IN PARTNERSHIP WITH:
CNRS

Institut polytechnique de Grenoble

Université Joseph Fourier (Grenoble)

Activity Report 2015

Project-Team NANO-D

Algorithms for Modeling and Simulation of Nanosystems

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Numerical schemes and simulations

Table of contents

Project-Team NANO-D

Creation of the Team: 2008 January 01, updated into Project-Team: 2014 July 01

Keywords:

Computer Science and Digital Science:

6. - Modeling, simulation and control
6.2. - Scientific Computing, Numerical Analysis & Optimization
8.2. - Machine learning

Other Research Topics and Application Domains:

1. - Life sciences
1.1. - Biology
2. - Health
5.3. - Nanotechnology and Biotechnology
5.5. - Materials

1. Members

Research Scientists

Stephane Redon [Team leader, Inria, Researcher]
Sergei Grudinin [CNRS, Researcher]
Leonard Jaillet [Inria, Starting Research position]

Engineers

Svetlana Artemova [Inria, granted by ERC (European Research Council Excecutive Agency)]
Marc Aubert [Inria, until Oct 2015]
Mohamed Nadhir Ben Hadj Abdellatif [Fondation nanosciences, until Oct 2015]
Jocelyn Gate [Inria, granted by ERC (European Research Council Excecutive Agency)]
Mohamed Yengui [Inria, until May 2015]

PhD Students

Semeho Edorh [Inria]
Alexandre Hoffmann [Univ. Grenoble I]
Minh Khoa Nguyen [Inria]
Petr Popov [Inria, until Feb 2015, granted by ANR PEPSI project]
Francois Rousse [Inria]
Krishna Kant Singh [Inria, granted by Conseil Regional Rhone-Alpes]
Zofia Trstanova [Inria, granted by ERC (European Research Council Excecutive Agency)]

Post-Doctoral Fellows

Silvia Dias Pinto [Inria, from Jul 2015]
Emilie Neveu [Inria]

Administrative Assistants

Julie Bourget [Inria]
Maria Immaculada Presseguer [Inria]

Others

Andreas Eisenbarth [Inria, Master student, from Nov 2015]
Aleksandr Katrutca [Inria, Master student, from Sep 2015]

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* [15], *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 [70], for neighbor search between large rigid molecules [14], and for bond-order reactive force-fields [19]. 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 [17] 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 [41]. 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 [66]. Therefore, we are developing fast search methods to perform exhaustive search.

3. Research Program

3.1. The need for practical design of nanosystems

Computing has long been an essential tool of engineering. 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, for the most part, designed and tested on computers. Digital prototypes have progressively replaced actual ones, and effective computer-aided engineering tools (e.g., CATIA, SolidWorks, T-FLEX CAD, Alibre Design, TopSolid, etc.) have helped cut costs and reduce production cycles of macroscopic systems [69].

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. The magazine Science, for example, recently featured a paper demonstrating an example of DNA nanotechnology, where DNA strands are stacked together through programmable self-assembly [32]. 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 *de-novo* computationally designed protein interface [33], a wheelbarrow molecule [44], a nano-car [73], a Morse molecule [10], etc. Typically, these designs are optimized using semi-empirical quantum mechanics calculations, such as the semi-empirical ASED+ calculation technique [11].

While impressive, these are but two examples of the nanoscience revolution already impacting numerous fields, including electronics and semiconductors [52], textiles [51], [37], energy [56], food [23], drug delivery [35], [77], chemicals [38], materials [24], the automotive industry [8], aerospace and defense [34], medical devices and therapeutics [28], medical diagnostics [79], etc. According to some estimates, the world market for nanotechnology-related products and services will reach one trillion dollars by 2015 [68]. 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 and, in Europe, Grenoble and the Rhone-Alpes area gather numerous institutions and organizations related to nanoscience.

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, such as the planetary gear designed by Eric Drexler at Nanorex. 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.

3.2. Challenges of practical nanosystem design

As with macrosystems, designing nanosystems will involve modeling and simulation within software applications: modeling, especially structural modeling, will be concerned with the creation of potentially complex chemical structures such as the examples above, using a graphical user interface, parsers, scripts, builders, etc.; simulation will be employed to predict some properties of the constructed models, including mechanical properties, electronic properties, chemical properties, etc.

In general, design may be considered as an "inverse simulation problem". Indeed, designed systems often need to be optimized so that their properties — predicted by simulation — satisfy specific objectives and constraints (e.g. a car should have a low drag coefficient, a drug should have a high affinity and selectivity to a target protein, a nano-wheel should roll when pushed, etc.). Being the main technique employed to predict properties, simulation is essential to the design process. At the nanoscale, simulation is even more important. Indeed, physics significantly constrains atomic structures (e.g. arbitrary inter-atomic distances cannot exist), so that a tentative atomic shape should be checked for plausibility much earlier in the design process (e.g. remove atomic clashes, prevent unrealistic, high-energy configurations, etc.). For nanosystems, thus, efficient simulation algorithms are required both when modeling structures and when predicting systems properties. Precisely, an effective software tool to design nanosystems should (a) allow for interactive physically-based modeling, where all user actions (e.g. displacing atoms, modifying the system's topology, etc.) are automatically followed by a few steps of energy minimization to help the user build plausible structures, even for large number of atoms, and (b) be able to predict systems properties, through a series of increasingly complex simulations.

3.3. Current simulation approaches

Even though the growing need for effective nanosystem design will still increase the demand for simulation, a lot of research has already gone into the development of efficient simulation algorithms. Typically, two approaches are used: (a) increasing the computational resources (use super-computers, computer clusters, grids, develop parallel computing approaches, etc.), or (b) simulating simplified physics and/or models. Even though the first strategy is sometimes favored, it is expensive and, it could be argued, inefficient: only a few supercomputers exist, not everyone is willing to share idle time from their personal computer, etc. Surely, we would see much less creativity in cars, planes, and manufactured objects all around if they had to be designed on one of these scarce super-resources.

The second strategy has received a lot of attention. Typical approaches to speed up molecular mechanics simulation include lattice simulations [82], removing some degrees of freedom (e.g. keeping torsion angles only [50], [75]), coarse-graining [80], [71], [12], [72], multiple time step methods [64], [65], fast multipole methods [29], parallelization [45], averaging [22], multi-scale modeling [21], [18], reactive force fields [20], [86], interactive multiplayer games for predicting protein structures [27], etc. Until recently, quantum mechanics methods, as well as mixed quantum / molecular mechanics methods were still extremely slow. One breakthrough has consisted in the discovery of linear-scaling, divide-and-conquer quantum mechanics methods [83], [84].

Overall, the computational community has already produced a variety of sophisticated simulation packages, for both classical and quantum simulation: ABINIT, AMBER, CHARMM, Desmond, GROMOS and GROMACS, LAMMPS, NAMD, ROSETTA, SIESTA, TINKER, VASP, YASARA, etc. Some of these tools are open source, while some others are available commercially, sometimes via integrating applications: Ascalaph Designer, BOSS, Discovery Studio, Materials Studio, Maestro, MedeA, MOE, NanoEngineer-1, Spartan, etc. Other tools are mostly concerned with visualization, but may sometimes be connected to simulation packages: Avogadro, PyMol, VMD, Zodiac, etc. The nanoHUB network also includes a rich set of tools related to computational nanoscience.

To the best of our knowledge, however, all methods which attempt to speed up dynamics simulations perform a priori simplification assumptions, which might bias the study of the simulated phenomenon. A few recent, interesting approaches have managed to combine several levels of description (e.g. atomistic and coarse-grained) into a single simulation, and have molecules switch between levels during simulation, including the adaptive resolution method [60], [61], [62], [63], the adaptive multiscale method [55], and the adaptive partitioning of the Lagrangian method [39]. Although these approaches have demonstrated some convincing applications, they all suffer from a number of limitations stemming from the fact that they are either ad hoc methods tuned to fix specific problems (e.g. fix density problems in regions where the level of description changes), or mathematically founded methods that necessitate to "calibrate" potentials so that they can be mixed (i.e. all potentials have to agree on a reference point). In general, multi-scale methods, even when they do not allow molecules to switch between levels of detail during simulation, have to solve the problem of rigorously combining multiple levels of description (i.e. preserve statistics, etc.), of assigning appropriate levels to different parts of the simulated system ("simplify as much as possible, but not too much"), and of determining computable mappings between levels of description (especially, adding back detail when going from coarse-grained descriptions to fine-grained descriptions).

3.4. 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*[15], *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 [70], for neighbor search between large rigid molecules [14], and for bond-order reactive force-fields [19]. 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 [17] 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 [41]. 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 [66]. Therefore, we are developing fast search methods to perform exhaustive search.

4. Application Domains

4.1. 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.

4.2. 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.

4.3. 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.

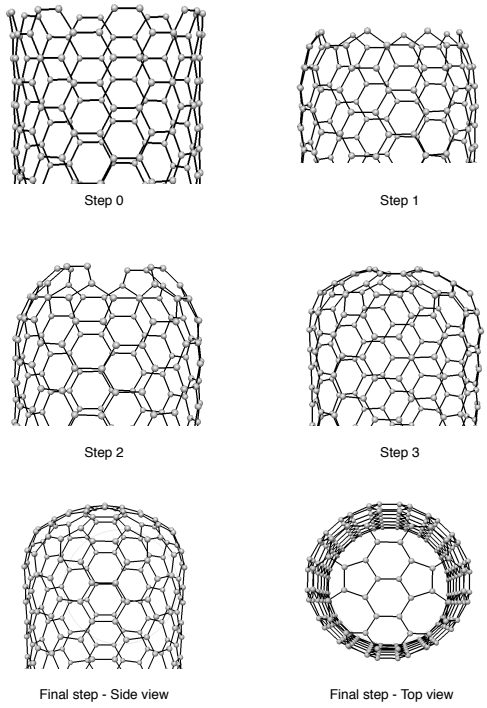

Figure 1. Snapshots of a nanotube capping process with the adaptive interactive modeler. Thanks to the adaptive methodology, this operation can be done in a few minutes.

4.4. Nano-engineering

In general, we want to develop methods to ease nano-engineering of artificial nanosystems, such as the ones described above (DNA nanotechnology, nano-mechanisms, etc.). We have shown, for example, 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). Please read more about the SAMSON software platform for more examples.

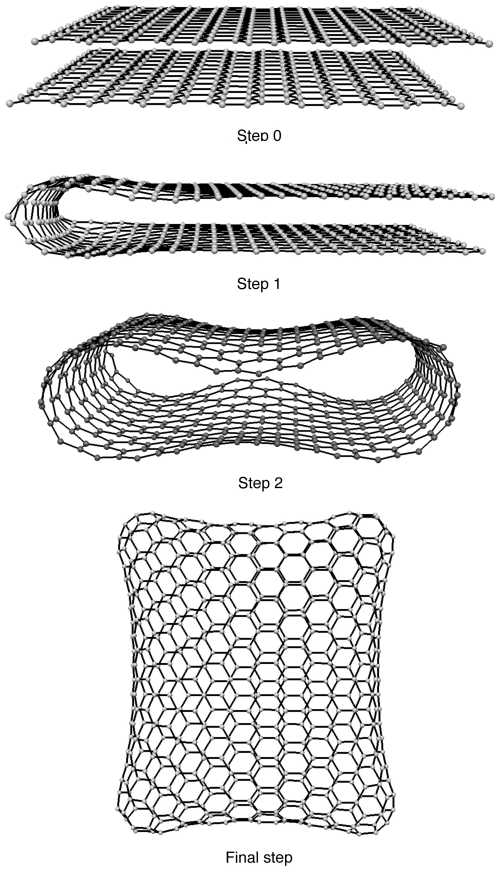

Figure 2. Different steps to prototype a "nano-pillow" with the adaptive interactive modeler.

5. Highlights of the Year

5.1. Highlights of the Year

We have released the first version of the SAMSON software platform for computational nanoscience on the SAMSON Connect website (http://www.samson-connect.net). Using the SAMSON Connect website, users may download SAMSON and choose which SAMSON Elements (modules for SAMSON) to add to their

configuration (e.g. a nanotube creator, for users interested in materials science). Developers may download the SAMSON Software Development Kit (SDK) to develop SAMSON Elements and upload them to the SAMSON Connect website. We are frequently releasing updates, on Windows, Linux, and Mac (Figure 3).

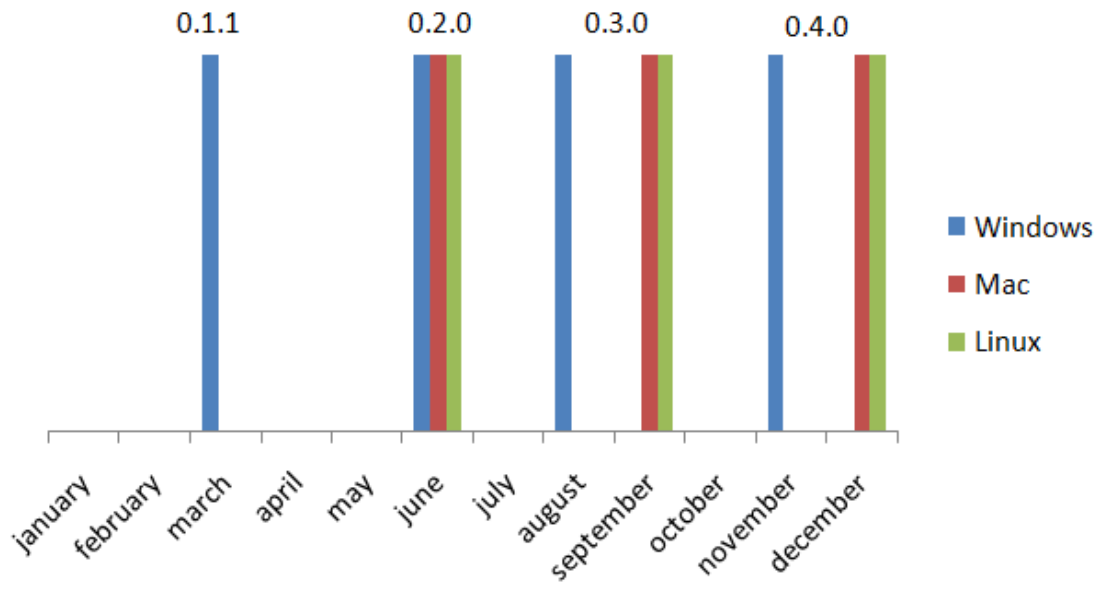

Figure 3. Release dates of the various versions of SAMSON

6. New Software and Platforms

6.1. SAMSON

SAMSON (Software for Adaptive Modeling and Simulation Of Nanosystems) is a software platform for computational nanoscience. SAMSON has a modular architecture that makes it suitable for different domains of nanoscience, including material science, life science, physics, electronics, chemistry, and education.

