



ELSEVIER

Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

SCIENCE @ DIRECT®

C. R. Biologies 326 (2003) 1105–1110



Molecular biology and genetics

# The macro-ethics of genomics to health: the Physiome Project

James B. Bassingthwaight

Department of Bioengineering, Box 35-7962, University of Washington, Seattle, WA 98195-7962, USA

Received 16 September 2003; accepted 23 September 2003

Presented by François Gros

## Abstract

Advances in pharmacology and genomics, and their intervention in human biology are beyond our abilities to understand their consequences. Therapeutic intervention in highly complex, non-linear, adaptive biological systems results in some unforeseen and undesirable consequences. To do the most good with the least harm, the information on biological systems should be gathered into databases, and into comprehensive quantitative models that can help to predict the long-range effects of proposed interventions. This is a societal or professional macro-ethical imperative. The Physiome Project helps to meet this imperative via databasing and creating models and tools for large-scale integration. **To cite this article:** *J.B. Bassingthwaight, C. R. Biologies 326 (2003).*

© 2003 Académie des sciences. Published by Elsevier SAS. All rights reserved.

## Résumé

**La macro-éthique, de la génomique à la santé : le projet Physiome.** Les avancées en pharmacologie et en génomique ainsi que leur intervention en biologie humaine vont au-delà de notre compréhension de leurs conséquences. L'intervention thérapeutique sur des systèmes biologiques hautement complexes, non linéaires et adaptables, provoque des conséquences imprévisibles et indésirables. De manière à faire le plus de bien possible et le moindre mal, l'information sur les systèmes biologiques doit être regroupée dans des modèles quantitatifs qui prédisent les effets à longue portée des interventions proposées. Ceci est un impératif macro-éthique sociétal ou professionnel, et le projet Physiome fournit un moyen d'y faire face en créant des outils d'intégration à grande échelle. **Pour citer cet article :** *J.B. Bassingthwaight, C. R. Biologies 326 (2003).*

© 2003 Académie des sciences. Published by Elsevier SAS. All rights reserved.

**Keywords:** biological databases; computational systems biology; control of gene expression; ethics and society; metabolic engineering; pharmaceuticals; physiome

**Mots-clés :** bases de données biologiques ; biologie systémique computationnelle ; contrôle de l'expression génique ; éthique et société ; ingénierie métabolique ; médicaments ; physiome

## 1. Introduction

In giving an overview of the state of engineering as we enter the new millennium, William Wulf [1] in-

troduced the concept of 'macro-ethical' behavior, behavior for which the incentive is to create intellectual pressure within our society to do 'the right thing' for the long-term betterment of the society. Examples abound: planning with respect to developing maintainable energy resources, preserving the environment, avoiding ecological disasters, educating *all* of our peo-

*E-mail address:* [jbb@bioeng.washington.edu](mailto:jbb@bioeng.washington.edu)  
(J.B. Bassingthwaight).

ple well. The macro-ethical issue we address here concerns practical means for advancing health care, while minimizing risk and maximizing benefits. The thesis is that it is our duty to ‘think as hard as we can’ to plan for the future.

The current approaches to developing new therapies are not sufficient to prevent unlikely calamitous results. There will be no sure way. What is needed is a great leap forward to allow ‘knowledgeable’ calculation of risk and benefit. We need information on which to make the wisest possible decisions. Gene insertion, stem cell infusion, and new pharmaceuticals are not within the purview of scientific funding agencies, but of regulatory agencies such as the Food and Drug Administration in the USA. They operate to protect us against advances that are too speculative or too risky. They are heavily dependent on large and expensive, clinical trials, which put human subjects at risk in the interest of protecting others. Moreover, most novel interventions are right at the edge: the possibilities are great, but the evidence is sparse and the risk of failure, and harm or damage, is high. When faced with such issues, Wulf’s admonition is to think hard, as mightily as one can, to move ahead as one must, but to minimize the likelihood of harm. My interpretation of this is that we must arm our intuition and insight to maximize our power to predict.

