In Silico Experiments in Scientific Papers on Molecular Biology

Sabrina Moretti

This article explores the role of the so-called *in silico* experiments used in molecular biology. It is based on the analysis of some papers that present scientific applications which rely on *in silico* experiments. By means of this study I found two basic ways of viewing them. According to the first view, the *in silico* experiment is a computer program that realizes some specific operations: it constitutes some particular experimental conditions, which allow us to investigate biological phenomena, and which complement those present in *in vivo* and *in vitro* experiments. According to the second view, *in silico* experimentation has a different meaning, which corresponds more closely to the meaning of "simulation": its identity is linked to that of the "model" used to construct such simulation. The authors of the analysed papers never express an intention to standardize a model, so its meaning remains contingent, and cannot be turned into a technical object.[0]

Keywords: Molecular biology, *in silico* experiment, simulation modelling

Introduction

In the last fifty years, the use of computer simulation in the field of scientific research has become more frequent. Simulation has been used for making predictions, for training purposes, for developing and verifying scientific theories – in particular, theories of non-linear, dynamic systems – by means of the creation of a new interdisciplinary scientific discipline for studying complex systems. The results of simulations appear more recurrent in scientific discussions, political decision-making processes and the mass media, which is why there is now a requirement for better understanding of their role in scientific practice. Simulation gives rise to many epistemological discussions, both at a general level (cf. Hughes, 1999; Sismondo, 1999; Dahan Dalmedico, 2000; Knuuttila et al., 2006) and in the various scientific disciplines (cf. Rohrlich, 1991; Galison, 1997; Merz, 1999; Keller, 2000; Sunberg, 2008; Wieber, 2009).

My work will focus on the role of *in silico* experimentation in the field of molecular biology. The term *in silico* was coined at the end of the 1980s, to refer to "virtual" experiments existing only "inside" computers. It complements the terms *in vivo* and *in vitro*, characterizing experiments that are accomplished, respectively, within a living

organism and outside the organism, in a controlled environment.

In the last two decades, the use of *in silico* experimentation has grown considerably in the field of biology. This growth has been further accelerated by the *Human Genome Project,* which has inspired the strategy of recording all structural and functional information regarding numerous biological phenomena, thus creating an integrated set of bioinformatics databases, many of which are made available to everybody via the World Wide Web. Molecular biologists now have at their disposal a huge mass of electronically stored data to which a biological meaning must be given. Thus bioinformatics has developed as a discipline, with the aim to furnish some important computational tools with which to analyse the genome, and to formulate new hypotheses about the evolutionary and functional relationships of biological elements (Valle et al., 2007).

The earliest databases containing sequences of nucleic acids date back to the 1980s. These were soon followed by databases of protein sequences. There is now a huge proliferation of databases gathering the results of many types of research being carried out both in the laboratory (*in vivo* and *in vitro* analysis) and through the use of computational tools (*in silico* analysis). The online review *Nucleic Acids Research*, which each year publishes a list of all databases containing molecular biology information, lists 179 of them in January 2009, as compared to 84 in January 2008 (Galperin & Cochrane, 2009).

My work is intended to help in understanding how the *in silico* experiment is regarded in the scientific literature of the field of molecular biology. It is based on the analysis of some papers that propose a scientific application which relies on *in silico* experimentation. I will try to understand the role of the virtual experiment by means of a systematic study of the textual organization; the positions that *in silico* experiments take up in the papers and the relationship they have with the other elements that form the text. The starting point of this analysis is Rheinberger's distinction (1997) between "epistemic things" and "technical objects", two different but inseparable components of experimental systems. For Rheinberger, experimental systems are the "smallest integral working units of research" (Rheinberger, 1997: 28). They are "irrevocably local and situated in space and time", and they allow us to "materialize" scientific events. They include "scientific objects and the technical conditions of their coming into existence", which Rheinberger calls, respectively, "epistemic things" and "technical objects" (Rheinberger, 1997: 28).

Epistemic things are those entities upon which the researchers' questions are concentrated. They function essentially as "question-generating machines" (Rheinberger, 1997: 32). Epistemic things appear with an irreducible vagueness, which is inevitable because they incorporate what we do not know. They are those hidden things that we must bring to light by means of some sophisticated manipulations. Technical objects are those objects that form experimental conditions: they function, in essence, as "answering machines" (Rheinberger, 1997: 32), and belong to the technical repertoire of an experimental setting.

In a particular experimental system, both types of things are linked by an interaction and inter-conversion that develop in both time and space. "The technical conditions determine the realm of the possible representations of an epistemic thing; and sufficiently stabilized epistemic things turn into the technical repertoire of the experimental arrangement" (Rheinberger, 1997: 29). Therefore, the difference between

an epistemic thing and a technical object is functional, not structural. It is not possible to define once and for all a distinction between these two components of a system. The function of an object depends on the position or "node" (Rheinberger, 1997: 30) that it takes up in an experimental setting. In scientific practice this distinction of roles is evidently organized, so the text of scientific papers is often divided into "materials and methods" (technical things), "results" (halfway-hybrids) and "discussions" (epistemic things) (Rheinberger, 1997: 30).

Knorr Cetina (1997) maintains that Rheinberger's equation of instruments with technological objects is not fair. Indeed, advanced technologies, such as computers and computer programs, "are simultaneously things-to-be-used and things-in-a-process-of-transformation" (Knorr Cetina, 1997: 10), thus placing them in the class of epistemic things. Knuuttila and Voutilainen argue, in turn, that "models can be treated as epistemic artifacts from the scientific practice point of view. As epistemic artifacts, scientific models are open to different interpretations and uses, functioning as both tools and objects of inquiry" (Knuuttila & Voutilainen, 2003: 1485).

Merz (1999) and Sundberg (2008) use the concepts of "epistemic things" and "technological objects" to analyse the role of computer simulations in particle physics and meteorology respectively. Merz (1999) focuses on "event generators" used by physicists: types of software that simulate mechanisms of particle production. She maintains that event generators accomplish different tasks depending on different work settings and different actors. Each local setting requires, of these objects, a specific function, meaning or viewpoint, to enable them to occupy different places. But, unlike Rheinberger, Merz insists that the oscillation among these different places does not follow a temporal dynamic. Event generators are not things that are initially epistemic things, which then turn into technical objects after their use is consolidated. They can be, at one and the same time, "the question generating machines" (Rheinberger, 1997: 32) in a particular setting, and "the technical repertoire of the experimental arrangement" in another (Rheinberger, 1997: 29). Like Knorr Cetina, Merz maintains that event generators are characterized by the "unfolding ontology of knowledge objects. [Because they] are always in a process of being materially defined, they continually acquire new properties and change the ones they have" (Knorr Cetina, 2001: 180). Event generators can be perceived by the physicists using them as objects with a multiplicity of aspects.

