Activity Report 2011

Team NANO-D

Algorithms for Modeling and Simulation of Nanosystems

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Computational models and simulation

Table of contents

Team NANO-D

Keywords: Modeling, Simulation, Nanosystems, Adaptive Algorithm

1. Members

Research Scientists
Sergei Grudinin [CNRS Junior Researcher]
Stéphane Redon [Team leader, INRIA Junior Researcher]

Technical Staff
Jocelyn Gaté [Software Engineer]

PhD Students
Svetlana Artemova [INRIA PhD Student]
Maël Bosson [PhD Student - Université Joseph Fourier]
Petr Popov [INRIA PhD Student]

Administrative Assistant
Françoise de Coninck [Assistant]

Other
Jelmer Wolterink [Intern (Utrecht University)]

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

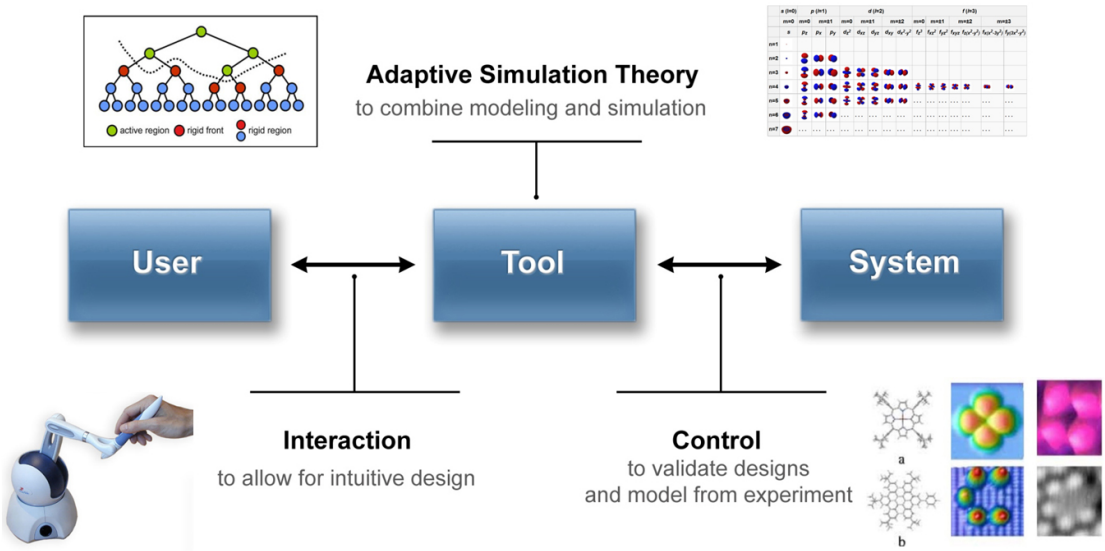

Figure 1. NANO-D's research axes.

The goal of NANO-D is to help current and future designers of nanosystems by developing some of the mathematical and computational foundations of a software application which will run on a desktop computer, and will allow for efficient analysis, design, modeling and simulation of complex nanosystems, whether they are artificial or natural, or a combination of both. For clarity, the research program of the NANO-D group is best introduced by referring to Figure 1. There:

- **User** is any person who wants to study, analyze, design, model, simulate, and control a nanosystem.
- **Tool** is the software application being built on the research performed within NANO-D. This tool will have functionalities that will be similar to those used to design macrosystems (e.g. CATIA, SolidWorks, etc.), but also some others which will be specific to nanoscience.
- **System** is the nanosystem being designed, simulated, controlled, etc., and potentially its environment. The environment can be any system in interaction with the given nanosystem: two electrodes between which a nanotube has been placed, the atomic force microscope which interacts with a nano-wheel, a protein interacting mechanically with a nano-drug or an engineered protein, etc.

This simple diagram makes it clear that a complete, coherent effort towards practical design of nanosystems should be organized in three research axes:

- **Adaptive Simulation Theory**. This first research axis deals with the core algorithms for modeling and simulation, which constitute the heart of the design functionalities of the tool. The main

paradigm in this axis is to rely on divide-and-conquer, hierarchical representations to design to adaptive algorithms for modeling and simulating nanosystems.

- **Interaction**. The second group of tasks deals with the interaction of the user with the tool. This is essential, because nanosystems may have complex topologies, kinematics and dynamics, which may make it difficult to edit and model.
- **Control**. The last group of tasks deals with the bidirectional relationships between the simulated nanosystems and the actual ones. Indeed, nanosystems may be characterized through a variety of experimental techniques (e.g. Atomic Force Microscopy, etc.). In order to help verify designs, the tool must thus be able to simulate characterization techniques. Conversely, the tool must be able to take advantage of existing knowledge (both experimental and computational) to help the user design nanosystems.

We believe that each of these three groups of tasks is essential. Fast modeling and simulation algorithms allow for relevant, efficient design ("Adaptive Simulation Theory"). However, software tools with powerful functionalities but which are difficult to use ("Adaptive Simulation Theory" without "Interaction") end up not being used at all. Finally, a tool for modeling and simulating nanosystems which would be "disconnected" from the reality of experimentation ("Adaptive Simulation Theory" without "Control") could only be used for theoretical designs with little assurance of their practicality.

2.3. Highlights

We wish to highlight three results:

- **Interactive Quantum Chemistry**: we have developed what appears to be the first method for *interactive quantum chemistry*, at the ASED-MO level of theory. This should be of significant help to *e.g.* analyze and design nanosystems, as well as in chemistry education.
- **Adaptively Restrained Particle Simulations**: we have developed a rigorous method for adaptive simulation of particle systems, with potential applications in many areas of nanoscience, and beyond (particle simulations are widely used in *e.g.* computational fluid dynamics, astrophysics, computer graphics, etc.). The method has numerous advantages, and allows for the first time to rigorously and smoothly trade between precision and cost when performing a particle simulation.
- **ANR PEPSI**: NANO-D obtained a new ANR grant, called PEPSI. The PEPSI project is coordinated by Sergei Grudinin from NANO-D, and is in collaboration with Dave Ritchie at Loria and Valentin Gordeliy at IBS (Grenoble). The goal of the PEPSI project is to develop new representations of 3D protein structures, in order to calculate protein interactions extremely efficiently.

More details are available below.

3. Scientific Foundations

3.1. Overview

The adaptive simulation algorithms we develop typically consist in two main components. The first one determines *which degrees of freedom are simulated* at a give time step, based on the current system's state, as well as user-defined precision or cost thresholds. The second component *incrementally updates the system's state* based on the set of active degrees of freedom. In particular, incremental algorithms update the system's potential energy and forces. This allows the user to smoothly trade between precision and cost.

We detail this approach in two important types of simulations: Cartesian quasi-statics and torsion-angle dynamics. A novel, very general approach for adaptive dynamics simulations of particles — that has a number of important benefits over previous approaches — is mentioned in more detail in Section 6.1.