## 2. Complexity, informatics and information flow

A problem in medicine and biology is that much relevant information is undiscovered, unreliable or difficult to retrieve. A completed human genome does not suffice to define human function, although it is a guide to a list of possible ingredients. The genetically derived ingredients, the proteins, are much more numerous than genes; in yeast there are about three proteins per gene and in humans more like ten proteins per gene. The concentrations of the proteins, the measure of their expression in the various cell types, is unpredictable from the genome, being governed also by environment, behavior and by the dynamic relationships amongst proteins, substrates, ionic composition, energy balance, and so on. The large variance in the ratios of protein concentrations to their mRNA levels attests to the insecurity in predicting individ-

ual protein levels from the message. Proteins are synthesized and broken down continuously. Pretranslational selection from different parts of the DNA sequence, post-translational slicing out of parts of the protein, splicing of two or more proteins together, and combining groups of proteins together into functional assembly line-like complexes, all contribute to increasing the variety of the products of gene expression.

Knowledge of the proteins, their locations and concentrations in various cell types under various conditions and the kinetics of the reactions in which each is involved, would comprise a magnificent database. The organizers of the enzyme handbooks and enzyme databases (for example, WIT of Selkov et al., at <http://www.anl.gov>), of the protein data banks such as PDB (<http://pdb.sdsc.edu>) and Swissprot (<http://www.expasy.ch>) have made giant strides toward such goals. They provide sequence and much structure, but little function. Selkov’s database is oriented toward providing kinetic or functional information, rather than structural, and it covers a wide variety of species, being more complete on single cell species than on more advanced species.

## 3. The combinatorial dilemma

The sorting out of the genome will leave us with a moderate number of genes to incorporate into our thinking. The estimates of the number have come down from the range of 60 to 100 000 to half of those. It is now apparent that the level of complexity in mammalian protein expression far exceeds that of *C. elegans* with its 19 536 genes and 952 cells: we might have only double or triple the number of genes, but have a much higher ratio of proteins per gene. If there are 10 proteins per gene resulting from pre-transcriptional divergence and from post-translational modifications, then we have on the order of half a million proteins, in widely varied abundance. If each interacts with five others (for example, exchanging substrates with neighbors in a pathway or modifying the kinetics of others in a signaling sequence), then each protein may be linked to all others in the cell through rather few steps, a kind of ‘6 degrees of separation’ from any other protein. If there were only two states per protein the numbers of possible

cellular states would be huge. Ours is not, however, a binary system of bits, but a composite of a stochastic and continuous systems in perpetual change, so that the numbers of possibilities become countless. Moreover, cells contain not just proteins, but substrates and metabolites, and are influenced by their environments. The complexity provides a basis for functionality not predictable from the function of any of the components, but ‘emergent’ behavior coming from interactions among larger integrated units, subcellular systems, or aggregates of cells, tissues, organs, and systems within an organism. Physiological systems are highly non-linear, higher-order, and provide a milieu in which chaotic dynamics are observed [2, 3]. Chaotic dynamics implies short-term predictability, long-term unpredictability, but a confined operating range. By the stability of the *milieu interieur*, Bernard [4,5] meant a mildly fluctuating state rather than a stagnant ‘homeostasis’. Real biological systems are ‘homeodynamic’, a word implying ‘fluctuation under control’.

What the idea of complexity implies is that even if we knew all proteins and all the rate constants for their reactions we would still have trouble predicting the outcome of an intervention that interfered with any one of these proteins. Nevertheless, remembering that even chaotic systems have behavior that is exactly predictable for some short time, it is clearly possible to predict *something* about the outcomes of a proposed therapy. Since proteins are building blocks or nodes on pathways, and each reacts with substrates or other proteins of a limited variety, there are road maps of reactions. The charts of biochemical reactions are a good start, and give the stoichiometry of the reactions, though information on the thermodynamics is absent. The charts have limitations: some reactions and some proteins are missing, and there is no indication of the fluxes along the pathways. There is little information on the substances controlling the activities of enzymes or transporters, or about the spatial arrangements of the proteins in pathways. We need information on the regulatory pathways for gene expression so we can begin to discover how adaptations occur over long times. ‘Thinking hard’, quantitatively, is difficult and not presently likely to yield accurate prediction of the effects of intervention at the single protein level.

