European Journal of Molecular & Clinical Medicine

ISSN 2515-8260 Volume 10, Issue 07, 2023

Review Article

The challenges of in-silico in Biotechnology

Nutan Kumari[1*], Abhishek Kumar Mishra[1], Rahul Pant[2]

1. Department of Biotechnology, Noida International University, Greater Noida (U.P.) India

2. Institute of Pharmacy, Bundelkhand University Jhansi (U.P.) India

Mail for correspondence: meet2nutan@gmail.com

Abstract

With the development of high-throughput techniques like genomics and proteomics, researchers are now able to examine cells as systems. These are not only does this produce a completely novel set of logistical challenges, but it also forces a philosophical reevaluation of the idea of cells as a grouping of distinct biological components. What are we going to do with this growing list of cellular components and their characteristics? These lists, as useful as they are, essentially provide us with the chemicals which make up cells and each one's unique chemical properties. How can we now translate these exhaustive lists of chemical constituents into the biological characteristics.

Keywords: In silico, Biotechnology, In silico modeling, Docking.

Globalisation follows reductionist theory

Bioscience became influenced by simplified methods in the second half of the 20th century, which were successful in revealing information regarding individual cellular components and their activities. The development of genetics has greatly accelerated up the entire process during the past decade. We are constantly defining the gene portfolios of creatures whose full DNA sequences are now available. We can soon anticipate the assignment and verification of function for the bulk of the genes on chosen genomes, despite the fact that functional assignment to these genes is now insufficient.

The development of what essentially amounts to a "parts catalogue" of cellular components in a wide number of animals will subsequently most likely be accelerated by extrapolation between genomes.We have the capacity because of technologies like expression arrays and proteomics.But it's now widely known.

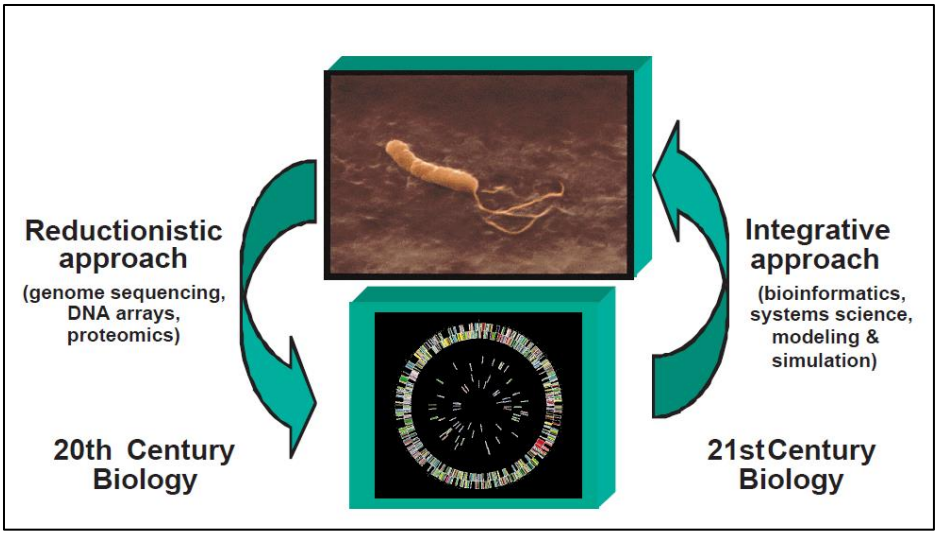

Figure 1. The shift in emphasis of biological research. Biology has traditionally followed a reductionist approach in which individual components of a living system are studied separately. It is becoming clear that we need to reverse the process and to study how these components interact to form complex systems using an integrative approach.

That integrating the activity of several gene products has grown to be a important topic for the advancement of biology. Bioinformatics and systems analysis techniques will be used in this integrated analysis. Thus, it is anticipated that the biological sciences will concentrate more on the systems aspects of cellular and tissue functioning throughout the ensuing years and decades. These are the "real" biological features that result from the system as a whole.

