Enterprise Biology Software: VI. Research (2005)

ROBERT P. BOLENDER

Enterprise Biology Software Project, P. O. Box 303, Medina, WA 98039-0303, USA
http://enterprisebiology.com

Summary

What would happen to the basic and clinical sciences if we had a **universal biology database**, one capable of storing and integrating research data from all disciplines? How would we design such a database and what could it do? On this the fifth anniversary of the **Enterprise Biology Software Project**, we have our first glimpse at the answers to these questions. The reader is invited to explore a world largely unfamiliar to most of us – a **data-driven biology**. Here answering questions consists of solving each piece on the way to solving the puzzle. If we identify our universal biology database as a puzzle, then the pieces become the literature, methods, change, complexity, prediction, discovery, and rules – assembled in that order. Easier puzzles, such as writing equations for organs, require fewer pieces. For example, unfolding and refolding the complexity piece was enough to generate diagnosis and prediction equations for the hippocampus. Perhaps the most important lesson learned thus far is that order in biology is everywhere and can be captured routinely with three common rules of nature - power, exponential, and optimization. Although the project offers a host of new insights, one was quite unexpected. By building an information system parallel to that of biology, we too can manage complexity and maybe – just maybe - learn to innovate by emulating biology. The report illustrates with examples how a data-driven biology can give us most of what we want – with surprising ease. The 2005 software package includes BIOLOGYtabs, solutions to three puzzles, updated databases and programs, and all previous releases.

Enterprise Biology

Discovery and innovation based on empirical equations define the research component of *enterprise biology*. The **Enterprise Biology Software Project** actively encourages this strategy by turning published research papers into database libraries and putting this new technology in the hands of those most qualified to make discoveries – the original authors. By bundling the data of one lab with those of many others, we begin the long process of transforming biology into a **data-driven science**. Notice what happens. Our individual research data actively support - and in turn are supported by - a community of many. Such a model offers great promise because it can deliver the critical mass of data needed to manage complexity in our research – now and in the future.

Introduction

Background

The **Enterprise Biology Software Project** is five years old. We will use this milestone as an opportunity to review what we can learn with this approach.

Advances in any field depend ultimately on what we want. We wanted to know the genetic code of humans, for example, and now we have the code. In fact, the **Human Genome Project** produces far more than genetic code. It generates new industries and changes the way we do science. The point is the strategy. What we have today can be explained by what someone wanted yesterday. Five years ago, the **EBS Project** set out

to anticipate what we - as biologists - might want now and in the future. Did it Work?
What did we want? What did we get?

Progress

Literature

***We want to sit down at our computer, click on an icon, and immediately view the best
available data of biological stereology – in standard format.*** Why? We want the option
of exploring the biology literature actively – with mathematics. We don't want the hassle
of having to find all the articles, check on all the methods, transform all the graphs back
into numbers, and make all the data tables needed to hunt for patterns, connections, and
rules.

What's wrong with the publications as they currently appear in journal articles? In short,
they are not user-friendly. Obtaining copies of research papers and extracting the data
can be both time-consuming and expensive. Often, data are "encrypted" in graphs,
"buried" in the body of the text, or reported in nonstandard ways. In any case, getting the
numerical data out of a paper requires a good deal of energy and patience. Getting the
data out of ten papers can be a serious undertaking. There is, however, a simple
alternative. If one person extracts the data from ten papers and shares the results with the
original authors, then together we enjoy free access to these data – quickly and easily. If
we expand our group of ten to include the larger community of biological stereology,
then we can become a data-driven discipline - one with the resources needed to explore
biology mathematically. Fortunately, the community has actively supported the project
by supplying thousands of reprints. ***Today, many in the stereology community can sit
down at their computer, click on the EBS icon, and…***

Methods

The best methods tend to produce the best data. Therefore, a research paper becomes a
candidate for data entry when it demonstrates unbiased sampling, applies design-based
methods, and satisfies the demands of the relational database model.

Unfortunately, there is a small problem. Research data invariably come with unwanted
baggage. All published data of biology carry unknown biases. Moreover, the animals
supplying the data add to the variability of our estimates – by differing in size, weight,
age, sex, environmental exposure, et cetera. Taken together, methodological biases and
animal variability routinely add *distracting noise* to our research data (Reports 2002-
2005).

