Thomas COKELAER
Systems Biology - Computational Science - Data Analysis - Software Engineering

born July 13, 1975, FRANCE
cokelaer@ebi.ac.uk
cokelaer@gmail.com
http://thomas-cokelaer.info

Current Research: Modelling in systems biology.

- My current research focuses on building predictive models of protein networks to understand deregulation of signalling pathways, and using machine learning and data mining to solve biological problems.

- I am also interested in the development of scientific-computing tools and best practices in software-engineering for computational sciences.

- I gained experience on big data in a variety of research projects spanning astronomy, plant science and biology.

EDUCATION

PhD
1999–2003
Sciences, gravitational waves. Université de Nice Côte d'Azur, France. **Subject:** *Detection of gravitational waves emitted by black hole-black hole coalescences.* Collaboration Thalès Marconi Sonar, Sophia Antipolis, France

MSc
1998–1999
Astronomy, Imaging – Université de Nice Sophia-Antipolis, France.

BSc
1996–1998
Physics – Université du Littoral, Dunkerque, France.

BSc
1995–1996
Mathematics and computer science – Université du Littoral, Calais, France.

SYSTEM BIOLOGY
Aug 2011—Present

Position

EMBL-EBI
Aug. 2011–Present
Bioinformatician/Research Staff at the *European Bioinformatics Institute* (EMBL-EBI), Cambridge, United Kingdom. Member of the BioMedecine group (leader: Julio Saez-Rodriguez).

Research

Data analysis and modelling in systems biology using or creating bioinformatics software

Phospho-proteomic (CellNOptR)
Developed and maintained **CellNOptR** [8, 9] a software dedicated to the logic modelling of networks of proteins (www.cellnopt.org). Implementation of various formalisms including boolean logic (synchronous/asynchronous). Analysis of phospho-proteomic data sets obtained under different perturbations (e.g., Luminex data). The models obtained can be mechanistic (understand deregulation) and predictive (novel therapies). Analysed colorectal cancer data sets (human).

Mass-spectrometry
Data analysis related to mass-spectrometry. Yeast data sets analysed to understand cross-talks in signalling pathways [1]. Pipeline developed to provide quality control and descriptive analysis.

DREAM Challenges
Member of the DREAM consortium (Dialogue for Reverse Engineering Assessments and Methods) [2]. DREAM proposes challenges in systems biology. I participated to the organisation of several challenges such as the parameter estimation of gene regulatory networks (DREAM6-7) [4], and the HPN-DREAM Breast cancer network inference challenge. Implementation of scoring functions and statistical tests used to compare participants.

BioServices
Author of BioServices [6], a software that eases programmatic access to over 30 web services related to life sciences (e.g., KEGG, UniProt, Ensembl).

Optimisation
Developed and used optimisation tools dedicated to logic modelling: genetic algorithms [8], heuristics [3], Monte Carlo methods (MCMC) or optimal methods [7].

Others
Sequence analysis to validate identifiers and peptide sequences from mass spectrometry data sets. Automatic identification of phosphorylated amino acids in peptide sequence. Sequence alignments. Manipulation of biological networks in various formats (SIF, SBMLqual [5]).

Teaching

BioPreDyn Tutorial given at a research level (BioPreDyn European grant) on logic modelling and programmatic access to web services using BioServices [6]- See Tutorials available on-line.

CoLoMoTo Logic modelling tutorial and lecture given to CoLoMoTo consortium (Available on line).

Python Series of Python lectures for beginners. Organised on the Wellcome Trust Campus, U.K.

Mentoring Mentor of a Masters student (6-months–2013, optimisation problems using MCMC methods) and PhD (3-months–2015, analysis of multi-dimensionnal phosphorylation data sets).

Working groups

DREAM Member of the DREAM community (organisation of challenges).

CoLoMoTo Participation to the CoLoMoTo consortium (logic modelling).

Python Organisation of a Python working group on the EMBL-EBI/ Wellcome Trust Campus.

Publications

Full list available on line http://thomas-cokelaer.info.

