Activity Report 2019

Team NANO-D

Algorithmes pour la Modélisation et la Simulation de Nanosystèmes

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Grenoble - Rhône-Alpes

THEME
Numerical schemes and simulations

Table of contents

Team NANO-D

Creation of the Team: 2018 October 01

Keywords:

Computer Science and Digital Science:

A3.4.1. - Supervised learning
A3.4.6. - Neural networks
A5.5.1. - Geometrical modeling
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A9. - Artificial intelligence

Other Research Topics and Application Domains:

B1.1.1. - Structural biology
B1.1.7. - Bioinformatics
B2.6.3. - Biological Imaging

1. Team, Visitors, External Collaborators

Research Scientists

Sergei Grudinin [Team leader, CNRS, Researcher]
Leonard Jaillet [Inria, Starting Research Position, until Jan 2019]

Post-Doctoral Fellow

Didier Devaurs [Univ Grenoble Alpes, Post-Doctoral Fellow]

PhD Students

Maria Kadukova [Univ Grenoble Alpes, PhD Student]
Guillaume Pages [Inria, PhD Student, until May 2019]

Interns and Apprentices

Ilia Igashov [Inria, from Nov 2019]
Khalid Mustafin [Inria, until Feb 2019]
Dmitrii Sergeevich Zhemchuzhnikov [Univ Grenoble Alpes, from May 2019 until Sep 2019]

Administrative Assistant

Maria Immaculada Presseguer [Inria, Administrative Assistant]

Visiting Scientists

Karina Dos Santos Machado [FURG Brazil, until Oct 2019]
Vadim Strijov [MIPT Moscow, from Jun 2019 until Aug 2019]

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* [20], *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 [62], for neighbor search between large rigid molecules [19], and for bond-order reactive force-fields [23]. 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 [21] and over 500,000 structures of small molecules stored in the Cambridge Structural Database [15]. We are developing algorithms for protein-protein interactions and protein-ligand interactions.
   - **Developing parametrization algorithms for interaction potentials**: Molecular models typically require their own potential energy function (or a *forcefield*) to be assigned. However, the development of a new potential function is a very difficult and sometimes challenging task [41]. Therefore, we are developing algorithms for automatic parametrization of new potential functions for some particular representations of a molecular system.
   - **Developing algorithms for exhaustive sampling**: Some application domains, such as computational docking, cryo-EM rigid-body fitting, etc., require sampling in a low-dimensional space. For such applications it is advantageous to perform an exhaustive search rather than accelerated sampling [59]. 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 [61].

The twenty-first century is most likely to see a similar development at the atomic scale. Indeed, the recent years have seen tremendous progress in nanotechnology. The magazine Science, for example, recently featured a paper demonstrating an example of DNA nanotechnology, where DNA strands are stacked together through programmable self-assembly [32]. In February 2007, the cover of Nature Nanotechnology showed a "nano-wheel" composed of a few atoms only. Several nanosystems have already been demonstrated, including a *de-novo* computationally designed protein interface [33], a wheelbarrow molecule [43], a nano-car [65], a Morse molecule [16], etc. Typically, these designs are optimized using semi-empirical quantum mechanics calculations, such as the semi-empirical ASED+ calculation technique [17].

While impressive, these are but two examples of the nanoscience revolution already impacting numerous fields, including electronics and semiconductors [50], textiles [48], [38], energy [52], food [27], drug delivery [37], [68], chemicals [39], materials [28], the automotive industry [14], aerospace and defense [34], medical devices and therapeutics [30], medical diagnostics [69], etc. According to some estimates, the world market for nanotechnology-related products and services will reach one trillion dollars by 2015 [60]. 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 [71], removing some degrees of freedom (e.g. keeping torsion angles only [46], [66]), coarse-graining [70], [63], [18], [64], multiple time step methods [57], [58], fast multipole methods [31], parallelization [45], averaging [26], multi-scale modeling [25], [22], reactive force fields [24], [74], interactive multiplayer games for predicting protein structures [29], 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 [72], [73].

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 [53], [54], [55], [56], the adaptive multiscale method [51], and the adaptive partitioning of the Lagrangian method [40]. 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).

4. Highlights of the Year

4.1. Highlights of the Year

- The work on first-principle simulation has been completed. The aim was to use the restrained dynamical model ARPS previously developed by the team to speed-up dynamical simulations using a first-principle interaction model. We have chosen Orbital-Free Density Functional Theory (OF-DFT), a fast scheme of DFT, as interaction model. We have developed a new OF-DFT code adapted to restrained particle simulations and have compared the accuracy and speed of our method to the state of the art OF-DFT code, PROFESS. The results were published in the Journal of Computational Chemistry [11] and the code is available in SAMSON. The thesis at the origin of this research has been defended in October.
- The proof-of-concept orientation-dependent potential for small molecules was developed and tested.
- With the advance of experimental procedures, obtaining sparse experimental data of proteins in solution (chemical crosslinking and small-angle scattering) is becoming a fast and routine practice. These can greatly enhance the accuracy of protein structure modeling. We participated in reviewing the current state of the art in modeling protein structures with the assistance of experimentally determined chemical crosslinks and small-angle scattering profiles within the framework of the 13th meeting of Critical Assessment of Structure Prediction approaches [2], [4].

5. New Software and Platforms

5.1. Knodle

KNOwledge-Driven Ligand Extractor

KEYWORDS: Bioinformatics - Machine learning

FUNCTIONAL DESCRIPTION: KNOwledge-Driven Ligand Extractor is a software library for the recognition of atomic types, their hybridization states and bond orders in the structures of small molecules. Its prediction model is based on nonlinear Support Vector Machines. The process of bond and atom properties perception is divided into several steps. At the beginning, only information about the coordinates and elements for each atom is available :

Connectivity is recognized. A search of rings is performed to find the Smallest Set of Smallest Rings (SSSR). Atomic hybridizations are predicted by the corresponding SVM model. Bond orders are predicted by the corresponding SVM model. Aromatic cycles are found. Atomic types are set in obedience to the functional groups. Some bonds are reassigned during this stage.

