
IN PARTNERSHIP WITH:
**CNRS**

Activity Report 2016

# Project-Team NANO-D

## Algorithms for Modeling and Simulation of Nanosystems

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

RESEARCH CENTER
**Grenoble - Rhône-Alpes**

THEME
**Numerical schemes and simulations**

# Table of contents

## 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
5.5. - Materials

# 1. Members

**Research Scientists**

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

**Engineers**

Svetlana Artemova [Inria, until Feb 2016, granted by ERC (European Research Council Excecutive Agency)]
Yassine Naimi [Inria, from Nov 2016]

**PhD Students**

Semeho Edorh [Inria]
Alexandre Hoffmann [Univ. Grenoble I]
Maria Kadukova [from Feb 2016]
Minh Khoa Nguyen [Inria]
Guillaume Pages [Inria]
Francois Rousse [Inria]
Krishna Kant Singh [Inria, granted by CONSEIL REGION AUVERGNE RHONE ALPES]
Zofia Trstanova [Inria, granted by ERC (European Research Council Excecutive Agency)]

**Post-Doctoral Fellows**

Clement Beitone [Inria, from Jun 2016]
Silvia Dias Pinto [Inria, until Jun 2016]
Dmitriy Marin [Inria, from Aug 2016]
Emilie Neveu [Inria, until Mar 2016]

**Visiting Scientists**

Dmytro Kozakov [Jan 2016]
Dzmitry Padhorny [Jan 2016]

**Administrative Assistants**

Julie Bourget [Inria, until Jul 2016]
Maria Immaculada Presseguer [Inria]

**Others**

Andreas Eisenbarth [Inria, until Apr 2016]
Jocelyn Gate
Mikhail Karasikov [Inria, from Aug 2016]
Aleksandr Katrutca [Inria, until Jan 2016]
Lucien Mathieu [Inria, until Jan 2016]
Alisa Patotskaya [Inria, from Feb 2016 until Jun 2016]

# 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* [22], *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 [67], for neighbor search between large rigid molecules [21], and for bond-order reactive force-fields [25]. 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 [23] and over 500,000 structures of small molecules stored in the Cambridge Structural Database [17]. 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 [43]. 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 [64]. 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 [66].

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 [35]. 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 [37], a wheelbarrow molecule [44], a nano-car [70], a Morse molecule [18], etc. Typically, these designs are optimized using semi-empirical quantum mechanics calculations, such as the semi-empirical ASED+ calculation technique [19].

While impressive, these are but two examples of the nanoscience revolution already impacting numerous fields, including electronics and semiconductors [53], textiles [52], [40], energy [55], food [29], drug delivery [39], [72], chemicals [41], materials [30], the automotive industry [16], aerospace and defense [38], medical devices and therapeutics [33], medical diagnostics [73], etc. According to some estimates, the world market for nanotechnology-related products and services will reach one trillion dollars by 2015 [65]. 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 [75], removing some degrees of freedom (e.g. keeping torsion angles only [51], [71]), coarse-graining [74], [68], [20], [69], multiple time step methods [61], [62], fast multipole methods [34], parallelization [46], averaging [28], multi-scale modeling [27], [24], reactive force fields [26], [78], interactive multiplayer games for predicting protein structures [32], 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 [76], [77].

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 [57], [58], [59], [60], the adaptive multiscale method [54], and the adaptive partitioning of the Lagrangian method [42]. 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*[22], *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 [67], for neighbor search between large rigid molecules [21], and for bond-order reactive force-fields [25]. 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 [23] and over 500,000 structures of small molecules stored in the Cambridge Structural Database [17]. 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 [43]. 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 [64]. 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.

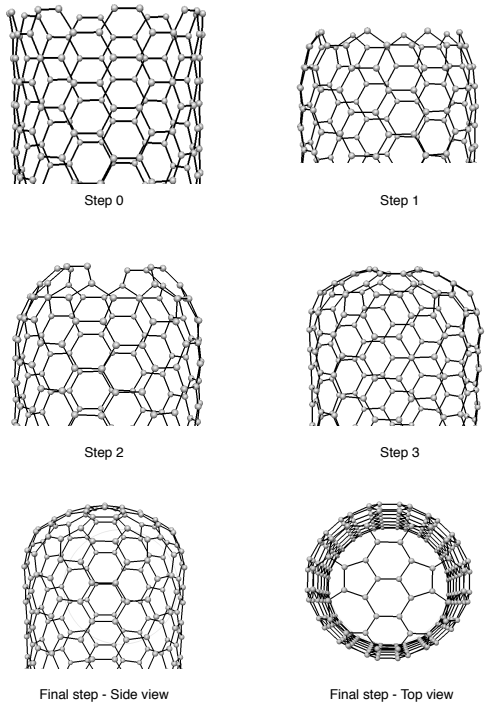


*Figure 1. Snapshots of a nanotube capping process with the adaptive interactive modeler. Thanks to the adaptive methodology, this operation can be done in a few minutes.*

## 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.

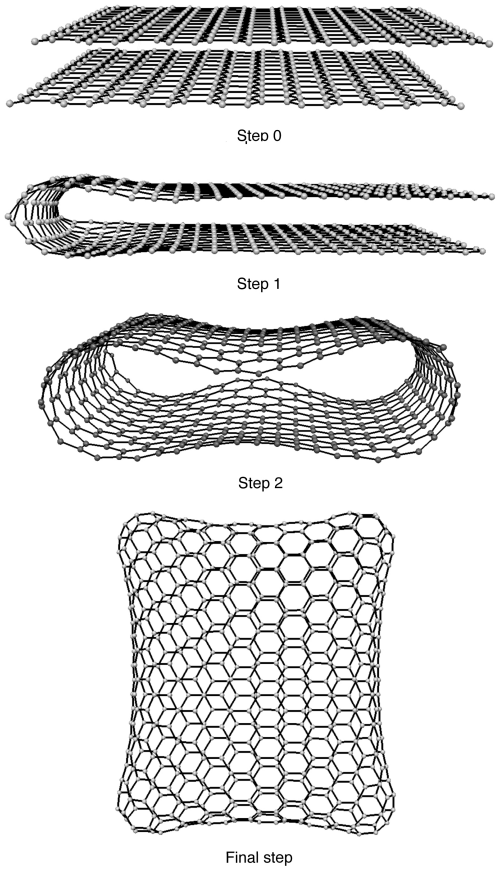


*Figure 2. Different steps to prototype a "nano-pillow" with the adaptive interactive modeler.*

# 5. New Software and Platforms

## 5.1. SAMSON

Software for Adaptive Modeling and Simulation Of Nanosystems
KEYWORDS: Bioinformatics - Simulation - Nanosystems - Structural Biology - Chemistry
SCIENTIFIC DESCRIPTION

Please refer to https://www.samson-connect.net

FUNCTIONAL DESCRIPTION

SAMSON is a software platform for real-time modelling and simulation of natural or artificial nanosystems. The objective is to make SAMSON a generic application for computer-aided design of nanosystems, similar to existing applications for macrosystem prototyping (CATIA, SolidWorks, etc.).

- Contact: Stéphane Redon
- URL: http://www.samson-connect.net/

# 6. New Results

## 6.1. Development of a novel minimization method

**Participants:** Clement Beitone, Stephane Redon.

Finding the optimized configuration of a system of particles so that it minimizes the energy of the system is a very common task in the field of particles simulation. More precisely, we are interested in finding the closest atomic structure located at a minima on the Potential Energy Surface (PES) starting from a given initial configuration. Achieving faster but reliable minimizations of such systems help to enhance a wide range of applications in molecular dynamics. To improve the efficiency of the convergence some authors have proposed alternative methods to the steepest descent algorithm; for example, the conjugate gradient technique or the Fast Inertial Relaxation Engine (FIRE).

In this work, we are developing a novel method that helps to increase the efficiency and the reliability of existing optimizers, *e.g.* FIRE and Interactive Modelling (IM).

We have implemented the modified versions of these algorithms along with others optimization algorithms like L-BFGS and Conjugate Gradient as state updaters in SAMSON. To assess the efficiency of the proposed approaches we have developed an App in SAMSON that allows us to reliably and conveniently probe several criteria during the minimization process (Figure 3).

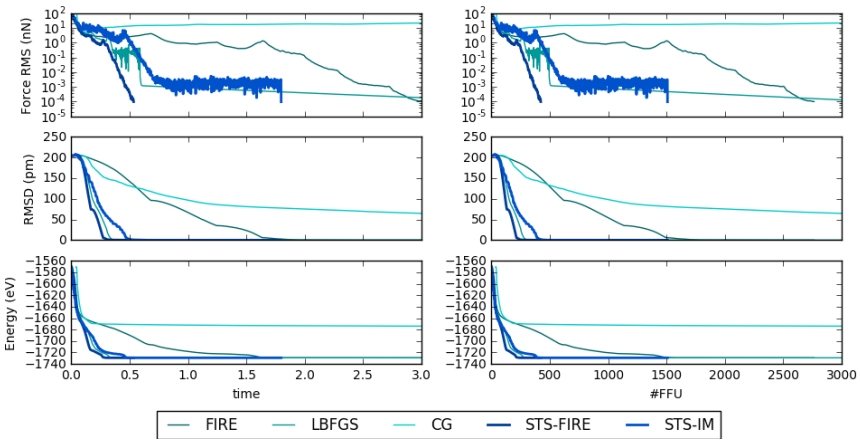


*Figure 3. Comparison of different optimizers with the proposed methods on the fullerene C240. For this experiment the force field used to model the interactions between the atoms is the Brenner potential.*

## 6.2. Parallel algorithms for adaptive molecular dynamics simulations

**Participants:** Dmitriy Marin, Stephane Redon.

We have developed a parallel implementation of Adaptively Restrained Particle Simulations (ARPS) in LAMMPS Molecular Dynamics Simulator with the usage of Kokkos [1] package. The main idea of the ARPS method [22] is to speed up particle simulations by adaptively switching on and off positional degrees of freedom, while letting momenta evolve; this is done by using adaptively restrained Hamiltonian. The developed parallel implementation allows us to run LAMMPS with ARPS integrator on central processing units (CPU), graphics processing units (GPU), or many integrated core architecture (MIC). We modified the ARPS algorithm for efficient usage of GPU and many-core CPU, e.g. all computations were parallelized for efficient calculations on computational device; communications between host and device were decreased.

