IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Joseph Fourier (Grenoble)

Activity Report 2013

Team NANO-D

Algorithms for Modeling and Simulation of Nanosystems

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Numerical schemes and simulations

Table of contents

Team NANO-D

Keywords: Modeling, Simulation, Nanosystems, Adaptive Algorithm, Knowledge-based Algorithms

Creation of the Team: 2008 January 01.

1. Members

Research Scientists

Stephane Redon [Team leader, Inria, Researcher]
Sergey Grudinin [CNRS, Researcher]
Leonard Jaillet [Inria, Starting Research position, from Mar 2013]

Engineers

Maria Werewka [Inria, Graphics Designer, from Sep 2013 until Nov 2013]
Svetlana Artemova [Inria, granted by ANR project, from January 2013]
Mael Bosson [Inria, granted by European Research Council, until Mar 2013]
Jocelyn Gate [Inria, granted by European Research Council, from Nov 2013]
Marc Piuzzi [Inria, granted by ANR project, from Feb 2013]
Mohamed Yengui [Inria, from Oct 2013]

PhD Students

Petr Popov [Inria, granted by ANR project, from January 2013]
Krishna Kant Singh [Inria, granted by Conseil Regional Rhone-Alpes, from Sep 2013]
Zofia Trstanova [Inria, granted by European Research Council, from Jun 2013]

Administrative Assistant

Maria Immaculada Presseguer [Inria]

Others

Astha Agarwal [Inria, intern, from May 2013 until Jul 2013]
Alexandre Hoffmann [Inria, intern, from Apr 2013 until Aug 2013]
Himani Singhal [Inria, intern, from May 2013 until Jul 2013]

2. Overall Objectives

2.1. Overview

During the twentieth century, the development of macroscopic engineering has been largely stimulated by progress in numerical design and prototyping: cars, planes, boats, and many other manufactured objects are nowadays designed and tested on computers. Digital prototypes have progressively replaced actual ones, and effective computer-aided engineering tools have helped cut costs and reduce production cycles of these macroscopic systems.

The twenty-first century is most likely to see a similar development at the atomic scale. Indeed, the recent years have seen tremendous progress in nanotechnology - in particular in the ability to control matter at the atomic scale. The nanoscience revolution is already impacting numerous fields, including electronics and semiconductors, textiles, energy, food, drug delivery, chemicals, materials, the automotive industry, aerospace and defense, medical devices and therapeutics, medical diagnostics, etc. According to some estimates, the world market for nanotechnology-related products and services will reach one trillion dollars by 2015. Nano-engineering groups are multiplying throughout the world, both in academia and in the industry: in the USA, the MIT has a “NanoEngineering” research group, Sandia National Laboratories created a “National Institute for Nano Engineering”, to name a few; China founded a “National Center for Nano Engineering” in 2003, etc. Europe is also a significant force in public funding of nanoscience and nanotechnology.

Similar to what has happened with macroscopic engineering, powerful and generic computational tools will be employed to engineer complex nanosystems, through modeling and simulation.

Modeling and simulation of natural or artificial nanosystems is still a challenging problem, however, for at least three reasons: (a) the number of involved atoms may be extremely large (liposomes, proteins, viruses, DNA, cell membrane, etc.); (b) some chemical, physical or biological phenomena have large durations (e.g., the folding of some proteins); and (c) the underlying physico-chemistry of some phenomena can only be described by quantum chemistry (local chemical reactions, isomerizations, metallic atoms, etc.). The large cost of modeling and simulation constitutes a major impediment to the development of nanotechnology.

The NANO-D team aims at developing efficient computational methods for modeling and simulation of complex nanosystems, both natural (e.g., the ATPase engine and other complex molecular mechanisms found in biology) and artificial (e.g., NEMS - Nano Electro-Mechanical Systems).

In particular, the group develops novel multiscale, adaptive modeling and simulation methods, which automatically focus computational resources on the most relevant parts of the nanosystems under study.

2.2. Research axes

The goal of the NANO-D group is to help current and future designers of *nanosystems*, i.e. systems studied or designed at the atomic scale (whether natural or artificial, independently of the application domain, including structural biology, material science, chemistry, etc.) by developing the **foundations of a software application which will run on a desktop computer, and will allow for efficient analysis, design, modeling and simulation of nanosystems**.

To achieve this, we will be developing a series of **adaptive methods and algorithms** that allow users to focus computational resources on the parts of the models that they want to simulate, and that allow to finely trade between speed and precision.

In parallel, we will develop the architecture of a new desktop application for virtual prototyping of nanosystems, and will integrate all our algorithms into this application. Furthermore, the architecture of this platform will be open, so that independent developers may add modules, for **multiple application domains** (physics, biology, chemistry, materials, electronics, etc.). With this open platform, we will attempt to federate the research performed in computational nanoscience throughout the world.

This application is called **SAMSON: "Software for Adaptive Modeling and Simulation Of Nanosystems"**.

Our two research axes are:

1. **Developing adaptive algorithms for simulating nanosystems**
   - **Defining adaptive Hamiltonians**: In order to be able to perform simulations with good mathematical properties, we are expanding on our recent work on *adaptively restrained Hamiltonians*[11], *i.e.* modified Hamiltonian representations of molecular systems that are able to switch degrees of freedom on and off during a simulation. These will allow us to finely trade between precision and computational performance, by choosing arbitrarily the number of degrees of freedom. Even though we have already obtained some promising results in this domain, our goal is to develop several different simplification methods.
   - **Developing algorithms for incremental potential update**: In order to benefit from performing adaptive particle simulations, we need to develop a series of algorithms that will take advantage of the fact that some (potentially relative) atomic positions are frozen. We have already demonstrated how this is possible for torsion-angle quasi-static simulation of classical bio-molecular force-fields [16], for neighbor search between large rigid molecules [10], and for bond-order reactive force-fields [13]. We are developing new algorithms for incremental neighbor search, energy and force updates corresponding to the adaptive Hamiltonians that we are defining.
2. **Developing algorithms for modeling molecular interactions**

- **Developing knowledge-driven methods, potentials and algorithms**: Over time, more and more experimental information becomes available. One can use this information to predict and discover new types of molecular interactions and various mechanisms or molecular organization. For example, currently there are more than 50,000 protein structures of a high resolution stored in the Protein Data Bank [12] and over 500,000 structures of small molecules stored in the Cambridge Structural Database [9]. We are developing algorithms for protein-protein interactions and protein-ligand interactions.
- **Developing parametrization algorithms for interaction potentials**: Molecular models typically require their own potential energy function (or a *forcefield*) to be assigned. However, the development of a new potential function is a very difficult and sometimes challenging task [14]. Therefore, we are developing algorithms for automatic parametrization of new potential functions for some particular representations of a molecular system.
- **Developing algorithms for exhaustive sampling**: Some application domains, such as computational docking, cryo-EM rigid-body fitting, etc., require sampling in a low-dimensional space. For such applications it is advantageous to perform an exhaustive search rather than accelerated sampling [15]. Therefore, we are developing fast search methods to perform exhaustive search.

3. Application Domains

3.1. Overview

NANO-D is *a priori* concerned with all applications domains involving atomistic representations, including chemistry, physics, electronics, material science, biology, etc.

Historically, though, our first applications have been in biology, as the next two sections detail. Thanks to the development of algorithms to efficiently simulate reactive force fields, as well as to perform interactive quantum mechanical calculations, however, we now have the possibility to address problems in chemistry, and physics.