We want to find distinct mathematical patterns in published data. This requires "quiet"
data, carrying a lightened burden of bias and variability and capable of giving regression
equations with coefficients of determination (r^2) close to 1.0. To get what we want, we
have to minimize both bias and variability. Since our research data now live in a
database where we can operate on them, mitigating the noise problem becomes routine.
Allow me to explain.

Whenever we don't want something, the simple mathematical solution is to throw it
away. Algebraic canceling makes unwanted things disappear completely, without a trace.
Let's see how this works. Absolute data characterize a structure – or one of its parts - as
a weight, volume, surface, length, or number. Each of these estimates carries a full load
of unknown bias and variability. Each part has units tied to the volume of its parent
structure, which typically carries most of the unwanted baggage. Divide two different
parts, each carrying similar units and sharing the same reference, and we get a
dimensionless ratio minimized for bias and variability. Moreover, the ratio defines a
quantitative connection between the two parts, one that often persists – by rule - within
and across animal species. This persistence – presumably reflecting the conservation of
animal genomes – routinely supplies regression equations with $r^2 \approx 1.0$.

The distinct mathematical patterns, which take the form of power and exponential equations with r² close to 1.0, can be readily viewed, filtered, and sorted by visiting an equation library (BIOLOGYtabs 2005). In turn, these equations become a convenient starting point when we want to hunt for clues or to explore a biological event.

Change

We want to interpret change correctly, as it occurs across the biological hierarchy of size – from molecules to organisms. Detecting change is the Achilles' heel of the basic and clinical sciences. It is a ready source of unimaginable trouble, anxiety, confusion, and complexity. Each year, the EBS project revisits the task of unraveling the complexity of change and tries to offer helpful guidelines (Reports 2002-2005). This year, a new puzzle program introduces the problem to advanced students in cellular, molecular, and systems biology (BIOLOGYtabs 2005;Puzzle 1; **Counting Molecules**).

The root of the change problem undoubtedly comes from the way we have been taught to deal with complexity. To wit: reduce each experiment to measuring only the "essential" variables and ignore - or hold constant - all the rest. This turns out to be a prescription for disaster if the biology refuses to accept our "ignore" and "hold constant" provisions. In truth, biology unwittingly finds itself in the embarrassing position of being unable to accommodate many of our experimental assumptions because it is - by nature - a vastly complex and totally interrelated system.

The simplest way of exploring the problem of change is to write a program and then use it to run experiments with and without complexity. By actually doing experiments on the computer, one quickly understands how information is connected in biology and why it becomes so difficult to interpret the results unless the experiment is expressed as one or more correctly balanced equations.

The **Counting Molecules** program also explores a promising link between biological stereology and molecular biology. Estimating a numerical density of molecules with stereology or measuring an optical density with biochemistry gives exactly the same thing - a molecular concentration. This fortunate overlap is duly noted in the program and allows the student to ask – and answer – a key question. Do the rules of change that we must obey in biological stereology also apply to molecular biology? In other words, does a "reference trap" exist in molecular biology as it does in stereology? If such a trap exists, what can a molecular biologist do to avoid falling into it? By using a simulator to answer these questions about change, the student quickly understands what is important to look for in an experimental design, a published paper, or a grant proposal.

Prediction

We – as biologists - want to be as clever as physicists and chemists at predicting results. For example, we want to collect data from a small part of a structure and use them – along with data already in the database - to predict the full structure and the animal to which the structure belongs. Although the request - at first reading – may appear somewhat excessive, the process of moving toward such a goal becomes the interesting part of the story. What would it take for us to do this? We would have to know quantitatively all the connections between all the parts that make up all the structures in that animal. Such connections would define a network of equations with which we could then interact.

Once again, we can make our job easier by breaking it into smaller, more manageable pieces. The first job is to write a set of equations that can generate an animal starting from a single data point by predicting structures both upstream and down. This can be accomplished by fitting published data to power equations and linking these equations across the biological hierarchy of size (Algorithms; BIOLOGYtabs 2003-2004). A comparison of predicted to observed values quickly confirmed the ability of the equations to reproduce the original data. Although this prediction exercise worked reasonably well

with a small data set, it became perfectly clear that reliable prediction systems based on the entire literature database would have to be built with power equations having r^2s very close to one (Reports 2002-2004). This provided a key clue.

