AAPG2022	BIOPSYN	Funding instrument
Coordinated by :	François FAGES	Duration 48m
H.14 Interfaces : mathématiques, sciences du numérique - biologie, santé		

BIOPSYN
Biochemical Program Synthesis

I. Proposal's context, positioning and objective(s)

a. Objectives and research hypothesis

Chemical Reaction Networks (CRN) are a standard formalism used in chemistry and biology to describe, analyze, and now also design, complex molecular interaction networks. In the perspective of systems biology, they are a central tool to analyze the high-level functions of the cell in terms of low-level molecular interactions. In the perspective of synthetic biology, they constitute a target programming language to implement in chemistry new functions in either living cells or artificial vesicles.

Based on a previous result showing the **strong Turing-completeness of finite CRNs** interpreted by Ordinary Differential Equations (ODE) [26] which was the last open computability problem for CRNs [6], and on further results [13, 14, 17] obtained in the framework of our former common ANR-MOST BIOPSY project, and also ANR-DFG SYMBIONT project, the first partner has implemented, in the open-source Biochemical Abstract Machine software BIOCHAM, an original pipeline for **compiling any elementary mathematical function** (either of time, or of the concentration of some input species) into an **abstract CRN** which implements that function (see Fig. 1). The result is given by the concentration of one distinguished output molecular species, either by its trace for a function of time, or by its stabilization value for an input/output function.

Based on original research in the past two decades [31], the second partner has shown his capabilities to design **concrete CRNs** to implement **Boolean circuits** in chemistry [28], to assemble them in artificial vesicles created by microfluidic devices [18,23,25,27,32], and to validate them for innovative biomedical diagnosis applications [21,28,29,25].

The objective of BIOPSYN project is to combine the unique skills of both partners to go beyond the synthesis of logical circuits for diagnosis applications, in order to implement more general continuous mathematical functions [16], such as antithetic integral controllers for the synthesis of chemical circuits ensuring homeostasis [5]. The approach investigated in BIOPSYN will consist in constraining our abstract compilation pipeline to use the ODE terms of a limited set of well-characterized real enzymatic reactions. To tackle this open challenge, we will focus on four main tasks:

1. the **static mapping** from the abstract CRN to a concrete CRN (Fig. 2 as backend of Fig. 1);
2. the **symbolic computation theory of our CRN compilation pipeline** in order to constrain it to the limited resources of a reaction catalogue in the early ODE transformation phases for the implementation of a mathematical function (output of Fig. 2 as input to Fig.1);
3. the **dynamic interaction** between the ODE term transformations and the extraction of concrete reactions (Fig. 1 and Fig 2. in interaction).
4. The design and making of **one study case with *in vitro* validation**.

The study cases of BIOPSYN will be twofold. First, the redesign using BIOPSYN tools of a combination of small CRNs implementing simple **logical circuits**, already built separately, well-characterized and validated, to combine them in a single vesicle with **affordable validation**. Second, the design of a concrete CRN ensuring **the homeostasis of pH/ions and partial O2 pressure in an artificial vesicle**, a crucial step before launching a major project we have in mind for the creation of artificial minimal "red blood vesicles" along these design principles.

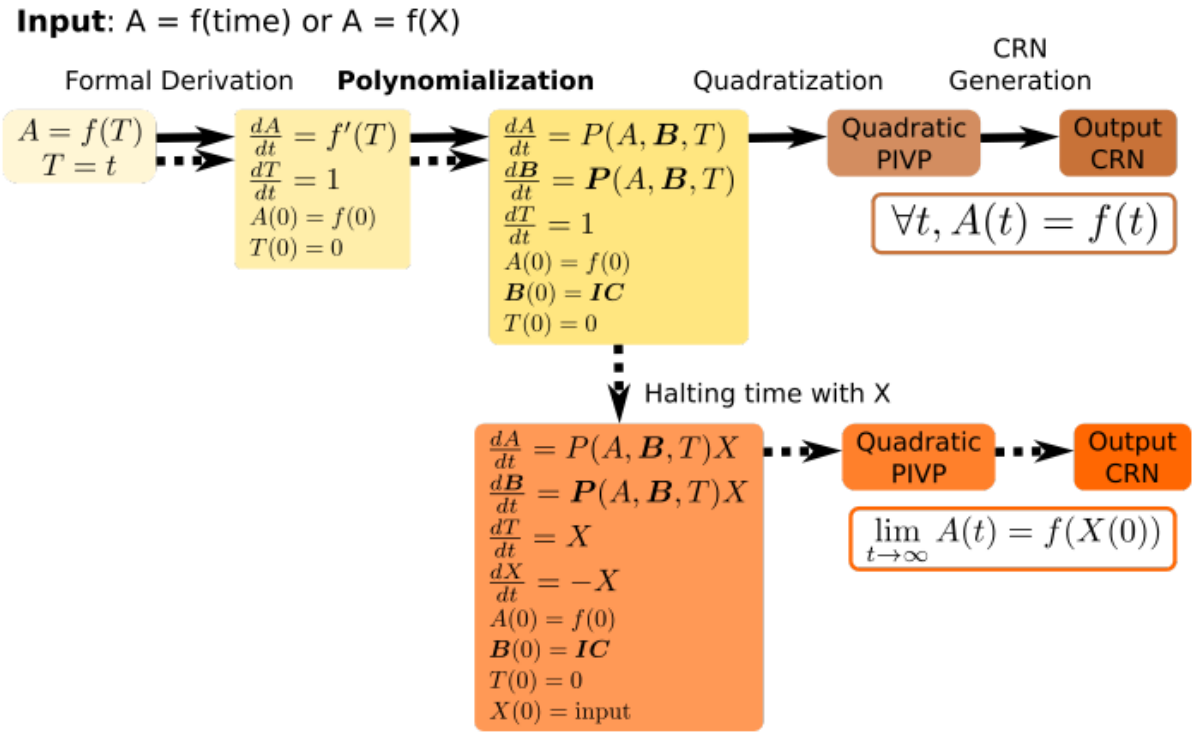

Figure 1. *Abstract compilation pipeline implemented in BIOCHAM v4 by the first partner to transform any elementary mathematical function f into an abstract CRN [13,17,26], either a function of time (plain arrows) or an input/output function (dashed arrows): P and **P** are polynomials, **B** denotes the set of abstract auxiliary species introduced by polynomialization given with initial conditions **IC**.*

