physical systems, and in selected cases, its convergence has been proved. The convergence speed of the algorithm is tightly tied to the connectivity properties of the underlying random walk.

In this work [19], we propose an efficient random walk that uses geometrical information to circumvent the following inherent difficulties: avoiding overstepping strata, toning down concentration phenomena in high-dimensional spaces, and accommodating multidimensional distributions. These improvements are especially well suited to improve calculations on a per basin basis – included anharmonic ones.

Experiments on various models stress the importance of these improvements to make WL effective in challenging cases. Altogether, these improvements make it possible to compute density of states for regions of the phase space of small biomolecules.

5.2.2. Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy

Participant: F. Cazals.

In collaboration with a group of colleagues led by J. M. Carazo, CSIC, Biocomputing Unit, National Center for Biotechnology, Spain.

Single-particle analysis by electron microscopy is a well established technique for analyzing the three-dimensional structures of biological macromolecules. Besides its ability to produce high-resolution structures, it also provides insights into the dynamic behavior of the structures by elucidating their conformational variability. In this work [17], the different image-processing methods currently available to study continuous conformational changes are reviewed.

5.3. Algorithmic foundations

Keywords: Computational geometry, computational topology, optimization, data analysis.

5.3.1. Comparing two clusterings using matchings between clusters of clusters

Participants: F. Cazals, D. Mazauric, R. Tetley.

In collaboration with R. Watrigant, University Lyon I.

Clustering is a fundamental problem in data science, yet, the variety of clustering methods and their sensitivity to parameters make clustering hard. To analyze the stability of a given clustering algorithm while varying its parameters, and to compare clusters yielded by different algorithms, several comparison schemes based on matchings, information theory and various indices (Rand, Jaccard) have been developed. In this work [15], we go beyond these by providing a novel class of methods computing meta-clusters within each clustering— a meta-cluster is a group of clusters, together with a matching between these.

Let the intersection graph of two clusterings be the edge-weighted bipartite graph in which the nodes represent the clusters, the edges represent the non empty intersection between two clusters, and the weight of an edge is the number of common items. We introduce the so-called D -Family matching problem on intersection graphs, with D the upper-bound on the diameter of the graph induced by the clusters of any meta-cluster. First we prove NP -completeness and APX -hardness results, and unbounded approximation ratio of simple strategies. Second, we design exact polynomial time dynamic programming algorithms for some classes of graphs (in particular trees). Then, we prove spanning-tree based efficient heuristic algorithms for general graphs.

Our experiments illustrate the role of D as a scale parameter providing information on the relationship between clusters within a clustering and in-between two clusterings. They also show the advantages of our built-in mapping over classical cluster comparison measures such as the variation of information (VI).

5.3.2. Low-Complexity Nonparametric Bayesian Online Prediction with Universal Guarantees

Participant: F. Cazals.

In collaboration with A. Lhéritier, Amadeus SA.

In this work [18], we propose a novel nonparametric online predictor for discrete labels conditioned on multivariate continuous features. The predictor is based on a feature space discretization induced by a full-fledged k -d tree with randomly picked directions and a recursive Bayesian distribution, which allows to automatically learn the most relevant feature scales characterizing the conditional distribution. We prove its pointwise universality, i.e., it achieves a normalized log loss performance asymptotically as good as the true conditional entropy of the labels given the features. The time complexity to process the n -th sample point is $O(\log n)$ in probability with respect to the distribution generating the data points, whereas other exact nonparametric methods require to process all past observations. Experiments on challenging datasets show the computational and statistical efficiency of our algorithm in comparison to standard and state-of-the-art methods.

5.3.3. How long does it take for all users in a social network to choose their communities?

Participant: D. Mazauric.

In collaboration with J.-C. Bermond (Coati project-team), A. Chaintreau (Columbia University), and G. Ducoffe (National Institute for Research and Development in Informatics, Bucharest).