Because they develop from the whole and are not inherent features of the individual parts, these features are sometimes referred to as "emergent" characteristics. Several fundamental scientific issues and consequences for this

In silico Lifescience

Systems mathematics will be used more frequently in biological sciences, as it has previously been utilised in other sciences as well as computer simulations. This trend has already started, and it will probably continue. Systems science and challenging mathematical simulations have advanced to a high level of sophistication in a wide variety of different scientific and engineering domains. These abilities have an impact on how we live. A telephone call enters a sophisticated and well-optimized network. The refineries and other highly integrated chemical processes with complex control systems that are comparable to those of living cells are the source of the chemicals that we all utilise. Pilots undergo training in simulators, and the aerospace sector no longer constructs prototypes since computer simulations of aircraft designs are now so exact. new task include. Only a few years ago, this would not have been possible. Thus, quick data

production, analysis, model construction, and computer simulation have been productive stages in many sectors of science and engineering.

Why not consider biology? Will it ever develop mathematical modelling and simulation to a level of sophistication comparable to other fields?

The opinions are divided. Several individuals are suspicious about the efficacy of such initiatives due to the complexity of biological systems and how they are always changing as a result of evolution. Of course, only time will be able to determine their level of success.

As the Human Genome Project nears completion and more and more expression data become available, in silico biology is receiving more and more attention.

In silico biology is the general phrase for using computers to conduct biological research. Currently, it is common practise to compute the structures of complex biomolecules. Many believe that in the coming decades, biology will be dominated by the mathematical description and computer simulation of the simultaneous activity of several gene products. This field is now gaining relevance.

What will we do next?

Building mathematical models in biology is likely to be different from doing so in physical sciences, at least at first. Basic rate equations like the diffusion equation, fundamental rate ideas like chemical potential, and the fundamentals of electrochemistry like the Nernst equations are where one should start while studying these fields. These equations have a huge number of parameters, the majority of which can be measured individually, and are based on fundamental physical ideas and principles. Computer models of complicated processes contain data on both the individual characteristics of each system component and their interrelationships.

Despite the great bioinformatic databases, we are unable to gather all the data required to create a computer model of a whole cell at this level of detail.

This objective might be reached in the future, but for now, if we want functional and practical computer models of complete cells, a different strategy is required. Currently, we can determine the network structure of multigenic processes (for example, using stoichiometry and yeast two-hybrid systems), but it is much more challenging to learn about the physicochemical characteristics of gene products, such as binding constants and turnover rates.

An alternate strategy can be developed in the absence of specific information and is based on the idea that cells are subject to limitations that restrict the behaviours that they can engage in. One can then decide what is and is not possible for a cell by imposing these limits. One can limit likely cellular behaviour by imposing a set of constraints, but one can never anticipate it with accuracy. Figure 2's left side is an illustration of this strategy.

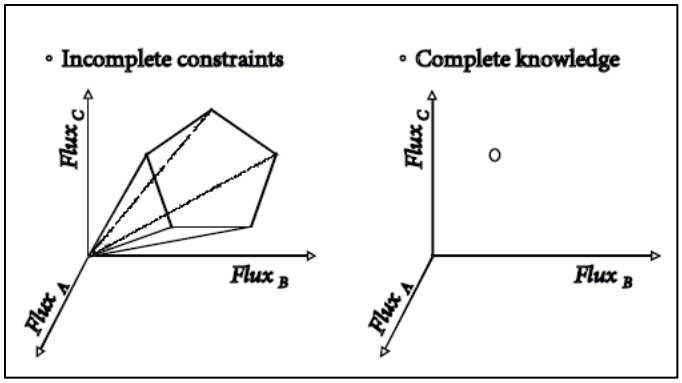

Figure 2. Constraining possible behaviors. Because biological information is incomplete, it is necessary to take into account the fact that cells are subject to certain constraints that limit their possible behaviors. By imposing these constraints in a model, one can then determine what is possible and what is not, and determine how a cell is likely to behave, but never predict its behavior precisely

