

PERSPECTIVES

subsequent events occur over an ever wider range of delays.

Single-molecule methods (8) overcome the latter difficulty. Like an orchestra, a molecular ensemble needs to be synchronized for the outcome to be comprehensible. In contrast, a single molecule is a soloist, directly imparting its intricate improvisations to us. The pathway it takes can, in principle, be tracked as a function of time, allowing rare and potentially crucial events to be captured under physiological conditions and at equilibrium.

This dream has now come true for a single protein molecule (1). Yang *et al.* have followed the fluorescence lifetime of the flavin cofactor of a single enzyme as a function of time (see the figure). The flavin fluorescence is partly quenched by reversible electron transfer to a tyrosine residue. The transfer is sensitive to the distance R between flavin and tyrosine ($R \approx 0.45$ nm). Fluctuations of this distance by 0.1 nm cause variations of the fluorescence lifetime and intensity by a factor of 4.

These fluctuations lead to nonexponential fluorescence decay in an ensemble. Nonexponential decay is also observed for a single complex upon time averaging. The statistics of fluorescence lifetimes, or

equivalently of distances R , relate to the thermodynamics of R , which cannot be distinguished from those of a harmonic oscillator at this stage. Much more interesting, however, are the correlation times of the fluctuations, which tell us how long a given conformation lives, rather than how often it occurs. The correlation times range from seconds to hundreds of microseconds (the latter limit being imposed by the weak fluorescence signal).

The broad distribution of the correlation times may not come as a surprise, but the quantitative fit of the correlation to a stretched exponential, and the absence of times longer than 1 s, are unexpected. Stretched kinetics are a signature of a broad distribution of relaxation rates, indicating that the potential has wrinkles and hills on scales finer than its overall harmonic shape.

Electron transfer probes donor-acceptor distances shorter than a few nanometers. It thus complements single-pair energy transfer (9), which is suited to longer distances (2 to 10 nm). The tyrosine-flavine distance monitored in the present work may or may not be relevant to the enzymatic reaction. More direct information on function may be obtained by labeling the active site. An

even more ambitious goal will be to correlate enzyme-substrate conformations with activity fluctuations, which will have to be measured in real time.

Furthermore, deeper analysis of time trajectories of single molecules, such as that of Fig. 3C in (1), may reveal much more information than just a distribution of characteristic times. For example, the temporal order in which different conformations succeed one another may shed light on the topography of the energy landscape.

Relating function to structure and dynamics of proteins remains a major challenge in molecular biology. By adding a new method to our toolbox, Yang *et al.* have improved our odds for success.

References

1. H. Yang *et al.*, *Science* **302**, 262 (2003).
2. H. Frauenfelder *et al.*, *Science* **254**, 1598 (1991).
3. R. A. Böckmann, H. Grubmüller, *Nature Struct. Biol.* **9**, 198 (2002).
4. R. Ishima, D. A. Torchia, *Nature Struct. Biol.* **7**, 740 (2000).
5. L. Columbus, W. L. Hubbell, *Trends Biochem. Sci.* **27**, 288 (2002).
6. F. G. Parak, *Rep. Prog. Phys.* **66**, 103 (2003).
7. M. Lim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 5801 (1993).
8. Special issue on Single Molecules, *Science* **283**, 1593–1804 (1999).
9. S. Weiss, *Nature Struct. Biol.* **7**, 724 (2000).

GENOMICS

Microarrays— Guilt by Association

John Quackenbush

From their inception, DNA microarrays (1, 2) have been touted as having the potential to shed light on cellular processes by identifying groups of genes that appear to be coexpressed (3). Although promising, this “guilt by association” approach has fallen into disfavor because of its many perceived limitations. On page 249 of this issue, Stuart *et al.* (4) weigh in with their study of the coexpression of groups of genes across species, which suggests that there may be merit to the guilt-by-association approach provided it is applied in an evolutionary context.