In this work [14], we consider a community formation problem in social networks, where the users are either friends or enemies. The users are partitioned into conflict-free groups (i.e., independent sets in the *conflict graph* $G^- = (V, E)$ that represents the enmities between users). The dynamics goes on as long as there exists any set of at most k users, k being any fixed parameter, that can change their current groups in the partition *simultaneously*, in such a way that they all strictly increase their utilities (number of friends i.e., the cardinality of their respective groups minus one). Previously, the best-known upper-bounds on the maximum time of convergence were $\mathcal{O}(|V|\alpha(G^-))$ for $k \leq 2$ and $\mathcal{O}(|V|^3)$ for $k = 3$, with $\alpha(G^-)$ being the independence number of G^- . Our first contribution in this paper consists in reinterpreting the initial problem as the study of a dominance ordering over the vectors of integer partitions. With this approach, we obtain for $k \leq 2$ the tight upper-bound $\mathcal{O}(|V| \min \{\alpha(G^-), \sqrt{|V|}\})$ and, when G^- is the empty graph, the exact value of order $\frac{(2|V|)^{3/2}}{3}$.

The time of convergence, for any fixed $k \geq 4$, was conjectured to be polynomial. In this paper we disprove this. Specifically, we prove that for any $k \geq 4$, the maximum time of convergence is in $\Omega(|V|^{\Theta(\log |V|)})$.

6. Partnerships and Cooperations

6.1. Regional Initiatives

– Frédéric Cazals is endowed chair within the 3IA Côte d’Azur (<http://3ia.univ-cotedazur.fr/>), within the focus area *Computational Biology and Bio-Inspired AI*.

6.2. International Research Visitors

6.2.1. Visits of International Scientists

6.2.1.1. Internships

- Internship of Maria Guramare, Harvard University, Cambridge, Massachusetts. Supervision: Frédéric Cazals and Dorian Mazauric. *Shortest Paths under Constraints Problem with Application for Structural Alignments*.
- Internship of Guilherme Santa Cruz, Polytech Nice. Supervision: Frédéric Cazals and Dorian Mazauric. *New method for assessing protein-phenotype relevance interpolating gene expression and biological networks*.
- Internship of Gabriel Djebbar, École Polytechnique de l’Université Nice Sophia Antipolis. Filière Sciences Informatiques, deuxième année du cursus ingénieur (niveau Master 1). Supervision: Frédéric Havet and Dorian Mazauric. *Graph Coloring Games with application on Algorithmic Geometry*.
- Projet de fin d’études de Youssef Benjelloun et Quentin Sautel, École Polytechnique de l’Université Nice Sophia Antipolis. Filière Mathématiques Appliquées et Modélisation, deuxième année du cursus ingénieur (niveau Master 1). Supervision: Dorian Mazauric. *Algorithmes pour le calcul de motif commun d’un ensemble de conformations d’une protéine*.

7. Dissemination

7.1. Promoting Scientific Activities

7.1.1. Scientific Events: Organisation

7.1.1.1. General Chair, Scientific Chair

– Frédéric Cazals:

- F. Cazals co-organized, with P. Alliez and F. Chazal, the conference *New Horizons in Computational Geometry and Topology*, held at Sophia Antipolis (September 5-6, 2019) in honor of J-D. Boissonnat. See <https://project.inria.fr/jdb2019/en/>.

7.1.2. Scientific Events: Selection

7.1.2.1. Member of the Conference Program Committees

– Frédéric Cazals was member of the following program committees:

- Symposium on Computational Geometry
- Symposium On Geometry Processing
- International Conference on Computational Systems-Biology and Bioinformatics:
- Intelligent Systems for Molecular Biology (ISMB), PC member of Protein Interactions & Molecular Networks

7.1.2.2. Reviewer

– Frédéric Cazals reviewed for the following journals:

- Bioinformatics
- Discrete and applied mathematics
- PLOS Computational Biology
- PLOS One

7.1.3. Invited Talks

– Frédéric Cazals gave the following invited talks:

- *Multiscale molecular flexibility analysis: novel insights*. MPI Frankfurt, July 2019.
- *Multiscale molecular flexibility analysis: novel insights*. Conference of the Polish Bioinformatics Society, Torun, June 2019.
- *Bio-molecules: on the role of geometry in the tryptic structure - dynamics -function*. Meeting of the French Academy of Sciences, French Riviera, June 2019.
- *Leveraging structural data by decoupling structure, thermodynamics and dynamics*, Challenges in large scale biomolecular simulation, Inst. Etudes Sc. de Cargese, France, May 2019
- *Mining molecular flexibility: novel tools, novel insights*, University College, London, April 2019.
- *Novel structural parameters of IG-Ag complexes yield a quantitative description of interaction specificity and binding affinity*, Journées scientifiques de la SFHI (Soc. Fr. d'histocompatibilité et d'immunogénétique), Paris, Mars 2019.

