**BioPreDyn**

From data to models: new bioinformatics methods and tools for data-driven predictive dynamic modelling in biotechnological applications

FP7 THEME [KBBE.2011.3.6-01]

Grant number 289434

**WP3: Integrated Software Tools for the Modelling Cycle**

**DETAILED IMPLEMENTATION PLAN (DIP) FOR THE M1-M12 PERIOD**

**DIP period: October 1st, 2011 to September 30, 2012**

**Date of preparation: 9 Nov 2011**

**Revision: 4.1** (last revised 22-Dec-2011)

**Start date of the project:** October 1st, 2011 **Duration:** 36 months

One form per Work Package

**Work package number: WP3 Type of activity: RTD**

**Work package title: Integrated Software Tools for the Modelling Cycle**

Start month 1 / End month 36

WP leader: CSIC (partner no. 2)

**Persons-months**

**Number Partner short name Person-months per partner**

1 CRG 18.00

2 CSIC 30.00

3 EMBL 3.00

4 UvA 15.00

5 CWI 30.00

6 FTELE.IGM 12.00

7 UNIMAN 6.00

8 USFD 17.00

9 CSM 10.00

Total 141.00

**Principal investigators and postodcs**

**Number Partner short name P.I. Postdoc/other**

1 CRG J. Jaeger Anton Crombach, postdoc

Damjan Cicin-Sain,programmer

2 CSIC J. R. Banga Alex F. Villaverde

3 EMBL J. Saez Martijn van Iersel

4 UvA J.A. Kaandorp Lotte Huisman PhD

Daniel Botman PhD

5 CWI J. Blom T.B.Appointed (PostDoc)

6 FTELE.IGM D. di Bernardo G. Gambardella (PhD)

Luca Cardone (PostDoc)

7 UNIMAN P. Mendes Kieran Smallbone

8 USFD N.Lawrence/M.Rattray R. Andrade Pacheco (PhD)

T.B.Appointed (PostDoc)

9 CSM E. Boix Thomas Lacroix (CTO),

Bertrand Moreau (Jan. 2012)

O

bjectives

**Objectives of WP3**

To develop novel methods to support the **model building cycle**, and to integrate them into a unified, powerful and easy-to-use **software framework**, which can be applied to a wide range of **modelling** activities and processes.

WP3 forms one large, indivisible unit at the core of our project by specifically pooling the expertise of the academic partners to produce an **integrated suite of methods and software tools for model identification, optimization, and analysis, as well as for optimal experimental design**. The synergistic and complementary expertise that we accumulate within our consortium will ensure that algorithm development will be up to the most stringent quality standards possible, and will enable novel combinations and algorithmic developments.

**Progress beyond the state of the art (from DoW p 62)**

Regarding the model-building cycle, currently there are diverse tools, often incompatible with each other for model identification, fitting, analysis etc.

The **expected outcome** is to deliver an integrated suite of tools and newly developed/improved algorithms to support the entire model building cycle.

The **performance indicator** will be the public availability of tools as part of the software package to be developed by CSM.

**Description of work and role of partners (details for M1-M12 period)**

**Task 3.1**: We will implement **Bayesian approaches to model building** for tractable models based on differential equations and Gaussian processes, as well as for less tractable models based on non-linear differential equations and probabilistic modelling where Markov Chain Monte Carlo methods are required for parameter inference.

**Task leader**: USFD.

**Participants and roles**: USFD, in collaboration with FTELE.IGM who will provide additional technical expertise, will develop the methods and implement the models required for this task.

**Objectives for M12**:

FTELE.IGM, in collaboration with USFD will explore probabilistic models to infer modulators of gene expression (i.e. kinases, phosphatases, microRNA, etc.) from massive gene expression profiles (both dynamic and steady-state data). Specifically, we will develop a set of algorithms able to predict post-transcriptional regulators of gene activityfrom . This effort will complement the work of USFD which is specifically addressing the problem of modeling and inferring transcriptional regulation from dynamic gene expression profiles.

**Subtasks**:

**Task 3.1.1**- Identification of post-transcriptional modulators of gene expression. We will investigate the use of multi-dimensional mutual information estimation from the field of image analysis to identify the modulator(s) whose level of expression is predictive of coordinated regulation of a set of genes. We will test the approach first on simulated data and then on real data set (modulators of MYC-target expression).

**Task 3.1.2**- Modify the Gaussian process framework to incorporate the action of kinases or other modulators in the initiation of transcription.

**Planned interactions and visits**:

We plan the visit of Mr Gennaro Gambardella (FTELE.IGM) to the lab of Prof. Neil Lawrence in March-April 2012. Prof Lawrence will also make regular site visits to the lab of di Bernardo.

**Deliverables linked**:

None in the period M1-M12. At a later time:

D3.1) Bayesian Inference Tools: New algorithms based on a Bayesian approaches to identify genome-wide regulatory network topologies from heterogeneous information [month 24]

**Task 3.2**: We will develop **new parameter estimation strategies**, based on stochastic global optimisation algorithms. These will be paired with fast local search algorithms to yield powerful hybrid search strategies.

