

updated December 2nd, 2015

CONTRIBUTIONS TO MATHEMATICAL BIOLOGY

JACQUES NINIO

(included in the web site <http://www.lps.ens.fr/~ninio>)

TOPICS DISCUSSED HERE :

General comments on mathematical biology

How to be useful

The biologist's understanding of mathematics

Contributions to biomathematics

Probabilistic tools in enzyme kinetics

Probabilistic derivations

The peelback problem

The Russian school

First passage times in enzymology

RNA topology

A link between vision and molecular biology

Appendix: on models in times of consensus

GENERAL COMMENTS ON MATHEMATICAL BIOLOGY

HOW TO BE USEFUL

I do not belong to the community of mathematical biologists. Often, I meet students or researchers trained in mathematics, physics or informatics, who tell me about their desire to apply their competence to a pertinent biological problem. Invariably, I tell them that the first motivation in a research work should be the interest of the subject itself. So, they have to determine first which biological question is of the greatest interest to them, then work with humility in a laboratory in which the question is addressed. Later, at one time or another, the fact that they have a different background from that of the other researchers in the field could inspire to them questions that their colleagues did not raise, or let them see features, or implications in the data that the

colleagues did not notice, or allow them to devise new tools. But they should not put the tool before the subject.

Mathematicians or physicists often make a deep mistake when they try to "model" biological data. Actually, the published data are heavily contaminated by the implicit or explicit models of the time. The experimenters design an experiment, selecting what they believe are the important questions, and the pertinent procedures and controls. They decide to discard incomprehensible results, and they focus on the results that make sense, with which they will be able to present a coherent story that will be acceptable by the reviewers. Actually, they are constructing a piece of reality that fits their prejudices. Therefore, the mathematicians or the physicists do not model the data. They make instead a formal model of the experimenters' conceptual world. More on this topic in Appendix 1 below.

Some models, it is true, are made in complete ignorance of experimental publications, which is worse. As long as their authors stand alone in defending their models, the consequences are small. But when an army of modelling specialists comes next, and explores all the consequences of the initial model, still without seeking any contact with experimental reality, this becomes damaging to science. I have witnessed the development of such a situation for the first time in relation to Manfred Eigen's 1971 initial papers on the origins of life [1]. But now, vacuous "virtual reality" science is becoming a common pathology in many disciplines.

THE BIOLOGIST'S UNDERSTANDING OF MATHEMATICS

Of course, much of the biological work performed today is not motivated by scientific questions, it is merely technology-driven research. "Using fashionable tools to address fashionable questions" could be the generic title for most grant applications. Accordingly, this type of work produces results devoid of real conceptual or practical interest.

Most biologists have an extremely poor understanding of mathematics, or even logics:

— *Confusion between a proposition and the reciprocal proposition.* In logical thinking, the statement "A implies B" is not equivalent to the reciprocal statement "B implies A". In biological writings, if A implies B then when B is observed, "it is likely" that A is true.

— *Deductive versus inductive reasoning.* Whereas in Latin, Slavic, Arabic or Indian cultures, deductive reasoning is valued, it is rejected in the English-speaking biological tradition, which puts faith, instead, in inductive reasoning (there are a few fortunate exceptions, among which Crick's "wobble hypothesis" article, which uses deductive reasoning [2]). Otherwise a perfectly rigorous piece of work, from the point of view of deductive reasoning, may appear to a native English-speaking biological reviewer as a piece of "complete and utter nonsense which has no place in a supposedly serious scientific article" - a kind of comment my articles often receive (See the web chapter on contributions to memory). On the other side, many highly valued theoretical papers use inductive reasoning, and appear to me like dog's breakfast.

In practice, there is in a theoretical work, a cluster of starting ideas a, b, c, d ... and a cluster of relevant experimental facts u, v, w, x In a deductive mode, one would

work out the theoretical consequences of a, b, c, d ... and reach a theoretical proposal P. Then the whole would be confronted with the set of facts u, v, w, ... and the author would determine how to modify the theory, or how to re-evaluate the facts to seek a better agreement between facts and theory. In inductive reasoning, one takes idea a and fact u, and combines them into a hybrid idea i, then one takes ideas i and b, combines them with fact v, and derives the half-baked idea j, and so on. This is what most reviewers demand.

As a consequence of this state of affairs, I have had, in a number of cases, to hide to the readers the true heuristic line of reasoning that led me to a (valid) conclusion, and present instead some inductive stinking substitute, more palatable to the reviewer.

