

**BioPreDyn INTERIM REPORT**

**Grant Agreement number: FP7-KBBE-2011-5-289434**

**Project acronym: BioPreDyn**

**Project title:** From Data to Models: New Bioinformatics Methods and Tools for Data-Driven Predictive Dynamic Modeling in Biotechnological Applications

**Start date of project: 01.12.2011**

**End date: 01.12.2016**

**Interim report: 2st**

**Period covered:** from **01.04.2012** to **01.10.2012**

### Work package number: WP3: Integrated Software Tools for the Modelling Cycle

**This interim report is a deliverable**

Core of the report for the period: Project objectives, work progress, and achievements

### I. Work progress and achievements during the period

Please provide a concise overview of the progress of the work in line with the structure of Annex I to the Grant Agreement:

1. A summary of progress towards objectives and details for each task of the work package;

**Task 3.1**

USFD has continued working on their Python implementation of a Gaussian process toolkit. These efforts are feeding into this work package as well as WP5 below. The toolkit has now been extended by James Hensman to incorporate clustering of gene expression time series across different replicates (and even species). USFD has submitted a publication at BMC Bioinformatics considering a Drosophila development example and a submission to IEEE TPAMI describing the methodology underpinning the approach. These ideas come out of fundamental work that has been accepted for publication at the Neural Information Processing Systems conference in December 2012.

Contemporaneously with these developments, PhD student Ricardo Andrade Pacheco has been expanding the toolbox to deal with non-Gaussian data, using sparse Gaussian process methods and the expectation propagation approximation. USFD foresee that this may be useful in augmenting the differential equation model to deal with count data, for example deriving from RNA-seq experiments.

FTELE.IGM, in collaboration with USFD, has developed a novel probabilistic model to infer signalling pathways from massive gene expression profiles (both dynamic and steady-state data). Specifically, the method is able to identify modulators (i.e. kinases and phosphatases) regulating the activity of transcription factors. The approach is based on an innovative application of Multi-Information and it has been already validated on simulated datasets generated from conditional multivariate Gaussian distributions. Preliminary results obtained on the in-silico datasets prompted us to proceed testing the approach on real gene expression profiles collected from public databases. We are now completing validation of the method on these datasets.

**Task 3.2**

CSIC has developed a new parameter estimation strategy, based on an enhanced scatter search method (eSS, which belongs to the class of stochastic global optimisation algorithms). This technique makes use of diversification (global search) and intensification (local search) methods so it can be considered as an advanced hybrid strategy. Currently it can handle non-linear programming (NLP) and, to some extent, mixed-integer nonlinear programming (MINLP) problems. Several innovative mechanisms have been implemented in eSS in order to enhance its efficiency and robustness when solving large-scale optimization problems. CSIC has also implemented another stochastic method for integer programming (combinatorial optimization) based on extensions of the variable neighbourhood search (VNS) metaheuristic. These methods were initially implemented in Matlab, and later, in collaboration with EMBL, have also been implemented in R. CSIC has tested this novel method with benchmark problems, including parameter estimation in large-scale model of microorganisms (see also task 3.3 below).

UNIMAN supplied two benchmark parameter estimation problems that were shared with the CSIC team. These problems are particularly hard parameter estimation problems: a yeast metabolic model with 306 reactions, and a protein translation model for yeast with 56 parameters to be estimated from data from 150 steady states.

The work in this task proceeds as planned and is closely related with deliverable D3.2 (Parameter Estimation Tools), due on month 18.

CRG...

**Task 3.3**

CSIC has implemented parallel cooperative metaheuristics which automatically favour the specific optimization strategies developed in T3.2 according to the measured current efficiency of each algorithm. The resulting cooperative parallel techniques, named CeSS and CVNS, were initially implemented in Matlab and tested with several large-scale parameter estimation problems, including two challenging problems related with *E.coli*. This work has been recently published in BMC Systems Biology.