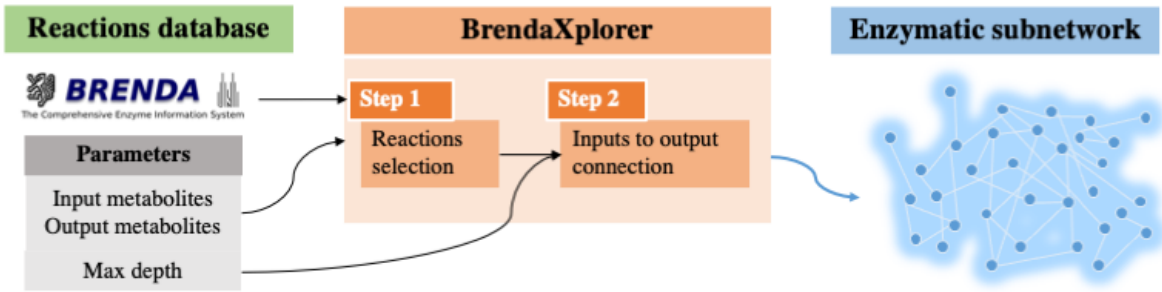

Figure 2. *Concrete database pipeline implemented in BrendaXplorer software by the second partner to extract chemical reactions, currently used with SiliCell Maker software for implementing logical circuits [28,31,32].*

b. Position of the project as it relates to the state of the art.

Inria Saclay EPI Lifeware and Sys2Diag CNRS/ALCEN lab. are recognized key actors in the fields of, respectively, CRN computational theory and artificial non-genetic CRN design and making. They already joined their forces in early collaborations to create theoretical, computational and experimental foundations in this field.

BIOPSYN is a very original research follow-up project of our France-Taiwan ANR-MOST BIOPSY project, focusing on some fundamental research issues with highly innovative applications potential. The compilation pipeline implemented in Biocham by Partner 1 (Fig. 1) and the SiliCell Maker software developed by Partner 2 (Fig. 2) have been developed with partial support from the former ANR-MOST BIOPSY *Biochemical Programming Systems* project. The abstract compilation pipeline of Partner 1 also benefited from partial support for the French-German ANR-DFG SYMBIONT *Symbolic Methods for Biological Networks* project.

We are aware of only two systems performing similar CRN synthesis tasks:

- The CRN++ system [2] developed by David Soloveichik et al. at Austin Texas Univ. and Erik Winfree at Caltech USA
- The work of Luca Cardelli et al. [6] at the University of Oxford, UK.

Both work however in the perspective of making **concrete CRN implementations** with **DNA strand displacement** systems, whereas we target **DNA-free enzymatic reactions with proteins** for several reasons. Enzymatic reactions with proteins are indeed

- faster by several orders of magnitude,
- of the same type as natural CRNs in the cell,
- mainly based on analog computation as in natural cells [16],
- safer than DNA containing devices for many biomedical applications,

but also limited

- in variety,
- in reaction rates,
- and by the necessary presence of other non-removable extra reactions (e.g. reverse reactions).

These limitations are solved in practice by the second partner in the restricted setting of **logical circuits** [28] using essentially **rate-independent** [14,34] concrete CRNs. This approach led to the effective **making of biomedical chemical sensors in artificial vesicles,** as a *worldwide première* in [23].

Furthermore, the second partner has already made software tools that allows us to extract, from a reaction database, **a catalogue of connected enzymatic reactions** that relate any given metabolite inputs to outputs (see Fig. 2). The database contains kinetics information along with environment parameters (pH, temperature, ...) that are also extracted along with the output network.

c. Methodology and risk management

In our symbolic computation pipeline for compiling mathematical functions in abstract chemistry (Fig.1), the different steps are of different computational complexity and some of

them now require solving some open problems in order to scale-up towards concrete chemical implementations with more generality.

- The first step of symbolic differentiation is standard.
- The second step of polynomialization of the ODEs has been shown last year to be solvable for elementary mathematical functions by introducing a linear number of variables, with an algorithm of quadratic time complexity for functions with derivatives of linear size [13 plus presentation at CASC 2021, invited talk at SIAM Algebraic Geometry 2021]. This step can thus be considered as an easy symbolic computation task. Nevertheless, **different choices** for the introduction of different variables lead to **different PODEs** and may lead at the end to impossible concretization step (Task 1). Therefore it makes sense to investigate how the concrete catalog can be taken into account to constrain and guide the choices done as early as in that polynomialization step (Tasks 2 and 3).
- The third step of quadratization of the PODE in quadratic ODEs (QODE), i.e. polynomial ODEs with degree at most 2 (in order to restrict ourselves to elementary reactions with at most two reactants and mass action law kinetics) has the highest theoretical complexity and some open problems attached to it. It is folklore theorem that any solution to a PODE on some variable of interest is the solution of a QODE with potentially an exponential number of new variables.
    - The proof given by Carothers et al. in [35] shows that one solution exists by introducing variables for all the monomials that can be formed with the original variables and all degrees bounded by their maximum degree in the original system. We have shown in [17] that the associated decision problem to determine the existence of QODE with k variables is **NP-complete in the non-succinct representation** where the input is given in matrix form of (exponential) size containing all Carothers monomial. We conjecture in [17] that the quadratization minimization problem with auxiliary variables restricted to Carothers' monomials is **NEXP-hard**.
    - Nevertheless, we give in [17] a **MAXSAT encoding** of this optimization problem, improved in [13] by a **heuristics to choose some terms** for which auxiliary variables are introduced in order to efficiently compute solutions, at the cost of losing optimality in some cases. This is the algorithm currently implemented in Biocham which is shown in [17] capable of solving a benchmark of interesting CRN synthesis problems.
    - Interestingly, Bychkov and Pogudin in [1] have shown last year with a branch&bound algorithm that Carothers monomials may lead to suboptimal solutions, in the sense that QODE of smaller size may be obtained if one introduces variables for monomials of higher degrees than in the original system. Furthermore, Aluddin in [33] showed that even better solutions in term of size of the quadratic form can be obtained if one is allowed to **introduce variables for polynomials rather than just monomials**, without any hint however on how to find those polynomials, nor providing any algorithm for guiding the search. One ambition of BIOPSYN is to contribute to those theoretical complexity and algorithmic questions, especially by assuming given a catalog of allowed reactions and using it to guide the search for the polynomials that can be eliminated by introducing auxiliary variables (Tasks 2 and 3).
- The fourth step generates an abstract CRN with no guarantee that it can be implemented with a concrete one using real molecular compounds. In the case of rate-independent CRN, a limited but well understood class of piece-wise linear functions for which the