Next, we have to mitigate the negative effects of animal variability. If, for example, the connections between two different structures are not always the same within and across animals, then we need an additional layer of information to account for such differences. A convenient way of doing is to treat the connections between pairs of structures as a *repertoire*, one that includes all the possible connections occurring in nature. Although such an approach to building networks of equations requires an exhaustive search, it nonetheless provides the extra level of information (Report, 2004).

It turns out that a considerable amount of animal variability – within and across animals - comes from the multiplicity of connections that can occur between the same pair of structures. This becomes an essential piece of information because it puts our prediction goal of having $r^2 \approx 1.0$ within reach. When this method of analysis was applied to the hippocampus, for example, two distinct networks of equations emerged: global (all species) and local (individual species). This tells us that prediction in biology will depend largely on our ability to manage complexity at this level of detail.

Starting with a single experimental value, we can now use networks of equations to predict the structure of the hippocampus, as it occurs in five different animal species. The predicted values are roughly within five percent of the published (observed) values. In other words, we now know how to transform published data into equations and to use them to predict the structure of the hippocampus. Observe, if you will, that the success of this approach depended entirely on a ready access to data coming from a great many papers. One finding was quite unexpected. Controls provided a far richer source of information than experimentals. It took 119 equations to assemble the networks of the control hippocampus, but only 5 to account for all the observed changes. In fact, assembling prediction and diagnostic tools from the stereology literature database is a surprisingly easy task. See, for example, BIOLOGYtabs 2005;Puzzle 2;**Hippocampus**.

Complexity

In experimental biology, complexity is everywhere and everything. In comes from the biology, the animals, the experiments, the methods, the data, the papers, and the interpretations (Reports, 2001-2005). Moreover, the effects of complexity are cumulative. Each step - throughout the long process of producing a research paper - adds its own measure to a growing total. In such a research setting, it comes as no surprise that agreement can be accompanied so often by disagreement. One thing is certain. Uncontrolled, complexity quickly becomes a destabilizing force - usually with undesirable consequences. In contrast, complexity under control becomes a steady and reliable source of new information.

We want to use complexity to our advantage. Complexity becomes a thoroughly manageable commodity by minimizing the limiting factors and maximizing the intrinsic order. ***By simply recognizing that biology stores its most accessible mathematical secrets in the connections between its parts, we can engage complexity comfortably and with confidence.***

Recall that stoichiometry is the mathematics of chemistry, defining the relative proportion of constituents. In biology, our stoichiometry takes the form of repertoire equations that define the relative proportion of parts.

Here is a short list of recommendations for managing complexity by minimizing the noise from bias and variability.

Source of Complexity	Management
Biology	Find the rules and follow them.
Animals	Report data by structure, strain, age, sex, diet, and side.
Experiments	Express experiments as one or more balanced equations.

Methods	Use unbiased sampling techniques and design-based methods.
Data	Report original data, not just percentages.
Papers	Publish numerical data in tables; include original data units.
Interpretations	Avoid traps by testing the logic of experimental designs.

Principles

We want to know the basic principles of organization being used by biology. Such information becomes useful in developing prediction equations, in working out the relationships of genes to gene products, and in searching for the physical antecedents of biological order.

We know that biological data, when expressed as connected pairs, can be fitted routinely to power equations with r^2s approaching 1.0 – in both control and experimental settings (Reports 2002-2005). Moreover, an entire set of power equations can be fitted to a single exponential equation, one for controls and another for experimentals (Report 2004; BIOLOGYtabs 2004-2005). This means that all the structural data stored in the stereology literature database can be summarized by two intersecting exponentials, each having an r^2 approaching 1.0. Are the fundamental organizing principles of biology being expressed as power and exponential equations? Yes, most likely. Are the intersecting exponential equations signaling a biology ultimately driven by best solutions (e.g., linear programming)? Perhaps. ***We now know that the rules of organization in biology can be described with power and exponential equations that show a remarkable interdependence within and across hierarchical levels.***

There is a simple fact. Everything operates according to the rules of nature. These rules can be expressed overtly - as they are in the physical sciences - or in biology they can be wrapped in the subtleties of complexity and emergent properties. Once we begin to explore biology as a mathematical puzzle, however, the rules become the rules of the game. If we play to win, then we enjoy a distinct advantage by knowing the rules. The easiest way of discovering these rules is to hunt for them empirically, just as physical scientists do. The problem for us in biology is that the price of admission to the game is extremely high. We have to become a data-driven science. The only way of doing this is to rethink how we manage our research data. Should we continue to bury them in our libraries (journals on shelves) or are we ready to put our data to work by storing them in databases?