Further, CSIC and EMBL have collaborated implementing these methods in a software toolbox (MEIGO, metaheuristics for global optimization in bioinformatics and systems biology), which allow users to choose among a wide range of powerful global search methods, taking advantage of parallel high-performance computers. The MEIGO toolbox offers Matlab and R implementations, thus facilitating its use with many existing packages in bioinformatics. EBI has also developed a Python interface to the R version.

MEIGO can be executed in a wide range of computing architectures, ranging from low-cost multi-core PCs to larger and more sophisticated computing clusters. CSIC and EMBL are currently writing a paper describing the toolbox, which will be made available to the scientific community. MEIGO is currently being tested against problems of parameter estimation of dynamical models, as well as structural identification of discrete Boolean models, by EMBL and CSIC.

Similarly to task 3.2, the work in task 3.3 proceeds as planned and is closely related with deliverable D3.2 (Parameter Estimation Tools), due on month 18.

**Task 3.4**

UvA is currently working on applying multi-objective optimization in the gene regulatory network models which they have developed for *Nematostella vectensis* (see WP1 and WP2).

CRG ...

CSIC has developed and applied a multi-objective optimization framework to the problem of regulation in metabolic networks. A multi-criteria approach has been used to optimize parameters for the allosteric regulation of enzymes in a model of a metabolic substrate-cycle. This has been carried out by calculating the Pareto set of optimal solutions according to two objectives: the proper direction of flux in a metabolic cycle and the energetic cost of applying the set of parameters. The results indicate that final and optimal consensus set of parameters can be obtained, which is an indication of the possible existence of a universal regulation mechanism in metabolic networks. This work has been published in PLoS ONE.

**Task 3.5**

CWI has extended the software prototype developed to deal with multi-dimensional independent normally distributed uncertainties to handle correlated and non-normal uncertainties. First results have been obtained on problems from biochemistry and neurodynamics. The quantification of uncertainty in biochemistry predictions is an important goal in this project. For this reason CWI is also developing a method to look for optimal probability density functions in biochemistry in order to find the best probabilistic model for the quantities (parameters, initial conditions, inputs). We are applying Polynomial Chaos (PC) expansions for the representation of probabilities density functions and study their utility for the propagation of uncertainty in this context. We are also studying uncertainty quantification of (the delay parameter in) delay equations; with application models for developmental gene regulatory networks (WP6)

UvA has developed a multi-scale model of calcification in scleractinian corals in which the existing knowledge on the contributing sub-processes is integrated within a mathematical framework. UvA has developed a spatial representation of a single coral polyp surrounded by seawater. In this geometry we simulate the relevant chemical reactions in the external seawater, inside the organism and at the site of calcification as well as molecular transport processes, photosynthesis, respiration and calcium carbonate precipitation. UvA models the spatio-temporal dynamics of the relevant molecular fluxes as a set of coupled reaction-diffusion equations combined with biological transport modelled as Michaelis-Menten and Hill equations. Simulations can be employed to clarify the influence of different individual processes and reaction rates as well as changes in the chemical composition of the surrounding seawater. Details about the model can be found in a publication (C. Cronemberger, L. Huisman, D. Allemand and J.A. Kaandorp, A spatial model of calcification in scleractinian corals, submitted).

CRG and CSIC have been collaborating on the identifiability of a model of protein production from mRNA on the gap-gene circuit of the Drosophila blastoderm. Spatial expression patterns of mRNA and protein for gap genes (gt, Kr, and kni) are modelled using a simple linear delay model with the objective of investigating whether post-transcriptional regulation is required to explain the quantitative patterns. The model would also allow the determination of production, decay, and diffusion rates, as well as delays, for all three genes. CRG is interested in a sound statistical analysis of the model fit and its identifiability analysis. CSIC is performing a structural identifiability analysis of this model, which is non-trivial due to the delay and the presence of a boxcar function representing the shutdown of translation during mitosis. Currently, CSIC is using a non-standard approach based on approximating the boxcar and delayed control by suitable differentiable functions, and then applying the generating series method. Preliminary results indicate that the model parameters are globally identifiable provided some conditions on the measurements and the delayed control are met. CSIC will double check these results with another approach, and will also perform a practical (robust) identifiability analysis.