results is independent of the reaction rates [34,14], the tools presented in Fig.2 may solve the problem by using them as backend to the compilation pipeline (Task 1). When reaction rates matter however, the task may become difficult or even impossible to obtain a good approximation of the specified function (Task 1.2). Hence the idea of **taking the concrete catalog into account in the early phases of the compilation pipeline** (tasks 2 and 3), in a fashion meeting both the theoretical reasons and algorithmic open problem described above, and the practical needs to guide search in the early symbolic transformation steps towards concrete CRN implementations.

The Biomachines developed by Partner 2 are made of artificially designed biochemical networks encapsulated into liposomes, a type of vesicle. They are constituted of lipids bilayer containing an aqueous core. Because of their similarity with cell architecture regarding cellular membrane and size, as well as their capability of creating out-of-equilibrium. compartments, they have been used as models for artificial cells' studies and drug delivery systems.

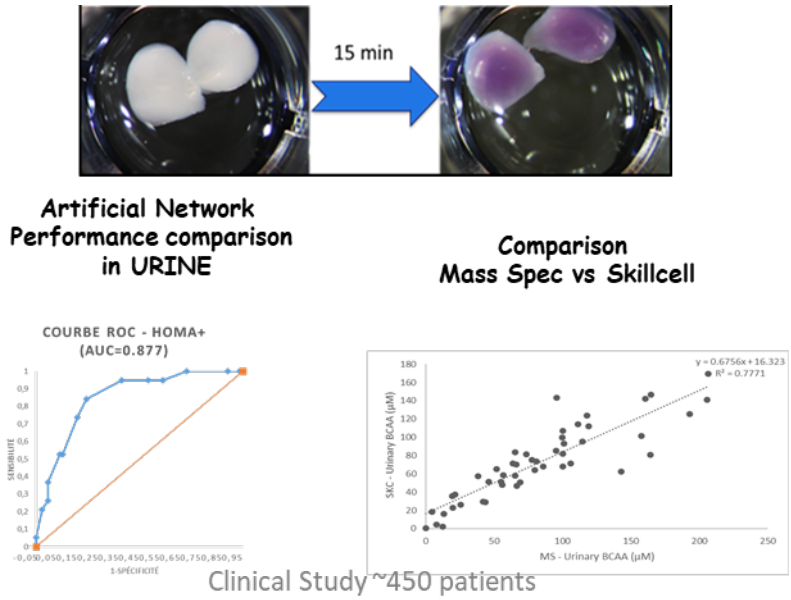

Figure 3*: Synthetic Artificial Biochemical network for biomarkers detection tested on raw clinical samples. The artificial networks are encapsulated in artificial vesicles (microfluidics) and entrapped in alginate gel beads to ease colorimetric response. The biomarkers detection compared with mass spectrometry is linear [32].*

Bulk methods present limited control of membrane formation process and liposomes produced by these methods are often polydisperse and multilamelar. Low encapsulation yield and identical inner and outer material are problems faced by bulk methods. Some examples of such methods are extrusion through porous membrane, freeze-drying, electro-formation and thin-film hydration [36, 37].

On other hand, microfluidics handles with fluids inside geometrically constrained channels in low Reynolds Numbers scale. Laminar flow, on the contrary to bulk method, allows precise control of lipid hydration process and thus production of micro/nano sized liposomes presenting monodispersity, membrane unilamelarity and high yield of encapsulation. Pulsed jetting and

flow focusing methods are widely used and developed amongst various methods for liposomes production. In our case, we use a specific two flow-focusing method to produce our "biomachines" [32] (see Fig. 3). At the end of the process solvent excess elimination is performed and the bilayers are stabilized. These are the methods well-mastered by Partner 2 that will be used to produce and validate our study cases *in vitro* (Tasks 4 and 5).

The BIOPSYN approach to those challenges will be evaluated on two study cases. The first study case will be the combination of a set of already designed, produced and well-characterized CRNs used in biomedical diagnosis, making their redesign in combination using BIOPSYN tools an innovative benchmark of our results with affordable validation. The second study case is a non-Boolean control system of homeostasis in an artificial vesicle which has never been done.

The tasks are organized as follows and summarized in the following Gantt chart:

Tasks	Partners		Year 1		Year 2		Year 3		Year 4	
	1 Inria Saclay	2 CNRS ALCEN	6	12	18	24	30	36	42	48
Task 1. Mapping from an abstract CRN to a concrete CRN (Fig 2 as backend to Fig 1)										
1.1	X	X		D1.1						
1.2	X	X				D1.2				
Task 2. Constrained polynomialization algorithms (Fig. 2 catalog as input to Fig.1)										
2.1	X					D2.1				
2.2	X	X						D2.2		
Task 3. Interactive concrete compilation pipeline										
3.1	X	X			D3.1					
3.2	X	X							D3.2	
Task 4. Study case design										
4.1		X	D4.1							
4.2	X	X				D4.2				
4.3	X	X						D4.3		
Task 5. In silico validation										
5.1		X						D5.1		
5.2		X								D5.2
Workshops					•				•	

Task 1 Mapping from an abstract CRN to a concrete CRN (Fig. 2 as backend of Fig. 1)

Objective:

This first task basically consists in connecting the concrete design tool of the second partner as a concretization backend to the abstract design tool of the first partner. That task is divided in two subtasks in order to distinguish between two levels of difficulty: first by relaxing reaction rate constraints, as possible in particular for rate-independent CRNs [14,34], then in a second step, by handling reaction rate requirements.