Discovery

We want to know how to discover new things, using a data-driven approach to biology. How does this work? Such a question invites us to look back on the history of science and ask a more general question first. If we imagine that discovery depends largely on a single truth, what might that truth be? In science, the road to discovery was, is, and always will be paved with equations.

The Enterprise Biology Software Project is learning how to use equations in biology by letting mathematics and technology unfold and refold complexity. Unfolding is used to standardize published data by reducing them to a simpler, universal format. By connecting data – two pieces at a time – we can generate a large pool of data with exactly the right properties. Forming these data pairs reduces complexity by automatically minimizing both bias and variability (Report, 2004).

Starting with more than 25,000 pieces of well-behaved control data, the process of refolding complexity can be reduced to the straightforward task of generating equations with regression analysis. For example, refolding data pairs into **repertoire equations** uncovers the design rules being used by biology to construct structures – all across the biological hierarchy (4.4-4.6; BIOLOGYtabs 2005). In turn these rules (expressed as equations) can be assembled into networks of equations for diagnosis and prediction, incorporated into our experimental designs, or just used to hunt for more rules. Rules,

especially when they point to generalizations, become the powerful clues to finding direct connections to the physical sciences.

The newest collection of equations is called the **decimal library** (BIOLOGYtabs 2005; 4.7). Its purpose is to refold all the individual data pairs into equations, according to decimal steps. This means that every data pair (connection), which is reported as a single value (the ratio Y/X), can now be attached to and interpreted as an equation. Here refolding conveniently reduces more than 25,000 data pairs to just 103 equations. Moreover, decimal steps can be chosen – and updated - to predict the original data values (X or Y) with a given level of error. The current configuration accepts prediction errors of up to 15% and has decimal steps ranging from 0.001 to 30. The overall mean of the predicted values is 100.12% with a standard deviation of 4.53 (BIOLOGYtabs 2005; Puzzle 3; tab 4).

What does this decimal library do? It demonstrates further advantages of a data-driven biology. Finding data that share similar connections (when looking for related parts), pinpointing data in agreement with those of an ongoing study (when writing a discussion), and figuring out how to analyze complex data sets (when clear-cut patterns are lacking) – all become surprisingly straightforward tasks.

There is an important point. The power of the decimal library comes from the fact that we can translate sparse *point data* into equations and then operate on the equations. Since we already know that power equations routinely fit to exponentials globally, the power equations of the decimal library extend our reach importantly by allowing us to hunt for exponential equations locally. Let's look at some specific applications of the decimal library - applied directly to the literature.

A Genetic Surprise: One of the most active areas of research today includes modifying genomes and looking for phenotypic effects (e.g., transgenic mutants, isogenic strains, et cetera). In a recent study by Seecharan, Kulkarni, Lu, Rosen, and Williams (2003), cells were counted in the lateral geniculate nucleus (neurons, glia, and endothelial cells) and in the retina (retinal ganglion cells) – for 58 isogenic strains of the mouse. The purpose of the study was to look for interconnections between neuronal populations in the visual system. Modest correlations were reported for cells within the nucleus (e.g., $r=0.44; r=0.33$), but no correlations were found between the neurons in the nucleus and those in the retina.

When the cell counts from the lateral geniculate nucleus are entered into the stereology literature database, standardized, and unfolded into data pairs, we find the same pattern of data clumping reported in the original paper (before).

