Cross-fertilization between Proteomics and Computational Synthesis

Ivan I. Garibay and Annie S. Wu
{igaribay,aswu}@cs.ucf.edu
University of Central Florida
School of Electrical Engineering and Computer Science
P.O. Box 162362, Orlando, FL 32816-2362

Abstract

The problem of designing basic building blocks that self-reproduce, self-assemble, and self-organize into increasingly complex functionalities has already been solved by nature. Proteins are the basic building blocks of biological functionality. The genome, proteome, transcriptome and metabolome interact via regulatory networks, protein interaction networks, and metabolic webs, to build functional modules and ultimately achieve large-scale functional biological organization. Since proteomics seeks to understand the functions of the proteins as well as their complex interaction, the fundamental question that we post is: does the science of proteomics have something to teach us about the design competent functional building blocks and about the procedures for combining them to achieve high-level complex functionality? We believe that, as in many other cases, cross-fertilization among computational synthesis and proteomics would be productive. In particular, we provide a brief discussion on a relatively new and surprising proteomic result that may provide some insights on how to effectively organize building blocks: biochemical interaction networks share large-scale topology and properties. These networks are scale-free, modular, hierarchical, error and attack tolerant, their elementary building blocks self-organize into small recurrent patterns (i.e., motifs in gene regulatory networks, and pathways in metabolic networks), and they follow a simple self-organization premise: new nodes, if possible, prefer to connect to highly connected existing nodes. These results suggest that along the road from basic building blocks to high level functionality, we may find it useful to make a stop in the town of proteomics.

Introduction

The problem of designing basic building blocks that self-reproduce, self-assemble, and self-organize into increasingly complex functionalities has already been solved by nature. DNA is the information storage structure, proteins are the basic building blocks of biological functionality, and the universal genetic code defines the mapping between them. Beyond the protein level, the problem becomes increasingly complex and less well understood. At this level, we start talking about gene regulatory networks and, in a wider sense, about metabolic pathways.

Recently, the success of the various genome sequencing projects and the overwhelming amount of information that they have produced have fueled the new fields of genomics and proteomics. While genomics is the study of the complete set of genetic material that defines a biological organism—its genome—, proteomics (Wasinger *et al.* 1995) is the study of an organism's complete set of functional building blocks, the protein complement of its genome—its proteome. Proteins cooperate, compete, and antagonize in order to ultimately produce complex high-level functions. Proteins can self-organize into bigger molecular machinery and can trigger the production of more proteins on demand. In isolation, proteins implement to a great extent the basic functionality of their corresponding genes and, in conjunction with other proteins, they contribute as basic building blocks to the construction of an organism's increasingly complex functionalities.

In a sense, computational synthesis and proteomics approach the same problem from two very different perspectives. On the one hand, computational synthesis seeks to design competent functional building blocks and the procedures for combining them to achieve high-level complex functionality. On the other, the field of proteomics tries to understand how proteins (functional building blocks) have been designed by nature to achieve necessary functionalities and how proteins are combined to produce more complex biological function.

The fundamental question that we post is: does the science of proteomics have something to teach us about computational synthesis? We believe that, as in many other cases, cross-fertilization between this two fields would be productive. The interesting part is to discover what are the general principles that apply to both fields, if any. In principle, computational models of functional building blocks could reproduce some biological phenomena, which could contribute to the understanding of biological processes. Reciprocally, biological proteomics results can potentially inspire better engineering designs of functional building blocks.

It is in our research agenda to build proteome-inspired computer models of functional building blocks in order to promote complex emergent behavior in evolutionary algorithms for open-ended problems.

Computational Synthesis and Evolutionary Computation

We postulate that evolutionary algorithms can benefit from emergent complex behavior at the representation level and that this kind of behavior is desirable as a way to increase the "innovativeness" of the evolutionary search for open-ended design problems. In order to achieve this, we envision a method for building emergent behavior at the representation level by introducing intermediate representation structures (see Figure 1). For this method, we found inspiration in the natural process of genome to proteome transformations and the interactions that lead to high-level biological function, as studied by the sciences of genomics and proteomics. These intermediate representation structures parallel biological proteins in the role of "complexity builders". The feasibility of this line of thought strongly depends on our ability to adequately design these representation structures. These structures should, as proteins do, effectively organize themselves hierarchically into increasingly complex structures and ultimately lead to complex functionalities. In this sense, we are also on the computational synthesis quest: "From basic building blocks to high level functionality"