Risk management:

Both partners will work in tight cooperation on this task. They fully master their respective software tools and their connection should not raise other difficulties than the intrinsic

limitations of the static backend approach which motivates the constraint-based approach developed in the other tasks 2 and 3.

Subtask 1.1 Rate-independent CRN concretization

BrendaXplorer is a software tool that extract sub-networks from the reaction database BRENDA (https://www.brenda-enzymes.org). BrendaXplorer selects reactions from a preprocessed image of the whole BRENDA database, that connect a given set of input metabolites to a given set of output metabolites. The distance, number of reactions, between the inputs and the outputs is also a given parameter, that cannot be greater than 4. Other parameters, such as the pH or the temperature, can be used to get a subset of reactions that could work in the same environment. The kinetics of the reactions are present in BRENDA but can be hardly used because of the very large range of their values, depending on the organism from which the enzymes are extracted for example.

To obtain implementations of an abstract CRN, a first step will be to use specific sets of input and output metabolites in order to use BrendaXplorer as a filter to reduce the size of the output reactions set. Then search in this output network one or more subnetworks that match the input abstract CRN. Although finding an isomorphic subgraph in a graph is known to be NP complete, we plan to find specific characteristics of enzymatic reactions networks to speed up the process.

Subtask 1.2 Rate-equivalent CRN concretization

In order to take into account rate constraints in the abstract CRN produced, we will first select the most appropriate implementations found in subtask 1.1, and obtain the rates needed by optimizing the enzyme concentrations, either manually, using extensive simulations, or by using BIOCHAM optimization procedures as done in [23]. For the simulation approach, we plan to develop a completely new version of Partner 2 simulator HSIM, using highly parallel simulation algorithms, on *General-Purpose Graphics Processing Unit*, GPGPU (graphics processing unit that is programmed for purposes beyond graphics processing). This new simulation tool will be dedicated to run thousands of simulations of the same model with varying concentrations of enzymes and reactants.
This subtask will also investigate a theoretical notion of input/output rate-preserving CRN equivalence, possibly using CRN rewriting techniques, and generalizing model reduction techniques as studied in the former ANR-DFG Symbiont project.

Deliverables:
D1.1. Rate-independent catalog-based backend (M12).
D1.2. Rate-equivalent catalog-based backend (M24).

Task 2 Constrained polynomialization algorithms (Fig. 2 catalog as input to Fig.1)

Objective:

This task will study how the limited resources of a catalog of concrete reactions can be used to constrain the early ODE transformation of our compilation pipeline for the implementation of a mathematical function. It will involve one foundational symbolic computation subtask about one open computability and complexity question about the quadratization problem, and one implementation subtask fitting the need of our compilation pipeline towards concrete CRNs.

Risk management:

The first subtask ambitions to contribute to one open problem in symbolic computation and is thus risky. The idea is to get some intuition from our applicative context to design a novel quadratization algorithm based on the introduction of new variables for polynomials instead of just monomials as currently done. In case of failure to generalize this approach in an interesting algorithm, the second subtask will not implement such a general purpose algorithm but will adopt a more pragmatic attitude to guide the search towards a concrete quadratization as needed by our compilation pipeline.

This task will be mainly investigated by the first partner but in tight collaboration with the second partner to provide inputs and examples of solutions previously found by experience.

Subtask 2.1 Quadratization minimization algorithm

This subtask will tackle the challenge of designing a quadratization algorithm minimizing the number of QODE variables, without restricting to Carother's monomials, nor higher degree monomials, possibly in a restricted setting for constraining the introduction of auxiliary variables.

Motivated by the BIOPSYN context, the input of that quadratization algorithm might be not restricted to a PODE but an ODE in quite general form, including Michaelis-Menten terms or Hill terms of any order, making it possible to apply it right after the formal derivation step and before the polynomialization step of the compilation pipeline which could be integrated in a general quadratization algorithm.

Subtask 2.2 Catalog-restricted compilation pipeline

This subtask will formalize the constraints coming from the use of a fixed catalog of reaction with fixed kinetics, in terms of constraints for the symbolic transformation step of the abstract compilation pipeline and the quadratization minimization algorithm. A benchmark of CRN synthesis problems will be developed to illustrate the benefits of this approach on artificial examples.

Deliverables:
D2.1. Quadratization minimization algorithm (M24).
D2.2. Catalog-restricted compilation pipeline (M36).

Task 3 Interactive concrete compilation pipeline (Fig. 1 and Fig 2. in interaction)

Objective:

This task concerns the dynamic interaction between the compilation pipeline and the extraction of concrete reactions.

Risk management:

Both partners will work in tight cooperation on this task. They fully master their respective software tools. The design of an appropriate communication protocol for use in an interaction compilation pipeline should thus succeed once the constraint-based symbolic transformation steps are well understood.

Subtask 3.1 Query-answer protocol

This subtask concerns the definition of a query-answer protocol between BrendaXplorer and BIOCHAM in which the compilation pipeline is implemented. Up to now, BrendaXplorer has been designed to answer graph connectivity queries between molecular compounds. The reaction rate constraints considered in Task 1.2 and the polynomial algorithms investigated in Task 2 will lead to new kinds of queries that will be specified and implemented in this subtask.

Subtask 3.2 Interactive compilation pipeline

This subtask will automate the interactive concrete compilation pipeline using the query-answer protocol defined in the previous subtask.

Deliverables:
D3.1. Communication protocol between BIOCHAM and BrendaXplorer (M18)
D3.2. Interactive compilation pipeline (M42)

Task 4 Study case design

Objective:

This task will concern the design of two study case CRNs. The first study case is a set of CRNs already designed, produced and well experimentally studied in Sys2diag partner laboratory. It

will serve as tuning and benchmarking system to characterize theoretical and computational innovations of the project. In addition since already existing incremental CNR complexity will be tested as well as affordable experimental validation.

The second study case aims at designing a CRN capable of implementing control functions for ensuring the homeostasis of ph/Ions and O2 in an artificial vesicle.

