

Mathematical Modeling: Epidemiology Meets Systems Biology

Cornelia M. Ulrich,^{1,2} H. Frederik Nijhout,³ and Michael C. Reed⁴

¹Cancer Prevention Program, Fred Hutchinson Cancer Research Center; ²Department of Epidemiology, University of Washington, Seattle, Washington; and Departments of ³Biology and ⁴Mathematics, Duke University, Durham, North Carolina

"For every complex problem there is a simple, easy to understand, incorrect answer." –Albert Szent-Gyorgy

This issue of *Cancer Epidemiology, Biomarkers, and Prevention* includes a study on mathematical modeling of biological processes. Sokhansanj and Wilson describe a mathematical model that mimics the kinetics of base excision repair and thus permits them to investigate *in silico* the effects of genetic variation in this important DNA repair pathway (1).

Why should epidemiologists have interest in this type of mathematical modeling? Although perhaps not obvious at first glance, there are several benefits of this approach to cancer epidemiology. First, as we all know, the quality of cancer epidemiology is much improved by a thorough understanding of the underlying biological processes. Molecular epidemiology has increasingly moved away from considering isolated genes towards a pathway-based approach (2). Most grant applications in molecular cancer epidemiology these days delineate specific biological pathways that are targeted for investigation and provide schemata that describe the main reactions and connections. However, these pathway depictions are usually simplified: they often ignore (for parsimony) intricate intermediate steps or interactions with other pathways, and often, we have incomplete knowledge about the specific interplay of the many elements in the system. Biological systems are immensely complex. In our simplified descriptions of pathways, we tend to disregard the fact that changes in inputs typically affect many different cellular biochemical components (genes, substrates, and enzymes); thus, the net effect on the entire system is difficult to anticipate. There are many mutual regulatory interactions among genes, proteins, and the myriad organic molecules in a cell. Even more daunting is that many genes that code for enzymes are up-regulated or down-regulated by the cellular inputs or by changes in the levels of various biochemical substrates. The cell is a dynamic system that is always changing in response to the fluctuating cellular environment. In addition, of course, it is not just the metabolism and molecular biology of healthy cells and diseased cells that is of interest. We also need to understand how cellular changes lead to large-scale disease processes and to the biomarkers and disease outcomes that epidemiologists measure.

One way to study such complicated systems is by mathematical modeling. Typically, one develops differential equations for the gene activities or substrate concentrations in

small or medium-sized networks that contain the molecular phenomena that one wants to understand. The first difficult issue is to decide how to limit the network to be modeled. Each local network is connected to the whole of cell physiology by shared substrates, genes, and enzymes. By studying a small piece of the network and oversimplifying the rest as "constant," one may make simplifications that give non-physiologic results. Second, one must decide on the level of detail to be used for each individual interaction. Is it important to include the sequences of steps by which an enzyme catalyzes the reaction or will a simple Michaelis-Menten formula do? Is it necessary to model the time delays that occur in both transcription and translation? Can one treat the cell as a well-mixed biochemical reactor, or should one take into account compartmentalization and the localization of many processes? There are no simple or universal answers to such questions. They must be answered on a case-by-case basis depending on the biological or biochemical questions being investigated.

If one succeeds in constructing a mathematical model that reasonably represents the biochemical reality, the payoff is large. One can experiment with the model by increasing or decreasing inputs (corresponding, say, to changes in diet) or by raising or lowering activities of enzymes (corresponding to genetic polymorphisms), or eliminating entire reactions completely (corresponding to gene-knockout experiments). One can take apart and put back together the biochemical network piece by piece to determine how it works. In contrast to biological experiments, these *in silico* experiments are quick and inexpensive and, if done well, can give real insight into the genetic and molecular network.

How can epidemiologists benefit from these *in silico* experimentations? The article by Sokhansanj and Wilson (1) provides an excellent illustration of the use of a mathematical model to predict the functional effect of genetic polymorphisms on biologically relevant outputs. The authors explored base excision repair and used as main end points the number of unrepaired lesions in the cell and clearance time of lesions (focusing on 8-hydroxyguanine, a prominent component of oxidative DNA damage). Results from this model illustrate that this pathway is remarkably robust, as one would expect after millions of years of evolution, given the need to preserve the integrity of DNA in the face of the most common form of DNA damage (3, 4).