Instead of computing a single solution, it results in the formation of solution spaces. It is possible to exhibit behaviours within this range, each of which essentially represents a different phenotype depending on the component list, the biochemical characteristics of the constituents, and the enforced limitations. When every restriction is known, As stated on the right in Figure 2, the solution space condenses to a single point. Will we ever have this level of understanding of biological processes, then? Most certainly not, at least not in the near future, unless there are exceptional circumstances, like in the case of the human red blood cell9 or simple viruses10. This method does, however, produce models that are useful for deciphering, understanding, and even forecasting the genotype-phenotype association.

Types of restrictions and their application

The former can be used to define a range of potential actions. The latter can be used to further restrict acceptable behaviour, although these restrictions can change as a result of evolution. The changeable restrictions, such as kinetic constants, will also differ from person to person. When analysing metabolic fluxes, a series of sequential constraints can be used to reduce the range of possible flow distributions for a given metabolic genotype (see Fig. 3).

In the first section of Figure 3, fluxes through each individual reaction in the metabolic network are represented by axes in a space.

Due to the interdependence of the fluxes, not all of the points in this space can be reached. The steady-state fluxes are constrained by the stoichiometric matrix to a subspace, and since metabolic transients are quick, any deviations from this subspace are transient. Convex analysis is used to transform this plane into a cone if the reactions are defined such that all fluxes are positive. The points on the interior of the cone can be viewed as positive combinations of the edges of the cone, which become a collection of distinct, systemically defined metabolic pathways (see review in ref. 11). The length of each edge is constrained due to the capacity

restrictions on the individual routes steps. By closing the cone (step 3 in Fig. 3), these capacity limitations create a closed solution space in which all feasible metabolic flux maps are located. Through the use of linear optimisation, this space can be explored for the best phenotypes12,13. Recent experiments conducted in my group have demonstrated that Escherichia coli growth occurs along an edge that corresponds to ideal growth on minimum media.

With this knowledge, one can look for the kinetic restrictions that push the solution to the closed cone's edge.

The use of consecutive limitations on metabolism is most likely only the first such instance. By allowing for time-varying or adjustable restrictions, it is a strategy that connects the usage of clear physicochemical limits to the evolutionary change present in biological processes. This method presents a compelling alternative to quantitative modelling of biological systemic functions. Numerous textbooks14,15 have developed and discussed the more traditional physicochemical method to analysing biological dynamics, and specialised theories, including metabolic control analysis16, have been produced for the analysis of biological systems.

The iterative process to develop models

Iterative mathematical modelling of intricate biological processes and computer simulation of those processes will be used.

We'll start creating "in silico organisms"—computer simulations of their real-world counterparts. Genomic, biochemical, and physiological data will be used to synthesise the initial versions. These models will be able to anticipate and interpret some things.

However, these first models will only be able to accurately reflect portion of the organism's functions due to limited knowledge of limitations and incorrect annotation.We must develop the ability to accept failure as we go through this iterative model-building process.

The primary distinction between in silico and the in silico version lacks some traits that are present in the in vivo creature. As a result, we must conduct the tests, update the models, and generate experimentally testable hypotheses based on the in silico analysis (see Fig. 4). Curiously, this repetitive procedure for creating organisms in silico is expected to have two feedback loops. One is an in silico experiment, which is shown on the left in Figure 4, and the other is a traditional experimental loop. It is anticipated that many of the corrections and modifications made to these models will come from examining and searching the bioinformatic databases that are becoming more and more widely available.

With these in silico models, what will we do?

They most certainly have some fundamental scientific applications, such as comparative genomics and evolutionary research. The earliest metabolic models are probably useful for designing and running industrial bioprocesses as well as studying human infections. We shall

transition from discussing the genetic engineering of individual genes to what may eventually be referred to as "genome engineering," where the entire organism serves as the framework for the design. Some preliminary investigations in this area are now available

In this process of iterative model construction, there is one more issue that deserves discussion. A "need to know everything" mentality is being produced by high-throughput technology.

