UNIVERSITÉ PARIS DESCARTES
MEMBRE DE
USPC Université Sorbonne Paris Cité

LABORATOIRE DE CRISTALLOGRAPHIE ET RMN BIOLOGIQUES
U M R 8 0 1 5
Faculté de pharmacie
4, avenue de l'Observatoire
75270 Paris cedex 06

Master internship proposal

Nucleic acids modeling

The study of non-coding RNA has been a very active area of research in recent years, as new functions of RNAs were reported. In order to perform their correct biological activities, RNA molecules must adopt one or more specific three-dimensional structures, and the equilibrium between the various possible configurations is regulated mainly by thermal fluctuations, pH, ionic conditions, and the presence of ligands. Despite advances in experimental methods, many questions remain open concerning the structure, the dynamics and thermodynamics of these molecules. As a first step, understanding how non-coding RNA works implies being able to determine its tertiary structure and understanding how the molecule adapts to the changes of the environment. Computer studies can complement the experimental information, but they are limited by the size of the systems and by the time-scales characteristic of the processes. In recent years, new coarse-grained models allow to simulate these molecules.
Among them we have developed HiRE-RNA/DNA to study RNA folding and DNA assembly. With this model we have been able predict folding of RNA molecules of complex architectures up to sizes of 50 nucleotides from sequence only, and of structures of larger systems when coupled to limited or low-resolution experimental data. The unique feature of the model is to account for a large set of non-canonical and multiple base pairs (triplets and quadruplets), allowing to study systems as complex as G-quadruplexes.
According to the applicant's interests and background the internship can focus on one of the following axes:

1. Introducing explicit ions in the coarse-grained model HiRE-RNA
RNA molecules are highly charged and the presence of ions is a crucial element in the folding process and to stabilize the overall architecture. Our RNA coarse-grained model implicitly takes into account the presence of ions at large distances through a screened electrostatic potential, but does not account explicitly for structural ions. To move toward a more realistic description of RNA molecules, we need to develop an RNA model including a finer description of electrostatic interactions.
The aim of this internship is to set the ground for introducing explicit ions in our coarse-grained model to go toward a model that can describe DNA condensation, structural ions of RNA folds, and eventually the response of the system to changes in pH. The work will be composed of two phases: definition and parameterization of a force field for ions only (in the framework of HiRE-RNA) reproducing well known properties of ionic solutions, benchmarking of folding of small RNA molecules and DNA double helical assembly with explicit ions. This work will be done also in collaboration with Pr. Fernando Barroso da Silva at University of Sao Paolo, providing expertise on simulations of ionic systems and of proteins and RNA responses to pH variations.

2. Integrating HiRE-RNA with the coarse-grained protein model OPEP.

The coarse-grained OPEP force field for protein is the "twin" model of HiRE-RNA. It has proven very successful in predicting correct structures for amyloid and non-amyloid. Both force fields are coupled to molecular dynamics, simulated tempering, and Monte Carlo simulations. Up to now the two models have been developed independently from one another, allowing to perform simulations either on proteins alone, or on nucleic acids alone, but not of the two systems together.

For RNA, we have also recently developed UnityMol+HiRE-RNA, a program based on a game engine, to interactively fold RNA molecules (http://www.baaden.ibpc.fr/umol/ ; https://hirerna.galaxy.ibpc.fr/).

In this internship we propose to take the first steps in the direction of a unified force-field for proteins/nucleic acids complexes developing a simple interface accounting for occupied volume, electrostatics and Van der Waals interactions. This first implementation will be imported into UnityMol+HiRE-RNA to allow for interactive simulations including both nucleic acids and proteins. This work will be done in collaboration with Pr. Philippe Derreumaux, principal developer of OPEP, and Dr. Marc Baaden, principal developer of UnityMol, both at Laboratoire de Biochimie Théorique in Paris.

3. Application of HiRE-RNA to the study of triple helices

Nucleic bases can form pairs other than the canonical G-C and A-T or A-U pairs. Following the description introduced by E. Westhof, bases can form pairs on each of their three sides: Watson-Crick, Hoogsteen and Sugar. These non-canonical and multiple pairings are crucial in the formation of single stranded RNA structures and of multiple strand structures for DNA or DNA/RNA mixed systems.

Double stranded DNA in the B-form helix can form triple helices with the insertion of an additional DNA or RNA strand in its major groove through Hoogsteen pairs. These structures play important roles in vivo and are inherently prones to mutations and recombination, making the triple helix a possible interesting pharmacological target. HiRE-RNA is at present the only coarse-grained RNA model including a description for triplets and quadruplets, and it is therefore in a unique position to study the properties of nucleic acids triple helices.

In this internship we propose to determine a proper parametrization of the coarse-grained model to study the DNA triple helix and to investigate the dynamical and thermodynamical properties of its formation as well as the role of specific sequences. The work will be done also in collaboration with Pr. Alessandra Villa at Karoninska Institut in Stockholm, studying these same systems in silico at atomistic resolution.

4. Coarse-grained simulations of G-quadruplexes

HiRE-RNA is currently the only coarse-grained NA representation that accounts for a large set of non-canonical pairings with the possibility of forming multiple base interactions, which allowed the study of G-quadruplexes, observing, for the first time in a simulation, a complete folding event.