Risk management:

Both partners will work in tight cooperation on this task. The first design subtask will be specified by Partner 2 and developed by both partners together using the tools developed in the previous tasks to ensure their usability and applicability in a presumably well-mastered setting.

The second design subtask for a non-logical circuit is more ambitious and risky because innovative, but also more representative of the contributions expected in BIOPSYN. Partner 1 will take inspiration of natural CRN found in cells for solving similar homeostasis functions in biochemistry, in order to specify the problem in abstract terms for the compilation pipeline. This precise specification will be done in tight collaboration with Partner 2 for the relevant choice of biochemical compounds that are most amenable to experimental validation.

Subtask 4.1 Study case #1 artificial CRNs for insulin-resistance in human urines

This subtask will be devoted to the definition and preparation of the study case #1 to evaluate experimentally BIOPSYN project innovations over a realistic time frame and to perform experimental validation of incremental CRN complexity and optimization. This study case #1 consists in two separated artificial CRNs which were already design to perform urine biomarkers detection to diagnose insulin-resistance in human normal (non-diabetic) population [18]. Insulino-resistance subjects are considered at risk to develop diabetes in the future. One CRN assays the fasting urine status of the tested subjects (return white color if fasting urine, red if not or if already diabetic). The second CRN assay by measuring quantitatively 3 biomarkers and according to programmed thresholds address the insulin-resistant status of the subject (blue color if insulin-resistant white of not). Figure 3 shows experimental results on clinical samples. Thanks to the project these two separate CRNs can be redesigned in order to be replaced by a unique larger CRN returning the same qualitative clinical colorimetric results. The great advantage here is that all the compounds of the CRNs, all the kinetics and experimental setup are well controlled by the partners. Hence these study case #1 will serve as innovation benchmark with an affordable validation system.

Subtask 4.2 Study case #2 design of O2 homeostasis CRN design ph/Ions homeostasis CRN design

Red blood cells, or erythrocytes, enable the transport of gases such as CO2 and O2 in the blood, taking up O2 in the lungs and releasing it into the tissues, where they uptake CO2 to be released in the lungs. Conjointly with their respiratory function, red blood cells also assume a buffer role in the blood, via both the Jacob-Stewart cycle and the haemoglobin protein. This participates in maintaining the homeostasis of the red blood cell. Red blood cell homeostasis also comprises a regulation of the osmotic pressure to prevent an excessive swelling or shrinking of the cell. Mechanisms involved in regulation of osmotic pressure include ion content and volume regulation. A few mathematical models of those mechanisms already exist. The aim of this task is to take inspiration from those mechanisms to specify and design a concrete CRN performing similar homeostasis function in a purely synthetic artificial vesicle.

Deliverables:

D4.1. Study case #1 input CRNs experimentally ready to test (M6)
D4.2. Study case #1 combined CRN design (M24)
D4.3 Study case #2 design (M36)

Task 5 In vitro validation.

Objective:

This task will concern experimental use the well known study case #1 to evaluate and validate experimentally theoretical and computational innovations of the project. As well we will demonstrate here the efficiency of the CRN design solutions proposed during the project. To this end incremental CRN design complexity will address with the CRNs combinations of study case #1. As well as CRNs design from scratch will be demonstrated with study case #2. Partner #2 (Sys2Diag) is specialized in controlled artificial cell design and production using microfluidics. CRNs are encapsulated into vesicles in order to perform biomarkers detection in complex human fluids (here urine) without disturbing CRNs functionalities while the relevant biomarkers cross the membrane to be processed by the CRN inside the vesicle [23, 25]. The CRNs perform the programmed tasks (including quantitative biomarkers measurements, threshold controls etc.) and return a final qualitative response either colorimetric or fluorescent. Microfluidics setup allow high throughput encapsulated CRN production and detection thanks to microscopes and camera. If required a set of encapsulated CRNs can be trap in large beads of gel to perform advantageously eye colorimetric evaluation.

Risk management:

The *in vitro* validation of the second study case has a high level of risk because it is driven by software propositions and common experimental conditions must be identified for all the compounds of the designed CRN. The Sys2Diag laboratory has however already a long experience to mitigate such problem in particular using CRN modeling to identify alternatives solutions compatible to CRN programing specifications. Furthermore, Partner 1 will keep ready to revisit the automated design and guide the search toward alternative designs taking into account more constraints.

Subtask 5.1 Artificial encapsulated study case #1 CRNs fabrication and functional validation

This subtask will perform metabolites and enzymes sourcing for further CRN experimental production. It will as well use microfluidics/microscopy/high-speed-camera setup to encapsulate and observe the functionalities of various version of the CRNs from study case #1. In addition, incremental CRN complexity validations will be done according to the specifications provided by the early tasks of the project.

Subtask 5.2 O2 homeostasis vesicle making and validation

This subtask will ensure metabolites and enzymes sourcing for CRN from study case #2. If requested, some enzymatic compounds can be synthesized and produced on demand. In case of failure of the experiments, some iterations with the compilation pipeline can be investigated to synthesize alternative designs more amenable to experimental validation.

Deliverables:
D5.1. Innovation validation on study case #1 (M36)
D5.2. Study case #2 making and validation (M48)

II. Organisation and implementation of the project

a. Scientific coordinator and its consortium / its team

Project-team Lifeware, Inria Saclay Ile de France, will partly assign to the BIOPSYN project

- François Fages, *Coordinator*,
  - Directeur de Recherche de 1ère classe Inria,
  - resp. EPI Lifeware,
  - part-time Prof. Chargé de cours at Ecole Polytechnique
  - *Prix Michel-Monpetit - Académie des Sciences 2014,*
  - *Prix la Recherche - Sciences et Technologie de l'Information 2019.*
- Sylvain Soliman,
  - Chargé de Recherche Hors Classe Inria Saclay.

- Mathieu Hemery,
  - Ingénieur qualifié, Inria Saclay.

