**Benefits package**
- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- French courses

---

**Generale Information**

**Theme/Domain:** Data and Knowledge
- Representation and Processing
- Statistics (Big data) (BAP E)
- Inria Center: CRII Nancy - Grand Est

**Starting date:** 2018-09-01
**Duration of contract:** 1 year, 6 months
**Deadline to apply:** 2018-06-30

**Contacts**

- **Inria Team:** ORPAILLEUR (DGD-S)
- **Recruiter:** Napoli Amedeo / amedeo.napoli@inria.fr

**About Inria**

Inria, the French National Institute for computer science and applied mathematics, promotes "scientific excellence for technology transfer and society". Graduates from the world's top universities, Inria’s 2700 employees rise to the challenges of digital sciences. With its open, agile model, Inria is able to explore original approaches with its partners in industry and academia and provide an efficient response to the multidisciplinary and application challenges of the digital transformation. Inria is the source of many innovations that add value and create jobs.

**Conditions for application**

**Defence Security:**
This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorization to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

**Recruitment Policy**:
As part of its diversity policy, all Inria positions are accessible to people with disabilities.

---

**Warning:** you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

---

**Context**

**Scientific Context**

The exploratory research project HyGraMi is defined in the contest of the fight against the resistance of bacteria to antibiotics. The consequences of this resistance are multiple, with a management of patients becoming more and more difficult and costly, in humans and financial terms. Accordingly, the objective of the HyGraMi project is to design hybrid data mining strategies for discovering new antibacterial agents for the future. These strategies will rely on a combination of numeric and symbolic classifiers, where the first include SVM, Random Forests, neural networks, while the second include Bayesian networks, clustering, graph mining and Formal Concept Analysis. Moreover, the classifiers can also be guided by domain knowledge and be complemented with specialized systems developed by biologists and chemists.

**Objectives**

The design of new drugs should take into account large volumes of heterogeneous data in an efficient way (molecular structures, genomes, documents...) and needs a adapted data mining process. Following a supervised learning paradigm, there is a need for specialized databases of molecules known for being antibacterial agents (training set); databases of candidate molecules (test set); and as well knowledge bases about the resistance of bacteria to antibiotics from a biological and chemical points of view. Such databases should be integrated for being deeply analyzed and for discovering relevant information on antibacterial candidates. Active molecules include substructures holding specific functionalities. Such substructures should be detected in the collection of available molecules and will qualify potential antibacterial candidates.

**Assignment**

In this research work, we intend to combine numeric and symbolic classifiers for carrying the mining of molecular graphs w.r.t. domain knowledge. One approach could be the following. The structures of active antibacterial molecules can be analyzed thanks to graph mining techniques. Then numerical classifiers select potential antibacterial candidates in a supervised way, based on exact or approximate matching. These candidates hold substructures related to known antibacterials. In turn, symbolic classifiers can be used for ranking candidates w.r.t. a set of domain constraints and preferences, or to some specific biologic and chemical objectives (non supervised classification). Such strategies where numeric and symbolic data mining operations are intertwined will be thoroughly studied in the HyGraMi research project.

The postdoctoral fellow will work with domain specialists and will benefit from past experiences in the domain. Researchers in the team have already worked on the design of molecule databases and on the combination of classifiers. In addition, the Orpailleur team collaborates with experts in antibacterials, who have built the ABC database and the ABCMune platform aimed at guiding the search for new antibacterials. The ABC platform includes a collection of more than 4,000 bacteria, which thus covers a large part of the bacteria responsible for human diseases, and as well the associated description of the resistance to antibiotics phenomena.

**Main activities**

**Main activities**
- Study and analysis of state of the art literature
- Design and adaptation of new algorithms
- Design of the associated computer programs
- Writing of articles

**Skills**

The postdoctoral fellow will work in the Orpailleur Team, which is a research team at Inria Nancy Grand Est Research Center and at LORIA laboratory. He will work within the scientific environment of Inria and LORIA, and will be able to reuse proper software and hardware related to this research project.

Ideally, the postdoctoral fellow should have a thesis in computer science, or applied mathematics, or bioinformatics. Moreover, the fellow should have a good knowledge of machine learning and data mining techniques. Additional knowledge in biology and/or chemistry will be very appreciated.
Remuneration
Salary: 2653€ gross/month