CHALLENGES, LIMITS AND OPPORTUNITIES IN SYSTEMS BIOLOGY

OCTOBER 2021

FACING THE COMPLEXITY OF LIFE

Rather than another review of systems biology, this first paper is a personal statement of my transition from engineering to biology, two fields I am passionate about. However far apart, the two worlds have many points in common and can benefit greatly from each other. The key to success is to find the right strategy to cope with the complexity of living systems. Let's bridge the complexity gap together.

ARTICLE BY :
Arnaud Bonnaffoux - CSO Vidium

WHITE PAPER

LIFE COMPLEXITY AND PROMISES OF SYSTEMS BIOLOGY

"

"Hard sciences" used in engineering (mathematics, physics, informatics, etc.) and biology are in a mutual and ambiguous relation of admiration/repulsion. They hardly manage to work together efficiently. This segregation is due to the gap of complexity. Vidium is our answer, our commitment, to bridge this complexity gap. We bring engineering to biology to master life complexity

"

Biology is the science of the century. Biology needs to continuously upgrade in order to face most pressing health and environmental issues such as climate change, the arrival of new pandemics, the increasing incidence of cancer cases, the aging, the detection of hitherto incurable diseases.... In such a scenario, state of the art technologies is emerging and being developed at a fast pace (cell therapies, RNA vaccines, single cell sequence…), but the development of new treatments is still long and uncertain. This is due to life complexity.

An organism is composed of billions of heterogeneous cells interacting between themselves, each cell influencing each other, some cells having the ability to become another cell type.

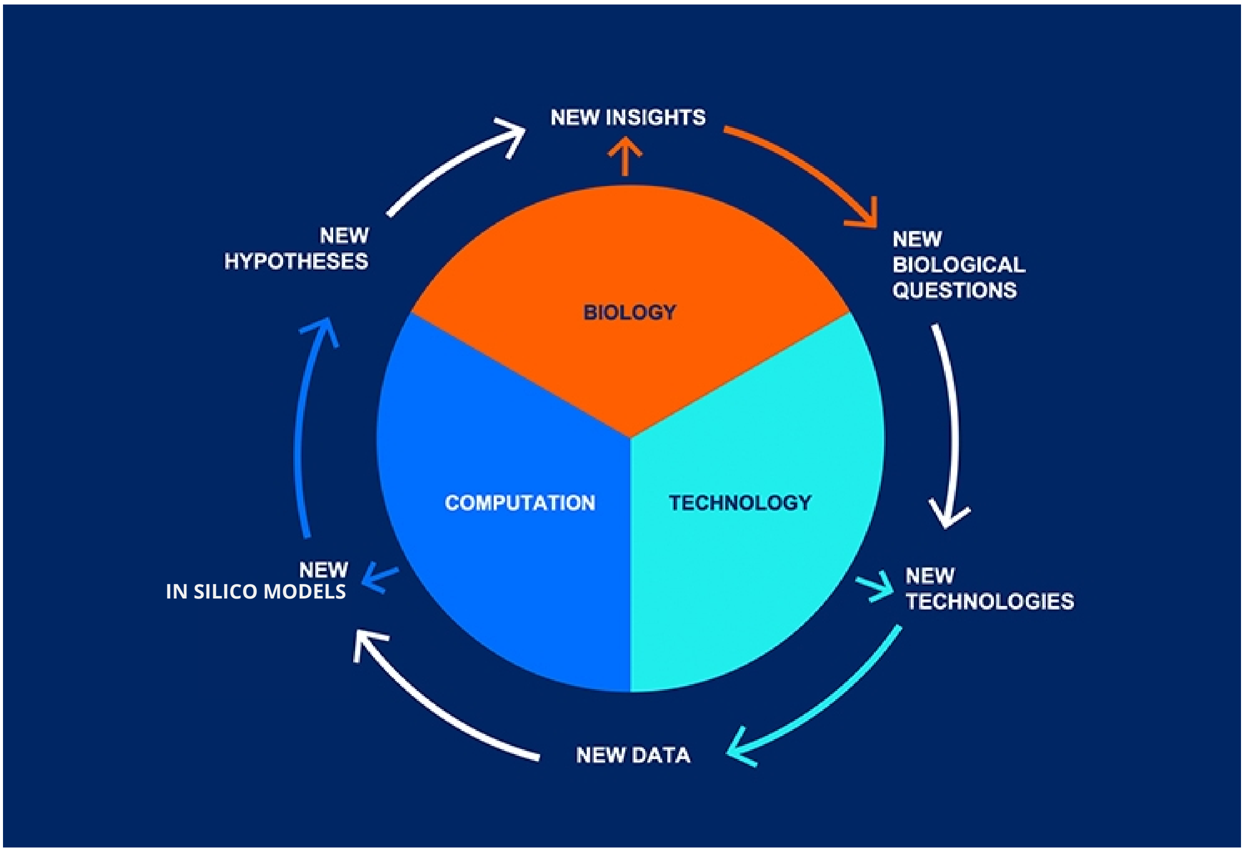
Each cell is *per se* an autonomous living system, composed of billions of interacting molecules in an ordered and stochastic process. In other words, cell function result from the collective behavior of many molecular parts, all acting together.