Sys2Diag UMR9005 CNRS-ALCEN lab. will partly assign to the BIOPSYN project

- Franck Molina,
  - Directeur de Recherche de 1ère classe CNRS,
  - Dir. Lab. Sys2Diag
  - *Médaille de Bronze CNRS 2004*
  - *IAAM Scientist of the year 2019,*
  - *Prix de l'innovation University de Montpellier 2020,*
  - *Médaille de l'innovation CNRS 2020.*
- Patrick Amar,
  - Maître de Conférence Classe Exceptionnelle, LISN, Université Paris-Saclay,
  - en délégation CNRS a Sys2Diag,
  - convention de collaboration long terme entre le LISN et Sys2Diag.
- Martin Davy,
  - Doctorant en deuxième année, contrat CIFRE avec SkillCell.

Implication of the scientific coordinator and partner's scientific leader in on-going project(s)

Name of the researcher	Person.month	Call, funding agency, grant allocated	Project's title	Name of the scientific coordinator	Start - End
François Fages	7	ANR Difference	Complexity theory with discrete ODEs	Olivier Bournez	2020-2023
Sylvain Soliman	7	ANR Difference	Complexity theory with discrete ODEs	Olivier Bournez	2020-2023
Partners 1 and 2		Bilateral industrial contracts are not mentioned here			

b. Implemented and requested resources to reach the objectives

Partner 1: Inria Saclay EPI Lifeware

Staff expenses

Permanent staff (not funded by the project):

- François Fages, coordinator, will be assigned to the project for a total implication estimated to 11.8 months
  - He will co-supervise with Franck Molina the PhD thesis of the recruited PhD candidate dedicated to the project
  - will co-supervise with Sylvain Soliman the Post-Doc recruited for 12 months
  - and will co-supervise with Sylvain Soliman, Mathieu Hemery for the subsequent software developments in BIOCHAM

- Sylvain Soliman will be assigned to the project for a total implication estimated to 11.8 months
    - He will co-supervise with François Fages the Post-Doc recruited for 12 months
    - and will co-supervise with François Fages, Mathieu Hemery for the subsequent software developments in BIOCHAM
- Mathieu Hemery will be assigned to the project for a total implication estimated to 7.2 months
    - He will supervise the software developments made by the non-permanent staff recruited by BIOPSYN to ensure their best integration in BIOCHAM
    - possibly revisiting the architecture of BIOCHAM implementation to make it easy to maintain

One PhD candidate (funded by BIOPSYN)) will be recruited for 36 months (120ke)

- at the start of the project or during the first year of the project
- to work on Tasks 1, 2, 3 and 4
- with a strong background in fundamental computer science and computer algebra or symbolic computation
- and with an interest for computational systems biology
- work will be done in tight collaboration with Partner 2 on tasks 1 and 4

That PhD thesis will be fully focused on the objectives of BIOPSYN but with an effort to generalize the algorithmic advances made in the project to transform them in contributions in computer algebra of independent interest.

One Post-doc researcher (funded by BIOPSYN) will be recruited for 12 months (48ke)

- to work on Task 2
- with a strong background in computer algebra
- in order to generalize the symbolic computation algorithms developed for the compilation pipeline of the project

6 months of internship (funded by BIOPSYN, 3.6ke)

- will be assigned to some specific research questions of interest for the project
- related to tasks 2, 3 or 4 with high pedagogical value

Instruments and material costs

One laptop will be acquired for the non-permanent staff of the project (2ke)

Building and ground costs

None

Outsourcing / subcontracting

None

General and administrative costs & other operating expenses

Missions of the funded staff 4400e

The Inria administrative costs are of 23140e, i.e. 13% of the funded costs.

Partner 2:Sys2diag UMR 9005 CNRS / ALCEN

Staff expenses 182700€

Permanent staff (not funded by the project):
-Franck Molina, DR CNRS, will cosupervise phD thesis with François Fages to be recruited at INRIA. He will co supervise with Patrick Amar the research engineer to be recruited at Sys2Diag . In Particular for in silico CRN implementation andbiochemical and microfluidics methods in order to build real CRN experimentally. These CRN will be evaluated accordingly to project issues.
-Patrick Amar MdC will cosupervise IR to be recruited at sys2diag with Franck molina in particular in CRN in silico design. He will interact with IRIA team to design hybrid computational approaches between BIOCHAM and computing tools already design by him.

Staff recruited for the project :
-One research engineer biochemistry/biophysic for 40 months (182700€) will be in charge of implementing and experimentally validating the functionalities of the biological networks of interests (including microfluidics). He/she will be in charge of molecular source of the project (enzyme production for instance)

Instruments and material costs 32000€

Consumable: 30000 €
Biochemical and chemical reactive compounds, to build experimental networks (metabolites, enzymes, synthetic chemical compounds if needed), the vesicles (oils, microfluidic chips design), specific enzymes production and purification
Microfluidics consumables: Chips design compounds, Connectivity and pressure tools and joins, light filters, seringes, etc.
One PC computer (2000€)

Building and ground costs
None

Outsourcing / subcontracting
None

General and administrative costs & other operating expenses

-Travel: 4000€

Internationals meetings, consortium meetings

-Administrative cost from CNRS 13% (28431€)

Requested means by item of expenditure and by partner*

	Partner *Inria Saclay EPI Lifeware*	**Partner** *CNRS ALCEN Sys2Diag*
Staff expenses	171600	182700
Instruments and material costs (including the scientific consumables)	2000	32000
Building and ground costs	0	0
Outsourcing / subcontracting	0	0
General and administrative costs & other operating expenses	4400	4000
Administrative management & structure costs**	23140	28431
Sub-total	**201140**	**247131**
Requested funding 448271e		

III. Impact and benefits of the project

Dissemination of Results

The results obtained in BIOPSYN will be disseminated by several means:

- through journal publications (e.g. PLOS Comp. Biol, Natural computing, Bioinformatics, Theor. Comp. Sc., perhaps JACM, ...)
- through communication in international conferences of several disciplines, in particular in computational systems biology (e.g. CMSB, ECCB, ICSB, ...), synthetic biology (ACS, IWBDA, ICCAD…) and computer algebra (CASC, SIAM AG, ...)
- through the organization of one or two workshops on the topic of BIOPSYN , in association with an international conference and open to external participants,
- through the development of tool releases and most notably the open-source software BIOCHAM, and also BrendaXplorer;

- to students, through the different Master's courses taught by the partners,
- to the industrial collaborators of the partners, through direct contacts after agreement of the consortium;
- to general public through *videos on youtube, fête de la science, interstices, blog binaire*, and possibly beyond.