To measure speed up of the developed parallel implementation we used several benchmarks and heterogeneous computational systems with next parameters: 2x CPU Intel Xeon E5-2680 v3 (24 cores in total), GPU Nvidia Quadro K4200, GPU Nvidia Tesla K20c. Results on the speed up in comparison with serial ARPS code for one of the benchmarks (Lennard–Jones liquid, 515K atoms, $\sim$1% of particles switches their state at each timestep from active to restrained or from restrained to active) are shown in Figure 4. It can be seen, that for small number of CPU cores the speed up is almost constant for all the percentage of active atoms in the system. But for large number of CPU cores and for GPUs the speed up is decreasing with decreasing percentage of active atoms, because of divergence of threads and limited occupancy. The achieved speed up on 20 CPU-cores is up to 14 times, on GPU Nvidia Tesla K20c is up to 24 times.

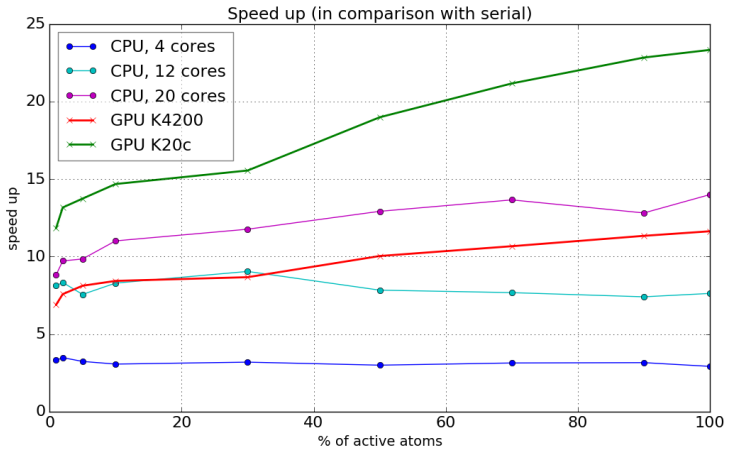


*Figure 4. The parallel ARPS results*

## 6.3. Adaptive Algorithms for Orbital-Free Density Functional Theory

**Participants:** Francois Rousse, Stephane Redon.

[1] The Kokkos package is based on Kokkos library, which is a templated C++ library that provides two key abstractions: it allows a single implementation of an application kernel to run efficiently on different hardware, such as a many-core CPU, GPU, or MIC; it provides data abstractions to adjust (at compile time) the memory layout of basic data structures — like 2d and 3d arrays — for performance optimization on different platforms. These abstractions are set at build time (during compilation of LAMMPS).

The SAMSON App developed to simulate molecular systems with an adaptive version of OF-DFT has been continued. It has been tested on several small systems : atoms, dimers, etc. The errors found on the energies and the bond length found were coherent with the predictive characteristics of OF-DFT and with other OF-DFT softwares like PROFESS.

The pseudopotentials computed by the Carter Group of Princeton (who developed PROFESS) have been implemented in the SAMSON App. The electronic densities became smoother and the predictions were improved, but it restricted the applicability of the SAMSON App since the pseudopotentials were computed only for the elements of the columns III (like aluminum) and V (like Potassium) of the periodic table.

Several optimization algorithms have been tried : projected gradient, Primal-Dual, Lagrangian multiplier improved with a penalization, different nonlinear conjugate gradient minimization algorithms ... None of them showed a clear superiority on the other in both stability and speed. Currently, we use the projected gradient since it is the most stable.

We have implemented an interaction model in SAMSON based on the OF-DFT code and tested its ability to predict the geometry of system on a small crystal of aluminum. The crystal contracted itself, which is coherent with the OF-DFT theory, since it tends to underestimate bond lengths, and with the surface tension, since it tends to minimize the surface of the system. The next step will be to make this interaction model adaptive and measure how much time is gained.

## 6.4. A crystal creator app

**Participants:** Francois Rousse, Stephane Redon.

We developed a new SAMSON Element able to generate models of crystals. The user can either write its own unit cell or load it from a CIF file ("Crystallographic Information File"). Once written or imported, this unit cell can be replicated again in every direction to generate a whole crystal. As the important characteristics of crystals often comes from the defects, the replacements and the insertions, these repetitions of unit cells are not mere copies but are whole new unit cells generated again each time. Thus a crystal with enough unit cells shall have the right proportion of elements, with the right amount of defects, replacements and insertions, randomly disposed. In the document view, the unit cells are separated to ease the manipulation of the crystal. Last, it allows the user to cut the crystal on the planes given by Miller indexes.

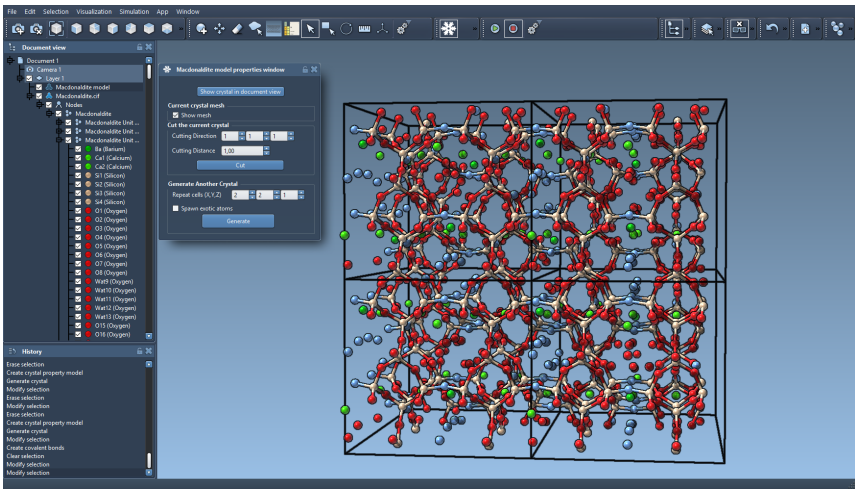

*Figure 5. A Macdonaldite crystal generated in SAMSON*

## 6.5. Software development process improvements

**Participants:** Jocelyn Gate, Stephane Redon.

We set up a Jenkins server on a virtual machine at Inria. The server is accessible to the team and is able to build and generate everything related to SAMSON. This Jenkins server is linked to diferents slaves, located in our offices:

- Window 7 / Windows 10
- Fedora 21 / Fedora 25 / Ubuntu 16.04
- MacOs 10.10.5

Slave machines are used by the Jenkins server to build the specified version of SAMSON, generate the associated SDK, build all SAMSON elements that are specified on Jenkins and upload everything to our private version of SAMSON Connect. Thanks to this, the team has access each day to the latest developments.

In order to efficiently upload everything from slaves nodes, Jenkins uses a private helpers that is able to communicate with SAMSON-Connect, and that knows every SAMSON files format.

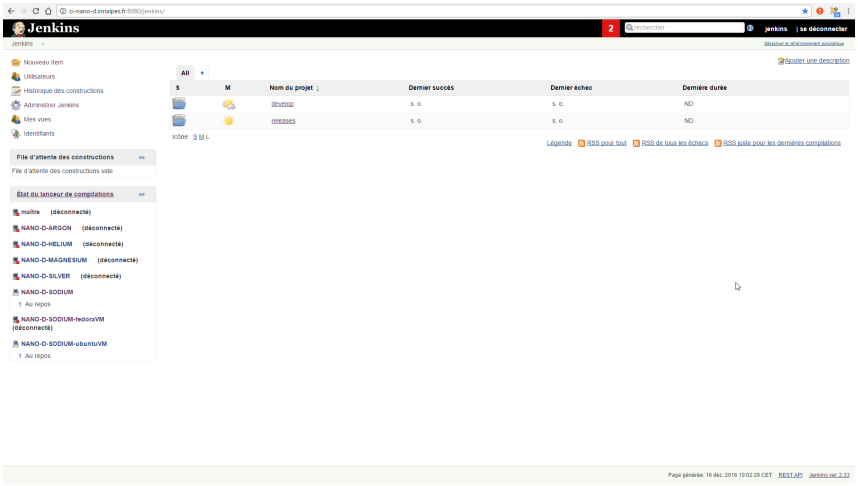


*Figure 6. The jenkins interface*

We developed a private, command line SAMSON helper that is able to do everything concerning the packing and the uploading of new versions of SAMSON, the SAMSON SDK and the installer to SAMSON Connect. It can:

- Upload the SAMSON or SAMSON-SDK packaged file to SAMSON-Connect (adding a new version of SAMSON/SAMSON-SDK).
- Upload the SAMSON or SAMSON-SDK Setup executable to SAMSON-Connect.
- Package the SAMSON elements of a developer to .element files.
- Upload .element files to SAMSON Connect.

## 6.6. Updates to SAMSON and SAMSON Connect

**Participants:** Jocelyn Gate, Stephane Redon.

To be able to know if SAMSON works well on users computers, we added some logging features inside SAMSON, SAMSON installers and SAMSON Helpers. Thanks to this functionality, users may accept to send logs when bugs are found. For example, if SAMSON crashes on a user computer, a log is generated, anonymized, and automatically sent to the SAMSON Connect web service. If SAMSON crashes because of a SAMSON Element, an email is sent to the author of the SAMSON Element. If a new user tries to install SAMSON or the SAMSON SDK, a log is sent to the SAMSON Connect web service.

We also added the possibility for users to configure proxy access to SAMSON Connect.

These functionalities will be part of the upcoming 0.6.0 release of SAMSON.

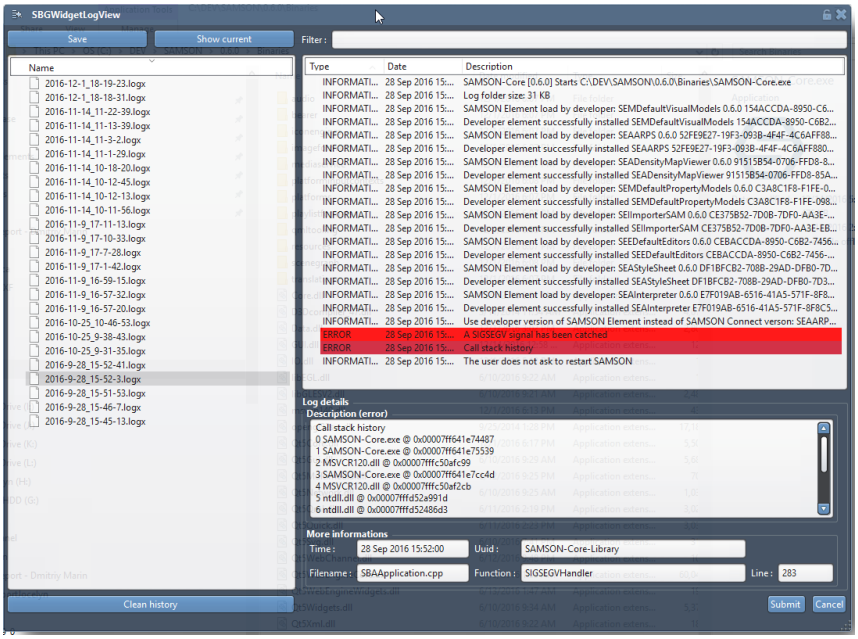


*Figure 7. The SAMSON log interface*

## 6.7. As-Rigid-As-Possible molecular interpolation paths

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

We submitted a paper describing a new method to generate interpolation paths between two given molecular conformations. It applies the As-Rigid-As-Possible (ARAP) from the field of computer graphics to manipulate complex meshes while preserving their essential structural characteristics. The adaptation of ARAP interpolation approach to the case of molecular systems was presented. Experiments were conducted on a large set of benchmarks and the performance was compared between ARAP interpolation and linear interpolation. They show that ARAP interpolation generates more relevant paths, that preserve bond lengths and bond angles better.