- Participants: Maria Kadukova and Sergei Grudinin
- Partner: MIPT Moscow
- Contact: Sergei Grudinin
- Publication: Knodle: A Support Vector Machines-Based Automatic Perception of Organic Molecules from 3D Coordinates
- URL: https://team.inria.fr/nano-d/software/Knodle/

5.2. DeepSymmetry

KEYWORDS: Bioinformatics - 3D modeling - Machine learning - Neural networks

FUNCTIONAL DESCRIPTION: DeepSymmetry is a method based on three-dimensional (3D) convolutional networks that detects structural repetitions in proteins and their density maps. It identifies tandem repeat proteins, proteins with internal symmetries, their symmetry order, and also the corresponding symmetry axes.

- Participants: Guillaume Pages and Sergei Grudinin
- Contact: Sergei Grudinin
- Publication: DeepSymmetry : Using 3D convolutional networks for identification of tandem repeats and internal symmetries in protein structures
- URL: https://team.inria.fr/nano-d/software/deepsymmetry/

5.3. Ananas

Analytical Analyzer of Symmetries

KEYWORDS: Bioinformatics - Structural Biology

FUNCTIONAL DESCRIPTION: Analytical Analyzer of Symmetries is a software for detection and assessment of the quality of symmetry in a protein assembly.

This software can : Detect the best axes of symmetry for any symmetry group in an assembly containing the right amount of chains, Provide the symmetry-aware RMSD for these axes, Detect the best axis of symmetry for cyclic assemblies with missing subunits, Compute the axes of symmetry with user-provided correspondences.

RELEASE FUNCTIONAL DESCRIPTION: Version 0.7 from 3 Sep 2018: mmCIF support added. A file can be gzipped. Version 0.8 from 30 July 2019: Checks the similarity between chains. Explicitly output when an incomplete assembly is detected as symmetric. Version 0.9 from 8 Nov 2019: Added lists of excluded chains for both PDB and CIF formats.

- Participants: Guillaume Pages and Sergei Grudinin
- Contact: Sergei Grudinin
- Publications: Analytical symmetry detection in protein assemblies. I. Cyclic symmetries - Analytical symmetry detection in protein assemblies. II. Dihedral and Cubic symmetries
- URL: https://team.inria.fr/nano-d/software/ananas/

5.4. Pepsi-SAXS

KEYWORDS: Bioinformatics - Structural Biology - Data modeling

FUNCTIONAL DESCRIPTION: Pepsi-SAXS (PEPSI stands for Polynomial Expansions of Protein Structures and Interactions) is new implementation of the multipole-based scheme initially proposed by Stuhrmann (Stuhrmann, 1970). Overall, our method is significantly faster with a similar accuracy compared to Crysol, FoXS, and the 3D-Zernike implementation from the SAStbx package.

RELEASE FUNCTIONAL DESCRIPTION: Version 1.0 from 6th March 2018: Added computation of P(r). Version 1.1 from 19th April 2018: Excluded solvent radii updated. Many more ions added. Version 1.2 from 24th April 2018: Added automatic identification of labile and nonlabile explicit hydrogens. Added bulk SLD option. Version 1.3 from 30th July 2018: Added initial prototype of 2D dcattering images. Version 1.4 from 1st August 2018: Added docking module. Version 2.0 from January 2019: Added flexible optimization of conformations. A bug with absolute contrast fixed. Default behaiviour of Pepsi-SAXS changed, now it reads all the hydrogen atoms if the –hyd flag is set. Otherwise, it reads only explicit hydrogens on non-standard residues. Version 2.1 from February 2019: First implimentation of the multi-structure fitting.

- Participant: Sergei Grudinin
- Partner: MIPT Moscow
- Contact: Sergei Grudinin
- Publication: Pepsi-SAXS : an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles
- URL: https://team.inria.fr/nano-d/software/pepsi-saxs/

5.5. NOLB

NOn-Linear rigid Block NMA method

KEYWORDS: Structural Biology - Bioinformatics - Elasticity - Proteins - Motion analysis

FUNCTIONAL DESCRIPTION: It's a new conceptually simple and computationally efficient method for non-linear normal mode analysis of macromolecules.

RELEASE FUNCTIONAL DESCRIPTION: Version 1.2 from January 2019: Exclusion interaction list added. Version 1.3 from April 2019: Sequence and structural alignment fixed thanks to Guillaume Pages. More advanced iterative structural alignment added. Version 1.4 from June 2019: Firt version of domain identification added. Please use the "–nDomains" option Version 1.5 from July 2019: Rigid block format updated. JSON and Text output formats added. Version 1.6 from December 2019: Added initial Voronota support.

- Participants: Sergei Grudinin and Alexandre Hoffmann
- Contact: Sergei Grudinin
- Publications: NOLB: Nonlinear Rigid Block Normal Mode Analysis Method - RapidRMSD: Rapid determination of RMSDs corresponding to motions of flexible molecules
- URL: https://team.inria.fr/nano-d/software/nolb-normal-modes/

5.6. SBROD

KEYWORDS: Bioinformatics - Machine learning

FUNCTIONAL DESCRIPTION: Smooth orientation-dependent scoring function (SBROD) for coarse-grained protein quality assessment uses only the conformation of the protein backbone, and hence it can be applied to scoring the coarse-grained protein models.

The workflow of SBROD consists in two stages. First, the method extracts features from each protein model in the dataset. Then, the scoring function assigns a score to each processed protein model depending on its features extracted at the first stage. Figure above schematically shows the workflow of SBROD. Here, four types of inter-atomic interactions, described in details below, are taken into account when extracting the features. After these features have been extracted and preprocessed, a Ridge Regression model is trained on them to predict the GDT-TS of protein models.

- Participants: Mikhail Karasikov, Guillaume Pages and Sergei Grudinin
- Contact: Sergei Grudinin
- Publication: Smooth orientation-dependent scoring function for coarse-grained protein quality assessment
- URL: https://team.inria.fr/nano-d/software/sbrod/

5.7. Ornate

KEYWORDS: Bioinformatics - Machine learning - Neural networks

FUNCTIONAL DESCRIPTION: Oriented Routed Neural network with Automatic Typing is a method for protein quality assessment. Ornate is a residue-wise scoring method. It first constructs a three dimensional map representing the structure of the residue, and its neighborhood.