A Web site will be created at the beginning of the project with both a public space for external communication, and a private space (wiki) for collaborative work between the partners.

Exploitation of Results

The partners have complementary interests and beneficiate from the already existing industrial partnership of the Sys2Diag lab. to develop BIOPSYN solutions under industrial environments. A consortium agreement will be signed between the partners including INRIA CNRS and ALCEN since Sys2diag laboratory is a join lab between CNRS and ALCEN. See next section for the terms.

Partner 1, Inria Saclay EPI Lifeware, is directly involved in open-source software development for systems and synthetic biology, especially with the development of the BIOCHAM platform. The results of BIOPSYN will be valorized by integration in these software developments in BIOCHAM. Furthermore, the first partner has a long experience of collaboration with the pharma industry, initially through its active participation to the OSEO/BPI Biointelligence project (2005-2010), coordinated by Dassault-Systèmes, with a consortium composed of SOBIOS, SANOFI, IPSEN, SERVIER, PIERRE FABRE, and BAYER, and INSERM Genopole Evry; and recently renewed with two CIFRE PhD theses with SERVIER and Johnson&Johnson respectively. The first partner, further encouraged by the creation of the common Inria-APHP Bernouilli lab., thus benefits from privileged communication channels to valorize the results of the BIOPSYN project in compliance with the term of the consortium agreement and the position of the Sys2Diag lab. for biosensor exploitation. By the end of the BIOPSYN project, the development of the biochemical system design language and its use to engineer synthetic biosensors should thus lead to unique software developments of great interest on the international scene for the bioengineering, biomedical and pharmacological domains.

Partner 2, Sys2diag UMR9005 CNRS/ALCEN, is directly involved in the design of biosensors in particular for application in medical diagnosis. Sys2diag has industrial partnerships in this field (main ALCEN). In addition, Sys2diag is involved in synthetic biology for future clinical applications. The CNRS group has for instance the capacity to work on human samples and in tight relationships with clinicians. The BIOPSYN project is of key interest to enhance their capacity to design stable and non-toxic systems for further clinical use under ethics regulation compliance. CNRS and ALCEN already patented (April 2016; EU patent and US provisional patent CNRS/ALCEN) the principle of artificial biochemical network engineering for clinical application such as medical diagnosis. Today ALCEN group and CNRS develops together clinical diagnostic products based on this technology. Then ALCEN will naturally develop biotechnological applications issued from our BIOPSYN project.

Management of intellectual property

To protect the interests of all partners, a consortium agreement will be signed at the beginning of the project. This will settle any remaining organizational issues such as the role of the main investigators, the existing knowledge and intellectual property, as well as remaining issues of intellectual property and exploitation of results, according to the following general principles:

- The institution generating results and IPR is the owner of such results and IPR and is responsible for their legal protection and transfer. Results jointly developed shall be jointly owned.
- If the use of background IPR of a partner of the project is necessary to exploit the results and IPR generated in the project, the owner of such background IPR, if free to do so (no previous incompatible commitments), shall negotiate in good faith, at favourable conditions, access rights to allow the use of such background.

To this end a consortium agreement will be signed in between all the institutions involved in the project meaning CNRS, Sys2Diag, INRIA.

IV. References related to the project

Main external references:

1. Andrey Bychkov and Gleb Pogudin. Optimal monomial quadratization for ode systems. In Proceedings of the IWOCA 2021 - 32nd Int. Workshop on Combinatorial Algorithms, July 2021.
2. Marko Vasić, David Soloveichik & Sarfraz Khurshid. CRN++: Molecular programming language. Natural Computing volume 19, pages391–407 (2020)
3. Luca Cardelli , Mirco Tribastone, Max Tschaikowski. From electric circuits to chemical networks. Nat. Comput. 19(1): 237-248 (2020)
4. Bournez, O., Graça, D.S., Pouly, A.: Polynomial Time corresponds to Solutions of Polynomial Ordinary Differential Equations of Polynomial Length. Journal of the ACM (2017).
5. Corentin Briat, Ankit Gupta, Mustafa Khammash. Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. Cell Systems, 2(1), 2016.
6. L. Rizik, Y. Ram, and R. Danial. Noise tolerance analysis for reliable analog and digital computation in living cells. J Bioengineer & Biomedical Sci, 6(186), 2016.
7. Ramiz Daniel, Jacob R. Rubens, Rahul Sarpeshkar, and Timothy K. Lu. Synthetic analog computation in living cells. Nature, 497(7451):619–623, 05 2013.
8. Ho-Lin Chen, David Doty, and David Soloveichik. Deterministic function computation with chemical reaction networks. Natural computing, 7433:25–42, 2012.
9. K. Oishi and E. Klavins. Biomolecular implementation of linear i/o systems. IET SYstems Biology, 5(4):252–260, 2011.
10. Matthew Cook, David Soloveichik, Erik Winfree, and Jehoshua Bruck. Programmability of chemical reaction networks. In Anne Condon, David Harel, Joost N. Kok, Arto Salomaa, and Erik Winfree, editors, Algorithmic Bi processes, pages 543–584. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
11. Olivier Bournez, Manuel L. Campagnolo, Daniel S. Graça, and Emmanuel Hainry. Polynomial differential equations compute all real computable functions on computable compact intervals. Journal of Complexity, 23(3):317–335, 2007.
12. D.S. Graça and J.F. Costa. Analog computers and recursive functions over the reals. Journal of Complexity, 19(5):644–664, 2003.