## 6.8. As-Rigid-As-Possible molecular interpolation paths

**Participants:** Krishna Kant Singh, Stephane Redon.

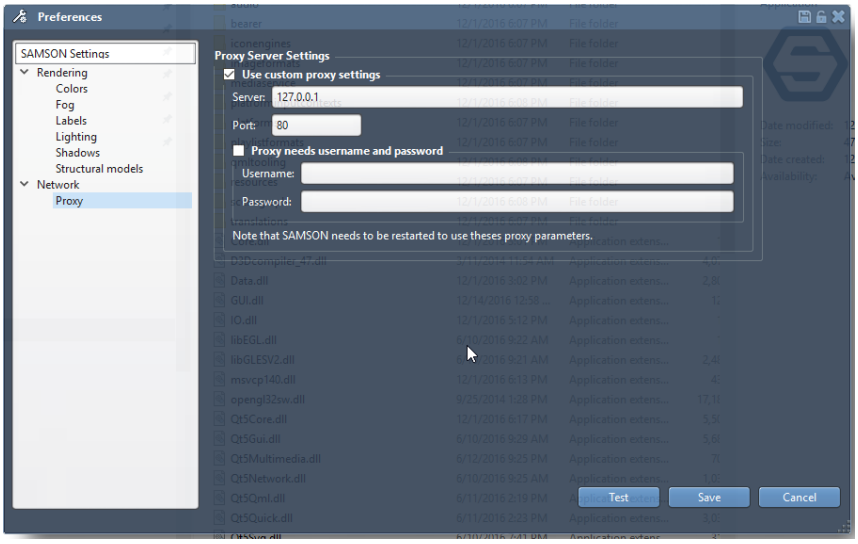


*Figure 8. The proxy setting interface*

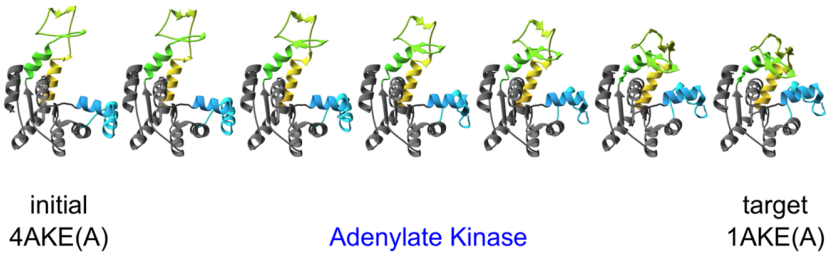


*Figure 9. The morphing path for Adenylate Kinase from 4AKE (chain A) to 1AKE (chain A) by ARAP interpolation:*

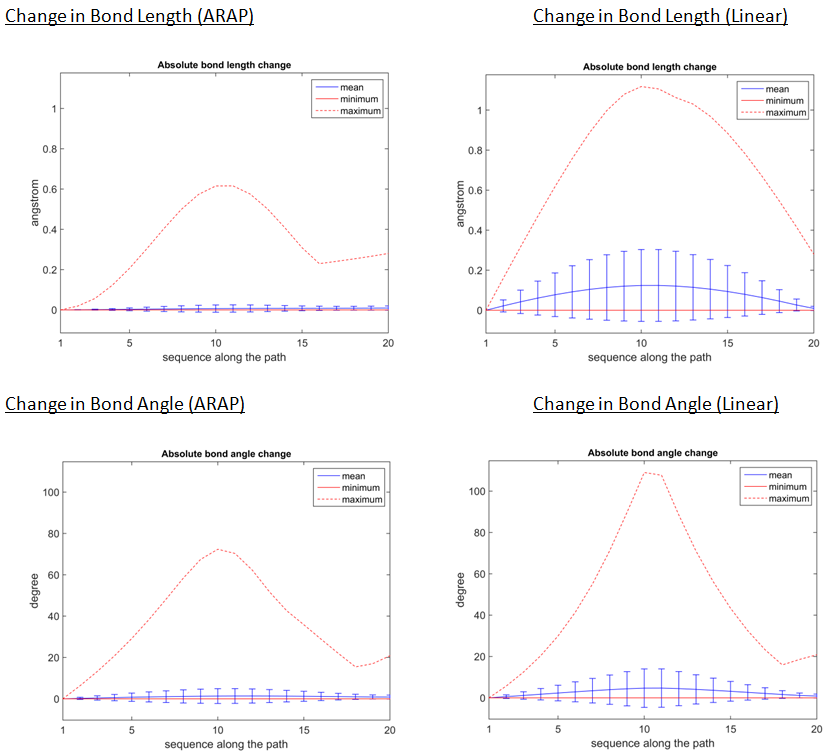


*Figure 10. Comparison of ARAP and linear interpolation for preserving structural characteristics of adenylate kinase*

We have continued our work on the development of parallel adaptively restrained particle simulations. We proposed new algorithms to compute forces involving active particles faster. These algorithms involved construction of the Active Neighbor List (ANL) and incremental force computations. These algorithms have advantages over the state-of-the-art methods for simulating a system using Adaptively Restrained Molecular Dynamics (ARMD). Previously proposed algorithms required at-least $60\%$ restrained particles in order to achieve speed up. In new algorithms, we overcome this limitation and speed up can be achieve with $10\%$ retrained particles. We implemented our algorithm in the popular molecular dynamics package LAMMPS and submitted our results in the *Computer Physics Communications* Journal [2]. Figure 11 show that speed-up can be achieved for more than $10\%$ of the particles are restrained. We also achieved significant speed up in constructing the ANL (figure 12).

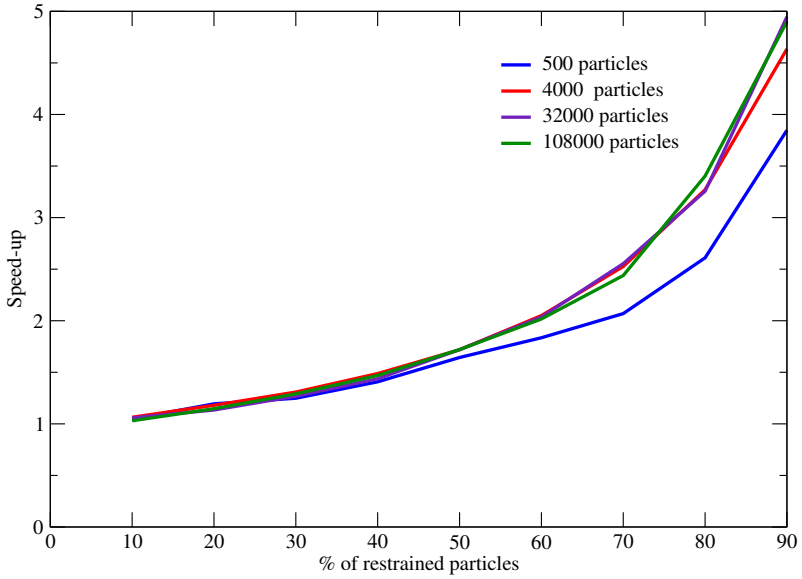


*Figure 11. speed up using ARMD on different benchmark*

## 6.9. Refining the energy landscape sampling of protein-protein associations

**Participants:** Dmytro Kozakov, Leonard Jaillet.

PIPER is a FFT-based protein docking program with pairwise potentials. It combines a systematic sampling procedure with an original pairwise potential that provides an energy landscape representation through a set of samples [48].

In [49], an experimental validation of the complexes obtained with PIPER, has been made possible thanks to the PRE method [31]. PRE (NMR paramagnetic relaxation enhancement) is an experimental technique used to characterize the states present for a given system. Hence, it characterizes the accessible region of the energy landscape corresponding to a given protein. For this, it introduces paramagnetic labels (tags) one at a time at few sites on one protein. The method then relies on measures of the transverse paramagnetic relaxation enhancement rates of the backbone amide protons (HN) of the partner protein. These value correspond to the weighted averages of the values for the various states present. One advantage of PRE is that it is nicely sensitive to lowly populated states.

[2] K.K. Singh, S. Redon, Adaptively Restrained Molecular Dynamics in LAMMPS, Submitted to *Computer Physics Communications*.

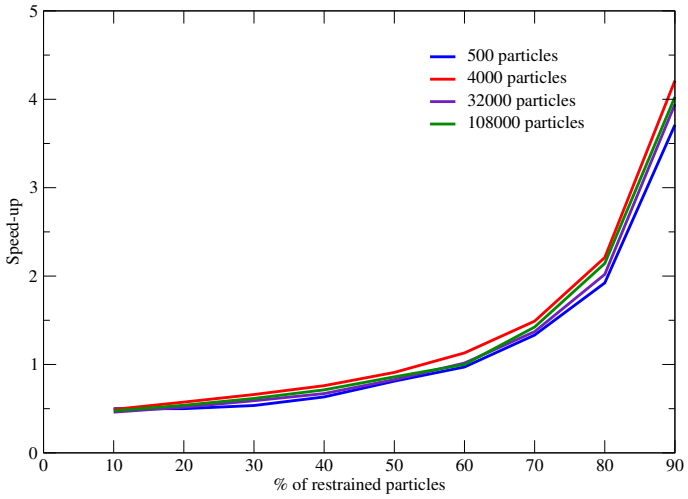


*Figure 12. Obtained speed up in constructing the ANL.*

In [49] the values measured obtained from a set of PIPER output have been compared to those obtained when using only the native state. It appears that using all the PIPER states give a better correlation respect to experimental results than when using only the native state.