- Participants: Guillaume Pages, Benoit Charmettant and Sergei Grudinin
- Contact: Sergei Grudinin
- Publication: Protein model quality assessment using 3D oriented convolutional neural networks
- URL: https://team.inria.fr/nano-d/software/ornate/

6. New Results

6.1. Reconstructing molecular shapes from SAXS data

We are working on a novel method to reconstruct the three-dimensional shape of a molecule from low-resolution experimental data. The structural data we are currently focusing on is obtained through small-angle X-ray scattering (SAXS) experiments, but we plan to also consider small-angle neutron scattering (SANS) data. Our *ab initio* reconstruction method is inspired by iterative phase-retrieval algorithms that can produce an image for an object when only the amplitudes of its Fourier transform are known and the phases are unknown. In our context, the X-ray scattering amplitudes associated with a molecule are the Fourier transform of its electron density. The novelty of our approach resides in the use of spherical harmonics expansions, which will allow performing the whole reconstruction process in Fourier space—contrary to existing methods that iterate between Fourier space and real space—for an improved computational efficiency. Our method is being implemented within the software PEPSI (polynomial expansions of protein structures and interactions).

6.2. Docking of cyclic molecules

In 2018 we participated in a docking Grand Challenge 4, which was organized by the Drug Design Data Resource (D3R) group. The goal was to predict correct poses and affinities of ligands binding the beta secretase 1 (BACE) receptor. Most of the ligands were macrocyclic. The challenge answers and results were fully released in February 2019. Upon that we started analyzing our protein-ligand docking strategy and wrote a corresponding paper [5].

6.3. Convex-PL and entropy

During the 2019 we continued developing our knowledge-based scoring function for protein-ligand interactions called Convex-PL [44]. We introduced a new descriptor characterizing side-chain entropy, tried two ways of computing the solvent-accessible surface area (SASA) descriptors with either using buried SASA, or SASA of those atoms that are contributing to the protein-ligand interaction according to the scoring function design (i.e which are within the cutoff distance of an interaction potential). Using these descriptors alongwith Convex-PL score and ligand side chain entropy descriptor, we trained a ridge regression model to predict binding affinities. This modification of Convex-PL was assessed on the CASF Benchmark 2016 released in the end of 2018, and on a subset of the DUD benchmark [42]. We are currently preparing the corresponding manuscript for submission.

6.4. Orientational potential for small molecules

With Pablo Chacon and Karina Dos Santos Machado we have developed a novel statistical protein-ligand scoring function called KORP-PL. KORP-PL is based on the coarse grained backbone-only representation of protein that is very common in protein structure modeling. It is based on ideas implemented by Pablo Chacon in the preceding scoring function for protein quality assessment called KORP [49], where each amino acid residue was characterized with a 3D oriented frame. We kept this representation for protein residues and computed the distances and angles between the oriented frames and points that describe ligand atoms. Using this data, we then derived the scoring function statistically in the same way, as it was done for protein-protein interactions in KORP. We also ran an optimization procedure to compute weights for each residue-atom interaction that would minimize the difference between predicted and experimental binding constants. We have assessed KORP-PL on several benchmarks [47], [67], one of which was manually derived from the user-submitted data of the D3R Challenges 2 [35], 3 [36], and 4. On all of them it performed exceptionally in pose prediction despite being a coarse-grained scoring function. We are currently preparing the corresponding manuscript for submission.

6.5. Analysis of a deep-learning architecture for fold recognition

Deep learning has recently demonstrated outstanding capabilities in classical pattern recognition problems. It has also obtained a tremendous success in the very recent protein structure prediction tasks. This work studies recurrent structural patterns in protein structures recognized by a deep neural network. We demonstrated that neural networks can automatically learn a vast amount of chemo-structural features with only a very little amount of human supervision. For example, our architecture correctly learns atomic, amino acid, and also higher-level molecular descriptors. The network architecture and the results are available at https://team.inria.fr/nano-d/software/Ornate/.

6.6. Controlled-advancement rigid-body optimization of nanosystems

In this study, we proposed a novel optimization algorithm, with application to the refinement of molecular complexes. Particularly, we considered optimization problem as the calculation of quasi-static trajectories of rigid bodies influenced by the inverse-inertia-weighted energy gradient and introduce the concept of advancement region that guarantees displacement of a molecule strictly within a relevant region of conformational space. The advancement region helps to avoid typical energy minimization pitfalls, thus, the algorithm is suitable to work with arbitrary energy functions and arbitrary types of molecular complexes without necessary tuning of its hyper-parameters. Our method, called controlled-advancement rigid-body optimization of nanosystems (Carbon), is particularly useful for the large-scale molecular refinement, as for example, the putative binding candidates obtained with protein–protein docking pipelines. Implementation of Carbon with user-friendly interface is available in the SAMSON platform for molecular modeling at https://www.samson-connect.net. The method was published in [10].

6.7. SAXS- and SANS-assisted modeling of proteins

We collaborated on data-assisted modeling of a KRAB-domain associated protein 1. Our work sheds light on its overall organization and combins solution scattering diffraction data, integrative modeling and single-molecule experiments [3].

We also participated in a combination of coarse-grained molecular dynamics simulations with previously measured small-angle scattering data to study the conformation of three-domain protein TIA-1 in solution. More precisely, we contributed with a specifically developed version of the Pepsi-SANS code. Our results suggest a general strategy for studying the conformation of multi-domain proteins in solution that combines coarse-grained simulations with small-angle X-ray scattering data that are generally most easy to obtain [13].

6.8. Predicting protein functional motions

Large macromolecules, including proteins and their complexes, very often adopt multiple conformations. Some of them can be seen experimentally, for example with X-ray crystallography or cryo-electron microscopy. This structural heterogeneity is not occasional and is frequently linked with specific biological function. Thus, the accurate description of macromolecular conformational transitions is crucial for understanding fundamental mechanisms of life's machinery.