7.1.4. Leadership within the Scientific Community

– Frédéric Cazals:

- 2010-.... Member of the steering committee of the *GDR Bioinformatique Moléculaire*, for the *Structure and macro-molecular interactions* theme.
- 2017-2019. Co-chair, with Yann Ponty, of the working group / groupe de travail (*GT MASIM - Méthodes Algorithmiques pour les Structures et Interactions Macromoléculaires*, within the *GDR de Bioinformatique Moléculaire* (GDR BIM, <http://www.gdr-bim.cnrs.fr/>).

7.1.5. Scientific Expertise

– Frédéric Cazals:

- 2019: reviewer for ERC consolidator grants.

7.1.6. Research Administration

– Frédéric Cazals:

- 2017-.... President of the *Comité de suivi doctoral (CSD)*, Inria Sophia Antipolis - Méditerranée. The CSD supervises all aspects of PhD student's life within Inria Sophia Antipolis - Méditerranée.
- 2018-.... Member of the *bureau du comité des équipes projets*.

– Dorian Mazaauric:

- 2019-.... Head of *Commission Mastic* (Médiation et Animation des MATHématiques, des Sciences et Techniques Informatiques et des Communications), Inria Sophia Antipolis - Méditerranée.
- 2016-2019. Member of the *Comité de Centre*, Inria Sophia Antipolis - Méditerranée.
- 2018-2019. Member of the *Commission de Développement Technologique*, Inria Sophia Antipolis - Méditerranée.

7.2. Teaching - Supervision - Juries

7.2.1. Teaching

- Master: Frédéric Cazals (Inria ABS) and Frédéric Chazal (Inria Saclay), *Foundations of Geometric Methods in Data Analysis*, Data Sciences Program, Department of Applied Mathematics, Ecole Centrale Paris. (<http://www-sop.inria.fr/abs/teaching/centrale-FGMDA/centrale-FGMDA.html>)
- Master : Dorian Mazauric, Algorithmique et Complexité, 23h30 TD, niveau M1, École Polytechnique de l'Université Nice Sophia Antipolis, filière Sciences Informatiques, France.
- Bachelor : Dorian Mazauric, Langages Formelles et Automates, 24h TD, niveau L3, École Polytechnique de l'Université Nice Sophia Antipolis, filière Sciences Informatiques, France.
- Bachelor : Dorian Mazauric, Algorithmique, 3h de cours et 8 heures de TD, niveau licence, Université Côte d'Azur, Diplôme interuniversitaire (formation des enseignants de lycée), Sophia Antipolis, France.

7.2.2. Supervision

- **PhD, 4th year:** Augustin Chevallier, *Random walks for estimating the volume of convex bodies and densities of states in high dimensional spaces*. Defended on the 08/04/2019. Université Côte d'Azur.
- **PhD in progress, 3rd year:** Méliné Simsir, *Modeling drug efflux by Patched*. Université Côte d'Azur. Thesis co-supervised by Frédéric Cazals and Isabelle Mus-Veteau, IPMC/CNRS.
- **PhD in progress, 2nd year:** Timothée O'Donnel, *Modeling the influenza polymerase*. Université Côte d'Azur. Thesis co-supervised by Frédéric Cazals and Bernard Delmas, INRA Jouy-en-Josas.
- **PhD in progress, 2nd year :** Thi Viet Ha Nguyen, Graph Algorithms techniques for (low and high) resolution models of large protein assemblies, Frédéric Havet (Inria/I3S project-team Coati) and Dorian Mazauric.