Peer-reviewed journals

[1] S. Vaga, M. Bernardo-Faura, **T. Cokelaer,** J. Saez-Rodriguez, R. Aebersold *Phosphoproteomic analyses reveal novel cross-modulation mechanisms between two signaling pathways in yeast* Molecular Systems Biology (2014) 10: 767

[2] Costello JC et al. *A community effort to assess and improve drug sensitivity prediction algorithms* Nature Biotechnology 2014

[3] J. A Egea, D. Henriques, **T. Cokelaer**, A. F Villaverde, A. MacNamara, D.P. Danciu, J. R Banga, J.Saez-Rodriguez. *MEIGO: an open-source software suite based on metaheuristics for global optimization in systems biology and bioinformatics* BMC Bioinformatics 2014, **15**:136

[4] Meyer P. et al. *Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach BMC Systems* Biology vol. 2014 **8** (1) p. 13

[5] Chaouiya et al. *SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools* BMC Systems Biology 2013, **7**:135

[6] **T. Cokelaer** et al. *BioServices: a common Python package to access biological Web Services programmatically* Bioinformatics **29** (24) 3241-3242 (2013)

[7] C. Guziolowski et al. *Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming, 2013*

[8] Terfve et al. *CellNOptR: a flexible toolkit to train protein signaling networks to data using multiple formalisms,* BMC System Biology, 6, 1, 133 (2012)

Peer-reviewed conference proceedings

[9] **Cokelaer T.** and Saez-Rodriguez J. *Using Python to Dive into Signalling Data with CellNOpt and BioServices* Proceedings of the 7th European Conference on Python in Science (EuroSciPy 2014). http://arxiv.org/abs/1412.6386

PLANT MODELLING
Nov 2008 — Dec.2010

Position

INRIA
Nov. 2008–Dec.2010 Computer scientist/software developer at INRIA, (French National Institute of Research in Computer Science) in the Virtual Plants team, Montpellier, France.

Research

OpenAlea
- Software development within **OpenAlea**, a plant modelling platform [11], which allows the creation of biological models using a visual programming environment (similar to Galaxy). Creation of scientific workflows to re-use functional model of plants [13].

- Connection between OpenAlea and scientific tools from the Python community such as Matplotlib (visualisation) or SciPy (numerical analysis).

- Professional usage of tools dedicated to software development (test, documentation, code quality, bug tracking, ...). Managed Python and C++ libraries over 300,000 lines of code.

Modelling Used generative grammar (L-systems) to model plant growth [10]. Modelling of trees using mechanistic and statistical approaches [12]. Used statistical analysis and graph theory.

Publications

Full list available on line http://thomas-cokelaer.info.

Peer-reviewed journals

[10] Boudon, F., Pradal, C., **Cokelaer, T.**, Prusinkiewicz, P., Godin, C. – *L-py: an L-system simulation framework for modeling plant architecture development based on a dynamic language.* – Frontiers in plant science, **3** p76 (2012)

Peer-reviewed conference proceedings

[11] Pradal C., Barbeau D., **Cokelaer T.**, Moscardi E.– *VisuAlea, Towards a Scientific Modelling Environment using Visual Programming* – EuroSciPy 2010.

[12] Han, L. et al. – *Investigating the Influence of Geometrical Traits on Light Interception Efficiency of Apple Trees: a Modelling Study with MAppleT* – International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications, IEEE, pages 152-159, (2012)

[13] Chopard J., Pradal C., Barbeau D., **Cokelaer T.**, Godin C. – *Scientific workflow for reusing plant/FSPM models* – MODSIM2011. 19th International Congress on Modelling and Simulation (2011) 968-974

COMPUTER SCIENCE

Computing Skills

- Experienced developer, specialised in scientific computing, signal processing and visualisation.
- Experience in software management (design and project management).

Python Expert knowledge of Python as well as scientific libraries such as Matplotlib (visualisation), NumPy, NetworkX (graph theory), Pandas (data analysis), SciKit-learn (machine learning).

C/C++ Developed C/C++ libraries combined with Python or R languages.

Other languages R: maintainer of 5 libraries on BioConductor website. MATLAB/octave, Perl, sed/awk.

Web Knowledge of HTML/CSS/Javascript. Web site examples: home page, www.cellnopt.org.

Cluster Skills Deployment of code and pipelines on clusters based on Condor or LSF technologies.

Others Knowledge of Linux systems (administration).

Software contributions

github List of personal software on https://github.com/cokelaer.

BioServices
2013–Present Author of BioServices [6], that provides a programmatic access in Python to over 30 web services (e.g., UniProt, Ensembl). Documentation, tutorials and *iPython notebooks available online.*

CellNOptR
2011–Present Maintainer and developer of CellNOptR (R language) and its add-ons (4 other packages) that are updated on BioConductor web site.

OpenAlea
2009–2011 Participated to the development of OpenAlea (Python, PyQt) that contains graphical interface and over 150,000 lines of code.