Our proteomics approach to evolutionary computation proposes the following general premises:

- The proteome's role in the evolutionary computation process is the role of complexity builder: self-assembly and self-organization.
- The genome and proteome should have different structures with specialized roles: replication and variation for the former; and functionality and complexity for the later.
- Problem representation is based on the functional space not on the evolutionary space. The evolutionary space, in traditional evolutionary algorithms, is the space of all possible genome encoding for a given problem representation. The functional space, in our approach, is the space of basic functional building blocks.
- The mapping between evolutionary space and functional space is problem independent.

Preliminary results (Wu & Garibay 2002a; 2002b) using a restricted implementation of this ideas indicate that proteome based representations are, on average, at least as good as a Simple GA for optimization problems. Our current research interest is to test our hypothesis of the proteome as "complexity builder" and its possible benefits for open-ended design using evolutionary techniques.

From genetics to genomics and into proteomics

Genetics, since Mendel's early work on inheritance (Mendel 1865) , has focused on the study of genes [1]. When Watson and Crick discovered the double helix shape of the DNA molecule and its self-replicating properties in the early 1950's, the idea of genes found a concrete foundation at the molecular level as a unit of information storage and transmission.

[1] Gregor Mendel call them "factors"

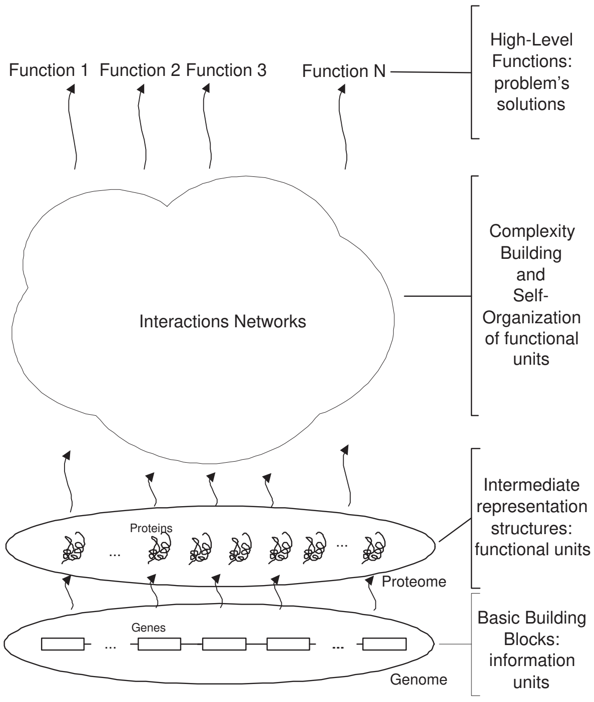

Figure 1: Evolutionary algorithms encode instances of solutions in artifical genomes. Designing this encoding is known as the problem representation. In the proteomics approach to evolutionary computation, we propose intermediate structures analogous to a proteome and related interaction networks. Basic building blocks are represented in the genome but the ultimate functionality of the representation will emerge from the building blocks network of interactions.

Relatively recent technological advances in molecular biology, particularly in DNA sequencing, have produced a shift of interest from the study of individual genes to a more ambitious endeavor: the study of the entire set of genes that define an organism—the genome. Genomics was born as a science devoted to genome analysis. Genome analysis has gone from genome sequencing to genome expression and function. In this sense, Genomics is a natural progression from Genetics.

Two additional reasons are believed to have contributed to the shift from the study of genes to the study of genomes. First, unlike genomes, genes are difficult to define. At the beginning of the 1960's, a gene was a DNA fragment that coded for a protein. Nowadays, this definition has become more flexible to accommodate various facts. Among them are that some proteins are made from several genes; that genes are fragmented over wide genome spaces (exons and introns); that those fragments, in some cases, may impact multiple genes (regulatory sequences); and that several different proteins, each of them with different functionality, can be made from the same gene (differential splicing, multiple reading frames, RNA editing (Scott 1995)). Genomes, by contrast, have a simple definition: the complete biological information of an organism (Morange 2001). The second reason is that no gene acts alone. A gene was originally thought to be a single hereditary unit that produces a single observable feature in an organism. This simple definition of gene functionality has since been abandoned in favor of a looser interpretation of gene function. Genes can act cooperatively or antagonistically, with the final observable traits being the result of a complex interaction of the entire set of genes that composes the genetic material of an organism (Ridley 2000). Based on these reasons, some researchers (Beurton & others 1995; Morange 2001) have argued for the use of the concept of bits of genomes instead of genes.