Happily, native English-speaking mathematicians and physicists do use deductive reasoning. Otherwise, they would not survive in their discipline.

— *Failure to see the threads.* Mathematicians strive for elegant proofs, that go to the heart of things, and can be expressed as concisely as possible. If a mathematician works out a proof that takes twenty pages of equations, then finds a way to go straight to the conclusion in three sentences without equations, then he is proud of having made this big step forward. In physics, there is the art of the "back of the envelope" calculations. Top physicists may reach important conclusions by doing mathematically elementary order-of-magnitude calculations. The simplicity of these calculations is deceptive. You need to really understand what you are doing to succeed. In mathematical and physical reasoning, there are threads: you know from where you start, you know where you are heading, and the succession of steps that will take you there. All this is foreign to most biologists. For most biologists, doing mathematics is writing down lots of equations, combining them to derive other equations, and continue the mixing and extraction procedures until, at some point, the desired result miraculously emerges. So they are impressed by mathematical treatments that cover pages and pages, although the treatments may be totally uninspired and if they happen to be correct, could be replaced with much shorter derivations.

Bonus. A nice example of mathematical illiteracy : « "The smallest prime number greater than one" uniquely identifies "three" (---) ». The authors of this delightful statement are Charles R. Gallistel, member of the National Academy of Sciences of the USA and Rochel Gelman. The quotation is from a book chapter in "Numerical cognition" (Stanislas Dehaene, editor, Blackwell, 1993, page 45).

CONTRIBUTIONS TO BIOMATHEMATICS

Mathematics is the discipline I liked best in my studies. My favourite branches in mathematics were geometry, analytical geometry, algebra and probabilities. I did not like integrals and differential equations. Although I do not have the level for doing mathematical research, and although my mathematical competence has declined over the years, I keep some lucidity in seeing the threads of a reasoning, and am quite perspicacious at detecting the mistakes. Whereas I work in a physics laboratory, I feel that I am more a mathematician, or an engineer than a physicist. My understanding of chemistry is extremely poor, and this restricted my capacity to contribute to the origins of life. Occasionally, I use mathematics in my work, or "back of the envelope" calculations. But once I get the biologically pertinent results, I stop there and do not try

to push the mathematics further.

While doing my thesis work, Piotr Slonimski submitted to me a problem in the interpretation of pictures taken by electron microscopy. It was, I think about having series of slices taken from yeast cells, and reconstructing the shape, assumed to be ellipsoidal, from the slices. I found within a week an elegant solution to the problem, equating (if I remember well) two expressions of the volume, one as a Riemann and one as a Lebesgue integral. But Slonimski lost interest in the problem, and the calculations remained in the drawers.

In the thesis work, I described a method (using Fourier transform) for calculating the theoretical X-ray scattering curve from a molecule in solution [3, 4]. The mathematical part is Luzzati's merit (see section on bio-informatics). The method is still considered as basically sound, but there has been progress on how to take into account the contribution of the hydration shell of the molecule [5, 6].

At the time of the thesis work, I also had the ambition to develop a classification of all possible RNA structures, and planned a theoretical work in four parts, entitled: "Properties of nucleic acid representations". Part 1 dealt with topological properties, and was published in 1971 [7] (see "RNA topology" below). Part 2 was about orientational properties (questions such as which side of the nucleotide is "up" in a structure, what makes a UA dinucleotide different from AU, what are the different kinds of triple-stranded associations). This is partly reflected in my thesis [3], pages 57-64, but otherwise unpublished. Having clear notions on these topics was however very useful in my subsequent bio-informatics work. Part 3 was supposed to deal with the metric properties of nucleic acids, for instance, in what environment a G.U pair could be present without introducing too much distortion?. Part 4 was supposed to deal with the influence of local sequence on the structure (e.g., what are the preferred sequences in stems or loops). In practice, the work on the energy models for predicting RNA secondary structures [8, 9] fulfils part of the goal of part 4.

When I started working on the kinetic theory of accuracy, I developed probabilistic methods to understand the outcome of the competitions between cognate and non-cognate substrates [10, 11] (see "probabilistic tools in enzyme kinetics" below). Later, I succeeded in calculating "absolute" rates in enzyme kinetics, from first passage times and pathway probabilities [12]. I consider [12] as my best contribution to mathematical biology.