We report on a real-time method to predict such transitions by extrapolating from instantaneous eigen-motions, computed using the normal mode analysis, to a series of twists. We demonstrate the applicability of our approach to the prediction of a wide range of motions, including large collective opening-closing transitions and conformational changes induced by partner binding. We also highlight particularly difficult cases of very small transitions between crystal and solution structures. Our method guaranties preservation of the protein structure during the transition and allows to access conformations that are unreachable with classical normal mode analysis. We provide practical solutions to describe localized motions with a few low-frequency modes and to relax some geometrical constraints along the predicted transitions. This work opens the way to the systematic description of protein motions, whatever their degree of collectivity. Our method is available as a part of the NOn-Linear rigid Block (NOLB) package at https://team.inria.fr/nano-d/software/nolb-normal-modes/ [12].

6.9. Protein structure prediction experiments

We participated in the CAPRI Round 46, the third joint CASP-CAPRI protein assembly prediction challenge. The Round comprised a total of 20 targets including 14 homo-oligomers and 6 heterocomplexes. Eight of the homo-oligomer targets and one heterodimer comprised proteins that could be readily modeled using templates from the Protein Data Bank, often available for the full assembly. The remaining 11 targets comprised 5 homodimers, 3 heterodimers, and two higher-order assemblies. These were more difficult to model, as their prediction mainly involved "ab-initio" docking of subunit models derived from distantly related templates [7].

7. Partnerships and Cooperations

7.1. Regional Initiatives

- An IDEX UGA grant is covering post-doc of Didier Devaurs, starting from December.
- Inria CORDI-S post-doctoral fellowship was obtained for Agnieszka Karczynska.

7.2. National Initiatives

7.2.1. ANR

In 2019, NANO-D had funding from one ANR program:

- **ANR PRCI**: covered the end of the PhD thesis of Guillaume Pages.

7.3. European Initiatives

7.3.1. Collaborations with Major European Organizations

The European Bioinformatics Institute (EMBL-EBI), Protein Data Bank in Europe (PDBe) team, Hinxton (UK)

We are collaborating on the integration of methods developed in the team into the PDBe web resource.

The Institute Laue-Langevin (ILL), the bioSANS team, Grenoble (France)

We are collaborating on the development of neutron small-angle scattering software

7.4. International Initiatives

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

7.4.1.1. FlexMol

Title: Algorithms for Multiscale Macromolecular Flexibility

International Partner (Institution - Laboratory - Researcher):

Rocasolano Institute of Physical Chemistry (IQFR-CSIC), Madrid, Spain (Spain) - Pablo Chacon

Start year: 2019

See also: https://team.inria.fr/nano-d/research/flexmol/

Molecular flexibility is essential to link structure and function of many biological macromolecules. Changes in protein conformation play a vital role in biochemical processes, from biopolymer synthesis to membrane transport. Many proteins can drastically alter their architecture and display considerable interdomain flexibility, as found in their 3D structures. For example, proteins rely on flexibility to respond to environmental changes, ligand binding and chemical modifications. Also, protein flexibility is tightly bound to their stability and is fundamental for drugs to exert biological effects.

Thus, one of the main challenges in the field of computational structural biology is to predict and explain molecular flexibility and corresponding conformational changes. For example, currently there are no methods that can reliably predict structural changes in proteins upon their binding. However, these are crucial to predict the structure of protein complexes with large conformational changes upon binding. To give another example, flexibility of the protein binding pocket is the major hurdle in reliable prediction of protein-ligand interactions for computer-aided drug design. Finally, intrinsic flexibility of macromolecules is nowadays the limiting factor for high-resolution experimental structure determination.

The partners of this associate team proposal comprise world-leading teams working with sound mathematical representations and techniques in the field of structural bioinformatics. These include spherical harmonics, normal modes analysis, high-order fast Fourier transforms, and more. The partners have very similar interests, but complimentary expertise. The goal of this collaboration is to mutually explore novel computational techniques for emerging problems in structural biology and bioinformatics related to molecular flexibility. This problem can be tackled at different scales. Large-scale flexibility of macromolecules can be efficiently described using collective coordinates. We will try to represent these in polynomial spaces, such that a practical flexible docking method can be based on this representation. Other applications include 3D shape reconstruction and scattering problems. Local molecular flexibility can be modelled using various techniques, including robotics-inspired methods, fragment libraries, etc. Here, our goal will be to rapidly sample the conformational space, and to construct a potential energy function applicable to flexible molecules. The ultimate goal of the project is to combine multiple levels of representation of molecular flexibility together. The project outcome will be built around innovative computer-aided drug-design algorithm with applications to prediction and computational design of important pharmaceutical targets such as antibodies.

7.4.2. Inria International Partners

7.4.2.1. Declared Inria International Partners : BIOTOOLS

Title: Novel Computational Tools for Structural Bioinformatics

International Partner (Institution - Laboratory - Researcher):

MIPT (Russia (Russian Federation)) - Department of Control and Applied Mathematics - Vadim Strijov

Duration: 2016 - 2020

Start year: 2016

Abstract : The general scientific objectives of the forthcoming collaboration are the new developments of computational tools for structural bioinformatics. In particular, we plan to collaborate on several subjects: 1. Development of tractable approximations for intractable combinatorial problems in structural biology. 2. Development of new computational tools for scattering experiments. 3. Machine learning for structural bioinformatics.

7.4.2.2. Informal International Partners

- University of Stony Brook, lab of Dima Kozakov (USA). We have been collaborating on the development of novel protein docking methods.
- University of Vilnius, department of Bioinformatics (Lithuania). We have been collaborating on the development of novel protein docking methods.
- KU Copenhagen (Denmark), department of Chemistry. We collaborated on the integrative structural biology approaches.
- Francis Crick Institute, London (UK), Biomolecular Modelling Laboratory. We collaborate on the development of flexible protein docking methods.
- University of Oslo. Ongoing collaboration on modeling protein systems guided by small-angle Xray and neutron small-angle scattering.
- University of Bergen, Norway. Ongoing collaboration on novel methods for normal mode analysis of protein structures.
- Nagoya University and RIKEN Center for Computational Science, Kobe, Japan. We collaborated on novel algorithms for scattering methods.
- University of Kansas, bioinformatics unit, USA. We have been collaborating on modeling protein-protein interactions.