SAMSON Elements are modules for SAMSON, developed with the SAMSON Software Development Kit (SDK). SAMSON Elements help users perform tasks in SAMSON, including building new models, performing calculations, running interactive or offline simulations, and visualizing and interpreting results.

SAMSON Elements may contain different class types, including for example:

- Apps - generic classes with a graphical user interface that extend the functions of SAMSON
- Editors - classes that receive user interaction events to provide editing functions (e.g., model generation, structure deformation, etc.)
- Models - classes that describe properties of nanosystems (see below)
- Parsers - classes that may parse files to add content to SAMSON's data graph (see below)

SAMSON Elements expose their functions to SAMSON and other Elements through an introspection mechanism, and may thus be integrated and pipelined.

SAMSON represents nanosystems using five categories of models:

- Structural models - describe geometry and topology
- Visual models - provide graphical representations
- Dynamical models - describe dynamical degrees of freedom
- Interaction models - describe energies and forces
- Property models - describe traits that do not enter in the first four model categories

Simulators (potentially interactive ones) are used to build physically-based models, and predict properties. All models and simulators are integrated into a hierarchical, layered structure that form the SAMSON data graph. SAMSON Elements interact with each other and with the data graph to perform modeling and simulation tasks. A signals and slots mechanism makes it possible for data graph nodes to send events when they are updated, which makes it possible to develop e.g., adaptive simulation algorithms.

SAMSON is developed in C++ and implements many features to ease development of SAMSON Elements, including:

- Managed memory
- Signals and slots
- Serialization
- Multilevel undo-redo
- Introspection
- Referencing
- Unit system
- SAMSON Element source code generators

SAMSON, SAMSON Elements and the SAMSON Software Development Kit are distributed via the SAMSON Connect website (http://www.samson-connect.net). The site acts as a repository for the SAMSON Elements being uploaded by developers, and users of SAMSON choose and add Elements from SAMSON Connect.

7. New Results

7.1. Algorithms for Orbital-Free Density Functional Theory

Participants: Francois Rousse, Stephane Redon.

The Schrödinger equation permits, in theory, to model and simulate every molecular systems exactly. Unfortunately it is not computationally doable to solve this equation even on really small systems (2 atoms). Density Functional Theory (DFT) gives a method to solve this equation, find the electronic structure and simulate molecules with the laws of physics on reasonably large system : from 1.000 to 10.000 depending on the basis chosen and the version of DFT used. Unfortunately, the computation of kinetic energy requires the orthogonalization of the basis, which consumes a lot of time and prevents the algorithm from being adaptive : one needs to recompute the whole system if a little change is done in the molecules position. One can deals with this issue by computing the kinetic energy directly with the electronic density and not anymore with the orbitals. That is the idea of Orbital-Free DFT (OF-DFT). It can models great systems (up to 1.000.000 atoms) and be turned adaptive. On the other hand, it looses a lot of accuracy and power to model different kind of systems on the other DFT.

We have already developed our own OF-DFT code. It runs on parallel cores, is implemented in the SAMSON platform as a SAMSON App and gives correct electron's densities. The electronic structures are computed in real space to preserve the possibility of incremental calculations. We are now going to test our implementation, and will then attempt to make the method adaptive. The difficulty will be the determination of the domain that needs to be recomputed when a part of the system has moved, and the criteria that will help to do so.

7.2. Parallel adaptively restrained particle simulations

Participants: Krishna Kant Singh, Stephane Redon.

We have continued our work on the development of parallel adaptively restrained particle simulations. We developed new algorithms for neighbor list and incremental force updates. These algorithms have advantages over the state-of-the-art methods for simulating a system using Adaptively Restrained Molecular Dynamics (ARMD). We have simulated systems with different number (500, 4000 and 108000) of LJ particles using adaptively restrained integrator and Lennard-Jones potential in NVE (constant number of particles, Volume and Energy) and NVT ensemble (constant number of particles, Volume and Temperature). All the particles were placed in an orthogonal box. We used periodic periodic boundary conditions with 8.5 angstrom cut-off for the Lennard-Jones potential. The system was simulated using 2 femtoseconds. We compared the LAMMPS algorithm to adaptive algorithm while using adaptively restrained integrator. Our results show that a significant speed-up can be achieved if more than 60% of the particles are restrained (Figure 4). Figure 5 shows that ARMD in NVT ensemble preserves the average temperature of the system (irrespective of number of restrained particles).

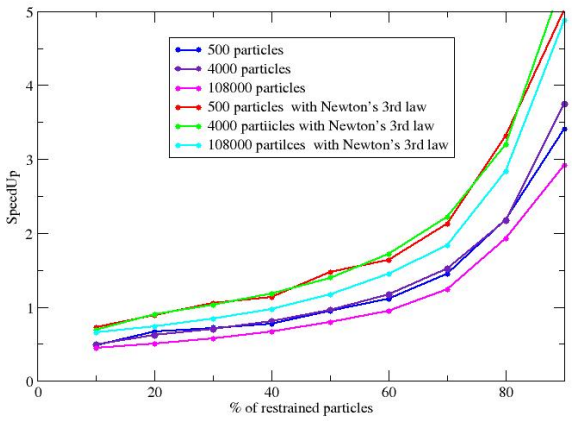

Figure 4. Speedup using ARMD on different benchmark

7.3. Incremental algorithms for long-range interactions

Participants: Semeho Edorh, Stephane Redon.

Numerical simulations of molecular dynamics (MD) are very expensive in terms of CPU resources. During Molecular dynamics simulations, the most CPU intensive task is the evaluation of the interaction potential [78]. Due to the large number of particles involved, updating this potential may have, at each time-step, a very high computational cost.

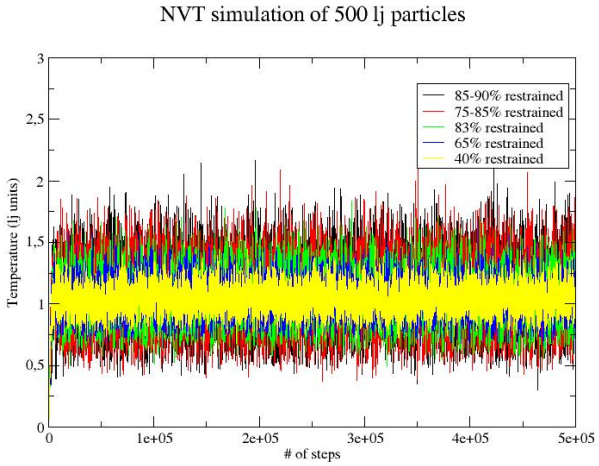

Figure 5. Temperature profile of 500 LJ particles in NVT ensemble using ARMD

In large crystalline ionic system, Ewald summation is the most popular method for computing electrostatic interactions. It rewrites the interaction potential ϕ as the sum of a short-range term and a long-range term. Ewald summation using optimal parameters requires $\mathcal{O}(N^{3/2})$ operations [47], [30] but it can be modified so that it involves only about $\mathcal{O}(N \log N)$ operations [31], [85] by using the Fast Fourier Transform.

We want to develop a new approach that can reduce the computational cost by using incremental algorithms. The key-idea is to use, for each time-step of the simulation, information that we have computed in previous steps.

The Particle Mesh Ewald (PME) algorithm developped by Darden et al. is the most successful approach for computing long range interactions. In the particle mesh method, just as in standard Ewald summation, the generic interaction potential is separated into two terms.The so-called short-range contribution can be easily calculated in a direct space by using truncation methods. Where as the long-range contribution is calculated using two Fast Fourier transforms ($N \log(N)$ algorithm). In pratice, the long range contribution algorithm boil than to [30]:

- Map particle charge density Q to a mesh
- Compute the forward Fast Fourier Transform of the approximation Qm of charge density on the mesh
- Multiply Qm by a green function (related to the choice of the mesh).
- Compute a backward Fast Fourier Transform of the result.
- Retrieve the long-range contribution potential by interpolating the previous result at particles positions.

We modified this algorithm to make it incremental. We started from the PME implementation in LAMMPS. Instead of mapping the charge density to the mesh, we mapped the increment of density dQ to the mesh. The FFT solver KissFFT is based on a divide-and-conquer algorithm. We built a sparse input solver as a modified version of FFT solver which computes only needed (non trivial) operations [74]. We built also a sparse output solver inspired by the algorithm proposed by Katabi et al. [36]. Unfortunately, we did not get significant speed-ups with these modifications.

We decided to compute the increment of the long-range contribution related to the increment of density dQ by using multi-resolution methods. These methods are slower than PME but have better adaptive behavior. The multigrid approach was chosen because of its $\mathcal{O}(N)$ behavior and its good scalability [13]. We are currently developing an adaptive multigrid method.

7.4. Motion planning architecture for nanosystems

Participants: Leonard Jaillet, Stephane Redon.

In the past, we have started the development of original quasi-static simulation methods for nano-scale systems, based on motion planning methods inspired from Robotics. In the continuity of this work, we have proposed an original Motion Planning architecture for nanosystems platform called planning. This platform offers a general framework for motion planning applied to nanosystems. In particular it includes:

- A flexible definition of the degrees of freedom that describe the system, allowing different levels of representation (e.g. Cartesian coordinates, internal coordinates, coarse grain representation, etc.).
- The possibility to define an arbitrary set of initial, final and intermediate states, guiding the search for a solution path.
- The possibility to define an arbitrary set of constraints on the intermediate states of the path (e.g. geometric constraints, energy constraints, etc.)
- Several modular functionalities specific to motion planning (e.g. conformational sampling, exploration strategy, nearest neighbor search, etc.)
- An adapted integration within SAMSON which allows using directly all the existing force fields and state updaters present in the platform.

The planning architecture has been the base of several SAMSON modules. In particular, it led to the Planner-Explore module, which regroups many of the functionalities proposed and that can by combined together through a graphical interface. This module has in particular been used to study two complex problems:

- To capture the transition paths between endiandric acids (see Figure 6).
- To find the global minima of Lennard-Jones clusters, for dimensions up to one hundred.

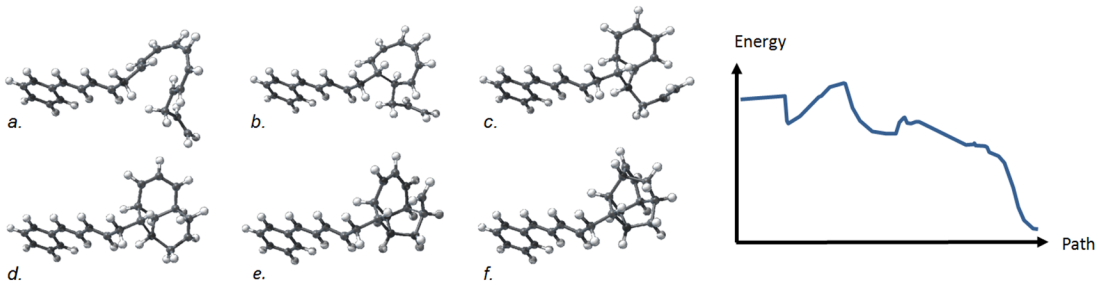

Figure 6. Transition path with its corresponding energy for an homolog of the endiandric acid and produced thanks to the Planner-Explore module.

7.5. Optimization of transition paths

Participants: Leonard Jaillet, Stephane Redon.

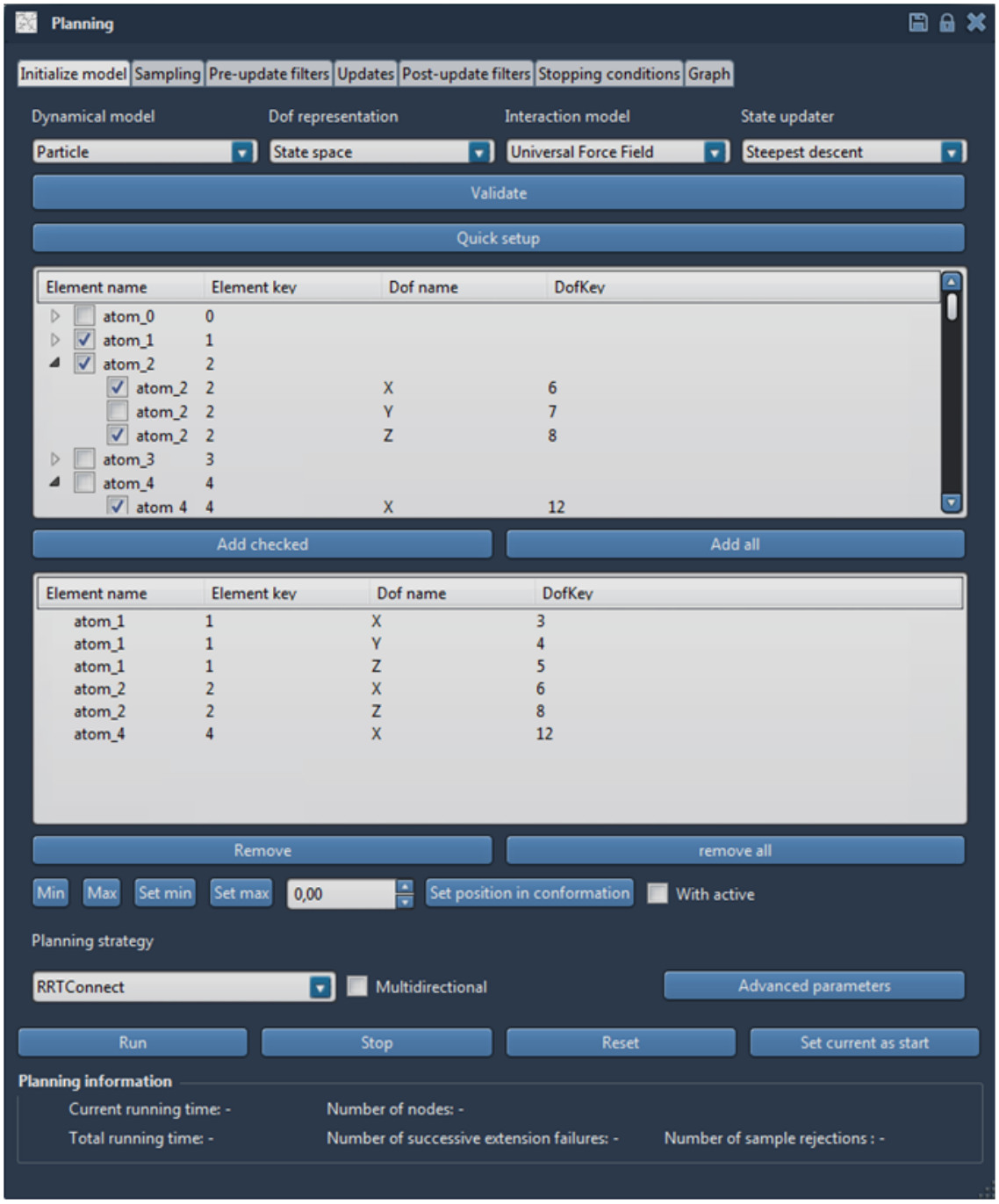

Figure 7. One tab of the graphical interface of the Planner-Explore modules, which allows initializing the model to be simulated.