7.2.3. Juries

– Frédéric Cazals:

- Maria Elisa Ruiz Echartea, Université de Lorraine, December 2019. Rapporteur on the PhD thesis *Multi-component protein assembly using distance constraints*. Advisors: David W. Ritchie - Marie-Dominique Devignes.
- Alba Chiara De Vitis, Université Côte d'Azur, May 2019. Committee member. *Kernel Methods for High Dimensional Data Analysis*. Advisors: D. Cohen-Steiner, J-D. Boissonnat.
- Juraj Michalik, Ecole Polytechnique, March 2019. Rapporteur on the PhD thesis *Non-redundant sampling in RNA bioinformatics*. Advisors: Yann Ponty and Hélène Touzet.

7.3. Popularization

This part mainly concerns Dorian Mazauric.

7.3.1. Internal or external Inria responsibilities

- 2019-.... Head of *Commission Mastic* (Médiation et Animation des MATHématiques, des Sciences et Techniques Informatiques et des Communications), Inria Sophia Antipolis - Méditerranée.
- 2019-.... Coordinator of *Terra Numerica – vers une Cité du Numérique*.
- 2018-2019. Member of the *Conseil d'Administration de l'association les Petits Débrouillards*.
- 2018-2019. Coordinator of *projet de médiation Galéjade* : Graphes et ALgorithmes : Ensemble de Jeux À Destination des Ecoliers... (mais pas que).

7.3.2. Articles and contents

See <https://galejade.inria.fr> and <http://terra-numerica.org/>.

7.3.3. Education

- 26/10/2019. Animation d'ateliers de GALEJADE au weekend d'initiation à la culture numérique organisé par Class'Code MED. Graphes, algorithmes et magie des mathématiques et de l'informatique.
- 03/06/2019. Formation d'enseignants à Draguignan (ÉSPÉ de l'Académie de Nice). Graphes et algorithmes.
- 05/04/2019. Coordination d'ateliers Galejade animés par des enseignants et des parents d'élèves de l'école Pasteur du Cannet (suite à une formation Galejade quelques mois plus tôt). Ateliers Galejade.
- 22/03/2019. Formation d'enseignants de REP+ à l'ÉSPÉ de l'Académie de Nice (site de Stéphen Liégeard).
- 26/02/2019. Formation des étudiants de l'ÉSPÉ de l'Académie de Nice (site de Stéphen Liégeard) à l'animation d'activités informatiques et mathématiques. En collaboration avec l'Inspection Académique. Dans le cadre de la semaine des mathématiques. Graphes, algorithmes, jeux et magie des mathématiques.

7.3.4. Interventions

- National events:
 - 19-20/10/2019. Ateliers au village des sciences et de l'innovation au Palais des Congrès d'Antibes Juan-les-Pins. Voyage en train au pays des graphes, des algorithmes et de l'intelligence artificielle.
 - 06/10/2019. Atelier au village des sciences de Nice. Fête de la Science 2019. Le plus grand réseau de tri humain du monde.
 - 05/10/2019. Ateliers et conférence au village des sciences de Villeneuve-Loubet. Fête de la Science 2019. Algorithmes grandeur nature ? La magie des graphes et du binaire.
 - 4/10/2019. Conférence à la maison des associations de Nice (114 élèves de première et terminal). Fête de la Science 2019. Pas besoin de réfléchir, les ordinateurs calculent tellement vite ?
 - 03/10/2019. Deux conférences et ateliers au collège de la Vesubie-Jean Salines à Roquebillière. Fête de la Science 2019. La magie des graphes et du binaire ? Algorithmes grandeur nature ? Pas besoin de réfléchir, les ordinateurs calculent tellement vite ? Théorie des graphes et algorithmique pour les réseaux.
 - 26/03/2019. Fête des maths à l'ÉSPÉ de l'Académie de Nice (site de Stéphen Liégeard). Graphes et Algorithmes.
 - 26/02/2019. Formation des étudiants de l'ÉSPÉ de l'Académie de Nice (site de Stéphen Liégeard) à l'animation d'activités informatiques et mathématiques. En collaboration avec l'Inspection Académique. Dans le cadre de la semaine des mathématiques. Graphes, algorithmes, jeux et magie des mathématiques.
- Public exhibitions (Futurs en Seine,...)
 - 24/05/2019. Présentation du projet Galéjade à l'AFINEF'Tour Sophia Antipolis (avec EducAzur).
 - 18/05/2019. Semaine du Numérique à Valbonne. Algorithmes pour toute la famille.
 - 27-28/04/2019. Atelier et conférences au festival Les Souffleurs d'Avenir. La magie des graphes, des algorithmes et du binaire.
- In educational institutions
 - 11/01/2019. Conférence au lycée Léonard de Vinci d'Antibes (deux classes de terminale). Pas besoin de réfléchir, les ordinateurs calculent tellement vite ? Théorie des graphes et algorithmique pour les réseaux ? Présentation du métier de chercheur.