But even without "knowing everything," one can build effective and useful computer models, as experience in other domains has demonstrated.

We wouldn't have refineries or aeroplanes if we insisted on creating computer models that take into account every aspect of a process being researched. In reality, figuring out what is required to create an informative and practical computer model is one of the arts of model construction. It is conceivable that model development in biology will benefit from the knowledge gained from other sciences

from complexity to simplicity

It is evident that although the genotype (or molecular makeup) of live cells is intricate, the number of distinct behaviours (or phenotypes) that they exhibit is significantly smaller. The singular value decomposition of gene expression data, which unequivocally demonstrates that many expressed gene products operate in a highly coordinated manner17,18, is the source of this crucial principle of simplicity from complexity. These findings, for instance, demonstrate that two fundamental motions drive the genome-wide expression pattern of yeast during its cell cycle. Similar characteristics can be seen in studies of mathematical models of complex biochemical reaction networks. Only a few governing dynamic factors are revealed by sophisticated metabolic and growth models' temporal decomposition and robustness analysis, respectively.

A number of recognised system identification and model reduction techniques that have been used in a variety of science and engineering domains will be used to clarify the underlying simplicity.

Similar to how was mentioned above, the method of consecutive application of restrictions results in few allowed behaviours depending on a high number of interacting components.

The main "genetic circuits"1 that underlie cell activity are anticipated to be clarified by using these analysis techniques on the massive amounts of biological data now being created.

Why do we rate limit?

Biodata are being produced by high-throughput experimental technologies at previously unheard-of rates, and this trend will continue. The bioinformatic infrastructure, such as WIT, EcoCyc, Mips, Kegg, Biology WorkBench, EMP, Swiss-Prot, is growing and tabulates, curates, and makes these data retrievable. For data analysis, a variety of early visualisation tools and

ISSN 2515-8260 Volume 10, Issue 07, 2023

statistical procedures, such clustering, are becoming accessible. Mathematical models are typically not available, with very few exceptions. However, portable versions of models like the human red blood cell and Mycoplasma genitilium are starting to be made available. The talent that goes into formulating these models, performing the numerical analysis, and interpreting the results is presently in short supply. The amount of processing power that can currently be used to solve these models has not proven a barrier.

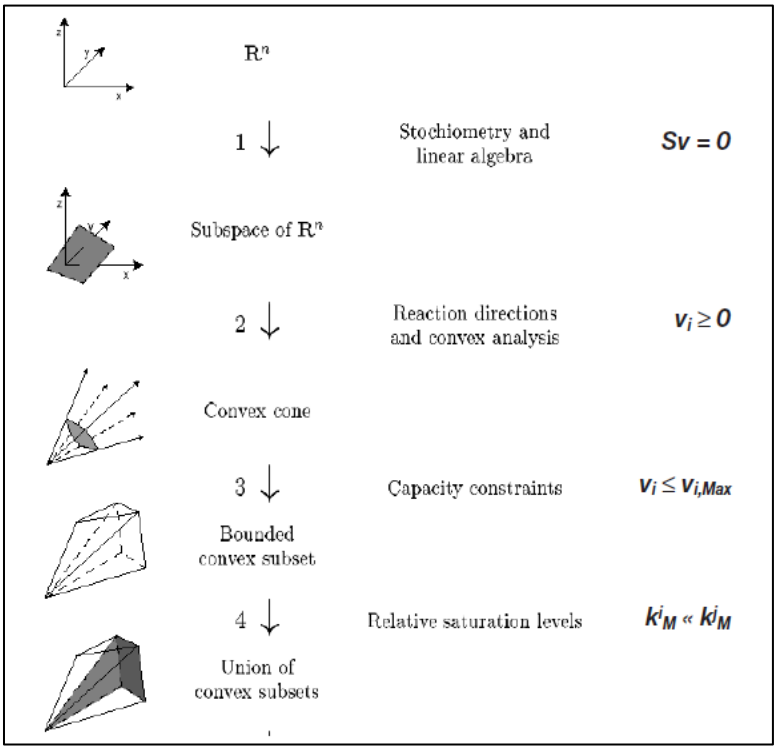

Figure 3. Narrowing down the alternatives. The application ofsuccessive constraints to a set of reactions in a pathway allows one to narrow down the attainable outcomes ("flux distributions") from a defined metabolic genotype (see text for further details).