I used my skills in analytical geometry to produce original stereoscopic images (see web chapter on stereoscopic vision) and my understanding of projective geometry to deal with the correspondence problem in stereoscopic vision (see web chapter on stereoscopic vision). Having good geometrical intuition allowed me to produce several interesting pieces of work in visual perception, but since most workers in this domain do not understand geometry, this work has had little impact. The level of geometric illiteracy in the field is, in fact, amazing. People are not even conscious of their limitations. It is just that geometry is not part of their intellectual world. This and other pathologies that are damaging to the field of visual perception will be described elsewhere.

Last, there are curious bits of mathematical insights which allowed me to solve simultaneously problems in quite different domains (see "a link between vision and molecular biology" below).

PROBABILISTIC TOOLS IN ENZYME KINETICS

PROBABILISTIC DERIVATIONS

When I started thinking seriously about the accuracy of molecular processes, the conceptual tools were not yet available. The enzymologists were describing the mechanics of enzyme activity using a steady-state treatment, in which they lined up equations saying that the concentrations of all the intermediates in an enzyme reaction did not vary with time. The set of simultaneous equations was solved and led to the expression of the concentration of the products of the reaction as ratios between two rather complicated expressions. A highly praised method was that of King and Altman [13], but it was just spelling out a type of calculation that was trivial for mathematicians. Nothing could be grasped intuitively in this approach.

So, I started thinking in terms of probabilities. I imagined a molecule of enzyme, its random encounters with molecules of substrate, the variable time during which the enzyme and the substrate remained bound, the random occurrence of an action that pushed the initial enzyme-substrate complex into another state, etc. All this was strictly equivalent, mathematically, to the steady-state picture, but I was not aware of that. In fact, chemical kinetics are intrinsically probabilistic. Macroscopic descriptions, such as those embodied in the steady-state treatments have their justification in the underlying probabilistic understanding of molecular processes.

Reasoning with probabilities (instead of applying the steady-state treatments) was very productive, and allowed me to derive all my important results on the accuracy of molecular processes [10, 11, 14-17]. The mathematics were simple and intuitive. One reason that made the mathematics simple is that, in accuracy problems, one deals with the competition between two or several substrates, and what matters is the ratio of the compounds of each kind which are formed. Using probabilities, most of the complexities are removed when taking the ratios. On the other hand, with the steady-state treatments, there is a whole set of additional equations for each competing substrate, and all equations must be solved simultaneously. Despite the elegant simplicity of the probabilistic approach, most people in enzyme kinetics remained glued to the steady-state treatments, to KM's and Kcat's and other horrors. And since they were not able, with their tools, to see the implications of the schemes they were studying, they used so-so semi-empirical arguments and often reached erroneous conclusions. The situation, over the last forty years has not improved. On the question of erroneous treatments, see the web chapters on the kinetic theory of accuracy, and on statistical illusionism.

The biologists with whom I discuss often consider probabilistic treatments as speculative, and do not try to invest time to understand them. They think it is a romantic view of enzyme mechanisms, yet it is chiefly a return to first principles. The mathematics of probabilities are subtle, not complex. One has to accept paradoxes such as the infinite sum " $p = 1/2 + 1/4 + 1/8 + 1/16 \dots = 1$ ". More strikingly, one has to accept that p can be computed, very easily, from " $p = 1/2 + p/2$ ". Do you see why?. All the power of the probabilistic treatment comes from the capacity to write such simple equations in which p is on both sides, because one explores a tree of events, and there

are, on some branches, a return to earlier conditions.

I gave most of the (probabilistic) recipes to calculate error-rates for various enzyme mechanisms in my review article in the book on accuracy by Kirkwood, Rosenberger and Galas [11]. Simple, general expressions are given there for several classes of mechanisms.

THE PEELBACK PROBLEM

The most advanced calculations were about the "peelback problem". Early in 1974, I had in my hands an issue of the *Journal of Molecular Biology*, containing an article by Goodman et al. [18], which contained a theoretical treatment of the interplay between the polymerization and excision functions of the DNA polymerases. At that time I was well-advanced in the mathematics of accuracy, and felt immediately that something was wrong with their treatment. They had completely overlooked the "next nucleotide effect" (which I knew, but had not yet published) according to which, when a polymerase has a proofreading function, the errors made at one position increase with the concentration of the nucleotide to be incorporated next (see web chapters on the kinetic theory of accuracy, and on DNA and RNA polymerase kinetics). Furthermore, they had derived their theoretical equation by combining several equations which were mere definitions that did not involve any physical insight.