- 11/10/2019. Deux conférences au collège la Chênaie à Mouans Sartoux. La magie des graphes et du binaire ? Jeux combinatoires.
- Année scolaire 2019/2020. Projets dans le cadre de MATH.en.JEANS au collège Jean Rostand de Nice. Avec Olivier Ginola.
- 03/10/2019. Deux conférences et ateliers au collège de la Vésubie-Jean Salines à Roquebille. Fête de la Science 2019. La magie des graphes et du binaire ? Algorithmes grandeur nature ? Pas besoin de réfléchir, les ordinateurs calculent tellement vite ? Théorie des graphes et algorithmique pour les réseaux.
- 26/03/2019. Conférence au collège Jules Romains de Nice (classes de sixième). La magie des graphes, des algorithmes et du binaire.
- 13/03/2019. Conférences et ateliers au collège Alphonse Daudet de Nice (deux classes de sixième et deux classes de cinquième). Magie mathématique et jeux combinatoires.
- 05/04/2019. Coordination d'ateliers Galejade animés par des enseignants et des parents d'élèves de l'école Pasteur du Cannet (suite à une formation Galéjade quelques mois plus tôt). Ateliers Galejade.
- Welcoming of schoolchildren or the general public in an Inria center: MathC2+ internship, open days,...
 - Semaine du 16 décembre 2019. Accueil de 12 stagiaires de troisième à Inria Sophia Antipolis - Méditerranée.
 - 28/06/2019. Activité pour une quarantaine d'étudiants de classes préparatoires scientifiques de Sophia Antipolis. Algorithmes de tri.
 - 14/06/2019. Activité pour une quarantaine de lycéens des Alpes-Maritimes (accueillis à Inria Sophia Antipolis - Méditerranée durant 4 jours dans le cadre du stage MathC2+). Algorithmes grandeur nature.
 - 26/04/2019. Visite d'une dizaine de jeunes de la mission locale de Grasse à Inria Sophia Antipolis - Méditerranée. La magie des graphes, des algorithmes et du binaire.

7.3.5. Internal action

- Internal meetings such as Café des sciences
 - 29/04/2019. Café-in à Inria Sophia Antipolis - Méditerranée (séminaire interne pour tous les membres du centre de recherche : chercheurs, ingénieurs, administratifs...). Ateliers Galejade.
- Training of colleagues on new contents or media (activités débranchées, Poppy Ergo, TensorFlow,...)
- Training follow-up (Media Training, new media such as Poppy ergo, with SIF, etc.)
- Science outreach towards services (DPEI, STIP...)

7.3.6. Creation of media or tools for science outreach

See <https://galejade.inria.fr> and <http://terra-numerica.org/>.

8. Bibliography

Major publications by the team in recent years

- [1] F. CAZALS, P. KORNPBST (editors). *Modeling in Computational Biology and Medicine: A Multidisciplinary Endeavor*, Springer, 2013 [DOI : 10.1007/978-3-642-31208-3], <http://hal.inria.fr/hal-00845616>