Coarse-grained simulations were able to give access to partially folded intermediates and competing folded states that are out of reach for atomistic simulations and helped reveal the complexity of the landscape of G-quadruplexes folding. In collaboration with J. Sponer at the Czech Academy of Science, coarse-grained simulations were coupled with atomistic studies to understand the details of the alternative folds and metastable states and the dynamic transitions between them.

Even with coarse-grained simulations, complete unfolding/folding events are very rare. We will pursue the study of the possible alternative stable states of G-quadruplexes through an optimization of the coarse-grained force field in order to be able to statistically reproduce complete folding events.

5. Atomistic structural analysis of the non-coding RNA 7SK

The ncRNA 7SK was an early discovery in the field of ncRNA and stands as a paradigm of RNA regulators. It plays a major role in activating transcription by the RNA-polymerase and in the maturation of messenger RNA. Its function is directly related to several human diseases, like in cancers such as leukemia. In recent years, different experimental techniques unveiled structural aspects of parts of 7SK obtained from crystallography, NMR, exposed surface probing SHAPE, SAXS, Cryo-EM and

biochemistry techniques. Taken together, 7SK appears to function as a platform for its various partners, with possibly varying secondary structure arrangements to allow for interactions with different proteins. of the structure of the non-coding RNA 7SK.
In a continuing collaboration between modeling and experiments, which already led to the submission of a article on the structure of HP1 and of three PDB structures on the PDB repository, we will to continue investigating the structure of 7SK both in terms of stability of experimentally resolved sub-parts, and of ab initio predictions for the other regions for which some experimental information on secondary structures or on the global shape is available. This project is based on atomistic simulation.

The work will be supervised by Pr. Samuela Pasquali at the Laboratoire de Cristallographie et RMN Biologiques (LCRB), UMR 8015 CNRS, Université Paris Descartes. SP has newly joined the lab and is in the process of starting her own modeling team in close contacts with the experimental teams in the lab.
The project will benefit from the environment at LCRB, where different biophysical techniques are used to study biomolecules and that will provide benchmark systems for modeling, and of the strong collaboration with the Laboratoire de Biochimie Theorique (UMR 9080 CNRS).
The student will receive an allowance of 550 euros/month to cover local expenses. Duration of the internship will depend on the applicant's master program requirements.

We are looking for a candidate holding an undergraduate degree in chemistry, physics, or bioinformatics, with good programming skills, and good analytical understanding of force fields.
If interested, please send your CV including a copy of your undergraduate records to Samuela Pasquali: samuela.pasquali@parisdescartes.fr

References

S. Pasquali, P. Derreumaux, « HiRE-RNA: a high resolution coarse-grained energy model for RNA », J Phys Chem B., 114, 11957-11966 (2010)

T. Cragnolini, P. Derreumaux, S. Pasquali, « Coarse-grained simulations of RNA and DNA duplexes », J Phys Chem B, 117, 8047-8060 (2013)

F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J-F. Saint-Pierre, M. Kalimeri, A. Barducci, Y. Laurin, A. Tek, M. Baaden, P.H. Nguyen, P. Derreumaux, « The OPEP coarse-grained protein model: from single molecules, amyloid formation, role of macromolecular crowding and hydrodynamics to RNA/DNA complexes », Chem Soc Reviews, 43, 4871-4893 (2014)

T. Cragnolini, Derreumaux, S. Pasquali, Ab initio RNA folding, Journal of Physics: Condensed matter, 2015, 23, 233102 T. Cragnolini, Y. Laurin, P. Derreumaux, S. Pasquali (2015), « Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base pairings », JCTC, 11, 3510-5322

F. L. Barroso da Silva, S. Pasquali, P. Derreumaux, L. Gustavo, « Electrostatic analysis of the mulational and pH effects of the N-terminal domain self-association of the Major Ampullate Spidroin »,, *Soft Matter*, **12**, 5600 (2016)

A. Azevedo Reis Teixeira, M. Lund, F. Barroso da Silva, « Fast proton titration scheme for multiscale modeling of protein solutions », *J. Chem. Theory Comput.*, **2010**, *6* (10), pp 3259–3266

F. Sterpone, S. Melchionna, P. Tuffery, S. Pasquali, N. Mousseau, T. Cragnolini, Y. Chebaro, J-F. Saint-Pierre, M. Kalimeri, A. Barducci, Y. Laurin, A. Tek, M. Baaden, P.H. Nguyen, P. Derreumaux, « The OPEP coarse-grained protein model: from single molecules, amyloid formation, role of macromolecular crowding and hydrodynamics to RNA/DNA complexes », Chem Soc Reviews, 43, 4871-4893 (2014)

S. Doutreligne, C. Gageat, T. Cragnolni, A. Taly, S. Pasquali, P. Derrumaux, M. Baaden, « UnityMol : interactive and ludic visual manipulation of coarse-grained RNA and other biomolecules », Virtual and Augmented Reality for Molecular Science (VARMS@IEEEVR), 2015

P. Stadlbauer, L. Mazzanti, T. Cragnolini, D. J. Wales, P. Derreumaux, **S. Pasquali**, J. Sponer, «Coarse-Grained Simulations Complemented by Atomistic Molecular Dynamics Provide New Insights into Folding of Human Telomeric G-Quadruplexes», *J. Chem. Theory Comput.*, in press. DOI: 10.1021/acs.jctc.6b0066