Before

$$y = -0.17x + 15216.71$$
$$R^2 = 0.03$$
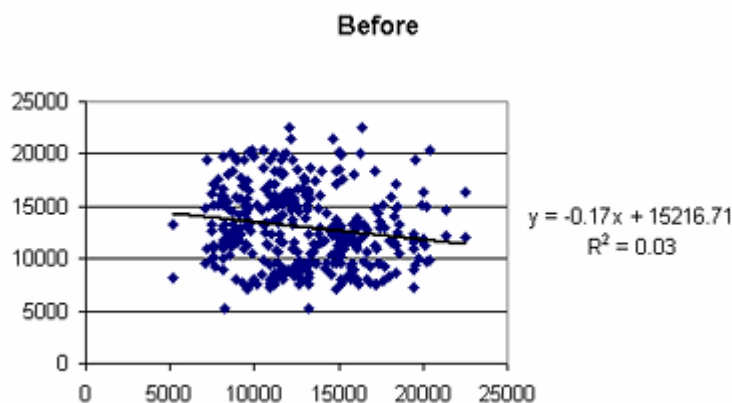
However, we now know how to turn these points into decimal equations and use them to look for patterns in these different mouse strains (after).
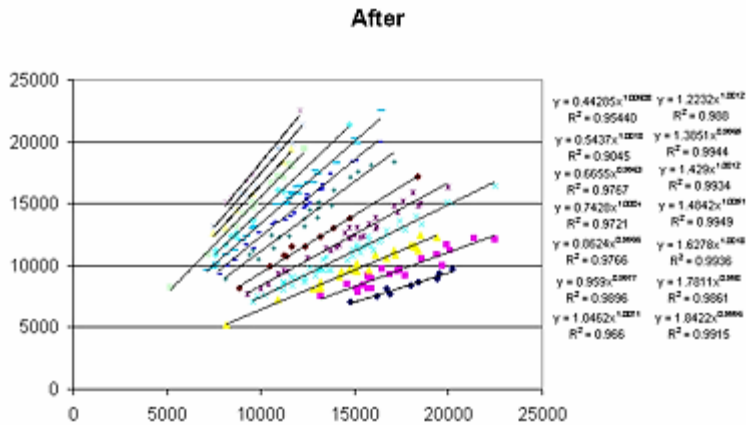
After

The after figure shows that the proportions of cells in the lateral geniculate nucleus – across the 58 isogenic strains - can be summarized by fourteen power equations. Each equation is defined as an individual decimal step in the new library. Notice that the curves (equations) have $r^2 > 0.9$ and identify a wide range of cell proportions. These equations clearly suggest that – for mice - there is more than one way to build a lateral geniculate nucleus. Indeed, many apparent "inconsistencies" in the biology literature might be reconciled merely by doing a repertoire analysis similar to the one shown here.

What do these equations tell us? It appears that operating anywhere on the genome – either by adding or subtracting genes – produces both intended and unintended consequences. In these isogenic strains, with genetic changes largely **unrelated** to the nervous system, the proportion of cells in the lateral geniculate nucleus shows extensive variability.

It takes fourteen equations to define the repertoire of all possible cell proportions that can occur in the lateral geniculate nucleus of these mice. When we translate these equations into an interactive connection matrix, similarities and differences among the strains can be quickly detected (4.7 Decimal Library; BIOLOGYtabs 2005). In biology, where multiple levels of control and redundancies abound, connection matrices may become a powerful tool for investigating complex relationships. They are already being used extensively in molecular biology.

Let's consider one implication of these unintended consequences. In the **Counting Molecules** program (BIOLOGYtabs 2005; Puzzle 1), recall that we explored the logic of our experimental model. How does this apply here? To study a biological variable (read gene or gene product), can we change one or more variables and hold the rest constant? Probably no. In biology, wherein everything is connected, variables may be **experimentally inseparable**. The standard experimental model – taken from the physical sciences - was intended for experiments with separable variables and may fail importantly when applied to the connected and therefore complex variables of biology. Indeed, we may need a new generation of experimental models in tune with biology.

Finding Growth Rules: Our second example from the literature of an application of the decimal library comes from the integument. Recall that we have already seen several

examples of exponential equations in biology, but only on the global level (Reports 2003-2004; BIOLOGYtabs 2005; 4.3; 5.4). If these exponential equations operate within and across hierarchical levels like the power equations, then we should be able to find them locally. Such a demonstration would be useful because it supplies the missing evidence needed to generalize order as a function of power and exponential distributions – within and across all levels.

