SiliCell initiative : "In Silico Cell Simulation".

Modelisation and simulation of functional pathways

Pierre Mazière[#], Claude Granier[#], Isabelle Mougenot, Thérèse Libourel*, Charles Auffray[§], Bernard Pau[#] and Franck Molina[#]*

[#] CNRS UMR 5094, Institut de Biotechnologie and Pharmacologie, Faculté de Pharmacie, 15 Av. Charles Flahault, Montpellier, France.
*** CNRS UMR 5506, Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier. 161 rue Ada, 34392 Montpellier, France.**
[§] CNRS ERS 1984, 15 rue Guy Moquet, Villejuif, France.
franck.molina@ibph.pharma.univ-montp1.fr

Key Words: functional pathways, annotation and simulation, insulin signaling.

The progress in the human genome sequencing initiatives and the explosion of the genomic and proteomic approaches lead to an accumulation of a large amount of complex and heterogenous biological data. In order to deal with that new kind of data and to help their understanding it is an urgent need to create and use new bioinformatic tools. Our goal is to build a bioinformatic environment for the modelisation of functional pathways involved in cell signaling and regulation : ***SiliCell.*** To simulate the biological function of each object, we associate our "object oriented" knowledge database with programmed "biological roles". Unlike biological experiments, this bioinformatic tool will offer to the biologist the unique way to study the virtual global functioning of a given pathway. On the other hand, our tool *SiliCell* will provide to the biologist a computer assisted generator of hypotheses. Interactive "virtual experiments" (knock out, activation inhibition, etc...) on functional pathways and checking their potential consequences will be possible. Application to insulin signaling : Starting from experimental results already obtained in our laboratory, we are currently developing a knowledge database on the main cellular mechanisms potentially involved in insulin signaling involved in diabetis. Several functional pathways like apoptosis, DNA repair, cell signaling etc ..., are closely related. Different functional pathways have in common some cellular factors and only a powerful bioinformatic tool will help us to integrate all the biological data and understand their meaning in a broad and complex biological context.

Introduction

The progress in the human genome sequencing initiatives and the explosion of genomic and proteomic approaches lead to an accumulation of a large amount of complex and heterogenous biological data. The main challenge is to try to understand at the molecular level what are the differences between normal and abnormal behaviors of biological cells. The cell behavior is the result of a large amount of inter-dependent biochemical reactions between various biomolecules (proteins, DNA, RNA etc....) which are organized into biological functional pathways. In order to deal with the complexity of these data and to help their understanding we need to create and use new bioinformatic tools for global system biology.

This field of global system biology is still largely unexplored although some initiatives in world show the growing interest of the scientific community (Weng et al. 1999). However, most of these works try to simulate metabolic networks of small molecules (for instance the glycolysis). Today, with the increasing discovery of genes and proteins involved in diseases, the biologist needs a tool to simulate proteins and genes functional pathways. Recent large projects are going into that direction (Karp et al. 2000 ; Tomita et al. 1999) but unfortunately are based on prokaryote models (*E. Coli* or *Mycoplasma genitalium*) with only a limited interest for human diseases.

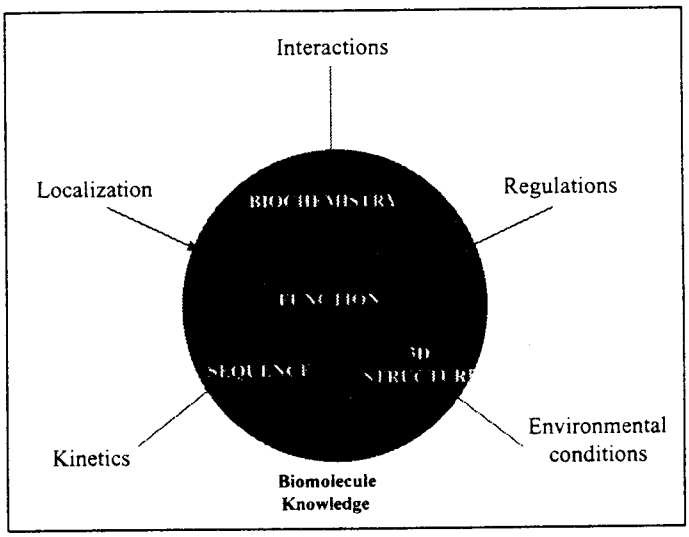
We describe herein, the strategy we chose in order to make available a maximum of biological and pharmacological knowledge and to extract from it the biological meaning. Our initiative is called *SiliCell* (in silico cell simulation). This environment for the modelisation and the simulation of functional pathways is dedicated to the biologists. This tool is designed as a generator of hypothesis and a decision helper for experimentalists. Thanks to the simulation of functional pathways provided by *SiliCell* it will become possible to test "in silico" the consequences of virtual biological experiments (knock out, mutation, inhibition etc...), therefore reducing the number of costly biological experiments by anticipating on the possible effects on functional pathways modifications.