In this context, our objective is to refine the energy landscape description by filtering some of the PIPER output complexes in order to improve even further the correlation with experimental measures. The method is developed as a module of the SAMSON software package (http://www.samson-connect.net/).

We have proposed a refinement from process of PIPER complexes based on two criterions: a RMSD-based filtering and an energy-based filtering.

The RMSD-filtering first creates a graph of connected component by connecting a pair of complexes if their distance is lower than a given RMSD threshold. Such a process forms clusters. Then, only the complexes that are in the cluster where belongs the native state are conserved. Since only rigid transforms are applied, RMSD are computed thanks to the fast RMSD computation method previously proposed in the team [56].

The energy-based filtering compares the energy of the complexes to the native state energy. The states for which the difference of energy is higher than a given threshold are discarded.

We have evaluated the results obtained when using our filtering scheme, for a distance threshold ranging form 3 to 9 and for an energy threshold ranging from 70 to $240 kJ.mol^{-1}$. Some setting of the filtering are able to improve the correlation (see figure 13), but the gain around $0.3\%$ remains limited (e.g. the correlation rising from $0.770$ to $0.773$). We are currently working on a more sophisticated state selection process to filter more precisely the PIPER states and hence to further improve the correlation.

## 6.10. CREST: Chemical Reactivity Exploration with Stochastic Trees

**Participants:** Leonard Jaillet, Stephane Redon.

We have proposed the CREST method (Chemical Reactivity Exploration with Stochastic Trees), a new simulation tool to assess the chemical reaction paths of molecular systems. First, it builds stochastic trees based on motion planning principles to search for relevant pathways inside a system's state space. This generates low energy paths transforming a reactant to a given product. Then, a nudged elastic band optimization step locally improves the quality of the initial solutions. The consistency of our approach has been evaluated through tests

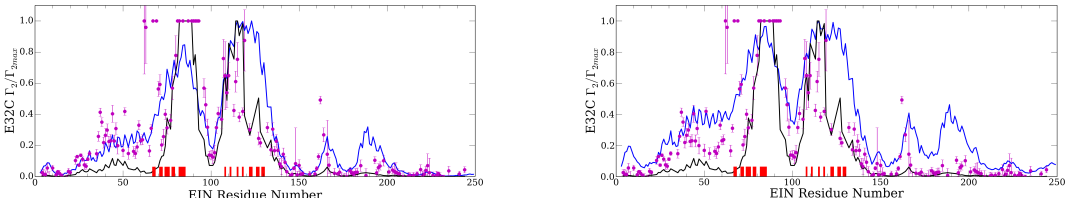


*Figure 13. The experimental PRE rates ($\Gamma 2$) are displayed as filled-in magenta circles. Theoretical intermolecular PREs, calculated only from the coordinates of the specific EIN/HPr complex, are shown as black lines. Calculated PRE values from PIPER output are shown as blue lines. The calculated PRE value obtained from the filtered complexes (left) gives a higher correlation with experimental ($c = 0.773$) than the correlation obtained from all the complexes generated with PIPER (right) ($c = 0.770$).*

in various scenarios. It shows that CREST allows to appropriately describe conformational changes as well as covalent bond breaking and formations present in chemical reactions (see figure 14).

This contribution appears in continuity of our previous work regarding the development of a geenric Motion planning architecture for nanosystems. Important features have been added to specifically treat the case of chemical reaction, such as structure alignment, exploration based on multiple trees, automatic resizing of the sampling volume, etc.

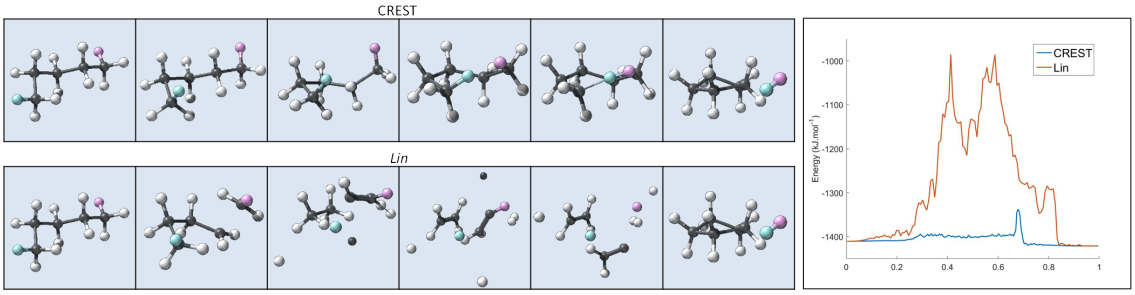


*Figure 14. Fictive chemical reaction transforming a pentane into a cyclopentane with a H2 molecule. Hydrogen atoms leading to the H2 molecule are colored. The path obtained with CREST (top) is able to capture the CH3 internal rotations that approaches the two H2 Hydrogens and thus, lead to a low energy barrier. By comparison, a method based on linear interpolation (Lin) gives intermediate broken structures after local path optimization (down). The plot on the right shows the respective energies along the paths. This represents Scenario 3 described in our Results section.*

## 6.11. IM-UFF: extending the Universal Force Field for interactive molecular modeling

**Participants:** Leonard Jaillet, Svetlana Artemova, Stephane Redon.

We have completed the development of IM-UFF (Interactive Modeling - UFF), an extension of UFF that 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. This approach introduces weighted atom types and weighted bonds, used to update topologies and atom parameterizations at every time step of a simulation. IM-UFF has been evaluated on a large set of benchmarks and is proposed as a self-contained implementation integrated in a new module for the SAMSON software platform for computational nanoscience.

This contribution has been submitted to the Journal of Molecular Modeling.

## 6.12. Incremental methods for long range interactions

**Participants:** Semeho Edorh, 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 low velocities, which allows to save computational time since particles do not move and forces need not be updated.

We continued our work on developing an extension of **ARPS** to electrostatic simulations.

We have decided to compute the electrostatic contribution by using Multigrid method. This choice have been made because of its $\mathcal{O}(N)$ behavior and its good scalability. In systems containing point charges, Multigrid can't be applied directly because of the discontinuous distribution created by these charges. To overcome this problem, one can replace this distribution by a smooth charge distribution. This charge distribution will be the source term of a Poisson equation which will be solved by Multigrid method. By doing so we retrieve an approximative electrostatic contribution which can be corrected by a near field correction. Concretely each charge will be smeared by a smooth density function.This function is chosen with a compact support. The accuracy of the method is related to the degree of smoothness and the size of the support $r_{cut}$ of the chosen function Fig(15). The bottleneck of this method is often the time spent building the smooth charge distribution. To overcome this issue, We've introduced an interpolation scheme in the near field correction. This leads to a significant reduction of the support required to achieve a specified accuracy. The time spent building the smooth charge distribution is also reduced. Conversely the near correction is slowed down. Nevertheless, the introduction of the interpolation scheme speeds up the method in most of cases Fig(16).

Finally we modified our algorithm to take advantage of ARPS dynamics. This leads to a speed up related to the amount of restrained particles. According to our benchmarks our method can challenge Particle Particle Particle Mesh(**PPPM**), the traditional fast method to compute electrostatics Fig(17). Our algorithm is implemented in LAMMPS.

## 6.13. Error Analysis of Modified Langevin Dynamics

**Participants:** Zofia Trstanova, Gabriel Stoltz, Stephane Redon.

We performed a mathematical analysis of modified Langevin dynamics. The aim of this work was first to prove the ergodicity of the modified Langevin dynamics (which fails to be hypoelliptic), and next to analyze how the asymptotic variance on ergodic averages depends on the parameters of the modified kinetic energy. Numerical results illustrated the approach, both for low-dimensional systems where we resorted to a Galerkin approximation of the generator, and for more realistic systems using Monte Carlo simulations.

## 6.14. Estimating the speed-up of Adaptively Restrained Langevin Dynamics

**Participants:** Zofia Trstanova, Stephane Redon.

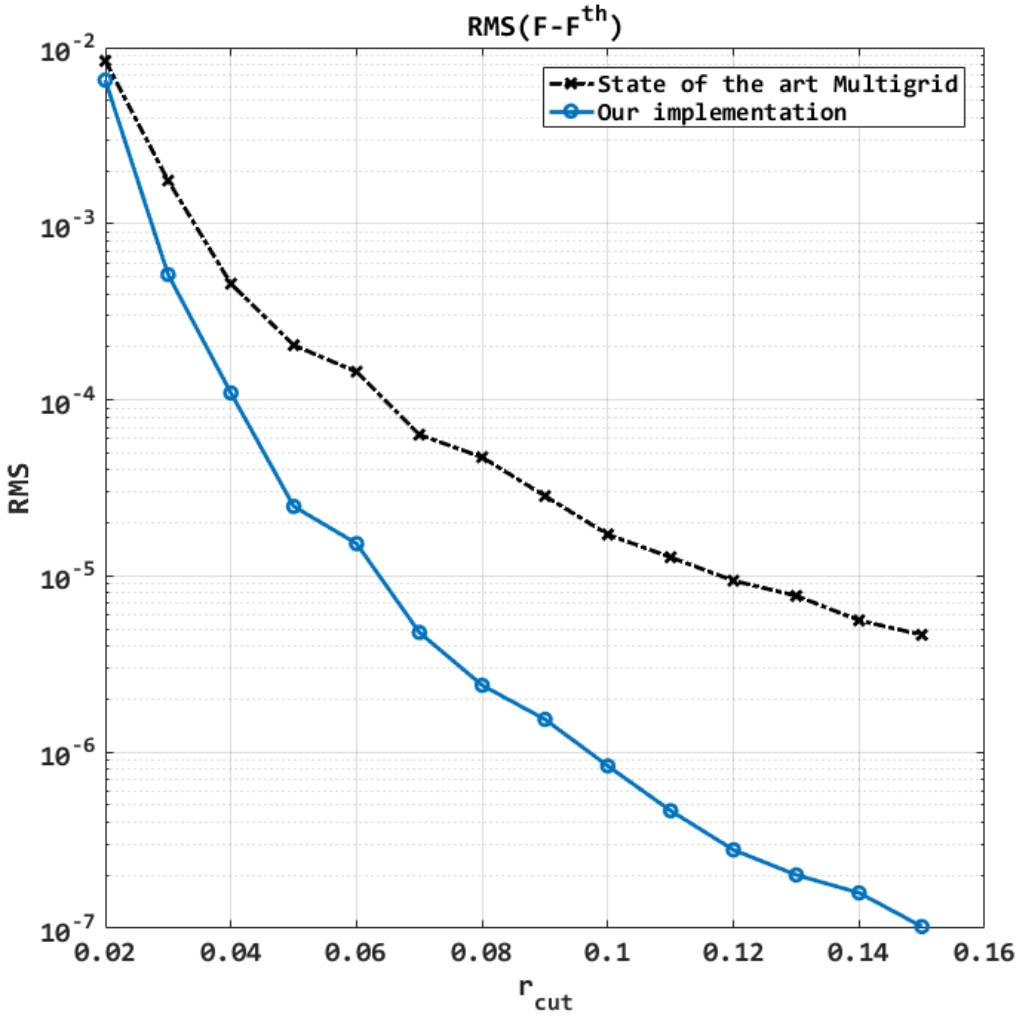


*Figure 15. Accuracy in forces for the state of the art multigrid and our implementation : 125000 charged particles randomly distributed in a cubic box. $r_{cut}$ represents the width of the chosen function.*