The primary factor for the flourishing of genomics, however, is not the gene's identity crisis, but the advent of the Genome Era heralded by the remarkable success of the numerous genome sequencing projects and the wealth of information they have produced. Sequencing an organism's genome means to spell out the information encoded on its genome in terms of DNA's four-letter alphabet. Projects of this nature were made possible by novel DNA sequencing techniques that became available near the end of the twenty-first century. The first whole genome of a free-living organism was published on 1995 (Fleischmann *et al.* 1995) and, as of today, there are 118 complete genome sequences in the public domain and 588 genomes under analysis (Bernal, Ear, & Kyrpides). The most remarkable genome sequenced, the *homo sapiens* genome, was published in 2001 independently by Venter and Lander (Venter & Others 2001; Lander 2001). The success of these genome sequencing projects and the public availability of the data have triggered efforts to analyze the information now available. This information is already being exploited in a number of ways. Comparative genomics compares the gene sequences of a particular organism with all other genomes in order to identify differences that could account for important properties;

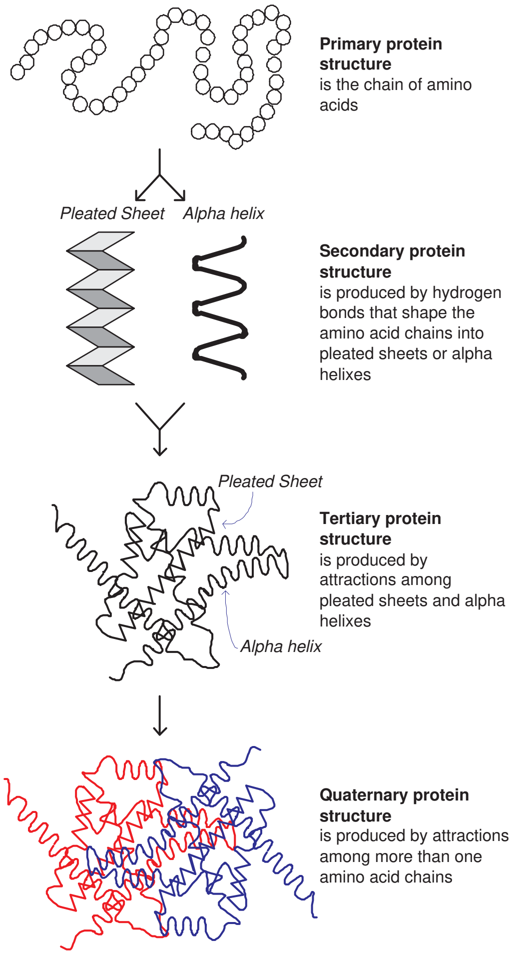

Figure 2: Protein schematic showing the three-dimentional folding on primary, secondary, tertiary and quaternary protein structures. Proteins are the primary functional building blocks.

structural genomics seeks to relate protein structure to gene sequences; and functional genomics tries to elucidate gene expression and function.

The huge amount of information produced by these projects, as well as the new techniques developed, have fueled not only the new science of Genomics but also the emerging field of Proteomics.

According to Palzkill (Palzkill 2002), proteomics is:

> "...a branch of functional genomics that has risen in response to the inevitable question posted by the genome sequencing projects, i.e., what are the functions of all the proteins? Proteomics can be defined as the large-scale study of protein properties such as expression levels, post-translational modifications and interactions with other molecules to obtain a global view of cellular processes at the protein level."

Proteomics focuses on analyzing the entire protein complement or *proteome* of an organism, as opposed to the genomics focus on the entire genetic content. The term 'proteome' was first used circa 1995 by various researchers (Wasinger *et al.* 1995) to describe the protein complement of a genome [2]

The proteome, unlike the genome, is not static. Each cell contains all the information necessary to build a new organism, but only a subset of the proteins encoded on a genome is expressed. Genes that code for proteins that are essential for cell function tend to be expressed in almost all cells, while more specialized proteins are only expressed in specific cell types (Liebler 2002). Thus, every organism has one genome but multiple proteomes. We can say that the proteome is the genomic expression for a given set of environmental factors such as developmental state—time—and cell differentiation—space.