Motion planning methods allow producing initial paths which represent transitions from one given conformation to another. However, these paths are typically suboptimal because of the probabilistic nature of the search strategy. Hence, it is necessary to develop tools to locally increase the path quality of the solution generated during the first phase. We have proposed several methods to address such a problem. One method developed is a variant of a state-of-the-art approach called nudged elastic band (NEB). It optimizes a set of intermediate images along the path, such that each image finds the lowest energy possible while maintaining equal spacing to neighboring images. Another technique we proposed is to rely on an equivalent of the shortcutting technique developed in Robotics motion planning, but applied to the context of energy landscape. Finally, we also have complemented these methods with additional tools to do simple path edition such as cutting or thinning paths.

7.6. As-rigid-as-possible shape interpolation for molecular modeling

Participants: Minh Khoa Nguyen, Leonard Jaillet, Stephane Redon.

Computer-aided methods play an important role in the study of molecular structures and interactions. Inspired by the as-rigid-as-possible approaches in the field of computer graphics, we created a tool for studying large deformation of molecular structures. This tool generates interpolated structures between two known conformations of a molecule while satisfying physical constraints. The users may use it for exploring, preprocessing, or combining their model with other biological algorithms. The developed method is flexible and can be extended to include physical properties of molecular structures.

We tested our method on a graphene sheet folding into a nanotube (Figure 8) and a few biological molecules, one of which is shown in Figure 9. The results show realistic transition motions compared to those from the linear interpolation approach.

The ARAP interpolation method has two main advantages: simplicity and preservation of local rigidity. The method is totally geometrical, yet can be extended to include physical or biological properties such as bond strength. It will be proposed as a SAMSON Element for the SAMSON software platform for computational nanoscience.

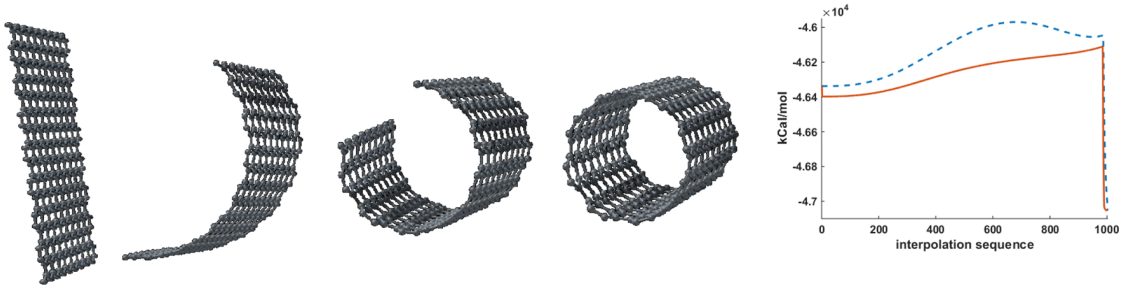

Figure 8. ARAP interpolation to generate graphene sheet folding into a nanotube. The last figure plots the energy for a sequence of 1000 interpolated images. The energy of ARAP interpolation is shown by the dotted blue curve and the optimal energy after applying NEB is shown by the solid red curve.

7.7. Automatic parameterization for the Universal Force Field

Participants: Svetlana Artemova, Leonard Jaillet, Stephane Redon.

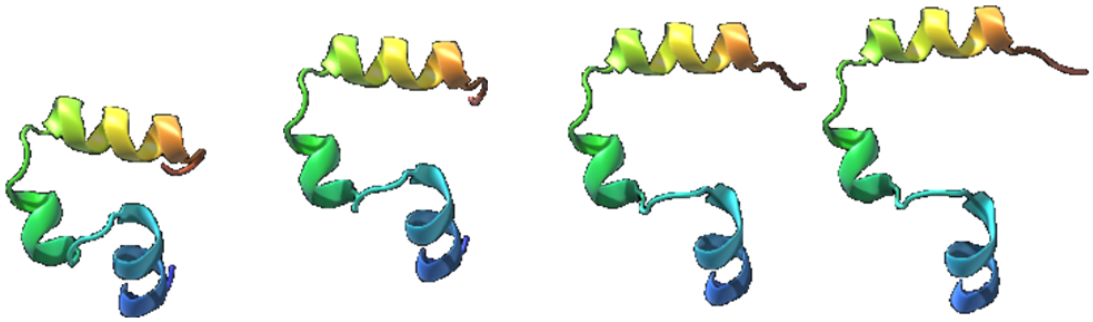

Figure 9. Transition obtained by the ARAP method of a subdomain of the villin headpiece (protein ID: 1YFR) into its distorted shape generated manually.

We have continued working on the integration of the Universal Force Field in SAMSON. This force field is a classical non-reactive force field that has parameterizations for all atoms of the periodic table with atomic number lower than 103. Our implementation of this force field includes a new automatic perception scheme for molecular systems that is specifically-tailored for UFF, as well as several corrections and refinements that have been lately proposed in the literature. We have tested this implementation on more benchmarks and improved its computational performance. Additionally, we have compared our implementation to that of the OpenBabel toolbox. As a result, our self-contained implementation was integrated in a new module for SAMSON and is now available on SAMSON-Connect website (see Figure 10). The paper describing the obtained results will appear in the Journal of Computational Chemistry.

7.8. Interactive modeling with the Universal Force Field

Participants: Leonard Jaillet, Svetlana Artemova, Stephane Redon.

In parallel with the classical Universal Force Field, we have continued working on an extension of this force field that we call Interactive Modeling UFF (IM-UFF). In classical UFF topologies and atoms' typizations are set in the initialization phase and remain fixed for the entire simulation. IM-UFF, on the contrary, allows soft transitions for both topologies and atoms' typizations. This new approach, thus, combines the possibility to significantly modify molecular structures (as with reactive force fields) with a broad diversity of supported systems thanks to the universality of UFF. Such an extension lets the user easily build and edit molecular systems interactively while being guided by physically-based inter-atomic forces. The validity of this extended version of UFF was tested on the same large set of benchmarks as those used to test classical UFF, and the results of both approaches were compared.

7.9. Error Analysis of Modified Langevin Dynamics

Participants: Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

Adaptively Restrained Particles Simulations (ARPS) were recently proposed with the purpose of speeding up molecular simulations. The main idea is to modify the Hamiltonian such that the kinetic energy is set to zero for small velocities, which allows to save computational time since particles do not move and forces need not be updated. ARPS can be combined with Langevin dynamics in order to speed up the computation of macroscopic quantities.

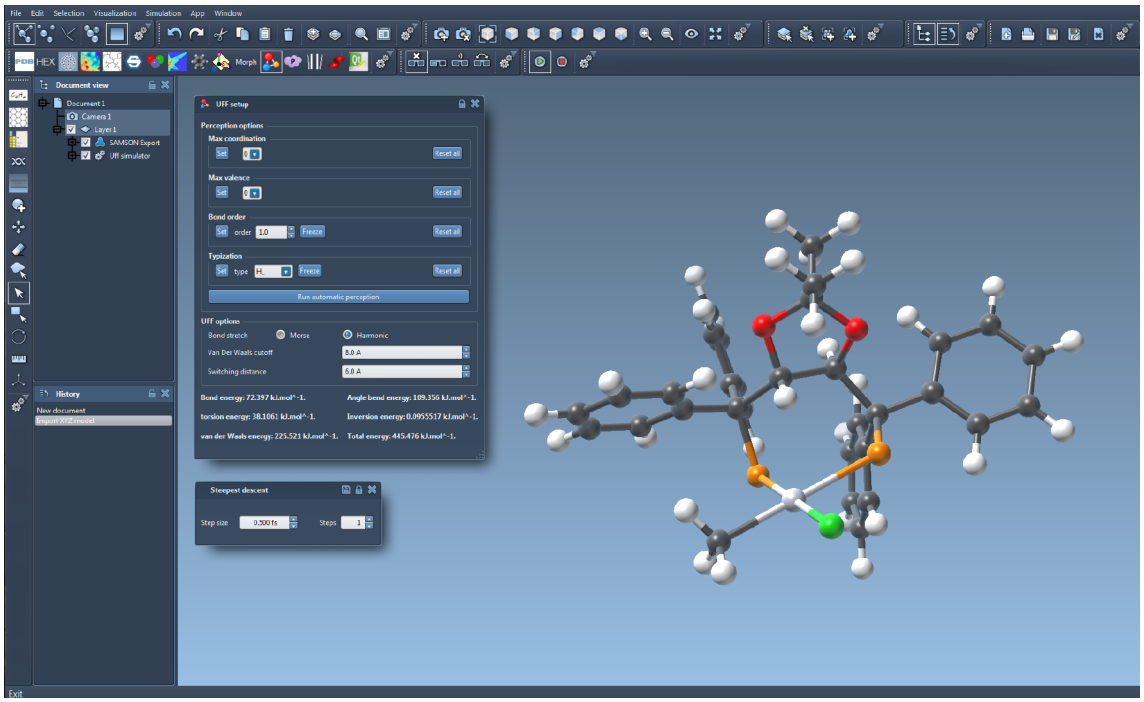

Figure 10. A molecule being interactively manipulated in SAMSON thanks to the UFF module. The interface of the UFF module allows to setup UFF. The upper part of the interface proposes options to manually adjust the perception of the molecular system. The middle part proposes the UFF options. The lower part prints out each energy contribution with the resulting total energy.

The aim of this work is to understand how simulation errors depend on the parameters of the method. We distinguish the statistical error and the systematic error related to the finiteness of the time step Δt. The statistical error is controlled by variance, that is given by

$$\sigma^2 = -2 \left\langle A - \mu(A), \mathcal{L}^{-1} \left(A - \mu(A) \right) \right\rangle_{L^2(\mu)} \tag{1}$$

where μ is the invariant measure, $\mathcal{L}$ is the generator of the stochastic process and A an observable. First we demonstrate by use of weighted L^∞ estimations that the ARPS-Langevin dynamics are well defined. In the main part of this work, we quantify the increase of variance of the ARPS-Langevin process as a function of the ARPS parameters. For small parameters, we express the generator of the ARPS-Langevin dynamics as a perturbed generator of the Langevin dynamics, and study the asymptotic expansions of the variance (1) in the restrained dynamics parameter ε.

$$\sigma_\varepsilon^2 = \sigma^2 + \mathcal{O}(\varepsilon)$$

For large values of ε, we perform numerical simulations. For a simple 1D system we approximate $\mathcal{L}^{-1}$ by Galerkin approach and for higher-dimensional systems we discretize the stochastic differential equations by a second order method and analyze a model of a dimer surrounded by solvent particles.

7.10. Algorithmic speed up of the ARPS method

Participants: Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

Adaptively Restrained Particles Simulations (ARPS) allow to save computational time at each time step since particles do not move and forces need not be updated. The associated gain can be quantified by an algorithmic speed-up factor $S_{\text{algo}} \geq 1$. Intuitively, freezing more particles leads to larger algorithmic speed-ups, but also larger correlations in time.

We analyzed the algorithmic speed up with respect to the standard methods. Since the ARPS algorithm is based on adding and subtracting of the forces between active particles, the gain with respect to the standard method, where only one complete computation of all interactions is performed at each time step, is achieved only if the percentage of restrained particles is big enough. Hence we studied the necessary conditions, under which the computational complexity of the forces updating in the ARPS method is lower than the one of the standard method. This allows to achieve an algorithmic speed up that is always bigger than one.

We also propose a simple strategy for choosing optimal simulation parameters.

7.11. Numerical analysis for the ARPS method

Participants: Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

Previous works have led to understanding of the choice of optimal parameters for the ARPS dynamics. The interest lies in achieving the highest percentage of restrained particles, while minimizing the modification of the variance and the systematic error. We study discretization schemes of the ARPS-Langevin dynamics, such that the systematic error remains of second order in the time step size and we introduce a Metropolis step in order to stabilize the simulations and hence to allow "a sharper" choice of the ARPS parameters, which lead to better algorithmic speed-ups.

7.12. New rendering algorithm for secondary structures

Participants: Marc Aubert, Stephane Redon.

We developed a new algorithm for rendering secondary structures of proteins (Figure 11). The method relies on the determination of the most probable secondary structure elements (e.g. alpha helices and beta sheets) based on geometrical features of a protein. After construction of control points on the CPU, the method generates triangles directly on the Graphics Processing Unit (GPU) through geometry shaders. The number of generated triangles may be adaptively chosen based on e.g. the camera distance and the desired resolution. The secondary structure algorithm and the rendering algorithm are both fast enough to allow for interactive modification of the protein (e.g. thanks to As-Rigid-As-Possible editing algorithms).