3.2. Structural Biology

Structural biology is a branch of molecular biology, biochemistry, and biophysics concerned with the molecular structure of biological macromolecules, especially proteins and nucleic acids. Structural biology studies how these macromolecules acquire the structures they have, and how alterations in their structures affect their function. The methods that structural biologists use to determine the structure typically involve measurements on vast numbers of identical molecules at the same time, such as X-Ray crystallography, NMR, cryo-electron microscopy, etc. In many cases these methods do not directly provide the structural answer, therefore new combinations of methods and modeling techniques are often required to advance further.

We develop a set of tools that help biologists to model structural features and motifs not resolved experimentally and to understand the function of different structural fragments.

- Symmetry is a frequent structural trait in molecular systems. For example, most of the water-soluble and membrane proteins found in living cells are composed of symmetrical subunits, and nearly all structural proteins form long oligomeric chains of identical subunits. Only a limited number of symmetry groups is allowed in crystallography, and thus, in many cases the native macromolecular conformation is not present on high-resolution X-ray structures. Therefore, to understand the realistic macromolecular packing, modeling techniques are required.
- Many biological experiments are rather costly and time-demanding. For instance, the complexity of mutagenesis experiments grows exponentially with the number of mutations tried simultaneously. In other experiments, many candidates are tried to obtain a desired function. For example, about 250,000 candidates were tested for the recently discovered antibiotic Platensimycin. Therefore, there is a vast need in advance modeling techniques that can predict interactions and foresee the function of new structures.

- Structure of many macromolecules is still unknown. For other complexes, it is known only partially. Thus, software tools and new algorithms are needed by biologists to model missing structural fragments or predict the structure of those molecule, where there is no experimental structural information available.

3.3. Pharmaceutics and Drug Design

Drug design is the inventive process of finding new medications based on the knowledge of the biological target. The drug is most commonly an organic small molecule which activates or inhibits the function of a biomolecule such as a protein, which in turn results in a therapeutic benefit to the patient. In the most basic sense, drug design involves design of small molecules that are complementary in shape and charge to the biomolecular target to which they interact and therefore will bind to it. Drug design frequently relies on computer modeling techniques. This type of modeling is often referred to as computer-aided drug design.

Structure-based drug design attempts to use the structure of proteins as a basis for designing new ligands by applying accepted principles of molecular recognition. The basic assumption underlying structure-based drug design is that a good ligand molecule should bind tightly to its target. Thus, one of the most important principles for designing or obtaining potential new ligands is to predict the binding affinity of a certain ligand to its target and use it as a criterion for selection.

We develop new methods to estimate the binding affinity using an approximation to the binding free energy. This approximation is assumed to depend on various structural characteristics of a representative set of native complexes with their structure solved to a high resolution. We study and verify different structural characteristics, such as radial distribution functions, and their affect on the binding free energy approximation.

3.4. Nano-engineering

The magazine Science has recently featured a paper demonstrating an example of DNA nanotechnology, where DNA strands are stacked together through programmable self-assembly. In February 2007, the cover of Nature Nanotechnology showed a “nano-wheel” composed of a few atoms only. Several nanosystems have already been demonstrated, including a wheelbarrow molecule, a nano-car and a Morse molecule, etc. Typically, these nanosystems are designed in part *via* quantum mechanics calculations, such as the semi-empirical ASED+ calculation technique.

Of course, not all small systems that currently fall under the label “nano” have mechanical, electronic, optical properties similar to the examples given above. Furthermore, current construction capabilities lack behind some of the theoretical designs which have been proposed. However, the trend is clearly for adding more and more functionality to nanosystems. While designing nanosystems is still very much an art mostly performed by physicists, chemists and biologists in labs throughout the world, there is absolutely no doubt that fundamental engineering practices will progressively emerge, and that these practices will be turned into quantitative rules and methods. Similar to what has happened with macroscopic engineering, powerful and generic software will then be employed to engineer complex nanosystems.

We have recently shown that our incremental and adaptive algorithms allow us to easily edit and model complex shapes, such as a nanotube (Fig. 1) and the “nano-pillow” below (Fig. 2).

4. Software and Platforms

4.1. SAMSON

A major objective of NANO-D is to try and integrate a variety of adaptive algorithms into a unified framework. As a result, NANO-D is developing SAMSON (Software for Adaptive Modeling and Simulation Of Nanosystems), a software platform aimed at including all developments from the group, in particular those described below.

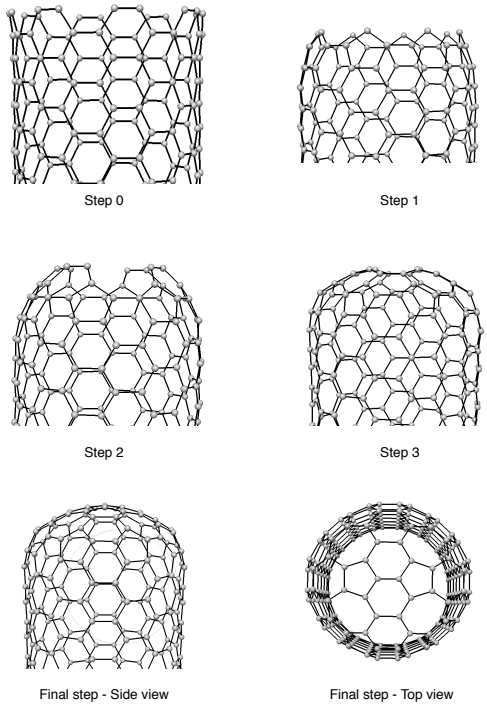

Figure 1. Snapshots of a nanotube capping process with the adaptive interactive modeler. Thanks to the adaptive methodology, this operation can be done in a few minutes.

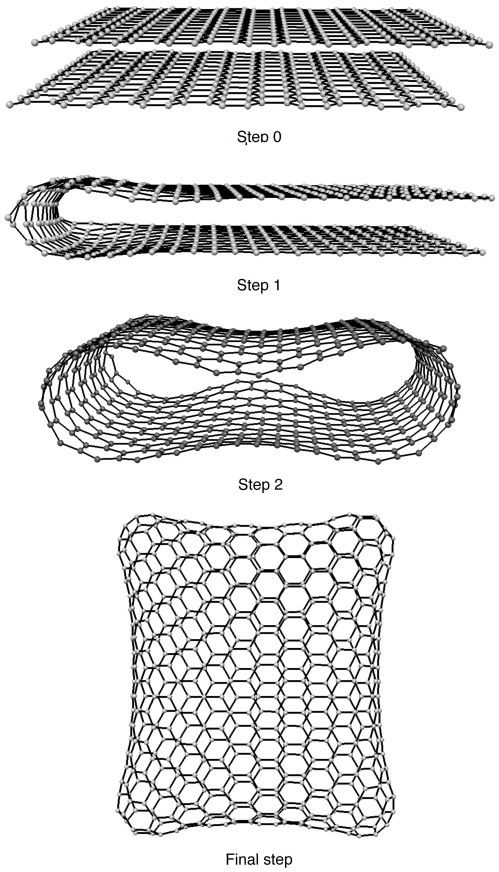

Figure 2. Different steps to prototype a "nano-pillow" with the adaptive interactive modeler.