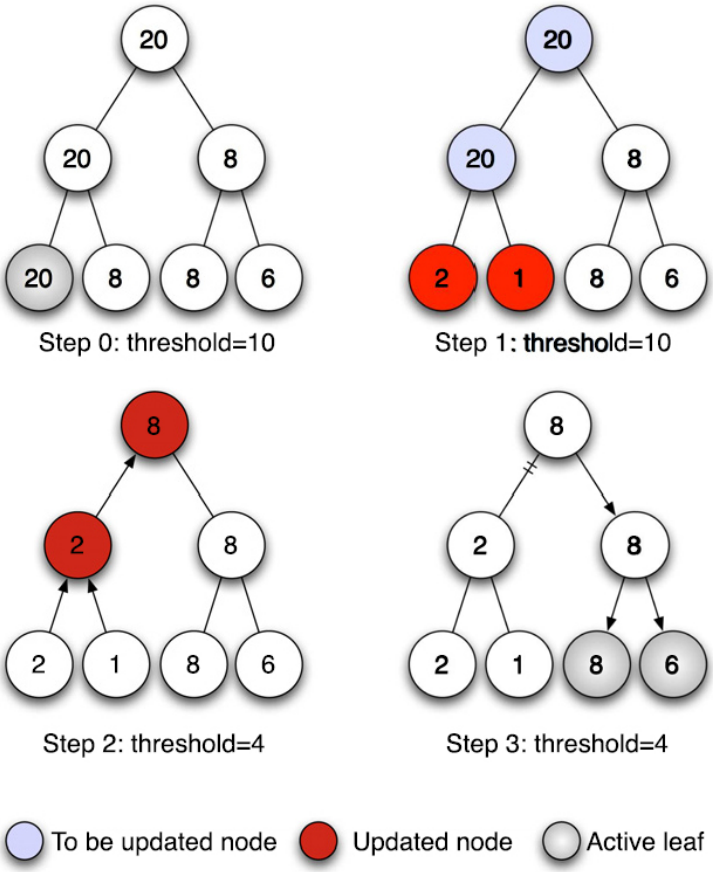

Figure 2. Adaptive Cartesian mechanics.

3.2. Adaptive Cartesian mechanics

In order to focus computations on a specific set of atoms, when performing quasi-static simulations (minimizations), we have developed an adaptive Cartesian mechanics algorithm, which decides which atoms should move at each time step.

In the simplest approach, we simply examine the force applied on each atom. When the norm of the force is above a user-defined threshold, the atom is active. Else the atom position is frozen. A slightly more elaborate version consists in defining the threshold automatically based on the system state (it might be *e.g.* the average applied force, a percentage of the maximum norm, etc.).

In order to avoid the linear cost of determining the set of active atoms at each time step, a binary tree is used to represent the system. Each leaf node represents an individual atom, while each internal node represents a set of atoms. Each leaf node stores the norm of the force applied to the corresponding atom. Each non-leaf node stores the maximum of the two force norms of its children, as illustrated in Figure 2. We use two tree passes in order to update tree nodes' values and to determine the new active atoms. In the first, bottom-up pass, force norms are updated in a sub-tree of the binary tree (only some atoms have moved since the previous time step, so only some forces have been updated), starting from the leaves with modified norms, in $O(k^{old}(log(\frac{n}{k^{old}})+1))$ times where k^{old} is the number of active atoms and n the total number of atoms. In the second, top-down pass, the new active atoms (i.e. the atoms with the force norms which are now the largest), are determined in $O(k^{new}(log(\frac{n}{k^{new}})+1))$ times where k^{new} is the new number of active atoms. This process is illustrated in Figure 2 as well.

Precisely, Figure 2 illustrates the procedure to determine the active zone, when the threshold is automatically set to half the largest atomic force norm. In this example, the four leaves correspond to atoms 1 to 4. The value indicated in each leaf node is the norm of the force applied to its corresponding atom. For internal nodes, this value is the maximum of the norms of the forces applied to atoms in the corresponding group. In step 0, the threshold is automatically set to 10. As a result, only atom 1 moves. In step 1, the potential is incrementally updated, and the norms of the forces applied to atoms 1 and 2 are updated. In step 2, the values associated to the tree nodes are incrementally updated through a bottom-up pass that starts from the modified leaf nodes values. Because of this bottom-up update, the adaptive threshold becomes equal to 4. In step 3, the new active atoms are determined through a top-down pass, by visiting only the nodes that have a value larger than the adaptive threshold.

3.3. Adaptive torsion-angle mechanics

In many situations, it is preferable to represent molecular systems as articulated bodies, and perform so-called *torsion-angle* mechanics. This may be to allow for larger time step sizes in a simulation, or because the user wants to focus to *e.g.* protein backbone deformations.

We have also developed an adaptive mechanics algorithm in the case of torsion-angle representations. In this case, a molecular system is recursively defined as the assembly of *two* molecular systems connected by a *joint* (when connecting two subassemblies which belong to the same molecule) or, more generally, by a *rigid body transform* (to assemble several molecules).

As in the Cartesian mechanics case, the complete molecular system is thus also represented by a binary tree, in which leaves are rigid bodies (a rigid body can be a single atom), internal nodes represent both sub-assemblies and connections between sub-assemblies, and the root represents the complete molecular system (see Figure 3 on the right, which shows an assembly tree associated to a short polyalanin). This hierarchical representation handles any branched molecule or groups of molecules, since the connections between two sub-molecular systems can be a rigid body transformation. In this representation, the positions of atoms are thus represented as superimposed rigid transformations: the position of any atom is obtained from the position of the whole set, to which is "added" the transformation from the complete set to the sub-set the atom belongs to, and so on until we reach the leaf node representing the atom. Similarly, the atomic motions are superimposed rigid motions.

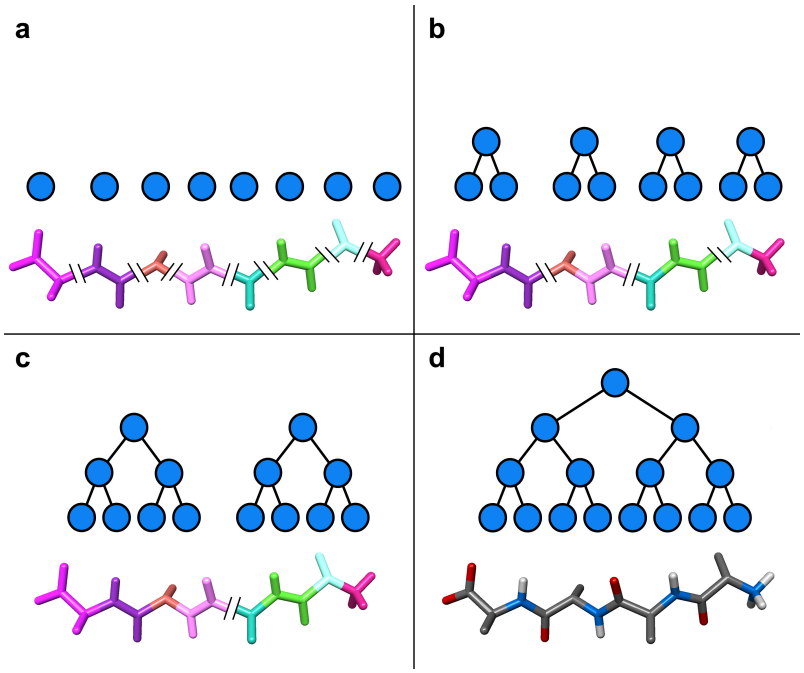

Figure 3. The assembly tree associated to a short polyalanin.