Regarding deliverable 3.4 (CSM – Integrated software suite), CSM is waiting for tools developed in WP3 to be made available for integration. CSM created one page in the project intranet for each use case identified during the CRG internal workshop (June 2012); the aim is to gather information about those cases analysis pipelines. Finally, the pLSA (parallel Lam Simulated Annealing) code was transmitted by CRG to CSM; this code is currently being analysed by CSM.

CRG...

**Task 3.6**

EMBL has developed CellNOptR, a package for the optimization of boolean logic networks of signalling pathways based on a previous knowledge network and a set of data upon perturbation of the nodes in the network. This package will be able to make use of the solvers implemented in R and developed in task 3.3 above.

1. clearly highlight significant results;

CSIC has successfully developed a first version of the parameter estimation tools planned for deliverable D3.2 (due on month 18), successfully testing them with several challenging case studies dealing with large-scale *E.coli* models. These results have been published in BMC Systems Biology. CSIC and EMBL have closely collaborated in the development of a software toolbox, MEIGO, implementing these methods in two popular environments (Matlab and R), plus an interface with Python. A joint publication describing the toolbox is in preparation.

UvA: by our knowledge most computational physiology models are based on ODEs and there are only a few examples of (published) computational physiological models which include a spatial component. Potentially these spatio-temporal physiological models can be coupled with spatio-temporal models of gene regulation and biomechanical models of cells and cell movement and are very useful in multi-scale modelling.

1. if you have had deviations from Annex I, explain the reasons for them and how this has impacted other tasks and/or available resources and planning;
2. if applicable, explain the reasons for failing to achieve critical objectives and/or not being on schedule, and explain the impact on other tasks as well as on available resources and planning;
3. have you have hired people for the project as planned? Explain any deviations between actual and planned actions;

EMBL: Martijn van Iersel, the hired scientist, left group in July 2012. Currently a search is open (likely he will be replaced by PhD student + partial support for staff scientist).

CWI: -deviation: no person months during the first six months. The candidate who accepted the position after the first application round finally did not take the position. A second application round was open.

-action: Finally, the position was split; one part-time position (0.2FTE) has been filled by a postdoc (Aldemar Torres Valderrama) who moved to CWI on 1 April 2012; the second, full-time position has been taken up by the candidate chosen in the second application round (Maria Navarro) who was hired on 1 August 2012.

UvA: L. Huisman has been working since 1 October 2011, but resigned in August 2012. She has been replaced by P. Silva (we hired her since 15 June 2012). D. Botman is working, since 1 February 2012, in the BioPreDyn project.

1. if applicable, propose corrective actions.

**II. Partner/project dissemination**

1. Meetings attended by WP members since 01/04/2012

CSIC:

* BioPreDyn Workshop, Barcelona, Spain, 11-15 June 2012. Instructors from CSIC: Julio R. Banga and Eva Balsa-Canto. Attendants from CSIC: Alex F. Villaverde and David Henriques.

Visits of CSIC members to other BioPreDyn partners:

* Julio R. Banga: research stay at EBI-EMBL (Saez group) during 3-30 July 2012
* Alex F. Villaverde: visit to UniMan (Mendes group), 19-20 April 2012
* Alex F. Villaverde: visit to InSilico Biotechnology (Stuttgart), 25-27 Sept 2012

UNIMAN:

Pedro Mendes attended the BioPreDyn Workshop in Barcelona (CRG), from 11/6/2012 to 14/6/2012, and delivered a lecture on optimisation. Together with Jean-Marie Muillon (FS) he presented the yeast benchmark model to the rest of the consortium.