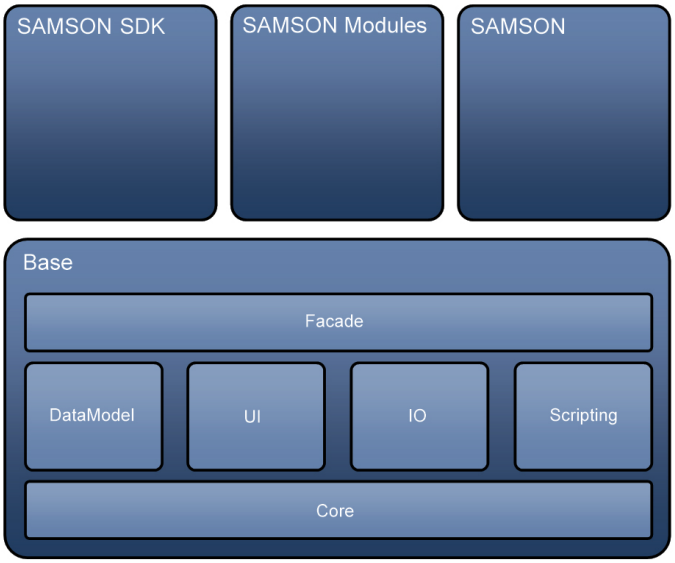

Figure 3. SAMSON's architecture.

The objective is to make SAMSON a generic application for computer-aided design of nanosystems, similar to existing applications for macrosystem prototyping (CATIA, SolidWorks, etc.).

The current architecture of SAMSON is visible in Figure 3. The code is organized into four main parts: a) the Base (in which "Core" contains, in particular, the heart of the adaptive algorithms: signaling mechanisms specifically designed for SAMSON), b) the Software Development Kit (SDK: a subset of the base that will be provided to module developers), c) Modules, and d) the SAMSON application itself.

Similar to the concept of Mathematica *toolboxes*, for example, the goal has been to make it possible to personalize the user interface of SAMSON for potentially many distinct applications. For example, we may want to personalize the interface of SAMSON for crystallography, drug design, protein folding, electronics, material science, nano-engineering, etc., by loading different modules at startup, depending on the user application domain.

5. New Results

5.1. Adaptively Restrained Particle Simulations for Isobaric-Isothermal Ensemble

Participants: Zofia Trstanova, Stephane Redon.

We continued working on the Adaptively Restrained Particles Simulations (ARPS) approach that was proposed by Svetlana Artemova and Stephane Redon [11] and that was designed to speed up the particles simulations by switching on and off the degrees of freedom based on the kinetic energy of the particle.

It has been shown, for the NVE and the NVT ensemble, that this method has many advantages [11]. We want to extend ARPS for the isobaric-isothermal ensemble (NPT) since this ensemble is very often used in particle simulations, because many chemical reactions happen under constant pressure. An adaptive method for this ensemble with advantages of ARPS might be very useful in many scientific domains (physics, biology, chemistry).

We combined the ARPS method with an existing method that describes the NPT ensemble. We already obtained very promising analytical and numerical results that support the main characteristically advantages of ARPS shown by Svetlana Artemova and Stephane Redon. For instance, Figure 4 shows preservation of the radial distribution function.

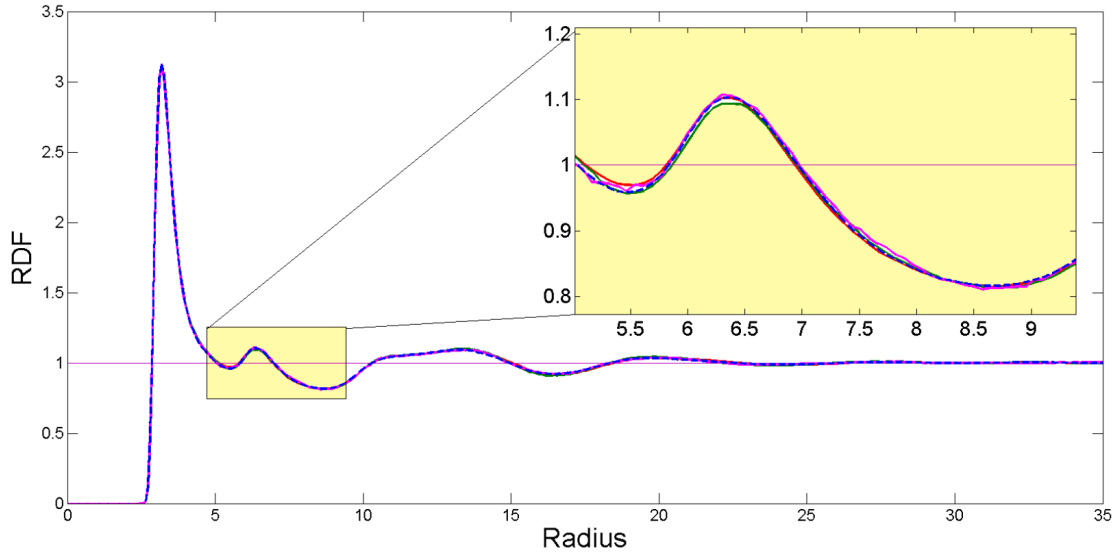

Figure 4. Radial distribution function obtained with different ARPS simulations compared to the full dynamics (blue dash line)

5.2. Interactive large-scale deformations of molecular structures

Participants: Jelmer Wolterink, Himani Singhal, Marc Piuzzi, Stephane Redon.

We have developed new interaction methods for large-scale deformation of molecular structures. These new methods allow a user to attach control points to molecules, and use these control points to easily deform the structures while preserving their realism (e.g. local interactions, etc.). The new methods may be applied to any type of molecule (e.g. proteins, carbon nanotubes, etc.), and may be used in combination with interactive simulation.

5.3. Towards parallel adaptive molecular simulations

Participants: Krishna Kant Singh, Benjamin Bouvier, Jean-Francois Mehaut, Stephane Redon.

The adaptive algorithms that we are developing have two main components. The first component determines when and how degrees of freedom can be deactivated and reactivated during a simulation. The second component takes advantage of the frozen degrees of freedom to accelerate the calculation of the potential energy and interatomic forces. Indeed, the potential energy and forces can often be expressed as a (potentially complex) sum of terms which only depend on relative atomic positions. When the relative positions do not change, it is not necessary to update the corresponding terms, which reduces the computation time. We have shown that it is possible to significantly speed up simulations using this approach, while being able to recover static equilibrium statistics [11].

We have now begun to study the possibility of developing adaptive *parallel* simulation algorithms, and have begun to review and benchmark popular simulation packages (GROMACS, NAMD, OpenMM, etc.), depending on the number of atoms, the number of available cores, etc.

5.4. Protein secondary structure prediction for dynamic simulations

Participants: Marc Piuzzi, Sergei Grudinin, Stephane Redon.

There is a tight link between a protein's function and its molecular structure. Hence, global stability is essential for a protein to keep its role inside the cell. Various chemical interactions helps stabilizing the structure (covalent bonds, hydrogen bonds, etc.) but not all parts of a protein present the same stability. The most stable regions of a protein present numerous hydrogen bonds on backbone atoms composing geometrically distinguishable secondary structures (the primary structure being the amino acid sequence): helices and beta sheets.

These structures have been well studied and although important properties have been defined, there is no absolute definition of what is a helix or a beta sheet. Thus, various methods have been developed to predict the secondary structure of a protein using the amino acid sequence and/or the protein structure using different parameters and structural descriptors.