DreamTools
2015–Present Created Common library used to score the DREAM challenges. Available on DreamTools.

Spectrum
2011–present Author of Spectrum, spectral analysis library written in Python. Can be used to apply Fourier transform or parametric methods (ARMA, BURG) to create spectral estimation of time-series.

PHYSICS – GRAVITATIONAL WAVES
April 2003– July 2008

Positions

Research Associate/Assistant
April 2003–April 2008 Cardiff University, Physics and Astronomy Department. Gravitational Physics group (Prof B.S. Sathyaprakash).

PhD
Sept. 1999–Jan. 2003 Observatoire de la Côte d'Azur, France Topic: *Detection of gravitational waves emitted by black hole-black hole coalescences.*

Internship
Mars–Sept. 1999 Research project (6 months) on detection of spacecrafts (image processing, statistics).

Research

Detection theory During my PhD, I developed tools to make possible the detection of small signal-to-noise ratio signals using detection theory, parameter estimation and spectral analysis. I participated to the implementation of data analysis pipelines for the VIRGO collaboration [21].

Collaborations
Matched filtering While at Cardiff University, I joined the GEO600 [23] and LIGO collaborations 17 (600 members). I was involved in the data analysis working group dedicated to the detection of gravitational waves using matched filtering and signal processing techniques [14, 15, 16, 18, 19, 20, 22].

Image processing Time-frequency analysis (burst detection in 2D plane), contour detection, segmentation.

Software Collaborative work with 50 members working on a common code (C/C++/Python).

Others

Teaching University level: computer science at BSc level (20 hours).

Mentoring Mentoring: physics and mathematics at BSc level (20 hours).

Publications

Full list available on line http://thomas-cokelaer.info.

Peer-reviewed journals

[14] Babak et al. *Searching for gravitational waves from binary coalescence* – Phys. Rev. D **87**, 024033 (2013)

[15] C. Van Den Broeck, D. A. Brown, **T. Cokelaer**, I. Harry, G. Jones, B.S. Sathyaprakash, H. Tagoshi, H. Takahashi – *Template Banks to Search for Compact Binaries with Spinning Components in Gravitational Wave Data* – Phys. Rev. D **80**, 024009 (2009)

[16] **T. Cokelaer** and D. Pathak *Searching for Gravitational-Wave Signals Emitted by Eccentric Compact Binaries Using a non-Eccentric Template Bank: Implications for Ground-Based Detectors* – Class. Quant. Grav. **26**, 045013 (2009)

[17] The LIGO Scientific Collaboration & The Virgo Collaboration – *An Upper Limit on the Stochastic Gravitational-Wave Background of Cosmological Origin* – Nature **460**, 990-994 (2009)

[18] B. Abbott *et al.* [**Corresponding author** for the LIGO Scientific Collaboration] – *Search for Gravitational Waves from Binary Inspirals in S3 and S4 LIGO data* – Phys. Rev. D **77**, 062002 (2008)

[19] **T. Cokelaer** *Gravitational Waves from Inspiralling Compact Binaries: Hexagonal Template Placement and its Efficiency in Detecting Physical Signals* – Phys. Rev. D. **76** 102004 (2007)

[20] S. Babak, R. Balasubramanian, D. Churches, **T. Cokelaer** and B. S. Sathyaprakash – *A Template Bank to search for Gravitational Waves from Inspiralling Compact Binaries. I: Physical models* – Class. Quant. Grav. **23**, 5477 (2006)

[21] F. Acernese *et al.* Virgo Collaboration *The commissioning of the central interferometer of the Virgo Gravitational Wave detector* – Astropart. Phys. **21**, 1 (2004).

Peer-reviewed conference proceedings

[22] **T. Cokelaer** – *Parameter Estimation of Inspiralling Compact Binaries in Ground-Based Detectors: Comparison Between Monte Carlo Simulations and the Fisher Information Matrix* – Class. Quant. Grav. **25**, 184007 (2008).

[23] H. Luck *et al.* – *Status of the GEO600 Detector* – 6th Edoardo Amaldi Conference on Gravitational Waves (Amaldi6), Kise Nago, Okinawa, Japan, 20-24 Jun 2005. – Class. Quant. Grav. **23**, S71 (2006).

CONTACT

Thomas Cokelaer
BioMedecine Team, EMBL-EBI
Cambridge, United Kingdom

email:cokelaer@gmail.com
email:cokelaer@ebi.ac.uk
☎ on request