**Task leader**: CSIC.

**Participants and roles**: CSIC, UvA, CWI, CRG and UNIMAN will combine their expertise to develop new, and improve their existing optimisation algorithms.

**Objectives for M12**: to develop and test new parameter estimation methods based on global optimization which are significantly (i.e. one order of magnitude) more efficient that the current state of the art while guaranteeing a satisfactory level of robustness.

**Subtasks**:

**Task 3.2.1**- Development and testing of novel global optimization methods for large-scale parameter estimation problems. CSIC will focus on extensions of metaheuristics based on the scatter search framework, paying special attention to local methods which can handle large-scale problems in an efficient way. These extended and improved metaheuristics will be tested on various benchmark problems (see next Task). UvA will focus on the development of stochastic global optimization methods (evolutionary algoritms, simulated annealing et.c) and multi-objective optimization (see task 3.4)

**Task 3.2.2**- Development of a set of representative benchmark problems to evaluate the new methods. All the participants will contribute with a set of realistic parameter estimation problems which can be used as a test bed for measuring the performance of the novel methods being developed. These benchmark problems consist of measured data and the corresponding model(s) presented in a uniform way, similarly to existing benchmarks, like e.g. the DREAM6 benchmark for parameter estimation. The standard for this will be developed during the first year.

- UNIMAN will supply benchmark problems related with large-scale kinetic models of yeast and *E. coli*.

- CRG will provide a spatial test problem of a pattern-forming network in early fly development.

- UvA will provide a spatial test problem of a pattern forming network

**Task 3.2.3**- As a stepping stone for task 3.3, implementation of the new methods on high performance computing environments. Both traditional clusters (CSIC, CRG) and new GPGPU-based alternatives (CWI) will be considered.

**Planned interactions and visits**: the postdocs of these groups will keep in touch via email, meeting via skype every two months and reporting progress to the PIs. The postdocs will organize an internal workshop in ?location? around M5 where each group will present the current status of the methods and their performance.

Organization of a Lorentz workshop on this topic by CWI, with participation of all the partners in this task plus other selected external participants.

On the 16 and 17th of April 2012, UvA is organizing the 9th International Conference on Networks in Bioinformatiucs in Amsterdam and would like to invite BioPreDyn partners to present their work, we can reserve the 18th of April for technical meetings

**Deliverables linked**:

None in the period M1-M12. At a later time:

D3.2- Parameter Estimation Tools: New software tools for parameter estimation via global non-linear optimisation (including co-operative parallel meta-heuristics making use of high performance computing facilities [month 18]

**Task 3.3**: We will implement **parallel meta-heuristics**, which automatically favour specific optimisation strategies developed in T3.2 according to the measured current efficiency of each algorithm. These techniques will be implemented in software toolboxes which allow the user to choose among a wide range of powerful global search methods, taking advantage of parallel high-performance computers (including GPU-based architectures), as well as distributed/cloud computing on variable architectures.

**Task leader**: CSIC.

**Participants and roles**: CSIC will develop the cloud-/parallel-computing code framework required for combining optimization algorithms developed by CSIC, as well as those provided by CRG, UvA, CWI and UNIMAN.

**Objectives for M12**: to develop and implement prototypes of parallel meta-heuristics which are able to run on Linux clusters, and to evaluate them using the benchmark set developed in 3.2.2

**Subtasks:**

**Task 3.3.1**- Development of parallel metaheuristics. CSIC will implement a cooperative strategy based on extensions of scatter search and selected local methods. Both homogeneous and heterogeneous cooperating strategies will be evaluated. Initially, prototypes will be developed in Matlab. Porting of the best strategies to Python/C will be done after M12. Partners such as CRG will test these algorithms in their research context.

**Task 3.3.2**- Evaluation of GPU-based architectures as alternatives/complement to standard parallel clusters. CSIC will explore possible workflows for porting the metaheuristics developed in 3.3.1 to clusters, and CWI will explore the use of accelerators for GPUs.

**Planned interactions and visits**: same as in T3.2 above

**Deliverables linked**:

None in the period M1-M12. At a later time:

D3.2- Parameter Estimation Tools: New software tools for parameter estimation via global non-linear optimisation (including co-operative parallel meta-heuristics making use of high performance computing facilities (incl. GPU-based architectures) [CSIC, month 18]

**Task 3.4**: We will develop efficient algorithms for **parameter estimation via multi-objective optimisation** (for example, maximising both goodness of fit and robustness of the resulting network models).

**Task leader**: UvA.

**Participants and roles**: UvA will co-ordinate integration of multi-objective methods into existing search strategies provided by CSIC, CRG, CWI and UNIMAN.