Of course, model predictions will be, although qualitatively correct, not necessarily quantitatively exact. Nevertheless, they do provide an initial understanding of where this system is robust to changes in enzyme function and where it is sensitive (see Table 3 in Sokhansanj and Wilson). Results from this type of sensitivity analysis allow epidemiologists to target their investigations of genetic variability towards components of the system in which an alteration in function would have consequences (i.e., result in a notable disturbance of the system). Based on their Table 3, one would suggest that Lig1 and perhaps Ogg1 could be such critical

Cancer Epidemiol Biomarkers Prev 2006;15(5):827–9

Received 3/28/06; accepted 3/29/06.

Grant support: This research was supported by NIH grant R01-CA105437 (PI: Ulrich).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Cornelia Ulrich, Cancer Prevention Program, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, M4-B402, Seattle, WA 98109-1024.

Phone: 206-667-7617; Fax: 206-667-7850. E-mail: nulrich@fhcrc.org.

Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-06-0252

elements of the base-excision repair network and thus should undergo more scrutiny in molecular epidemiologic investigations (1).

A second key advantage of mathematical modeling is that it provides a tool for quick and inexpensive (perhaps also quick and dirty) examinations of gene-gene interactions. Again, a mathematical model per se will never provide a complete answer and will certainly not replace the need for experimental or epidemiologic studies. However, results from modeling can guide an epidemiologist's choice of gene-gene interactions that are most promising for further investigation. Model predictions may provide information that can be incorporated in the statistical analysis in form of a "probability that an interaction will exist," comparable with the false-positive report probability (5). Sokhansanj and Wilson's findings illustrate that the investigation of interactions in a complex system can yield evidence for nonintuitive consequences of multiple simultaneous variants (1). Their findings suggest the next steps of targeted experimental investigations, including gene knockout studies.

Other mathematical models of biological networks may supply novel information on gene-environment interactions: for example, we are in the process of developing such a model of folate-mediated one-carbon metabolism, including its links to methylation capacity and nucleotide synthesis (6-8). A number of nutritional factors play a role in this biological pathway, both as substrates and cofactors, including folate; vitamins B₂, B₆, and B₁₂; methionine; and choline. This model will facilitate investigation of the independent and combined effects of dietary factors and genetic variability on the pathway's components and on specific biomarkers (e.g., thymidine synthesis). Although the model's results on gene-nutrient interactions will not substitute for *in vivo* studies, we expect that they will guide human feeding studies towards the combination of nutrients and genetic factors most relevant for disease processes.

A third benefit of mathematical modeling relates to our need for incorporating information on biological pathways in the statistical data analysis. D. Thomas has previously discussed some of the approaches that can be used in pursuing this goal (2). Results from mathematical modeling can yield information of direct use to epidemiologists, by giving values for prior covariates (factors that help explain variability between variables of interest) that can be used in a hierarchical modeling structure: for example, the modeling of folate metabolism is motivated by the strong evidence that folate intakes and polymorphisms in folate-metabolizing enzymes are related not only to neural tube defects but also several types of cancer (9-13), in the absence of well-defined biological mechanisms. Folate metabolism is integral for the provision of S-adenosylmethionine, a substrate for methylation reactions, including DNA methylation (14). At the same time, thymidine and purine syntheses require folate cofactors; thus, the integrity of DNA can be severely impaired under folate deficiency (15). Finally, homocysteine increases with a low folate status, and this compound may directly participate in disease processes as part of its role in redox signaling (16). There are now a number of hypotheses regarding the folate-related mechanisms that are relevant to carcinogenesis for specific polymorphisms or under specific conditions (17-21). An example of a prior covariate that could be derived from this mathematical modeling would be a quantitative value for the relative change in a specific biomarker that results from a polymorphic change in enzyme function (22, 23). Attaching such prior knowledge to specific polymorphisms in a

standard case-control analysis will help identify which specific biological mechanisms (e.g., effects on purine synthesis) are the primary links of folate to the disease outcome investigated.

In addition to the biological pathways of relevance to cancer biology discussed here, mathematical modeling has proven extremely useful in understanding complex molecular and biochemical mechanisms, such as the control of the cell cycle (24) and mitochondrial metabolism (25). Furthermore, the difficulties inherent in the analysis of complex, nonlinear gene-metabolic networks have stimulated the development of novel mathematical methodologies (26, 27).⁵

Mathematical modeling of cancer-relevant biological pathways provides epidemiologists with a new powerful tool. However, the strengths and limitations of this instrument need to be well understood to gain the most benefit from its use and avoid misinterpretation. A model will never be perfect, in part because it is usually based on incomplete and simplified biological information. Nevertheless, models can be used as platforms to test hypotheses that may be experimentally difficult or expensive. They can investigate the role that specific components (such as nutritional inputs or genetic factors) play in the overall behavior of the system. Such mathematical modeling will allow us to link epidemiology to systems biology and will help drive our understanding of cancer etiology and cancer biology.