Whereas an integrated complex system such as that of a modern autopilot can be understood from its engineering design and detailed plans, attempting to understand the integrated system of a complex living system requires to infer all the interactions; all of which must be deduced *a posteriori* from the behavior of the system. Thus, the global behavior of living systems cannot be reduced to the sum of its parts. This reductionist vision has been replaced by a global and systemic approach, the systems biology.

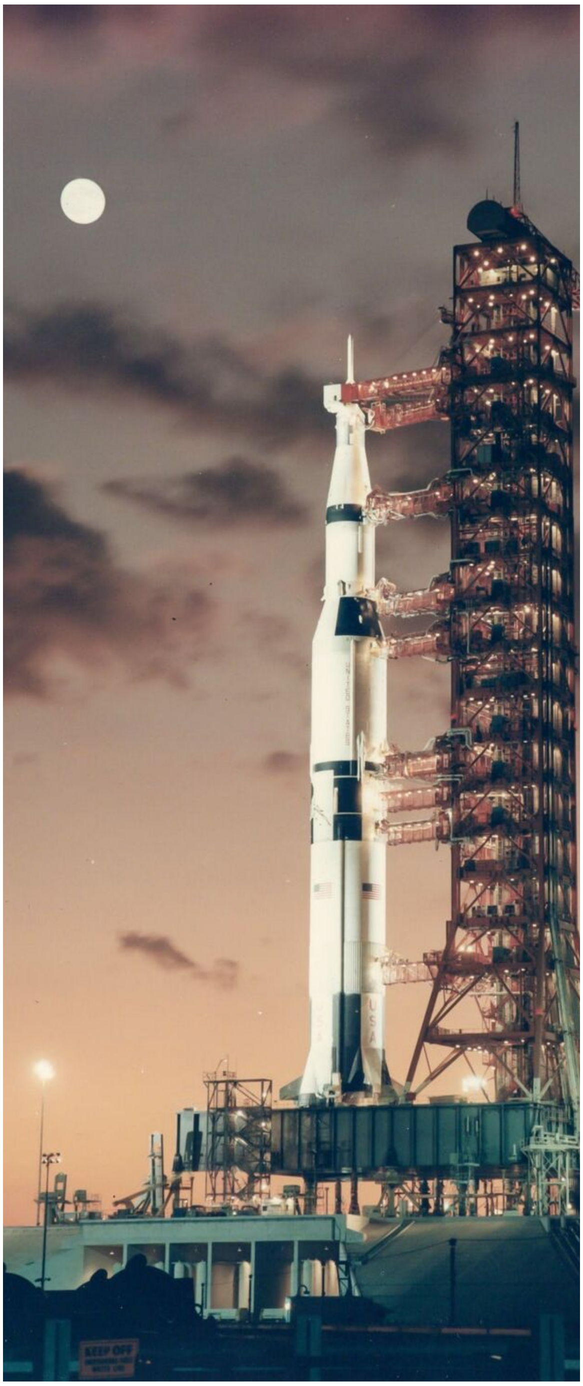
The goal of systems biology is to analyze a biological system considering all its parts, from an "integrated systems" point of view. This is possible with the arrival of omic data, which corresponds with the development of systems biology. Systems biology aims to be a rational iterative approach which addresses a biological question, analyzing and integrating a set of omic data, with the use of mathematical and computational models. These models can be used to make predictions or validate/invalidate biological assumptions and to propose new experiments to guide the biological research.

In spite of some successes in the science of human health and environmental sustainability, systems biology is still limited in its applications. Systems biology has not yet developed its full potential, but it should be the unique and natural way to study biological systems. Life's complexity is VERY important, we do not have enough experimental data or computational means to analyze it. Nonetheless, to Arnaud "the main limit of system biology comes from our scientific culture".