I found rapidly that their equation was in fact a concealed identity — something having the same status as $(a+b)(a-b) = (a)(a) - (b)(b)$. So this "theoretical treatment" was a fantastic mathematical blunder. And the error had escaped the attention of four authors, two reviewers and an editor! Although the equation was an identity (therefore satisfied by any set of results, real or imaginary), the authors managed to plot experimental points that were not exactly on the curve representing their equation, some of the points being above, and others below the curve. They achieved this trick by plotting, not the individual results, but averages of the results. An identity which is verified by any arbitrary set of values, need not be verified by separate averages on the values. This article was, therefore, a piece of anthology in bio-mathematical blunders. Yet, the authors never retracted it.

I solved the problem they had addressed - deriving the error-rate for a DNA polymerase having a proofreading activity, taking into account all the forward and backward steps that the polymerase could make on the template. The calculations were performed under the simplifying condition that the incorporation and excision kinetic constants at one position were independent of the nature of the nucleotide at the preceding position. The resulting equation was given, without proof, in the 1975 *Biochimie* paper [10], and with the derivations in [15]. In 1975 I had a correspondence with David Galas on this topic (see web chapter on the kinetic theory of accuracy). He and Elbert Branscomb had detected the mistake in [18] and they ultimately published their (correct) equation in [19]. Later, I made a simpler treatment and extended the work to study the effect of pyrophosphorolysis [20]. The peelback problem was solved under increasingly general conditions by Malygin and Yashina [21] then by Jean Durup [22]. These treatments have remained mathematical curiosities. Recently, Cady and Qian proposed a "new stochastic thermodynamic analysis of DNA replication" [23]. I have not been able to determine whether their results agree with the previous results or not.

THE RUSSIAN SCHOOL

In my search for the most elegant expressions, I was inspired by work from the Russian school of chemical kinetics. I had access to this work, owing to a visit Ernst Malygin made in Chapeville's laboratory in Paris. I saw that the Russians had developed a powerful tool to write down at once the equations corresponding to a kinetic scheme. So they published papers in Russian with ten different equations for ten different schemes, each equation being given without comment, as though it was obvious. At the same time, authors from the Western block were publishing papers in *The Biochemical Journal* or *The Journal of Theoretical Biology*, with pages of derivations, merely to deal with a single scheme. Knowing the elegant forms of the expressions helped me to derive my own expressions in [11]. On the other hand, I did not really understand the Russian method. It was introduced by M. I. Temkin in [24]. In this paper, the reasonings appear to be elementary ones, and it seems that there is mainly a great talent for rearranging algebraic expressions, making them beautifully simple. In 2004, I was discussing with my colleague Bernard Derrida of the Statistical Physics Laboratory at Ecole Normale Supérieure (he was awarded the Boltzman medal in 2010). Having shown to him a "Russian" formula (Equation 2-59 in "Non-formal enzyme kinetics" by Schmid and Sapunov [25]) Derrida saw at once how it could have been obtained, and derived the formula on the spot, before my eyes. The trick was to use a Laplace transform of the steady-state equations. I do not know whether or not Derrida's shortcut had been used by the Russian school. The answer, perhaps, is in the book "Chemical kinetics" (which I have not seen) by Emanuel' and Knorre [26].

FIRST PASSAGE TIMES IN ENZYMOLOGY

Having developed the probabilistic tool for calculating error-rates (which are ratios of rates) I attempted to see, to make the story complete, whether or not this type of thinking could be extended to have a direct access to "absolute" rates of reactions. Ultimately, I succeeded, and managed to derive reaction rates for enzyme mechanisms, using first passage times and pathway probabilities [12]. Among other advantages, this treatment allows one to make general statements about classes of kinetic mechanisms. While the steady-state treatment is "unstable" (one needs to compute everything from scratch when one changes a detail in a reaction mechanism) my method allows one to leave many details, including topological ones, unspecified. Furthermore, my treatment allows one to incorporate, in a reaction scheme, diffusive steps that cannot be simulated with any small-sized classical sub-scheme. I showed that enzymes for which there was a single state allowing the entry of the substrate followed Michaelis kinetics whatever the intricacies of the subsequent steps.

