

Molecular Prodigality

Dennis Bray

Genome studies reveal that we share about 95% of our DNA sequence with chimpanzees (1). This confirms something we have known for years: We are not solely the products of our genes, but of our genes interacting with each other and with the environment. The

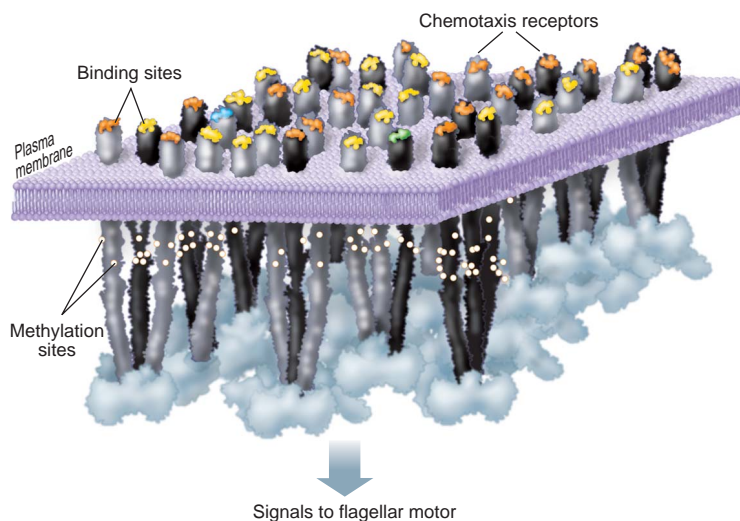
Enhanced online at
www.sciencemag.org/cgi/
content/full/299/5610/1189

outcome is a prodigiously complicated chemical system that not only renders humans different from monkeys, but also makes each human being (and probably each cell of each human being) unique. In living cells, macromolecules, and especially proteins, exist in a myriad of forms. But how many of these molecular variants are accidental products of RNA splicing and posttranslational modification, and how many are genetically designed and of evolutionary significance?

Consider, for example, the receptor used by the bacterium *Escherichia coli* to detect temperature. This is the very same protein—the chemotaxis receptor, Tar—that *E. coli* uses to detect the amino acid aspartate. Not only that, but the *sense* of the temperature response depends upon the concentration of aspartate in the environment. In a medium lacking this amino acid, Tar is a detector of warmth and *E. coli* swims toward a distant source of heat. But if the surrounding medium contains dissolved aspartate, then Tar becomes a cold detector and the bacterium swims away from heat. The molecular basis for this change in behavior is methylation of the cytoplasmic domains of the Tar protein (2). The receptor becomes highly methylated as it adapts to aspartate, and this apparently changes the conformational response of the receptor to a rise in temperature. But is it simply an accident that methylation causes Tar to change its re-

sponse to temperature? Or could it be that in *E. coli*'s natural environment—whether a gut, or a hamburger, or a drain—it is an advantage to titrate aspartate against temperature? In other words, could this effect have been selected for during evolution?

E. coli has three other types of methyl-accepting receptors in addition to Tar, and between them they mediate attractant and repellent responses to perhaps 50 distinct chemicals, as well as to pH and temperature (see the figure). Not only are the receptors multifunctional, but they also cluster together on the cell surface and have been shown to influence each other's re-



Combinatorial possibilities of receptor clusters. Several thousand transmembrane chemotaxis receptors of *E. coli* aggregate together in the plasma membrane by binding to downstream signaling proteins. Each receptor has eight possible sites of methylation (white circles) and can exist in at least two conformational states (gray or black). Four homologous types of receptor (distinguished by the color of their binding sites) are randomly mixed within the cluster and interact in groups of three.

sponses (3). So, the phenomenon observed with aspartate and temperature could be applied to many other combinatorial pairs, such as serine and aspartate, or ribose and pH. To what extent does the bacterium prioritize its responses to multiple environmental stimuli, and how much of this is genetically programmed?

It is difficult to answer this question because of the complexity of the clusters of *E. coli* chemotaxis receptors (see the figure). Each individual receptor can be methylated at up to eight positions and can adopt at least two possible conformations. Within a cluster, several thousand recep-

tors, of four different types, interact with each other in sets of three. Thus, the number of distinct receptor trimers—each having a different response to multiple environmental influences—is extremely large.