However, none of these methods have been made in the context of interactive simulation where the shape of the protein is dynamic: here the prediction has to be done at each time step on the whole protein. Moreover, the result is deterministic and returns only the type of structure without any information about the accuracy. We are developing a new approach that is appropriate in an interactive context, where secondary structure assignment has to continuously change during interaction.

5.5. Motion Planning for Quasi-Static Simulation

Participants: Leonard Jaillet, Stephane Redon.

Recently, motion planning methods inspired from Robotics have been applied to the study of biological molecular systems [8]. These approaches rely on compact graph representations that aim to capture large amplitude motions more efficiently than classic simulation techniques, despite their lower resolution.

We developed within the SAMSON's architecture a new motion planning strategy to perform quasi-static simulation at the nano-scale.

The user provides as inputs the initial and final state of the system he or she wants to simulate. Then, the method searches a transition path that follows the low-energy valleys of the conformational landscape (see figure 5).

The adaptation of motion planning approaches to quasi-static simulation at the nano-scale comes with several challenges. First, these approaches must be adapted to tackle the high dimensionality involved in the case of nanosystems, dimensionality that is directly related to the number of atoms considered. Second, these approaches must be extended to face the complexity of the underlying physics that comes from the various types of interactions between atoms.

The method we propose is able to perform simulations involving bonds breaking. This is, up to our knowledge, the first motion planning approach able to simulate chemical reactions.

5.6. Molecular Modeling

5.6.1. Rapid determination of RMSDs corresponding to macromolecular rigid body motions

Participants: Petr Popov, Sergei Grudinin.

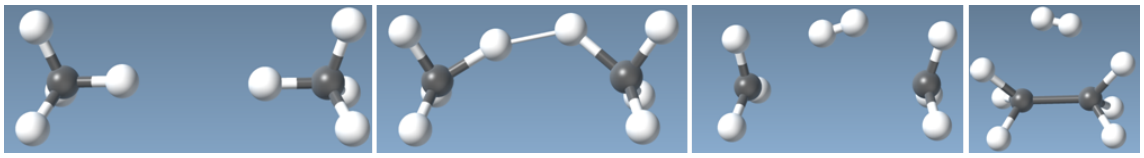

Figure 5. Snapshots of the transition path obtained with our motion planning simulation method. It represents a chemical reaction where two molecules of methanes interact to form a an ethane and a dihydrogen.

Finding the root mean sum of squared deviations (RMSDs) between two coordinate vectors that correspond to the rigid body motion of a macromolecule is an important problem in structural bioinformatics, computational chemistry and molecular modeling. Standard algorithms compute the RMSD with time proportional to the number of atoms in the molecule. We developed *RigidRMSD*, a new algorithm that determines a set of RMSDs corresponding to a set of rigid body motions of a macromolecule in constant time with respect to the number of atoms in a molecule. Our algorithm is particularly useful for rigid body modeling applications such as rigid body docking, and also for high-throughput analysis of rigid body modeling and simulation results. A C++ implementation of our algorithm will be available at http://nano-d.inrialpes.fr/software/RigidRMSD.

To demonstrate the efficiency of the RigidRMSD library, we compared the clustering application implemented with our algorithm to the one from the Hex software. We chose Hex for the comparison because it is a very fast rigid body docking tool and also because it explicitly provides the clustering time. For the comparison, we collected a small benchmark of 23 protein dimers of various size. After, we launched Hex version 6.3 on this benchmark and collected docking solutions before clustering, sizes of clusters, and clustering time. We then also clustered these solutions using the *RigidRMSD* library. Figure 6 shows the clustering time of the HEX clustering algorithm with respect to our clustering using two rotation representations as a function of the number of atoms in the smaller protein (left) and the number of docking solutions before the clustering (right). We can clearly see that our implementation of the clustering algorithm is more than an order of magnitude faster compared to the Hex implementation. Also, the quaternion representation of rotation is on average three times more efficient compared to the matrix representation.

5.6.2. *Fast fitting atomic structures into a low-resolution density map using 3D orthogonal Hermite functions*

Participants: Georgy Derevyanko, Sergei Grudinin.

We developed a new algorithm for fitting protein structures into a low resolution electron density (e.g. cryo-electron microscopy) map. The algorithm uses 3D orthogonal Hermite functions for fast operations on the electron density.

Orthogonal Hermite function of order n is defined as:

$$\psi_n(x;\lambda) = \frac{\sqrt{\lambda}}{\sqrt{2^n n! \sqrt{\pi}}} \exp\left(-\frac{\lambda^2 x^2}{2}\right) H_n(\lambda x), \tag{1}$$

where $H_n(x)$ is the Hermite polynomial and λ is the scaling parameter. In Fig. 7 we show several orthogonal Hermite functions of different orders with different parameters λ. These functions form an orthonormal basis set in $L^2(\mathbb{R})$. A 1D function $f(x)$ decomposed into the set of 1D Hermite functions up to an order N reads

$$f(x) = \sum_{i=0}^{N} \widehat{f}_i \psi_i(x;\lambda) \tag{2}$$

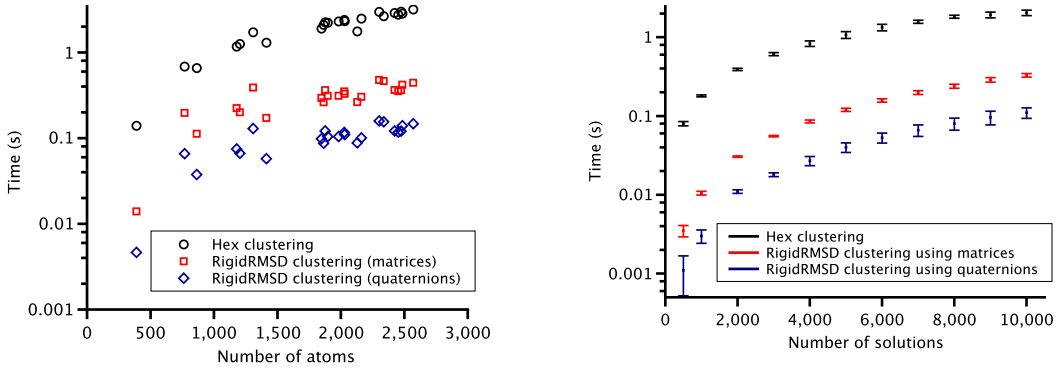

Figure 6. Left: Time spent on clustering by Hex and RigidRMSD with respect to the number of atoms in the ligand protein. Number of considered solutions and the RMSD threshold was fixed to 10,000 and 10.0 Å, respectively. Right: Average time spent on clustering by Hex and RigidRMSD with respect to the number of docking solutions. For this plot we chose five structures with the number of atoms of about 2,000 and plotted the standard deviation of the running time. For both plots, the RMSD threshold was fixed to 10.0 Å.