7.4.3. Participation in Other International Programs

Our team has obtained the PHC Gilibert grant for a 2-year collaboration with the Vilnius University (Lithuania). Our partner is the Department of Bioinformatics, http://www.bti.vu.lt/en/departments/department-of-bioinformatics.

7.5. International Research Visitors

7.5.1. Visits of International Scientists

- Karina Dos Santos Machado, lecturer at the Federal University of Rio Grande (FURG, Brazil), Oct 2018 - Oct 2019.
- Vadim Strijov, professor at the department of Intelligent Systems, MIPT Moscow MIPT Moscow, July-August 2019.

7.5.1.1. Internships

- Khalid Mustafin (MIPT Moscow, Russia), Sep 2018 - Feb 2019.
- Ilia Igashov (MIPT Moscow, Russia), Nov 2018 - Apr 2020.

- Dmitrii Zhemchuzhnikov (UGA Grenoble), May 2019 - Sep 2019.

7.5.2. Visits to International Teams

- Sergei Grudinin visited the team of Ilia Vakser at Kansas University, Oct 15-31, 2019.

7.5.2.1. Explorer programme

- Sergei Grudinin visited Florence Tama and Osamu Miyashita, Nagoya University and RIKEN Center for Computational Science, Kobe, Japan. This was supported by the Exploration Japon 2019 program.
- Sergei Grudinin visited the team of Reidar Lund at University of Oslo, and the team of Nathalie Reuter at University of Bergen, Norway. Supported by the ÅSGARD 2019 program.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events: Organisation

8.1.1.1. Member of the Organizing Committees

- Sergei Grudinin coorganized CECAM workshop simSAS 2019, April 8-11, 90 participants, at ILL Grenoble, with Jean-Louis Barrat (UGA), Anne Martel (ILL), Sylvain Prévost (ILL), https://workshops.ill. fr/event/143/overview.

8.1.2. Scientific Events: Selection

8.1.2.1. Member of the Conference Program Committees

International Conference on Bioinformatics and Computational Biology (Devaurs)

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

- Biophysical Reviews (Devaurs)
- Current Proteomics (Devaurs)

8.1.3.2. Reviewer - Reviewing Activities

- Applied Sciences (Devaurs)
- Autonomous Robots (Devaurs)
- Bioinformatics (Oxford Press) (Devaurs, Grudinin)
- Current Proteomics (Devaurs)
- IEEE Transactions on Automation Science and Engineering (Devaurs)
- Journal of Bioinformatics and Computational Biology (Devaurs)
- Sensors (Devaurs)
- PLOS Computational Biology (Grudinin)
- Journal of Computational Chemistry (Grudinin)
- BMC Bioinformatics (Grudinin)
- Computational and Structural Biotechnology Journal (Elsevier) (Grudinin)
- Accounts of Chemical Research (ACS) (Grudinin)
- Computational Biology and Chemistry (Grudinin)
- Journal of Chemical Information and Modeling (Grudinin)
- Journal of Computer-Aided Molecular Design (Grudinin)

- Nature (Grudinin)
- Proteins: Structure, Function, and Bioinformatics (Grudinin)

8.1.4. Invited Talks

- D. Devaurs. Efficient strategies to explore the conformational space of proteins and molecular complexes. Seminar of the Functionality and Protein Engineering Unit, University of Nantes, France, 09/2019
- D. Devaurs. Efficient strategies to explore the conformational space of proteins and molecular complexes. CAPSID Seminar, Loria (Inria, CNRS, University of Lorraine), Nancy, France, 09/2019
- D. Devaurs. Efficient strategies to explore the conformational space of proteins and molecular complexes. ABS Seminar, Inria Sophia Antipolis - Méditerranée, France, 11/2019
- S. Grudinin. What does deep learning see in 3d protein structures? In 5th Korean Polish Conference on Protein Folding, September 16-18, 2019, Seoul, 2019.
- S. Grudinin. What does artificial intelligence see in 3d protein structures? In Workshop on Artificial Intelligence Applied to Photon and Neutron Science, 12-14 November, Grenoble, France, https://workshops.ill.fr/event/209/, 2019.
- S. Grudinin. Symmetry in protein complexes. In 7th CAPRI evaluation meeting, EMBL-EBI, Hinxton, UK, April 3-5, 2019, 2019.
- S. Grudinin. SAXS/SANS-assisted flexible fitting and docking. In EMBO Practical course on small angle neutron and X-ray scattering from biomolecules in solution, http://meetings.embo.org/event/19-small-angle-scattering, 2019.
- S. Grudinin. The least constraint approach for automatic coarse-graining of macromolecules. In The First International Conference on Mathematical Multiscale Modeling in Biology, Guanacaste, October 21-25, Costa Rica, 2019.
- S. Grudinin. Deep learning and artificial intelligence applied to the prediction of protein structure and interactions. In Aviesan ITMO Molecular and Structural Basis of Life Sciences : Deciphering The Functional Mechanisms Of Biological Macromolecules: Upcoming Challenges In Bioinformatics, Modelling And Experimental Validations, October 7-8, Paris, France, 2019.
- S. Grudinin. Challenges of modern structural bioinformatics. At the biological department, MSU Minsk, December 26, Minsk, Belarus, 2019.
- S. Grudinin. Novel software tools for small angle scattering. Seminar at the Synrotron SOLEIL, 13th May, 2019.
- S. Grudinin. On the analysis of macromolecular flexibility. Seminar at STFC Rutherford Appleton Laboratory, UK, March 2019.
- S. Grudinin. Novel algorithms for rapid modeling and analysis of flexibility and symmetry in macromolecules. Seminar at Cambridge Institute for Medical Research, UK, April 2019.
- S. Grudinin. Modeling and analysis of flexibility and symmetry in macromolecules. Seminar at Computational Biophysics department, Nagoya University, Japan, March 2019.
- S. Grudinin. On the nonlinear normal mode analysis. Seminar at the department of Chemistry, University of Bergen, Norway, August 2019.
- S. Grudinin. Novel computational tools for biomolecular small-angle scattering. Seminar at the department of Chemistry, University of Oslo, Norway, August 2019.
- S.Grudinin. Integrative Approaches for Current Problems in Structural Biology. Seminar at Laboratory of Theoretical Biochemistry, Institut de Biologie Physico-Chimique, Paris, 31 Jan 2019.
- S.Grudinin. New methods for protein-protein docking. Seminar at the bioinformatics unit, Kansas University, Lawrence, USA, October 2019.

