A Time to Model and a Time to Experiment

MICHEL MORANGE*

ABSTRACT

The nature and role of models have been amply discussed by philosophers of science. They have emphasized the diversity of models and their functions. Biological sciences in general, and molecular and cellular biology in particular, are no exceptions. The nature and role of models in molecular and cellular biology are also a legacy of the different disciplines that contributed to its formation. Models can be a step toward abstraction, or the opposite, a step toward a material representation of an—to date—abstract phenomenon. Models can also help to collect information and knowledge. I will consider different models that played a highly important role in molecular and cellular biology, up to the Gene Regulatory Network model. There is a right time to model, and a right way to do it. I will try to understand why a model is well received (or not), and what kind of relationship it may or must have with experiments and experimental data.

Keywords: biological model, cell biology, gene regulation, synthetic biology, systems biology.

THIS ARTICLE GIVES A VERY GENERAL OVERVIEW of models in molecular and cellular biology—the roles they have played in various fields and their relationship with experiment—as an introduction to this special issue. I begin by describing what a model is. I then characterize models of gene regulation and present their evolution over the past 50 years. Finally, I give an even more global picture of the evolution of models in molecular biology and in what emerged from the molecular revolution of the 1960s up to the present, concluding with the changes that are occurring in the current transformation of biology.

1. WHAT IS A MODEL?

For philosophers, a model is a kind of intermediary between laws and general principles, theories, and experiments; it allows the transition from general theory to the more concrete ground of experiment. There are many different types of models such as computational and mathematical models and more simple models. In my opinion, there is a continuity between these models, whatever their difference. To establish a computational or mathematical model, you have to start from a simpler model that represents what you want to finally put in your computational or mathematical model.

A model is a representation and simplification. As a child, you might want a car that is a perfect model of your parents' car down to the very last detail. But when you create a model as a scientist, you know that normally you have to simplify—you have to extract some elements and eliminate others, otherwise it would

---

IHPST, UMR 8590, Université Paris 1, 13 rue du Four, 75006 Paris, France.

*Michel Morange's article was compiled and edited on the basis of a transcript of his presentation at the workshop, as he was prevented from writing it himself.

be absolutely impossible to establish the model. So a model is always a simplification, but a simplification that normally does not lack an innate essential element necessary for precisely understanding the phenomena under study.

There are mathematical and computational models (or computational simulations), and there are chemical models in molecular biology as well as organism models such as *Escherichia coli* and bacteriophage.

Among the people who were very important for chemical models in molecular biology were Linus Pauling, who originated important research on the structure of macromolecules, and more recently, Jane Richardson (Richardson et al., 1980), who proposed the ribbon model for the structure of proteins.

A model may have different functions in a work. It can be the objective of the work, as was the case with Monod and Jacob when they built the operon model: the objective was to create a precise model representing how the bacterium *E. coli* adapted to changing concentrations of lactose.

A model might not actually begin as one, but then later become one when chosen by a scientist of great reputation. This is particularly true for animal models. A case in point is the animal model chosen by the great scientist Thomas Hunt Morgan, who chose *Drosophila* as an experimental object for genetics. Not actually a model in the beginning, it later became one after many scientists worked on this object obtaining important results. Subsequently, a large part of the genetics community began working on this object.

A model can be used to summarize the state of knowledge on an issue. But the opposite can also be true; a model can be created at the beginning of the work with the goal of helping the scientist to imagine experiments for testing this model and eventually modifying it. And a model can be used to demonstrate that a type of explanation can be valuable even if there are no experimental proofs that it works this way. This is precisely what Max Delbrück did at the end of the 1940s when he proposed the nongenetic multiple steady state model in which memory was not dependent on the genetic content of the cells. He imagined two interacting metabolic pathways with mutual retroinhibition. Depending on the order in which one or the other of the substrates of these two pathways were added, a different state of functioning would be transmitted to the progeny. This model (which Delbrück also wanted to use to contradict the plasmagene hypothesis) did not open the way to precise experiments, but only demonstrated the possibility of this type of nongenetic memory.

One important point that I will return to later is that the position and role of models are not the same in all disciplines. For instance, it was more common in physics to have a model at the beginning of the work; in biology models were proposed more at the end of the experimental work. But it can also be different in the same discipline at different times in its development.

Creating a model is risky because you can be wrong and the model can have nothing to do with truth. This has been given as an explanation for the reason that tacit models—models that are shared by scientists, but are never clearly expressed—are so common. Model creation is risky because at the time that you are proposing the model, you may be missing essential information on the system. Only in retrospect can you clearly see that your efforts to build the model were premature.