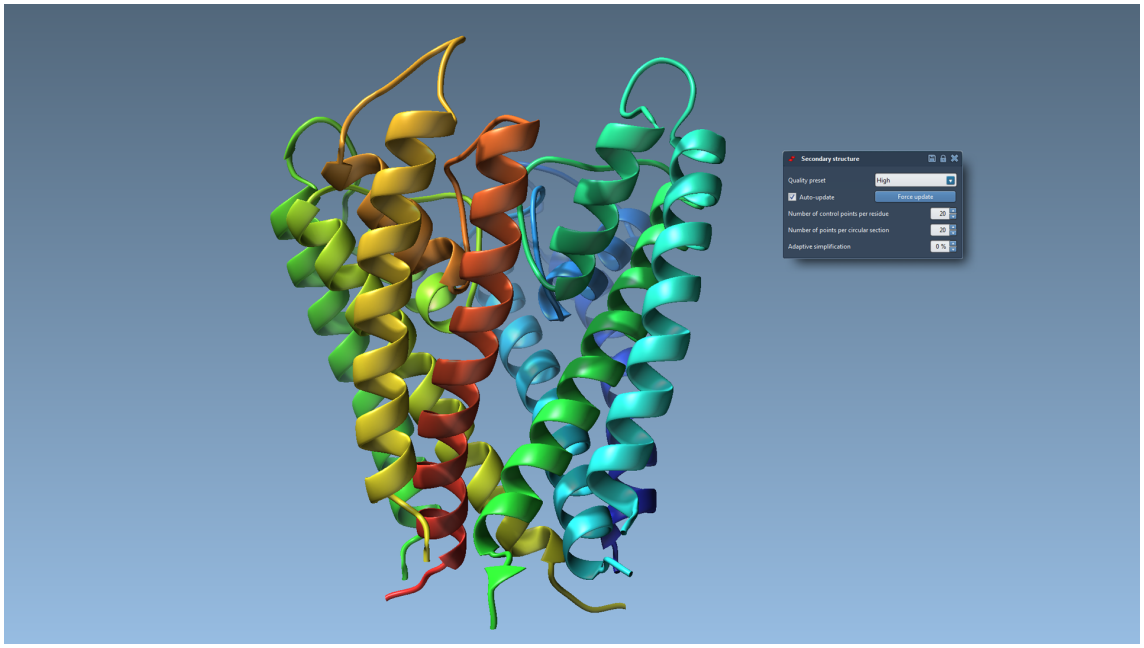

Figure 11. Protein secondary structure rendering on GPUs.

7.13. Property models

Participants: Marc Aubert, Stephane Redon.

We extended the hierarchy of classes in SAMSON for *property models*. Property models are one of the five categories of models in SAMSON, with structural models (for geometry and topology), visual models (for custom graphical representations), dynamical models (to describe degrees of freedom) and interaction models (to represent energies and forces). We have added classes to easily represent in SAMSON various functions, fields (e.g. scalar fields and vector fields), etc. These property models are template classes which may rely on the unit system of SAMSON to perform dimensional analysis at compile time.

7.14. Integration of tools in SAMSON

Participants: Nadhir Abdellatif, Stephane Redon.

Thanks to funding from the Nanosciences Foundation in Grenoble, we developed SAMSON Elements (modules for SAMSON) that integrate existing tools. In particular, we integrated OpenBabel, a tool to convert between numerous molecular formats (Figure 12), ClustalW, a tool for sequence alignment (Figure 13), and Pepsi-SAXS, a tool for SAXS developed in the group (Figure 14).

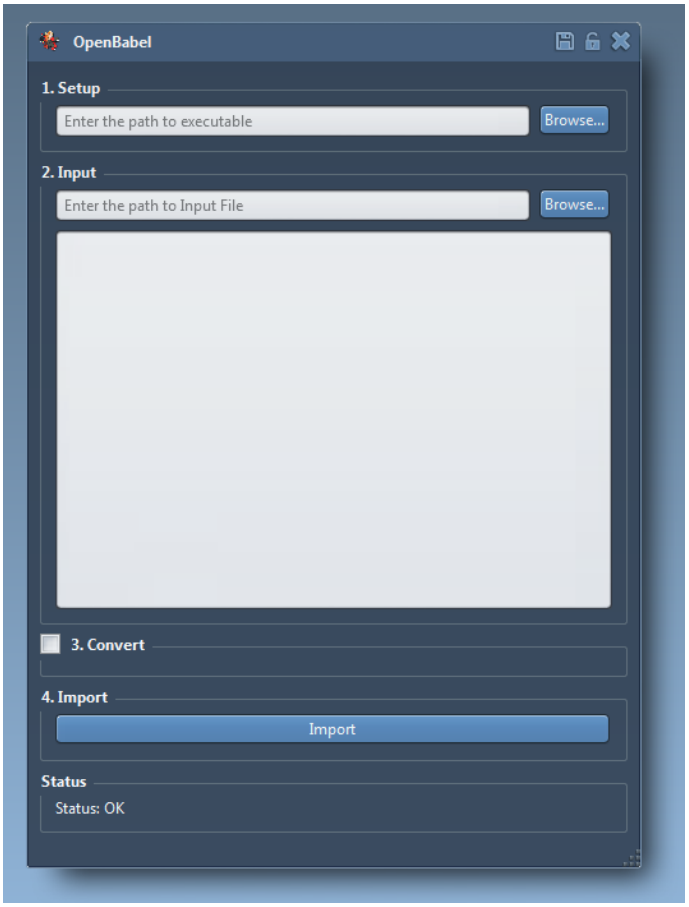

Figure 12. The OpenBabel connector in SAMSON

Figure 13. ClustalW in SAMSON

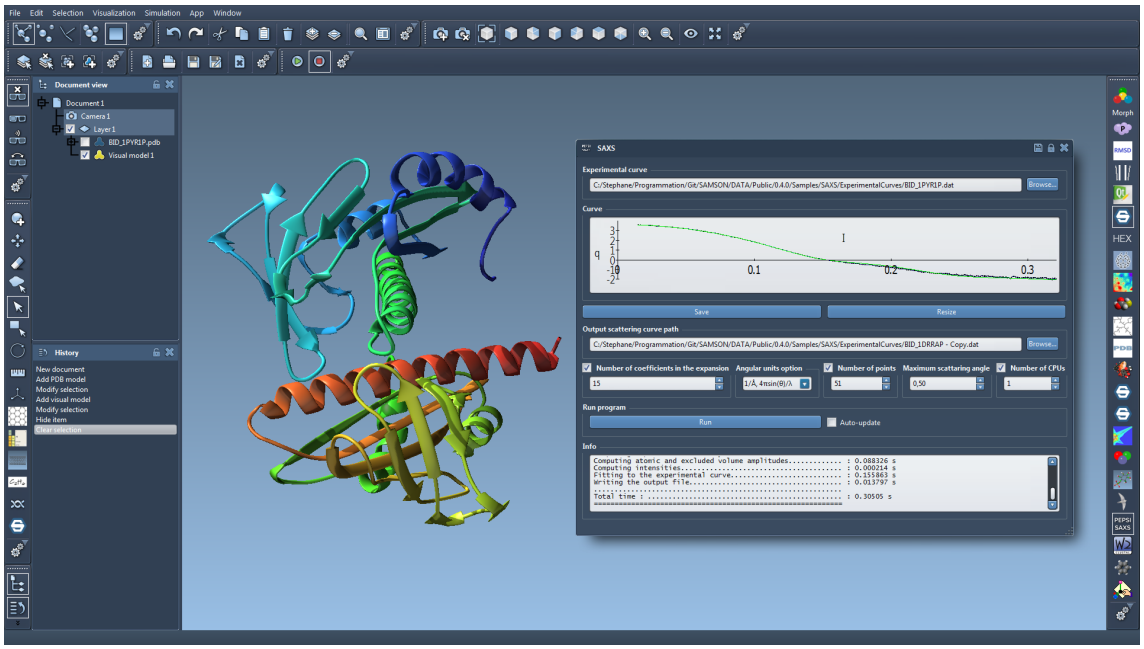

Figure 14. Our Pepsi-SAXS tool integrated in SAMSON

7.15. Development of SAMSON Connect

Participants: Mohamed Yengui, Jocelyn Gate, Stephane Redon.

We have continued the development of SAMSON Connect (Figure 15, https://www.samson-connect.net), the online platform for distributing SAMSON and SAMSON Elements (modules for SAMSON). SAMSON Connect is a web application, associated to a database, that functions as the well-known stores for mobile Apps (e.g. Google Play, the Apple App store, etc.). Users may create an account, download SAMSON, and add SAMSON Elements to their configuration based on their needs (Figure 16). Adding a SAMSON Element is performed in just one click (to the Add button of the corresponding SAMSON Element), and the SAMSON Element is installed when the user restarts SAMSON. Users may also request an upgrade to a Developer status, after which they can download the SAMSON Software Development Kit used to develop SAMSON Elements. They may then upload their SAMSON Elements to SAMSON Connect in order to share them. The platform opened in March 2015 to release the first beta version of SAMSON. We also produced some video tutorials for SAMSON (Figure 17).

On the back-office of SAMSON Connect, we added several functionalities that facilitate publishing new versions of SAMSON and SAMSON Elements (e.g. choosing default SAMSON Elements), email users based on their account type (user, developer, etc.), OS, etc. We also turned to automatic acceptance of new user accounts (once they validate their email address). We also updated the SAMSON web service to enable more message types and retrieve information about the server, the database, etc.

7.16. Documenting the SAMSON Software Development Kit

Participants: Stephane Redon, Jocelyn Gate, Svetlana Artemova.

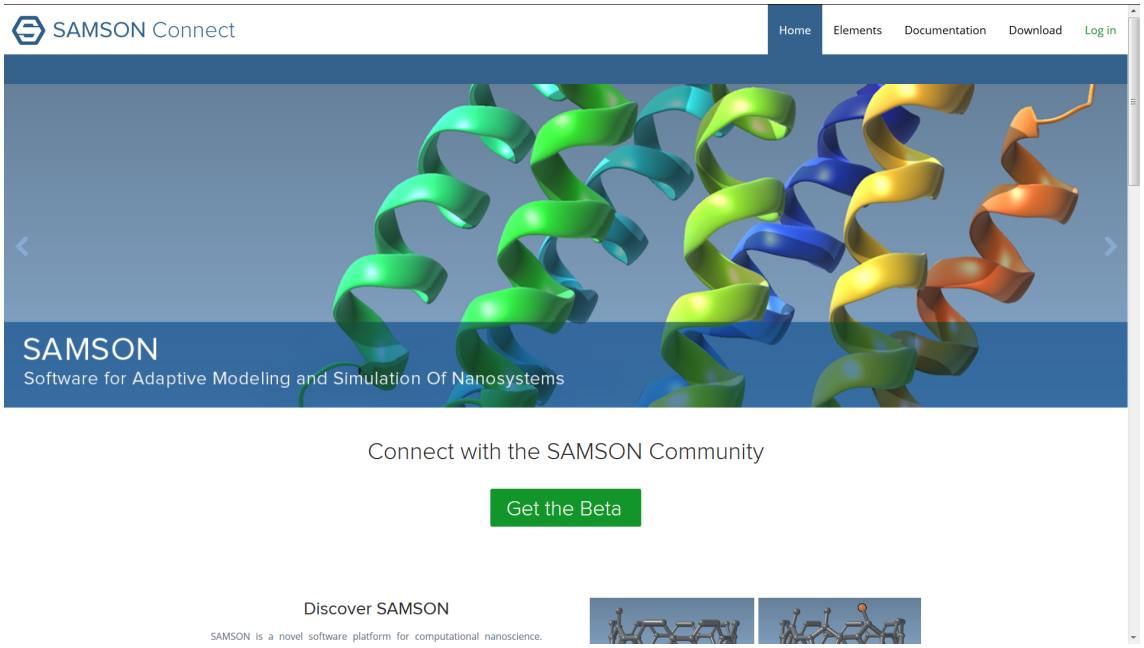

Figure 15. The home page of SAMSON Connect (https://www.samson-connect.net)

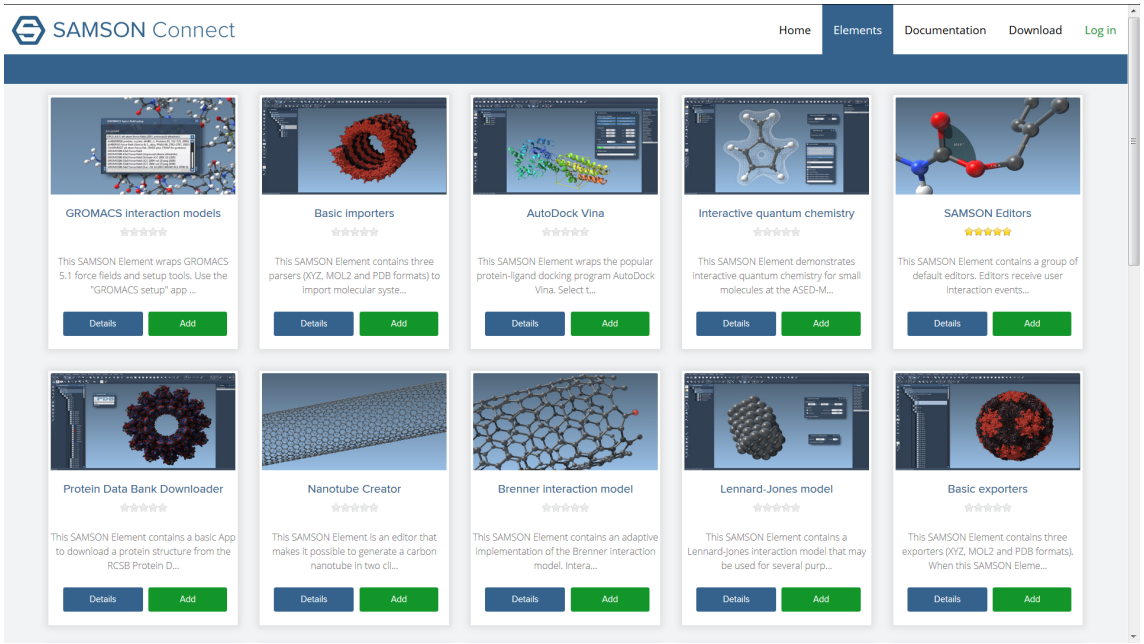

Figure 16. The Elements page, where users may add SAMSON Elements (modules) to their configuration (https://www.samson-connect.net)