Here is the abstract of this paper :

An alternative method for deriving rate equations in enzyme kinetics is presented. An enzyme is followed as it moves along the various pathways allowed by the reaction scheme. The times spent in various sections of the scheme and the pathway probabilities are computed, using simple rules. The rate equation obtains as a function of times and probabilities. The results are equivalent to those provided by the steady-state formalism. While the latter applies uniformly to all schemes, the formalism

presented here requires adaptation to each additional class of schemes. However, it has the merit of allowing one to leave unspecified many details of the scheme, including topological ones. Furthermore, it allows one to decompose a scheme into subschemes, analyze the parts separately, and use the intermediate results to derive the rate equation of the complete scheme. The method is applied here to derive general equations for one- and two-entry sites enzymes.

This paper should have been taken as a very significant progress in enzymology. Alexey Mazur discussed the foundations of this method [27], and Janos Südi used it in a work on control coefficients in enzymatic networks [28], but on the whole, it has not yet set (formal) enzymology in motion. The paper was channelled to PNAS by John Hopfield, and the reviewers made valid constructive suggestions. In particular, one reviewer drew my attention to an earlier paper by José Manuel Olavarriá [29] on a stochastic model of the enzyme-substrate reaction, in which he had derived the correct probabilistic equation for the Michaelis-Menten scheme. However, his article was more concerned with computer simulations than with analytical expressions. The reviewer who was so well informed about work from Latin America must have been Athel Cornish-Bowden, whose Chilean wife, Maria Luz Cardenas co-signed papers with J.-M. Olavarriá.

While the article was under review, I sent a copy to Leslie Orgel, and he sent me immediately a draft of a paper he had in his drawers, expounding the same philosophy (how a rate could be computed from times and probabilities) and applying it to a famous issue [30] in enzyme mechanism optimization. The title of the manuscript was "Albery and Knowles made easy". However, it was not as advanced, mathematically, as mine, and Orgel declined the invitation to co-sign my article.

In 2005, Ernst Malygin wrote to me, and we initiated a substantial correspondence on how to make enzymological calculations easier. This was an occasion for me to attack rather complex schemes with loops. Our results agreed, and in some cases, Malygin's derivations were very elegant. However, his approach had two limitations compared to mine : (i) the schemes needed to satisfy the constraints of detailed balance [31] thus excluding cases with energetic couplings – such as, e.g., kinetic proofreading mechanisms (ii) the derivations could not be made on classes of reactions, for which some details were not specified. Malygin applied his method to experimental results on a bacteriophage T4 DNA methyltransferase [32], and he expounded his calculation methods in [33].

It is perhaps during the period of training with Malygin on complex schemes that I noticed that my treatment in the PNAS 1987 article [12] could be considerably simplified. On several occasions, I had replaced in this article infinite series such as $p = 1/2 + 1/4 + 1/8 + \dots$ by the compact expression $p = 1/2 + p/2$. Actually, the same simplification applies to time series. Perhaps writing equations in which a time on the left side equals something plus a function of the time on the right side was too iconoclastic for me. Anyway, it works, and makes the calculations simpler. Whenever there is an equation with the p's, there is in parallel a very similar equation with the times. I used this simplification in [34].

Finally, it is a pleasure to note that with the current developments in single molecule enzymology, my calculation method in which I follow a single molecule of enzyme over successive cycles of reactions is becoming quite natural, conceptually. On the

other hand, numerical simulations with interactive graphics are becoming very easy, and analytical results are less needed.

RNA TOPOLOGY

My article on RNA topology [7] was part of an ambitious program, in Bourbaki's axiomatic style, to describe the space of possible nucleic acid structures. It did not have any impact on the field, and I do not take it as major work. It could be fun, for people who would like to write about the early papers on nucleic acid topology, to look at this one. One reason, perhaps, for the low impact of this article, was that very soon much work was to be invested in the topology of DNA, not RNA molecules, and in particular, in the problems raised by the action of DNA topoisomerases. It is only much more recently, with the determinations of the structures of complex RNA molecules, that this article might take value. One thing I discuss, in this article, is what I called "the principle of contiguity" which would, in current, looser terminology, be restated as a principle of prevalence of "unknotted" secondary structures. I pointed out that structures which obey this principle do not need to be formed sequentially, while the structures which do not obey the principle may fold according to different 3d configurations, depending upon the sequence of formation of the secondary structure segments. This is a valid point, and I do not know whether or not it has been brought to the surface again, now that it is really useful.

I give now the abstract of the article:

"Our general purpose in undertaking this study was to develop a conceptual tool which might be convenient to understand the structures of nucleic acids presenting much lower regularity than DNA fibres, and might be of some help for the conception and the construction of molecular models.