- [2] D. AGARWAL, J. ARAUJO, C. CAILLOUET, F. CAZALS, D. COUDERT, S. PÉRENNES. *Connectivity Inference in Mass Spectrometry based Structure Determination*, in "European Symposium on Algorithms (Springer LNCS 8125)", Sophia Antipolis, France, H. BODLAENDER, G. ITALIANO (editors), Springer, 2013, pp. 289–300, <http://hal.inria.fr/hal-00849873>
- [3] D. AGARWAL, C. CAILLOUET, D. COUDERT, F. CAZALS. *Unveiling Contacts within Macro-molecular assemblies by solving Minimum Weight Connectivity Inference Problems*, in "Molecular and Cellular Proteomics", 2015, vol. 14, pp. 2274–2282 [DOI : 10.1074/MCP.M114.047779], <https://hal.archives-ouvertes.fr/hal-01078378>
- [4] J. CARR, D. MAZAURIC, F. CAZALS, D. J. WALES. *Energy landscapes and persistent minima*, in "The Journal of Chemical Physics", 2016, vol. 144, n^o 5, 4 p. [DOI : 10.1063/1.4941052], <https://www.repository.cam.ac.uk/handle/1810/253412>
- [5] F. CAZALS, F. CHAZAL, T. LEWINER. *Molecular shape analysis based upon the Morse-Smale complex and the Connolly function*, in "ACM SoCG", San Diego, USA, 2003, pp. 351-360
- [6] F. CAZALS, T. DREYFUS. *The Structural Bioinformatics Library: modeling in biomolecular science and beyond*, in "Bioinformatics", 2017, vol. 7, n^o 33, pp. 1–8 [DOI : 10.1093/BIOINFORMATICS/BTW752], <http://sbl.inria.fr>
- [7] F. CAZALS, T. DREYFUS, D. MAZAURIC, A. ROTH, C. ROBERT. *Conformational Ensembles and Sampled Energy Landscapes: Analysis and Comparison*, in "J. of Computational Chemistry", 2015, vol. 36, n^o 16, pp. 1213–1231 [DOI : 10.1002/JCC.23913], <https://hal.archives-ouvertes.fr/hal-01076317>
- [8] F. CAZALS, T. DREYFUS, S. SACHDEVA, N. SHAH. *Greedy Geometric Algorithms for Collections of Balls, with Applications to Geometric Approximation and Molecular Coarse-Graining*, in "Computer Graphics Forum", 2014, vol. 33, n^o 6, pp. 1–17 [DOI : 10.1111/CGF.12270], <http://hal.inria.fr/hal-00777892>
- [9] T. DREYFUS, V. DOYE, F. CAZALS. *Assessing the Reconstruction of Macro-molecular Assemblies with Toleranced Models*, in "Proteins: structure, function, and bioinformatics", 2012, vol. 80, n^o 9, pp. 2125–2136
- [10] T. DREYFUS, V. DOYE, F. CAZALS. *Probing a Continuum of Macro-molecular Assembly Models with Graph Templates of Sub-complexes*, in "Proteins: structure, function, and bioinformatics", 2013, vol. 81, n^o 11, pp. 2034–2044 [DOI : 10.1002/PROT.24313], <http://hal.inria.fr/hal-00849795>
- [11] N. MALOD-DOGNIN, A. BANSAL, F. CAZALS. *Characterizing the Morphology of Protein Binding Patches*, in "Proteins: structure, function, and bioinformatics", 2012, vol. 80, n^o 12, pp. 2652–2665
- [12] S. MARILLET, P. BOUDINOT, F. CAZALS. *High Resolution Crystal Structures Leverage Protein Binding Affinity Predictions*, in "Proteins: structure, function, and bioinformatics", 2015, vol. 1, n^o 84, pp. 9–20 [DOI : 10.1002/PROT.24946], <https://hal.inria.fr/hal-01159641>
- [13] A. ROTH, T. DREYFUS, C. ROBERT, F. CAZALS. *Hybridizing rapidly growing random trees and basin hopping yields an improved exploration of energy landscapes*, in "J. Comp. Chem.", 2016, vol. 37, n^o 8, pp. 739–752 [DOI : 10.1002/JCC.24256], <https://hal.inria.fr/hal-01191028>