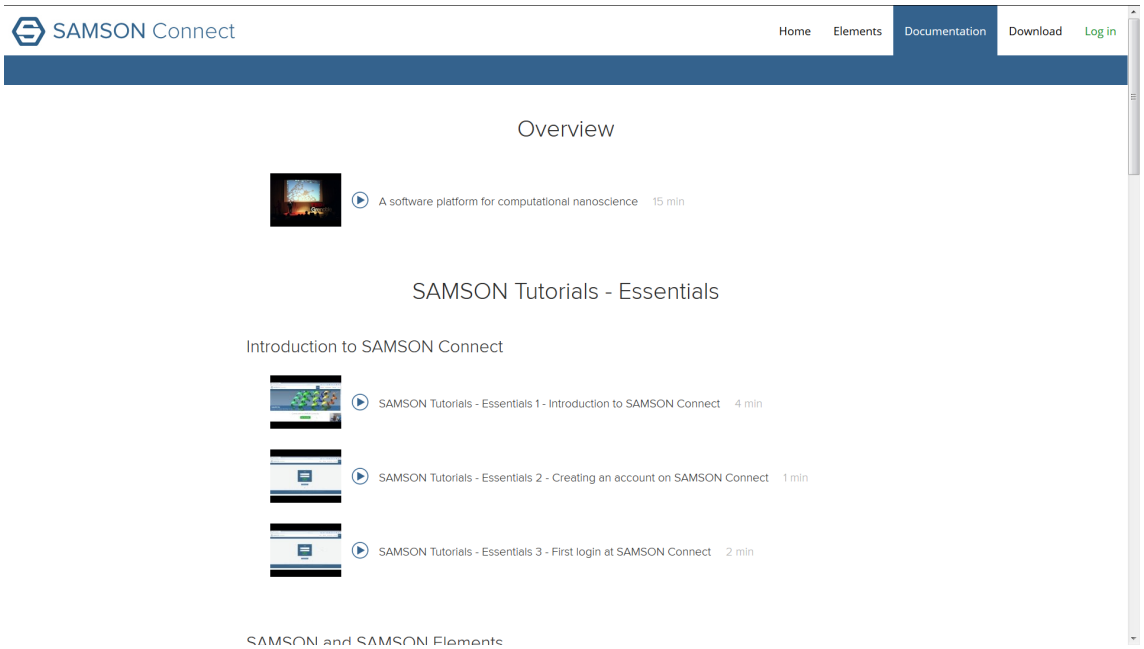

Figure 17. The documentation page on SAMSON Connect (https://www.samson-connect.net)

The SAMSON Software Development Kit (SDK) is at the core of the SAMSON platform and makes it possible to develop SAMSON Elements (modules). The API of SAMSON contains numerous classes and allows for a variety of modules types (e.g. parsers, force fields, visual models, integrators, apps, editors, etc.), and provides several non-elementary mechanisms (e.g. a unit system, a signals and slots mechanism, memory management, data structures for incremental calculations, etc.). We continued writing the SDK documentation accessible to SAMSON Elements developers. The current PDF version for the beta 0.4.0 version has passed 500 pages.

7.17. SAMSON SDK Helpers

Participants: Jocelyn Gate, Stephane Redon.

We have developed a number of Helpers in the SAMSON SDK, in order to facilitate the development of SAMSON Elements (modules for SAMSON). For example, the SAMSON Element generator (Figure 18) generates code that immediately compiles and runs, and that developers may complete, for a number of SAMSON Classes.

We have also developed for the group a helper able to upload numerous SAMSON Elements to SAMSON Connect at the same time (Figure 19), which is especially useful given the rapidly growing number of modules being developed in the team.

7.18. Pepsi-SAXS : an adaptive method for rapid and accurate computation of small angle X-ray scattering profiles

Participant: Sergei Grudinin.

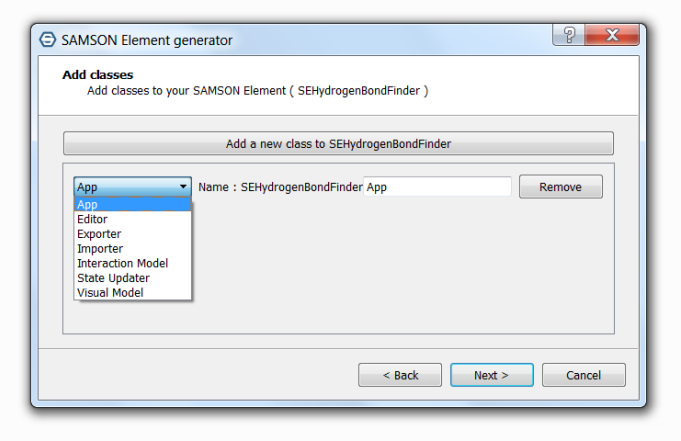

Figure 18. The SAMSON Element Generator makes it easy to develop modules for SAMSON

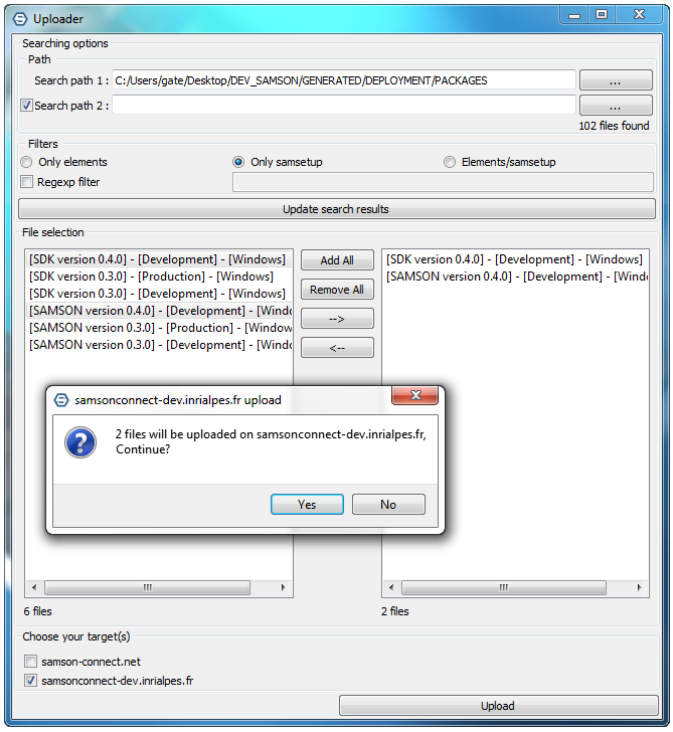

Figure 19. The SAMSON Element uploader eases the transfer of multiple SAMSON Elements to SAMSON Connect

We developed a new method called Pepsi–SAXS that calculates small angle X-ray scattering profiles from atomistic models. Our method is based on the multipole expansion scheme and is significantly faster and more precise compared to other tested methods. In particular, using the Nyquist–Shannon–Kotelnikov sampling theorem, we adapt the multipole expansion order to the size of the model and the resolution of the experimental data. We argue that using the adaptive expansion order, our method has the same quadratic dependence on the number of atoms in the model as the Debye-based approach, however, with a much smaller prefactor in the computational complexity.

We have systematically validated our method on an excessive set of over fifty models collected from the BioIsis and SASBDB databases. Using a laptop, we demonstrated that Pepsi-SAXS is about 9, 33 and 43 times faster compared to CRYSOL, FoXS and the 3D-Zernike method in SAStbx, correspondingly, when tested on data from the BioIsis database, and is about 5, 18 and 23 times faster compared to CRYSOL, FoXS and SAStbx, correspondingly, when tested on data from SASBDB. On average, Pepsi-SAXS achieves 17% smaller value of χ compared to CRYSOL and 15% smaller value of χ compared to FoXS for BioIsis profiles, and 6% smaller value ofχ compared to CRYSOL and 19% smaller value of χ compared to FoXS for SASBDB profiles.

7.19. Knodle: a Support Vector Machines-based automatic perception of organic molecules from 3D coordinates

Participants: Maria Kadukova, Sergei Grudinin.

We addressed the problem of the assignment of atom types and bond orders in low molecular weight compounds. For this purpose, we developed a prediction model based on nonlinear Support Vector Machines (SVM), implemented in a KNOwledge-Driven Ligand Extractor called *Knodle*, a software library for the recognition of atomic types, hybridization states and bond orders in the structures of small molecules. We trained the model using an excessive amount of structural data collected from the PDBbindCN database. Accuracy of the results and the running time of our method is comparable with other popular methods, such as NAOMI, fconf, and I-interpret. More precisely, on the popular Labute's benchmark set consisting of 179 protein-ligand complexes, *Knodle* makes five to six perception errors, NAOMI makes seven errors, I-interpret makes nine errors, and fconv makes thirteen errors. On a larger set of 3,000 protein-ligand structures collected from the PDBBindCN general data set (v2014), *Knodle* along with NAOMI have a comparable accuracy of approximately 6 % of errors, whereas fconv produces approximately 13 % of errors. Overall, our study demonstrates the efficiency of nonlinear SVM in structure perception tasks.

7.20. Symmetry Detection Method

Participants: Silvia Dias Pinto, Sergei Grudinin.

We developed an algorithm for automatic recognition of the point group symmetry in electron density maps of biological objects. More precisely, the method operates on cryo-Electron Microscopy (cryoEM) data, which typically contain 3D structures of multi-domain proteins and their complexes. We represent the shape using a spherical harmonic decomposition and then operate on the expansion coefficients to quantify the structural symmetry thanks to a mismatch function. Overall, we developed new mathematical and computational frameworks for symmetry detection using the polynomial expansion approach.

7.21. Pepsi-Dock: fast predictions of putative docking poses using accurate knowledge-based potentials functions to describe interactions between proteins

Participants: Emilie Neveu, Sergei Grudinin, David W. Ritchie, Petr Popov.

Many biological tasks involve finding proteins that can act as an inhibitor for a virus or a bacteria, for example. Such task requires knowledge on the structure of the complex to be formed. Protein Data Bank can help but only a small fraction of its proteins are complexes [16]. Therefore, computational docking predictions, being low-cost and easy to perform, are very attractive if they describe accurately the interactions between proteins while being fast to find which conformation will be the most probable. We have been developing a fast and accurate algorithm that combines the FFT-accelerated docking methods [67] with a precise knowledge-based potential functions [58] describing interactions between the atoms in the proteins .
Interactions between proteins follow complex and non-linear laws which computation is time-consuming. It is of common usage to start the predictions with a simple, approximated, expression of these interactions to then reduce the space search in order to use more complex laws. However we think it is important to use the most accurate free energy not to miss some important docking solutions. Thus, our aim is to integrate the very-detailed knowledge-based potentials into the *Hex* code and to take advantage of its exhaustive search, which is by now still the most efficient and reliable search algorithm [67] .
Last year, we adapted the machine learning process so that the knowledge-based potentials describing atom interactions can be translated into the polynomial basis used in *Hex*. The current evaluations of the knowledge-based scores takes more time than a shape+electrostatic representation but is still fast: exploring 10^9 conformations of a complex takes on average 5-10 minutes on a regular laptop computer.
This year, we run cross-validation experiments and tested different data sets in order to improve the predictions. Using bound conformations of each proteins to make the predictions, we retrieve up to 70% correct complexes of about 200 complexes. Results show that the knowledge-based potentials, while being general, correctly predict the interactions. Even better results could be achieved without the limitations in the search range by the spherical sampling grid which lacks of precision far away of its origin. Because many complexes have separation distances greater than 30 Å, we are now working on a multi-centre definition of the potentials in order to correctly predict the structures of protein complexes starting from their unbound structures.

7.22. Pepsi-Piper: rigid docking predictions using Pepsi potentials into Piper code

Participants: Sergei Grudinin, Emilie Neveu, Dima Kozakov, Dzmitry Podgorny.

This work is the continuation of the Pepsi-Dock project that aims to develop fast predictions of putative docking poses using accurate knowledge-based potentials functions to describe interactions between proteins. The goal is to integrate the precise, and yet easy to compute, distance-based pairwise knowledge-based potentials [58] into the Piper search code [48] in order to compare its exhaustive search with the *Hex* one. The former samples the conformations using a cartesian grid while the latter, a spherical one. We proved our potential used in *Hex* can predict the structures of complexes with a really good success rate, the main limitation being the lack of precision of the spherical sampling when the separation distance of the two proteins is too large. We think predicting docking combining our potential and a sampling search based on a cartesian grid as in Piper will achieve greater results, but will require more computational time.
We first adapt our potential to the Piper code and showed that the ranking results on the data set used for training are better than the ranking provided by Piper [25]: when the potential is used to sort the conformations, the correct solution is found in the first ten for 85% cases, while Piper found it in only 25% cases. The next step is to use the cartesian sampling to make docking predictions. When the Piper code will be ready to integrate our potential, we will be able to confront with other knowledge-based potentials such as the one initially used in Piper, DARS.

7.23. Flex-Dock: towards flexible docking predictions using metaheuristics optimisation methods

Participants: Emilie Neveu, Sergei Grudinin, Alexandre Hoffman, Angelo Migliosi, Xavier Besseron, Grégoire Danoy, Pascal Bouvry.

Docking numerical methods are used to predict the preferred location of one molecule with respect to the second when bound to each other. This is particularly useful for the design of drugs that inhibit the effects of viruses or bacteria. However proteins change their conformation upon binding and searching for flexible conformations involves enormous degrees of freedom and complex physics. Thus, the prediction of realistic interactions with full flexibility of the two partners is an intractable global optimisation problem.
There are currently several algorithms that produce high quality predictions of molecular complexes [43]. But very few manages to deal with the flexibility of the proteins. A common method is to refine the most probable predicted rigid complexes with a scoring allowing for flexibility [81]. Here, we want to tackle flexibility and sampling all together. Exhaustive search methods, which were by now the most accurate optimisation method for relatively small molecules [53] will be too time-consuming when it comes to large proteins. There is a strong need to explore and define new optimisation algorithms such as metaheuristic ones that can deal with several local minima and a large minima and a large search space. The main goal of this project is to define the problem and find for the optimisation method that will potentially give better results than the actual reference, SwarmDock [54].
We worked on a first comparison of several evolutionary-based algorithms (Genetic Algorithm [40], Differential Evolution [76], Particle Swarm Optimisation [46]) using rigid proteins only and on the use of multi-objective algorithms when the proteins are flexible.
To take into account flexibility, we approximate large-scale deformations of each proteins using an elastic network model combined with a low-frequency approximation called normal mode analysis such as in [81] or in [54]. Combined with the rigid transformation between the two proteins, it defines a complete while reduced set of degrees of freedom to search for.
The scoring function has to discriminate correct conformations from impossible ones. Our scoring is the main difference with SwarmDock. It takes into account the energy gained by docking using the precise knowledge-based potentials derived in [58], whereas only a simple physics-based energy is used in SwarmDock. We also want to explore another scoring that will also add the energetic cost of each moves of the proteins. To do so, we started to develop multi-objective algorithms. Combined with a Pareto Front analysis, this will help us to validate the scoring and to compare different evolutionary-based algorithms.Tests will be directly made on the Protein-Protein Benchmark [42] so that we can compare with other docking methods.