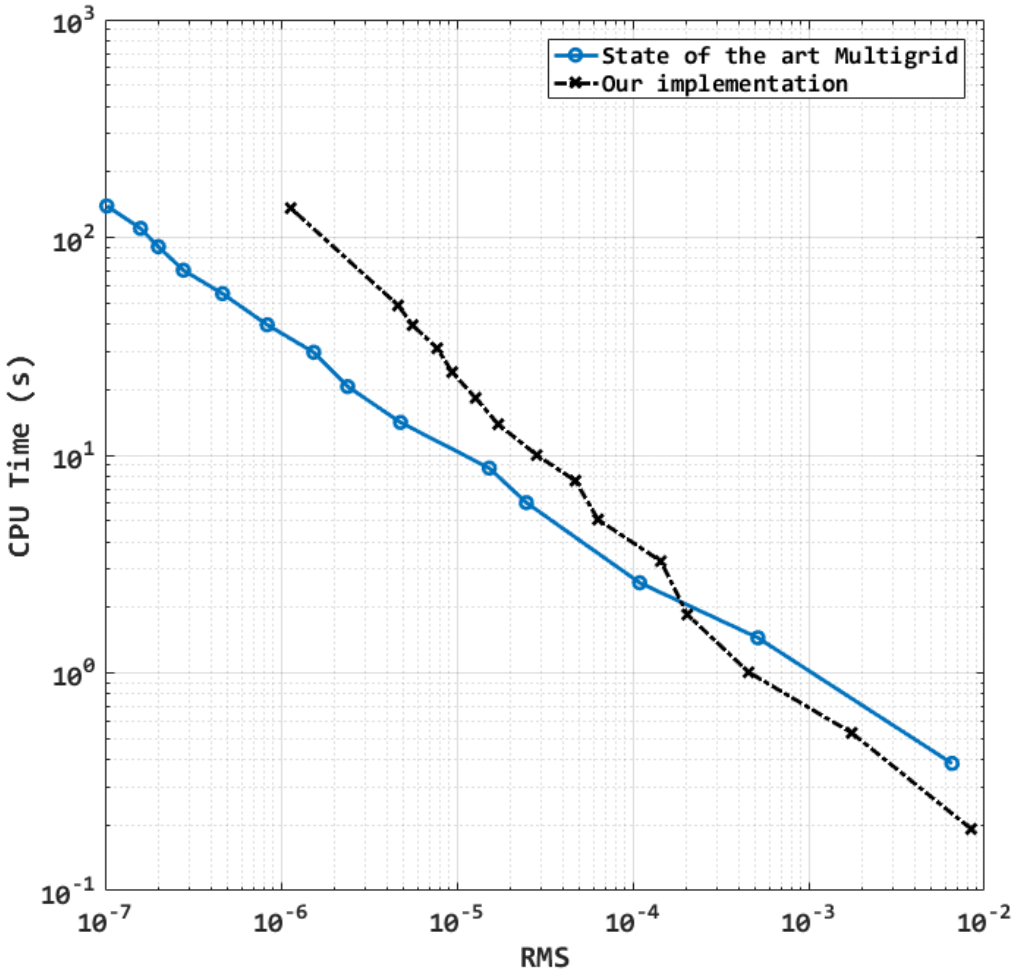


*Figure 16. Comparison in terms of CPU time between the state of the art multigrid and our implementation : 125000 charged particles randomly distributed in a cubic box. $r_{cut}$ represents the width of the chosen function.*

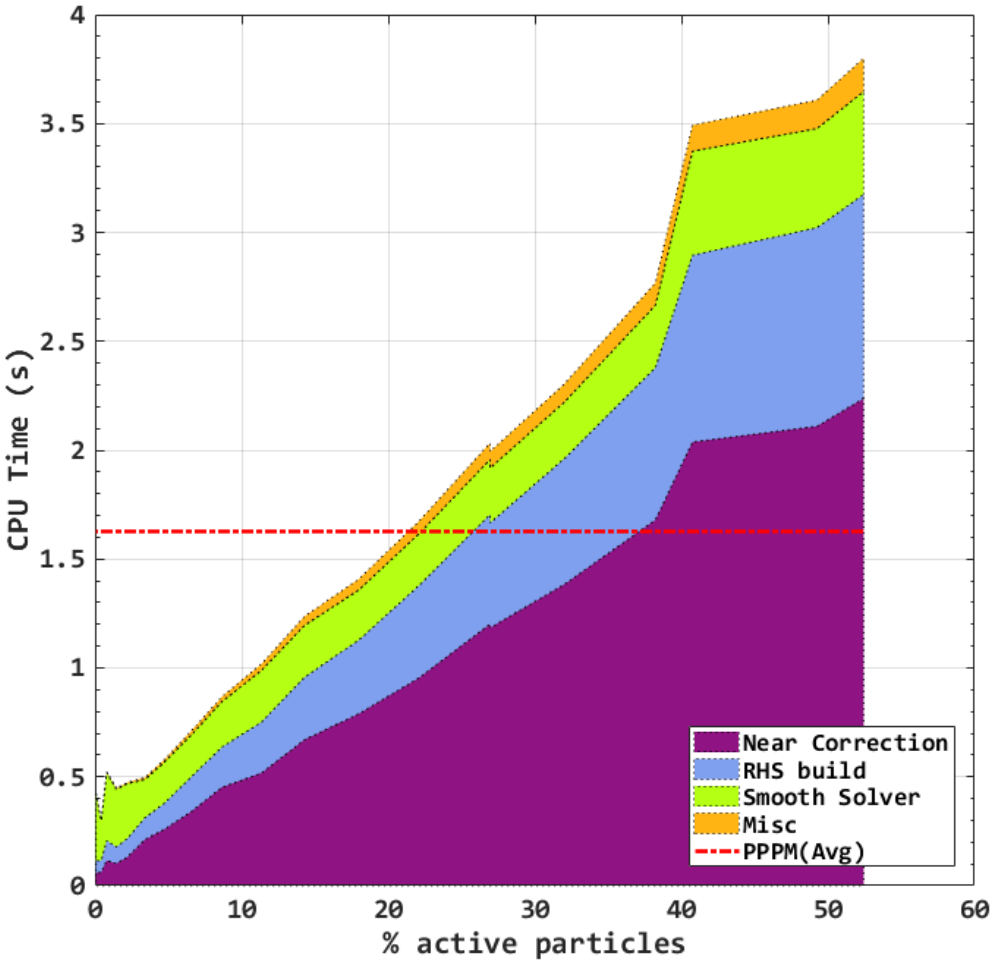


*Figure 17. Comparison in terms of CPU time between PPPM and our implementation for a fixed accuracy : 64000 charged particles randomly distributed in a cubic box. Some particles are in restrained dynamics. Colored areas show the associated contribution of each part of our multigrid algorithm. Red dash-dot line represents CPU Time of Particle Particle Particle Mesh needed for this system.*

We performed a computational analysis of Adaptively Restrained Langevin dynamics, in which the kinetic energy function vanishes for small velocities. Properly parameterized, this dynamics makes it possible to reduce the computational complexity of updating inter-particle forces, and to accelerate the computation of ergodic averages of molecular simulations. We analyzed the influence of the method parameters on the total achievable speed-up. In particular, we estimated both the algorithmic speed-up, resulting from incremental force updates, and the influence of the change of the dynamics on the asymptotic variance. This allowed us to propose a practical strategy for the parametrization of the method. We validated these theoretical results by representative numerical experiments on the system of a dimer surrounded by a solvent.

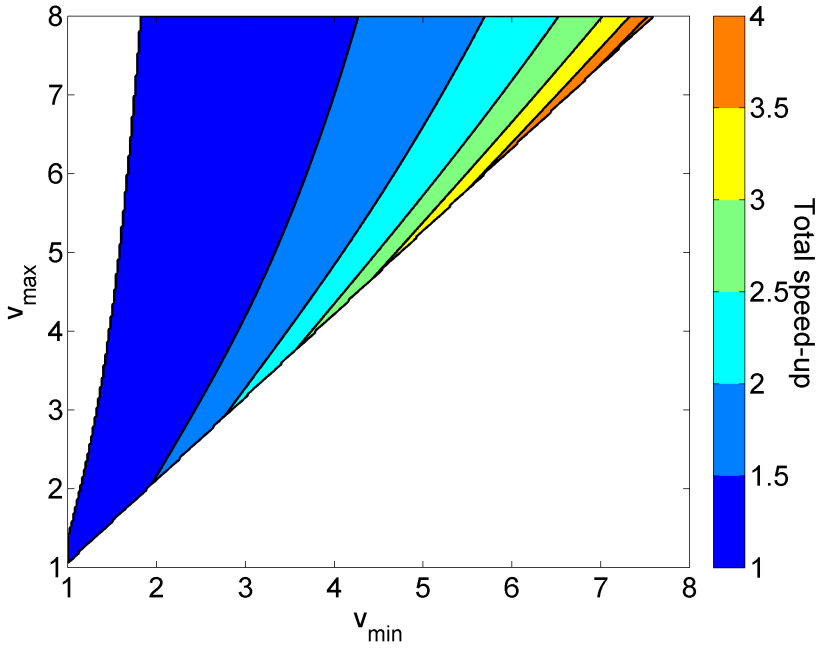


*Figure 18. Analytical estimation of the total speed-up of the 3D simulation of the dimer in solvent. Only the solvent particles are restrained by the AR-method. We estimated the expected total speed-up $S_{\mathrm{total}}$ for the observable dimer distance $A_D$ with respect to the restraining parameters $v_{\mathrm{min}}$ and $v_{\mathrm{max}}$ ($v_{\mathrm{max}} \leq 0.95 v_{\mathrm{max}}$). The variance was estimated from three points as a linear function of $v_{\mathrm{min}}$ and $v_{\mathrm{max}}$ and we used the analytical estimation of the algorithmic speed-up $S_{\mathrm{a}}$. Only $S_{\mathrm{total}} > 1$ is plotted.*

## 6.15. Stable and accurate schemes for Langevin dynamics with general kinetic energies

**Participants:** Zofia Trstanova, Gabriel Stoltz.

We studied integration schemes for Langevin dynamics with a kinetic energy different from the standard, quadratic one in order to accelerate the sampling of the Boltzmann–Gibbs distribution. We considered two cases: kinetic energies which are local perturbations of the standard kinetic energy around the origin, where they vanish (this corresponds to the so-called adaptively restrained Langevin dynamics); and more general non-globally Lipschitz energies. We developed numerical schemes which are stable and of weak order two, by considering splitting strategies where the discretizations of the fluctuation/dissipation are corrected by a

Metropolis procedure. We used the newly developed schemes for two applications: optimizing the shape of the kinetic energy for the adaptively restrained Langevin dynamics, and reducing the metastability of some toy models with non-globally Lipschitz kinetic energies.

