



Jean-Philippe Vert  
Chef d'équipe  
jean-philippe.vert@curie.fr  
Tél : +33 1 56 24 69 17

De nombreuses technologies nouvelles permettent d'observer les organismes vivant à une échelle jusque-là inexplorée. Par exemple, les techniques de biopuces ou de séquençage massivement parallèle permettent de lire l'information génétique complète d'un échantillon biologique, la spectrométrie de masse permet de quantifier le protéome, la vidéo-microscopie à haute résolution permet de suivre les évolutions de populations de cellules, et les méthodes de criblage à haut débit sont utilisées pour caractériser les effets biologiques de grandes quantités de molécules.

**Toutes ces technologies génèrent de très grandes quantités de données, qu'il est souvent difficile d'interpréter directement. Afin de mieux exploiter ces masses de données et d'en extraire des informations biologiques et médicales utiles, notre équipe développe des méthodes mathématiques et des algorithmes innovants, basés sur notre expertise en modélisation mathématique, statistique, machine learning, bioimage informatics et biologie structurale.**

Nous développons de nouveaux outils pour examiner différentes questions biologiques et médicales, notamment:

- Biologie et biologie des systèmes *in silico*. Nous développons des approches innovantes pour la reconstruction statistique de réseaux biologiques, l'analyse et la modélisation de la progression tumorale au niveau génomique, transcriptomique et épigénétique, l'annotation automatique de gènes et d'éléments fonctionnels par intégration de données hétérogènes et complexes, en particulier les données provenant de séquençage à haut débit ou de vidéo-microscopie.
- Médecine prédictive et de pré Nous développons des méthodes pour la classification de

# Apprentissage statistique et modélisation des systèmes biologiques

## U900 - Cancer et génome : bioinformatique, biostatistiques et épidémiologie

tumeurs et l'identification de biomarqueurs diagnostiques, pronostiques et prédictifs de réponse à un traitement. Ces classifications s'appuient sur un grand nombre de paramètres mesurables comme les données cliniques du patient, l'ensemble des mutations somatiques de la tumeur, l'expression de gènes et de transcrits alternatifs, ou les modifications structurales de l'ADN, et fait appel à des méthodes d'apprentissage statistique en grande dimension.

- Développement de médicaments : nous développons de nouvelles méthodes pour le criblage virtuel et la chémogénomique *in silico*, afin d'aider à identifier ou créer de petites molécules capables de se lier spécifiquement à des cibles thérapeutiques par modélisation informatique. Ces méthodes combinent des modélisations 3D des molécules et des procédures statistiques innovantes, et peuvent par exemple aider à prédire les profils d'activité ou de toxicité de candidats médicaments.