Our adaptive framework relies on two essential components. First, we associate a hierarchical set of reference frames to the assembly tree. Precisely, each node is associated to a local reference frame, in which all dynamical coefficients are expressed. This allows us to avoid updating these coefficients when a sub-assembly moves rigidly. Second, we have demonstrated that it is possible to determine a priori, at each time step, the set of joints which have the largest accelerations. Precisely, when going down the tree to compute joint accelerations, we are able to compute the weighted sum of the (squared) norms of joint accelerations in a sub-assembly C before computing joint accelerations themselves:

$$\mathcal{A}(C) = (\mathbf{f}^C)^T \Psi^C \mathbf{f}^C + (\mathbf{f}^C)^T \mathbf{p}^C + \eta^C, \tag{1}$$

where the right part is a quadratic form of the spatial forces applied on the "handles" of node C. This allows us to cull away those sub-assemblies with (relatively) lower internal accelerations, and focus on the most mobile joints. Thus, at each time step, we can thus predict the set of joints with highest accelerations without computing all accelerations, and we simulate only a sub-tree of the assembly tree (the green nodes in the assembly tree, as in the figure above), based on an user-defined error threshold or computation time constraints. This sub-tree is called the active region, and may change at each time step.

We have exploited these two characteristics - hierarchical coordinate systems and adaptive motion refinement - to develop data structures and algorithms which enable adaptive molecular mechanics. The key observation in our approach is the following: all coefficients which only depend on relative atomic positions do not have to be updated when these relative positions do not change. We can thus store in each node of the assembly tree partial system states which hold information relative only to the node itself.

Precisely, each time step involves the following operations:

1. Adaptive acceleration update
    1. Determination of the acceleration update region: we determine the acceleration update region, i.e. the subset of nodes of the full articulated body which matter the most according to the acceleration metric, as indicated above. The union of the previous active region and the acceleration update region is the transient active region, i.e. the region temporarily considered as the active region.
    2. Joint accelerations projection: the acceleration is projected on the reduced motion space defined by the transient active region (to ensure that joint accelerations are consistent with both motion constraints and applied forces).
2. Adaptive velocity update
    1. Determination of the new active region: we update the joint velocities and the velocity metric values of the nodes in the transient active region. We then determine the set of nodes which are considered to be important according to the velocity metric (which is similar to the acceleration metric). This set becomes the new active region.
    2. Joint velocities projection: if one or more nodes become inactive due to the update of the active region, we determine a set of impulses that we must apply to the transient hybrid body to perform the rigidification of these nodes. This amounts to projecting joint velocities to the reduced motion space defined by the new active region.
3. Adaptive position update
    1. Position update: we update joint positions based on non-zero joint velocities in the active region.
    2. State update: once joint positions have been updated, we update the rest of the system's state: inverse inertias, acceleration metric coefficients, partial neighbor lists, partial force tables, etc.

Again, each of these steps involves a limited sub-tree of the assembly tree, which enables a fine control of the compromise between computation time and precision.

We have showed that our adaptive approach allows for a number of applications, some of which that were not possible for classical methods when using low-end desktop workstations. Indeed, by selecting a sufficiently small number of simultaneously active degrees of freedom, it becomes possible to perform interactive structural modifications of complex molecular systems.

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.

4.4. Nano-engineering

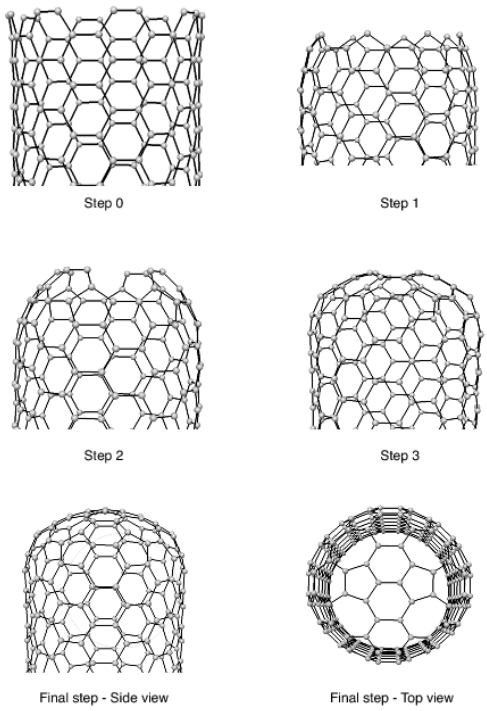

Figure 4. Snapshots of a nanotube capping process with the adaptive interactive modeler. Thanks to the adaptive methodology, this operation can be done in a few minutes.

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

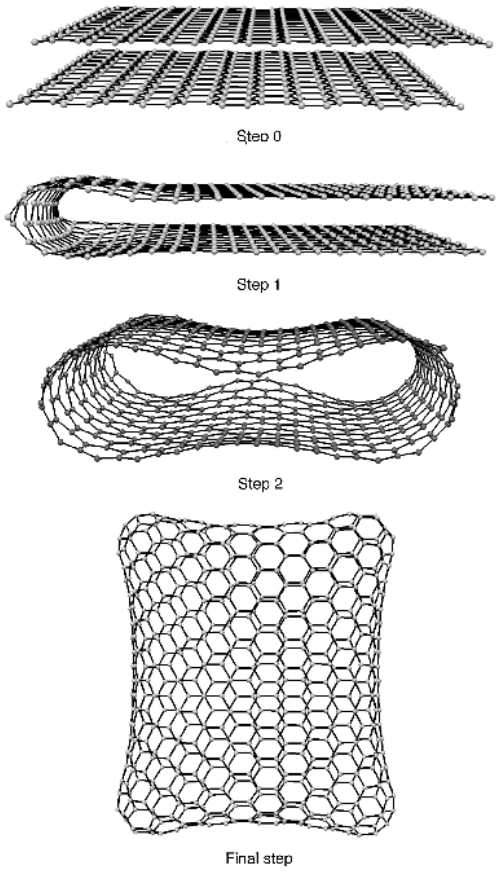

Figure 5. Different steps to prototype a "nano-pillow" with the adaptive interactive modeler.

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

We have recently shown that our incremental and adaptive algorithms allow us to easily edit and model complex shapes, such as a nanotube (Fig. 4) and the "nano-pillow" below (Fig. 5).

5. Software

5.1. SAMSON

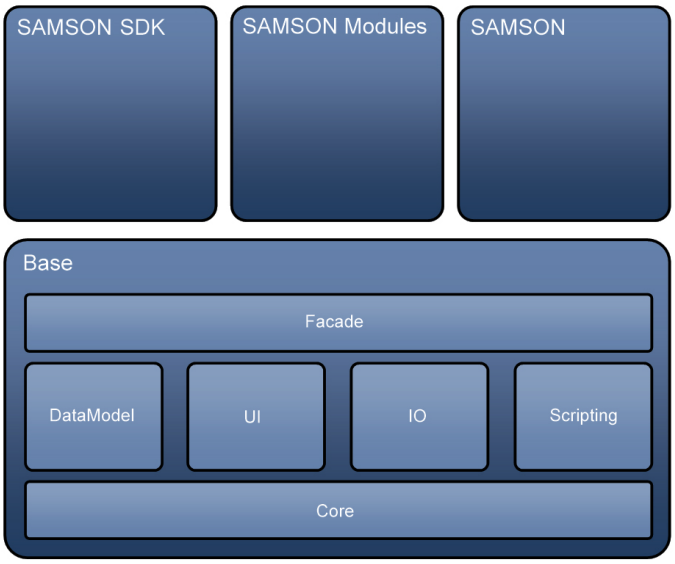

Figure 6. SAMSON's architecture.

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

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

The current architecture of SAMSON is visible in Figure 6. The code is organized into four main parts: a) the Base (in which "Core" contains, in particular, the heart of the adaptive algorithms: signaling mechanisms specifically designed for SAMSON), b) the Software Development Kit (SDK: a subset of the base that will be provided to module developers), c) Modules, and d) the SAMSON application itself.