Here, $\widehat{f_i}$ are the decomposition coefficients, which can be determined from the orthogonality of the basis functions $\psi_i(x;\lambda)$. Decomposition in Eq. 2 is called the *band-limited decomposition* with $\psi_i(x;\lambda)$ basis functions. To decompose the electron density map and the protein structures, we employ the 3D Hermite functions:

$$\psi_{n,l,m}(x,y,z;\lambda) = \psi_n(x;\lambda)\psi_l(y;\lambda)\psi_m(z;\lambda), \tag{3}$$

which form an orthonormal basis set in $L^2\left(\mathbb{R}^3\right)$. A function $f(x,y,z)$ represented as a band-limited expansion in this basis reads

$$f(x,y,z) = \sum_{i=0}^{N}\sum_{j=0}^{N-i}\sum_{k=0}^{N-i-j}\widehat{f}_{i,j,k}\psi_{i,j,k}(x,y,z;\lambda) \tag{4}$$

Our algorithm accelerates rotation of the Fourier image of the electron density by using 3D orthogonal Hermite functions. As a part of the new method, we presented an algorithm for the rotation of the density in the Hermite basis and an algorithm for the conversion of the expansion coefficients into the Fourier basis. We implemented the program of fitting a protein structure to a low-resolution electron density map, which uses the cross-correlation or the Laplacian-filtered cross-correlation as the fitting criterion. We demonstrated that in the Hermite basis, the Laplacian filter has a particularly simple form. To assess the quality of density encoding in the Hermite basis, we uses two measures, the R-factor and the cross-correlation factor. Finally, we validated our algorithm using two examples and compare its efficiency with two widely used fitting methods, ADP_EM and *colores* from the Situs package.

5.6.3. Fast Rotational-Translation Matching of Rigid Bodies by Fast Fourier Transform Acceleration of Six Degrees of Freedom

Participants: Alexandre Hoffmann, Sergei Grudinin.

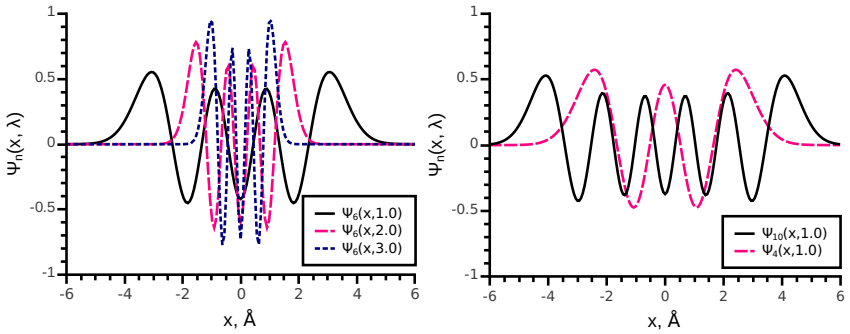

Figure 7. Left: 1D Hermite functions of order 6 for three different scaling parameters λ. Right: 1D Hermite functions of two different orders for the scaling parameter $\lambda = 1$.

We introduced a new method for rigid molecular fitting. This problem is usually solved by maximizing the Cross Corelation Function (CCF), which is computed using the Fast Fourier Transform (FFT) algorithm. Our method handles six degrees of freedom at once and requires only one computation of the Cross Corelation Function, with the six-dimensional Fast Fourier Transform. Our method only requires a low pre-processing time ($O(N^7)$), which is comparable to the cost of the subsequent 6D FFT ($O(N^6 \log(N^6))$). It also uses a dual Hermite-Fourier representation, which allows to represent a small molecule with a fewer number of coefficients in the Hermite basis.

5.6.4. Prediction of complexes with point group symmetry using spherical polar Fourier docking correlations

Participants: David W. Ritchie, Sergei Grudinin.

Many proteins form symmetric homo-oligomers that perform a certain physiological function. We present the first point group symmetry docking algorithm that generates perfectly symmetrical protein complexes for arbitrary point group symmetry types (C_n, D_n, T, O, and I). We validate the algorithm on proteins from the 3D-Complex database, where it achieves on average the success rate of 55%. The running time of the algorithm is less than a minute on a modern workstation.

Many of the protein complexes in the protein Data bank (PDB) are symmetric homo-oligomers. According to the 3D-Complex database, C_2 homo-dimers comprise the majority of known homo-oligomers. However, many complexes have higher order rotational symmetry (i.e. C_n>2), and a good number have multiple rotational symmetry axes, namely those with dihedral (D_n), tetrahedral (T), octahedral (O), and icosahedral (I) point group symmetries. Although symmetrical complexes are often solved directly by X-ray crystallography, it would still be very useful to be able to predict whether or not a given monomer might self-assemble into a symmetrical structure. We present a new point group symmetry docking algorithm. In the last few years, several protein-protein docking programs have been adapted to predict symmetrical pair-wise docking orientations for C_n and D_n symmetries. However, to our knowledge, there does not yet exist an algorithm which can automatically generate perfectly symmetrical protein complexes for arbitrary point group symmetry types.

We introduce the notion of a "docking equation" in which the notation $A(\underline{x}) \longleftrightarrow B(\underline{x})$ represents an interaction between proteins A and B in 3D space. It is also useful introduce the operators $\widehat{T}(x, y, z)$ and $\widehat{R}(\alpha, \beta, \gamma)$, which represent the actions of translating an object by an amount (x, y, z) and rotating it according to the three Euler rotation angles (α, β, γ). Then, guided by Figure 8, and assuming that we start with two identical monomers at the origin, we can write down a C_n docking equation for the two monomers as

$$\widehat{T}(0,y,0)\widehat{R}(\alpha,\beta,\gamma)A(\underline{x}) \longleftrightarrow \widehat{R}(0,0,\omega)\widehat{T}(0,y,0)\widehat{R}(\alpha,\beta,\gamma)B(\underline{x}). \tag{5}$$

Then, we perform a series of fast Fourier transform (FFT) correlation searches using the Hex spherical polar Fourier docking algorithm to determine the four parameters (y,α,β,γ). For higher symmetries, D_n, T, O, and I, we introduce two more parameters and perform a series of FFT in a similar way. The calculation for each structure takes less than a minute on a modern workstation.

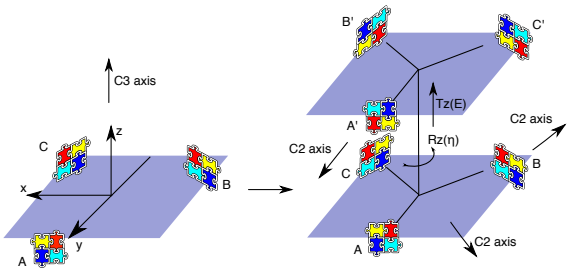

Figure 8. Illustrations of the C_3 and D_3 point group symmetries.

We validated our method on protein structures from the 3D-Complex database, which contains 17,183 protein complexes with assigned biological unit and symmetry type. It mostly contains cyclic and dihedral proteins, as well as 86 tetrahedral, 47 octahedral, and 6 icosahedral complexes (excluding all viral structures). Starting from the structures of monomers, we generated symmetric biological units based on the symmetry type for each complex provided by 3D-Complex. Figure 9 summarizes the performance of our method on these proteins, where we say that the model is correct if all pair-wise RMSDs are smaller than 10 Ångstroms. On average, we found about 55% of correct predictions ranked first.

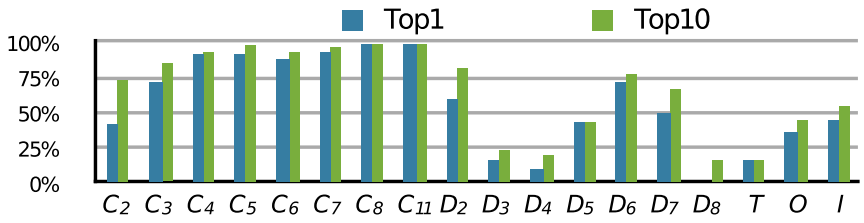

Figure 9. Summary of the correctly predicted complexes found on the first place (blue) and in the top ten solutions (green).