Treatment and organization of biological data.

The first part of our project consist to develop an original "object oriented" knowledge database to combine already available and experimental biological knowledge. The object oriented structure of such a database combines the flexibility of manipulation and the

ease to handle the great complexity of biological data. For a given biological functional pathway we consider each protagonist as a discrete biological object.

Each biological object is considered through its biological function. A categorization of the proprieties and biological functions allow us to reduce the number of object classes described in our knowledge database. In order to standardize our representation of the functional knowledge we try to develop new tools based on XML standard. However, experimental data are often partial, not always reliable and not always fully understood. We take in account such a problem by providing a confidence scale on which we can evaluate the error associated with the biological data.



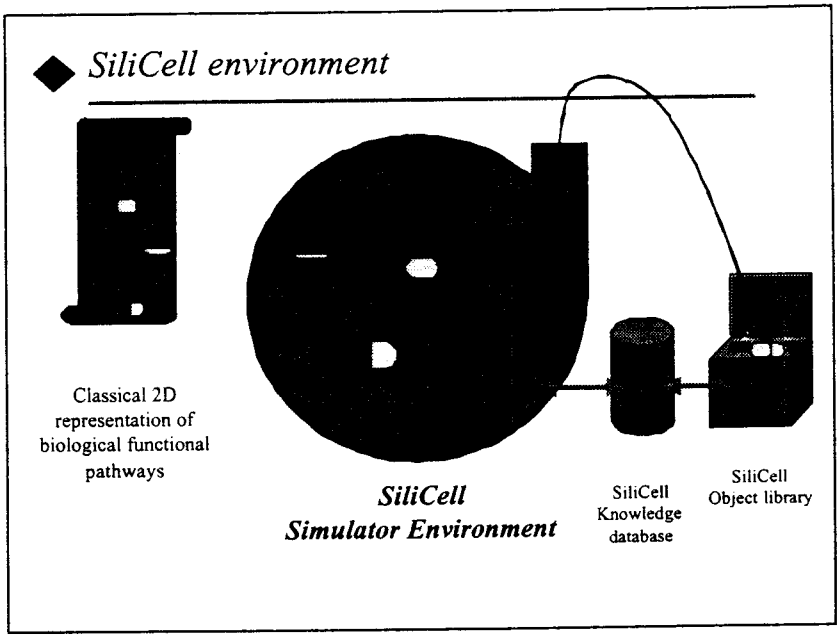
Knowledge representation in the SiliCell database.

In order to implement and annotate our functional database under a controlled process we are using a semi-automatic procedure combining systematic parsing from other databases (Pietu et al. 1999a ; Pietu et al. 1999b) and manual annotations. In order to guaranty the relevance of the data implemented in our knowledge database, we rely on a group of biologists expert in their field.

Modelisation and simulation of functional pathways and cell signaling

To simulate the biological function of each biological object, we associate our "object oriented" knowledge database with programmed "biological roles". In that respect, we split complex biological function into basic elements.

A biological function could be considered as a complex series of basic biochemical events. This decomposition could be applied to a large number of biological functions and the parameters needed by each basic biochemical event are directly accessible through the knowledge database.