CSM:

• Eric Boix, Bertrand Moreau – Internal CRG workshop, Barcelona, 11-15th of June 2012

• Bertrand Moreau – WP7 Yeast kinetic model definition, Manchester, 12-13th of September 2012

• Bertrand Moreau – WP7 Bioproduction software requirements, Teleconference, 28th of September 2012

FTELE.IGM:

* BioPreDyn Workshop, June, 11-15th , CRG, Barcelona (FTELE.IGM)

1. Presentations that acknowledged BioPreDyn at open

conferences/workshops since 01/04/2012

CSIC:

Alex F. Villaverde attended:

- XXXIII Jornadas de Automática, Vigo, Spain, 5-7 September 2012.

- Joint EMBL-EBI-Wellcome Trust Course In silico Systems Biology, 23-27 April 2012, Hinxton, Cambridge, UK.

EMBL:

Julio Saez-Rodríguez:

- EFMC course on medicinal Chemistry, April 2012 Leiden, Netherlands

- Joint EMBL-EBI and Wellcome Trust In Silico Systems Biology, Hinxton UK

- Computational Proteomics School, June 2012, Munich, Germany

- Computational Plant Biology, August 2012, Sainsbury Institute, Cambridge

UNIMAN:

Pedro Mendes presented a seminar entitled “Modelling biochemical networks: What is there to do?” at the University of Edinburgh, 25/4/2012.

Pedro Mendes presented a seminar entitled “From networks to models – large scale kinetic models of metabolism” at the University of Saarbrucken, 23/7/2012.

CWI:

Poster entitled “BioPreDyn” at the CWI Business day (5 October 2012) and Open day (6 October 2012).

UvA:

Jaap Kaandorp (invited lecture) Modelling gene regulation of morphogenesis, International Young Scientists Conference “High Performance Computing and Simulation’’, April 2012, Amsterdam

Jaap Kaandorp (invited lecture) Modelling gene regulation of morphogenesis, Turing Centenary Conference and 8th conference on Computability in Europe, June 2012, Cambridge, UK

Jaap Kaandorp (invited lecture) Modelling gene regulation of morphogenesis in the sea anemone *Nematostella vectensis*, The 8th International Conference on Bioinformatics of Genome Regulation and Structure Systems Biology, June 2012, Novosibirsk, Russia

Jaap Kaandorp, (oral presentation) A spatial model of calcification in scleractinian corals, 12th International Coral Reef Symposium, July 2012, Cairns, Australia

Jaap Kaandorp (oral presentation) Complexity of morphogenesis, European Conference on Complex Systems, September, 2012, Brussels

Paula Ramos-Silva, Isabelle Zanella-Cleon, Lotte Huisman, Benjamin Marie, Jaap Kaandorp and Frédéric Marin "Acropora ‘skeletome’ reveals coral specific proteins and common functional domains in biomineralization", COST action TD0903 workshop, 18th-20th September 2012, Aarhus, Denmark

1. Submission of publications that acknowledge BioPreDyn since 01/04/2012

CSIC:

Alejandro F Villaverde, Jose A Egea, Julio R Banga (2012). A cooperative strategy for parameter estimation in large scale systems biology models. BMC Systems Biology 2012, 6:75.

Higuera C, Villaverde AF, Banga JR, Ross J, Morán F (2012) Multi-Criteria Optimization of Regulation in Metabolic Networks. PLoS ONE 7(7):e41122.

Egea JA, Henriques D, Cokelaer T, Villaverde AF, Banga JR, Saez-Rodriguez J. (2012). MEIGO: a software suite based on metaheuristics for global optimization in systems biology and bioinformatics. In preparation.

EMBL:

F. Eduati, J. de las Rivas, B. di Camillo, G. Toffolo, J. Saez-Rodriguez. Integrating literature-constrained and data-driven inference of signalling networks. Bioinformatics, Jun 25 2012.

A. MacNamara, C. Terfve, D. Henriques, B. Pealver Bernab, J. Saez-Rodriguez. State-time spectrum of signal transduction logic models. Physical Biology 9(4):045003, 2012.

F. Iorio, J, Saez-Rodriguez, & D. Di Bernardo,. Network based elucidation of drug response: from modula- tors to targets. submitted to Open Network Biology

C. Terfve, C. et al. CellNOptR: a flexible toolkit to train protein signaling net- works to data using multiple logic formalisms. Submitted to BMC SysBio.