"The matter dealt with is that of topological properties, that is, the properties for which the nucleic acid is considered as a thread, regardless of the sequence of the bases, their orientation or the configuration of the sugarphosphate backbone. The thread can be closed or open and the associations of the various regions of the molecule can result in various patterns. After study of these patterns, the two concepts of parallelism and contiguity in the association are defined with precision, and the problems entailed by circularity are briefly reviewed. Under the chapter of knots, some topological singularities peculiar to tRNA models or other nucleic acids are discussed. The distinction is made between true topological knots, and slip knots. The physical presence of such situations is discussed in relation to the properties of contiguity in the associations of the molecule."

A LINK BETWEEN VISION AND MOLECULAR BIOLOGY

In 1975-76, I was deeply thinking about the geometry underlying visual illusions, and saw that much of the phenomenology could be expressed in terms of relationships between the lengths of two segments a and b, of the form:

perceived length of $(a+b) >$ perceived length of $(a) +$ perceived length of (b) ,

or, when $b > a$,

perceived length of (b) / perceived length of (a) > length of (b) / length of (a)

or the reverse. Such relationships are typical of convex functions that take the value zero at the origin.

I was simultaneously working on kinetic amplification mechanisms, and it is apparent that the second condition expresses some amplification of the b/a ratio. Understanding that "concavity" was the most general mathematical formulation of "amplification" allowed me to produce a subtle proof of (quasi) impossibility of amplification mechanisms that would operate on a forward step in a reaction scheme. For a more precise statement, see [8].

It was fun to put the same figure on convex functions in my Journal of Theoretical Biology article on illusions ([35], Figure 7, see also [36], Figure 5a) and in the Biochimie 1977 article on kinetic amplification on forward steps ([8], Fig. 1). In the Biochimie 1977 article, I wrote cautiously that in order to have amplification on a forward rate "one would need a mechanism in which the substrate has to go more than once through the same step, and it would go more often through that step when it is the correct one". Actually, a scheme fulfilling these constraints was proposed in 1999 for DNA topoisomerases by Yan, Magnasco and Marko [37]. Going "more than once through the same step" implied that the enzyme is doing something, then undoing it, then doing it again.

APPENDIX : ON MODELS IN TIMES OF CONSENSUS

What follows is a summary of a talk at a meeting on "Biologie théorique, modélisation et enseignement de la biologie", organized by Anne-Marie Leseney, Université Paris VI, November 10th, 1998. This summary is preceded by its English translation.

ON MODELS IN TIMES OF CONSENSUS. Often, scientific progress is brought by the discovery of an anomaly: a very slight deviation in a phenomenon that is expected to be fully mastered. The difficulty, for the theoretician, is to recognise the essential anomalies, and distinguish them from deviations due to secondary effects. More to the point: a good theoretician is capable to perceive, behind the structure of classical results that are well explained by classical models, the possibility of alternative models that would equally well account for known results, but would also predict undetected anomalies.

Too often, modelling is considered as an outfit to give lustre to a body of results that everyone accepts. I will provide a few examples, taken from my own experience in biology to show that modelling can also be used to reveal phenomena that are well hidden behind the appearances.

Science administrators encourage work around a common theme by laboratories belonging to remote fields. This bureaucratically programmed interdisciplinarity produces perverse effects. An artificial consensus is created by researchers from different disciplines, around an oversimplified body of knowledge that impoverishes the phenomena, and by its normative effects, eliminates the anomalies that would have been the source of future progress.

LA MODELISATION A L'HEURE DE LA PENSEE UNIQUE. Souvent, le progrès scientifique est amené par la découverte d'une anomalie: une très légère déviation dans un phénomène qu'on croyait parfaitement maîtrisé. La difficulté, pour le théoricien, est de savoir reconnaître ces anomalies essentielles, et les distinguer des écarts dûs à des effets secondaires. Mieux encore, un bon théoricien est capable de percevoir, derrière la structure des résultats classiques, bien expliqués par des modèles classiques, la possibilité de modèles alternatifs rendant également compte des résultats connus, mais prédisant aussi des anomalies non encore recensées.

Trop souvent, la modélisation est conçue comme une parure construite pour valoriser après coup un corps de résultats admis par tous. Je donnerai quelques exemples, tirés de ma propre expérience en biologie, pour montrer qu'elle peut être également utilisée pour dévoiler des phénomènes bien dissimulés derrière les apparences.