Turning to “real” cells, consider the potassium channels of a squid giant axon. These exist in different forms as a result of RNA editing, in which specific adenosine residues are converted to inosine, and then to guanosine. In one specific potassium channel gene, RNA editing produces amino acid replacements at up to 13 distinct locations (4). The altered amino acids produce various functional modifications to the channel, altering its voltage sensitivity, conductance, and even levels of expression. Each substitution site has a different efficiency, so that any individual protein molecule will have a distinct combination of substitutions out of a total of 8192 possibilities. Because each channel is composed of four subunits—and these appear to be combined in a random fashion—there should be 4.5×10^{15} distinct tetrameric channels.

These are not isolated examples. Many proteins of the mammalian myofibril exist in multiple forms as a result of alternative RNA splicing. Troponin T, which affects the calcium sensitivity of muscle, occurs in more than 80 different forms, which change their relative abundance depending on the type of muscle and on physiological processes such as growth and exercise (5). Because RNA splicing takes place in the nucleus and muscle myofibrils are multinucleate, the troponin T complement of

any one myofibril is unique. An even more dramatic case is that of the histone proteins in chromatin, whose amino-terminal tails are modified at multiple sites by acetylation, phosphorylation, and methylation. The enzymes responsible for these histone-tail modifications are highly specific for particular amino acid positions and are themselves influenced by a confluence of signaling pathways and by the local concentrations of other modified histones. These modifications are so extensive that they have been proposed to constitute a “histone code” that extends the information contained in the DNA sequence itself (6).

The author is in the Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. E-mail: d.bray@zoo.cam.ac.uk

CREDIT: KATHARINE SUTUIFF/SCIENCE

Mechanisms of genetic variation in proteins are abundant, especially in higher organisms. Up to one-half of all structural genes in humans may be alternatively spliced, and perhaps one-half of all proteins are modified by phosphorylation. The brain seems to be an especially rich source of molecular diversity, possibly reflecting its enormously complex anatomy and physiology. For example, humans have just three genes encoding neuronal cell surface proteins called neurexins. However, through alternative promoters and alternative splicing, these three genes yield thousands of different isoforms expressed in distinct combinations on the surfaces of different neuronal cell types (7, 8). Taken together with all other sources of molecular variation, it seems likely that each cell in an organism—even the estimated 10^{11} nerve cells of the human nervous system—is chemically unique.

Most variations in protein structure are likely to result from chance encounters between enzymes and their substrates. Many other variations are probably caused by environmental influences: temperature, mechanical stress, and the instantaneous local concentrations of ions and small molecules. But there also must be a purely genetic component underlying these structural variations. Each gene, by itself or in collaboration with other genes, is capable of elaborating a complex pattern of protein products in a manner reminiscent of cellular automata—computer simulations that generate astonishingly complex patterns by the repeated application of a few simple rules (9). But the extent of this genetically specified variation is completely unknown. How many of the variant forms of a given protein result from natural selection? Which modifications of protein structure have evolved so that the organism operates optimally under many different environmental conditions?

In the immune system, variations in protein structure are clearly functional. Mammals manufacture millions of lymphocytes, each able to make a distinct type of randomly generated antibody. Clonal selection allows the proliferation of those few lymphocytes that recognize a specific antigen; the remaining lymphocytes undergo cell death. Mechanisms for generating antibody diversity include the combinatorial joining of gene segments, random variations during gene segment joining, combinatorial joining of antibody light and heavy chains, and somatic hypermutation in the antibody gene. These mechanisms, taken together, allow a human to make perhaps 10^{12} different antibody molecules. The astonishing success of this strategy raises the question of whether echoes of this strategy

exist in nonimmune processes. Perhaps there are mechanisms of protein selection in cells that enhance some protein isoforms at the expense of others.