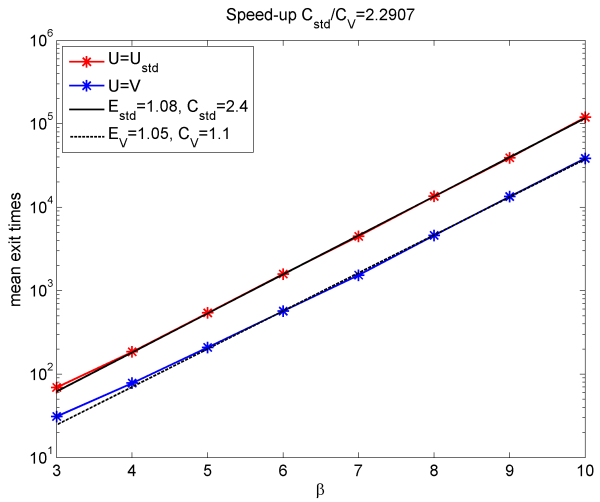


*Figure 19. Comparison of the mean exit times for 2D double-well potential with the standard and the modified kinetic energy function (2000 realizations) as a function of the inverse temperature* $\beta \in \{3, 4, 5, 6, 7, 8, 9, 10\}$. *Thanks to the modified kinetic energy, the transition between two metastable states occurs on average three times faster.*

## 6.16. Quadratic Programming Approach to Fit Protein Complexes into Electron Density Maps

**Participants:** Alexander Katrutsa, Sergei Grudinin.

We investigated the problem of simultaneous fitting protein complexes into electron density maps of their assemblies. These are represented by high-resolution cryo-EM density maps converted into overlapping matrices and partly show a structure of a complex. The general purpose is to define positions of all proteins inside it. This problem is known to be NP-hard, since it lays in the field of combinatorial optimization over a set of discrete states of the complex. We introduced quadratic programming approaches to the problem. To find an approximate solution, we converted a density map into an overlapping matrix, which is generally indefinite. Since the matrix is indefinite, the optimization problem for the corresponding quadratic form is non-convex. To treat non-convexity of the optimization problem, we use different convex relaxations to find which set of proteins minimizes the quadratic form best.

## 6.17. Inverse Protein Folding Problem via Quadratic Programming

**Participants:** Mikhail Karasikov, Sergei Grudinin.

We presented a method of reconstruction a primary structure of a protein that folds into a given geometrical shape. This method predicts the primary structure of a protein and restores its linear sequence of amino acids in the polypeptide chain using the tertiary structure of a molecule. Unknown amino acids are determined according to the principle of energy minimization. This study represents inverse folding problem as a quadratic optimization problem and uses different relaxation techniques to reduce it to the problem of convex optimizations. Computational experiment compares the quality of these approaches on real protein structures.

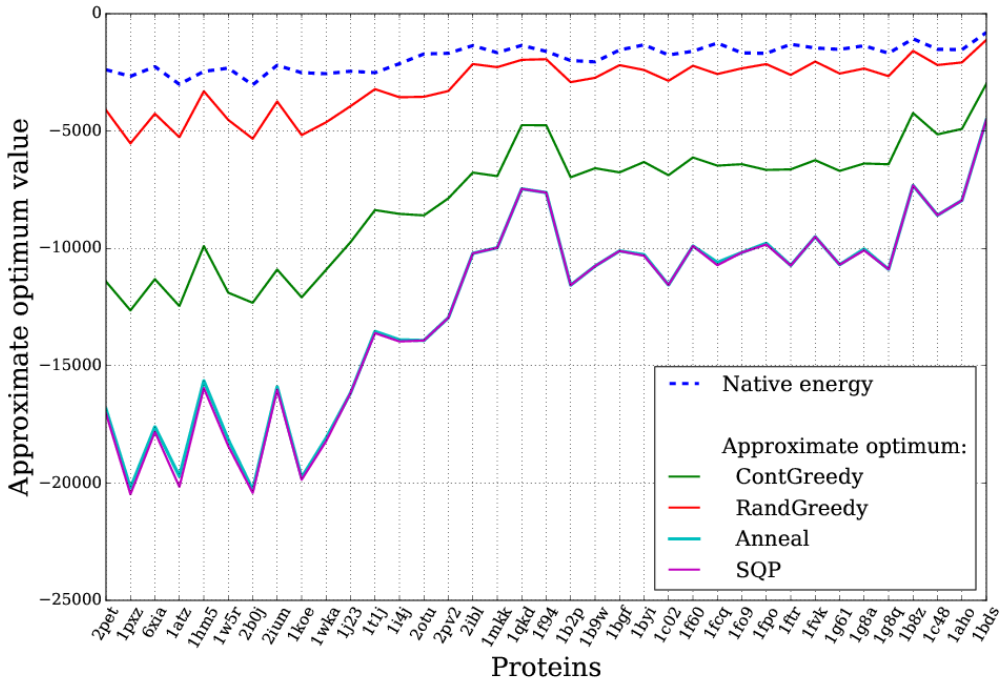


*Figure 20. Approximate energy optimum for different relaxations computed on the test set*

## 6.18. Coarse-Grained Protein Scoring Based on Geometrical Features

**Participants:** Mikhail Karasikov, Sergei Grudinin.

We learnt a scoring function to score protein structures with application to highly important problems in structural biology, namely, protein design, side-chain prediction, and selection of mutations increasing protein stability. For each native structure $P_0$ a set of ordered decoy structures $\mathcal{D}$ is given:

$$\mathcal{D} = \{P_1, \cdots, P_m\} \subset \mathcal{P},$$

$$(i_1, \cdots, i_m) : \; P_{i_m} \preceq \cdots \preceq P_{i_1} \prec P_0.$$

The problem is to train protein scoring function

$$S : \; \mathcal{P} \to \mathbb{R},$$

such that

$$S(P_0) < S(P_{i_1}) \leq \cdots \leq S(P_{i_m}).$$

We proposed a residue-based scoring function, which uses not the positions of protein's atoms separately, but configurations of the entire residues. The proposed method requires artificially generated decoy structures for the training process and provides high quality scoring functions, which are efficient to compute. Several types of scoring functions are considered according to restrictions imposed by the specific application. For the prediction problems where the whole domain should be searched for the best prediction, we use functions that allow the reduction of emerging optimization problem

$$\sum_{k=1}^{m} \sum_{l=1}^{m} E_{kl}(a_k, a_l) \to \min_{(a_1, \cdots, a_m) \in \mathcal{A}^m} \tag{1}$$

to quadratic binary constrained optimization

$$\begin{array}{ll} \underset{\overrightarrow{x} \in \{0,1\}^n}{\text{minimize}} & \overrightarrow{x}^{\mathsf{T}} \mathbf{Q} \overrightarrow{x} \\ \text{subject to} & \mathbf{A} \overrightarrow{x} = \overrightarrow{1}_m. \end{array} \tag{2}$$

## 6.19. Development of a Normal Modes Analysis element for SAMSON platform

**Participants:** Yassine Naimi, Alexandre Hoffmann, Sergei Grudinin, Stephane Redon.

We are currently developing an element for the SAMSON platform for the calculation of normal modes based on the Normal Modes Analysis method. This element will be based on the program developed by Alexandre Hoffmann and Sergei Grudinin on Linux and Mac operating systems. First, we have ported the initial program from Linux and Mac operating systems to Windows and linked the program to the libraries needed for the calculations. These libraries consist in: an optimized version of BLAS (Basic Linear Algebra Subprograms) library called OpenBLAS for basic vector and matrix operations; LAPACK (Linear Algebra PACKage) library for solving systems of simultaneous linear equations, least-squares solutions of linear systems of equations, eigenvalue problems, and singular value problems; ARPACK library for solving large scale eigenvalue problems and ARMADILLO library which is a linear algebra library for the C++ language. We will also compare the performances of our program using these libraries to the Intel MKL (Math Kernel Library) libraries. The ultimate goal is to develop the interface for the SAMSON platform using the SAMSON SDK and Qt software.

## 6.20. Pairwise distance potential for protein folding

**Participants:** Maria Kadukova, Guillaume Pages, Alisa Patotskaya, Sergei Grudinin.

We have developed a new knowledge-based pairwise distance-dependent potential using convex optimization. This method uses histogram of distances repartition between each different pair of atom types as feature to feed an SVM-like algorithm. We then obtained a potential for each pair of atom types that can be used to score protein conformations. This method have been extensively used during the CASP12 blind assesement.

## 6.21. Knowledge-based scoring function for protein-ligand interactions

**Participants:** Maria Kadukova, Sergei Grudinin.

We have developed a knowledge-based pairwise distance-dependent scoring function based on the similar physical principles, as the protein folding potentials. It was trained on a set of protein-ligand complexes taken from the PDBBindCN database and validated on the CASF 2013 benchmark [50]. The corresponding paper submitted to Journal of Chemical Information and Modeling is currently under revision. We used this scoring function while participating in the 2015-2016 D3R Challenge.