> *"The main limit of systems biology comes from our scientific culture"*



THE PARADOX OF COMPLEXITY VS MEANS OF ANALYSIS



On one side, systems biology is still poorly applied in biology. Many biologists still do not apply this systemic approach, preferring a reductionist one. The main reason is that biologists are often not trained nor comfortable with “hard” sciences (mathematics, physics, computation, ...). On the other side, many laboratories which apply systems biology counts on statisticians or physicists rather than biologists, that have found in biology a panoply of possibilities in which to apply their own science.

There exists this opposition between “hard” sciences with their rigid formalism and powerful theories, against the “soft” science of biology (and in fact the very opposite of soft), where life cannot easily be reduced to a set of equations. There is in fact a gap of method where there should be a continuum based on the synchronization of cross-disciplinary science. However, despite this gap, there is a mutual attraction and fascination from both sides

To design and build an aircraft, thousands of engineers are needed to work on the same project, organized per disciplines, using the latest technologies in mathematics and computation, but not only. All these people can work and interact efficiently thanks to specific processes and the intensive use of computational modelling. The domains of systems engineering and computational models were developed during the space conquest era.

The excellent movie “Hidden figures” perfectly illustrates the first use of digital computation to overcome the limits of mathematics to solve a complex problem (at that time) i.e. the estimation of the critical trajectory to return to earth. Since then, these domains were extended to others industries, like cars, aircraft or energy.

In reaction to the paradox of complexity vs means of analysis, many biologists will argue that and aircraft, or even a space station, is certainly very complicated, but not complex. A complicated system is the sum of all its parts; each sub-system has been designed to play a specific role. In other words, there is no emergent property coming from the interaction of its parts. It is a reductionist approach.

Contrary, attempting to understand the integrated system that is a biological organism, a complex system, is far more difficult. Understanding the individual parts alone is insufficient to understand or predict the biological system’s behavior. A fly is far more complex than a space station. But, if so, why biologists are split in hundreds of isolated and small research teams with an important turn over to study a complex system as a fly (drosophila)? Why analysis of omic data is generally run on a simple laptop? Why there is no well shared and formalized model for the formation of a fly's wing? Can we imagine engineers sharing their results writing and reading hundreds of papers?

The fact is that biology, which certainly study the most complex systems in the universe, is the most backward science in the use of analysis means. This paradox must be resolved.

LIMITS AND OPPORTUNITIES IN SYSTEMS BIOLOGY

"

We are drowning in a sea of data, and starving for knowledge!

"

Sydney Brenner formulated this paradox with other words: “We are drowning in a sea of data, and starving for knowledge!” This quote sums up the limits and opportunities in biology.

Hereinafter, we are going to develop 3 related aspects:

- **Comprehensive & integrative data analysis beyond statistics**
- **Knowledge capitalization**
- **Integrative trans-disciplinary process**

AI GLASS CEILING

Pure statistical approaches, which generally are aggregated in buzz words like "data sciences" or "AI", are the mainstream approach to analyze complex experimental data. This is a consequence of multi-disciplinary teams mainly composed of biologists and statisticians. These descriptive approaches have the big advantage to require no prior knowledge to cluster, classify, mix, order and compare experimental results. For example, we can find significant differences in gene expressions caused by a perturbation or being linked to a specific condition or phenotype.

These methods are particularly well suited for diagnostic or the identification of markers. However, statistical approaches rely on pure correlation analysis, which is not causation. Thus, they do not create knowledge, only black boxes. They cannot predict what has not been already observed. They are qualitative, not quantitative. T**his is the AI glass ceiling**. No matter the amount of big data or computational power, statistic cannot predict the effect of a new treatment, despite many claims (1).

KNOWLEDGE DRIVEN DATA ANALYSIS

Knowledge driven data analysis is an alternative to pure statistic approaches to capitalize knowledge. The fundamental difference is the analysis of **smart** (vs big) data with prior knowledge to generate more knowledge. Causalities of mechanistic processes, such as biochemical reactions, are **formalized** in mathematical equations and implemented in **executable** in silico models.

These in silico models, named "digital twins", are used to integrate and cross-validate heterogeneous data, to assert biological assumptions or to make predictions in new conditions.