The process shown in Figure 1 has repercussions beyond only a significant change in scientific attitude and emphasis. The biological sciences' educational infrastructure must adapt. Future biological scientists will need to have a better level of training in maths and informatics, as well as become more computer literate. The necessary fundamental modifications may be challenging to implement within the framework of the current biology departments. With the existing peer review system in place, it might not be viable to change the way faculty members approach their research and teaching.

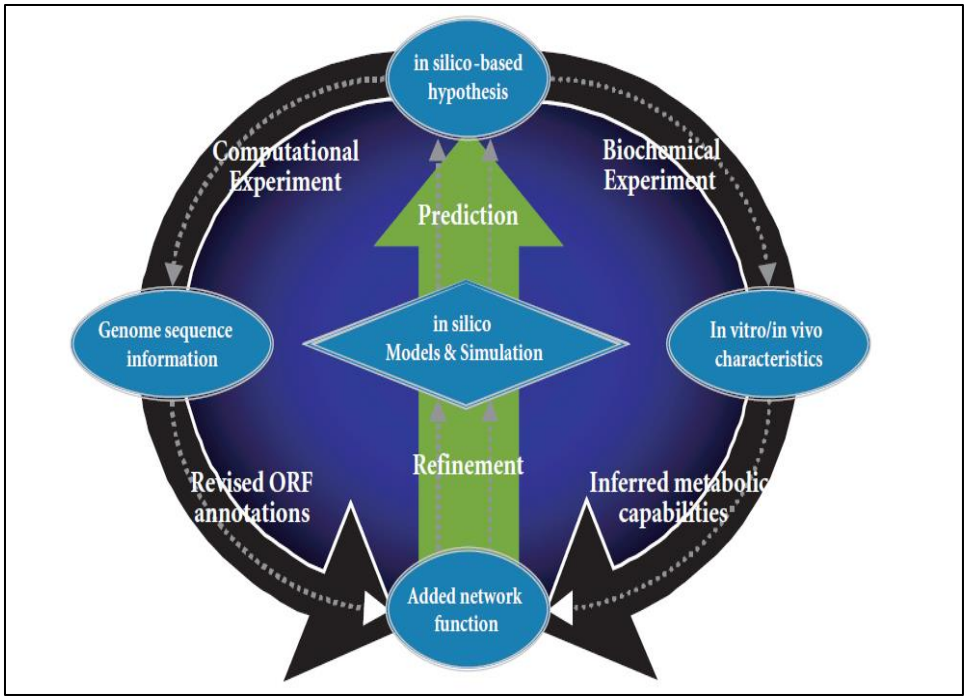

Figure 4. If at first you don't succeed. Iterative in silico model building in biology involves the formulation of experimentally testable hypotheses based on the in silico analysis, collection of experimental data, and subsequent refinement of the models based on these data.

New educational programs and department will develop. In addition to computer science and biology, the new curricula that must be created will also include mathematical modelling, numerical analysis, and systems science. Similar to how chemical engineering developed from chemistry and mechanical engineering at the beginning of the last century, new biologically based engineering programmes are expected to appear.

Conclusions

In addition to requiring researchers to adopt the systems point of view in cellular and molecular biology, high-throughput experimental techniques also allow us to investigate cells as systems. This capability necessitates the creation of mathematical models and computer simulations to examine the concurrent function of numerous gene products, given the complexity of even the most basic cellular function. These models will most likely be created for well-researched biological model systems and species (such as E. coli, yeast, and Drosophila) and used to interpret, predict, and analyse the genotype-phenotype link. The term "phenomics22," which is similar to the term "genomics," refers to the study of phenotypes with knowledge of the genotypes. Using computer simulation and the construction of mathematical models, phenomics will have a significant theoretical component. The complexity and unique characteristics of biological systems, such as time-varying constants (evolution), resilience, and redundancy are likely to distinguish model construction from other branches of science and engineering.