Les administrateurs de la science encouragent les relations de travail autour d'un sujet commun, entre laboratoires de disciplines éloignées. Cette interdisciplinarité, bureaucratiquement programmée, produit des effets pervers: un consensus factice se crée entre chercheurs de différentes disciplines, autour d'un corps de savoir minimal qui appauvrit les phénomènes et, par ses effets normalisateurs, fait disparaître les anomalies sources des progrès futurs.

REFERENCES

- [1] Eigen, M. (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58, 465-523. One of the earliest vacuous papers in the exploding field of « virtual reality » science.
- [2] Crick, F.H.C. (1966) Codon-anticodon pairing : The wobble hypothesis. *Journal of Molecular Biology* 19, 548-555. A superb theoretical paper. Crick's proposals were taken for granted. Textbooks described codon-anticodon recognition as though the wobble hypothesis was a proven experimental fact.
- [3] Ninio, J. (1971) Etude de la structure de l'ARN de transfert par diffusion centrale des rayons X, et de ses implications biologiques. Thèse d'Etat, Université Paris 7.
- [4] Ninio, J., Luzzati, V. and Yaniv, M. (1972) Comparative small-angle X-ray scattering studies on unacylated, acylated and cross-linked *Escherichia coli* transfer RNA Val. 1. *Journal of Molecular Biology* 71, 217-229. The article in which we introduced the « cube method » for calculating the theoretical X-ray scattering by a macromolecule in solution.
- [5] Pavlov, M.Y., and Fedorov, B.A. (1983) Improved technique for calculating X-ray scattering intensity of biopolymers in solution: evaluation of the form, volume, and surface of a particle. *Biopolymers* 22, 1507-1522.
- [6] Koch, M. H. J., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering : a view on the properties, structures and structural changes of biological macromolecules in solution. *Quarterly reviews of biophysics* 36, 147-227.
- [7] Ninio, J. (1971) Properties of nucleic acid representations. 1. Topology. *Biochimie*

53, 485-494.

[8] Ninio, J. (1979) Prediction of pairing schemes in RNA molecules. Loop contributions and energy of wobble and non-wobble pairs. *Biochimie* 61, 1133-1150.

[9] Papanicolaou, C., Gouy, M. and Ninio, J. (1984) An energy model that predicts the correct folding of both the tRNA and the 5S RNA molecules. *Nucleic Acids Res.* 12, 31-44.

[10] Ninio, J. (1975) Kinetic amplification of enzyme discrimination. *Biochimie* 57, 587-595.

[11] Ninio, J. (1986) Kinetic and probabilistic thinking in accuracy. In *Accuracy in Molecular Processes* (Kirkwood, T.B.L., Rosenberger, R. and Galas, D.J., eds) Chapman & Hall, London, pp. 291-328.

[12] Ninio, J. (1987) Alternative to the steady-state method : Derivation of reaction rates from first passage times and pathway probabilities. *Proc. Nat. Acad. Sci. USA* 84, 663-667. This is my best contribution to bio-mathematics.

[13] King, E. L., and Altman, C. (1956) A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *The Journal of Physical Chemistry* 60, 1375-1378. This is a famous article, an enzymologist's bible, quoted over 800 times. However, it merely explains how to solve a set of linear equations, using the ratio of two determinants. The proposed 'schematic method' spells out, for people ignorant in mathematics, how to select the non-zero terms in the determinants.

[14] Ninio, J. (1974) A semi-quantitative treatment of missense and nonsense suppression in the strA and ram ribosomal mutants of *Escherichia coli*. Evaluation of some molecular parameters of translation *in vivo*. *Journal of Molecular Biology* 84, 297-313. This article introduces kinetic thinking in a molecular recognition problem. It explains why there must be an efficiency/accuracy trade-off.

[15] Ninio, J. (1977) Are further kinetic amplification schemes possible? *Biochimie* 59, 759-760. A mathematically subtle, almost miraculous proof of impossibility.

[16] Bernardi, F., Saghi, M., Dorizzi, M. and Ninio, J. (1979) A new approach to DNA polymerase kinetics. *Journal of Molecular Biology* 129, 93-112.

[17] Ninio, J. (1986) Fine tuning of ribosomal accuracy. *FEBS Letters* 196, 1-4.

[18] Goodman, M.F., Gore, W.C., Muzyczka, N. and Bessman, M.J. (1974) Studies on the biochemical basis of spontaneous mutation. III. Rate model for DNA-polymerase-effected nucleotide misincorporation. *Journal of Molecular Biology* 88, 423-435. This paper contains one of the most fantastic mathematical blunders of all times. It was never retracted.