Acknowledgments

We thank Dr. Duncan Thomas for his valuable input and our colleagues who are participating in the mathematical modeling of one-carbon metabolism project.

References

- Sokhansanj BA, Wilson DM. Estimating the impact of human base excision repair protein variants on the repair of oxidative DNA base damage. *Cancer Epidemiol Biomarkers Prev* 2006;15:1000-8.
- Thomas DC. The need for a systematic approach to complex pathways in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2005;14:557-9.
- Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. *Nature* 2001;411:366-74.
- Lu AI, Li X, Gu Y, Wright PM, Chang DY. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochemistry & Biophysics* 2001;35:141-70.
- Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434-42.
- Reed MC, Nijhout HF, Sparks R, Ulrich CM. A mathematical model of the methionine cycle. *J Theor Biol* 2004;226:33-43.
- Nijhout HF, Reed MC, Budu P, Ulrich CM. A mathematical model of the folate cycle: new insights into folate homeostasis. *J Biol Chem* 2004;279:55008-16.
- Nijhout HF, Reed M, Anderson D, Mattingly J, James SJ, Ulrich CM. Long-range allosteric interactions between the folate and methionine cycles stabilize DNA methylation rate. *Epigenetics* 2006;1:81-7.
- Little J, Sharp L, Duthie S, Narayanan S. Colon cancer and genetic variation in folate metabolism: the clinical bottom line. *J Nutr* 2003;133:3758-66S.
- Ulrich CM. Nutrigenetics in cancer research-folate metabolism and colorectal cancer. *J Nutr* 2005;135:2698-702.
- Giovannucci E. Epidemiologic studies of folate and colorectal neoplasia: a review. *J Nutr* 2002;132:2350-5S.
- Rampersaud GC, Bailey LB, Kauwaf GP. Relationship of folate to colorectal and cervical cancer: review and recommendations for practitioners. *J Am Diet Assoc* 2002;102:1273-82.
- Robien K, Ulrich CM. 5¹⁰-methylene tetrahydrofolate reductase polymorphisms and leukemia risk. *Am J Epidemiol* 2003;157:571-82.
- Wagner C. Biochemical role of folate in cellular metabolism. Folate in health and disease. New York: Marcel Dekker; 1995.
- Duthie SJ, Narayanan S, Brand GM, Pirie L, Grant G. Impact of folate deficiency on DNA stability. *J Nutr* 2002;132:2444-9S.
- Zou C-G, Banerjee R. Homocysteine and redox signaling. *Antioxid Redox Signal* 2005;7:547-59.
- Choi SW, Mason JB. Folate status: effects on pathways of colorectal carcinogenesis. *J Nutr* 2002;132:2413-8S.
- Ulrich CM, Potter JD. Folate supplementation: too much of a good thing? *Cancer Epidemiol Biomarkers Prev* 2006;15:189-93.
- Chen J, Giovannucci E, Kelsey K, et al. A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res* 1996;56:4862-4.

⁵ D. Anderson et al. Propagation of fluctuations in biochemical systems. Linear SSC networks, 2006, submitted for publication.

20. Ulrich CM, Bigler J, Bostick R, Fosdick L, Potter JD. Thymidylate synthase promoter polymorphism, interaction with folate intake, and risk of colorectal adenomas. *Cancer Res* 2002;62:3361–4.
21. Ulrich CM, Curtin K, Potter JD, Bigler J, Caan B, Slattery ML. Polymorphisms in the reduced folate carrier, thymidylate synthase, or methionine synthase and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14:2509–16.
22. Hung RJ, Brennan P, Malaveille C, et al. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:1013–21.
23. De Roos AJ, Poole C, Teschke K, Olshan AF. An application of hierarchical regression in the investigation of multiple paternal occupational exposures and neuroblastoma in offspring. *Am J Ind Med* 2001;39:477–86.
24. Novak B, Tyson JJ. A model for restriction point control of the mammalian cell cycle. *J Theor Biol* 2004;230:563–79.
25. Thiele I, Price ND, Vo TD, Palsson BO. Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J Biol Chem* 2005;280:11683–95.
26. Craciun G, Tang Y, Feinberg M. Understanding bistability in complex enzyme-driven reaction networks. *Proc Natl Acad Sci U S A*. In press 2006.
27. Ball K, Kurtz TG, Popovic L, Rempala G. Asymptomatic analysis of multiscale approximations to reaction networks. *Ann Appl Prob*. In press 2006.