Sundberg (2008) analyses the use of simulation models in meteorology. She maintains that the role of simulation is related to the relationship between the object and the people engaging with it. She distinguishes two types of relationship: the *development* and the *use* of a simulation model. In the first case, such a model takes the role of an epistemic thing, because it is in a process of transformation, and is consistently changed and improved. In the phase of utilization, the simulation can take the role of either a technological object or an epistemic thing.

Keller (2000) discusses the role of models in molecular biology, considering the dichotomy between theoretical and experimental science. She focuses, in particular, on a computational model for gene regulation, derived from the experimental work of Eric Davidson and his colleagues at Caltech (Yuh et al., 1998). The model is based on the metaphor of the "genetic computer". Such a metaphor has a conceptual role, not only in the sense of

directing the attention and perception of the researchers, but also in the sense of guiding the material manipulation in the different types of laboratories (biological, computational and industrial), and for various types of aims (theory development, laboratory tools and commercial products). Keller concludes that the theoretical propositions in the applied sciences, such as molecular biology, concern mainly means and ends rather than the "truth". Such means and ends are specific, local and specialized. Attention is shifted away from the validity of the different representations of reality, and towards questions regarding the preference for certain types of intervention rather than others (Keller, 2000: S85).

Wieber (2009) deals with simulation in the field of protein structures. He calls simulation technologies "theoretical technologies", and maintains that, thanks to their nature, they have been added to the traditional toolbox used by experimenters to analyse and interpret empirical data relating to molecular structures, leading to a focus on dynamic, as opposed to static, structures.

Hine (2006) describes an ethnographical research on the use of databases in molecular biology. The research has the objective of understanding how databases influence the scientific practices and modes of communication. She concludes that databases do not cause a radical transformation in scientific practices, but instead make small-scale changes to working practices. They do not give rise to a new and distinct epistemic culture, but can be considered as additional resources that enter into an already existent research culture.

In what follows I will study the link between epistemic questions and needs of application, suggested by the aforementioned studies, in the field of computational biology. I will focus on the various functions of the so-called *in silico* experiments. In particular, my objective is to understand the different roles of *in silico* experimentation as they are presented in the scientific papers on molecular biology. These roles depend on the objectives of each paper, and on the way in which *in silico* experiments relate to all other elements of the discussion. I am interested in the identity of *in silico* experiments emerging in these papers. In particular, I wish to analyse the level of detail in which they are described, how they are related to more traditional experiments, and how their usage is justified. Usually, a scientific paper is subdivided into various parts, each of which has a precise role. The central part generally shows results, while the final part focuses on discussions. The "methods" section lists all methods and techniques used in the work. Some supplementary files may report certain technical details, figures, tables, etc. The analysis of the structure of a paper can be useful for understanding the position of the given *in silico* experiment in relation to other elements, and the meaning that it can assume in the presented work.

Obviously, document analysis has its limits, and it does not allow us to evaluate fully the use of simulation in biology. A scientific paper is not a research report, even though it is often considered as such. That is to say, it is not a faithful and detailed account of the activities in the laboratory. It hides the "contingent situational logic" of the laboratory. Indeed, it represents a genuine process of transformation and recontextualization that misrepresents the reality of the research (Knorr Cetina, 1981). Document analysis does not reveal how this transformation process is realized, and therefore does not allow us to understand completely just how *in silico* experiments are incorporated in the contingent day-to-day activities of the laboratory. However, the scientific paper is paramount

in scientific practice, not because it collects and imparts information, but because it represents "institutionally authorized enunciation of scientific truth" (Frohmann, 1999: 72). The scientific paper represents the public face of scientists, and so, through the analysis of its formal communicative modalities, it is possible to identify these aspects. Through my work, I intend to understand how the *in silico* experiment is received and accepted by the scientific community, and what role it plays in the public conception of molecular biology; what is assumed in a non-problematic way within a scientific culture, and what are the aspects that lead us to consider an assertion valid (in the sense of being shared by the scientific community).

My analysis shows that the *in silico* experiment is accepted whenever it is in keeping with the practices already in use, and that there are essentially two different ways in which it is linked to the traditional activities of the laboratory. Moreover, *in silico* experiments allow scientists to work on particular "disposable" arrangements of epistemic things that are used locally to generate new theoretical reflections considered as useful in a specific context. We thus see a continuous de-contextualization and re-conceptualization of epistemic things that generate a multiplicity of new and provisional epistemic formations.

In Silico Experiments

This work is based on the analysis of 48 papers, published in the years 2000–2008, presenting some applications in the field of molecular biology based on *in silico* experiments. Because I wish to analyse how they are received by scientists, I have chosen, as my source, the journal *Nature*, which is considered prestigious and authoritative by the scientific community, and used by scientists who wish to be recognized widely.

My initial intention was to analyse all papers in the field of biology that make reference to *in silico* experiments, but I discovered that most of these papers relate to molecular biology, and that the use of *in silico* experiments in other branches of biology – such as systems biology and genetics – is presented in a very different way. This seems to be due to the fact that, compared with other fields, molecular biology is much more closely linked to the epistemic culture based on conventional laboratory experimentation. Because of this, I took into consideration only molecular biology.

The papers were selected by means of the search facility provided by the website of the journal: www.nature.org. I used the following search expressions: "molecular biology" and "in silico", and I chose only those papers published between January 2000 and October 2008. The search facility revealed 120 papers that matched the criteria. I initially selected only "articles" and "letters", before eliminating all papers in which the expression "in silico" appeared only in the bibliography or in captions. I then added another four papers found by means of bibliographical references. I was thus left with 48 papers actually presenting applications that use so-called "*in silico* experiments".

Most of these papers can be subdivided into two classes:

Type 1 applications using bioinformatics technologies, in order to analyse the DNA of some organisms;
Type 2 applications aiming to study metabolic networks and networks of genomic regulation.

The papers of Type 1 present some applications that analyse the DNA of some specific chromosomes of living organisms, in order to find or compare genomic sequences for analysing the causes of given diseases. In

simple terms, these applications are developed on the basis of the results of a series of *in vivo* and *in vitro* experiments on some fragments of biological DNA that produce some sequences of nucleic acids represented by a set of symbolic codes. These sequences of symbols are explored heuristically by means of algorithms that find patterns, motifs or statistical regularities, or similarities and homologies with some genomic sequences of other organisms already recorded in public databases. In spite of the complexity of these processes, some methods of exploration have been defined, based on statistical algorithms or some typical artificial intelligence techniques, giving rise to a set of programs used by numerous researchers. The choice among different methods depends upon the results that they produce; often almost all of the methods are used simultaneously.