Figure 10 shows correctly predicted examples from each of the symmetry types. Each complex is perfectly symmetrical, although due to the soft docking function in Hex it is possible that some interfaces might contain minor steric clashes.

5.7. Software Engineering

5.7.1. SAMSON User interface

Participants: Jocelyn Gate, Maria Werewka, Stephane Redon.

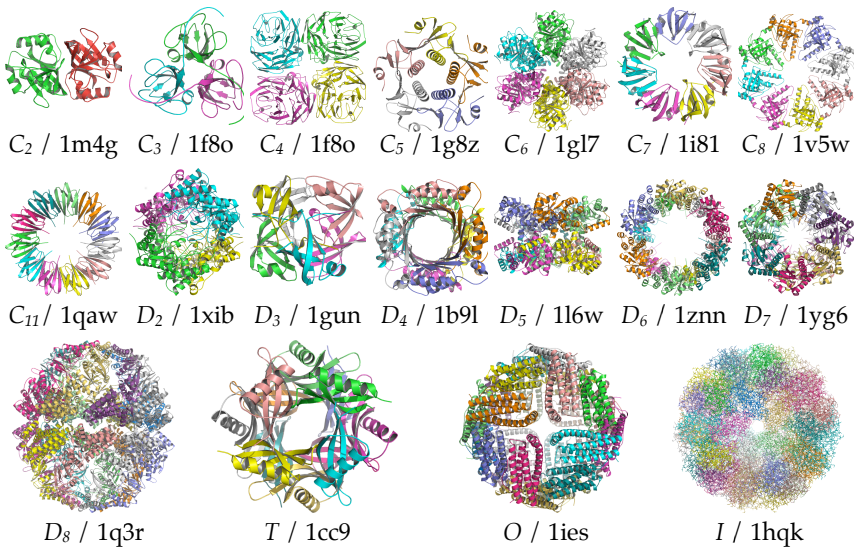

Figure 10. Illustrations of the correctly predicted complexes. For each complex, the group symbol and the PDB code are shown.

We have continued the development of SAMSON, our open-architecture platform for modeling and simulation of nanosystems (SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems):

- We have moved to Qt5 to handle the Graphical User Interface.
- We now compile SAMSON is 64 bits only. This removes limitations of 32 bits applications, in particular concerning memory limits.
- We now have a complete installer mechanism for both users and developers.
- We extended the set of development tools (action generators, UUID generators, etc.)
- We have changed the windowing system to allow windows to move outside SAMSON.
- We have designed a coherent style for icons, windows, menus, etc.
- We have added 3D rendering.

There are now more than 40 modules in SAMSON (parsers, editors, models, apps, etc.).

The current user interface of SAMSON is visible in Figure 11.

5.7.2. SAMSON Elements

Participants: Svetlana Artemova, Stephane Redon.

We have added new SAMSON Elements (modules).

We have been working on input and output for SAMSON. Precisely, we now have the possibility to download molecules to SAMSON and save them to external files in three possible formats:

- pdb (Protein Data Bank format, containing experimentally determined 3d structures and widely used for applications in biology);
- mol2 (Sybyl chemical modeler input file, containing chemical compounds and small ligands);
- xyz (basic format, containing atoms coordinates).

Basic properties of atoms, residues, and molecules have been determined and structures storing these properties were implemented.

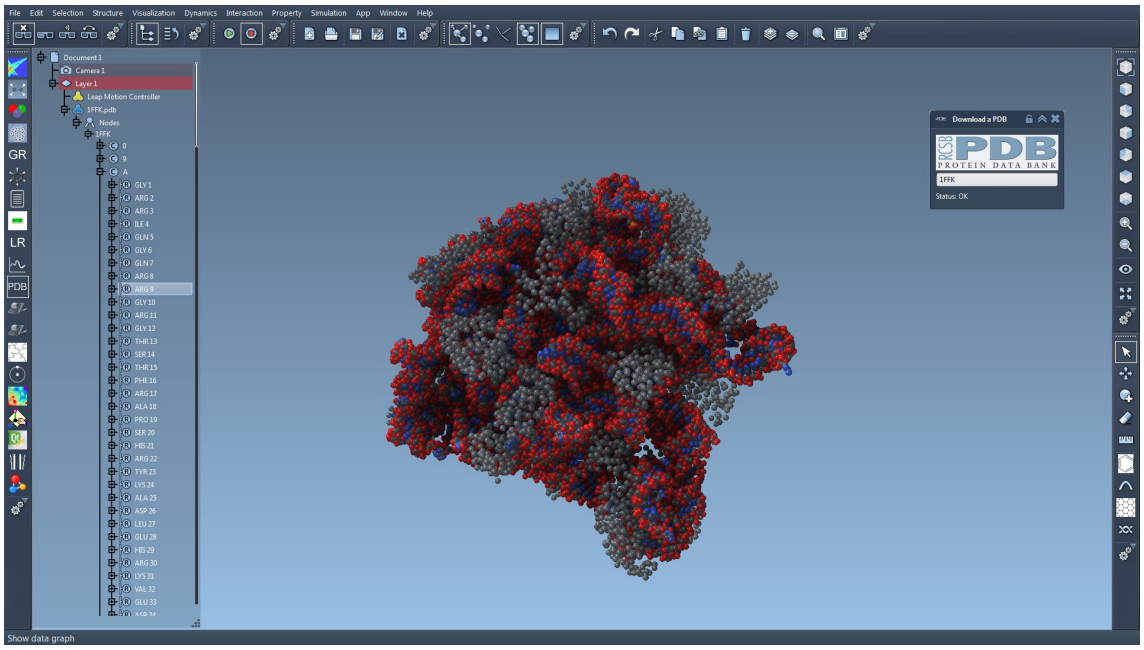

Figure 11. The current user interface of SAMSON showing an app to download molecules directly from the Protein Data Bank. The data graph on the left shows the hierarchical structure of the structural model.

Finally, since energy minimization is crucial for providing physically-correct structures while interactively editing molecules in SAMSON, we have implemented several fast and stable algorithms to perform such energy minimization in SAMSON.

5.7.3. SAMSON Website

Participants: Mohamed Yengui, Jocelyn Gate, Stephane Redon.

We are developing a web application aiming at distributing and valorizing SAMSON and SAMSON Elements (modules). The goal of the website is to develop a community of users and developers in all areas of nanoscience (physics, biology, chemistry, electronics, etc.). The website will:

- allow users and developers to create and manage accounts on the website.
- allow visualizing, searching and downloading SAMSON and SAMSON Elements.
- allow the creation, validation and dissemination of SAMSON Elements.
- provide tracking requests for the arrival of new SAMSON Elements or the modification of an existing SAMSON Element.

To achieve this, we have designed the architecture in a way that speeds development effort for a faster product release, while keeping in mind scalability, security and high reliability.

We have also implemented and tested locally the account validation process. A user can now sign up, confirm the registration from the received email and authenticate with the registered account to download SAMSON and SAMSON elements from the website. We will make the site public when we release SAMSON.

6. Partnerships and Cooperations

6.1. Regional Initiatives

- **ARC 2012**: This grant from the Rhone-Alpes region (http://www.arc.rhonealpes.fr/) has been provided to S. Redon, Jean-François Méhaut (LIG - Laboratoire d'Informatique de Grenoble) and Benjamin Bouvier (IBCP - Institut de Biologie et Chimie des Proteines) to develop adaptive, parallel algorithms for molecular simulation. The grants is for a PhD student.

