

Thomas COKELAER

Data Scientist - Computational Science -
System Biology - Software Engineering

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

SUMMARY of Qualifications

Interdisciplinary researcher (physics, astrophysics, system biology, bioinformatics) Experienced developer specialised in data mining, signal processing and interpretation of big data sets. Quick to assimilate new ideas, concepts and cutting-edge technologies whilst demonstrating a logical and analytical approach to solving complex problems and issues. Valuable member of collaborative working groups who encourages communication and sharing knowledge amongst colleagues.

WORKING EXPERIENCES

EMBL-EBI Aug.2011–Present

Research Staff / Bioinformatician at the *European Bioinformatics Institute* (EMBL-EBI), Cambridge, United Kingdom. Member of the *BioMedecine* group.

phospho-proteomics, logic modelling

Developping and maintaining *CellNOptR* software (R language). Used graph theory to model protein signalling networks trained to data using boolean steady-states formalism (or fuzzy or differential equations). Used Pandas for multi-dimensional data analysis and visualisation.

mass-spectrometry, data mining

Analysis of mass-spectrometry data sets (yeast) to understand deregulation of signalling pathways. Pipelines and software developped to provide quality control and statistical analysis. Designed tools to perform database-like requests and annotations on peptide sequences.

DREAM community, statistics

Member of the *DREAM* consortium. Participated to the organisation of several challenges such as the *HPN-DREAM Breast cancer network inference* challenge. Developped statistical analysis to score and rank participants using scikit-learn software (machine learning).

Web Services

Author of *BioServices*, a software that ease programmatic access in Python to more than 30 web services (REST or SOAP) related to life sciences (e.g., KEGG, UniProt, Ensembl ...).

Optimisation

Implementation of optimisation tools dedicated to logic modelling: from genetic algorithms and heuristics to Monte Carlo methods (MCMC).

Bioinformatics

Sequence analysis to validate identifiers and peptide sequence from mass spectrometry data sets, automatic identification of phosphorylated amino acid in sequences, sequence alignment, manipulation of biological networks in various formats (SIF, SBMLqual, ...).

INRIA

Nov. 2008–Dec.2010

Computer Scientist at *INRIA*, the French National Institute of Research in Computer Science. Member of the *virtual plants* team. Involved in the *OpenAlea* project, Montpellier, France.

Software Engineering

Software developer on the *OpenAlea* project, a visual programming environment (Python/PyQt) dedicated to the creation of biological models and scientific workflows. Created interface with Python libraries such as as Matplotlib or SciPy. Professional usage of tools dedicated to software development from integration test (nosetests) to documentation (sphinx/restructured text) and bug tracking. Managed large Python and C++ libraries (300,000 lines of code).

Modelisation

Used generative grammar (L-systems) to model plant development and proliferation over time. Modelisation using mecanistic and stochastic processus (e.g., Markov chain).

CARDIFF UNI.

April 2003–April 2008

Research Associate/Assistant at Cardiff University (U.K.), in the Physics and Astronomy Department. Member of the *Gravitational Physics* group.

Detection theory, signal processing

Developped modelisation tools and software pipelines to detect small signal-to-noise ratio signals from astrophysical sources. I used detection theory, parameter estimation and spectral analysis to guarantee optimal detection. Analysed and manipulated large data sets (TeraBytes).

Collaborations

Member of large scientific collaborations such as *LIGO* (600 members). Involved in working groups dedicated to the detection of astrophysical signals using software engineering and contributed to large algorithm libraries written in C/Python.

EDUCATION

- PhD (1999-2003)** **Sciences, gravitational waves.** Université Nice Côte d'Azur, FRANCE. **Title:** *Detection of gravitational waves emitted by black hole-black hole coalescences.* Collaboration Thalès Marconi Sonar, Sophia Antipolis.
- MSc (1998-1999)** **Astronomy, Imaging** – University of Nice Sophia-antipolis, France.
- BSc (1996-1998)** **Physics** – Université du Littoral, Dunkirk, France.
- BSc (1994-1996)** **Mathematics and computer science** – Université du Littoral, Calais, France.

COMPUTING SKILLS

Languages Experienced developer specialised in scientific computing, signal processing, data mining and visualisation. Experience in software management (architecture and project management).

Python Expert knowledge. From standard libraries to scientific ones: Matplotlib (visualisation 2D/3D), numpy (multi-dimensionnal arrays manipulation), SciPy (numerical analysis), NetworkX (graph theory), Pandas (data analysis), SciKit-learn (machine learning). Organisation of a [Python working group](#) on the EMBL-EBI/ Wellcome Trust Campus

C/C++ Knowledge of C/C++ languages combined with Python or R languages.

Others R: maintainer of libraries on BioConductor website (e.g., CellNOptR). MATLAB/octave, Perl, sed/awk. Linux administrator.

Web Knowledge of HTML/CSS/Javascript. Web site examples: [home page](#), [CellNOptR](#).

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

Software

github Exhaustive list of personal software or contributions on [GitHub github.com/cokelaer](#).

BioServices Author of [BioServices](#) (Python), which provides programmatic access to 30 web services (UniProt, KEGG, ChEMBL, ...) covering life sciences from proteomic to gene expression and sequence alignment. Extensive [documentation](#) using Sphinx. Numerous [iPython notebooks](#).

OpenAlea Participated to development of OpenAlea (Python). Complex software that uses PyQt as a graphical interface, with more than 150,000 lines of code.

Spectrum Author of [Spectrum](#), spectral analysis library (Python). Can be used to apply Fourier transform, or parametric methods (ARMA, MA, BURG)) to create spectral estimation of time-series.

COMMUNICATION


Lectures Tutorial given during the European project BioPreDyn (2014, research level) on (1) logic modelling thanks to the CellNOptR software and (2) programmatic access to web services related to life science using BioServices. Lecture given on logic modelling (2014) CoLoMoTo consortium. Series of lectures on Python for beginners on using standard and scientific libraries (10 hours). Organised on the EMBL-EBI/Sanger Wellcome Trust Campus.

Mentoring Mentor of a master student (6 months, 2013) on optimisation techniques (applied to protein networks / phosphorylation data using MCMC methods). Mentor of a PhD student (3 months, 2015) (multi-dimensional data set analysis). Mentoring in physics and mathematics at BSc level (20 hours).

Publications Peer-to-peer publications in international journals (Physical Review, Nature Biotech., Bioinformatics journals). Full list of publications available [on line](#) or on [Research Gate](#).

CONTACT

THOMAS COKELAER
BioMedecine Team, EMBL-EBI
CAMBRIDGE, UNITED KINGDOM

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