Publications of the year

Articles in International Peer-Reviewed Journals

- [14] J.-C. BERMOND, A. CHAINTREAU, G. DUOFFE, D. MAZAURIC. *How long does it take for all users in a social network to choose their communities?*, in "Discrete Applied Mathematics", 2019, vol. 270, pp. 37-57 [DOI : 10.1016/J.DAM.2019.07.023], <https://hal.inria.fr/hal-02264327>
- [15] F. CAZALS, D. MAZAURIC, R. TETLEY, R. WATRIGANT. *Comparing Two Clusterings Using Matchings between Clusters of Clusters*, in "ACM Journal of Experimental Algorithmics", December 2019, vol. 24, n^o 1, pp. 1-41 [DOI : 10.1145/3345951], <https://hal.inria.fr/hal-02425599>
- [16] F. CAZALS, R. TETLEY. *Characterizing molecular flexibility by combining least root mean square deviation measures*, in "Proteins - Structure, Function and Bioinformatics", March 2019, vol. 87, n^o 5, pp. 380-389 [DOI : 10.1002/PROT.25658], <https://hal.inria.fr/hal-02425596>
- [17] C. O. SORZANO, A. JIMÉNEZ, J. MOTA, J. L. VILAS, D. MALUENDA, M. MARTÍNEZ, E. RAMÍREZ-APORTELA, T. MAJTNER, J. SEGURA, R. SÁNCHEZ-GARCÍA, Y. RANCEL, L. DEL CAÑO, P. CONESA, R. MELERO, S. JONIC, J. VARGAS, F. CAZALS, Z. FREYBERG, J. KRIEGER, I. BAHAR, R. MARABINI, J. M. CARAZO. *Survey of the analysis of continuous conformational variability of biological macromolecules by electron microscopy*, in "Acta crystallographica. Section F, Structural biology communications", January 2019, vol. 75, n^o 1, pp. 19-32 [DOI : 10.1107/S2053230X1801510], <https://hal.inria.fr/hal-01968170>

International Conferences with Proceedings

- [18] F. CAZALS, A. LHÉRITIER. *Low-Complexity Nonparametric Bayesian Online Prediction with Universal Guarantees*, in "NeurIPS 2019 - Thirty-third Conference on Neural Information Processing Systems", Vancouver, Canada, December 2019, <https://hal.inria.fr/hal-02425602>

Research Reports

- [19] A. CHEVALLIER, F. CAZALS. *Wang-Landau Algorithm: an adapted random walk to boost convergence*, Inria Sophia Antipolis, France, November 2019, n^o RR-9223, pp. 1-34, <https://hal.archives-ouvertes.fr/hal-01919860>
- [20] F. HAVET, D. MAZAURIC, V.-H. NGUYEN, R. WATRIGANT. *Overlaying a hypergraph with a graph with bounded maximum degree*, Inria Sophia Antipolis, February 2019, n^o RR-9258, <https://hal.inria.fr/hal-02025469>

Other Publications

- [21] J.-C. BERMOND, D. MAZAURIC, V. MISRA, P. NAIN. *Distributed Link Scheduling in Wireless Networks*, January 2019, working paper or preprint, <https://hal.inria.fr/hal-01977266>

References in notes

- [22] F. ALBER, S. DOKUDOVSKAYA, L. VEENHOFF, W. ZHANG, J. KIPPER, D. DEVOS, A. SUPRAPTO, O. KARNI-SCHMIDT, R. WILLIAMS, B. CHAIT, M. ROUT, A. SALI. *Determining the architectures of macromolecular assemblies*, in "Nature", Nov 2007, vol. 450, pp. 683-694