Most microarray experiments focus on identifying patterns of gene expression in a particular system—for example, comparing tumor tissue with normal tissue, analyzing gene expression responses to a

stress stimulus, or comparing expression patterns in a particular tissue over time. Conceptually, differences between the various states should be reflected in changes in particular cellular pathways. Microarray data should allow users to identify the products of new genes and their contributions to these pathways. It is usually possible to identify cellular processes where most of the genes associated with a particular biological function are up- or down-regulated in a similar way. However, some genes known to be involved in a particular pathway invariably are missed, whereas other apparently unrelated genes exhibit expression profiles that are strikingly similar to bona fide pathway components. There is no consensus about how to interpret the gene expression patterns of hypothetical genes, genes of unknown function, or transcripts identified only by expressed sequence tags.

This perceived failure of microarrays has led some to portray the technique as

“noisy” or “unreliable.” Yet increasingly, when results from microarray studies are subjected to independent validation using other techniques (such as quantitative reverse transcription polymerase chain reaction), the confirmation rate—at least at the level of recapitulating the observed gene expression pattern if not the absolute magnitude—is well over 90%. So why aren't pathways emerging from microarray data that reveal groups of coexpressed genes?

One potential answer lies in the observation that many microarray studies fail to sufficiently sample the biological variability within a system. Increasingly persuasive arguments from statisticians (5–7), combined with improvements in the underlying protocols and technology used to collect expression data, have led to more sophisticated experimental designs encompassing increasingly broad surveys of diversity within a system. As a result, genes whose patterns of expression are identified as being statistically significant can be assigned a greater degree of confidence. Further, the validation rate—even among independent biological samples—is greater than in more naïve experimental designs. But has this resulted in the identification of new pathways? Not surprisingly, the answer is no. It seems that statistical significance is not always identical to biological significance.

The author is at The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA. E-mail: johnq@tigr.org

Stuart *et al.* attempt to address this problem by taking the idea of sampling biological variability one step further. They devised a computational method to look for correlated patterns of gene expression in more than 3182 DNA microarrays of tissues from humans, fruit flies, worms, and yeast. They postulated that genes with conserved biological functions in different species (orthologs) would be likely to retain similar patterns of expression while other associations occurring by chance would be filtered out by analysis of such a large multispecies data set. Stuart and colleagues predicted that conserved functions of groups of genes should be reflected in similar patterns of gene expression among yeast, worms, fruit flies, and humans.

The starting point for their analysis was the use of sequence homology to identify likely orthologs—genes that have retained their functions through evolutionary history. For example, Stuart and co-workers grouped together the human gene *Psmc4*, the worm gene *rpn-10*, the fruit fly gene *Pros54*, and the yeast gene *Rpn10*, all of which encode a non-ATPase subunit of the 19S proteasome cap. They used the expression values for these “metagene” groups taken from the entire data set to construct gene expression vectors. The authors then calculated pairwise Pearson correlation coefficients between these expression vec-

tors. The resulting correlation matrix served as the basis for constructing an interaction network of metagenes, subnets of which might be associated with particular biological functions. The results are both surprising and encouraging.

From the data, the investigators identified metagenes not previously associated with the cell cycle or proliferation. Yet these new metagenes exhibited significant association with these cellular processes. Such metagenes included several encoding proteins of unknown function as well as one encoding a small nuclear ribonucleoprotein (snRNP) involved in splicing and a second encoding a protein that interacts with nucleoporin. To test the hypothesis that these genes are in fact involved in the cell cycle, Stuart *et al.* first looked at expression of these metagenes in a previously published study comparing human pancreatic tumors with normal pancreatic tissue (8). As might be expected, the data showed that these genes were markedly up-regulated in rapidly dividing cancer cells compared with normal cells. Although this harks back to the guilt-by-association approach, the prediction of the association from an independent data set suggests that there might be real biology here. To lend further support to the involvement of these genes in the cell cycle and proliferation, the authors performed an RNA interference experiment in the worm

to down-regulate the worm gene *ZK652.1*, which encodes the snRNP splicing protein. They analyzed the resulting loss-of-function phenotype and discovered that the germline cells of treated worms contained extra nuclei. This suggested that the product of the *ZK652.1* gene normally suppresses proliferation of germline cells in the worm.