Do, for example, cell organelles developing across an epithelium fit the log phase of a growth curve? Recall that this segment of a growth curve corresponds to an exponential equation. Using data from a single paper (Klein-Szanto AJ, 1977) at a single hierarchical level, the decimal library gives us our answer in the figure below (see also BIOLOGYtabs 2005; 4.7).

Keratinocyte - Organelle Growth Curves

$$y = 0.07e^{0.7249x} \quad R^2 = 0.9957$$
$$y = 2.9984e^{0.7064x} \quad R^2 = 0.9614$$
$$y = 0.269e^{0.6119x} \quad R^2 = 0.9995$$
$$y = 0.0376e^{0.6561x} \quad R^2 = 0.9914$$
$$y = 0.1077e^{0.6931x} \quad R^2 = 0.9666$$
$$y = 3.0257e^{0.6675x} \quad R^2 = 0.8447$$
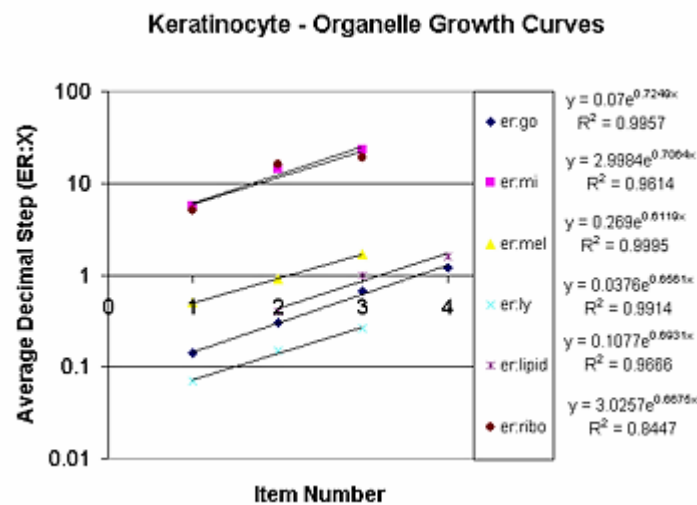
Over time, connections between the endoplasmic reticulum (er) and associated organelles (mitochondria, melanosomes, lysosomes, lipid, and ribosomes) all fit exponentials. As keratinocytes journey across the stratified squamous epithelium of the skin, their organelles grow according to an exponential rule. Moreover, this suggests that the growth of organelles in keratinocytes is being optimized. Why? Recall that maintaining a log growth phase for cells *in vitro* requires optimal growth conditions. It now appears that the analogy holds for organelles *in vivo* as well.

Future Opportunities and Challenges

As the published data of a discipline find their way into a database, that discipline becomes data-driven. The advantage of being data-driven is that it creates opportunities for finding empirical equations that can take us – more or less - to wherever we want to go. On the way to these equations, clues constantly surface. So many clues, in fact, that it becomes hard to know which ones to follow. Let's consider some of the more obvious ones.

The Concentration Trap

In basic and clinical research, most machines and methods produce concentration data. However, using just concentration data to detect change in biology is highly problematic in that we can expect such data to be correct roughly 50% of the time (Reports 2001-2005). In such a setting, it would be useful to know when a given concentration was telling us the truth or setting a trap. We can do this routinely with stereological data by comparing concentrations (densities) to structure or average cell data (BIOLOGYtabs 2005; 3.4). Elsewhere, such internal checks are often absent or seldom applied. A solution to the problem of detecting changes correctly - using just concentrations – therefore becomes an interesting problem. But, what are the clues and how do we find them? The question remains open.

In the meantime, an immediate solution to the concentration problem would be to develop new stereological methods that relate the concentrations of biochemistry and molecular biology to structures or average cells (BIOLOGYtabs 2005; **Counting**

Molecules). This integration of methods could contribute importantly to building the mathematical foundations of systems biology. Here the challenge is to know that an optical density represents a molecular count expressed as a concentration and to link this optical density reading to the supporting stereological equation. Moreover, these new integrated methods would define – automatically - experiments as balanced equations. Such equations would also satisfy the requirement of demonstrating the logic of an experimental design.

A Universal Database for the Basic and Clinical Sciences

If we want biology to become a data-driven science, then published data must be stored digitally in a database. The task of building a specialized database for each biology discipline and then linking them all together seamlessly would seem an impossible undertaking.