In the text of these papers some phases of sequencing are described, and all techniques used both *in vivo* and *in silico* are listed. Generally, there is some discussion about specific characteristics of the genome, and about how this knowledge could be used in a medical field. In the "methods" section, authors refer to some web sites for technical details – for example, methods for the construction of sequence-ready maps and for sequencing large insert clones by shotgun cloning, use of selectable markers – where it may also be possible to find some programs used to conduct *in silico* experiments.

In spite of the frequent use of the words *in silico* and *simulation* in Type 1 papers, these applications cannot be said to be true simulations as such, if by "simulation" we intend the reproduction of a process or a mechanism that causes a specific empirical phenomenon. The authors of these papers mention *in silico* experiments because their analysis is based on an "artificial" element, rather than directly on the biological material. The plausibility of the use of these techniques rises directly from the assumption that working with sequences of nucleic acids is exactly the same thing as working with sequences of corresponding symbolic codes. The fact that a gene can be considered a sequence of information-carrying symbols automatically enables the application of some operations for working on symbols. These operations are defined and used in non-biological contexts; they are usually instruments supplied by mathematics, statistics and computer science. All experiments on genome sequences work simultaneously both on a living organism, or a part of it, and on its digital representation. So, we see that the outputs of an *in vivo* or *in vitro* manipulation, using markers, X-rays or other particular instruments, often become the inputs of a program that will filter the data, delivering the final results in an electronic format. *In silico* experiments allow scientists to work on a "symbolic" gene, whereas *in vivo* and *in vitro* experiments work directly on a material object. Indeed, the function of digital computation in an experimental context is precisely to infer new knowledge from symbolic sequences constituting DNA.

All the papers of Type 2 that I have analysed are based on the definition of a model of the metabolism of some living organisms – generally Escherichia Coli, Drosophila and yeasts. Metabolic structures are conceived, in the models, as networks composed of nodes and links. The elements of these networks are certain selected elements of molecular biology – genes, enzymes, proteins, etc. – linked together by means of mathematical relationships. Often, these models are constraint-based, derived from the laws of thermodynamics, the law of mass action, or other deterministic or stochastic models. Sometimes these models are defined

by a set of differential equations expressing the rate of change of a given concentration as the sum or product of more or less empirical terms (for example, the power law terms, or the law of mass action terms).

The definition of a model always begins with an exploration of various databases and of the specific literature, in order to identify those biological elements that are interesting, and to define all the logical or mathematical functions representing their links. For example, if we need to define a metabolic network of a particular organism, we must find the list of metabolites and enzymatic reactions, and also identify and consult the literature that deals with flux balance analysis. Generally, several sources are used, sometimes of various types, which must always be readapted to construct the specific model.

The construction of a model is based on a series of choices that depend upon the application in hand. This process might be likened to the solving of a jigsaw puzzle, putting together theoretical concepts and pieces of other models that sometimes have very different properties. Each researcher has at his or her disposal a large quantity of databases and models describing different aspects of biological processes by means of stochastic techniques, differential equations, Boolean networks, neural networks, and so on. There are several formalisms and meta-models available, so the final work is a hybrid originating from various different approaches.

In essence, there are no established rules to define the construction of a model. All choices made by researchers depend on various, often contingent, factors: the objectives, the available data, the previous research, the results of other experiments. Frequently no justification is put forward for these choices.

An *in silico* experiment is the simulation of a model. In these papers there are two different modalities to refer to simulations:

2.1 Some papers base their discussion on comparison between the results of simulations and the results of *in vivo* or *in vitro* experiments. What is important in a model is its forecasting capability by means of interpreting results obtained from simulations, and their agreement with some data from *in vivo* or *in vitro* experiments: if there are some inconsistencies between predictions and experimental data, it means that this model, which furnishes a theoretical interpretation of biological systems functioning, is wrong, and so also are the theoretical concepts that form the basis of its definition; if its results are consistent, this model can be used to make some counterfactual experiments and, more generally, to infer new knowledge. On the basis of its predictive capability, it is possible to indicate knowledge gaps and identify previously unknown components and interactions in the regulatory and metabolic networks that the model represents. These papers, then, focus above all on the performance of a model.

2.2 In other papers, authors focus on the structure and the characteristics of a model. In essence, a model is explored; all resulting remarks are then used to obtain new information about the functioning of cells. As we have already seen, all models use the concept of "network" to represent the fundamental functions of metabolism. The characteristics of this abstract structure, formed by nodes, links, pathways, flows, etc., can help scientists to better understand some biological functions. All papers

assigned to this category describe the model by means of a graphical visualization of a network. Therefore, in these papers, the models' *structure* is of principal interest, and not their *performance*. In essence, we have moved one step forward from case 2.1. Here, the representative capability of a model is not discussed directly, although it may be mentioned briefly. The authors begin with the presupposition that the model is able to represent a given biological object. So, results obtained from its simulations can be meaningfully associated with the system that it represents.

These two distinct ways of referring to *in vivo* or *in vitro* experiments sometimes appear together in the same paper. In such cases, at the beginning they focus on the relationship between *in silico* and *in vivo*, or *in vitro*, results, and then, when the model is validated, they show some virtual experiments used to obtain new knowledge.

Epistemic Things and Technical Objects

From the analysis of these papers, it emerges that the expression "*in silico* experiment" denotes a scientific practice that performs some manipulations on a computational representation, rather than directly on a material object. It is a virtual experiment that does not substitute, however, but instead supports traditional experiments.

An *in silico* experiment is fundamentally a computational object, and it is worthwhile examining the meaning of some of its underlying concepts.

We may begin with the concept of "data". In computer science, data are objects that codify information that can be described and handled by a computer. Beginning in the 1970s, cloning and sequencing techniques led to the representation of molecular structures in general, and genes in particular, as a "literal open reading frame of sequence – digitized data, in other words" (Searls, 2010: 2). The possibility of representing biological data in digital form has allowed their storage and circulation through databases that have caused a "change in how biological knowledge is constructed"; a database makes "data accessible to other research contexts and therefore potentially reusable as evidence for new claims, and it associates data with a broader range of phenomena" (Leonelli, 2009: 746).

Another important concept is "model", in the specific sense of a simulation model. According to Morgan and Morrison, models serve "both as a means to and as a source of knowledge" (Morgan & Morrison, 1999: 35). Scientists work on certain representations defined by means of some tools that impose an interpretation, from a particular point of view, of their objects of study: "the model functions as *representative of* one or more phenomena as well as *representative for* a given theory" (Leonelli, 2007: 17). Both data and model are representations. Data represent a structure, something that can be identified with a list of all properties that characterize it, whereas a model represents also a mechanism that is able to transform data. A model is made in such a way that it can be translated into a code by means of a programming language.

The model not only has a representative function, but also provides "the kind of information that allows us to intervene in the world" (Morgan & Morrison, 1999: 23). Indeed, models have numerous practical functions, such as suggesting possible experiments, helping to predict consequences of given interventions and sug-

gesting new questions to ask (Keller, 2000: S78).