#### 4. Bioinformatics includes the databasing of physiological and anatomic data

Databases, well maintained and easily accessible, are greatly needed in order to develop behaviorally realistic models of systems. There are cogent arguments for building large databases to capture biological data [6,7]. US federal support has gone into genomic and the proteomic databases, but not into storing higher-level physiological information. The US National Library of Medicine’s effort to preserve anatomic information (The Visible Human project: <http://www.nlm.nih.gov>) is a part of the morphome (which we define as anatomic and morphometric), just as are the genomic and proteomic databases. They concern structure, an essential prerequisite to function, but are not enough by themselves.

Physiology and pathophysiology concern processes, the dynamics, kinetics, and the functioning of those structures. Statistical descriptions and correlations of physiological variables is not enough. The need is for explanation and mechanism, and for understanding regulation and control. The potential for models is that they can, at their best, account for mechanisms, and thereby define cause and effect. At their lesser levels, they can be used to summarize statistics on a system.

Models can be used to capture information and integrate it into a comprehensive, self-consistent conceptual framework, conveying understanding. Models come in varied forms, sketches of concepts, diagrams of relationships or schemas of interactions, mathematical models defined by sets of equations, and computational models (from analytical mathematical solutions or from numerical solutions to differential or algebraic equations) which can be explored by a user to develop insight into how the model and the real system (if the model is reasonably correct) behave. The Physiome Project, described below, combines these two goals, databasing and integrative modeling. Without data there is nothing to model; without models, there is no source of deep predictive understanding and no guide as to what next sets of information are needed. Without models our intuition is not fully armed.

## 5. Therapy is risky, but therapeutic risks will be taken

One cannot predict with certainty, but one has to go ahead anyway when failure to do something leads only to the death of a patient. The designer of pharmaceuticals to alleviate multiple sclerosis, the developer of stem cells modified to cure diabetes, the developer of materials for prolonged, controlled release of drugs, all have immense pressures to move forward into the unknown, even knowing there is risk. More information would help, if it were integrated and presentable in forms that advance understanding and allow one to examine conjectures, to answer our ‘What if?’. Ideally, computer models should run “at the speed of thought”, responding faster than we can think of the next question. Highly efficient computation lends itself to broad exploration of a system that has been well modeled and thereby allows us to evaluate contemplated plans and also to critique the behavior of the models.

## 6. The macro-ethical imperative

In order to do the right thing, to move ahead with the best assurance, the most insight, with the bets best hedged, one must do one’s utmost to predict accurately. The databasing, the development, archiving, and dissemination of simple and complex systems models, the exploration and critiquing, and the evaluation (and, necessarily, the rejection or improvement) of models become part of this moral imperative to think with the greatest possible depth into the problems accompanying and created by intervening into biological systems, either that of the individual or that of an ecosystem.

The information needs to be in the public domain, and maintained in forms that can be accessed over the web. Our government agencies need to address the issue of providing funding for databases that are larger than any previously developed for biology, databases which will probably challenge the skills of even the experts who have done this in atmospheric sciences and astronomy. The low reliability of physiological and pharmacological information and the impediments to its maintenance, the control of its quality, and the complexities of providing an assessment of

its reproducibility, these are issues that are more difficult than sequencing and sorting the genome. These too need to be addressed as a part of the macro-ethical imperative.

## 7. The Physiome Project

Undertaking the Physiome Project is a response to the macro-ethical imperative to minimize risk while advancing medical science and therapy [8,9] and <http://www.physiome.org/files/Petrodvoret.1997/abstracts/>. Other responses to be taken in parallel include animal experimentation, epidemiological data acquisition and analysis in human disease. The ‘physiome’ is defined as the quantitative description of the functional state of the organism. Like the genome, it is an object to be defined for each species and each individual within the species. Quantitative representation of the composite and integrated system behavior of the living organism is accomplished through the development of a hierarchical set of mathematical models representing the behavior of the system. The models are linked to databases of information from a multitude of experiments and observations: an explicitly quantitative model is the means of bringing compromise and self-consistency into the understanding of the system.