8.1.5. Research Administration

- Sergei Grudinin is a member of IA working group at LINX, Lund Institute Of Advanced Neutron And X-Ray Science, Sweden.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

- Sergei Grudinin gave 2 public lecture and tutorials for Master and PhD level students on structural bioinformatics at the biology department of BSU Minsk, Belarus in December 2019.
- Sergei Grudinin prepared 2 tutorials for the EMBO school and CECAM workshop on small-angle scattering.

8.2.2. Supervision

PhD : Phd thesis defence of Guillaume Pagès, Université Grenoble Alpes, 12 septembre 2019
Title: Novel computational developments for protein structure analysis and prediction.
Thesis committee: Sergei Grudinin, Pablo Chacón, Česlovas Venclovas, Elodie Laine, Konrad Hinsen, Stéphane Redon, Arne Elofsson.
Summary: Proteins are ubiquitous for virtually all biological processes. Identifying their role helps to understand and potentially control these processes. However, even though protein sequence determination is now a routine procedure, it is often very difficult to use this information to extract relevant functional knowledge about system under study. Indeed, the function of a protein relies on a combination of its chemical and mechanical properties, which are defined by its structure. Thus, understanding, analysis and prediction of protein structure are the key challenges in molecular biology.

Prediction and analysis of individual protein folds is the central topic of this thesis. However, many proteins are organized in higher-level assemblies, which are symmetric in most of the cases, and also some proteins contain internal repetitions. In many cases, designing a fold with repetitions or designing a symmetric protein assembly is the simplest way for evolution to achieve a specific function. This is because the number of combinatorial possibilities in the interactions of designed folds reduces exponentially in the symmetric cases. This motivated us to develop specific methods for symmetric protein assemblies and also for individual proteins with internal repeats. Another motivation behind this thesis was to explore and advance the emerging deep neural network field in application to atomistic 3-dimensional (3D) data.

This thesis can be logically split into two parts. In the first part, we propose algorithms to analyse structures of protein assemblies, and more specifically putative structural symmetries. We start with a definition of a symmetry measure based on 3D Euclidean distance, and describe an algorithm to efficiently compute this measure, and to determine the axes of symmetry of protein assemblies. This algorithm is able to deal with all point groups, which include cyclic, dihedral, tetrahedral, octahedral and icosahedral symmetries, thanks to a robust heuristic that perceives correspondence between asymmetric subunits. We then extend the boundaries of the problem, and propose a method applicable to the atomistic structures without atom correspondence, internal symmetries, and repetitions in raw density maps. We tackle this problem using a deep neural network (DNN), and we propose a method that predicts the symmetry order and a 3D symmetry axis.

Then, we extend the DNN architecture to recognise folding quality of 3D protein models. We trained the DNN using as input the local geometry around each residue in a protein model represented as a density map, and we predicted the CAD-scores of these residues. The DNN was specifically conceived to be invariant with respect to the orientation of the input model. We also designed some parts of the network to automatically recognise atom properties and robustly select features. Finally, we provide an analysis of the features learned by the DNN. We show that our architecture correctly learns atomic, amino acid, and also higher-level molecular descriptors. Some of them are rather complex, but well understood from the biophysical point of view. These include atom partial charges,

atom chemical elements, properties of amino acids, protein secondary structure and atom solvent exposure. We also demonstrate that our network learns novel structural features.

This study introduces novel tools for structural biology. Some of them are already used in the community, for example, by the PDBe database and CASP assessors. It also demonstrates the power of deep learning in the representation of protein structure and shows applicability of DNNs to computational tasks that involve 3D data.

PhD : Phd thesis defence of François Rousse, Université Grenoble Alpes, 2019
Title: Incremental Algorithm for Orbital-Free Density Functional Theory.
Thesis committee: Stéphane Redon, Jean Clérouin, Reinhold Schneider, Johannes Dieterich, Philippe Blaise, Florent Calvo, Xavier Bouju.
Summary: The ability to model molecular systems on a computer has become a crucial tool for chemists. Indeed molecular simulations have helped to understand and predict properties of nanoscopic world, and during the last decades have had large impact on domains like biology, electronic or materials development. Particle simulation is a classical method of molecular dynamic. In particle simulation, molecules are split into atoms, their inter-atomic interactions are computed, and their time trajectories are derived step by step. Unfortunately, inter-atomic interactions computation costs prevent large systems to be modeled in a reasonable time. In this context, our research team looks for new accurate and efficient molecular simulation models. One of our team's focus is the search and elimination of useless calculus in dynamical simulations. Hence has been proposed a new adaptively restrained dynamical model in which the slowest particles movement is frozen, computational time is saved if the interaction calculus method do not compute again interactions between static atoms. The team also developed several interaction models that benefit from a restrained dynamical model, they often updates interactions incrementally using the previous time step results and the knowledge of which particle have moved.

In the wake of our team's work, we propose in this thesis an incremental First-principles interaction models. Precisely, we have developed an incremental Orbital-Free Density Functional Theory method that benefits from an adaptively restrained dynamical model. The new OF-DFT model keeps computation in Real-Space, so can adaptively focus computations where they are necessary. The method is first proof-tested, then we show its ability to speed up computations when a majority of particle are static and with a restrained particle dynamic model. This work is a first step toward a combination of incremental First-principle interaction models and adaptively restrained particle dynamic models.

PhD in progress : Maria Kadukova, "Novel computational approaches for protein ligand interactions", Sep 2016-, supervisors: Sergei Grudinin (France) and Vladimir Chupin (MIPT, Russia).