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

6. New Results

6.1. Algorithms for molecular modeling and simulation

6.1.1. Interactive quantum chemistry

Participants: Maël Bosson, Caroline Richard, Antoine Plet, Sergei Grudinin, Stéphane Redon.

We have proposed what appears to be the first algorithm for *interactive quantum chemistry simulation* at the Atom Superposition and Electron Delocalization Molecular Orbital (ASED-MO) level of theory. When drawing and editing molecular systems, interactive quantum chemistry provide immediate, intuitive feedback on chemical structures. Our method is based on the divide-and-conquer (D&C) approach, which we show is accurate and efficient for this non-self-consistent semi-empirical theory. The errors induced by the D&C approach have been studied empirically and *via* a theoretical study of two toy models. With this method, we have demonstrated *interactive* quantum chemistry simulations for systems up to a few hundred atoms on a current multicore desktop computer. As the number of cores on personal computers increases, and larger and larger systems can be dealt with, we believe such interactive simulations – even at lower levels of theory – should thus prove most useful to effectively understand, design and prototype molecules, devices and materials. This result has been published in Journal of Computational chemistry [4]. Figure 7 illustrates an interactive modeling session with a benzene molecule.

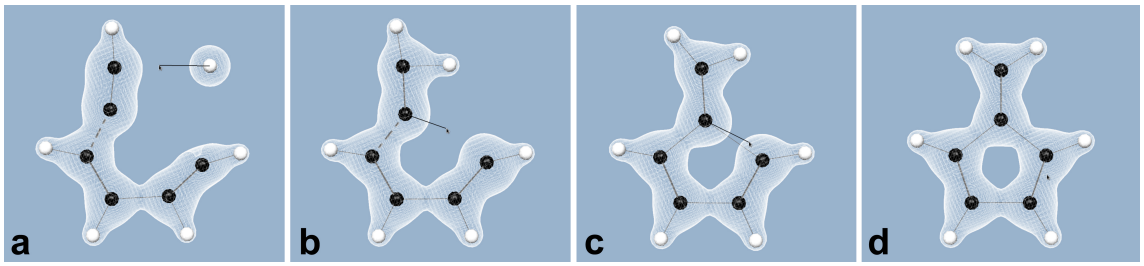

Figure 7. ***Interactive modeling session****.After breaking a benzene cycle, the user moves a hydrogen atom closer to the top carbon atom to force them to bond (**a**). Then, the user pulls on a carbon atom to form a fulvene molecule (**b**-**d**). Interactive electronic structure calculations allow the user to easily build plausible topologies, and get immediate feedback on the chemical structure.*

6.1.2. Adaptively Restrained Particle Simulations

Participants: Svetlana Artemova, Stéphane Redon.

Particle simulations are widely used in physics, chemistry, biology [17], [21], and even computer graphics [13]. However, many important problems still constitute significant computational challenges, including molecular docking, protein folding, diffusion across bio-membranes, fracture in metals, ion implantation, etc. Numerous methods have been developed to accelerate particle simulations, by *e.g.* increasing the simulation's time step [18], [9], [14], [26], [27], [24], improving the computational complexity of the simulation [32], [15], [8], [10], [31], or simplifying the system under study [19], [29], [28], [16], [11], [31], in particular *via* coarse-graining methods [20], [33], [30] or multiscale and multiresolution methods [25], [22], [23], [12].

We have introduced a novel, general approach to speed up particle simulations that we call Adaptively Restrained Particle Simulations (ARPS).

Our approach works by adaptively switching *positional* degrees of freedom on and off repeatedly during a simulation, while letting *momenta* evolve. The benefits of this approach are that (a) it is mathematically grounded and is able to produce stable, energy-preserving simulations; (b) it does not requires modifications to the simulated interaction potential, so that any suitable existing force-field can be directly used with ARPS; (c) under frequently-used assumptions on the interaction potential, ARPS make it possible to reduce the number of forces that have to be updated at each time step, which may significantly speed up simulations; (d) when performing constant-energy simulations, ARPS allow users to finely and continuously trade between precision and computational cost, and rapidly obtain approximate trajectories; (e) the trade-off between precision and cost may be chosen for each particle independently, so that users may arbitrarily focus ARPS on specific regions of the simulated system (*e.g.* a polymer in a solvent); (f) most important, when performing Adaptively Restrained Molecular Dynamics (ARMD) in the canonical (NVT) ensemble, unbiased statistics can be obtained.

We have illustrated ARPS on several numerical experiments, including a shock propagation example and a polymer-in-solvent study.

The shock propagation example demonstrates how ARPS make it possible to smoothly trade between precision and speed (Fig. 8).

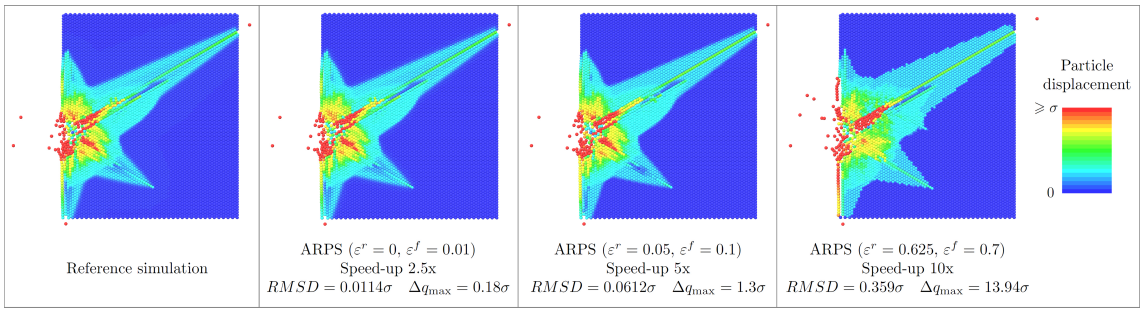

Figure 8. ***Simulating a shock propagation with controlled precision****. Adaptively restrained simulations allow us to smoothly trade between precision and speed. Even for large speed-ups (up to 10x) the features of the shock are extremely well preserved.*

The polymer-in-solvent study shows how one may collect unbiased statistics with ARPS, and demonstrates that it can be done faster than with usual (reference) simulations, The results are shown in Fig. 9.

These results have been submitted for publication.

6.1.3. Adaptive interactive quantum chemistry

Participants: Maël Bosson, Sergei Grudinin, Stéphane Redon.

We are now working on applying the adaptive paradigm to the quantum chemistry methods, to allow for interactive editing of systems of any sizes and shapes. We are developing new methods and criteria to adaptively focus the computational resources on the most relevant parts of the system. Figure 10 illustrates our recent results in this direction. In this framework, we can already achieve interactive rates and efficient virtual prototyping for systems of size up to thousand atoms on a current desktop computer.

6.1.4. Interactive molecular modeling with haptic feedback

Participants: Aude Bolopion, Barthelemy Cagneau, Stéphane Regnier, Stéphane Redon.

In collaboration with ISIR in Paris and LISV in Versailles, we have developed a new approach for haptic interaction with molecular systems.

