Part I

1

Introduction

Lucio G. Costa and David L. Eaton
University of Washington, Seattle, WA

Although risks from exposure to exogenous chemicals depend on the intrinsic properties and dosage of the chemical, there is growing recognition that these risks may be significantly influenced by variations in target sites, biotransformation enzymes, and repair responses in the host. The host responses and target molecules are often specified genetically, and may have significant genetically determined variations. Significant discoveries relating genetically determined enzyme variations to adverse drug reactions were made in the 1950s, when the term "pharmacogenetics" was coined. The extrapolation that genetic variations would be expected to affect responses to any kind of environmental and xenobiotic agent, not just drugs, lead to "ecogenetics," which can be defined as the study of critical genetic determinants that dictate susceptibility to environmentally influenced adverse health effects.

With more recent advances in molecular sequencing technology, mutations [e.g., single-nucleotide polymorphisms (SNPs)] can be identified. Biochemical, clinical, and epidemiologic studies can then follow to assess whether such genetic polymorphisms have phenotypic consequences, and whether associations exist with specific disease outcomes and environmental exposures.

Gene-Environment Interactions: Fundamentals of Ecogenetics, edited by Lucio G. Costa and David L. Eaton

1.1. GENE–ENVIRONMENT INTERACTIONS: FROM PHARMACOGENETICS TO ECOGENETICS

In the 1950s, in a seminal paper, Arno G. Motulsky stated that "genetically conditioned drug reactions not only are of practical significance, but may be considered pertinent models for demonstrating the interaction of heredity and environment in the pathogenesis of disease" (Motulsky 1957). The first examples provided (e.g., heritable variations of plasma cholinesterase or primaquine-induced hemolysis, now known as *glucose-6-phosphate dehydrogenase deficiency*) paved the way for the rapid development of the field of "pharmacogenetics" (Vogel 1959; Kalow 1962; 1990), which studies individual differences in response to pharmacologic treatment as exhibited by drug toxicity or lack of therapeutic effect. [For a more recent review of the history and current state of the field of pharmacogenetics, see Meyer (2004).]

Today, the field of pharmacogenetics is transforming the way that the pharmaceutical industry assesses the efficacy and safety of potential new drugs. Alan Roses, a pharmaceutical industry leader in genetics and genomics, gained public notoriety for stating that "The vast majority of drugs—more than 90 percent—only work in 30 or 50 percent of the people" (Connor 2003). Roses (2004) notes that pharmacogenetics "is becoming the first drug-discovery pipeline technology to affect the structure and economics of the pharmaceutical industry." He adds that laboratory testing for common genetic differences in people (genotyping) is already being used to identify human disease-associated drug targets, and that "A new age for the treatment of diseases with safer and more targeted medicines is now beginning" (Roses 2004).

The concept that genetic variations would be expected to affect response to any kind of environmental and xenobiotic agent, not just drugs, was first made by Brewer (1971), who introduced the term "ecogenetics" (Motulsky 1991; Costa et al. 1993; Costa 2000; Omenn 2001). As discussed more recently (Eaton et al. 1998; Omenn 2001), there are clinical, policy, and scientific reasons for our increased attention to individual susceptibility. Susceptibility genes, unlike major genes associated with specific diseases (e.g., phenylketonuria or cystic fibrosis), are neither necessary nor sufficient to cause disease, but modify the risk when there is appropriate exposure (Eaton et al. 1998). Indeed, in most cases, the genetic difference does not result in a qualitatively different response, but rather induces a shift in the dose–response relationship (Kelada et al. 2004).

1.2. GENETIC POLYMORPHISMS

The term "genetic polymorphism" defines monogenic traits that exist in the normal population in at least two phenotypes, neither of which is rare. When the frequency of a specific genetic variation reaches 1% or more in the population, it is referred to as a *polymorphism*. A polymorphism may have no effect (i.e., be "silent") or may be considered functional if it results in altered function, stability, and/or level of expression of the resulting protein (Kelada et al. 2004). Functional polymorphisms include point mutations in the coding region of the gene, resulting in amino acid substitutions that in turn affect protein function or stability; duplicated genes, which result in higher protein levels; completely or partially deleted genes, resulting in no gene product; or splice-site variants that result in truncated or alternatively spliced protein products (Kelada et al. 2004).

Mutations in the regulatory regions of genes may affect the level of protein expression, and mutations in other noncoding regions may affect mRNA stability or splicing. Furthermore, even nonfunctional or silent mutations may be important, because they may be either in linkage disequilibrium with other functional polymorphisms or associated with still unknown functions (Kelada et al. 2004).

Most DNA polymorphisms occur as single-nucleotide polymorphisms (SNPs). As of June 2004 there were nearly 20 million SNPs identified in the NIH dbSNP database, of which about 800,000 are considered frequent (http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi).