8.2.3. Juries

Sergei Grudinin served as an opponent at the defence of David Menéndez Hurtado's PhD thesis entitled 'Structured Learning for Structural Bioinformatics'. The defence took place at the Department of Biochemistry and Biophysics, Stockholm University, Sweden on the 11th of October.

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] F. ROUSSE. *Incremental algorithms for Orbital-Free Density Functional Theory*, Université Grenoble - Alpes, October 2019, https://hal.archives-ouvertes.fr/tel-02435781

Articles in International Peer-Reviewed Journals

[2] J. E. FAJARDO, R. SHRESTHA, N. GIL, A. BELSOM, S. CRIVELLI, C. CZAPLEWSKI, K. FIDELIS, S. GRUDININ, M. KARASIKOV, A. KARCZYŃSKA, A. KRYSHTAFOVYCH, A. LEITNER, A. LIWO, E. LUBECKA, B. MONASTYRSKYY, G. PAGÈS, J. RAPPSILBER, A. SIERADZAN, C. SIKORSKA, E. TRABJERG, A. FISER. *Assessment of chemical-crosslink-assisted protein structure modeling in CASP13*, in "Proteins - Structure, Function and Bioinformatics", 2019, forthcoming [*DOI :* 10.1002/PROT.25816], https://hal.inria.fr/hal-02315542

[3] G. FONTI, M. MARCAIDA, L. BRYAN, S. TRÄGER, A. KALANTZI, P.-Y. J. HELLEBOID, D. DEMURTAS, M. TULLY, S. GRUDININ, D. TRONO, B. FIERZ, M. DAL PERARO. *KAP1 is an antiparallel dimer with a functional asymmetry*, in "Life Science Alliance", August 2019, vol. 2, n° 4, e201900349 [*DOI :* 10.26508/LSA.201900349], https://hal.archives-ouvertes.fr/hal-02291553

[4] G. HURA, C. HODGE, D. ROSENBERG, D. GUZENKO, J. DUARTE, B. MONASTYRSKYY, S. GRUDININ, A. KRYSHTAFOVYCH, J. TAINER, K. FIDELIS, S. TSUTAKAWA. *Small angle X-ray scattering-assisted protein structure prediction in CASP13 and emergence of solution structure differences*, in "Proteins - Structure, Function and Bioinformatics", 2019, pp. 1-17, forthcoming [*DOI :* 10.1002/PROT.25827], https://hal.inria.fr/hal-02315292

[5] M. KADUKOVA, V. CHUPIN, S. GRUDININ. *Docking rigid macrocycles using Convex-PL, AutoDock Vina, and RDKit in the D3R Grand Challenge 4*, in "Journal of Computer-Aided Molecular Design", November 2019, pp. 1-10 [*DOI :* 10.1007/S10822-019-00263-3], https://hal.archives-ouvertes.fr/hal-02434514

[6] M. KARASIKOV, G. PAGÈS, S. GRUDININ. *Smooth orientation-dependent scoring function for coarse-grained protein quality assessment*, in "Bioinformatics", August 2019, vol. 35, n° 16, pp. 2801–2808 [*DOI :* 10.1093/BIOINFORMATICS/BTY1037], https://hal.inria.fr/hal-01971128

[7] M. LENSINK, G. BRYSBAERT, N. NADZIRIN, S. VELANKAR, R. A. CHALEIL, T. GERGURI, P. BATES, E. LAINE, A. CARBONE, S. GRUDININ, R. KONG, R. LIU, X. XU, H. SHI, S. CHANG, M. EISENSTEIN, A. KARCZYŃSKA, C. CZAPLEWSKI, E. LUBECKA, A. LIPSKA, P. KRUPA, M. MOZOLEWSKA, Ł. GOLON, S. SAMSONOV, A. LIWO, S. CRIVELLI, G. PAGÈS, M. KARASIKOV, M. KADUKOVA, Y. YAN, S. HUANG, M. ROSELL, L. A. RODRÍGUEZ-LUMBRERAS, M. ROMERO-DURANA, L. DÍAZ-BUENO, J. FERNANDEZ-RECIO, C. CHRISTOFFER, G. TERASHI, W. SHIN, T. ADERINWALE, S. RAGHAVENDRA MADDHURI VENKATA SUBRAM, D. KIHARA, D. KOZAKOV, S. VAJDA, K. PORTER, D. PADHORNY, I. DESTA, D. BEGLOV, M. IGNATOV, S. KOTELNIKOV, I. MOAL, D. RITCHIE, I. CHAUVOT DE BEAUCHÊNE, B. MAIGRET, M. E. R. ECHARTEA, D. BARRADAS-BAUTISTA, Z. CAO, L. CAVALLO, R. OLIVA, Y. CAO, Y. SHEN, M. BAEK, T. PARK, H. WOO, C. SEOK, M. BRAITBARD, L. BITTON, D. SCHEIDMAN-DUHOVNY, J. DAPKŪNAS, K. OLECHNOVIČ, Č. VENCLOVAS, P. J. KUNDROTAS, S. BELKIN, D. CHAKRAVARTY, V. BADAL, I. A. VAKSER, T. VREVEN, S. VANGAVETI, T. M. BORRMAN, Z. WENG, J. D. GUEST, R. GOWTHAMAN, B. G. PIERCE, X. XU, R. DUAN, L. QIU, J. HOU, B. RYAN MERIDETH, Z. MA, J. CHENG, X. ZOU, P. KOUKOS, J. ROEL-TOURIS, F. AMBROSETTI, C. GENG, J. SCHAARSCHMIDT, M. TRELLET, A. S. MELQUIOND, L. XUE, B. JIMÉNEZ-GARCÍA, C. NOORT, R. HONORATO, A. M. BONVIN, S. J. WODAK. *Blind prediction of homo- and hetero- protein complexes: The CASP13-CAPRI experiment*, in "Proteins - Structure, Function and Bioinformatics", October 2019, vol. 87, n° 12, pp. 1200-1221 [*DOI :* 10.1002/PROT.25838], https://hal.inria.fr/hal-02320974