[19] Galas, D.J. and Branscomb, E.W. (1978) Enzymatic determinants of DNA polymerase accuracy. *Journal of Molecular Biology* 124, 653-687.

[20] Herbomel, P. and Ninio, J. (1980) Fidélité d'une réaction de polymérisation selon la proximité de l'équilibre. *Comptes-Rendus Acad. Sci. Paris, Série D*, 291, 881-884.

[21] Malygin, E.G. and Yashina, L.N. (1980) Kinetic description of the error-correcting mechanism of bifunctional DNA polymerases (in Russian). *Doklad. Akad. Nauk. SSSR* 250, 246-250.

- [22] Durup, J. (1982) On the relation between error rates in DNA replication and elementary chemical rate constants. *Journal of Theoretical Biology* 94, 607-632.
- [23] Cady, F., and Qian, H. (2009) Open-system thermodynamic analysis of DNA polymerase fidelity. *Physical Biology* 6 036011 doi : 10.1088/1478-3975/6/3/036011. The equations proposed in this article mix thermodynamic with kinetic parameters, in a way that, for me, make the results opaque. I was not able to determine whether or not the equations were mathematically equivalent to those proposed in earlier work [16, 19-22].
- [24] Temkin, K. I. (1963) The kinetics of stationary reactions. Original article in Russian : *Doklad Akad. Nauk. Physical Chemistry section* 152, 156-159. English cover to cover translation, 152 (?) 782-785.
- [25] Schmid, R., and Sapunov, V.N. (1982) Non-formal kinetics. In search for chemical reaction pathways. Verlag Chemie, Weinheim. Volume 14 in the series "Monographs in modern chemistry".
- [26] Emanuel', N.M., and Knorre, D.G. (1973) Chemical kinetics. Homogeneous reactions. J. Wiley & Sons, New-York.
- [27] Mazur, A.K. (1990) A probabilistic view on steady-state enzyme reactions. *Journal of Theoretical Biology* 148, 229-242.
- [28] Südi, J. (1997) Control analysis of enzyme mechanisms in terms of the classical steady-state description. *Biochimica et Biophysica Acta* 1341, 108-136.
- [29] Olavarriá, J. M. (1973) Simulacion directa de sistemas enzimáticos, utilizando un modelo estocástico de reacción enzima-sustrato. *Annales Acad. Sci. Exact ; Fis. Nat.* 25, 129-165.
- [30] Albery, W.J., and Knowles, J.R. (1976) Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* 15, 5631-5640.
- [31] Wegscheider, R. (1901) Über simultane Gleichgewichte und die Beziehungen zwischen Thermodynamik und Reaktionskinetik homogener Systeme. *Monatshefte für Chemie* 22, 849-906.
- [32] Evdokimov, A. A., Zinoviev, V., Malygin, E. G., Schlagman, S. L., and Hattman, S. (2002) Bacteriophage T4 Dam DNA-[N6-adenine] methyltransferase. Kinetic evidence for a catalytically essential conformational change in the ternary complex. *The Journal of Biological Chemistry* 277, 279-286.
- [33] Malygin, E. G. and Hattman, S. (2006). A probabilistic approach to compact steady-state kinetic equations for enzymatic reactions. *Journal of Theoretical Biology* 242, 627-633.
- [34] Ninio, J. (2006) Multiple stages in codon-anticodon recognition : double-trigger mechanisms and geometric constraints. *Biochimie* 88, 963-992. In the Appendix B, I used my method [12] to analyse the consequences of the Gromadski and Rodnina scheme for ribosome kinetics, and I wrote parallel expressions for the times and the probabilities (page 987). I thus obtained in a few lines, it seems, results that no other person in the field could have derived.
- [35] Ninio, J. (1979) An algorithm that generates a large number of geometric visual

illusions. *Journal of Theoretical Biology* 79, 167-201. See also Fig. 5a in [36].

[36] Ninio, J. (2014) Geometrical illusions are not always where you think they are: a review of some classical and less classical illusions, and ways to describe them. *Frontiers in Human Neurosciences* 8, article 856, doi: 10.3389/fnhum.2014.00856

[37] Yan, J., Magnasco, M. O., and Marko, J. F. (1999) A kinetic proofreading mechanisms for disentanglement of DNA by topoisomerases. *Nature* 401, 932-935.