Acknowladgement

I great thank to my Supervisor Dr. Abhishek Kumar Mishra, Mr. Radheshyam Sharma and Dr. Rahul Pant for preparing the figures.

1. Eisenberg, D., Marcotte, E.M., Xenarios, I. & Yates,T.O. Protein function in the post-genomic era. *Nature* **405**, 823-826 (2000).
2. Palsson, B.O. What lies beyond bioinformatics? *Nat. Biotechnol.* **15**, 3–4 (1997).
3. Strothman, R.C. The coming Kuhnian revolution in biology. *Nat. Biotechnol.* **15**, 194–199 (1997).
4. Hartwell, L.H., Leibler, S. & Murray, A.W. From molecular to modular cell biology. *Nature* **402,** C47–C52 (1999).
5. Evans, G.A. Designer science and the "omic" revolu- tion. *Nat. Biotechnol.* **18**, 127 (2000).
6. Bailey, J.E. Lessons from metabolic engineering for functional genomics and drug discovery. *Nat. Biotechnol.* **17**, 616–618 (1999).
7. Aebersold, R, Hood, L.E., & Watts, J.D. Equipping sci- entists for the new biology. *Nat. Biotechnol.* **18**, 359 (2000).
8. McAdams, H.H. & Arkin, A. Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199–224 (1998).
9. Lee, I.-D. & Palsson, B.O. A comprehensive model of human erythrocyte metabolism: extensions to include
10. pH effects. *Biomed. Biochim. Acta* **49**, 771–789 (1991).
11. McAdams, H.H. & Shapiro, L. Circuit simulation of genetic networks. *Science* **269,** 651–656 (1995).
12. Schilling, C.H. et al. Metabolic pathway analysis: basic concepts and scientific applications in the post- genomic era. *Biotechnol. Prog.* **15**, 296–303 (1999).
13. Varma, A. & Palsson, B.O. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/Technology* **12**, 994–998 (1994).
14. Bonarius, H.P.J., Schmid, G. & Tramper, J. Flux analy- sis of underdetermined metabolic networks: the quest for the missing constraints. *Trends Biotechnol.* **15**, 308–314 (1997).
15. Reich, J.G. & Sel'kov, E.E. *Energy metabolism of the cell* Edn. 2. (Academic Press, New York, NY; 1981).
16. Heinrich, R. & Schuster, S. *The regulation of cellular systems.* (Chapman & Hall, New York, 1996), p. 372.
17. Fell, D. *Understanding the control of metabolism*. (Portland Press, London, UK; 1996).
18. Alter, O., Brown, P.O, & Botstein, D. Singular value decomposition for genome-wide expression data processing and modeling. *PNAS* **97**, 10101–10106 (2000).
19. Holter, N.S., Mitra, M., Martian, A., Cieplak, M., Banavar, J.R. & Fedoroff, N.V. Fundamental patterns underlying gene expression profiles. *PNAS* **97**, 8409–8414 (2000).
20. Palsson, B.O. Joshi, A., & Ozturk, S. Reducing com- plexity in metabolic networks. *Fed.*

Proc. **46**, 2485- 2489 (1987).

21. Alon, U., Surette, M.G., Barkai, N. & Leibler, S. Robustness in bacterial chemotaxis. *Nature* **397**, 168–171 (1999).
22. von Dassow, G., Meir, E., Munro, E.M., & Odell, G.M. The segment polarity network is a robust develop- mental module. *Nature* **406**, 188–192 (2000).
23. Tomita, M. et al. E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**, 72–84 (1999).
24. Schilling, C.H., Edwards, J.S. & Palsson, B.O. Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* **15**, 288–295 (1999).