From basic building blocks to high-level biological function

The most elementary biological building blocks of functionality are proteins. Proteins are involved in almost every biological process that takes place in living organisms and are ultimately responsible for all of an organism's properties.

Proteins are long molecules composed of one or more strings of amino acids chained together. Each of these individual chains is called a *polypeptide chain*. The encoded sequence of amino acids produces the necessary forces or molecular attractions that fold an amino acid string into a protein's characteristic three-dimensional shape (see Figure 2). These amino acid chains are assembled directly from information encoded in segments of the DNA chain of basepairs called *genes*.

In Mendelian genetics, a gene creates a phene (i.e., a "visible" character). Because alteration of a gene by mutation may alter its phene, different types of the same gene—genotypes—usually create different phenotypes. Nowadays, however, it has been established that high-level biological functions are not, generally, caused by any single gene or component. It is the rich and complex interactions among an organism's genome, transcriptome, proteome and metabolome that cooperate and antagonize to ultimately produce high-level biological function. There are numerous detailed accounts of particular metabolic pathways and of complete single gene regulatory maps , but what is known about large-scale biological organization, or about how all those pieces fit together in a coordinated way to produce high-level biological function, is far from understood.

[2]The term genome was coined by H. Winkler in 1920. Winkler created it from the words GENes and chromosOME to denote the complete set of chromosomes and their genes. While artificial at its beginnings, genome has since found its way into the English language. The American Heritage Dictionary of the English Language defines it as "The total genetic content contained in a haploid set of chromosomes in eukaryotes, in a single chromosome in bacteria, or in the DNA or RNA of viruses" or simply as "An organism's genetic material". The word proteome, however, is still artificial and denotes, in accordance with the term genome, the complete set of chromosomes and their encoded proteins.

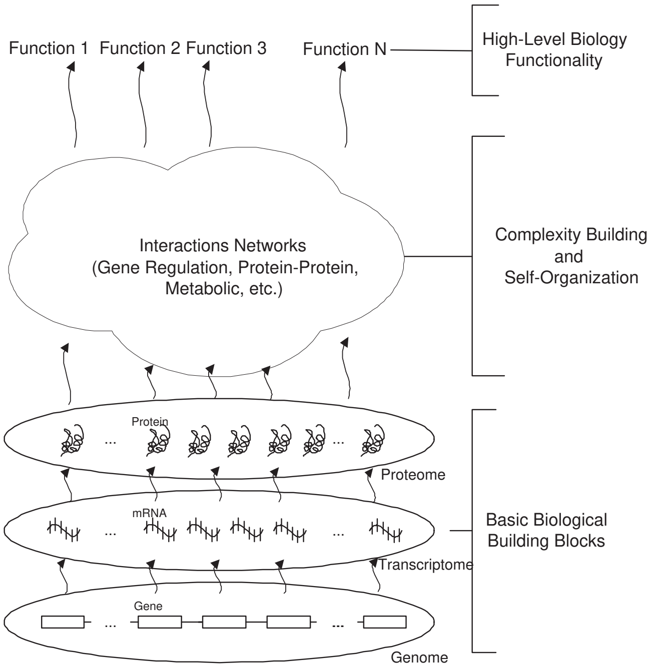

Figure 3: Genes, mRNA, and proteins as the basic building blocks of biological function. These basic components interact at many levels. Gene regulation networks, protein-protein interaction networks, and metabolic networks among others describe their interactions. From these interactions, complex high-level biological functionality emerge. Recently, it has been shown that these basic biological components self-organize into scale-free hierarchical networks with embedded modularity. It has also been shown that these type of networks are robust and error tolerant. These results seem to suggest that scale-free hierarchical networks with embedded modularity are a potentially good architecture for combining building blocks into high-level functionality in general.