[8] G. PAGÈS, B. CHARMETTANT, S. GRUDININ. *Protein model quality assessment using 3D oriented convolutional neural networks*, in "Bioinformatics", September 2019, vol. 35, n° 18, pp. 3313–3319 [*DOI :* 10.1093/BIOINFORMATICS/BTZ122], https://hal.inria.fr/hal-01899468

[9] G. PAGÈS, S. GRUDININ. *DeepSymmetry : Using 3D convolutional networks for identification of tandem repeats and internal symmetries in protein structures*, in "Bioinformatics", June 2019, pp. 1-24, https://arxiv.org/abs/1810.12026 , forthcoming [*DOI :* 10.1093/BIOINFORMATICS/BTZ454], https://hal.inria.fr/hal-01903624

[10] P. POPOV, S. GRUDININ, A. KURDIUK, P. BUSLAEV, S. REDON. *Controlled-advancement rigid-body optimization of nanosystems*, in "Journal of Computational Chemistry", October 2019, vol. 40, n° 27, pp. 2391-2399 [*DOI :* 10.1002/JCC.26016], https://hal.inria.fr/hal-02315276

[11] F. ROUSSE, S. REDON. *Incremental solver for Orbital-Free Density Functional Theory*, in "Journal of Computational Chemistry", September 2019, vol. 40, n° 23, pp. 2013-2027 [*DOI :* 10.1002/JCC.25854], https://hal.inria.fr/hal-02135603

Other Publications

[12] S. GRUDININ, E. LAINE, A. HOFFMANN. *Predicting protein functional motions: an old recipe with a new twist*, September 2019, working paper or preprint [*DOI :* 10.1101/703652], https://hal.archives-ouvertes.fr/hal-02291552

[13] A. H. LARSEN, Y. WANG, A. BOTTARO, S. GRUDININ, L. ARLETH, K. LINDORFF-LARSEN. *Combining molecular dynamics simulations with small-angle X-ray and neutron scattering data to study multi-domain proteins in solution*, January 2020, working paper or preprint [*DOI :* 10.1101/2019.12.26.888834], https://hal.archives-ouvertes.fr/hal-02434585

References in notes

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

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

[16] 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

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

[18] 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

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

[20] 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

[21] 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

[22] 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

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

[24] 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

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

[26] E. CANCES, F. CASTELLA, P. CHARTIER, E. FAOU, C. LE 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

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

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

[29] 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

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

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

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

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

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

[35] Z. GAIEB, S. LIU, S. GATHIAKA, M. CHIU, H. YANG, C. SHAO, V. A. FEHER, W. P. WALTERS, B. KUHN, M. G. RUDOLPH, S. K. BURLEY, M. K. GILSON, R. E. AMARO. *D3R Grand Challenge 2: blind prediction of protein–ligand poses, affinity rankings, and relative binding free energies*, in "Journal of Computer-Aided Molecular Design", 2018, vol. 32, n^{o} 1, pp. 1–20, https://doi.org/10.1007/s10822-017-0088-4

[36] Z. GAIEB, C. D. PARKS, M. CHIU, H. YANG, C. SHAO, W. P. WALTERS, M. H. LAMBERT, N. NEVINS, S. D. BEMBENEK, M. K. AMERIKS, T. MIRZADEGAN, S. K. BURLEY, R. E. AMARO, M. K. GILSON. *D3R Grand Challenge 3: blind prediction of protein–ligand poses and affinity rankings*, in "Journal of Computer-Aided Molecular Design", 2019, vol. 33, n^{o} 1, pp. 1–20, https://doi.org/10.1007/s10822-018-0180-4

[37] 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], https://doi.org/10.1163/156856207779996931

[38] 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

[39] 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

[40] 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

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

[42] A. JAHN, G. HINSELMANN, N. FECHNER, A. ZELL. *Optimal assignment methods for ligand-based virtual screening*, in "Journal of Cheminformatics", 2009, vol. 1, n^{o} 1, pp. 1–23, https://doi.org/10.1186/1758-2946-1-14

[43] 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

[44] M. KADUKOVA, S. GRUDININ. *Convex-PL: a novel knowledge-based potential for protein-ligand interactions deduced from structural databases using convex optimization*, in "Journal of Computer-Aided Molecular Design", 2017, vol. 31, n^{o} 10, https://doi.org/10.1007/s10822-017-0068-8

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

[46] 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

[47] Y. LI, M. SU, Z. LIU, J. LI, J. LIU, L. HAN, R. WANG. *Assessing protein–ligand interaction scoring functions with the CASF-2013 benchmark*, in "Nature protocols", 2018, vol. 13, n^{o} 4, 666 p.

[48] 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

[49] J. R. LOPEZ-BLANCO, P. CHACON. *KORP: Knowledge-based 6D potential for fast protein and loop modeling*, in "Bioinformatics", 2019, vol. 35, n^{o} 17, pp. 3013–3019, https://doi.org/10.1093/bioinformatics/btz026

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

[51] 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

[52] 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

[53] 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

[54] 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

[55] 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

[56] 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

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

[58] 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

[59] D. W. RITCHIE, G. J. KEMP. *Protein docking using spherical polar Fourier correlations*, in "Proteins: Structure, Function, and Bioinformatics", 2000, vol. 39, n° 2, pp. 178–194

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

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

[62] 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° 13 [*DOI :* 10.1093/BIOINFORMATICS/BTM191], http://bioinformatics.oxfordjournals.org/content/23/13/i408.abstract

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

[64] 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

[65] 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° 14, pp. 4854-4864, PMID: 16594722, http://pubs.acs.org/doi/abs/10.1021/ja058514r

[66] 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° 1, pp. 154–164, http://www.sciencedirect.com/science/article/pii/S1090780796910277

[67] M. SU, Q. YANG, Y. DU, G. FENG, Z. LIU, Y. LI, R. WANG. *Comparative assessment of scoring functions: the casf-2016 update*, in "Journal of chemical information and modeling", 2018, vol. 59, n° 2, pp. 895–913

[68] 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

[69] 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

[70] 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

[71] 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

[72] 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

[73] 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, http://www.sciencedirect.com/science/article/pii/016612809285024F

[74] 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