The *Physiome Project* is the effort to define the physiome of individual species from bacteria to man, through databasing, systematic organization of the information and integrative modeling to make predictions from the information. At this point the Project is developing spontaneously as a multinational collaborative effort. It began through collaborations established among groups of scientists in a small number of fields, and encompasses only a small fraction of physiological fields [10]. In each field, the investigators have begun with finite, achievable goals, and have proceeded to put the pieces together into impressive edifices of thoughtful integration. A website, <http://www.physiome.org>, serves to provide some history and is a guide to other sites. See, for example, the Cardiome efforts of McCulloch (<http://cardiome.ucsd.edu>) and of Hunter et al. (<http://www.bioeng.auckland.nz/physiome/physiome.php>), the cardiac electrophysiology efforts of Winslow et al. (<http://perspolis.bme.jhu.edu>), Rudy et al. (<http://www.cwru.edu/med/>

CBRTC/faculty.htm) and Noble et al. (<http://noble.physiol.ox.ac.uk>), the Microcirculation Physiome of Popel et al. (<http://www.bme.jhu.edu/news/microphys/>) and our own efforts in the processes of transport and exchange (<http://nsr.bioeng.washington.edu>). The models, via iteration with new experimentation, remove contradictions and demonstrate emergent properties. They are part of the tool kit for the ‘reverse engineering’ of biology.

Large-scale models of cell systems or overall regulatory systems such as endocrine systems are being built up of modular elements. A cell metabolism model, for example, is composed of transporters, channels, and enzymes, binding sites. These are organized to handle ionic currents, pathways fluxes, reaction sequences, and volume fluxes. They are constrained by balances of mass, volume, charge and membrane potential, redox potential, and limited by conservation laws for key elements and chemical constituents. The modularity is vital, for at the lowest levels it allows for the reusability of the code for specific components, and at higher levels for the maintenance of model code after it is proved out for numerical accuracy and as an adequate representation of the biology. The models serve as working hypotheses, temporary versions of what is believed to be the ‘correct’ biology until replaced. Each one invites dissenting opinion, the alternative hypothesis, and therefore, as proposed by Platt [11] the designing of the critical experiment that distinguishes between the alternatives: the experiment must disprove at least one of the hypotheses and so advance science.

Although the Physiome Project is well enough defined to begin to serve as a contributor to the macroethical imperative, at this point it is only an idealistic summary of good intentions. Its full realization will come as many small projects develop and become useful for those thinking about systems and analyzing data. The concepts about how to do biological modeling are changing, and will change more rapidly as appreciation that modeling is a requirement for biological research, and for the development of predictive capability, just as happened in physics and chemistry.

The hierarchical nature of biological systems is a guide to the development of hierarchies of models. While models at the molecular level can be based on biophysics, chemistry, energetics, and molecular

dynamics, it is obviously impractical to use molecular dynamics in describing the fluxes through sets of biochemical pathways, just as it is not practical to use the full set of biochemical reactions when describing force-velocity relationships in muscle, nor to use the details of myofilament crossbridge reactions when describing limb movement and athletic performance. By analogy, one does not use quarks to build trucks.

Quantitative considerations provide the keys to understanding systems. By observing model behavior, one gains much insight about the real system. To avoid the risk of being misled by an erroneous model, one must assess its closeness to reality. The databases of observations and experimental results are the basis from which to synthesize the schema and quantitative models that link the data from many experiments in a self-consistent manner. This is in the style of ‘thinking as hard as one can’, maximizing the integrated knowledge to gain as much foresight as one can muster. By such means we will be able to improve how we think about metabolic control, genetic regulatory networks, error correction and its energy cost in translation of mRNA to protein, signal transmission, cell recognition, etc., *and* we will begin to be able to predict, not very much at first, but gradually more.