	ARTIFICIAL INTELLIGENCE	KNOWLEDGE DRIVEN
DIAGNOSTIC	++	+
PREDICTION OF NEW TREATMENTS	-	+
GENERATION OF NEW KNOWLEDGE	-	+
IIERATIVE (EXPERIMENT ↔ MODELING)	-	+
SCALABLE	++	+
PRIOR KNOWLEDGE REQUESTED	NO	YES
QUANTITATIVE (VS QUALITATIVE)	-	+
KNOWLEDGE CAPTIALIZATION	-	+
USE OF DATA	BIG DATA	**SMART DATA**
BIOLOGICAL INTERPRETATION	BLACK BOX	**WHITE BOX**

WHAT WE LACK

HOW TO FORMALIZE KNOWLEDGE

It is crucial to formalize knowledge to enable its sharing, capitalization and exploitation. Currently, knowledge in biology is widely formalized with text or pictures (like for pathways). Texts and pictures are subjective and can hardly be capitalized. In engineering, computational models are shared and can be assembled like building blocks, to simulate a full aircraft model. Though, engineers do not "know" all the aircraft model parts, but they can use and aggregate sub-models from the other teams into their own model to check consistency and to test their sub-model in a realistic environment.

As a matter of fact, biologists spent a lot of time reading papers. The good news is that solutions already exist! The best Systems biology solution to formalize biological systems is SBML (2). This language can be used for either descriptive or executable use. It is linked with a graphical representation to have an intuitive and user-friendly representation of data. Note that graphical coding has been developed and extensively used 40 years ago in aeronautics (3). Intuitive, ergonomic and efficient tools exist (1) in biology and should be more used to formalize biological knowledge.

Formalization enables knowledge capitalization, which is crucial. Aeronautic started with aircraft models based on 4 equations implemented electronically with capacitors and resistances, way before the arrival of first computers. Now, aircraft models are made of millions of code lines. This is the result of years of knowledge capitalization. Biology shall start now!

Formalization of biological processes requires the ability to read and write mathematics, but NOT to solve equations. Equations will be solved digitally by computers. This is good news for biologists who are not comfortable with mathematics. Their responsibility is to write, or at least read and validate, mathematical equations from their systems of interest. As a biologist, we cannot let this responsibility to mathematicians, otherwise they would over-simplify the model to solve it mathematically, since this is exactly their job! When biologists can formalize, criticize and up-grade their white-box in silico model, they are back on the center of the game. Biologist should be aware that almost all mathematical problems are not solvable. By instance, in an aerospace engineering department there are (almost) no mathematicians, but certainly software developers.

"

When biologists can formalize, criticize and up-grade their "white-box in silico" model, they are back on the center of the game

"

Once again, the first application using a computer to solve equations was during space conquest, as shown in the movie “Hidden Figures”. The main advantage of these solutions is that we don’t need to over-simplify our fundamental assumptions to solve them. It just requires a certain computational power, and today we have a lot of computational power. Biologists have the power and they should use it.

Another important consequence of the knowledge driven approach is the capability to generate smart data with a design of experiment (DoE) iterative process. The goal of DoE is to design the best informatif, cost-effective and simplest next experiment to validate the in silico model.Knowledge drives data generation. Again, this is a fundamental difference with opportunistic AI approaches applied to pre-existing data. An efficient DoE requires tight cooperation between scientists, experimental platforms and in silico modelers. Experimental technologies in biology are developing at a fast pace. They provide unprecedented accurate information, but at the same time they are more complex and only biological platforms can master them. Vidium brings all its expertise and experience to connect biologist and platforms for a rational and successful research.

CHALLENGES, LIMITS AND OPPORTUNITIES IN SYSTEMS BIOLOGY

VIDIUM SOLUTIONS

We have a very talented and human international team, where everyone has their say, to contribute to the continuous improvement of our solutions. Each and every one of our members are PhDs with deep trans-disciplinary expertise in the areas of engineering, data-science, biology, genomics, virology, immunology, medicine... We are united towards a common goal: to shift the paradigm of drug discovery and to bring new horizons in the area of medicine and life science in a sustainable way.

A QUESTION?
AN INFORMATION?
AN IDEA TO SHARE?

Let us guide you through the "Life conquest"

E-MAIL US AT CONTACT@VIDIUM-SOLUTIONS.COM

REFERENCES

1. Adam, G., Rampášek, L., Safikhani, Z. et al. Machine learning approaches to drug response prediction: challenges and recent progress. npj Precis. Onc. 4, 19 (2020).

2. www.sbml.org

3. https://en.wikipedia.org/wiki/History_of_CAD_software

OCTOBER 2021

/vidiums

/vidium-solutions