6.2. National Initiatives

6.2.1. ANR

In 2013, NANO-D had funding from four ANR programs:

- **ANR Jeunes Chercheurs Jeunes Chercheuses (JCJC)**: 340,000 Euros over three years (2011-2014). This grant has been provided to S. Redon by the French Research Agency for being a finalist in the ERC Starting Grant 2009 call, and is for two PhD students and an engineer.
- **ANR Modeles Numeriques (MN)**: 180,000 Euros over four years (2011-2015). This project, coordinated by NANO-D (S. Grudinin), gathers biologists and computer scientists from three research groups: Dave Ritchie at LORIA, Valentin Gordeliy at IBS (total grant: 360,000 Euros).
- **ANR PIRIBio**: 25,000 Euros over four years (2010-2013). We are participating in this project coordinated by Michel Vivaudou at IBS, with Serge Crouzy at CEA/LCBM and Frank Fieschi at IBS.

6.2.2. PEPS

Sergei Grudinin participates in the Cryo-CA PEPS project. Cryo-CA (Computational algorithms for biomolecular structure determination by cryo-electron microscopy) is a 2-years project, supported by the Projets Exploratoires Pluridisciplinaires (PEPS) program in the panel Bio-Maths-Info provided by CNRS (French National Centre for Scientific Research). The project started on the 01/09/2012. Its main goal is to develop computational algorithms for cryo-electron microscopy (cryo-EM).

The partners of the Cryo-CA project are: Inria Nancy / Team Orpailleur (David Ritchie); Inria Grenoble / Team NANO-D (Sergei Grudinin); and INSERM IGBMC/ Team Integrated structural Biology (Annick Dejaegere, Patrick Schultz, and Benjamin Schwarz).

The main scientific aim of this cross-disciplinary project is to develop computational algorithms to help experimentalists and molecular modelers to solve more rapidly and accurately the structures of macromolecular complexes using cryo-electron microscopy (cryo- EM) and integrative structural biomolecular modeling techniques. More specifically, this PEPS initiative aims to address two important challenges in single particle cryo-EM, namely particle picking and multi-dimensional structure fitting. In the longer term, a further driving aim of this project is to develop strong collaborations amongst the participating teams to position ourselves for a larger project proposal to ANR or ERC.

6.3. European Initiatives

6.3.1. FP7 Projects

6.3.1.1. ADAPT

Type: IDEAS

Defi: NC

Instrument: ERC Starting Grant

Objectif: Theory and algorithms for adaptive particle simulation

Duration: September 2012 - August 2017

Coordinator: Stephane Redon

Inria contact: Stephane Redon

6.4. International Initiatives

6.4.1. Inria International Partners

6.4.1.1. Informal International Partners

NANO-D has an ongoing collaboration with the research group of Pr. Dr. Markus Reiher in ETH Zürich, to develop interactive quantum chemistry methods assisted with haptic feedback.

6.5. International Research Visitors

6.5.1. Visits of International Scientists

- Pr. Dr. Markus Reiher, from ETH Zürich, visited NANO-D in January 2013
- Pr. Eric Polizzi, from the University of Massachusetts Amherst, visited NANO-D in March 2013
- PhD students Moritz Haag and Arndt Finkelmann, from the Reiher group at ETH Zürich, visited NANO-D in October 2013

6.5.1.1. Internships

Astha Agarwal

Subject: Development of a Coarse-Grained Potential Function for Protein Folding and De Novo Design

Date: from May 2013 until Jul 2013

Institution: IIT Bombay (India)

7. Dissemination

7.1. Teaching - Supervision - Juries

7.1.1. Teaching

Licence : Stephane Redon, “Introduction to computer science”, INF311, 80h, Ecole Polytechnique, Paris, France

7.1.2. Supervision

PhD in progress : Petr Popov, Computational methods for protein structure prediction, November 2011, Sergei Grudinin

PhD in progress : Krishna Kant Singh, Adaptive parallel algorithms for molecular simulation, Grenoble University, September 2013, Jean-François Méhaut and Stéphane Redon

PhD : Zofia Trstanova, Adaptive Hamiltonians, Grenoble University, May 30, 2012, Stephane Redon

7.2. Participation to conferences, seminars

- S. Grudinin gave a talk "Efficient Boundary Element Method with curved elements" at the workshop "Computational Electrostatics for Biological Applications, CEBA 2013", Jul 1-3, Genova, Italy, 2013.

- S. Grudinin gave a talk "Predicting Multi Protein Assemblies" at the Second workshop on Computational Structural Biology: Integrative Approaches for Modeling Biomolecular Complexes, IAMB 2013, May 29-31, Nice, France, 2013.
- M. Piuzzi participated in ISMB/ECCB in Berlin, 2013.
- M. Piuzzi and S. Grudinin participated in XVIIIe congrès du Groupe de Graphisme et Modélisation Molèculaire, GGMM 2013, May 21-23, St Pierre d'Oléron, France, 2013.
- S. Grudinin and P. Popov participated at the Fifth CAPRI Evaluation Meeting, Apr 17-19, Utrecht, Netherlands, 2013.
- P. Popov gave a talk "Docktrina : Docking Protein Trimers" at the Fifth CAPRI Evaluation Meeting, Apr 17-19, Utrecht, Netherlands, 2013.
- S. Grudinin attended a school on Analysis of Diffraction Data in Real Space, ADD2013, Mar 18-22, Grenoble, France, 2013.

7.3. Popularization

NANO-D presented SAMSON to high school students during the 2103 “Fete de la Science” (Science Fair). The students were using SAMSON to interactively simulate chemical reactions, interact with models of nanotubes, build molecules, etc.

8. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journals

[1] M. BOSSON, S. GRUDININ, S. REDON. *Block-Adaptive Quantum Mechanics: An Adaptive Divide-and-Conquer Approach to Interactive Quantum Chemistry*, in "Journal of Computational Chemistry", March 2013, vol. 34, n^{o} 6, pp. 492-504 [*DOI :* 10.1002/JCC.23157], http://hal.inria.fr/hal-00755521

[2] I. Y. GUSHCHIN, V. I. GORDELIY, S. GRUDININ. *Two Distinct States of the HAMP Domain from Sensory Rhodopsin Transducer Observed in Unbiased Molecular Dynamics Simulations*, in "PLoS ONE", July 2013, vol. 8, n^{o} 7 [*DOI :* 10.1371/JOURNAL.PONE.0066917], http://hal.inria.fr/hal-00881143

[3] A. ISHCHENKO, E. ROUND, V. BORSHCHEVSKIY, S. GRUDININ, I. Y. GUSHCHIN, J. P. KLARE, T. BALANDIN, A. REMEEVA, M. ENGELHARD, G. BÜLDT, V. I. GORDELIY. *Ground state structure of D75N mutant of sensory rhodopsin II in complex with its cognate transducer*, in "Journal of Photochemistry and Photobiology B: Biology", June 2013, vol. 123, pp. 55-58 [*DOI :* 10.1016/J.JPHOTOBIOL.2013.03.008], http://hal.inria.fr/hal-00881553