- [23] F. ALBER, S. DOKUDOVSKAYA, L. VEENHOFF, W. ZHANG, J. KIPPER, D. DEVOS, A. SUPRAPTO, O. KARNI-SCHMIDT, R. WILLIAMS, B. CHAIT, A. SALI, M. ROUT. *The molecular architecture of the nuclear pore complex*, in "Nature", 2007, vol. 450, n^o 7170, pp. 695–701
- [24] F. ALBER, F. FÖRSTER, D. KORKIN, M. TOPF, A. SALI. *Integrating Diverse Data for Structure Determination of Macromolecular Assemblies*, in "Ann. Rev. Biochem.", 2008, vol. 77, pp. 11.1–11.35
- [25] O. BECKER, A. D. MACKERELL, B. ROUX, M. WATANABE. *Computational Biochemistry and Biophysics*, M. Dekker, 2001
- [26] A.-C. CAMPROUX, R. GAUTIER, P. TUFFERY. *A Hidden Markov Model derived structural alphabet for proteins*, in "J. Mol. Biol.", 2004, pp. 591-605
- [27] M. L. CONNOLLY. *Analytical molecular surface calculation*, in "J. Appl. Crystallogr.", 1983, vol. 16, n^o 5, pp. 548–558
- [28] R. DUNBRACK. *Rotamer libraries in the 21st century*, in "Curr Opin Struct Biol", 2002, vol. 12, n^o 4, pp. 431-440
- [29] A. FERNANDEZ, R. BERRY. *Extent of Hydrogen-Bond Protection in Folded Proteins: A Constraint on Packing Architectures*, in "Biophysical Journal", 2002, vol. 83, pp. 2475-2481
- [30] A. FERSHT. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Freeman, 1999
- [31] M. GERSTEIN, F. RICHARDS. *Protein geometry: volumes, areas, and distances*, in "The international tables for crystallography (Vol F, Chap. 22)", M. G. ROSSMANN, E. ARNOLD (editors), Springer, 2001, pp. 531–539
- [32] H. GOHLKE, G. KLEBE. *Statistical potentials and scoring functions applied to protein-ligand binding*, in "Curr. Op. Struct. Biol.", 2001, vol. 11, pp. 231-235
- [33] J. JANIN, S. WODAK, M. LEVITT, B. MAIGRET. *Conformations of amino acid side chains in proteins*, in "J. Mol. Biol.", 1978, vol. 125, pp. 357–386
- [34] V. K. KRIVOV, M. KARPLUS. *Hidden complexity of free energy surfaces for peptide (protein) folding*, in "PNAS", 2004, vol. 101, n^o 41, pp. 14766-14770
- [35] 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
- [36] I. MIHALEK, O. LICHTARGE. *On Itinerant Water Molecules and Detectability of Protein-Protein Interfaces through Comparative Analysis of Homologues*, in "JMB", 2007, vol. 369, n^o 2, pp. 584–595
- [37] J. MINTSERIS, B. PIERCE, K. WIEHE, R. ANDERSON, R. CHEN, Z. WENG. *Integrating statistical pair potentials into protein complex prediction*, in "Proteins", 2007, vol. 69, pp. 511–520

-
- [38] M. PETTINI. *Geometry and Topology in Hamiltonian Dynamics and Statistical Mechanics*, Springer, 2007
- [39] 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", 2007, vol. 67, n^o 4, pp. 897–907
- [40] D. RAJAMANI, S. THIEL, S. VAJDA, C. CAMACHO. *Anchor residues in protein-protein interactions*, in "PNAS", 2004, vol. 101, n^o 31, pp. 11287-11292
- [41] 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", 2005, vol. 102, n^o 1, pp. 57-62
- [42] F. RICHARDS. *Areas, volumes, packing and protein structure*, in "Ann. Rev. Biophys. Bioeng.", 1977, vol. 6, pp. 151-176
- [43] G. RYLANCE, R. JOHNSTON, Y. MATSUNAGA, C.-B. LI, A. BABA, T. KOMATSUZAKI. *Topographical complexity of multidimensional energy landscapes*, in "PNAS", 2006, vol. 103, n^o 49, pp. 18551-18555
- [44] G. SCHREIBER, L. SERRANO. *Folding and binding: an extended family business*, in "Current Opinion in Structural Biology", 2005, vol. 15, n^o 1, pp. 1–3
- [45] M. SIPPL. *Calculation of Conformational Ensembles from Potential of Mean Force: An Approach to the Knowledge-based prediction of Local Structures in Globular Proteins*, in "J. Mol. Biol.", 1990, vol. 213, pp. 859-883
- [46] C. SUMMA, M. LEVITT, W. DEGRADO. *An atomic environment potential for use in protein structure prediction*, in "JMB", 2005, vol. 352, n^o 4, pp. 986–1001
- [47] S. WODAK, J. JANIN. *Structural basis of macromolecular recognition*, in "Adv. in protein chemistry", 2002, vol. 61, pp. 9–73