Finally, a simulation is the elaboration of a program that translates a model. Its input is a set of data, and it produces other output data that constitute the results of this simulation. According to Parker (2009), computer simulation studies – studies based on a series of simulations – have all the characteristics of the traditional experiment in as much as they imply an intervention activity. They allow us to study the consequences of changes applied to a system. For Parker, they are "material" experiments in the real sense of the word, because they involve the behaviour of a physical/material system – namely, the programmed digital computer. The observed behaviour of the system is the result of such experiments. That is to say, the material/physical system of the computer constitutes the experimental setting.

If we consider the relationship between data, model, program and simulation, we can say that for each model we may have several programs, because the translation of a model into a code can occur in different ways. A program, in its turn, can give rise to numerous simulations, each of which is its elaboration with a distinct set of input data. The description of each model specifies the types of data on which the model works, whereas each simulation has specific data as its input.

My analysis of the papers reveals that papers of Type 1 never refer to simulation models. Data are the representation of the epistemic thing; they are interpreted as a sequence of DNA. *In silico* experimentation involves manipulation of data by means of some computer operations. Programs are chosen on the basis of some particular objectives. Generally these programs are not based on a model of a target system, but are, rather, algorithms for analysing data; they are black boxes that transform data. Papers refer to them only by means of their name and the type of processing they are able to make. We know the type of calculation they make, but nobody is interested in all the modalities through which they realize their result. They constitute some particular experimental conditions, more or less standardized, that allow us to investigate the structure of DNA. Such conditions complement those present in *in vivo* and *in vitro* experiments. The identity of an *in silico* experiment is determined by programs that realize some specific operations. Elements that come into play are input data and functions that allow us to acquire new output data. The definition of these functions agrees with the software that realizes them; in fact, whenever papers mention some computational operations, they always refer to the name of a computer program, both in their "methods" sections, where experimental conditions are described in detail, and in their main part. The validity of these programs is not questioned. There are no bibliographic references to papers that describe the model that forms the basis of their definition, or which discuss their potentiality or limitations. For the most part there is only a reference to a web site where it is possible to download the program[1]. These programs do not define actual experimental processes, but are, rather, types of tool, used in experimental practice, along with other tools necessary for material manipulation.

In all papers of Type 2, the expression "*in silico* experiment" has a different meaning, and it agrees with the meaning of "set of simulations". Each paper refers to a particular model, different from all the others. In reality, each model is complex, because it is a composition of other, simpler, models, following a hierarchic structure. Discussion mostly concerns the definition of the complex model, whereas the simpler models that compose it are only mentioned,

and their description is deferred to other papers. For example, in a model describing a series of metabolic pathways, chemical reactions are represented by means of stoichiometry, but often stoichiometric equations are not shown; only reagents and products that constitute the nodes of this network are specified. In these papers, models are representations of some metabolic networks, and are therefore inscriptions that allow us to visualize the invisible. Unlike simple "maps", simulation models are "dynamic".

As above, a reference to a model can follow two different modalities. In case 2.1. of the foregoing paragraph, a model represents some assumptions of a theory, and simulation has to verify them by means of a comparison with data obtained through other experimental conditions – namely, *in vivo* or *in vitro* experiments. Papers that follow this modality focus mostly on an analysis of the relationship between simulation results and experimental results.

> [T]he simulations show significant agreement with experimental results, thus establishing the utility of the model (Alvarez-Vasquez et al., 2005: 429).

> As can be seen in Fig. 3b–g, the agreement between the behaviour predicted by the model and the experimental results is very good, validating the bottom-up approach to understanding gene regulatory networks (Guido et al., 2006: 858).

> Comparison with the growth phenotypes showed that experimental and computational outcomes agreed in 10,828 (78.7%) of the cases examined, which is roughly the same success rate achieved in previous studies in E. coli and yeast that considered only a few hundred phenotype (Covert et al., 2004: 94).

The core of the paper does not regard the theoretical principles that form the basis of each model, but is concerned, instead, with the degree of the model's agreement with inscriptions obtained in other ways. Usually, in this type of discussion the model is not described in a detailed manner; some bibliographic references are included, where it is possible to find some details and justifications of the choices of authors. Therefore these papers focus on "simulations" rather than on "models". They show those parameters associated with a model that represents some specific initial conditions; they obtain outputs and then compare them with data originating from *in vivo* or *in vitro* experiments. This comparison is realized by means of some charts that contain, simultaneously, the representations of results of both a simulation and some *in vivo* or *in vitro* experiments. In essence, the objective is to establish a relationship between the "*in silico* experiment", on the one hand, and the "*in vivo* or *in vitro* experiment", on the other, so as to consider the former a valid substitution of the latter and then to render the theoretical model underlying the "*in silico* experiment" legitimate. In "methods" sections and in "supplementary" files, the structure and the construction of a model are described, but often in a very generic manner. Normally, databases, mathematical and statistical functions and programs are listed, or some bibliographic references are given, to enable the reader to find a detailed description elsewhere. In these sections, descriptions are always found, both of simulations with regard to initial conditions and some particular situations of each application of model, and of *in vivo* or *in vitro* experiments.

In modalities shown in case 2.2., models again represent some assumptions of a theory, but simulations aim to infer new theoretical propositions. Papers focus on some intrinsic characteristics of models.

Their elements are always described in a rather detailed manner in the central part of the text, almost always with accompanying graphical visualizations that render their understanding more intuitive. In "methods" sections and, above all, in "supplementary" files, some details of models are described – functions used to represent the links between the components of a network, objective functions, linear equations for constraints systems, logical rules in logico-mathematical systems (such as Boolean networks), and so on – and the list of databases from which authors have extracted data and programs to define and simulate models. Simulations assume different roles in cases 2.1 and 2.2. In case 2.1., authors focus on "results" that the simulation produces, whereas in case 2.2. they concentrate on its internal mechanisms. It becomes an experiment on the theory that underlies the definition of the model.

> Using an in silico representation of the metabolic network of Escherichia coli, we examine the role of contingency by repeatedly simulating the successive loss of genes while controlling for the environment (Pál et al., 2006: 667).

> Under laboratory conditions 80% of yeast genes seem not to be essential for viability. This raises the question of what the mechanistic basis for dispensability is, and whether it is the result of selection for buffering or an incidental side product. Here we analyse these issues using an in silico flux mode of the yeast metabolic network (Papp et al., 2004: 661).

> Here, we devise a theoretical method for simultaneously predicting key aspects of network functionality, robustness and gene regulation from network structure alone (Stelling et al., 2002: 190).