From an evolutionary standpoint, a protein's structure is significant only if it affects the survival of the organism. But many protein variants are likely to produce only subtle changes in cellular function. One chemical modification might slightly increase a cell's tendency to swim in a certain direction when confronted with a given combination of attractants. Another might marginally increase gene transcription at a specific stage of development. Investigators using genetic techniques to knock out specific genes have been puzzled to find that removal of even major components of cells can have little discernible effect on development. For example, mice lacking an isoform of creatine kinase, one of the most important enzymes in energy metabolism, grow like their wild-type relatives, are fertile, and have a normal life expectancy. It required a systematic study to show that these animals have abnormalities in brain circuitry and defects in spatial learning and habituation (10). Organisms have an enormous capacity to compensate and adapt to genetic changes. Individual cells can buffer small changes in protein structure so that they become apparent only during periods of stress (11). How can we evaluate the significance of subtle variations in protein structure?

SIGNAL TRANSDUCTION

Capturing Polo Kinase

Herman H. W. Silljé and Erich A. Nigg

The reversible addition of phosphate groups (phosphorylation) to proteins is one of the principal ways in which cells regulate protein activity. Phosphorylation not only directs the allosteric regulation of enzyme activity, but also is important for controlling protein-protein interactions, particularly those that assemble the protein complexes of signaling pathways. Phosphorylation creates docking sites in the phosphorylated protein to which other proteins then bind through specific phosphopeptide-binding domains. Prominent examples of phosphopeptide-binding domains include SH2 and PTB domains, which interact with phosphoty-

Fifty years ago, in the early days of molecular biology, Seymour Benzer drove the genetic analysis of a bacteriophage gene—the RII region of phage T2—down to the irreducible limit of a single nucleotide. Perhaps we need a similar enterprise today to solve the issue of molecular prodigality. Select a protein component, preferably one that is part of a large complex, and subject it to an exhaustive analysis. Find out precisely what forms this protein can take and use genetic or other methods to install each of these, separately, into an organism. Then subject this organism to a battery of high-throughput, high-resolution analyses in which every possible nuance of phenotypic behavior is examined. Such an approach would require huge populations of identical individuals and screens that can detect minor changes in survival rate under many different conditions. Something only a microbiologist could contemplate.

References

1. R. J. Britten, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13633 (2002).
2. T. Nara *et al.*, *J. Biol. Chem.* **271**, 17932 (1996).
3. P. Ames, C. A. Studdert, R. H. Reiser, J. S. Parkinson, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7060 (2002).
4. J. J. C. Rosenthal, F. Bezanilla, *Neuron* **34**, 743 (2002).
5. K. Nakada, J.-I. Mizazaki, T. Hirabayashi, *Differentiation* **70**, 92 (2002).
6. T. Jenuwein, C. D. Allis, *Science* **293**, 1074 (2001).
7. M. Missler, T. C. Sudhof, *Trends Genet.* **14**, 20 (1998).
8. K. Tabuchi, T. C. Sudhof, *Genomics* **79**, 849 (2002).
9. S. Wolfram, *A New Kind of Science* (Wolfram Media, Champaign, IL, 2002).
10. C. R. Jost *et al.*, *Eur. J. Neurosci.* **15**, 1692 (2002).
11. C. Queitsch, T. A. Sangster, S. Lindquist, *Nature* **417**, 618 (2002).

rosine motifs, and Forkhead-associated and WW (“double tryptophan”) domains, which interact with phosphoserine (pS) and phosphothreonine (pT) motifs (1, 2). On page 1228 of this issue, Elia and co-workers (3) describe a clever proteomics screen that they use to identify a new phosphopeptide-binding domain specific for pS and pT motifs. This new domain, the Polo box domain (PBD), was identified in Polo-like kinase 1 (Plk1), a kinase involved in the onset of mitosis. The PBD module appears to be specific to members of the PLK family and may help to regulate both the activity and subcellular localization of these kinases.

The novelty of Elia *et al.*'s approach is their use of an immobilized library of degenerate phosphopeptides to screen protein pools (prepared by *in vitro* translation of a cDNA library) for interacting proteins. Any

The authors are in the Department of Cell Biology, Max Planck Institute of Biochemistry, Am Klopferspitz 18a, D-82152 Martinsried, Germany. E-mail: sillje@biochem.mpg.de, nigg@biochem.mpg.de