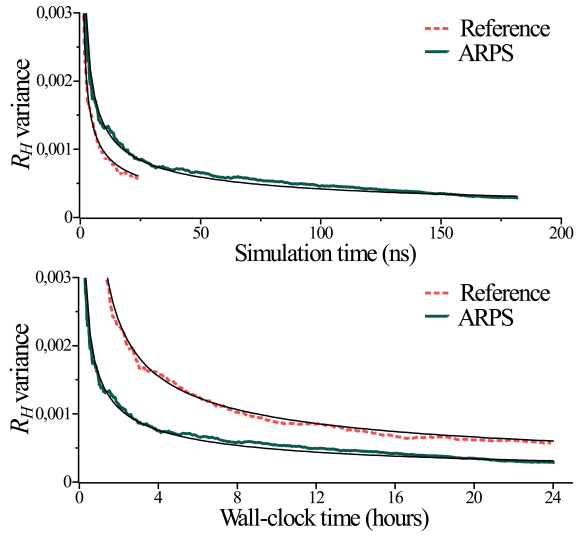

Figure 9. ***Computing the hydrodynamic radius*** R_H ***of a solvated polymer****. Traditional simulations reduce the variance more at each time step (top), but adaptively restrained (AR) simulations perform many more time steps, so that they reduce the variance faster in wall-clock time (bottom). In this example, for any target precision, AR simulations compute the hydrodynamic radius four times faster than reference simulations.*

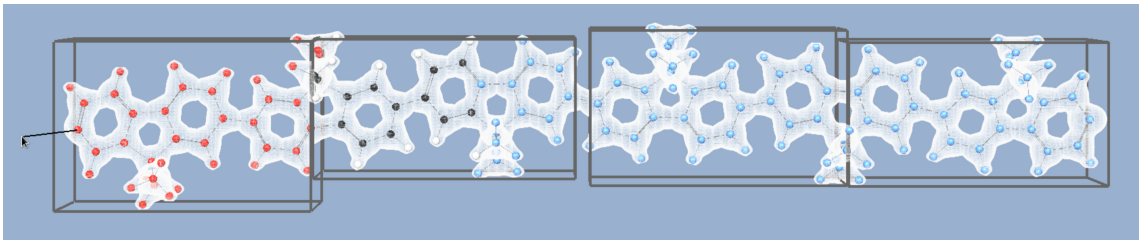

Figure 10. ***Interactive electronic structure calculations in SAMSON****. In this example the system is divided in four subsystems (for which the bounded boxes are displayed). The electronic structure is adaptively updated and the geometry is being optimized while the user edits the molecular system. Because the user pulls on one atom in the left part of the system, the electronic structure is accurately recomputed for the most left subsystem (atoms in red). In the neighboring subsystem, the electronic structure is updated with a cheaper model (carbon atoms are in black and hydrogen in white). In the right part of the system, the user force do not have a sufficiently large impact and atoms as well as the electronic structure are frozen (frozen atoms are displayed in blue).*

Molecular interactions typically have a high dynamic range (HDR), combining short-range stiff repulsive effects with long-range, soft attractive and repulsive terms. As a result, faithful haptic rendering of such molecular interactions is both important and difficult, in particular in applications where the precise perception of molecular forces is necessary (e.g. in molecular docking simulations). Traditionally, teleoperation coupling using constant gain control schemes have limited applications since they are unable to transmit to users low attractive forces without truncating repulsive ones. Furthermore, constant scaling displacement induces either instability or time-consuming experiments (displacements are slow), which deteriorates the ease of manipulation. We have described a variable gain haptic coupling method specifically designed to render high dynamic range (molecular) forces. The proposed method has been evaluated by user tests on an experiment involving two water molecules. We have observed that variable force amplification is widely appreciated, whereas variable displacement scaling is appropriate only for users that are already familiar with haptic manipulation. A complex experiment on a HIV molecule has been carried out using this variable gain system. This approach has been published in the proceedings of the 2011 World Haptics Conference [7]. Figure 11 shows SAMSON being used with a haptic interface at ISIR.

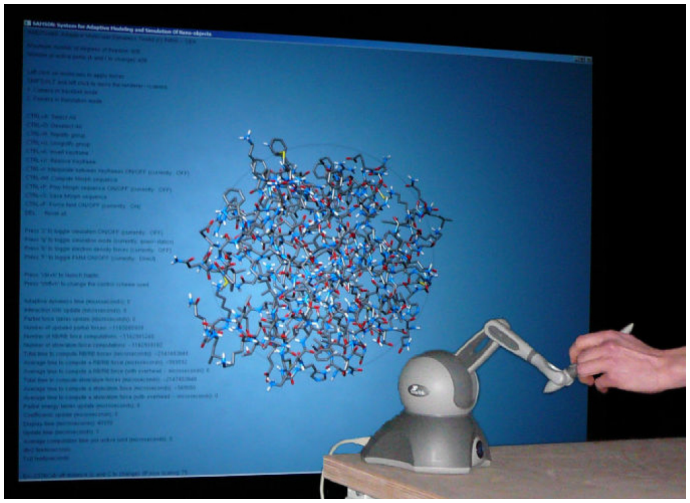

Figure 11. Haptic interaction with the HIV protease in SAMSON. Virtual environment setup at ISIR.

6.2. Algorithms for molecular docking

6.2.1. Prediction of Interface Water Molecules Using a Knowledge Base

Participants: Georgiy Derevyanko, Sergei Grudinin.

We developed a method to predict positions for interface water molecules as part of the predicted protein-protein complex. For this purpose we used a previously developed knowledge-base scoring methodology. First, we constructed a training set of non-homologous protein complexes with interfacial water. Then, we deduced the water-protein interaction energy using this training set. And finally, we positioned water molecules around a test protein complex on a regular grid and optimized their positions according to the knowledge-based water-protein interaction energy. This method was validated in a recent CAPRI competition. Figure 12 illustrates our method on a test protein.

6.2.2. Development of a Knowledge-Based Scoring Function

Participants: Georgiy Derevyanko, Sergei Grudinin.

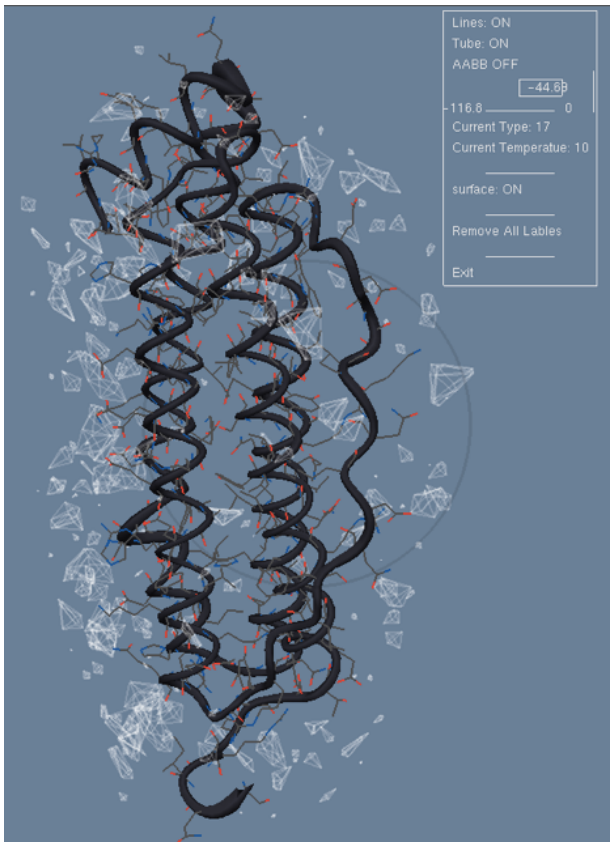

Figure 12. Densities of interface water molecules around a test protein computed using our water-protein knowledge-based potential.