This can be exemplified with Russian geneticist Nikolai Koltsov, who was one of the first to propose a physicochemical model of the chromosome in the 1930s. In this theoretical model, chromosomes were long chains of amino acids with branched substances like hormones. This model is now considered obsolete and has often been presented as ridiculous—it was too simple, and so on. But at the time when he proposed it, the role of DNA in the chromosome and gene structure was not yet known, or rather had been forgotten, so he could not have included DNA in his model. However, this does not mean that these models were useless: somehow they contributed to convincing biologists that determining the structure of genes and chromosomes was an important objective.

2. MODELS OF GENE REGULATION

I focus here on some important characteristics of these models, starting with the operon model of 1961 (Jacob and Monod, 1961), where I want to underline two aspects, including the way in which it was progressively transformed and modified. I then depict another model by Jacob, the regulon model.

The operon model in its representation in the *Journal of Molecular Biology* in 1961 contains very few structural details; there are arrows but there is not a good representation of the macromolecules (see fig. 6 in Jacob and Monod, 1961).^a A closer view, however, reveals that in these figures the repressor is represented

^aThe journal unfortunately did not grant us free access to use the figure.

in the same way as RNA, because at the time Jacob and Monod thought that the repressor was RNA. They changed it later. This famous article also shows that the authors were imagining two possible levels of regulation—transcription and translation. In subsequent publications, translation disappeared.

It is interesting to see how this model was progressively transformed in the following years. The next picture shows how it was represented in a lecture by Francois Jacob in 1965 (Fig. 1; see Jacob, 1966).

There are more details of the components of the system, for instance the structure of ribosome, and the structure of proteins is also somehow represented.

The model evolved toward far better representation in the 1980s. In many molecular biology text books published at that time, you now see a more beautiful representation with much more detail (Mathews and Van Holde, 1990).

A more recent representation of the operon model was produced by the group of Ozbudak (see fig. 1 in Ozbudak et al., 2004).^a It is interesting because it is very different in how it was designed and, in fact, this model was the introduction to the mathematical model of the lactose operon—maybe not the first, but the first to be of a sufficient quality to allow presentation of the system and to allow clear predictions from the model. I will come back to another point regarding this model later, but you can appreciate the huge transformation of the model.

FIG. 1. The operon model in a lecture by Jacob (1966).

I turn now to the regulon model that was proposed by Jacob et al. (1963). Highly speculative in nature, it was an extension of the operon model to the regulation of the replication of the bacterial cell, and it was generated in collaboration with Sydney Brenner and François Cuzin. It shows a regulatory gene coding for a protein that is an activator called “initiator.” It binds some DNA on a specific sequence to activate initiation of transcription. The model was proposed by Jacob, Brenner, and Cuzin to guide experiments on replication to test it.

In this context, it is, therefore, surprising that Jacob considered the model by Britten and Davidson of gene regulation for higher cells that I briefly present hereunder, too speculative! So, in fact, it was not a discussion about whether a model was speculative, but rather about whether a model was too speculative compared with already-established models or to existing experimental results. The operon model, you might say, was a model derived from experimental results obtained in the lactose system in *E. coli*. But the operon model was more than that; it was supposed to be a model for genetic regulation in all organisms. And from that point of view, it was not a model evolving from experiments, but rather a speculative model preceding experiments and helping to test this hypothesis. The value of the operon model was thus ambiguous, and this ambiguity was rarely mentioned in the publications.

I now come to Roy Britten and Eric Davidson’s 1969 model of gene regulation in higher cells (Britten and Davidson, 1969; see Figure 2). I want to underline a few points.

The first is that it was very different from the Jacob and Monod operon model of 1961, but similar in its representation to the 2004 Ozbudak et al. model. The Britten–Davidson model was clearly an anticipation of the future models in systems biology.

Second, the authors clearly stated in their introduction to the model that it was built specifically to stimulate experiments that would most likely show that some of the elements of the model were not valid, forcing changes in the concepts presented in the model. So it was a type of model that guided experiments more than represented results, as was stated by the authors:

Our purpose in presenting an explicit theory is to describe the regulatory systems proposed in terms of elements and processes which are capable of facing exact experimental test. It is hoped that our relatively detailed commitment will induce discussion and experiment, and it is expected that major modifications in concept will result. (Britten and Davidson, 1969)

A few words about the evolution of the representation of the Britten–Davidson model. In 2001, Davidson and his collaborators published the now famous description of the *endo16* promoter, a regulatory gene in the development of sea urchins (Yuh et al., 2001). Only a very simple representation of the proteins was given, together with their position on the promoter (Fig. 3).

FIG. 2. Britten and Davidson’s model of gene regulation for higher cells (1969).

FIG. 3. Model of the Endo 16 promoter by Yuh et al. (2001).

A year later, Davidson et al. (2002) proposed one of the first representations of a developmental gene regulatory network (Figure 4).