7.24. FastRMS: rapid determination of RMSDs corresponding to macromolecular rigid body motions, adding flexibility via collective motions

Participants: Sergei Grudinin, Petr Popov, Emilie Neveu.

Computing the root mean sum of squared deviations (RMSDs) between two sets of coordinates each describing a different conformation of a macromolecule is a necessary step in many structural bioinformatics and molecular modelling technics to assess structural predictions [43], identify biding sites [49] or structurally classify proteins. A straightforward and universally-used method determines the RMSD with a computational complexity proportional to the number of atoms in the molecule. We recently presented RigidRMSD, a fast algorithm that determines RMSDs corresponding to a set of rigid body motions of a macromolecule in constant time with respect to the number of atoms in the molecule [57]. Here, we extend it to proteins with flexibility modelled with collective motion such as an elastic network model combined with normal mode analysis.
With these new assumptions, the complexity of the algorithm depends linearly or quadratically with the number of collective motion vectors selected to approximate the flexibility. The typical number of vectors needed to have accurate flexible movements being much lower than the number of atoms composing the molecules, we prove our algorithm is still faster than the common method. Our algorithm is particularly useful for rigid body modelling applications such as rigid body docking procedures allowing for flexibility via collective motions: clustering, high-throughput analysis and simulation results [49], [26], [59]. A C++ implementation of our algorithm will be soon available at http://nano-d.inrialpes.fr/software/RigidRMSD.

7.25. SAM : Spherical Polar Fourier Assembly of Protein Complexes with Arbitrary Point Group Symmetry

Participants: David W. Ritchie, Sergei Grudinin.

We presented a novel FFT-based *ab inito* docking algorithm called "SAM" for building perfectly symmetrical models of protein complexes with arbitrary point group symmetry. The basic approach uses a novel and very fast 1D symmetry-constrained spherical polar Fourier search to assemble cyclic C_n systems from a given protein monomer. Structures with higher order (D_n, T, O, and I) point group symmetries may be built using a subsequent symmetry-constrained Fourier domain search to assemble trimeric sub-units. Our results show that the SAM algorithm can correctly assemble monomers of up to around 500 residues to produce a near-native complex structure with the given point group symmetry in 17 out of 18 test cases. The SAM program may be downloaded for academic use at *http://sam.loria.fr/*.

7.26. KSENIA : Knowledge of Native Protein-Protein Interfaces is Sufficient to Construct Predictive Models for the Selection of Binding Candidates

Participants: Petr Popov, Sergei Grudinin.

Selection of putative binding poses is a challenging part of virtual screening for protein-protein interactions. Predictive models to filter out binding candidates with the highest binding affinities comprise scoring functions that assign a score to each binding pose. Existing scoring functions are typically deduced collecting statistical information about interfaces of native conformations of protein complexes along with interfaces of a large generated set of non-native conformations. However, the obtained scoring functions become biased toward the method used to generate the non-native conformations, i.e. they may not recognize near-native interfaces generated with a different method.

Present study demonstrates that knowledge of only native protein-protein interfaces is sufficient to construct well-discriminative predictive models for the selection of binding candidates. Here, we introduce a new scoring method that comprises a knowledge-based potential called *KSENIA* deduced from the structural information about the native interfaces of 844 crystallographic protein-protein complexes. We derive *KSENIA* using convex optimization with a training set composed of native protein complexes and their near-native conformations that are obtained using deformations along the low-frequency normal modes. As a result, our knowledge-based potential has only marginal bias toward a method to generate putative binding poses. Furthermore, *KSENIA* is smooth by construction, which allows to use it along with a rigid-body optimization to refine the binding poses. Using several test benchmarks we demonstrate that our method discriminates well native and near-native conformations of protein complexes from the non-native ones. Our methodology can be easily adapted to the recognition of other types of molecular interactions, such as protein-ligand, protein-RNA, etc. *KSENIA* will be made publicly available as a part of the SAMSON software platform at https://www.samson-connect.net.

7.27. Predicting Binding Poses and Affinities in the CSAR 2013–2014 Docking Exercises Using the Knowledge-Based Convex-PL Potential

Participants: Sergei Grudinin, Petr Popov, Emilie Neveu, Georgy Cheremovskiy.

The 2013–2014 CSAR docking exercise was the opportunity to assess the performance of the novel knowledge- based potential we are developing, named Convex-PL. The data used to derive the potential consists only of structural information from protein-ligand interfaces found in the PDBBind database. As expected, our potential proved to be very efficient in the near- native pose detection exercises, where we correctly predicted two near-native poses in the 2013 exercise and also ranked 22 near- native poses first and 2 second in the 2014 exercise. Somewhat more surprisingly, we obtained a fair performance in some of the CSAR affinity ranking exercises, where the Spearman correlation coefficients between our predictions and the experiments are greater than 0.5 for several protein–ligand sets. Nonetheless, affinity prediction exercises turned out to be a challenge, and significant progress in the development of our method is needed before we can successfully predict binding constants.

7.28. Prediction of homo- and hetero-protein complexes by ab-initio and template-based docking: a CASP-CAPRI experiment

Participants: Sergei Grudinin, Petr Popov, Emilie Neveu.

We present the results for CAPRI Round 30, the first joint CASP-CAPRI experiment, which brought together experts from the protein structure prediction and protein-protein docking communities. The Round comprised 25 targets from amongst those submitted for the CASP11 prediction experiment of 2014. The targets included mostly homodimers, a few homotetramers, and two heterodimers, and comprised protein chains that could readily be modeled using templates from the Protein Data Bank. On average 24 CAPRI groups and 7 CASP groups submitted docking predictions for each target, and 12 CAPRI groups per target participated in the CAPRI scoring experiment. In total more than 9500 models were assessed against the 3D structures of the corresponding target complexes. Results show that the prediction of homodimer assemblies by homology modeling techniques and docking calculations is quite successful for targets featuring large enough subunit interfaces to represent stable associations. Targets with ambiguous or inaccurate oligomeric state assignments, often featuring crystal contact-sized interfaces, represented a confounding factor. For those, a much poorer prediction performance was achieved, while nonetheless often providing helpful clues on the correct oligomeric state of the protein. The prediction performance was very poor for genuine tetrameric targets, where the inaccuracy of the homology-built subunit models and the smaller pair-wise interfaces severely limited the ability to derive the correct assembly mode. Our analysis also shows that docking procedures offer a clear advantage over standard homology modeling techniques and that highly accurate models of the protein components are not always required to identify their association modes with acceptable accuracy.

Most of the targets in Round 30 of CAPRI were homodimers and homotetramers, thus it was a good opportunity to test our novel symmetry assembling docking method. To do so, we imposed C2 symmetry constraints for all the homodimers and we imposed C4 and D2 symmetry constraints for all the homotetramers from the target complexes. Below, we present the new fast multi-resolution method for docking both symmetric and non-symmetric protein complexes that was used in Round 30 of CAPRI. First, the structures of the individual subunits were taken from the stage two predictions of the CASP10 assessment experiment. More precisely, starting from 150 available CASP 3D models of monomers, we predicted models of symmetric multimers using the novel symmetry docking method, which performs symmetry-induced protein docking using the shape-complementarity scoring function computed as spherical polar Fourier correlations. Specifically, this method performs exhaustive search over the available (four in case of cyclic symmetries or six otherwise) degrees of freedom for the given point group symmetry type. For the targets of Round 30 of CAPRI we imposed three types of symmetry, C2, C4, and D2. For the case of heterodimers, we used the standard Hex docking method.

For the input of the docking methods, we generated the scaffolds of initial models of monomers by cutting-off the side chains. More specifically, we mutated all side-chains except for the glycines to alanines. Compared to the standard all-atom rigid-body docking methods, we expect the scaffold docking approach to produce binding poses that are less sensitive to the flexibility of the side-chains. We clustered the solutions with the threshold ligand-RMSD value of 8 Åusing the RigidRMSD library. Finally, we ranked the clusters by the value of the best score and kept 50 best clusters for the refinement stage. In total, for each target we proceeded to the refinement with 7,500 modeled structures of protein complexes.

On the next step, we optimized each putative binding interface of the all-atom representation of a protein complex by means of a rigid-body first-order minimization scheme. Specifically, after each rigid-body minimization step we proceeded with the optimization of side-chains described by the rotameric representation using the SCWRL4 package. We computed the interactions between the subunits in a protein complex using the novel reference state-free knowledge-based scoring function KSENIA, which is smooth by construction and is thus very suitable for a gradient-based minimization protocol. Finally, we ranked the predictions by the value of the KSENIA potential of the optimized structure and selected ten best candidates for the submission.

7.29. Convex relaxation for non-convex quadratic optimization problems with applications to side-chain prediction in protein structures

Participants: Aleksandr Katrutsa, Sergei Grudinin.

The side-chain prediction problem is the major part of the more general protein structure prediction problem, which is very important for drug design and in the prediction of stable protein mutations. Formally, the side-chain prediction problem states in the form of discrete quadratic optimization problem with an indefinite matrix in the quadratic term,

$$\mathbf{x}^\top \mathbf{Q}\mathbf{x} + \mathbf{b}^\top \mathbf{x} \to \min_{\mathbf{x}\in\{0,1\}^n} \tag{2}$$

This problem is NP-hard, so to get a good approximation solution we used convex semidefinite relaxation with different types of constraints. This approach is the powerful optimization technique that helps to reformulate the initial non-convex problem as a convex one and sometimes even gives the exact solution. The important step is to operate with precise energy function, which is used to compute the energy of different interactions in proteins. To obtain this, we used the machine-learning procedure, which extracts the parameter vector for the potential from the training set of protein structures. After the training step, we used this vector to compute the energy of a protein and to find the side-chains corresponding to the minimal total energy of the protein. The current accuracy in side-chain prediction is about 80%, which is achieved using the spectrum relaxation of the matrix in the quadratic term. Also, this approach is very fast, precisely, it requires less than 1 second per protein to predict the positions of its side-chains.

7.30. Critical assessment of protein-ligand docking methods via Drug Design Data Resource (D3R) Challenge 2015

Participants: Andreas Eisenbarth, Sergei Grudinin.

We participated at the Drug Design Data Resource (D3R) Challenge 2015. In the challenge, we were given protein structures and sets of ligand molecules in order to detect the putative binding poses. The aim was to find the energetically most favourable pose of each ligand relative to a protein. To do so, we first performed the docking simulations using the state-of-the-art software AutoDock Vina, then explored sets of parameters that produced chemically reasonable poses, and finally did the re-scoring using the ComvexPL potential. Later, we critically examined AutoDock Vina sampling method and detected points where it can be improved and also assessed the integration of our inhouse developed ConvexPL scoring algorithm.

7.31. Towards the development of FFT-accelerated flexible fitting methods

Participants: Alexandre Hoffmann, Valerie Perrier, Sergei Grudinin.

We studied a set of new methods for non-rigid molecular fitting. The problem can be formulated as follows : Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two molecular structures (e.g. proteins). We are given $d_1 : \mathbb{R}^3 \mapsto \mathbb{R}$, the electron density of $\mathcal{P}_1$ and $(Y_k \in \mathbb{R}^3)_{k=1\cdots N_{atoms}}$, the average positions of the atoms of $\mathcal{P}_2$. Assuming we can generate an artificial electron density $d_2 : \mathbb{R}^3 \mapsto \mathbb{R}$ from $(Y_k \in \mathbb{R}^3)_{k=1\cdots N_{atoms}}$, our problem is to find a transformation of the atoms $T : \mathbb{R}^{3N_{atoms}} \mapsto \mathbb{R}^{3N_{atoms}}$ that minimizes the L^2 distance between d_1 and d_2.

In image processing, this problem is usually solved using the optimal transport theory, but this method assumes that both densities have the same L^2 norm, which is not necessarily the case for the fitting problem. To solve this problem, one instead starts by splitting T into a rigid transformation T_{rigid} (which is a combination of translation and rotation) and a flexible transformation T_{flex}. Two classes of methods have been developed to find T_{rigid} :

- the first one uses optimization techniques such as gradient descent,
- the second one uses the Fast Fourier Transform (FFT) to compute the Cross Correlation Function (CCF) of d_1 and d_2.

We have developed several algorithms based on the FFT to find T_{rigid} and we have developed two algorithms for flexible molecular fitting that are based on convex and non-convex optimization and the trust region methods. Our tests demonstrate that while one method gives good results for small deformations, the other gives good results for bigger deformations.

We have been also improving the current NMA method (which is essentially a model reduction technique), that is used in other tools such as the flexible fitting to small angle scattering profiles. Finally, we started the development of a method for a harder fitting/docking problem in which only electron density would be known. The basic idea would be to find the C^1-diffeomorphism $T : \mathbb{R}^3 \mapsto \mathbb{R}^3$ that minimizes the L^2 distance between d_1 and d_2.

We developed several stand-alone C++ libraries to solve some of our problems including:

- a non-convex optimization library,
- a normal mode analysis library,
- a fitting library that implements our new methods.

8. Partnerships and Cooperations

8.1. Regional Initiatives

We have funding from the Rhone-Alpes region through an ARC6 grant for the development of parallel algorithms for adaptively restrained particle simulations. This grant is funding Krishna Kant Singh's PhD project.

8.2. National Initiatives

8.2.1. ANR

In 2015, NANO-D had funding from one ANR program:

- **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).

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

8.3.1.1. ADAPT

Type: ERC Starting Grant

Title: Theory and Algorithms for Adaptive Particle Simulation

Programm: FP7

Duration: September 2012 - August 2017

Coordinator: Inria

Inria contact: Stephane Redon

8.4. International Initiatives

8.4.1. Inria Associate Teams not involved in an Inria International Labs

8.4.1.1. PPI-3D

Title: Structure Meets Genomics

International Partner (Institution - Laboratory - Researcher):

Boston University (United States) - Dima Kozakov

Start year: 2015

See also: https://team.inria.fr/nano-d/research/ppi-3d-structure-meets-genomics/

Protein–protein interactions are integral to many mechanisms of cellular control, and therefore their characterization has become an important task for both experimental and computational approaches in systems biology. Genome-wide proteomics studies provide a growing list of putative protein-protein interactions, and demonstrate that most if not all proteins have interacting partners in the cell. A fraction of these interaction has been reliably established, however, one can only identify whether two proteins interact and, in the best cases, which are the individual domains mediating the interaction. A full comprehension of how proteins bind and form complexes can only come from high-resolution three-dimensional structures. While the most complete structural characterization of a complex is provided by X-ray crystallography, protein-protein hetero-complexes constitute less than 6%Â§ of protein structures in the Protein Data Bank. Thus, it is important to develop computational methods that, starting from the structures of component proteins, can determine the structure of their complexes.