Biological models can be defined at many hierarchical levels from gene to protein to cell, to organs and intact organisms. Practical models represent only one or two levels in the hierarchy. At a given hierarchical level, the strategy is not to model the details of events occurring in several underlying layers of the system, but to capture the essence of their behavior kinetically with good exactitude. This will give the proper dynamics at the chosen level for the circumstances specified. This leaves a problem: mono-hierarchical models have limited capability to adapt to changes in conditions. To handle transient conditions the modeler must make sure that the simplifications used in representations of events at the lower hierarchical levels are still correct: when changes are consequential, the modeler must adjust the computations at the lower level and incorporate them into the higher integrative level. (This is analogous to using adjustable time steps in systems of stiff equations.)

## 8. Application of physiome-style approaches to the design of therapy

Consider the futuristic state when well-developed models are available. The goal is to minimize risk while providing benefit. The models can then be explored for projected efficacy, for side effects, and for other long-range effects. This is why the modeling is imperative. To do this well, the models must extend from gene regulation to the functions of the organism.

Industrial and academic competitiveness will be enhanced for those who have the biological information and models most readily available. Like genomic information, biological information and models should be in the public domain. Policy needs to be made clear at federal and international levels so that parochial attitudes and pecuniary interests do not handicap open access to knowledge and scientific developments that depend on that knowledge.

Embarking on the Physiome Project has its own risks and benefits. While the expected benefit is better therapy and the more rapid development and testing of therapies, both pharmaceutical and genomic, the downside is the expense: immense databases, their maintenance, and the integrative modeling through multiple levels is costly. The costly investment is predicted to recompense society by a reduction in catastrophic error and the pharmaceutical industry by a reduction in the costs of bringing a drug to market. Mind you, really effective risk reduction is a long way off for it requires very deep and very broad integration; for example, the effects of thalidomide on embryonic development would only be prevented through models extending all the way from adult cell functioning to the regulation of expression.

No individual investigator or group can make build huge databases and construct large sets of models. If it is unethical to take large risks with human subjects, and if offsetting the risks requires national and international investment, then undertaking large scale biosys-

tems analyses such as make up the Physiome Project is a societal macro-ethical imperative.

## Acknowledgements

This work has been supported by NIH grant RR1243 from the National Center for Research Resources, through funds to the National Simulation Resource for Circulatory Mass Transport and Exchange (<http://nsr.bioeng.washington.edu>), which provides simulation systems XSIM and JSIM, and transport models, for public use.

## References

- [1] W.A. Wulf, Great achievements and grand challenges, *The Bridge* 30 (2000) 5–10.
- [2] J.B. Bassingthwaighte, L.S. Liebovitch, B.J. West, *Fractal Physiology*, Oxford University Press, New York, London, 1994.
- [3] A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms*, Cambridge University, 1996.
- [4] C. Bernard, *Introduction à l'étude de la médecine expérimentale*, J.-B. Baillière & fils, Paris, 1865.
- [5] C. Bernard, *An Introduction to the Study of Experimental Medicine*, Macmillan Company, New York, 1927.
- [6] N. Dao, P.J. McCormick, C.F. Dewey Jr., The human physiome as an information environment, *Ann. Biomed. Eng.* 28 (2000) 1032–1042.
- [7] H. Weissig, P.E. Bourne, An analysis of the protein data bank in search of temporal and global trends, *Bioinformatics* 15 (1999) 807–831.
- [8] J.B. Bassingthwaighte, R. Friesner, B. Honig, C.F. Starmer, V.Z. Marmarelis, Modeling and Simulation, in: NIH/NCRR Workshop on Technologies for the Future: Biomedical Computing for Visualization, Modeling and Decision Support, Bethesda MD, 12–13 April, 1991, pp. 1–15.
- [9] J.B. Bassingthwaighte, Toward modeling the human physiome, in: S. Sideman, R. Beyar (Eds.), *Molecular and Subcellular Cardiology, Effects on Structure and Function*, *Adv. Exp. Med. Biol.* 382 (1995) 331–339.
- [10] J.B. Bassingthwaighte, Strategies for the Physiome Project, *Ann. Biomed. Eng.* 28 (2000) 1043–1058.
- [11] J.R. Platt, Strong inference, *Science* 146 (1964) 347–353.