Graphing building blocks interactions: interaction networks

The relatively recent availability of actual genomic sequences has opened a new perspective: the study of a biological organism as a whole. Genomics and proteomics are dedicated not only to produce but also to analyze all this newly obtained data. Thanks to this wealth of information, new approaches for unraveling large-scale biological organization are possible. One of the simplest is to chart interaction graphs from the available data and then topologically analyze these graphs in order to discover hidden properties or underlying organizational principles that could potentially improve understanding of these complex systems. The nodes of these interaction graphs (or interaction networks) are typically basic biological building blocks (genes, mRNAs, proteins, and metabolites) and the edges represent functional relationships. For example, gene regulation networks relate genes with the proteins that promote their expression; protein-protein interaction networks chart functionally related proteins; and metabolic networks relate reactant and product substrates of metabolic reactions. These interaction networks are generated from information accumulated over years of painstaking biological work coupled with the new automatically generated data from the analysis of genomic sequences. As a result, the information provided by these networks is organism-wide and their analysis provide information about the large-scale organization of the interactions among basic biological building blocks.

From the perspective of systems biology, these organizational principles can help us to understand how basic biological building block produce organism-level function. In contrast, from the perspective of computational synthesis, these principles can help us to understand how to construct high-level function from basic functional building blocks.

Survey of properties of biomolecular interaction networks

Recent results have shown that various kinds of biomolecular interaction networks exhibit similar properties. Such properties include the presence of small recurrent patterns (Lee & others 2002; Milo & others 2002; Shen-Orr *et al.* 2002), small-world (Wagner & Fell 2002; Barabási & Albert 1999), scale-free (Barabási & Albert 1999; Jeong *et al.* 2000; 2001; Wuchty 2002), hierarchical organization (Oltvai & Barabási 2002; Ravasz & others 2002), modular organization (Hartwell *et al.* 1999; Barabási *et al.* 2003; Csete & Doyle 2002), and robustness (Albert, Jeong, & Barabási 2000; Jeong & others 2001).

Small-world

Small-world networks are sparse and highly clustered and are characterized by the property that there is at least one short path between any two nodes in the network (friendship networks follow this pattern, hence the term *small work*) (Watts & Strogatz 1998; Watts 1999).

Wagner et al. (2002) discovered that the metabolic network of the *Escherichia coli* is a small-world network. Currently, we know that all the cell's biomolecular interaction networks are not only small-world, but also that they belong to a more restricted subclass: *scale-free* networks.

Scale-free

Recently, Barabási & Albert (1999) found that biomolecular interaction networks are small-world and with connectivity following the power law. By power law, we mean that the probability $P(k)$ that one node is connected with k other nodes decays with the power law: $P(k)\ k^{-\gamma}$. This property implies that biomolecular interaction networks self-organize into a scale-free state. Furthermore, this scale-free feature has been found to be a consequence of two general self-organization principles: (i) networks are in continuous expansion via the addition of new nodes, and (ii) new nodes prefer to connect into highly connected existing nodes (Barabási & Albert 1999) as shown in figure 4 (right). Interestingly, deterministic models of this scale-free networks show hierarchical organization and an apparent self-similarity (Barabási, Ravasz, & Vicsek 2001)(figure 6). In contrast, Figure 4 (right) shows a random network. In random networks, the connectivity follows a Poisson distribution peaked at $< k >$; hence the probability to find a node with more than $< k >$ connections decays exponentially. For instance, in metabolic networks—networks in which nodes represent reactant and product substrates and edges represent metabolic reactions—a systematic study of metabolic networks large-scale properties of 43 different organisms representing all three domains of life has revealed that all of them are scale-free networks (Jeong *et al.* 2000). Furthermore, the same study shows that the diameter of these metabolic networks (the shortest path averaged over all pairs of nodes) is the same for all 43 organisms independent of the number of nodes, that the most connected nodes are common for all 43 organisms, and that only 4% of all nodes are common for all 43 species. Additionally, in similar studies over the same 43 organisms, Jeong *et al.* (2001) found that the entire biochemical reaction network of the cell (including: metabolism, bioenergetics, information pathways, electron transport, transmembrane transport, and signal transduction) is also a scale-free network with similar properties as the metabolic networks alone. Recently, Wuchty (2002) report that protein-protein interaction networks, domain sequence network and domain interaction networks of the *Saccharomyces cerevisiae* are also scale-free networks.