The basic problem of predictive protein docking is to start with the structures (or sequences) of unbound component proteins A and B, and to obtain computationally a model of the bound complex AB, as detailed structural knowledge of the interactions facilitates understanding of protein function and mechanism. Our current docking approaches performs ab initio docking of the two structures without the use of any additional information. The goal of this proposal is to speed up docking approaches to tackle genome-scale problems, and utilize additional information on interactions, sequences, and structures that is available for virtually any protein.

This project includes several methodological and application research directions: 1) Developing fast sampling approaches; 2) Development of new scoring functions; 3) Integrative approaches for structure determination.

Overall, during the course of the project we will (i) jointly develop new methodology and algorithms in the field of genomic-scale protein complex prediction; (ii) provide server-based applications built upon services of the Boston team; (iii) and finally develop modular applications coded inside the SAMSON software platform created by the Inria team.

8.4.2. Inria International Partners

8.4.2.1. Informal International Partners

- The Reiher group at ETH Zurich
- The Cherezov Lab, UCS USA
- The Katritch Lab, UCS USA
- ICS-5 FZJ Juelich, Juelich, Germany
- Laboratory for Advanced Studies of Membrane Proteins, MIPT, Moscow, Russia Laboratory of Structural Biology of G-protein Coupled Receptors, MIPT Moscow, Russia

8.5. International Research Visitors

8.5.1. Visits of International Scientists

8.5.1.1. Internships

Aleksandr Katrutsa.

Subject: Convex relaxation for non-convex quadratic optimization problems with applications to side-chain prediction in protein structures.

Institution: MIPT Moscow, Russia.

8.5.2. Visits to International Teams

8.5.2.1. Research stays abroad

- Emilie Neveu visited the Kozakov group at Stony Brook University, NY, USA for three weeks in November 2015.
- Alexandre Hoffmann visited the Kozakov group at Stony Brook University, NY, USA for two weeks in November 2015.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific events organisation

9.1.1.1. Member of the organizing committees

Stephane Redon is a member of the organizing committee of JOBIM 2016

9.1.2. Scientific events selection

9.1.2.1. Reviewer

- Leonard Jaillet was a reviewer ISRR (International Symposium on Robotics Research)

9.1.3. Journal

9.1.3.1. Reviewer - Reviewing activities

- Leonard Jaillet was a reviewer for T-RO (Transactions on Robotics)

9.1.4. Invited talks

- S. Grudinin gave an invited talk titled "Using Machine Learning and Polynomial Expansions to Predict Protein-Protein Interactions " at the 3rd International Conference on Protein and RNA Structure Prediction, 14th - 18th of December 2015, at Punta Cana, Dominican Republic.
- Stephane Redon gave a talk at IMPMC.
- Stephane Redon gave a talk at TEDx Grenoble 2015

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Stephane Redon is teaching INF585 (Introduction to C++) at Ecole polytechnique
- Stephane Redon is part of the teaching team of INF442 (Big data and high-performance computing) at Ecole polytechnique

9.2.2. Supervision

- Leonard Jaillet is advising the PhD of Minh Khoa Nguyen
- Sergei Grudinin is advising the PhD of Alexandre Hoffmann
- Sergei Grudinin advised the PhD of Petr Popov (defended February 2015)
- Stephane Redon is co-advising the PhD of Krishna Kant Singh in collaboration with Jean-Francois Mehaut
- Stephane Redon is advising the PhD of Francois Rousse
- Stephane Redon is advising the PhD of Semeho Edorh
- Stephane Redon is co-advising the PhD of Zofia Trstanova in collaboration with Gabriel Stoltz

9.2.3. Juries

- Stephane Redon was in the PhD committee of Matthieu Dreher

9.3. Popularization

- NANO-D participated to the 2015 Fete de la Science (science fair): SAMSON was used by high school students to perform various activities related to computational nanoscience.
- Emilie Neveu participated with the AirSea team to the 2015 Fete de la Science (science fair).
- Emilie Neveu is part of the group “Cafes Sciences et Citoyens de l’Agglomeration Grenobloise”, which organises public roundtables every month.
- Stephane Redon gave a talk at TEDx Grenoble https://www.youtube.com/watch?v=Atpigmv529E

9.4. Participation to conferences, seminars

- L. Jaillet, and S. Grudinin attended 2015 Inria Scientific Days, 16th June – 19th June, Nancy, France.
- E. Neveu attented the Basel Computational Biology Conference, 7th June – 10th June, in Basel Switzerland where she presented a poster about Pepsi-Dock.
- S. Redon, L. Jaillet, E. Neveu, Minh-Khoa Nguyen, and S. Grudinin attended the 2015 GT Enzymes / GGMM workshop, 25-28 Mai 2015, Séte France. They gave two talks and presented four posters.
- S. Redon, L. Jaillet, and S. Grudinin attended the AlgoSB winter school, 29th Nov – 04th Dec, Cargése, France.

10. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] P. POPOV. *Novel computational methods to predict protein-protein interactions on the structural level*, Université Grenoble Alpes, January 2015, https://tel.archives-ouvertes.fr/tel-01167112

Articles in International Peer-Reviewed Journals

[2] P. BUSLAEV, V. I. GORDELIY, S. GRUDININ, I. Y. GUSHCHIN. *Principal component analysis of lipid molecule conformational changes in molecular dynamics simulations*, in "Journal of Chemical Theory and Computation", January 2016 [*DOI :* 10.1021/ACS.JCTC.5B01106], https://hal.inria.fr/hal-01258167

[3] L. DEBREU, E. NEVEU, E. SIMON, F.-X. LE DIMET, A. VIDARD. *Multigrid solvers and multigrid preconditioners for the solution of variational data assimilation problems*, in "Quarterly Journal of the Royal Meteorological Society", September 2015 [*DOI :* 10.1002/QJ.2676], https://hal.inria.fr/hal-01246349

[4] S. GRUDININ, P. POPOV, E. NEVEU, G. CHEREMOVSKIY. *Predicting Binding Poses and Affinities in the CSAR 2013__2014 Docking Exercises Using the Knowledge-Based Convex-PL Potential*, in "Journal of Chemical Information and Modeling", 2015 [*DOI :* 10.1021/ACS.JCIM.5B00339], https://hal.inria.fr/hal-01258022

[5] P. POPOV, S. GRUDININ. *Knowledge of Native Protein–Protein Interfaces Is Sufficient To Construct Predictive Models for the Selection of Binding Candidates*, in "Journal of Chemical Information and Modeling", September 2015, vol. 55, n^{o} 10, pp. 2242–2255 [*DOI :* 10.1021/ACS.JCIM.5B00372], https://hal.inria.fr/hal-01229886

[6] D. W. RITCHIE, S. GRUDININ. *Spherical polar Fourier assembly of protein complexes with arbitrary point group symmetry*, in "Journal of Applied Crystallography", February 2016, vol. 49, n° 1, pp. 158-167 [*DOI :* 10.1107/S1600576715022931], https://hal.inria.fr/hal-01261402

Other Publications

[7] S. REDON, G. STOLTZ, Z. TRSTANOVA. *Error Analysis of Modified Langevin Dynamics*, January 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01263700

References in notes

[8] B. AHMADI, M. KASSIRIHA, K. KHODABAKHSHI, E. R. MAFI. *Effect of nano layered silicates on automotive polyurethane refinish clear coat*, in "Progress in Organic Coatings", 2007, vol. 60, n° 2, pp. 99 - 104 [*DOI :* 10.1016/J.PORGCOAT.2007.07.008], http://www.sciencedirect.com/science/article/pii/S0300944007001464

[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° 3 Part 1, pp. 380–388, http://dx.doi.org/10.1107/S0108768102003890

[10] F. AMPLE, S. AMI, C. JOACHIM, F. THIEMANN, G. RAPENNE. *A Morse manipulator molecule for the modulation of metallic shockley surface states*, in "Chemical Physics Letters", 2007, vol. 434, pp. 280-285 [*DOI :* 10.1016/J.CPLETT.2006.12.021], http://www.sciencedirect.com/science/article/pii/S0009261406018148

[11] F. AMPLE, C. JOACHIM. *A semi-empirical study of polyacene molecules adsorbed on a Cu(1 1 0) surface*, in "Surface Science", 2006, vol. 600, n° 16, pp. 3243 - 3251 [*DOI :* 10.1016/J.SUSC.2006.06.015], http://www.sciencedirect.com/science/article/pii/S003960280600700X

[12] A. ARKHIPOV, P. FREDDOLINO, K. IMADA, K. NAMBA, K. SCHULTEN. *Coarse-grained molecular dynamics simulations of a rotating bacterial flagellum*, in "Biophysical Journal", 2006, vol. 91, pp. 4589-4597

[13] A. ARNOLD, F. FAHRENBERGER, C. HOLM, O. LENZ, M. BOLTEN, H. DACHSEL, R. HALVER, I. KABADSHOW, F. GÄHLER, F. HEBER, J. ISERINGHAUSEN, M. HOFMANN, M. PIPPIG, D. POTTS, G. SUTMANN. *Comparison of scalable fast methods for long-range interactions*, in "Phys. Rev. E", December 2013, vol. 88, n° 6, 063308 [*DOI :* 10.1103/PHYSREVE.88.063308]

[14] 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° 13, pp. 2865–2877, http://dx.doi.org/10.1002/jcc.21868

[15] 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

[16] H. M. BERMAN, J. WESTBROOK, Z. FENG, G. GILLILAND, T. BHAT, H. WEISSIG, I. N. SHINDYALOV, P. E. BOURNE. *The protein data bank*, in "Nucleic acids research", 2000, vol. 28, n° 1, pp. 235–242

[17] 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

[18] X. BLANC, C. LE BRIS, F. LEGOLL. *Analysis of a prototypical multiscale method coupling atomistic and continuum mechanics*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2005, vol. 39, n^o 04, pp. 797-826, http://dx.doi.org/10.1051/m2an:2005035

[19] 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

[20] D. W. BRENNER. *Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films*, in "Phys. Rev. B", Nov 1990, vol. 42, pp. 9458–9471, http://link.aps.org/doi/10.1103/PhysRevB.42.9458

[21] T. CAGIN, G. WANG, R. MARTIN, G. ZAMANAKOS, N. VAIDEHI, D. T. MAINZ, W. A. GODDARD III. *Multiscale modeling and simulation methods with applications to dendritic polymers*, in "Computational and Theoretical Polymer Science", 2001, vol. 11, n^o 5, pp. 345 - 356 [*DOI :* 10.1016/S1089-3156(01)00026-5], http://www.sciencedirect.com/science/article/pii/S1089315601000265

[22] E. CANCES, F. CASTELLA, P. CHARTIER, E. FAOU, C. L. BRIS, F. LEGOLL, G. TURINICI. *Long-time averaging for integrable Hamiltonian dynamics*, in "Numerische Mathematik", 2005, vol. 100, pp. 211-232, 10.1007/s00211-005-0599-0, http://dx.doi.org/10.1007/s00211-005-0599-0

[23] Q. CHAUDHRY, M. SCOTTER, J. BLACKBURN, B. ROSS, A. BOXALL, L. CASTLE, R. AITKEN, R. WATKINS. *Applications and implications of nanotechnologies for the food sector*, in "Food Additives & Contaminants: Part A", 2008, vol. 25, n^o 3, pp. 241-258, http://www.tandfonline.com/doi/abs/10.1080/02652030701744538

[24] X. CHEN, S. S. MAO. *Titanium Dioxide Nanomaterials: Synthesis, Properties, Modifications, and Applications*, in "ChemInform", 2007, vol. 38, n^o 41, http://dx.doi.org/10.1002/chin.200741216

[25] G.-Y. CHUANG, D. KOZAKOV, R. BRENKE, S. R. COMEAU, S. VAJDA. *DARS (Decoys As the Reference State) potentials for protein-protein docking*, in "Biophysical journal", 2008, vol. 95, n^o 9, pp. 4217–4227

[26] S. R. COMEAU, C. J. CAMACHO. *Predicting oligomeric assemblies: N-mers a primer*, in "Journal of structural biology", 2005, vol. 150, n^o 3, pp. 233–244

[27] S. COOPER, F. KHATIB, A. TREUILLE, J. BARBERO, J. LEE, M. BEENEN, A. LEAVER-FAY, D. BAKER, Z. POPOVIC, F. PLAYERS. *Predicting protein structures with a multiplayer online game*, in "Nature", 2010, vol. 466, pp. 756-760

[28] M. CURRELI, A. H. NADERSHAHI, G. SHAHI. *Emergence of nanomedical devices for the diagnosis and treatment of cancer: the journey from basic science to commercialization*, in "International Journal of Technology Transfer and Commercialisation", 2008, vol. 7, n^o 4, pp. 290-307

[29] E. DARVE. *The Fast Multipole Method: Numerical Implementation*, in "Journal of Computational Physics", 2000

[30] M. DESERNO, C. HOLM. *How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines*, in "J. Chem. Phys.", 1998, vol. 109, pp. 7678 – 7693

[31] M. DESERNO, C. HOLM. *How to mesh up Ewald sums. II. An accurate error estimate for the Particle-Particle-Particle-Mesh algorithm*, in "J. Chem. Phys.", 1998, vol. 109, pp. 7694 – 7701

[32] H. DIETZ, S. M. DOUGLAS, W. M. SHIH. *Folding DNA into Twisted and Curved Nanoscale Shapes*, in "Science", 2009, vol. 325, n^{o} 5941, pp. 725-730 [*DOI :* 10.1126/SCIENCE.1174251], http://www.sciencemag.org/content/325/5941/725.abstract

[33] S. J. FLEISHMAN, T. A. WHITEHEAD, D. C. EKIERT, C. DREYFUS, J. E. CORN, E.-M. STRAUCH, I. A. WILSON, D. BAKER. *Computational Design of Proteins Targeting the Conserved Stem Region of Influenza Hemagglutinin*, in "Science", 2011, vol. 332, n^{o} 6031, pp. 816-821 [*DOI :* 10.1126/SCIENCE.1202617], http://www.sciencemag.org/content/332/6031/816.abstract