We developed a new method to obtain a knowledge-based potential function for protein-protein interactions. To derive such a potential, we formulated a convex quadratic programming problem with about 1,000,000 of linear constraints and developed a fast iterative solver to solve it. We validated this scoring function in the CAPRI competition Round 24, where our prediction of the Target 50 was ranked 4th.

Figure 13 shows the use of Legendre polynomials to fit statistics obtained on the knowledge base.

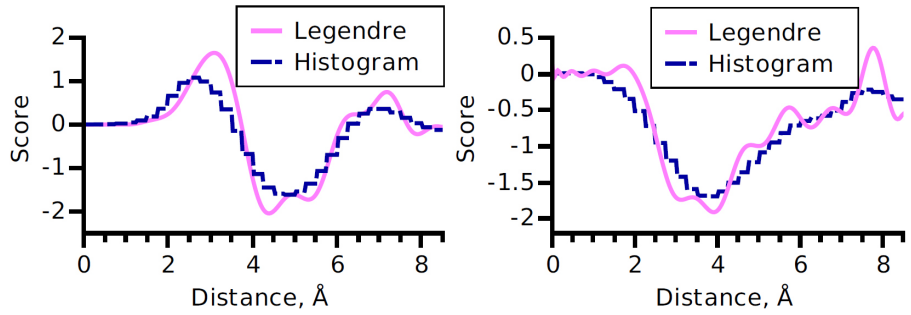

Figure 13. Using Legendre polynomials to fit protein-protein interaction statistics.

6.2.3. Development of a Local Knowledge-Based Potential for Structure Optimization and Prediction of Point Mutations in a Protein

Participants: Petr Popov, Sergei Grudinin.

We developed and validated a method that reconstructs the shape of the binding potential function between two proteins by locations of its global minima. After, we used the obtained potential function for optimization of positions of two docking partners. We demonstrated that using our method we can significantly improve the quality of predictions of such widely-used docking algorithms as HEX and ZDOCK. We validated this method in the CAPRI competition Round 26, where our re-scoring prediction of the Target 53 was ranked 3rd.

We have also developed a method to predict the influence of point mutations on the binding affinity constant of a protein complex. First, we made point mutations and reconstructed the sidechain of the mutated residue. Then, we repacked the sidechains that are within a certain cutoff distance from the mutated residue. After, we optimized the structure of two proteins using a smooth pair-additive knowledge-based potential function. We iteratively repeated the two previous steps until convergence of the binding energy. Finally, we converted the obtained binding energy into the binding affinity constant of two proteins. We validated this method in the CAPRI competition Round 26 with the Targets 55 and 56.

Figure 14 shows an interactive docking session using a knowledge-based potential for CAPRI Round 26 Target 53.

6.3. Software engineering

6.3.1. SAMSON's architecture

Participant: Stéphane Redon.

The data model of SAMSON has been expanded. The goal is to represent a nanosystem as the union of several interacting models: *structural models* (geometrical and topological information, to define relationships between structural elements), *dynamical models* (to define kinematical and dynamical relationships between structural components), *interaction models* (to define physical interactions between dynamical components, e.g. forces between atoms or rigid bodies), and *visual models* (visual representations, for user interaction).

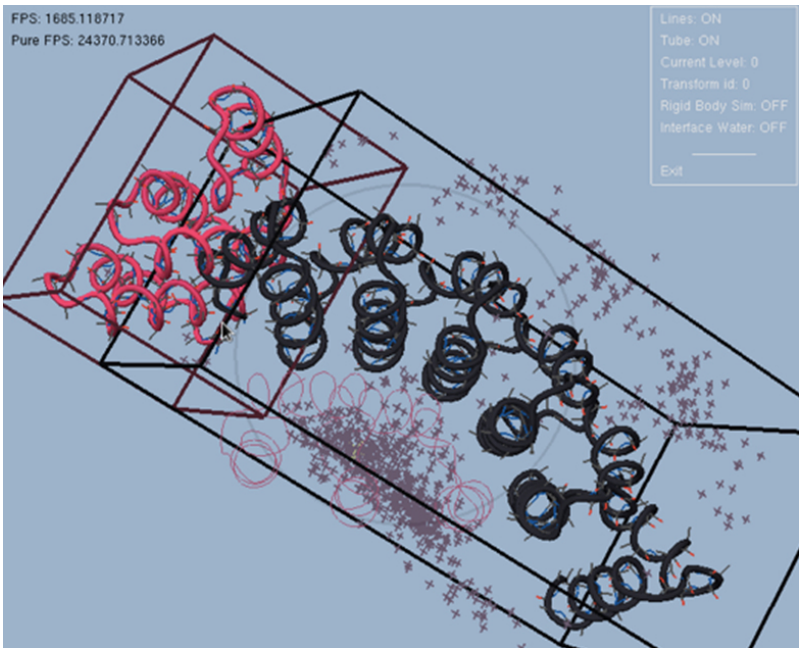

Figure 14. Interactive docking session using a knowledge-based potential for CAPRI Round 26 Target 53.

All models are part of the *data graph*, which contains all the information related to the system being modeled. The referencing system has been significantly expanded, with data structures to safely handle objects creation and deletion. An *event mechanism* has been designed and added to SAMSON, so that nodes of the data graph may send messages to each other. These messages can be related to topological changes, structural changes, dynamical changes, etc.

6.3.2. SAMSON's software engineering process

Participants: Jocelyn Gaté, Stéphane Redon.

SAMSON's software development process has been much improved. CMAKE is used to ensure that all parts of SAMSON may easily be built on several platforms (Windows, Mac and Linux). Thanks to CMAKE, a variety of Integrated Development Environments may be used (Visual Studio, Eclipse, XCode, etc.).

CTEST and CDASH are used to test SAMSON, and the Pipol platform has been used to perform nightly builds.

6.3.3. Graphical User Interface design

Participants: Jocelyn Gaté, Stéphane Redon.

Several functionalities have been added to the graphical interface of SAMSON, including customizable toolbars (that plug-in developers will be able to modify), as well as a data graph view (Figure 15).

Also, because plug-ins might have complex interfaces and settings, a mechanism to save and load custom presets has been developed (Fig. 16).

6.4. Applications

Methods and tools developed in our group have been used in the following studies:

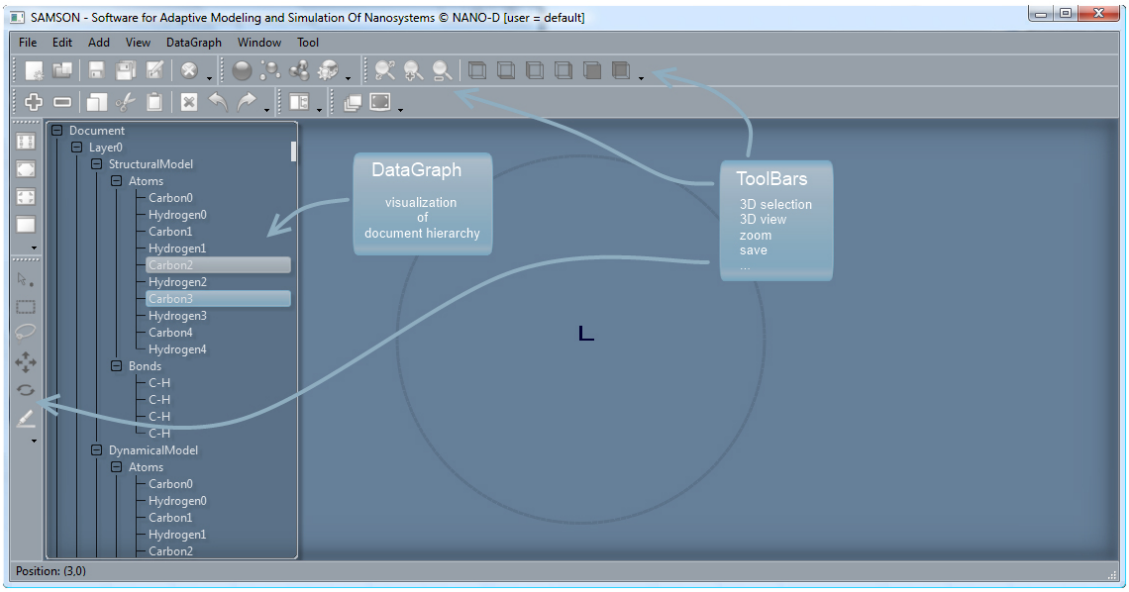

Figure 15. The current look of SAMSON's user interface.