References of the partners:

13. Mathieu Hemery, François Fages, Sylvain Soliman. Compiling Elementary Mathematical Functions into Finite Chemical Reaction Networks via a Polynomialization Algorithm for ODEs. In *CMSB'21: Proceedings of the nineteenth international conference on Computational Methods in Systems Biology*, *Lecture Notes in BioInformatics*. Springer-Verlag, 2021. [preprint]
14. Elisabeth Degrand, François Fages, Sylvain Soliman. Graphical Conditions for Rate Independence in Chemical Reaction Networks. In *CMSB'20: Proceedings of the eighteenth international conference on Computational Methods in Systems Biology*, *Lecture Notes in BioInformatics*. Springer-Verlag, 2020. [preprint] [slides] [video]
15. François Fages. Artificial Intelligence in Biological Modeling. In *A Guided Tour of Artificial Intelligence Research, Volume III: Interfaces and Applications of AI*. Springer-Verlag, 2020. [preprint]
16. François Fages, Franck Molina. La cellule, un calculateur analogique chimique. In *Approches symboliques de la modélisation et de l'analyse des systèmes biologiques*. ISTE, Ed Ltd, London, 2022. [preprint]
17. Mathieu Hemery, François Fages, Sylvain Soliman. On the Complexity of Quadratization for Polynomial Differential Equations. In *CMSB'20: Proceedings of the eighteenth international conference on Computational Methods in Systems Biology*, *Lecture Notes in BioInformatics*. Springer-Verlag, 2020. [preprint]
18. Francisco Santos Schneider, Patrick Amar, Asma Bahri, Julien Espeut, Julie Baptiste, Mellis Alali, François Fages, Franck Molina. Biomachines for Medical Diagnosis. *Advanced Materials Letters*, 11(4):1535–1558, 2020.
19. Elisabeth Degrand, Mathieu Hemery, François Fages. On Chemical Reaction Network Design by a Nested Evolution Algorithm. In *CMSB'19: Proceedings of the seventeenth international conference on Computational Methods in Systems Biology*, *Lecture Notes in BioInformatics*. Springer-Verlag, 2019. [preprint]
20. Huang, Wei-Chih, Jiang, Jie-Hong, Fages, François, Molina, Franck. Biochemical Threshold Function Implementation with Zero-Order Ultrasensitivity. In *15th IEEE Biomedical Circuits and Systems Conference*, 2019. [preprint]
21. Patent CNRS/SKILLCELL 2018-12-20 EP18214973,2 "Biodegradable biochemical sensor for determining the presence and/or the level of pesticides or endocrine disruptors: method and composition" F Molina, J Espeut
22. Adrien Baudier, François Fages, Sylvain Soliman. Graphical Requirements for Multistationarity in Reaction Networks and their Verification in BioModels. *Journal of Theoretical Biology*, 459:79–89, 2018. [preprint]
23. Alexis Courbet, Patrick Amar, François Fages, Eric Renard, Franck Molina. Computer-aided biochemical programming of synthetic microreactors as diagnostic devices. *Molecular Systems Biology*, 14(4), 2018. [preprint]
24. Fages, François, Soliman, Sylvain. On Robustness Computation and Optimization in BIOCHAM-4. In *CMSB'18: Proceedings of the sixteenth international conference on Computational Methods in Systems Biology*, *Lecture Notes in BioInformatics*. Springer-Verlag, 2018. [preprint]
25. Patent CNRS/SKILCELL « Method for the preparation of biosynthetic device and their uses in diagnostics EP16165189.8 , 13 avril 2017 PCT/IB2017/000683 17 avril 2018, Franck Molina, Alexis Courbet Francisco Schneider
26. Fages, François, Le Guludec, Guillaume and Bournez, Olivier, Pouly, Amaury. Strong Turing Completeness of Continuous Chemical Reaction Networks and Compilation of Mixed Analog-Digital Programs. In *CMSB'17: Proceedings of the fiveteen international conference on Computational Methods in Systems Biology*, pages 108–127, volume 10545 of *Lecture Notes in Computer Science*. Springer-Verlag, 2017. [preprint]
27. A. Courbet, E. Renard and F. Molina. Engineering next-generation diagnostics through synthetic biology. *EMBO Mol. Med.2016 Volume 8, Issue 9Pages 987–991*
28. Marc Bouffard. Conception, modélisation et simulation in silico d'un nanosystème biologique artificiel pour le diagnostic médical. PhD thesis. 2016
29. A Courbet, D Endy, E Renard, F Molina and J Bonnet. Detection of pathological biomarkers in Human clinical samples via amplifying genetic switches and logic gates. *Science Translational Medicine* 27May 2015:Vol.7,Issue289,pp.289

30. Tai-Yin Chiu, Hui-Ju K. Chiang, Ruei-Yang Huang , Jie-Hong R. Jiang, François Fages. Synthesizing Configurable Biochemical Implementation of Linear Systems from Their Transfer Function Specifications. *PLoS ONE*, 10(9):e0137442, 2015.
31. Stephanie Rialle, Liza Felicori, Camila Dias-Lopes, Sabine Peres, Sanaa El Atia, Alain R. Thierry, Patrick Amar, and Franck Molina. Bionetcad: design, simulation and experimental validation of synthetic biochemical networks. *Bioinformatics*, 26(18):2298–2304, 2010.
32. Francisco Santos Schneider, Patrick Amar, Asma Bahri, Julien Espeut, Julie Baptiste, Mellis Alali, François Fages, Franck Molina. Biomachines for Medical Diagnosis. *Advanced Materials Letters*, 11(4):1535–1558, 2020.

Further external references:

33. Alauddin, F.: Quadratization of ODEs: monomial vs. non-monomial. SIAM Undergraduate Res. Online 14 (2021). https://doi.org/10.1137/20s1360578 3.
34. Ho-Lin Chen, David Doty, David Soloveichik, Wyatt Reeves. Rate-Independent Computation in Continuous Chemical Reaction Networks. CoRR abs/2107.13681 (2021). Submitted to JACM.
35. Carothers, David C. ; Parker, G.Edgar ; Sochacki, James S. ; Warne, Paul G. Some properties of solutions to polynomial systems of differential equations. *Electronic Journal of Differential Equations (EJDE)*, 2005
36. van Swaay D.; deMello A.; Lab Chip., 2013, 13, 752. 6.
37. Sato, Y.; Takinoue M.; Micromachines (Basel), 2019, 10(4).