So have Stuart *et al.* provided evidence to vindicate the guilt-by-association strategy for analyzing microarrays? Not completely. Moving from associations to pathways will require additional work. And it remains unclear how this strategy can be applied to genes that are limited to a less evolutionarily diverse set of species. Yet this work represents an important conceptual step forward, extending comparative genomics to functional phylogenomics.

References

1. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995).
2. R. J. Lipshutz *et al.*, *Biotechniques* **19**, 442 (1995).
3. L. M. Staudt, P. O. Brown, *Annu. Rev. Immunol.* **18**, 829 (2000).
4. J. M. Stuart, E. Segal, D. Koller, S. K. Kim, *Science* **302**, 249 (2003); published online 21 August 2003 (10.1126/science.1087447).
5. G. A. Churchill, *Nature Genet.* **32** (suppl.), 490 (2002).
6. M. K. Kerr, G. A. Churchill, *Genet. Res.* **77**, 123 (2001).
7. R. Simon, M. D. Radmacher, K. Dobbin, *Genet. Epidemiol.* **23**, 21 (2002).
8. C. A. Iacobuzio-Donahue *et al.*, *Am. J. Pathol.* **162**, 1151 (2003).

DEVELOPMENTAL BIOLOGY

Rocks That Roll Zebrafish

Donna M. Fekete

There are relatively few places where the human geneticist, the developmental biologist, the protein biochemist, and the inorganic material scientist can find common ground. The field of biomineralization is such a place and the common ground is elegantly illustrated by the paper from Söllner, Nicolson, and colleagues (1) on page 282 of this issue. Examining the dysgenesis of inner-ear otoliths (so-called ear stones) in zebrafish, these authors show how alteration of a single protein called Starmaker influences the size, shape, and lattice structure of otolith salt crystals associated with these sensory organs. In humans, mutations of a related protein have been linked

to congenital deafness and defects in tooth mineralization.

Otoliths are large extracellular conglomerates in which organic biomolecules (proteins) are interleaved with inorganic crystals (usually calcium carbonate). In fishes, a single large otolith rests above each gravity-sensing organ, called a macula (see the figure). In mammals, birds, reptiles, and amphibians, each macula is blanketed instead by hundreds of tiny otoconia (so-called ear dust) enmeshed in a protein scaffold (2). Like the fish's otolith, the otoconia of higher vertebrates are dominated by microcrystals of calcium carbonate that together constitute more than 90% of the composite material. By virtue of their inertial mass, otoliths or otoconia are able to deflect the tiny sensory hair bundles beneath them to varying extents depending upon the orientation of the organ, and cor-

respondingly the head, with respect to gravity. In this way, these stealthy sensory organs are gravity sensors, although in fish they also sense sound.

Surprisingly, otoconia differ in size, shape, and crystal lattice structure (for example, the crystals can be composed of calcite, aragonite, or vaterite) depending upon the species, the type of macula (sacculus, utricle, or lagena), or even the developmental stage of a single organ (2–4). Different crystalline polymorphs correlate with differences in otoconial morphology across a wide range of vertebrates (2, 3). Not only is the functional significance of these otoconial morphotypes unclear, but little is known about what regulates the underlying biomineralization process that might contribute to this variation. One hypothesis is that both the crystalline polymorph and the resulting otoconial morphotype are controlled by a few prominent proteins found in the biomatrix (3). For example, the major proteins of calcitic otoconia in the utricles of amphibians, birds, and mammals have high molecular masses (90 to 100 kD), whereas the protein components of aragonite otoconia in amphibian sacculles have a lower molecular mass (3 to 22 kD). It is not known whether a single

The author is in the Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA. E-mail: dfekete@purdue.edu