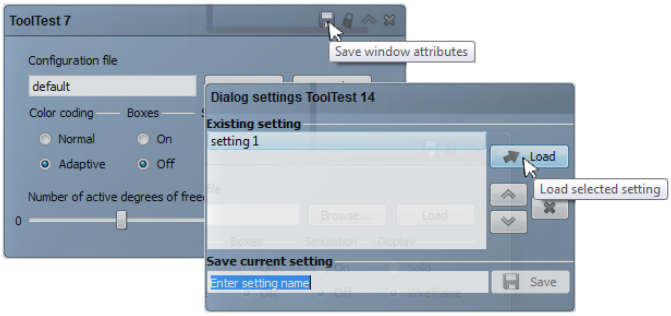

Figure 16. SAMSON allows users to save and load plug-in presets.

6.4.1. Building Blocks of Bacterial Chemoreceptor Arrays

Participant: Sergei Grudinin.

Bacterial chemoreceptors are known to cluster at the cell poles where they form partially hexagonally ordered arrays. This clusterisation is important for the function of chemotaxis system. In this study, we performed an analysis of the known structural and biochemical information on the components of chemoreceptor arrays: chemoreceptors themselves, histidine kinases and adapter proteins. Based on this analysis, we proposed a set of basic interactions within the chemotaxis system(the array building blocks) and constructed their atomistic models. The models resulting from these blocks are in agreement with experimental information and provide a basis for understanding the atomic-level structural organization of chemoreceptor arrays.

6.4.2. A Novel Dimerization Interface of Cyclic Nucleotide Binding Domain

Participant: Sergei Grudinin.

Cyclic nucleotide binding domain (CNBD) is a ubiquitous domain of effector proteins involved in signalling cascades of prokaryota and eukaryota. In this study, we described a novel CNBD dimerization interface found in crystal structures of bacterial CNG channel MlotiK1 and mammalian second messenger cAMP-activated guanine nucleotide-exchange factor Epac2. Using computational tools we demonstrated that the found interface is stable, in contrast to the dimerization interface reported previously. Comparisons with cN-bound structures of CNBD showed that the dimerization is incompatible with second messenger cAMP binding. Thus, the cAMP-dependent monomerization of CNBD may be an alternative mechanism of the cAMP sensing. Based on these findings, we proposed a model of the bacterial CNG channel modulation by cAMP.

7. Partnerships and Cooperations

7.1. National Initiatives

NANO-D is currently receiving funding from four ANR programs:

- **ANR JCJC**: 340,000 Euros over three years (2011-2014). This grant has been provided to S. Redon by the French Research Agency for being a finalist in the ERC Starting Grant 2009 call, and is for two PhD students and an engineer.
- **ANR MN**: 180,000 Euros over four years (2011-2015). This project, coordinated by NANO-D (S. Grudinin), gathers biologists and computer scientists from three research groups: Dave Ritchie at LORIA, Valentin Gordeliy at IBS (total grant: 360,000 Euros).
- **ANR COSINUS**: 85,000 Euros over four years (2009-2012). This project, coordinated by NANO-D (S. Redon), gathers physicists, biologists and computer scientists from five research groups: Xavier Bouju and Christian Joachim at CEMES, Martin J. Field at IBS, Serge Crouzy at CEA/LCBM, Thierry Deutsch and Frederic Lançon at CEA/SP2M (total grant: 380,000 Euros).
- **ANR PIRIBio**: 25,000 Euros over four years (2010-2013). We are participating in this project coordinated by Michel Vivaudou at IBS, with Serge Crouzy at CEA/LCBM and Frank Fieschi at IBS.

7.2. International Initiatives

7.2.1. Visits of International Scientists

7.2.1.1. Professors

- Matej Praprotnik. Department of Physics. Faculty of Mathematics and Physics. University of Ljubljana. Jadranska 19. SI-1000 Ljubljana. Slovenia.

- Andreas Heyden. Chemical Engineering. University of South Carolina. 301 Main St. Columbia, SC 29208. United States.

7.2.1.2. Internship

Petrus Popov

Subject: Conformational sampling strategies for macromolecules

Institution: M.M. Shemyakin & Yu.A. Ovchinnikov Institute of Bioorganic Chemistry (Russia (Russian Federation))

8. Dissemination

8.1. Animation of the scientific community

8.1.1. Program Committees

Stéphane Redon was a member of the program committee of the 2011 SIAM Conference on Geometric and Physical Modeling (GD/SPM 2011).

8.1.2. Steering Committees

Stéphane Redon is a member of the steering committee of the Nanoscience Foundation in Grenoble.

8.2. Participation to conferences, seminars

- M. Bosson attended the First Les Houches school in computational physics - soft matter (June 2011).
- S. Grudinin gave a talk on "Computer Modeling of Macromolecules" at a workshop "Workshop on Advanced Studies in Cell Biophysics", MIPT, Moscow, Russia (May 25, 2011).
- S. Grudinin participated in a conference "From Computational Biophysics to Systems Biology (CBSB11)", Juelich, Germany (July 20-22, 2011).
- S. Grudinin participated in a workshop "Innovative Approaches to Computational Drug Discovery", Lausanne, Switzerland (October 3-6, 2011).
- S. Grudinin gave a talk "Quadratic Optimization to Predict Protein-Protein Interactions" at a workshop "INRIA Workshop on Statistical Learning", Paris, France (December 5-6, 2011) where he participated with P. Popov.
- S. Redon gave a talk “Adaptive Algorithms for Modeling and Simulating Nanosystems” at CEA-LIONS (April 7, 2011).
- S. Redon gave a talk “Adaptive Algorithms for Modeling and Simulating Nanosystems” at McGill University, Montreal, Canada (August 28, 2011).

8.3. Teaching

8.3.1. Maël Bosson

- Licence: Analysis (Lebesgue theory, Fourier transform, distribution theory), 36h, INPG ENSIMAG, France
- Licence: Advanced numerical methods 12h, INPG ENSIMAG, France
- Licence: Numerical Analysis, 12h, INPG Pagora, France
- High School: Mobinet, 12h, INRIA Grenoble Rhone-Alpes, France
- High School: Introduction to chemistry, 3h, CIME Nanoschool, Grenoble, France

8.3.2. Sergei Grudinin

- Licence: Seminars on “Modeling and Simulations of Macromolecules”, 10h (May 23 - 37; September 19-26 2011), MIPT, Moscow, Russia

8.3.3. Stéphane Redon

- Grad: Lecturer in the “Hands-on Course: Coarse Grain Methods for Biomolecular Simulations”, 6h, Institut Pasteur, Montevideo, Uruguay (October 2011).
- Licence: “Introduction to computer science”, INF311 and INF321, 80h, Ecole Polytechnique, Paris, France

9. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] 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, p. 2865–2877.