[34] G. FOX-RABINOVICH, B. BEAKE, K. YAMAMOTO, M. AGUIRRE, S. VELDHUIS, G. DOSBAEVA, A. ELFIZY, A. BIKSA, L. SHUSTER. *Structure, properties and wear performance of nano-multilayered TiAl-CrSiYN/TiAlCrN coatings during machining of Ni-based aerospace superalloys*, in "Surface and Coatings Technology", 2010, vol. 204, pp. 3698 - 3706 [*DOI :* 10.1016/J.SURFCOAT.2010.04.050], http://www.sciencedirect.com/science/article/pii/S0257897210003178

[35] M. GOLDBERG, R. LANGER, X. JIA. *Nanostructured materials for applications in drug delivery and tissue engineering*, in "Journal of Biomaterials Science, Polymer Edition", 2007, vol. 18, n^{o} 3, pp. 241-268 [*DOI :* DOI:10.1163/156856207779996931], http://www.ingentaconnect.com/content/vsp/bsp/2007/00000018/00000003/art00001

[36] H. HASSANIEH, P. INDYK, D. KATABI, E. PRICE. *Simple and Practical Algorithm for Sparse Fourier Transform*, in "Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms", SODA '12, SIAM, 2012, pp. 1183–1194, http://dl.acm.org/citation.cfm?id=2095116.2095209

[37] J.-H. HE. *An elementary introduction to recently developed asymptotic methods and nanomechanics in textile engineering*, in "International Journal of Modern Physics B", 2008, vol. 22, n^{o} 21, pp. 3487-3578

[38] S. HELVEG. *Structure and Dynamics of Nanocatalysts*, in "Microscopy and Microanalysis", 2010, vol. 16, n^{o} Supplement S2, pp. 1712-1713, http://dx.doi.org/10.1017/S1431927610055005

[39] A. HEYDEN, D. G. TRUHLAR. *Conservative Algorithm for an Adaptive Change of Resolution in Mixed Atomistic/Coarse-Grained Multiscale Simulations*, in "Journal of Chemical Theory and Computation", 2008, vol. 4, n^{o} 2, pp. 217-221, http://pubs.acs.org/doi/abs/10.1021/ct700269m

[40] J. H. HOLLAND. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence.*, U Michigan Press, 1975

[41] 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

[42] H. HWANG, T. VREVEN, J. JANIN, Z. WENG. *Protein–protein docking benchmark version 4.0*, in "Proteins: Structure, Function, and Bioinformatics", 2010, vol. 78, n^{o} 15, pp. 3111–3114

[43] J. JANIN, K. HENRICK, J. MOULT, L. T. EYCK, M. J. STERNBERG, S. VAJDA, I. VAKSER, S. J. WODAK. *CAPRI: a critical assessment of predicted interactions*, in "Proteins: Structure, Function, and Bioinformatics", 2003, vol. 52, n^{o} 1, pp. 2–9

[44] C. JOACHIM, H. TANG, F. MORESCO, G. RAPENNE, G. MEYER. *The design of a nanoscale molecular barrow*, in "Nanotechnology", 2002, vol. 13, n^{o} 3, 330 p. , http://stacks.iop.org/0957-4484/13/i=3/a=318

[45] L. KALÉ, R. SKEEL, M. BHANDARKAR, R. BRUNNER, A. GURSOY, N. KRAWETZ, J. PHILLIPS, A. SHINOZAKI, K. VARADARAJAN, K. SCHULTEN. *NAMD2: Greater Scalability for Parallel Molecular Dynamics*, in "Journal of Computational Physics", 1999, vol. 151, n^{o} 1, pp. 283 - 312 [*DOI :* 10.1006/JCPH.1999.6201], http://www.sciencedirect.com/science/article/pii/S0021999199962010

[46] J. KENNEDY. *Particle swarm optimization*, in "Encyclopedia of machine learning", Springer, 2011, pp. 760–766

[47] J. KOLAFA, J. W. PERRAM. *Cutoff errors in the Ewald summation formulae for point charge systems*, in "Mol. Simul.", 1992, vol. 9, n^{o} 5, pp. 351–368, https://dx.doi.org/10.1080%2F08927029208049126

[48] D. KOZAKOV, R. BRENKE, S. R. COMEAU, S. VAJDA. *PIPER: an FFT-based protein docking program with pairwise potentials*, in "Proteins: Structure, Function, and Bioinformatics", 2006, vol. 65, n^{o} 2, pp. 392–406

[49] D. KOZAKOV, K. H. CLODFELTER, S. VAJDA, C. J. CAMACHO. *Optimal clustering for detecting near-native conformations in protein docking*, in "Biophysical journal", 2005, vol. 89, n^{o} 2, pp. 867–875

[50] Z. LI, H. A. SCHERAGA. *Monte Carlo-minimization approach to the multiple-minima problem in protein folding*, in "Proceedings of the National Academy of Sciences", 1987, vol. 84, n^{o} 19, pp. 6611-6615, http://www.pnas.org/content/84/19/6611.abstract

[51] L. LO, Y. LI, K. YEUNG, C. YUEN. *Indicating the development stage of nanotechnology in the textile and clothing industry*, in "International Journal of Nanotechnology", 2007, vol. 4, n^{o} 6, pp. 667-679

[52] W. LU, C. M. LIEBER. *Nanoelectronics from the bottom up*, in "Nature materials", 2007, vol. 6, n^{o} 11, pp. 841-850

[53] A. MAY, M. ZACHARIAS. *Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein–protein docking*, in "Proteins: Structure, Function, and Bioinformatics", 2008, vol. 70, n^{o} 3, pp. 794–809

[54] I. H. MOAL, P. A. BATES. *SwarmDock and the use of normal modes in protein-protein docking*, in "International journal of molecular sciences", 2010, vol. 11, n^{o} 10, pp. 3623–3648

[55] S. O. NIELSEN, P. B. MOORE, B. ENSING. *Adaptive Multiscale Molecular Dynamics of Macromolecular Fluids*, in "Phys. Rev. Lett.", Dec 2010, vol. 105, 237802, http://link.aps.org/doi/10.1103/PhysRevLett.105.237802

[56] A. NIKITIN, X. LI, Z. ZHANG, H. OGASAWARA, H. DAI, A. NILSSON. *Hydrogen Storage in Carbon Nanotubes through the Formation of Stable C-H Bonds*, in "Nano Letters", 2008, vol. 8, n^{o} 1, pp. 162-167, PMID: 18088150, http://pubs.acs.org/doi/abs/10.1021/nl072325k

[57] P. POPOV, S. GRUDININ. *Rapid determination of RMSDs corresponding to macromolecular rigid body motions*, in "Journal of computational chemistry", 2014, vol. 35, n° 12, pp. 950–956

[58] P. POPOV, S. GRUDININ. *Knowledge of Native Protein–Protein Interfaces Is Sufficient To Construct Predictive Models for the Selection of Binding Candidates*, in "Journal of chemical information and modeling", 2015, vol. 55, n° 10, pp. 2242–2255

[59] P. POPOV, D. W. RITCHIE, S. GRUDININ. *DockTrina: Docking triangular protein trimers*, in "Proteins: Structure, Function, and Bioinformatics", 2014, vol. 82, n° 1, pp. 34–44

[60] M. PRAPROTNIK, L. DELLE SITE, K. KREMER. *Adaptive resolution scheme for efficient hybrid atomistic-mesoscale molecular dynamics simulations of dense liquids*, in "Phys. Rev. E", Jun 2006, vol. 73, 066701, http://link.aps.org/doi/10.1103/PhysRevE.73.066701

[61] M. PRAPROTNIK, S. MATYSIAK, L. D. SITE, K. KREMER, C. CLEMENTI. *Adaptive resolution simulation of liquid water*, in "Journal of Physics: Condensed Matter", 2007, vol. 19, n° 29, 292201, http://stacks.iop.org/0953-8984/19/i=29/a=292201

[62] M. PRAPROTNIK, L. D. SITE, K. KREMER. *A macromolecule in a solvent: Adaptive resolution molecular dynamics simulation*, in "The Journal of Chemical Physics", 2007, vol. 126, n° 13, 134902 [*DOI :* 10.1063/1.2714540], http://link.aip.org/link/?JCP/126/134902/1

[63] M. PRAPROTNIK, L. D. SITE, K. KREMER. *Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution*, in "Annual Review of Physical Chemistry", 2008, vol. 59, n° 1, pp. 545-571, http://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.59.032607.093707

[64] P. PROCACCI, T. DARDEN, M. MARCHI. *A Very Fast Molecular Dynamics Method To Simulate Biomolecular Systems with Realistic Electrostatic Interactions*, in "The Journal of Physical Chemistry", 1996, vol. 100, n° 24, pp. 10464-10468, http://pubs.acs.org/doi/abs/10.1021/jp960295w

[65] X. QIAN, T. SCHLICK. *Efficient multiple-time-step integrators with distance-based force splitting for particle-mesh-Ewald molecular dynamics simulations*, in "Journal of Chemical Physics", 2002, vol. 116, pp. 5971-5983

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

[67] D. W. RITCHIE, D. KOZAKOV, S. VAJDA. *Accelerating and focusing protein–protein docking correlations using multi-dimensional rotational FFT generating functions*, in "Bioinformatics", 2008, vol. 24, n° 17, pp. 1865–1873

[68] M. C. ROCO. *The long view of nanotechnology development: the National Nanotechnology Initiative at 10 years*, in "Journal of Nanoparticle Research", 2010

[69] B. ROOKS. *A shorter product development time with digital mock-up*, in "Assembly Automation", 1998, vol. 18, n° 1, pp. 34-38 [*DOI :* DOI:10.1108/01445159810201405], http://www.ingentaconnect.com/content/mcb/033/1998/00000018/00000001/art00004

[70] 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

[71] R. E. RUDD. *Coarse-Grained Molecular Dynamics for Computer Modeling of Nanomechanical Systems*, in "International Journal for Numerical Methods in Engineering", 2004

[72] A. SHIH, P. FREDDOLINO, A. ARKHIPOV, K. SCHULTEN. *Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations*, in "Journal of Structural Biology", 2007, vol. 157, pp. 579-592

[73] Y. SHIRAI, A. J. OSGOOD, Y. ZHAO, Y. YAO, L. SAUDAN, H. YANG, C. YU-HUNG, L. B. ALEMANY, T. SASAKI, J.-F. MORIN, J. M. GUERRERO, K. F. KELLY, J. M. TOUR. *Surface-Rolling Molecules*, in "Journal of the American Chemical Society", 2006, vol. 128, n^{o} 14, pp. 4854-4864, PMID: 16594722, http://pubs.acs.org/doi/abs/10.1021/ja058514r

[74] H. V. SORENSEN, C. S. BURRUS. *Efficient computation of the DFT with only a subset of input or output points*, in "IEEE Transactions on Signal Processing", 1993, vol. 41, n^{o} 3, pp. 1184–1200, http://dx.doi.org/10.1109/78.205723

[75] E. G. STEIN, L. M. RICE, A. T. BRÜNGER. *Torsion-Angle Molecular Dynamics as a New Efficient Tool for NMR Structure Calculation*, in "Journal of Magnetic Resonance", 1997, vol. 124, n^{o} 1, pp. 154 - 164 [*DOI :* 10.1006/JMRE.1996.1027], http://www.sciencedirect.com/science/article/pii/S1090780796910277

[76] R. STORN, K. PRICE. *Differential evolution-a simple and efficient adaptive scheme for global optimization over continuous spaces*, ICSI Berkeley, 1995, vol. 3

[77] X. SUN, Z. LIU, K. WELSHER, J. ROBINSON, A. GOODWIN, S. ZARIC, H. DAI. *Nano-graphene oxide for cellular imaging and drug delivery*, in "Nano Research", 2008, vol. 1, pp. 203-212, 10.1007/s12274-008-8021-8, http://dx.doi.org/10.1007/s12274-008-8021-8

[78] G. SUTMANN. *Molecular Dynamics - Vision and Reality*, in "Computational Nanoscience: Do It Yourself!", Jülich, J. GROTENDORST, S. BLÜGEL, D. MARX (editors), NIC Series, John von Neumann Institute for Computing, February 2006, vol. 31, pp. 159 – 194, Lecture Notes

[79] D. TOMALIA, L. REYNA, S. SVENSON. *Dendrimers as multi-purpose nanodevices for oncology drug delivery and diagnostic imaging*, in "Biochemical Society Transactions", 2007, vol. 35, pp. 61-67

[80] N. VAIDEHI, W. A. GODDARD. *Domain Motions in Phosphoglycerate Kinase using Hierarchical NEIMO Molecular Dynamics Simulations*, in "The Journal of Physical Chemistry A", 2000, vol. 104, n^{o} 11, pp. 2375-2383, http://pubs.acs.org/doi/abs/10.1021/jp991985d

[81] V. VENKATRAMAN, D. W. RITCHIE. *Flexible protein docking refinement using pose-dependent normal mode analysis*, in "Proteins: Structure, Function, and Bioinformatics", 2012, vol. 80, n^{o} 9, pp. 2262–2274

[82] T. VETTOREL, A. Y. GROSBERG, K. KREMER. *Statistics of polymer rings in the melt: a numerical simulation study*, in "Physical Biology", 2009, vol. 6, n^{o} 2, 025013 p. , http://stacks.iop.org/1478-3975/6/i=2/a=025013

[83] W. YANG. *Direct calculation of electron density in density-functional theory*, in "Phys. Rev. Lett.", Mar 1991, vol. 66, pp. 1438–1441, http://link.aps.org/doi/10.1103/PhysRevLett.66.1438

[84] W. YANG. *Electron density as the basic variable: a divide-and-conquer approach to the ab initio computation of large molecules*, in "Journal of Molecular Structure: THEOCHEM", 1992, vol. 255, n^{o} 0, pp. 461 - 479 [*DOI :* 10.1016/0166-1280(92)85024-F], http://www.sciencedirect.com/science/article/pii/016612809285024F

[85] S. W. DE LEEUW, J. W. PERRAM, E. R. SMITH. *Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants*, in "Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences", 1980, vol. 373, n^{o} 1752, pp. 27–56 [*DOI :* 10.1098/RSPA.1980.0135], http://rspa.royalsocietypublishing.org/content/373/1752/27

[86] A. C. T. VAN DUIN, S. DASGUPTA, F. LORANT, W. A. GODDARD. *ReaxFF: A Reactive Force Field for Hydrocarbons*, in "The Journal of Physical Chemistry A", 2001, vol. 105, n^{o} 41, pp. 9396-9409, http://pubs.acs.org/doi/abs/10.1021/jp004368u