The identity of the *in silico* experiment depends on the model rather than on the simulation; it is the referent of the metabolic network of a particular living system. Papers of Type 2.2 discuss, above all, some elements that constitute the model, and how the *in silico* experiment allows us to analyse the theoretical mechanisms that underlie a particular process. In papers of Type 2.1, by contrast, the discussion centres on the performance of the simulation, on its results, and on the correspondence with results obtained by means of *in vivo* and *in vitro* experiments.

To conclude, we maintain that in papers of Type 1 the *in silico* experiment is a "technical object", an experimental condition that is added and mixed with other conditions that are part of traditional experiment.

By contrast, in all situations described in papers of Type 2, the *in silico* experiment is used to analyse metabolic networks in order to find new knowledge on them. The experiment consists of the manipulation of a model, in an experimental setting provided by the computer system. A model arises from a reconfiguration of some epistemic things in a new context. It allows scientists to generate new questions, and thus becomes a new "question-generating machine".

In case 2.1., authors wish to create a relationship between a new object and another well-known and well-established object; they seek to prove that the *in silico* experiment is representative of the *in vivo* or *in vitro* experiment. In essence, they aim to represent an epistemic thing in a new way, and then to define the shift from a graphematic space to another one (cf. Rheinberger, 1997: 102–13). Mathematical, statistical and computational tools – such as linear equations, the Montecarlo method, stoicometric equations, programming language, and so on – allow the defi-

nition of models and their simulation, in turn allowing this new representation.

In case 2.2., the *in silico* experiment has to analyse metabolic networks and find new knowledge about them. The resulting epistemic thing, now reconfigured in a new context, allows the generation of new questions. The representation of a metabolic network by means of a model involves a reconfiguration of this object, so that it can then be manipulated by computer simulations. The *in silico* experiment allows scientists to query epistemic things in new contexts.

As observed above, the authors in my sample never express an intention to standardize a model. It therefore remains contingent and short-lived, and it cannot be turned into a technical object.

Each model is a partial point of view of a specific phenomenon; one among all possible points of view. Therefore a way of identifying a modality of selection that is convincing to the scientific community is needed. The pragmatic function accomplishes this task. The rhetorical strategies used by authors to justify the use of *in silico* experiments try to show that a simulation has some meaning if it pursues a pragmatic objective. A model presents a particular interpretation of a target system. It does not have the same aim as that of the universal laws of physics, for example, but it is valid if it is able to intervene usefully in empirical reality. The pragmatic objective determines the level of observation, allowing the isolation of its essential function (Negrotti, 1997). In particular, frequent references appear regarding the contribution to medical and pharmaceutical research:

> [I]t is possible that discriminating between date and party hubs might also help to define new therapeutic drug targets (Han et al., 2004: 92).

> The network that resulted is a functional description of the eukaryotic proteome at a higher level of organization. Such higher-order maps will bring an increasing quality to our appreciation of biological systems. It is expected that this may provide drug discovery programmes with a molecular context for the choice and evaluation of drug targets (Gavin et al., 2002: 146).

> The reprogramming of DNA-binding specificity is an important challenge for computational protein design that tests current understanding of protein-DNA recognition, and has considerable practical relevance for biotechnology and medicine (Ashworth et al., 2006: 656).

> [W]e hope to elucidate biological function as well as predict the effect of internal perturbations (for example, genetic mutations) or external perturbations (for example, drugs) so that disease treatments are more precise and effective. Similarly, understanding biological modules and being able to engineer new ones will pave the way for re-engineering of organisms and cells for numerous applications (including medical, agricultural and ecological situations) (Di Ventura et al., 2006: 532)[2].

The epistemic value of models derives from their capacity to increase the future "manipulability" of biological material. Each model and its simulations allow us to analyse some particular aspect of an epistemic thing, and to obtain new information, some of which can then be used in *in vivo* or *in vitro* experiments. We may add that the validity of a model or a simulation does not depend on its capacity to represent an exemplar well, but rather on its ability to manage and modify living matter. That is why authors often try to demon-

strate that the manipulation of models by means of simulations can be reproduced in the corresponding physical objects, or can be a useful guide to physical handling (in the sense that the simulation gives some information about structures and processes of epistemic things).

Concluding Remarks

The studies on the use of databases and simulation models in biology indicate that these new technologies need not lead to the birth of new epistemic cultures, they rather add new resources inserted into the already consolidated research practices. The use of the metaphor of "information" has led to the construction of new representations of nature, which involve also the discursive and material practices (Kay, 2000). This transformation, though, does not spark a revolution that brings about new visions of the world, but instead tends to reflect the natural and social orders that already exist (Hine, 2006; Wieber, 2009). The use of simulation models is, however, always orientated towards material manipulation, following practical aims. For this reason, scientists are interested in demonstrating, not the "validity" of the models, but their "utility" (Keller, 2000). Scientists choose their models on the basis of their ability to satisfy particular practical and theoretical exigencies, and to reach specific objectives. The models are tools to aid "thinking" and also "doing", allowing scientists to reflect upon and modify the theory, and also to intervene materially. In my analysis, this aspect is confirmed by the fact that the authors of the papers justify the validity of their models by their ability to manage and modify living systems.

Analysing the scientific papers on molecular biology may help us to understand better how the new computational technologies may be inserted into research practices, considering, in particular, the case of the *in silico* experiment. To do this, I considered the concepts of "epistemic thing" and "technical object" defined by Rheinberger (1997). I discovered that *in silico* experiments are regarded in several different ways that are fundamentally reducible to two basic types. In papers of Type 1, the *in silico* experiment allows us to operate on it by means of some tools that, in some cases, are simply new experimental conditions that we can place alongside traditional conditions, in order to obtain new data of a functionally equivalent type. In essence, the scientists use new tools in order to be able to work on data presented in digital form. Such tools supplement those that characterize the traditional experimental setting. As Wieber (2009) says, the *in silico* experiment has become a part of the toolbox used by molecular biologists to elaborate and interpret empirical data.

In Type 2 papers, the *in silico* experiment is introduced with many different functions: it is the set of simulations that manipulate a model, given the experimental conditions provided by the computer system. A model has the role of locally reshaping some different epistemic things and creating new ones. The meaning of such a model is contingent, as is, *a fortiori*, the meaning of "simulation". A model is the product of a theoretical reflection of researchers regarding a particular biological phenomenon. This reflection can give rise to a visible object that can act and create new question-generating machines. In this way, we observe a proliferation of epistemic things that make the scientific context most varied and non-standardized.

As a matter of fact, each model is only partial. It embodies only one particular vision of a biological phenomenon. The objective of researchers is not to create a standard model that they can use and re-

use in different contexts, but only to rearrange locally various epistemic things, following a theoretical reflection. Indeed, only rarely are models recorded in a shared database. They always constitute some partial visions of an object, and we do not come across any work aiming to standardize models. Each paper begins with a particular problem, and it defines a model relating to this problem, starting from a series of assumptions that will render the model just one of many possible models able to represent the object in question.