**Objectives for M12**: to develop and test new parameter estimation methods based on global optimization. Usually, the accuracy with which a model reproduces observed expression patterns is measured by a cost function based on the sum of squared differences between model and data (single-objective optimisation). Here, we propose to take advantage of the fact that pattern formation must proceed reliably in the presence of molecular fluctuations, genetic variability and environmental perturbations. In other words, realistic patterning mechanisms are robust, and robustness should be considered when fitting models to data. This is achieved by adding additional optimisation criteria to the fitting procedure (multi-objective optimisation)

**Subtasks:**

**Task 3.4.1**- UvA will focus on the development of stochastic global optimization methods (evolutionary algoritms, simulated annealing et.c) and develop tools for the analysis of solutions (robustness, stability, etc.) and test new multi-criteria cost functions which can be used in multi-objective optimization

**Task 3.4.2**- CSIC will implement and test new multicriteria global optimization methods based on extensions of the normal boundary intersection (NBI) framework.

**Planned interactions and visits**: On the 16 and 17th of April 2012 the UvA is organizing the 9th International Conference on Networks in Bioinformatiucs in Amsterdam and would like to invite BioPreDyn partners to present their work, we can reserve the 18th of April for technical meetings, furthermore the symposium will provide the possibility to invite more researchers working on optimization methods and related topics

**Deliverables linked**:

None in the period M1-M12. At a later time:

D3.3) Multi-objective Optimisation Tools: New software tools for multi-objective optimisation, implementing a wide range of cost functions [UvA, month 36]

**Task 3.5**: Development of **novel methods, protocols and software tools for model building,** with a special focus on multi-scale modelling, model selection and discrimination, parameter identifiability analysis (both theoretical and practical), model validation and uncertainty quantification.

**Task leader**: CWI

**Participants and roles**: CWI, CSIC, CRG, UvA and UNIMAN will contribute and integrate novel as well as existing algorithms for these tasks to the integrated software framework to be developed by CSM. INSIL will contribute additional algorithms, and also integrate these algorithms into their own software framework.

**Objectives for M12**:

To integrate the steps in the SB-cycle by developing an SB-cycle protocol, optimization criteria that include data-information, model validation strategies. To develop uncertainty quantification methods to incorporate the data- and parameter distributions in the model simulation results. To develop strategies for rigorous multi-scale modeling.

**Subtasks**:

**Task 3.5.1**- CWI will develop an SB-cycle protocol with input from all participants.

**Task 3.5.2**- CWI will develop methods for uncertainty quantification tailored for biochemical deterministic models

**Task 3.5.3**- CSIC will develop methods for model selection and discrimination based on mixed-integer dynamic optimization (MIDO) formulations

**Task 3.5.4**- CWI will study the integration of biochemical models at deterministic (ODE/PDE) and probabilistic level (CME/RDME, queueing theoretical models)

**Task 3.5.6**- CSM will illustrate how to develop a simulator when given some reconstructed multi-scale model (chosen within the Applications). CSM will then illustrate how this simulator could be used for model validation, uncertainty analysis and conducing prospective numerical experiments.

**Planned interactions and visits**:

As in Task 3.2.

**Deliverables linked**:

None in the period M1-M12. At a later time:

D3.4) Integrated Suite of Tools: Integrated software-suite for iterative multi-scale model building providing tools for all the steps in the modelling cycle; documentation describing the suite, incl. algorithm comparison & applications [CSM, month 36]

**Task 3.6:** Integration of the above methods with the CellNOpt platform for **large-scale logic modelling.**

**Task leader:** EMBL

**Participants and roles**: EMBL will adapt CellNOpt to integrate the methods developed in the package, and for use with the integrated suite to be developed by CSM within this project.

**Objectives for M12**:

The methods to be developed in this WP can be applied to the large scale logic models that EMBL develops. Given the size and complexity of these models, and the amount of data they are trained to, these will serve as a realistic benchmark for those methods too.

**Subtasks**:

**Task 3.6.1**- Application of the approaches of Tasks 3.1., 3.2, and 3.3. to the logic models of EMBL. Exploration on various case studies. EMBL has access to a large cluster where these methods can be tested.

CSM will participate to the establishment of the regression tests, the demonstration scenarios and the relevant use cases to be integrated within treatment pipelines (for the Applications)

**Task 3.6.2**- Linking of these methods to the CellNOpt tool, both for its matlab and R implementation.

**Task 3.6.3**- CSM will assert the usage of the R version of CellNOpt (library) mode within the treatment framework.

**Planned interactions and visits**:

Close collaborations are expected between EMBL and the groups developing the methods, in particular with CSIC. These will be based on regular videoconference and some meetings in person. Later on, meeting with CSM to integrate in the platform.

CSM will visit EMBL CellNOpt experts in order to transfer the required know-how (that CSM will document for the other partner to access)

**Deliverables linked**:

None in the period M1-M12. At a later time:

D3.4) Integrated Suite of Tools: Integrated software-suite for iterative multi-scale model building providing tools for all the steps in the modelling cycle; documentation describing the suite, incl. algorithm comparison & applications [CSM, month 36]