In fact, quite the opposite is true (BIOLOGYtabs 2005; Puzzle 3). A universal biology database is surprisingly easy to design and may require as few as four tables – citation, author, method, and data ratio. Forming the data ratio (Y/X) produces a universal data format by removing units and references and by minimizing bias. Integrating data across disciplines is accomplished - automatically - by putting all the data ratios into the same database table. The motivation for such a simple design becomes apparent when we realize that many authors publish just concentration data anyway, and that may be all we can expect to get. Moreover, this data format is already widely accepted. Data are routinely being reported - by most disciplines - as data ratios.

What is the likelihood of a universal biology database on the grand scale envisioned by the Biomatrix (Morowitz and Smith, 1987)? One year ago the answer was probably near zero - today it approaches 100%. What happened?   We now have free access to many research papers over the Internet (highwire.com; pubmed.gov) – given a six months delay after publication. This means that all the data needed to build such a database are currently available, or will be shortly. In effect, the long-standing barrier to such a project no longer exists.

Emulating Nature

Nature creates emergent properties by combining elements - or parts - to form entirely new things with new properties. Sodium and chlorine combine to give salt and hydrogen and oxygen become water. This process, which occurs continuously with parts of all sizes, is nature's method of discovery, innovation, and evolution. Not surprisingly, we too are largely a function of emergent properties and have become very good at producing them on our own. For example, we know how to take millions of parts, put them together, and fly from one place to another in the final product. Indeed, the richness of modern life can be attributed largely to emergent properties. This takes us to the point of the paragraph. If we – as biologists – elect to follow nature's example, then what can we do to generate emergent properties within our research environment? The answer is wonderfully simple. Collect the parts – in our case research data – fit them together, and see what happens. Sound familiar? This is the Enterprise Biology Software Project. We started out with a collection of parts (Stereology Literature Database), tried fitting them together (equation libraries), and produced a data-driven tool for discovery and innovation (Universal Biology Database). Today we can do things we couldn't do yesterday. What will tomorrow offer as the database grows and welcomes contributions from other disciplines? What new emergent properties will appear? In truth, we have absolutely no idea. Emergent properties are entirely unpredictable and serendipitous. When at their best, however, they become delightful surprises.

Concluding Comments

Complexity and simplicity stand at opposite ends of a continuum. Complexity in biology can be managed with great success by simplifying a given complexity to sets and subsets of equations. By identifying previously hidden layers of complexity, the repertoire

equations are providing rare insights into the rules of organization and how they are being applied by nature to expand the diversity of phenotypic expression. Such information becomes critical when we seek to modify genes and gene expression. Gene therapy performed within the framework of reductionism may miss – perhaps entirely – an enormous spectrum of unintended consequences. If we can write a set of prediction equations for the hippocampus, then we can also write them for the organism. Performing gene therapy within the framework of connectionism therefore becomes not only possible, but would seem to offer a safer and more prudent approach. Like it or not, we are rapidly becoming a major driving force in the process of our own evolution. We need to protect ourselves.

Acknowledgement

References

Bolender, R. P. 2001a Enterprise Biology Software I. Research (2001) In: Enterprise Biology Software, Version 1.0 © 2001 Robert P. Bolender

Bolender, R. P. 2002 Enterprise Biology Software III. Research (2002) In: Enterprise Biology Software, Version 2.0 © 2002 Robert P. Bolender

Bolender, R. P. 2003 Enterprise Biology Software IV. Research (2003) In: Enterprise Biology Software, Version 3.0 © 2003 Robert P. Bolender

Bolender, R. P. 2004 Enterprise Biology Software V. Research (2004) In: Enterprise Biology Software, Version 4.0 © 2004 Robert P. Bolender

Klein-Szanto, A.J. 1977 Stereologic baseline data of normal human epidermis. J Invest Dermatol 68: 73-78.

Morowitz, H. J., and T. Smith. 1987 Report of the Matrix of Biological Knowledge Workshop. Santa Fe, NM. Santa Fe Institute.

Seecharan, D.J., Kulkarni, A.L., Lu, L., Rosen, G.D., and R.W. Williams. 2003 Genetic control of interconnected neuronal populations in the mouse primary visual system. Neurosci 23: 11178-88.