Egea JA, Henriques D, Cokelaer T, Villaverde AF, Banga JR, Saez-Rodriguez J. (2012). MEIGO: a software suite based on metaheuristics for global optimization in systems biology and bioinformatics. In preparation.

CWI:

Katja Rybakova, Aleksandra Tomaszewska, Simon van Mourik, Joke Blom, Hans V. Westerhoff, Carsten Carlberg, and Frank J. Bruggeman, Tracing the molecular basis of aberrant transcription in noisy data by using an experiment-based mathematical model. Under submission.

UNIMAN:

Heavner BD, Smallbone K, Barker B, Mendes P, Walker LP (2012) Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network *BMC Systems Biology* **6**:55 (submitted 3/3/2012, accepted 4/6/2012)

Kieran Smallbone, Natalie J Stanford (2012) Kinetic modelling of metabolic pathways: Application to serine biosynthesis. *Methods in Molecular Biology*. In Press.

UvA:

Paula Ramos-Silva, Isabelle Zanella-Cleon, Lotte Huisman, Benjamin Marie, Jaap Kaandorp, David Miller and Frédéric Marin "Acropora ‘skeletome’ reveals coral specific proteins and common functional domains in biomineralization" (in preparation).

C. Cronemberger, L. Huisman, D. Allemand and J.A. Kaandorp, A spatial model of calcification in scleractinian corals (submitted)

D. Botman and J.A. Kaandorp, Spatial gene expression quantification: a tool for analysis of in situ hybridization in the sea anemone Nematostella vectensis) (submitted, under review)

P. Ramos-Silva, S. Benhamada, N. Le Roy, B. Marie, N. Guichard, I. Zanella-Cléon, L. Plasseraud, M. Corneillat, G. Alcaraz, J. Kaandorp and F. Marin. Orphan proteins in molluscan shell biomineralization: a case study with Upsalin. ChemBioChem, 2012, 13:1¿13 DOI: 10.1002/cbic.201100708

FTELE.IGM;

Roberto Pagliarini and Diego di Bernardo. A genome-scale modelling approach to study inborn errors of liver metabolsim: towards an in-silico patient. Submitted to Special Issue J Comp Biol.

Gennaro Gambardella, Maria N. Moretti, Rossella De Cegli, Luca Cardone, Adriano Peron and Diego di Bernardo Differential Network Analysis for the identification of condition-specific pathway activity and regulation. Submitted to Bioinformatics

**Work package overview:**

**WP1**. Database Integration & Exploitation. WP leader: di Bernardo; Partners: Saez-Rodriguez, Jaeger, Kaandorp, Mendes, Lawrence, Boix

**WP2**. Visualisation Tools for Data & Model Building. WP leader: Lawrence; Partners: Saez-Rodriguez, Jaeger, Lawrence, Boix, Kaandorp, Rattray

**WP3**. Integrated Software Tools for the Modelling Cycle. WP leader: Banga; Partners: Jaeger, Saez-Rodriguez, Blom, di Bernardo, Mendes, Lawrence, Rattray, Boix, Kaandorp

**WP4**. Application: Large-scale Models of Microorganisms WP leader: Saez-Rodriguez; Partners: Banga, Blom, di Bernardo, Lawrence, Mattray, Mauch

**WP5**. Application: Signalling & Regulatory Networks in Cells WP leader: Saez-Rodriguez; Partners: Banga, Blom, di Bernardo, Lawrence, Mattray, Mauch

**WP6**. Application: Developmental Gene Regulatory Networks in Animals. WP leader: Kaandorp; Partners: Jaeger, Blom, Lawrence

**WP7**. Application: Biotechnological Production Processes. WP leader: Moullon; Partners: Banga, Blom, Mendes, Mauch, Mouillon

**WP8**. Dissemination, Exploitation & Training. WP leader: Boix; Partners: Jaeger, Banga