## 6.22. Updates for the atomic typization software

**Participants:** Maria Kadukova, Sergei Grudinin.

We have additionally validated Knodle – our atomic typization software – on an extensive set of more than 300,000 small molecules based on the LigandExpo database. Knodle workflow involves machine-learning based "models" for different atoms, this year we retrained several of them on the updated version of PDBBindCN database. These results were published in Journal of Chemical Information and Modeling [45]. We also added functions that add missing hydrogen atoms to the molecules. Knodle was used to classify ligand atoms into different types in our protein-ligand interactions scoring function.

## 6.23. FFT-accelerated methods for fitting molecular structures into Cryo-EM maps

**Participants:** Alexandre Hoffmann, Sergei Grudinin.

We have developed a set of new methods for fitting molecular structures into Cryo-EM maps. The problem can be formally written as follows, We are given two proteins $\mathcal{P}_1$ and $\mathcal{P}_2$, and we also have $d_1 : \mathbb{R}^3 \to \mathbb{R}$, the electron density of $\mathcal{P}_1$ and $(Y_k)_{k=0\cdots N_{atoms}-1}$, the starting positions of the atoms of $\mathcal{P}_2$. Assuming we can generate an artificial electron density $d_2 : \mathbb{R}^3 \to \mathbb{R}$ from $(Y_k)_{k=0\cdots N_{atoms}-1}$, our problem is to find a transformation of the atoms $T : \mathbb{R}^3 \to \mathbb{R}^3$ that minimize 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 of the 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 $T_{flexible}$. Two classes of methods have been developed to find $T_{rigid}$ :

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

The NANO-D team has already developed some algorithms based on the FFT to find $T_{rigid}$ and we have been developing an efficient extension of these to find $T_{flexible}$.

## 6.24. Protein sequence and structure aligner for SAMSON

**Participants:** Guillaume Pages, Sergei Grudinin.

Aligning sequences and structures of proteins is important to understand both the homologies and differences between them. We developed a SAMSON element for this purpose, that can perform both sequence and structure alignment. The sequence alignment is done thanks to the software MUSCLE [36]. The structural alignment is done by finding the transformation that minimize the RMSD between corresponding backbones atoms in both structures. We used the algorithm presented by Kearsley [47].

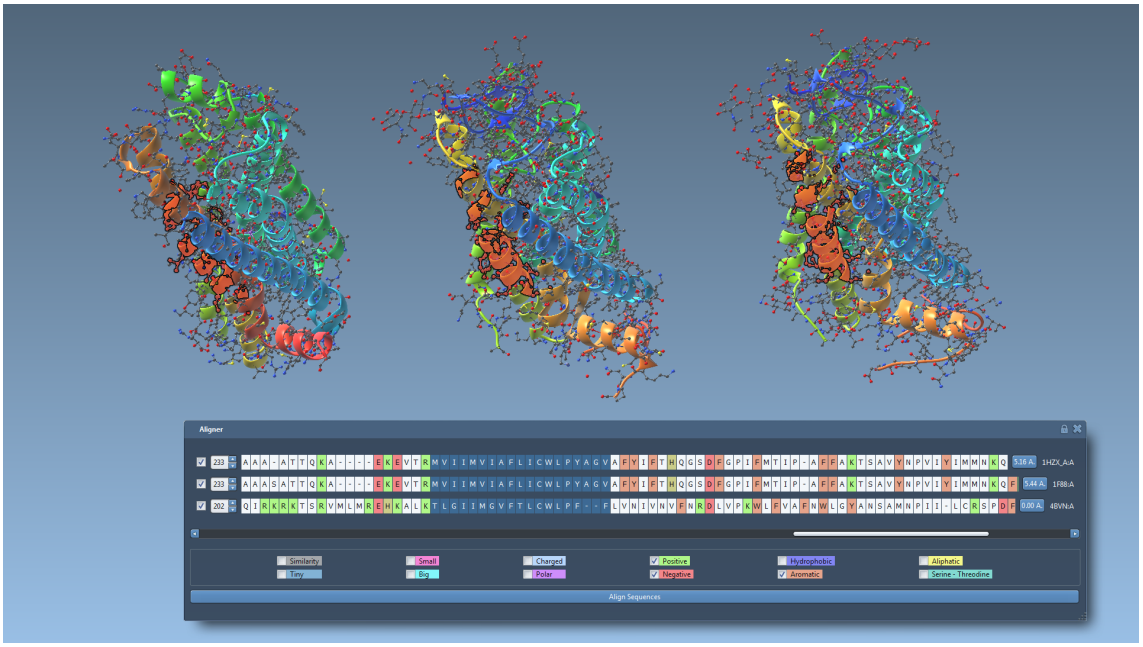


*Figure 21. The protein aligner element*

## 6.25. Implementation of an Interactive Ramachandran Plot Element for SAMSON

**Participants:** Guillaume Pages, Sergei Grudinin.

Each residue of a protein have two degrees of conformational freedom, described by the two dihedral angles of the backbone $\phi$ and $\psi$. Those two angles are crucial to visualize since they determine most of the protein backbone's overall conformation. A very useful way to represent them has been proposed by Ramachandran, Sasisekharan, and Ramakrishnan in 1963 [63].

We have developed a SAMSON element for displaying and editing the Ramachandran Plot of a protein. The favoured regions of the plot have been determined by analysing a database of high quality solved protein structure, provided on Richardson Lab's website (http://kinemage.biochem.duke.edu/databases/top8000.php).

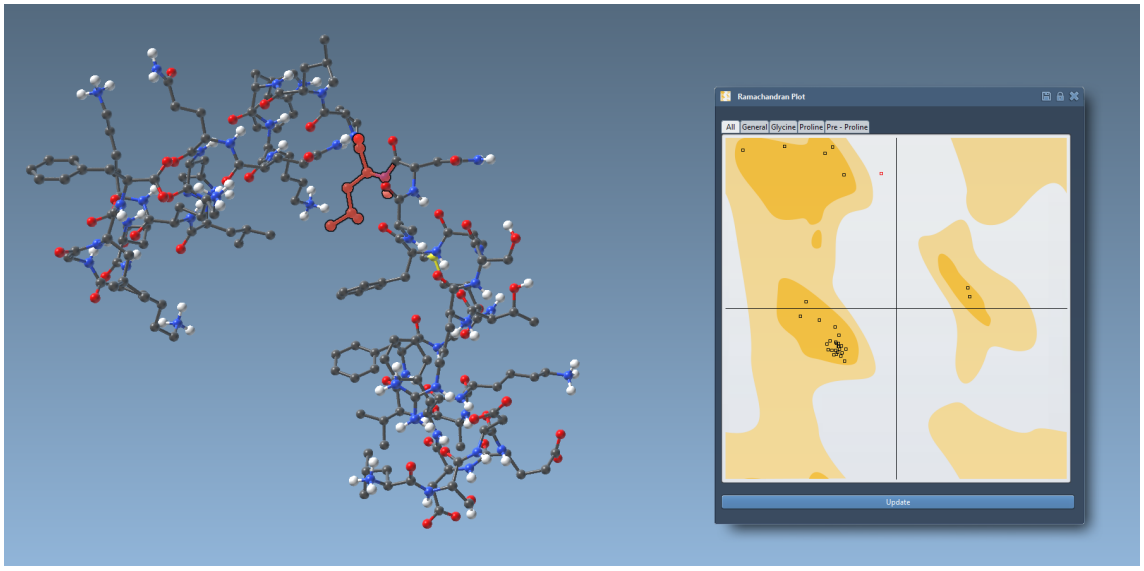


*Figure 22. The Ramachandran plot element*

# 7. Partnerships and Cooperations

## 7.1. Regional Initiatives

We have an ARC grant from the Rhone-Alpes region.

## 7.2. National Initiatives

### *7.2.1. 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).

## 7.3. European Initiatives

### *7.3.1. FP7 & H2020 Projects*

#### *7.3.1.1. ADAPT*

Title: Theory and Algorithms for Adaptive Particle Simulation

Programm: FP7

Duration: September 2012 - August 2017

Coordinator: Inria

Inria contact: Stephane Redon

'During the twentieth century, the development of macroscopic engineering has been largely stimulated by progress in digital prototyping: cars, planes, boats, etc. 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. Similar to what has happened with macroscopic engineering, powerful and generic computational tools will be needed to engineer complex nanosystems, through modeling and simulation. As a result, a major challenge is to develop efficient simulation methods and algorithms. NANO-D, the Inria research group I started in January 2008 in Grenoble, France, aims at developing efficient computational methods for modeling and simulating complex nanosystems, both natural and artificial. In particular, NANO-D develops SAMSON, a software application which gathers all algorithms designed by the group and its collaborators (SAMSON: Software for Adaptive Modeling and Simulation Of Nanosystems). In this project, I propose to develop a unified theory, and associated algorithms, for adaptive particle simulation. The proposed theory will avoid problems that plague current popular multi-scale or hybrid simulation approaches by simulating a single potential throughout the system, while allowing users to finely trade precision for computational speed. I believe the full development of the adaptive particle simulation theory will have an important impact on current modeling and simulation practices, and will enable practical design of complex nanosystems on desktop computers, which should significantly boost the emergence of generic nano-engineering.'

## 7.4. International Initiatives

### *7.4.1. Inria Associate Teams Not Involved in an Inria International Labs*

#### *7.4.1.1. PPI-3D*

Title: Structure Meets Genomics

International Partner (Institution - Laboratory - Researcher):

Boston University (United States) - ___DEPARTMENT???___ - 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.