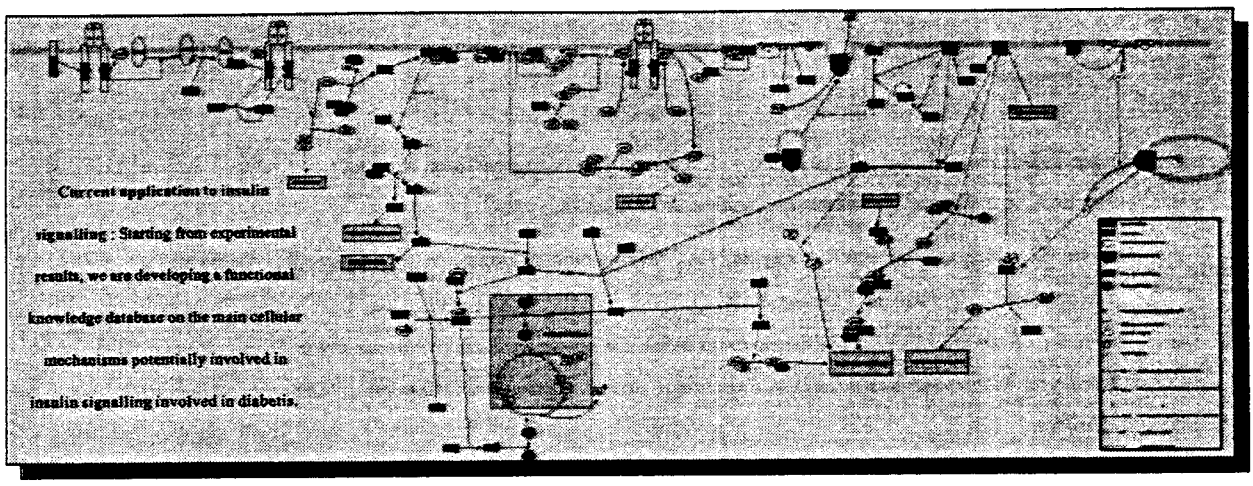


Global organization of SiliCell environnement compared to classical 2D representation of biological functional pathways.

The combination of a group of "functional biological objects" will allow us to simulate in a dynamic way one or several functional pathways in a separate cell compartment. Cross-talks between cell compartments will follow the cellular flux of the biological object populations. Our system will allow the user to follow the behavior of each individual object or population of objects. This approach allows us to simulate functional pathways into a specific biological environment without necessarily simulating the whole cell functioning.

Application to insulin receptor signaling

Starting from experimental results already obtained in our laboratory, we are currently developing a knowledge database oriented to the main cellular mechanisms potentially involved in insulin signaling through its receptor. We aggregate a great amount of functional knowledge on each biological protagonists and we standardize the way to provide an interactive 2D representation of the functional pathways.



This signaling cascade involves closely related functional pathways. Some cellular factors are in common between different functional pathways. Only a powerful bioinformatic environment could help us to integrate all the biological data and their associated dynamic behavior. This will allow us to understand their function in a broad and complex biological context and to identify new targets for future drug discovery.

Conclusions

Unlike biological experiments, this bioinformatic tool will offer biologist a unique way to study the virtual global functioning of a given pathway and meanwhile will make available a detailed historical trace of the events associated to each biological object.

On the other hand, our tool *SiliCell* will provide to the biologist a computer assisted

generator of hypotheses. The experimentalist will be able to do interactively "virtual experiments" (knock out, activation inhibition, etc...) on functional pathways and to check their potential consequences.

This new kind of bioinformatic environment for dynamic integratio n of biological data and simulation should become the most efficient tool for the understanding of functional pathways and the conception of strategy for pharmacological actions.

References :

Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Paley, S. M., Pellegrini-Toole, A. (2000)
The EcoCyc and MetaCyc databases ; Nucleic Acids Res 28, 1: 56-9

Piétu, G., Mariage-Samson,R., Fayein,N-A., Matingou,C., Eveno, E., Houlgatte,R., Decraene, C., Vandenbrouck, Y., Tahi, F., Devignes, M-D., Wirkner, U., Ansorge, W., Cox, D., Nagase, T., Nomura, N. and Auffray , C. (1999a) *The Genexpress IMAGE Knowledge Base of Human Brain Transcriptome : a Prototype Integrated Resource for Functional and Computational Genomics.* Genome Research 9, 195-209.

Piétu G, Eveno E, Soury-Segurens B, Fayein N-A, Mariage-Samson R., Matingou C, Leroy E, Dechesne C, Krieger S, Ansorge W, Reguigne-Arnould I, Cox D, Dehejia A, Polymeropoulos MH, Devignes M-D. and Auffray C (1999b) *The Genexpress IMAGE Knowledge Base of Human Muscle Transcriptome : A Resource of Structural, Functional and Positional Candidate Genes for Muscle Physiology and Pathologies.* Genome Research 9, 1313-1320.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J. C.Hutchison, C. A., 3^{d} (1999) E-CELL: software environment for whole-cell simulation ; Bioinformatics, 15, 1 : 72-84.

Weng, G., Bhalla, U. S., Iyengar, R (1999), Complexity in biological signaling systems ; Science 284, 5411 : 92-96.