In this sense, models do not follow the dynamics described by Rheinberger, who sees objects as oscillating between epistemic things and technical objects (Rheinberger, 1997). This depends not only on the open nature of the object, as Merz (1999) maintains, but also on its contingency; it is accepted for the moment for its pragmatic value, because it is useful in some particular type of application, but then it disappears, and never reappears as a technical object. The modelling of a system is not intended to lead to the discovery of the truth, and therefore to the construction of mutually coherent models, but must be useful for the material manipulation. The pragmatic function, as Keller (2000) says, guides the use of models. *In silico* experiments allow us to work on particular arrangements of epistemic things that we can define as "disposable", because they are partial and subjective "interpretations" of the object that we are studying. Consequently, *in silico* experiments are used locally to generate new theoretical reflections that are useful in a specific context. This new knowledge will successively be redefined and used in other contexts, through the definition of new models and new simulations. We thus see a continuous de-contextualization and re-conceptualization of epistemic things that generate a multiplicity of new and provisional forms when they are recombined.

The aforementioned points arise from the analysis of how scientists present their work for the scientific audience, reflecting the public face of molecular biology. A deeper understanding of *in silico* experiments in molecular biology should take into account also the processes that generate and define their criteria of inclusion, acceptance and validity. This would require the use of other methods that analyse the everyday practices of scientists, and study how *in silico* experiments are inserted into the existing social and epistemological contexts.

Notes

[0] This article was edited and approved for publication by Tarja Knuuttila.

[1] Only in one of the papers analysed did the authors use a program specifically created for their application. Yet even in this case there was no discussion about the definition of the algorithm that formed the basis of this program; in the "methods" section it was described very perfunctorily in a few lines.

[2] This paper is not one of the 48 papers analysed, because it does not describe an application using *in silico* experimentation, but presents, rather, a review of the use of *in silico* experiments in molecular biology.

References

Alvarez-Vasquez, F., K.J. Sims, L.A. Cowart, Y.O. Okamoto, E. Voit & Y.A. Hannun (2005) 'Simulation and validation of modelled sphingolipid metabolism in Saccharomyces cerevisiae', Nature 433: 425-30.

Ashworth, J., J.J. Havranek, C.M. Duarte, D. Sussman, R.J. Monnat, B.L. Stoddard

& D. Baker (2006) 'Computational redesign of endonuclease DNA binding and cleavage specificity', Nature 441: 656-9.

Covert, M.W., E.M. Knight, J.L. Reed, M.J. Herrgard & B.O. Palsson (2004) 'Integrating high-throughput and computational data elucidates bacterial networks', Nature 429: 92-6.

Dahan Dalmedico, A. (2000) 'Between Models as Structures and Models as Fictions: Computer Modeling Practices in post World War II', paper presented at the Princeton Workshop on Model Systems, Cases and Exemplary Narratives, 12 February 2000, in Princeton, N.J.

Di Ventura, B., C. Lemerle, K. Michalodimitrakis & L. Serrano (2006) 'From in vivo to in silico biology and back', Nature 443: 527-33.

Frohmann, B. (1999) 'The Role of the Scientific Paper in Science Information Systems', in M.E. Bowden, T.B. Hahn & R. Williams (eds), History of Information Science: Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems (Medford NJ: Information Today): 63-73.

Galison, P. (1997) Image and Logic: A Material Culture of Microphysics (Chicago, IL: University of Chicago Press).

Galperin, M.Y. & G.R. Cochrane (2009) 'Nucleic Acids Research annual Database Issue and the NAR online Molecular Biology Database Collection in 2009', Nucleic Acids Research 37, http://nar.oxfordjournals.org/cgi/content/abstract/37/suppl_1/D1 (accessed 20 December 2009).

Gavin, A., M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J.M. Rick, A.M. Michon, C.M. Cruciat et al. (2002) 'Functional organization of the yeast proteome by systematic analysis of protein complexes', Nature 415: 141-7.

Guido, N.J., X. Wang, D. Adalsteinsson, D. McMillen, J. Hasty, C.R. Cantor, T.C. Elston & J.J. Collins (2006) 'A bottom-up approach to gene regulation', Nature 439: 856-60.

Han, J.J., N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J.M. Walhout, M.E. Cusick, F.P. Roth et al. (2004) 'Evidence for dynamically organized modularity in the yeast protein–protein interaction network', Nature 430: 88-93.

Hine, C. (2006) 'Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work', Social Studies of Science 36(2): 269-98.

Hughes, R.I.G. (1999) 'The Ising Model, Computer Simulation, and Universal Physics', in M.S. Morgan & M. Morrison (eds), Models as mediators: perspectives on natural and social science (Cambridge: Cambridge University Press): 97-145.

Kay, L.E. (2000) Who Wrote the Book of Life?: A History of the Genetic Code (Stanford, Calif.: Stanford University Press).

Keller, E.F. (2000) 'Models of and Models for: Theory and Practice in Contemporary Biology', Philosophy of Science 67, Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association. Part II: Symposia Papers: S72-S86.

Knorr Cetina, K. (1981) The manufacture of knowledge: An essay on the constructivist and contextual nature of science (Oxford: Pergamon).

Knorr Cetina, K. (1997) 'Sociality with objects: Social relations in postsocial knowledge societies', Theory, Culture and Society 14(4): 1-30.

Knorr Cetina, K. (2001) 'Objectual practice', in T.R. Schatzki, K. Knorr Cetina & E. von Savigny (eds), The practice turn in contemporary theory (London: Routledge): 175-88.

Knuuttila, T. & A. Voutilainen (2003) 'A Parser as an Epistemic Artifact: A Mate-