This model was both a representation of experimental results and a basis for the subsequent elaboration of a computational model. The latter was open to experimental testing but not in the same way as the 1969 Britten–Davidson model (Britten and Davidson, 1969): what had to be tested now was the details, not the general value of the model.

3. A BRIEF HISTORY OF MODELS IN MOLECULAR BIOLOGY

Where does the way models are represented in molecular biology come from? It was borrowed from other disciplines, in particular from biochemistry and genetics. Biochemistry provided the representation of biochemical reactions and molecular structures, whereas genetics provided models of gene actions. It is interesting that the first encounter between the two types of representation occurred in the 1930s when

FIG. 4. Model of the regulatory gene network for endomesoderm specification: the view from the genome (Davidson et al., 2002).

Boris Ephrussi and George Beadle (and others) tried to look for the role of genes in development (Beadle and Ephrussi, 1937).

On the left side, their model (Fig. 5) shows the successive transformation of a substance that is important for determining eye color, something similar to a biochemical pathway. The different genes involved in the process are represented on the right, with arrows between the genes and the different steps. So there are two types of arrows—those representing chemical reactions and those indicating actions of the genes.

Exploring the relationships between the two types of arrows played an important role in the development of models in molecular biology.

Biochemistry and genetics were not the only disciplines that contributed to models of molecular biology, there was also cybernetics, in particular when the issue of gene regulation emerged. A significant aspect in the history of modeling in molecular biology was the conversion of arrows that pointed to chemical reactions into arrows of information transfer. It is particularly visible in a 1954 drawing by James Watson that depicted the DNA–RNA–protein relationships (Judson, 1996). Watson saw that there were two ways of conceiving the transition between DNA and RNA. It could be either a chemical transformation of the DNA molecule or a transfer of information between DNA and RNA consisting in the sequence of nucleotides. This drawing shows that there were doubts about the exact meaning of the arrow between DNA and RNA.

A further transformation in molecular biology models consisted of a progressive introduction of structural chemical details in the representation. I discussed the case of the operon model previously, and another example is the modeling of the behavior of the bacteriophage. It is a virus that infects bacteria either to introduce its genome within the chromosome and stay silent—it is lysogenic—or it lyses bacteria. This model by Lwoff (1953) has no structural details; the bacterium is a single rectangle, and the chromosome is a single thread. The only detail is the shape of the bacteriophage, which had been demonstrated by electron microscopy shortly before.

The trend toward realistically representing molecules became increasingly significant in the following years and was linked to various phenomena. The first was the publication of important textbooks popularizing the results of molecular biology, such as “Molecular Biology of the Cell,” first published by James Watson et al. in 1983 (Alberts et al., 1983). The authors made many efforts to include beautiful and significant representations of the way the structure of macromolecules (proteins) explains the functions that they fulfill

FIG. 5. Beadle's and Ephrussi's 1937 model of the action of developmental genes in *Drosophila*. Courtesy of the Genetics Society of America.

within cells. The second was that color printing became much easier, enabling the publishing of low cost books with beautiful color images. Also, some of the results of molecular biology, such as the description of signaling pathways, demanded a representation that enabled their functions to be fully understood.

In these models, such as those representing the signaling pathways of a growth factor, many cellular components are not shown, or are not represented with their correct size and abundance in the cell. For example, the size of the nucleus and receptor could be the same, stoichiometry could be disregarded, and many cell proteins not visible. These models are clear simplifications and deformations. In addition, there is not a single cell for which this representation is valid, since the models are summaries of the observations of the same pathways in different cells.

So what was the role of these models? They summarized the state of knowledge of a pathway in a way that was very simple to memorize, and these kinds of drawings became dominant in molecular biology in the 1980s and 1990s. This changed around 2000 as a consequence of the development of systems biology, which put more emphasis on the relationships than on the nature of the different components.

Also notable is the result of the research aiming at finding interacting proteins (interactomes). Interestingly, the first study on interactomes was carried out in yeast as a model, simply looking at the possible interaction between proteins. One would like to believe that the interactions shown in interactomes are real, but with the technique (double hybrid technique) used to build the interactome, it is impossible to demonstrate that the interaction was actually taking place in cells. Publications based on this technique show possible interactions without tests that they really take place.

This tendency in some of these models to abandon physical interaction in favor of possible poorly defined relationships has become progressively stronger. For instance, for the participation of two components in the same pathways, the existence of physical interactions was no longer required and a correlation in their expression and activity became sufficient.

The study of Charbord et al. is an example of a recent publication in systems biology. Their model represents a group of different genes that are linked by arrows. Arrows do not mean that they interact, they simply mean that there are “relations” between them (Charbord et al., 2014). For instance, they covary in different situations or they have been shown in other experiments to be related to the same biological process.