Small recurrent patterns

Primary building blocks such genes, mRNA, proteins, and metabolites organize into small recurrent patterns (Oltvai & Barabási 2002). These patterns are recurrent metabolic pathways in metabolic networks, recurrent *network motifs* in gene regulatory networks, or simple basic network building blocks that frequently reappear in various different biological networks. In a study of the most complete gene regulatory map of the yeast (*Saccharomyces cerevisiae*) to date with up to 35,000 genetic-regulatory interactions, Lee & others (2002) has identified six small recurrent patterns or motifs. These six frequently appearing patterns are *autoregulation*, the binding of a regulator to its own promoter (in gene regulatory networks, nodes represent regulators and gene-promoters while edges represent the binding of a regulator to a promoter.); *multicomponent loop*, regulator A binds to promoter B and regulator B binds to promoter A; *feedforward loop*, regulator A binds to promoters B and C, and regulator B binds also to promoter C; *single input motif*, a single regulator binds to three promoters; *multi input motif*, all three regulators bind to all three promoters; and *regulator chain* regulator A binds to promoter B, regulator B binds to promoter C, regulator C bind to promoters D_1 and D_2. Similar motifs where obtained in an analogous study of the bacterium *Escherichia coli* (Milo & others 2002; Shen-Orr *et al.* 2002).

Hierarchical organization and modularity

Small recurrent patterns and primary building blocks are organized into functional modules (Hartwell *et al.* 1999; Oltvai & Barabási 2002). These modules are spatially or chemically isolated and carry out single cellular functions that are combined hierarchically to obtain increasingly complex functionalities of large-scale cellular organization (Oltvai & Barabási 2002). Results based on the large-scale study of topological properties of metabolic networks of 43 distinct organisms show that these networks are organized into many small and highly connected topological modules (Ravasz & others 2002). These modules are organized hierarchically into larger, less clustered modules with the number of modules and their degree of clustering decreasing according with the power law. This study (Ravasz & others 2002) also presents experimental evidence that for the *Escherichia coli* these mathematically obtained topological modules correlate with real well known functional modules. Barabási *et al.* (2003) obtained similar results in the examination of protein interaction networks of the *Saccharomyces cerevisiae*.

As a result, biochemical interaction networks in general are scale-free, as shown in Figure 4, and are also hierarchically modular. In contrast with a plain modular network as shown in Figure 5, a hierarchically modular scale-free network (deterministic version) is shown in Figure 6 (Barabási, Ravasz, & Vicsek 2001).

Conclusions

The engineering problem of designing competent building blocks and methods to effectively combine them is analogous to the organization of living systems. In biology, the genome, proteome, transcriptome, and metabolome are basic building blocks and their complex interactions ultimately produce high-level biological functionalities; however our current understanding of this natural process is far from complete. Recent advances of emerging fields such as genomics and proteomics have produce an avalanche of data, stimulating new explorations and discoveries regarding the complex interactions that lead from building blocks to complex functions.

We have briefly surveyed some of the current results from the field of proteomics, especially those related to how basic building blocks organize. We have found that there is

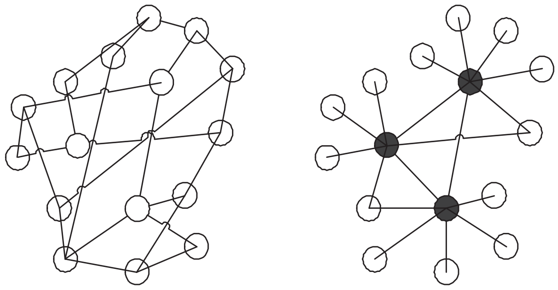

Figure 4: Random network (left) where every node is connected in average to k other nodes. Scale-free network (right) where the presence of *hubs* (black nodes in the picture) are the characteristic property, and the probability $P(k)$ that one node is connected with k other nodes decays with the power law: $P(k)\ k^{-\gamma}$. Scale-free networks naturally emerge from the premise that new nodes attach themselves to the network and that they strongly prefer to connect into highly connected nodes.

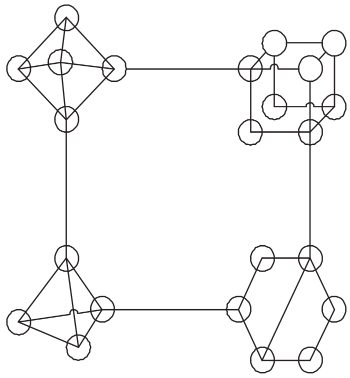

Figure 5: Modular network: highly connected clusters (modules) are connected together to form a network of modules.