rial View on Models', Philosophy of Science 70(5): 1484-95.
Knuuttila, T., M. Merz & E. Mattila (2006) 'Productive and Contested: Computer Models and Simulations in Scientific Practice', Science Studies 19(1): 3-11.
Leonelli, S. (2007) 'What Is In A Model?', in M. Laubichler & G.B. Müller (eds.), Modeling Biology. Structures, Behaviours, Evolution (Vienna Series: MIT Press): 15-36.
Leonelli, S. (2009) 'On the Locality of Data and Claims about Phenomena', Philosophy of Science 76(5): 737-49.
Merz, M. (1999) 'Multiplex and Unfolding: Computer Simulation in Particle Physics', Science in Context 12(2): 293-316.
Morgan, M.S. & M. Morrison (eds.) (1999) Models as mediators: perspectives on natural and social science (Cambridge: Cambridge University Press).
Negrotti, M. (1997) La terza realtà (Bari: Edizioni Dedalo).
Pál, C., B. Papp, M.J. Lercher, P. Csermely, S.G. Oliver & L.D. Hurst (2006) 'Chance and necessity in the evolution of minimal metabolic networks', Nature 440: 667-70.
Papp, B., C. Pál & L.D. Hurst (2004) 'Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast', Nature 429: 661-4.
Parker, W.S. (2009) 'Does Matter Really Matter? Computer Simulations, Experiments, and Materiality', Synthese 169(3): 483-96.
Rheinberger, H.J. (1997) Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube (Stanford, CA: Stanford University Press).
Rohrlich, F. (1991) 'Computer Simulation', Physical Sciences. Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association 2: 507-18.
Searls, D.B. (2010) 'The Roots of Bioinformatics', PLoS Computational Biology 6(6): 1-7.
Sismondo, S. (1999) 'Modeling and Simulation', Science in Context, Special Issue 12(2): 247-60.
Stelling, J., S. Klamt, K. Bettenbrock, S. Schuster & E.D. Gilles (2002) 'Metabolic network structure determines key aspects of functionality and regulation', Nature 420: 190-3.
Sunberg, M. (2008) 'The Everyday World of Simulation Modeling. The Development of Parameterizations in Meteorology', Science, Technology, & Human Values 20(10): 1-20.
Valle, G., M. Helmer Citterich, M. Attimonelli & G. Pesole (2007) Introduzione alla bioinformatica (Bologna: Zanichelli).
Wieber, F. (2009) 'Theoretical Technologies in an "Experimental" Setting: Empirical Modeling of Proteinic Objects and Simulation of their Dynamics Within Scientific Collaborations Around a Supercomputer', in: [2009] SPSP 2009: Society for Philosophy of Science in Practice (Minnesota, June 18-20, 2009).
Yuh, C.H., H. Bolouri & E.H. Davidson (1998) 'Genomic Cis-regulatory Logic: Experimental and Computational Analysis of a Sea Urchin Gene', Science 279: 1896-902.

Appendix: Lists of the analysed papers

1. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási (2000) 'The large-scale organization of metabolic networks', Nature 407: 651-4 (Letter).
2. Human Genome (2001) 'A physical map of the human genome', Nature 409: 934-41.
3. Human Genome (2001) 'Initial sequencing and analysis of the human genome', Nature 409: 860-921.
4. S.T. Cole, K. Eiglmeier, J. Parkhill, K. D. James, N.R. Thomson, P.R. Wheeler, N. Honoré, T. Garnier, C. Churcher, D.

Harris et al. (2001) 'Massive gene decay in the leprosy bacillus', Nature 409: 1007-11.
5. A.C. Gavin, M. Bösche, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.M. Michon, C.M. Cruciat et al. (2002) 'Functional organization of the yeast proteome by systematic analysis of protein complexes', Nature 415: 141-7.
6. S.D. Bentley, K.F. Chater, A.M. Cerdeño-Tárraga, G.L. Challis, N.R. Thomson, K.D. James, D.E. Harris, M.A. Quail, H. Kieser, D. Harper et al. (2002) 'Complete genome sequence of the model actinomycete Streptomyces coelicolor', Nature 417: 141-7.
7. R.U. Ibarra, J.S. Edwards & B.O. Palsson (2002) 'Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth', Nature 420: 186-9 (Letter).
8. J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster & E.D. Gilles (2002) 'Metabolic network structure determines key aspects of functionality and regulation', Nature 420: 190-3 (Letter).
9. P. Szabó, I. Scheuring, T. Czárán & E. Szathmáry (2002) 'In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity', Nature 420: 340-3 (Letter).
10. The mouse genome (2002) 'A gene expression map of human chromosome 21 orthologues in the mouse', Nature 420: 586-90.
11. R. Heilig, R. Eckenberg, J.L. Petit, N. Fonknechten, C. Da Silva, L. Cattolico, M. Levy, V. Barbe, Véronique de Berardinis, A. Ureta-Vidal et al. (2003) 'The DNA sequence and analysis of human chromosome', Nature 421: 601-7.
12. J.W. Thomas, J.W. Touchman, R.W. Blakesley, G.G. Bouffard, S.M. Beckstrom-Sternberg, E.H. Margulies, M. Blanchette, A.C. Siepel, P.J. Thomas, J.C. McDowell et al. (2003) 'Comparative analyses of multi-species sequences from targeted genomic regions', Nature 424: 788-93 (Letter).
13. E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai & A.L. Barabási (2004) 'Global organization of metabolic fluxes in the bacterium Escherichia coli', Nature 427: 839-43 (Letter).
14. M.W. Covert, E.M. Knight, J.L. Reed, M.J. Herrgard & B.O. Palsson (2004) 'Integrating high-throughput and computational data elucidates bacterial networks', Nature 429: 92-6 (Letter).
15. J.J. Han, N. Bertin, T. Hao, D.S. Goldberg, G.F. Berriz, L.V. Zhang, D. Dupuy, A.J.M. Walhout, M.E. Cusick, F.P. Roth et al. (2004) 'Evidence for dynamically organized modularity in the yeast protein–protein interaction network', Nature 430: 88-93 (Letter).
16. P. Oh, Y. Li, J. Yu, E. Durr, K.M. Krasinska, L.A. Carver, J.E. Testa & J.E. Schnitzer (2004) 'Subtractive proteomic mapping of the endothelial surface in lung and solid tumours for tissue-specific therapy', Nature 429: 629-35.
17. B. Papp, C. Pál & L.D. Hurst (2004) 'Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast', Nature 429: 661-4 (Letter).
18. J. Jaeger, S. Surkova, M. Blagov, H. Janssens, D. Kosman, K.N. Kozlov, M.E. Myasnikova, C.E. Vanario-Alonso, M. Samsonova et al. (2004) 'Dynamic control of positional information in the early Drosophila embryo', Nature 430: 368-71 (Letter).