[2] S. ARTEMOVA, S. GRUDININ, S. REDON. *Fast construction of assembly trees for molecular graphs*, in "Journal of Computational Chemistry", 2011, vol. 32, n^{o} 8, p. 1589-1598.

[3] M. BOSSON, S. GRUDININ, X. BOUJU, S. REDON. *Interactive physically-based structural modeling of hydrocarbon systems*, in "Journal of Computational Physics", 2011, vol. 231, n^{o} 6, p. 2581-2598.

[4] M. BOSSON, C. RICHARD, A. PLET, S. GRUDININ, S. REDON. *Interactive quantum chemistry: A divide-and-conquer ASED-MO method*, in "Journal of Computational Chemistry", 2012, vol. 33, n^{o} 7, p. 779-790.

[5] I. GUSHCHIN, V. GORDELIY, S. GRUDININ. *Role of the HAMP Domain Region of Sensory Rhodopsin Transducers in Signal Transduction*, in "Biochemistry", 2011, vol. 50, p. 574 - 580.

[6] I. GUSHCHIN, A. RESHETNYAK, V. BORSHCHEVSKIY, A. ISHCHENKO, E. ROUND, S. GRUDININ, M. ENGELHARD, G. BÜLDT, V. GORDELIY. *Active State of Sensory Rhodopsin II: Structural determinants for signal transfer and proton pumping* , in "Journal of Molecular Biology", 2011, vol. 412, p. 591 - 600.

International Conferences with Proceedings

[7] A. BOLOPION, B. CAGNEAU, S. REDON, S. REGNIER. *Variable gain haptic coupling for molecular simulation*, in "World Haptics Conference (WHC), 2011 IEEE", 2011, p. 469–474.

References in notes

[8] J. BARNES, P. HUT. *A hierarchical 0 (N log iV) force-calculation algorithm*, in "Nature", 1986, vol. 324, n^{o} 4.

[9] C. BENNETT. *Mass tensor molecular dynamics*, in "J. Comput. Phys.", 1975, vol. 19, n^{o} 3, p. 267–279.

[10] P. CALLAHAN, S. KOSARAJU. *A decomposition of multidimensional point sets with applications to k-nearest-neighbors and n-body potential fields*, in "Journal of the ACM (JACM)", 1995, vol. 42, n^{o} 1, p. 67–90.

[11] M. CHRISTEN, W. VAN GUNSTEREN. *Multigraining: an algorithm for simultaneous fine-grained and coarse-grained simulation of molecular systems*, in "J. Chem. Phys.", 2006, vol. 124, 154106.

[12] B. ENSING, S. NIELSEN, P. MOORE, M. KLEIN, M. PARRINELLO. *Energy conservation in adaptive hybrid atomistic/coarse-grain molecular dynamics*, in "J. Chem. Theory Comput.", 2007, vol. 3, n^{o} 3, p. 1100–1105.

[13] O. ETZMUSS, J. GROSS, W. STRASSER. *Deriving a particle system from continuum mechanics for the animation of deformable objects*, in "Visualization and Computer Graphics, IEEE Transactions on", 2003, vol. 9, n^{o} 4, p. 538–550.

[14] K. FEENSTRA, B. HESS, H. BERENDSEN. *Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems*, in "J. Comput. Chem.", 1999, vol. 20, n^{o} 8, p. 786–798.

[15] L. GREENGARD, V. ROKHLIN. *A fast algorithm for particle simulations*, in "J. Comput. Phys.", 1987, vol. 73, n^{o} 2, p. 325–348.

[16] A. GUNARATNE. *A penalty function method for constrained molecular dynamics*, Citeseer, 2006.

[17] R. HOCKNEY, J. EASTWOOD. *Computer simulation using particles*, Institute of Physics, 1992.

[18] J. IZAGUIRRE, S. REICH, R. SKEEL. *Longer time steps for molecular dynamics*, in "J. Chem. Phys.", 1999, vol. 110, 9853.

[19] H. LIN, D. TRUHLAR. *QM/MM: what have we learned, where are we, and where do we go from here?*, in "Theoretical Chemistry Accounts: Theory, Computation, and Modeling (Theoretica Chimica Acta)", 2007, vol. 117, n^{o} 2, p. 185–199.

[20] S. MARRINK, H. RISSELADA, S. YEFIMOV, D. TIELEMAN, A. DE VRIES. *The MARTINI force field: coarse grained model for biomolecular simulations*, in "J. Phys. Chem. B", 2007, vol. 111, n^{o} 27, p. 7812–7824.

[21] J. MONAGHAN. *Smoothed particle hydrodynamics*, in "Reports on Progress in Physics", 2005, vol. 68, 1703.

[22] S. NIELSEN, R. BULO, P. MOORE, B. ENSING. *Recent progress in adaptive multiscale molecular dynamics simulations of soft matter*, in "Phys. Chem. Chem. Phys.", 2010, vol. 12, n^{o} 39, p. 12401–12414.

[23] J. PARK, A. HEYDEN. *Solving the equations of motion for mixed atomistic and coarse-grained systems*, in "Mol. Simul.", 2009, vol. 35, n^{o} 10, p. 962–973.

[24] P. PLECHAC, M. ROUSSET. *Implicit Mass-matrix Penalization of Hamiltonian Dynamics with Application to Exact Sampling of Stiff Systems*, in "Multiscale Modeling & Simulation", 2010, vol. 8, n^{o} 2, p. 498-539.

[25] M. PRAPROTNIK, L. SITE, K. KREMER. *Multiscale simulation of soft matter: From scale bridging to adaptive resolution*, in "Annu. Rev. Phys. Chem.", 2008, vol. 59, p. 545–571.

[26] M. PRETO, S. TREMAINE. *A class of symplectic integrators with adaptive time step for separable Hamiltonian systems*, in "Astron J", 1999, vol. 118, 2532.

[27] F. RAO, M. SPICHTY. *Thermodynamics and kinetics of large-time-step molecular dynamics.*, in "J. Comput. Chem.", 2011.

[28] S. REICH. *Modified potential energy functions for constrained molecular dynamics*, in "Numerical Algorithms", 1998, vol. 19, n^{o} 1, p. 213–221.

[29] B. ROUX. *Implicit solvent models*, in "Eastern Hemisphere Distribution", 2001, 133.

[30] A. SHIH, A. ARKHIPOV, P. FREDDOLINO, K. SCHULTEN. *Coarse grained protein-lipid model with application to lipoprotein particles*, in "J. Phys. Chem. B", 2006, vol. 110, n^{o} 8, p. 3674–3684.

[31] M. TUCKERMAN, B. BERNE, G. MARTYNA. *Reversible multiple time scale molecular dynamics*, in "J. Chem. Phys.", 1992, vol. 97, n^{o} 3, 1990.

[32] A. VOTER. *Hyperdynamics: Accelerated molecular dynamics of infrequent events*, in "Phys. Rev. Lett.", 1997, vol. 78, n^{o} 20, p. 3908–3911.

[33] G. VOTH. *Coarse-graining of condensed phase and biomolecular systems*, CRC, 2009.