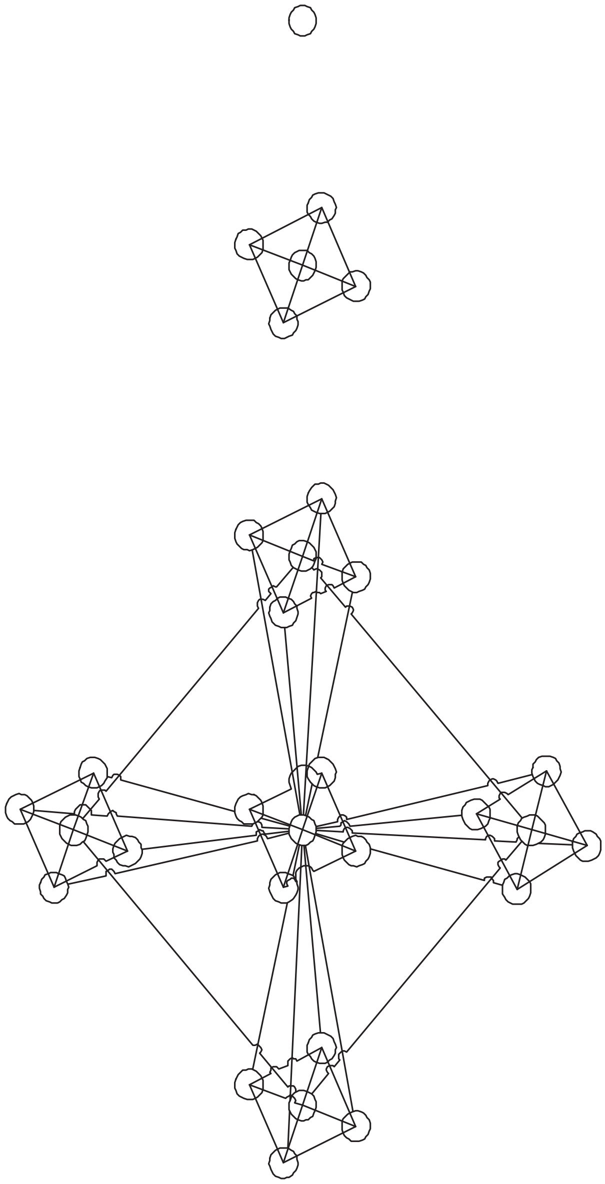

Figure 6: Scale-free hierarchical network with embedded modularity (deterministic version (Barabási, Ravasz, & Vicsek 2001)) presents properties of scale-free networks and of modular networks. Recently has been shown that this kind of networks closely resembles the large-scale topology of metabolic pathways, protein-protein interactions networks, and other cellular biochemical interaction networks. This results seems to suggest that there are common underlying principles of organization for biological functionality, and possibly for high-level funtionality in general.

a striking uniformity over all different biological networks of interaction (metabolic webs, gene regulatory networks, protein-protein interaction networks, etc.) and also over 43 different species. All of these networks seem to have the same underlying design principles at work. For instance, they all present very similar recurrent patterns and also they all share the same large-scale topological features: scale-free hierarchical networks with embedded modularity. All this suggests to us that the same underlying principles could be use to engineer competent systems to combine functional building blocks.

Overall these underlying principles are unknown, however, we have found that scale-free networks have two simple underlying self-organizing principles: (i) networks are in continuous expansion via addition of new nodes, and (ii) new nodes prefer to connect into highly connected existing nodes (Barabási & Albert 1999). We have also found that building blocks usually organize in recurrent patterns. Among these patterns are some very familiar engineering structures such as forward loops and backward loops.

We started by asking the question: does the science of proteomics have something to teach us about the design competent functional building blocks, and about procedures to combine them to achieve high-level complex functionality? What can be currently extrapolated from proteomics and related sciences into analogous computational systems is limited but promising; however, in this sequencing era, the fields of genomics and proteomics are in rapid growth as are the available genomic sequences and related data. All of this will enable the study of more and more properties of entire organisms from the uniquely perspective of their basic building blocks. As a result, we believe that the answer to our original posted question is a limited yes for the present, but a complete yes in the near future.