### *7.4.2. Inria International Partners*

*7.4.2.1. BIOTOOLS*

Title: Novel Computational Tools for Structural Bioinformatics

International Partner (Institution - Laboratory - Researcher):

MIPT (Russia (Russian Federation)) - Vadim Strijov

Duration: 2016 - 2020

## 7.5. International Research Visitors

### *7.5.1. Visits of International Scientists*

*7.5.1.1. Internships*

Sergey Kravchenko

Supervisor: Sergey Grudinin

### *7.5.2. Visits to International Teams*

*7.5.2.1. Research Stays Abroad*

Leonard Jaillet, Alexandre Hoffmann and Sergei Grudinin visited the lab of Dima Kozakov.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Selection*

*8.1.1.1. Reviewer*

- Leonard Jaillet was a reviewer for the ICRA (International Conference on Robotics and Automation) and IROS (International Conference on Intelligent Robots and Systems) conferences, the WAFR (International Workshop on the Algorithmic Foundations of Robotics) workshop and the T-RO (Transactions on Robotics) journal.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. 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

### *8.2.2. Supervision*

- Leonard Jaillet is advising the PhD of Minh Khoa Nguyen
- Sergei Grudinin is advising the PhD of Alexandre Hoffmann
- Sergei Grudinin is advising the PhD of Guillaume Pages
- 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 (defended in november 2016)

# 9. Bibliography

## Publications of the year

### Articles in International Peer-Reviewed Journals

[1] S. ARTEMOVA, L. JAILLET, S. REDON. *Automatic molecular structure perception for the universal force field*, in "Journal of Computational Chemistry", March 2016 [*DOI :* 10.1002/JCC.24309], https://hal.inria.fr/hal-01282433

[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", March 2016, vol. 12, n$^{o}$ 3, pp. 1019–1028 [*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", January 2016, vol. 142, n$^{o}$ 694, pp. 515–528 [*DOI :* 10.1002/QJ.2676], https://hal.inria.fr/hal-01246349

[4] M. EL HOUASLI, B. MAIGRET, M.-D. DEVIGNES, A. W. GHOORAH, S. GRUDININ, D. RITCHIE. *Modeling and minimizing CAPRI round 30 symmetrical protein complexes from CASP-11 structural models*, in "Proteins: Structure, Function, and Genetics", October 2016 [*DOI :* 10.1002/PROT.25182], https://hal.inria.fr/hal-01388654

[5] S. GRUDININ, M. KADUKOVA, A. EISENBARTH, S. MARILLET, F. CAZALS. *Predicting binding poses and affinities for protein-ligand complexes in the 2015 D3R Grand Challenge using a physical model with a statistical parameter estimation*, in "Journal of Computer-Aided Molecular Design", September 2016, vol. 30, n$^{o}$ 9, pp. 791–804 [*DOI :* 10.1007/S10822-016-9976-2], https://hal.inria.fr/hal-01377738

[6] 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", June 2016, vol. 56, n° 6, pp. 1053–1062 [*DOI :* 10.1021/ACS.JCIM.5B00339], https://hal.inria.fr/hal-01258022

[7] M. KADUKOVA, S. GRUDININ. *Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates*, in "Journal of Chemical Information and Modeling", July 2016, vol. 56, n° 8, pp. 1410–1419 [*DOI :* 10.1021/ACS.JCIM.5B00512], https://hal.inria.fr/hal-01381010

[8] M. F. LENSINK, S. VELANKAR, A. KRYSHTAFOVYCH, S.-Y. HUANG, D. SCHNEIDMAN-DUHOVY, A. SALI, J. SEGURA, N. FERNANDEZ-FUENTES, S. VISWANATH, R. ELBER, S. GRUDININ, P. POPOV, E. NEVEU, H. LEE, M. BAEK, S. PARK, L. HEO, G. R. LEE, C. SEOK, S. QIN, H.-X. ZHOU, D. W. RITCHIE, B. MAIGRET, M.-D. DEVIGNES, A. GHOORAH, M. TORCHALA, R. A.G. CHALEIL, P. A. BATES, E. BEN-ZEEV, M. EISENSTEIN, S. NEGI S., T. VREVEN, B. G. PIERCE, T. M. BORRMAN, J. YU, F. OCHSENBEIN, Z. WENG, R. GUEROIS, A. VANGONE, J. P. RODRIGUES, G. VAN ZUNDERT, M. NELLEN, L. XUE, E. KARACA, A. S. J. MELQUIOND, K. VISSCHER, P. L. KASTRITIS, A. M. J. J. BONVIN, X. XU, L. QIU, C. YAN, J. LI, Z. MA, J. CHENG, X. ZOU, Y. SHENG, L. X. PETERSON, H.-R. KIM, A. ROY, X. HAN, J. ESQUIVEL-RODRÍGUEZ, D. KIHARA, X. YU, N. J. BRUCE, J. C. FULLER, R. C. WADE, I. ANISHCHENKO, P. J. KUNDROTAS, I. A. VAKSER, K. IMAI, K. YAMADA, T. ODA, T. NAKAMURA, K. TOMII, C. PALLARA, M. ROMERO-DURANA, B. JIMÉNEZ-GARCÍA, I. H. MOAL, J. FERNÁNDEZ-RECIO, J. Y. JOUNG, J. Y. KIM, K. JOO, J. LEE, D. KOZAKOV, S. VAJDA, S. MOTTARELLA, D. R. HALL, D. BEGLOV, A. MAMONOV, B. XIA, T. BOHNUUD, C. A. DEL CARPIO, E. ICHIISHI, N. MARZE, D. KURODA, S. S. R. BURMAN, J. J. GRAY, E. CHERMAK, L. CAVALLO, R. OLIVA, A. TOVCHIGRECHKO, S. J. WODAK. *Prediction of homo- and hetero-protein complexes by protein docking and template-based modeling: a CASP-CAPRI experiment*, in "Proteins - Structure, Function and Bioinformatics", September 2016, vol. 84, n° S1, pp. 323–348 [*DOI :* 10.1002/PROT.25007], https://hal.inria.fr/hal-01309105

[9] P.-L. MANTEAUX, C. WOJTAN, R. NARAIN, S. REDON, F. FAURE, M.-P. CANI. *Adaptive Physically Based Models in Computer Graphics*, in "Computer Graphics Forum", 2016 [*DOI :* 10.1111/CGF.12941], https://hal.inria.fr/hal-01367170

[10] E. NEVEU, D. RITCHIE, P. POPOV, S. GRUDININ. *PEPSI-Dock: a detailed data-driven protein–protein interaction potential accelerated by polar Fourier correlation*, in "Bioinformatics", August 2016, vol. 32, n° 7, pp. i693-i701 [*DOI :* 10.1093/BIOINFORMATICS/BTW443], https://hal.archives-ouvertes.fr/hal-01358645

[11] S. REDON, G. STOLTZ, Z. TRSTANOVA. *Error Analysis of Modified Langevin Dynamics*, in "Journal of Statistical Physics", August 2016, vol. 164, n° 4, pp. 735–771 [*DOI :* 10.1007/S10955-016-1544-6], https://hal.archives-ouvertes.fr/hal-01263700

[12] 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

### International Conferences with Proceedings

[13] R. POGODIN, A. KATRUTSA, S. GRUDININ. *Quadratic Programming Approach to Fit Protein Complexes into Electron Density Maps*, in "Information Technology and Systems 2016", Repino, St. Petersburg, Russia, September 2016, https://hal.inria.fr/hal-01419380

[14] A. RIAZANOV, M. KARASIKOV, S. GRUDININ. *Inverse Protein Folding Problem via Quadratic Programming*, in "Information Technology and Systems 2016", Repino, St. Petersburg, Russia, September 2016, pp. 561-568, https://hal.inria.fr/hal-01419374

### Other Publications

[15] G. STOLTZ, Z. TRSTANOVA. *Stable and accurate schemes for Langevin dynamics with general kinetic energies*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01364821

## References in notes

[16] 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$^{o}$ 2, pp. 99 - 104 [*DOI :* 10.1016/J.PORGCOAT.2007.07.008], http://www.sciencedirect.com/science/article/pii/S0300944007001464

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

[18] 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

[19] 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$^{o}$ 16, pp. 3243 - 3251 [*DOI :* 10.1016/J.SUSC.2006.06.015], http://www.sciencedirect.com/science/article/pii/S003960280600700X

[20] 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

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

[22] 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

[23] 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

[24] 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

[25] 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° 6, pp. 2581 - 2598 [*DOI :* 10.1016/J.JCP.2011.12.006], http://www.sciencedirect.com/science/article/pii/S0021999111007042

[26] 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

[27] 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° 5, pp. 345 - 356 [*DOI :* 10.1016/S1089-3156(01)00026-5], http://www.sciencedirect.com/science/article/pii/S1089315601000265

[28] 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

[29] 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° 3, pp. 241-258, http://www.tandfonline.com/doi/abs/10.1080/02652030701744538

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

[31] G. M. CLORE. *Visualizing lowly-populated regions of the free energy landscape of macromolecular complexes by paramagnetic relaxation enhancement*, in "Molecular Biosystems", 2008, vol. 4, n° 11, pp. 1058–1069

[32] 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

[33] 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° 4, pp. 290-307

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

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

[36] R. C. EDGAR. *MUSCLE: multiple sequence alignment with high accuracy and high throughput*, in "Nucleic acids research", 2004, vol. 32, n° 5, pp. 1792–1797

[37] 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

[38] 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

[39] 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

[40] 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

[41] 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

[42] 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

[43] 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

[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] M. KADUKOVA, S. GRUDININ. *Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates*, in "J. Chem. Inf. Model.", 2016, vol. 56, n$^{o}$ 8, pp. 1410–1419

[46] 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

[47] S. K. KEARSLEY. *On the orthogonal transformation used for structural comparisons*, in "Acta Crystallographica Section A: Foundations of Crystallography", 1989, vol. 45, n$^{o}$ 2, pp. 208–210

[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. LI, D. R. HALL, D. BEGLOV, J. ZHENG, P. VAKILI, O. SCHUELER-FURMAN, I. C. PASCHALIDIS, G. M. CLORE, S. VAJDA. *Encounter complexes and dimensionality reduction in protein–protein association*, in "Elife", 2014, vol. 3, e01370 p.

[50] Y. LI, L. HAN, Z. LIU, R. WANG. *Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods an General Results*, in "J. Chem. Inf. Model.", Jun 2014, vol. 54, n$^{o}$ 6, pp. 1717-36, http://dx.doi.org/10.1021/ci500081m

[51] 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

[52] 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

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

[54] 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

[55] 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

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

[57] 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

[58] 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$^{o}$ 29, 292201, http://stacks.iop.org/0953-8984/19/i=29/a=292201

[59] 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$^{o}$ 13, 134902, http://aip.scitation.org/doi/abs/10.1063/1.2714540?journalCode=jcp

[60] 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$^{o}$ 1, pp. 545-571, http://www.annualreviews.org/doi/abs/10.1146/annurev.physchem.59.032607.093707

[61] 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$^{o}$ 24, pp. 10464-10468, http://pubs.acs.org/doi/abs/10.1021/jp960295w

[62] 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

[63] G. N. RAMACHANDRAN, C. RAMAKRISHNAN, V. SASISEKHARAN. *Stereochemistry of polypeptide chain configurations*, in "Journal of molecular biology", 1963, vol. 7, n$^{o}$ 1, pp. 95–99

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

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

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

[67] 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

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

[69] 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

[70] 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

[71] 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

[72] 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

[73] 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

[74] 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

[75] 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

[76] 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

[77] 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

[78] 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