19. X. She, Z. Jiang, R.A. Clark, G. Liu, Z. Cheng, E. Tuzun, D.M. Church, G. Sutton, A.L. Halpern & E.E. Eichler (2004) 'Shotgun sequence assembly and recent segmental duplications within the human genome', Nature 431: 927-30.
20. J.T. Wade, D.B. Hall & K. Struhl (2004) 'The transcription factor Ifh1 is a key regulator of yeast ribosomal protein genes', Nature 432: 1054-8 (Letter).
21. F. Alvarez-Vasquez, K.J. Sims, L.A. Cowart, Y. Okamoto, E.O. Voit & Y.A. Hannun (2005) 'Simulation and validation of modelled sphingolipid metabolism in Saccharomyces cerevisiae', Nature 433: 425-30 (Letter).
22. D.B. van Rossum, R.L. Patterson, S. Sharma, R.K. Barrow, M. Kornberg & D.L. Gill, S.H. Snyder (2005) 'Phospholipase Cγ1 controls surface expression of TRPC3 through an intermolecular PH domain', Nature 434: 99-104 (Letter).
23. L.W. Hillier, T.A. Graves, R.S. Fulton, L.A. Fulton, K.H. Pepin, P. Minx, C. Wagner-McPherson, D. Layman, K. Wylie, M. Sekhon et al. (2005) 'Generation and annotation of the DNA sequences of human chromosomes 2 and 4', Nature 434: 724-31.
24. L. Eichinger, J.A. Pachebat, G. Glöckner, M.A. Rajandream, R. Sucgang, M. Berriman, J. Song, R. Olsen, K. Szafranski, Q. Xu et al. (2005) 'The genome of the social amoeba Dictyostelium discoideum', Nature 435: 43-57.
25. C.S. Sullivan, A.T. Grundhoff, S. Tevethia, J.M. Pipas & D. Ganem (2005) 'SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells', Nature 435: 682-6 (Letter).
26. Y. Zhao, E. Samal & D. Srivastava (2005) 'Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis', Nature 436: 214-20.
27. E. Dekel & U. Alon (2005) 'Optimality and evolutionary tuning of the expression level of a protein', Nature 436: 588-92 (Letter).
28. International Rice Genome Sequencing Project (2005) 'The map-based sequence of the rice genome', Nature 436: 793-800.
29. N.J. Guido, X. Wang, D. Adalsteinsson, D. McMillen, J. Hasty, C.R. Cantor, T.C. Elston & J.J. Collins (2006) 'A bottom-up approach to gene regulation', Nature 439: 856-60 (Letter).
30. R. Broadhead, H.R. Dawe, H.Farr, S. Griffiths, S.R. Hart, N. Portman, M.K. Shaw, M.L. Ginger, SJ. Gaskell, P.G. McKean et al. (2006) 'Flagellar motility is required for the viability of the bloodstream trypanosome', Nature 440: 224-7 (Letter).
31. C. Pál, B. Papp, M.J. Lercher, P. Csermely, S.G. Oliver & L.D. Hurst (2006) 'Chance and necessity in the evolution of minimal metabolic networks', Nature 440: 667-70 (Letter).
32. J. Ashworth, J.J. Havranek, C.M. Duarte, D. Sussman, R.J. Monnat, B. L. Stoddard & D. Baker (2006) 'Computational redesign of endonuclease DNA binding and cleavage specificity', Nature 441: 656-9 (Letter).
33. K. Horikawa, K. Ishimatsu, E. Yoshimoto, S. Kondo & H. Takeda (2006) 'Noise-resistant and synchronized oscillation of the segmentation clock', Nature 441: 719-23.
34. M. Alleman, L. Sidorenko, K. McGinnis, V. Seshadri, J.E. Dorweiler, J. White, K. Sikkink & V.L. Chandler (2006) 'An

RNA-dependent RNA polymerase is required for paramutation in maize', Nature 442: 295-8 (Letter).
35. O. Shlyk-Kerner, I. Samish, D. Kaftan, N. Holland, P.S. Maruthi Sai, H. Kless & A. Scherz (2006) 'Protein flexibility acclimatizes photosynthetic energy conversion to the ambient temperature', Nature 442: 827-30 (Letter).
36. D. Chourrout, F. Delsuc, P. Chourrout, R.B. Edvardsen, F. Rentzsch, E. Renfer, M.F. Jensen, B. Zhu, P. de Jong, R.E. Steele et al. (2006) 'Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements', Nature 442: 684-7 (Letter).
37. S. Leininger, T. Urich, M. Schloter, L. Schwark, J. Qi, G.W. Nicol, J.I. Prosser, S.C. Schuster & C. Schleper (2006) 'Archaea predominate among ammonia-oxidizing prokaryotes in soils', Nature 442: 806-9 (Letter).
38. A. Sigal, R. Milo, A. Cohen, N. Geva-Zatorsky, Y. Klein, Y. Liron, N. Rosenfeld, T. Danon, N. Perzov & U. Alon (2006) 'Variability and memory of protein levels in human cells', Nature 444: 643-6 (Letter).
39. E.S. Lein, M.J. Hawrylycz, N. Ao, M. Ayres, A. Bensinger, A. Bernard, A.F. Boe, M.S. Boguski, K.S. Brockway, E.J. Byrnes et al. (2006) 'Genome-wide atlas of gene expression in the adult mouse brain', Nature 445: 168-76.
40. T.S. Mikkelsen, M.J. Wakefield, B. Aken, C.T. Amemiya, J.L. Chang, S. Duke, M. Garber, A.J. Gentles, L. Goodstadt, A. Heger et al. (2007) 'Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences', Nature 447: 167-77.
41. The Wellcome Trust Case Control Consortium (2007) 'Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls', Nature 447: 661-78.
42. B. Qian, S. Raman, R. Das, P. Bradley, A.J. McCoy, R.J. Read & D. Baker (2007) 'High-resolution structure prediction and the crystallographic phase problem', Nature 450: 259-64.
43. The International HapMap Consortium (2007) 'A second generation human haplotype map of over 3.1 million SNPs', Nature 449: 851-61.
44. Drosophila 12 Genomes Consortium (2007) 'Evolution of genes and genomes on the Drosophila phylogeny', Nature 450: 203-18.
45. K.A. Henzler-Wildman, M. Lei, V. Thai, S.J. Kerns, M. Karplus & D. Kern (2007) 'A hierarchy of timescales in protein dynamics is linked to enzyme catalysis', Nature 450: 913-16 (Letter).
46. E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall & U. Gaul (2008) 'Predicting expression patterns from regulatory sequence in Drosophila segmentation', Nature 451: 535-40.
47. J.M. Carlton, J.H. Adams, J.C. Silva, S.L. Bidwell, H. Lorenzi, E. Caler, J. Crabtree, S.V. Angiuoli, E.F. Merino, P. Amedeo et al. (2008) 'Comparative genomics of the neglected human malaria parasite Plasmodium vivax', Nature 455: 757-63.
48. H. Zheng, H. Ying, H. Yan, A.C. Kimmelman, D.J. Hiller, A.J. Chen, S.R. Perry, G. Tonon, G.C. Chu, Z. Ding et al. (2008) 'p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation', Nature 455: 1129-33 (Letter).

By means of the text analysis, these papers are classified in two types. The papers 2, 3, 4, 5, 6, 10, 11, 12, 16, 19, 20, 22, 23, 24, 25, 26, 28, 30, 34, 35, 36, 37, 39, 40, 41, 43, 44, 47, 48 belong to Type 1.

The papers 1, 7, 8, 9, 13, 14, 15, 17, 18, 21, 29, 31, 32, 33, 45, 46 belong to Type 2.

The papers 27, 38, 42 describe some particular applications that we can not classify in any defined typology.

Sabrina Moretti
Universiy of Urbino
Via Saffi, 15, 61029 Urbino, Italy
sabrina.moretti@uniurb.it