Therefore, we believe that computational synthesis and other computational sciences interested in the study of phenomena such self-organization and emergence would benefit from following the developments in proteomics and related fields, particularly in relation to their impact upon our current understanding of the organization of living systems.

References

Albert, R.; Jeong, H.; and Barabási, A.-L. 2000. Error and attack tolerance in complex networks. *Nature* 406:378–382.

Barabási, A.-L., and Albert, R. 1999. Emergence of scaling in random networks. *Science* 286:509–512.

Barabási, A.-L.; Dezsö, Z.; Ravasz, E.; Yook, S.-H.; and Oltvai, Z. 2003. Scale-free and hiearachical structures in complex networks. To appear in: Stiges Proceedings on Complex Networks.

Barabási, A.-L.; Ravasz, E.; and Vicsek, T. 2001. Deterministic scale-free networks. *Physica A* 299:559–564.

Bernal, A.; Ear, U.; and Kyrpides, N. Genomes online database (gold): a monitor of genome projects world-wide. http://wit.integratedgenomics.com/GOLD/.

Beurton, P. J., et al. 1995. Gene concepts and evolution. *Max-Plank-Institu für Wissenschaftsgeschichte* 18.

Csete, M. E., and Doyle, J. C. 2002. Reverse engineering of biological complexity. *Science* 295:1664.

Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; and Kerlavage, A. R. 1995. Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science* 269(5223):496–512.

Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; and Murray, A. W. 1999. From molecular to modular cell biology. *Nature* 402:C47–C52.

Jeong, H., et al. 2001. Lethality and centrality in protein networks. *Nature* 411:41.

Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; and Barabási, A.-L. 2000. The large-scale organization of metabolic networks. *Nature* 407:651–654.

Jeong, H.; Barabási, A.-L.; Tombor, B.; and Oltvai, Z. N. 2001. The global organization of cellular networks. In *Computation of biochemical pathways and genetic networks workshop*.

Lander, E. S. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860–921.

Lee, T. I., et al. 2002. Transcriptional ragulatory networks in *Saccharomyces cerevisiae*. *Science* 298:799–784.

Liebler, D. C. 2002. *Introduction to Proteomics: tools for the new biology*. Humana Press.

Mendel, G. 1865. *Experiments in plant hybridization*. ?

Milo, R., et al. 2002. Network motifs: Simple building blocks of complex networks. *Science* 289:824–827.

Morange, M. 2001. *The Misunderstood Gene*. Harvard University Press.

Oltvai, Z., and Barabási, A.-L. 2002. Life's complexity pyramid. *Science* 298:763–764.

Palzkill, T. 2002. *Proteomics*. Kluwer Academic Publishers.

Ravasz, E., et al. 2002. Hierarchical organization of modularity in metabolic networks. *Science* 297:1551–1555.

Ridley, M. 2000. *Genome*. Harper Collins Publishers.

Scott, J. 1995. A place in the world for RNA editing. *Cell* 81:833–836.

Shen-Orr, S. S.; Milo, R.; Mangan, S.; and Alon, U. 2002. Network motifs in the transcriptional regulation network of escherichia coli. *Nature Genetics* 31:64–68.

Venter, J. C., and Others. 2001. The sequence of the human genome. *Science* 291(5507):1304–1351.

Wagner, A., and Fell, D. 2002. The small world inside large metabolic networks. Technical Report SFI working paper: 00-07-041, Santa Fe Institute.

Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; and Humphery-Smith, I. 1995. Progress with gene-product mapping of the mollicutes: Mycoplasma genitalium. *Electrophoresis* 16(7):1090–1094.

Watts, D. J., and Strogatz, S. H. 1998. Collective dynamics of 'small-world' networks. *Nature* 393:440–442.

Watts, D. J. 1999. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton Reviews in Complexity. Princeton University Press.

Wu, A. S., and Garibay, I. 2002a. The proportional genetic algorithm: Gene expression in a genetic algorithm. *Genetic Programming and Evolvable Hardware* 3(2):157–192.

Wu, A. S., and Garibay, I. 2002b. The proportional genetic algorithm representation. In Langdon, W. B., et al., eds., *GECCO*, 703. Morgan Kaufmann Publishers.

Wuchty, S. 2002. Interaction and domain networks of yeast. *Proteomics* 2:1715–1723.