These general transformations have changed the position of models in articles and their relationship with experiments. For instance, in synthetic biology, models are now very often the beginning of research because it is necessary to model before starting experimental work. The famous repressilator, which was developed in the 2000s, was a dynamic system generating circadian rhythms in bacteria. Modeling was a necessary prerequisite for the study, otherwise it would have been impossible to design the system.

In systems biology, there is a back and forth rapid movement between models and experiments. For instance, in *drosophila*, the first model of genes acting early in development was shown not to work; it did not account for the development of *drosophila*, and for this reason it needed to be complemented by the addition of new elements—new genes—and with this addition, the model worked correctly. This rapid back and forth movement between models and experiments was not previously traditional in molecular biology.

So there are different places and roles for models today. Transformation is taking place not only in models of gene regulation but also in the representation of proteins. The representation of rigid molecules with beautiful shapes is now progressively losing their importance. Proteins in particular, but also RNAs, are seen as a population of different conformations in equilibrium, which is something that is very difficult to represent in a classical drawing.

To conclude, there is no universal way to progress in modeling. We might have the idea that we start from rough data, after which we create a kind of simple drawing, then create more precise drawings, and then do a computational or mathematical model of the system because we have enough knowledge of it. This gives us the illusion that modeling will always progress in this direction.

A counter-example is the famous Hodgkin–Huxley model of the propagation of nerve impulses, which was proposed at the beginning of the 1950s (Hodgkin and Huxley, 1952)—a famous model for which its authors were awarded a nobel prize. It is often quoted as one of the early mathematical models in biology.

Eventually this model lost its importance in the years that followed and was replaced by a precise description of the different channels in the nerve membranes that permit this exchange of ions and the generation of currents. This does not mean that the model was useless, rather it was only a step: it stimulated the characterization of the channels that was considered progress in comparison with the previous abstract model.

The model is no longer the last and final step in the evolution of a science. Biological models today are much more diverse than they were, for instance, at the end of the 1980s or 1990s. Models now play different roles in research, and they have different positions in the development of the work at hand.

If, in conclusion, we try to draw a global picture of molecular and cellular biology today, two different antagonistic trends can be distinguished. On the one hand, there is a move toward increasingly precise descriptions and models where all the components and their structural characteristics are described. You tend to have a more perfect representation, for instance, of a cell. The other direction is the opposite: toward abstract models that represent only relationships and do not aim at precise structural descriptions. It is unclear which of these two directions will prevail or whether the two will coexist in the coming years.

AUTHOR DISCLOSURE STATEMENT

The author declares there are no competing financial interests.

REFERENCES

- Alberts, B., Bray, D., Lewis, J., et al. 1983. *Molecular Biology of the Cell*. Garland Publishing.
- Beadle, G.W., and Ephrussi, B. 1937. Development of eye colors in *Drosophila*: Diffusible substances and their interrelations. *Genetics* 22, 76.
- Britten, R.J., and Davidson, E.H. 1969. Gene regulation for higher cells: A theory. *Science* 165, 349–357.
- Charbord, P., Pouget, C., Binder, H., et al. 2014. A systems biology approach for defining the molecular framework of the hematopoietic stem cell niche. *Cell Stem Cell* 15, 376–391.
- Davidson, E.H., Rast, J.P., Oliveri, P., et al. 2002. A genomic regulatory network for development. *Science* 295, 1669–1678.
- Hodgkin, A.L., and Huxley, A.F. 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.* 117, 500–544.
- Jacob, F. 1966. Genetics of the bacterial cell. *Science* 152, 1470–1478.
- Jacob, F., Brenner, S., and Cuzin, F. 1963. On the regulation of DNA replication in bacteria. *Cold Spring Harb. Symp. Quantit. Biol.* 28, 329–348.
- Jacob, F., and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.
- Judson, H.F. 1996. *The Eighth Day of Creation, Expanded Edition*. Cold Spring Harbor Press, Long Island, NY.
- Lwoff, A. 1953. Lysogeny. *Bacteriol. Rev.* 17, 269.
- Mathews, C.K., and Van Holde, K.E. 1990. *Biochemistry*. The Benjamin/Cummings Publishing Company, Inc., San Francisco.
- Ozbudak, E.M., Thattai, M., Lim, H.N., et al. 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427, 737.
- Richardson, J.S., Tainer, J.A., and Richardson, D.C. 1980. An illustrated museum of protein structures. *Biophys. J.* 32, 211.
- Yuh, C.H., Bolouri, H., and Davidson, E.H. 2001. Cis-regulatory logic in the *endo16* gene: Switching from a specification to a differentiation mode of control. *Development* 128, 617–629.

Address correspondence to:
 Prof. Michel Morange
 IHPST
 UMR 8590
 Université Paris 1
 13 rue du Four
 75006 Paris
 France

E-mail: michel.morange23@orange.fr