[4] M. LENSINK, I. MOAL, P. BATES, P. KASTRITIS, A. MELQUIOND, E. KARACA, C. SCHMITZ, M. VAN DIJK, A. BONVIN, M. EISENSTEIN, B. JIMÉNEZ-GARCÍ, S. GROSDIDIER, A. SOLERNOU, L. PÉREZ-CANO, C. PALLAR, J. FERNÁNDEZ-RECIO, J. XU, P. MUTHU, K. PRANEETH KILAMBI, J. GRAY, S. GRUDININ, G. DEREVYANKO, J. MITCHELL, J. WIETING, E. KANAMORI, Y. TSUCHIYA, Y. MURAKAMI, J. SARMIENTO, D. STANDLEY, M. SHIROTA, K. KINOSHITA, H. NAKAMURA, M. CHAVENT, D. RITCHIE, H. PARK, D. RITCHIE, J. KO, H. LEE, C. SEOK, Y. SHEN, D. KOZAKOV, S. VAJDA, P. KUNDROTAS, I. VAKSER, B. PIERCE, H. HWANG, T. VREVEN, Z. WENG, I. BUCH, E. FARKASH, H. WOLFSON, M. ZACHARIAS, S. QIN, H.-X. ZHOU, S.-Y. HUANG, X. ZOU, J. WOJDYLA, C. KLEANTHOUS, S. J. WODAK. *Blind prediction of interfacial water positions in CAPRI*, in "Proteins", October 2013 [*DOI :* 10.1002/PROT.24439], http://hal.inria.fr/hal-00880345

[5] R. MORETTI, S. J. FLEISHMAN, R. AGIUS, M. TORCHALA, P. A. BATES, P. L. KASTRITIS, J. P. G. L. M. RODRIGUES, M. TRELLET, A. M. J. J. BONVIN, M. CUI, M. ROOMAN, D. GILLIS, Y. DEHOUCK, I. MOAL, M. ROMERO-DURANA, L. PÉREZ-CANO, C. PALLARA, B. JIMÉNEZ-GARCÍ, J. FERNÁNDEZ-RECIO, S. C. FLORES, M. PACELLA, K. PRANEETH KILAMBI, J. J. GRAY, P. POPOV, S. GRUDININ, J. ESQUIVEL-RODRÍGUEZ, D. KIHARA, N. ZHAO, D. KORKIN, X. ZHU, O. N. A. DEMERDASH, J. MITCHELL, E. KANAMORI, Y. TSUCHIYA, H. NAKAMURA, H. LEE, H. PARK, C. SEOK, J. SARMIENTO, S. LIANG, S. TERAGUCHI, D. M. STANDLEY, H. SHIMOYAMA, G. TERASHI, M. TAKEDA-SHITAKA, M. IWADATE, H. UMEYAMA, D. BEGLOV, D. R. HALL, D. KOZAKOV, S. VAJDA, B. G. PIERCE, H. HWANG, T. VREVEN, Z. WENG, Y. HUANG, H. LI, X. YANG, X. JI, S. LIU, Y. XIAO, M. ZACHARIAS, S. QIN, H.-X. ZHOU, S.-Y. HUANG, X. ZOU, S. VELANKAR, J. JANIN, S. J. WODAK, D. BAKER. *Community-wide evaluation of methods for predicting the effect of mutations on protein-protein interactions*, in "Proteins", August 2013, vol. 81, n^{o} 11, pp. 1980 - 1987 [*DOI :* 10.1002/PROT.24356], http://hal.inria.fr/hal-00905794

[6] P. POPOV, D. RITCHIE, S. GRUDININ. *DockTrina: Docking triangular protein trimers*, in "Proteins: Structure, Function, and Genetics", January 2014, vol. 82, n^{o} 1, pp. 34-44 [*DOI :* 10.1002/PROT.24344], http://hal.inria.fr/hal-00880359

International Conferences with Proceedings

[7] P.-L. MANTEAUX, F. FAURE, S. REDON, M.-P. CANI. *Exploring the Use of Adaptively Restrained Particles for Graphics Simulations*, in "VRIPHYS 2013 - 10th Workshop on Virtual Reality Interaction and Physical Simulation", Lille, France, Eurographics Association, September 2013, http://hal.inria.fr/hal-00914653

References in notes

[8] I. AL-BLUWI, T. SIMÉON, J. CORTÉS. *Motion planning algorithms for molecular simulations: A survey*, in "Computer Science Review", 2012, vol. 6, n^{o} 4, pp. 125 - 143 [*DOI :* 10.1016/J.COSREV.2012.07.002], http://www.sciencedirect.com/science/article/pii/S157401371200024X

[9] F. H. ALLEN. *The Cambridge Structural Database: a quarter of a million crystal structures and rising*, in "Acta Crystallographica Section B", Jun 2002, vol. 58, n^{o} 3 Part 1, pp. 380–388, http://dx.doi.org/10.1107/S0108768102003890

[10] S. ARTEMOVA, S. GRUDININ, S. REDON. *A comparison of neighbor search algorithms for large rigid molecules*, in "Journal of Computational Chemistry", 2011, vol. 32, n^{o} 13, pp. 2865–2877, http://dx.doi.org/10.1002/jcc.21868

[11] S. ARTEMOVA, S. REDON. *Adaptively Restrained Particle Simulations*, in "Phys. Rev. Lett.", Nov 2012, vol. 109, http://link.aps.org/doi/10.1103/PhysRevLett.109.190201

[12] H. M. BERMAN, J. WESTBROOK, Z. FENG, G. GILLILAND, T. N. BHAT, H. WEISSIG, I. N. SHINDYALOV, P. E. BOURNE. *The Protein Data Bank*, in "Nucleic Acids Research", 2000, vol. 28, n^{o} 1, pp. 235-242 [*DOI :* 10.1093/NAR/28.1.235], http://nar.oxfordjournals.org/content/28/1/235.abstract

[13] M. BOSSON, S. GRUDININ, X. BOUJU, S. REDON. *Interactive physically-based structural modeling of hydrocarbon systems*, in "Journal of Computational Physics", 2012, vol. 231, n^{o} 6, pp. 2581 - 2598 [*DOI :* 10.1016/J.JCP.2011.12.006], http://www.sciencedirect.com/science/article/pii/S0021999111007042

[14] V. HORNAK, R. ABEL, A. OKUR, B. STROCKBINE, A. ROITBERG, C. SIMMERLING. *Comparison of multiple Amber force fields and development of improved protein backbone parameters*, in "Proteins: Structure, Function, and Bioinformatics", 2006, vol. 65, n^{o} 3, pp. 712–725, http://dx.doi.org/10.1002/prot.21123

[15] D. W. RITCHIE, G. J. KEMP. *Protein docking using spherical polar Fourier correlations*, in "Proteins: Structure, Function, and Bioinformatics", 2000, vol. 39, n^{o} 2, pp. 178–194, http://dx.doi.org/10.1002/(SICI)1097-0134(20000501)39:2<178::AID-PROT8>3.0.CO;2-6

[16] R. ROSSI, M. ISORCE, S. MORIN, J. FLOCARD, K. ARUMUGAM, S. CROUZY, M. VIVAUDOU, S. REDON. *Adaptive torsion-angle quasi-statics: a general simulation method with applications to protein structure analysis and design*, in "Bioinformatics", 2007, vol. 23, n^{o} 13 [*DOI :* 10.1093/BIOINFORMATICS/BTM191], http://bioinformatics.oxfordjournals.org/content/23/13/i408.abstract