Determining which of these SNPs are functional is a huge challenge. Functional SNPs located in exons [complementary SNPs (cSNPs)] are the easiest to locate and study. Of the 800,000 frequently occurring SNPs, over 120,000 are cSNPs, 40% of which (about 50,000) are expected to be functional. However, the number of SNPs affecting protein function, including level of expression, is believed to be much greater and, as mentioned earlier, will include many polymorphisms outside exons (Kelada et al. 2004). Many of the functional SNPs located outside exons are logically assumed to occur at exon–intron boundaries (splice-site variants), or in regulatory protein binding sites (*cis* elements). However, more recent studies have demonstrated that SNPs located in introns that seem far removed from any functional aspect of a gene (called "deep intronic variants") may also have functional significance, perhaps through alterations in splicing, or the activation of "cryptic exons," exons that are transcribed infrequently or selectively (Pagani and Baralle 2004). Even the redundant third base in many codons may have functional significance, even if it does not result in a change in amino acid (e.g., is a "synonymous" variant), by altering mRNA stability and/or splice efficiency. Thus, a major challenge in the field of eco-

genetics is the correct identification of genetic variation that is biologically meaningful.

1.3. THE ENVIRONMENTAL GENOME PROJECT

In 1997, as part of the Human Genome Project, the National Institute of Environmental Health Sciences (NIEHS) started a comprehensive effort to identify genetic polymorphisms in genes involved in environmentally induced disease, known as the "Environmental Genome Project" (EGP). Ken Olden, then director of NIEHS, liked to describe the relationship between genes and the environment this way: "Genes load the gun, the environment pulls the trigger." A loaded gun in itself causes no harm; it is only when the trigger is pulled that the potential for harm is released (Olden and Wilson 2000). Genetic susceptibility creates an analogous situation, where the loaded gun is equivalent to one or a combination of susceptibility genes, and the force pulling the trigger is an environmental exposure. The key objective of the EGP is to identify alleles that confer susceptibility to the adverse effects of environmental agents.

The EGP was proposed for development in three phases (Table 1.1). In the first phase, 554 candidate "environmental response" genes were identified, and are being resequenced to identify common genetic polymorphisms (SNPs). As of September 2004, 347 of these genes had been resequenced from 100 ethnically diverse human DNA samples, and hundreds of common SNPs had been identified. Such genes include those involved in xenobiotic

Table 1.1 Phases of the Environmental Genome Project

Phase 1	Develop a sample repository
	Resequence 200 candidate genes
	Develop new technologies for variant identification
	Develop databases of polymorphic variations
	Consider ethical, legal, and social implications
Phase 2	Functional studies of allelic variants
	Initiate population-based studies
	Refine databases
Phase 3	Implement genetic epidemiology studies
	Develop animal models
	Develop cell models
	Understand dose–response relationship
	Risk assessment

Source: Adapted from Olden and Wilson (2000).

metabolism, DNA repair, cell cycle regulation, cell death, oxidative metabolism, signal transduction, and immune and inflammatory response. (At the time of this writing, much of this resequencing is taking place at the University of Washington; see http://egp.gs.washington.edu for updated information.) In the second phase, the functional implications of polymorphisms in both the coding and the regulatory regions of these genes are being studied. A third phase involves the development of animal and cell models and the implementation of population-based genetic epidemiologic studies to investigate how a risk for a specific disease is altered by a particular genotype and environmental exposure. Considerations of the ethical, legal, and social implications (ELSI) of such research are an integral part of the EGP (see Chapters 23–26). Overall, the EGP will provide the ability to understand the combination of environmental and genetic components of important human diseases (cardiovascular disease, cancer, neurological disorders, asthma, etc.) and to identify and protect "at risk" subgroups.

1.4. A TEXTBOOK ON GENE–ENVIRONMENT INTERACTIONS

The growing realization that the vast majority of human diseases arise through a complex interplay of genetics and environment (gene–environment interactions) has advanced the field to the point that it has become a separate discipline in its own right. This volume is conceived as a textbook to be used primarily by students in the fields of toxicology, genetics, epidemiology, pharmacology, and other public health, medical, and general life science disciplines, as well as for practicing clinicians, public health officials, and regulators.

This book is divided into four sections. The first section covers fundamental aspects of ecogenetics, such as history of the discipline, a discussion of the molecular laboratory tools currently available to assess genotypes, the use of such measurements in molecular epidemiology studies, and the statistical issues involved in their analysis.

The second section focuses on a number of key genetic polymorphisms that are relevant for ecogenetics, including enzymes of phase I and phase II metabolism, enzymes involved in DNA repair, as well as receptors and ion channels. This section is by no means an exhaustive coverage of all possible polymorphisms relevant to gene–environment interactions, but is intended to highlight the characteristics of selected, widely studied genotypic–phenotypic differences, and discuss how these variations can influence responses to exogenous chemicals.

In the third section, gene–environment interactions are discussed through a disease-based approach to address the question of how genetic

polymorphisms can influence susceptibility to various diseases. The chapters in that section cover important disease conditions such as various types of cancer, neurodegenerative diseases, cardiovascular disease, chronic pulmonary diseases, infectious diseases, diabetes, and obesity. Diseases were selected that pose major health issues, causing high morbidity and mortality in the population, and for which information on the role of gene–environment interactions in the etiology of the disease is continuously emerging. Two chapters in this section also address the issue of diet and nutrition and their interaction with genetic backgrounds, as well as genetic determinants of addiction to alcohol, tobacco, and drugs of abuse.

The final section of this book discusses the ethical, legal, and social issues that must be considered when investigating and evaluating genetic polymorphisms in human populations, as well as the impact of ecogenetics on risk assessment, regulatory policies, and medicine and public health.