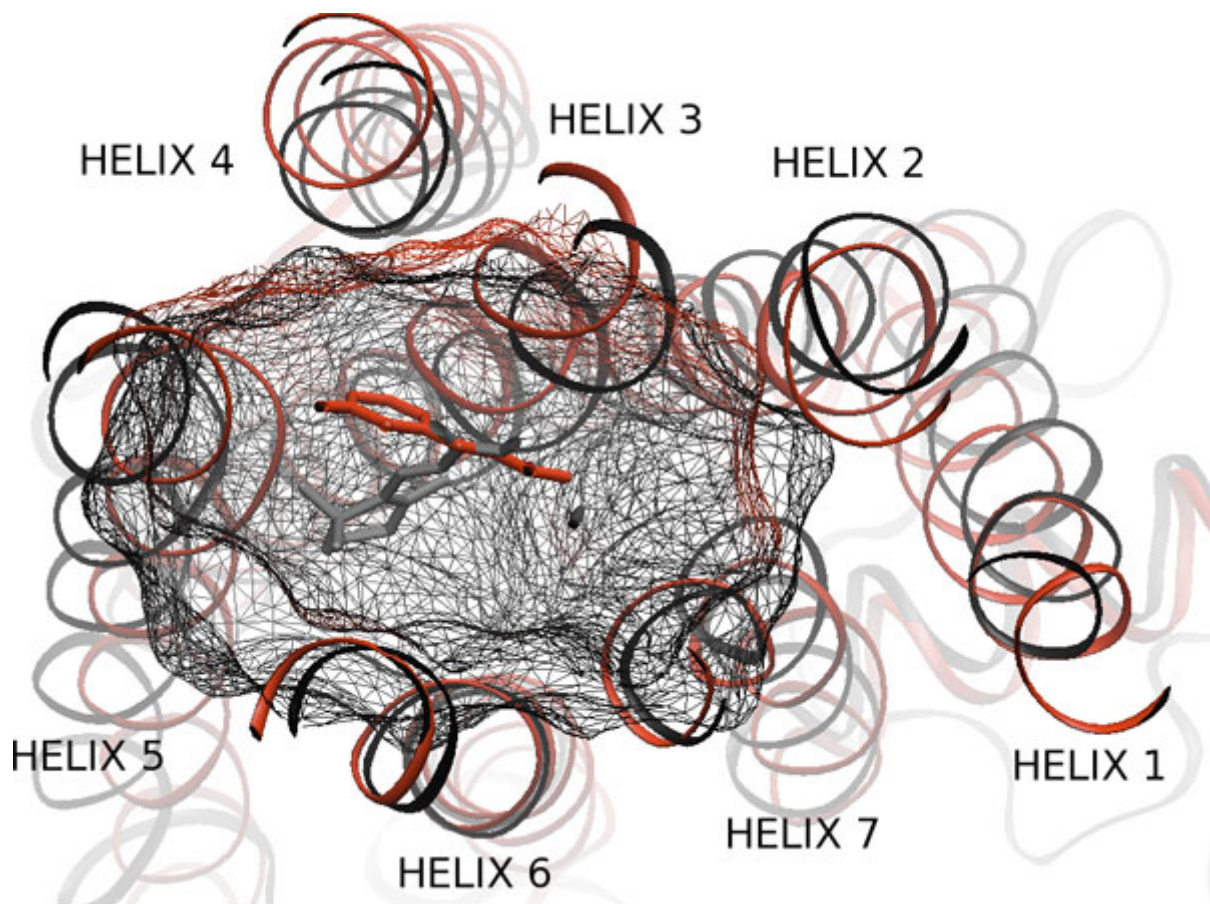


Figure 1 : Cette image montre la surface 3D de l'interaction entre une protéine membranaire et une petite molécule. La modélisation et l'étude statistique de ces interfaces nous permettent de prédire quelles molécules sont susceptibles d'interagir avec des protéines d'intérêt thérapeutique, notamment des récepteurs membranaires, de manière à proposer de nouvelles pistes dans la recherche de médicaments.

## Publications clés

### Année de publication : 2018

---

Tom Baladi, Jessy Aziz, Florent Dufour, Valentina Abet, Véronique Stoven, François Radvanyi, Florent Poyer, Ting-Di Wu, Jean-Luc Guerquin-Kern, Isabelle Bernard-Pierrot, Sergio Marco Garrido, Sandrine Piguel (2018 Nov 1)

**Design, synthesis, biological evaluation and cellular imaging of imidazo[4,5-b]pyridine derivatives as potent and selective TAM inhibitors.**

*Bioorganic & medicinal chemistry* : 26 : 5510-5530 : [DOI : 10.1016/j.bmc.2018.09.031](https://doi.org/10.1016/j.bmc.2018.09.031)

### Année de publication : 2015

---

Servant N., Varoquaux N., Lajoie B.R., Viara E., Chen C.J., Vert J.P., Heard E., Dekker J., Barillot E. (2015 Jan 1)

**HiC-Pro: an optimized and flexible pipeline for Hi-C data processing**

*Genome biology* : 16 : 1

### Année de publication : 2014

---

Elsa Bernard, Laurent Jacob, Julien Mairal, Jean-Philippe Vert (2014 May 9)

**Efficient RNA isoform identification and quantification from RNA-Seq data with network flows.**

*Bioinformatics (Oxford, England)* : 2447-55 : [DOI : 10.1093/bioinformatics/btu317](https://doi.org/10.1093/bioinformatics/btu317)

Ferhat Ay, Evelien M Bunnik, Nelle Varoquaux, Sebastiaan M Bol, Jacques Prudhomme, Jean-Philippe Vert, William Stafford Noble, Karine G Le Roch (2014 Mar 26)

**Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression.**

*Genome research* : 974-88 : [DOI : 10.1101/gr.169417.113](https://doi.org/10.1101/gr.169417.113)

### Année de publication : 2013

---

Veronika Graml, Xenia Studera, Jonathan L D Lawson, Anatole Chessel, Marco Geymonat, Miriam Bortfeld-Miller, Thomas Walter, Laura Wagstaff, Eugenia Piddini, Rafael E Carazo-Salas (2013 Nov 2)

**A genomic Multiprocess survey of machineries that control and link cell shape, microtubule organization, and cell-cycle progression.**

*Developmental cell* : 227-39 : [DOI : 10.1016/j.devcel.2014.09.005](https://doi.org/10.1016/j.devcel.2014.09.005)



# Apprentissage statistique et modélisation des systèmes biologiques

## **U900 - Cancer et génome : bioinformatique, biostatistiques et épidémiologie**

James C Costello, Laura M Heiser, Elisabeth Georgii, Mehmet Gönen, Michael P Menden, Nicholas J Wang, Mukesh Bansal, Muhammad Ammad-ud-din, Petteri Hintsanen, Suleiman A Khan, John-Patrick Mpindi, Olli Kallioniemi, Antti Honkela, Tero Aittokallio, Krister Wennerberg, , James J Collins, Dan Gallahan, Dinah Singer, Julio Saez-Rodriguez, Samuel Kaski, Joe W Gray, Gustavo Stolovitzky (2013 Jul 20)

### **A community effort to assess and improve drug sensitivity prediction algorithms.**

*Nature biotechnology* : 1202-12 : [DOI : 10.1038/nbt.2877](https://doi.org/10.1038/nbt.2877)