Acta Biochimica Polonica

Vol. 48 No. 2/2001

291–294

QUARTERLY

Understanding evolution

Jan Barciszewski[✉1] and Andrzej B. Legocki[✉2]

Institute of Bioorganic Chemistry of the Polish Academy of Sciences, Z. Noskowskiego 12/14, 61-704 Poznañ, Poland

At the end of the 20th century molecular genetics has been the central theme in biological thought. Nobody questions the importance of genetics, nor argues that DNA plays the role of the blueprint of life for all components of the living cell. However, we still do not understand the general mechanism of the cell function and development. The sequence of the human genome as well as other genomes is generally believed to be the starting point of biological and biomedical investigations in the following century and of the debate on the origin of life. Life depends on the interaction of thousands of genes and their protein products, orchestrated by the regulatory logic of each genome. If we are to comprehend this logic, we must hope that it can be dissected into a series of interlinked modules or networks, each of which can be studied in relative isolation. But even then, the complexity of a single module can be daunting. There is a hope that detailed annotation of the data base of genes and in-depth exploration of the physico-chemical principles of living systems will bring us closer to the understanding of the cell.

The basis of our insight into cell supramolecular structure is the doctrine of self assembly and self organisation which is a direct extension of the central dogma of molecular biology from the real sequence of linear information to the third dimension of protein and assemblies. The mechanism of self-assembly is very powerful and it operates far beyond atomic dimensions to form very complex structure like ribosomes or spliceosomes. At each level, novel laws can be found whose necessary, separate study has defined particular discipline. As Phil Anderson once said psychology is not applied biology, it is not applied chemistry or it is not applied physics. The features of the very same system depend on the scale of observation. This precludes the extrapolation of knowledge at one level to higher levels, where the “complexity” increases. Understanding why this is so and determining

✉1jbarcisz@ibch.poznan.pl; ✉2legocki@ibch.poznan.pl

how to formalize the problem of emergent features and multiscale description is one of the goals of the science of complex systems.

One can ask when we will be able to uncover universal principles working in living organisms and develop a set of design rules for biological activities, like those which the physical sciences already used to describe the mechanisms of the non-biological world. We should keep in mind that the organism is also a physical entity with geometric dimensions, subjected to the laws of macroscopic mechanics. Perhaps the most fundamental is to explain the immense diversity of life despite its deep and pervasively similar molecular architecture. The answer lies in conforming the presence of new genes, and then introducing them to a constantly changing ecological and physical environment. This can be possible through understanding the connection between phenotype and genotype. The implications of the revolution in molecular biology and developmental processes of the evolution are universally appreciated.

During the second half of the 20th century, biology was dominated by reductionist approaches that successfully generated information about individual cellular components and their functions. Understanding gene functions has to include knowledge about the hardware aspect. Cells are compartmentalized, and localization of proteins affects their function. Information transfer takes place not only through the specificity of protein binding. The cell responds to topological clues and mechanical forces that play a central role during morphogenesis and yet do not encode genetic information. Over the past decade, this process has been greatly accelerated by the emergence of genomics.

In the very near future, we will be overwhelmed by the exponential increase of biological data in terms of both volume and complexity. More and more powerful computers and computational tools for the understanding of the ever increasing number of databases will help to elucidate the lowest level compounds such as the structure and function of a molecule in biological networks. However, these tools may appear inadequate to uncover the complex system of control that characterize all living organisms. Evolution has produced families of proteins whose members share the same three-dimensional architecture and frequently have detectably similar sequences.

A time is coming when people will request more details and greater precision of the inferences drawn from complete genomes: how an enzyme performs its catalysis; why differences occur; what determines transcription differences, how cell induces changes in its neighbour and what shape will the organism be.

Structural genomics efforts have emerged in response to the fact that genome sequences encoding many proteins are often undetectable in the course of sequence comparison, and protein secondary and tertiary structures are highly coupled and difficult to predict accurately. The essence of structural genomics is to start from the gene sequence, produce a protein and determine its three-dimensional structure. The challenge, once the structure has been determined, is to extract useful biological information about the biochemical and biological role of the protein in the organism. This is a complete reversal of the classical structural biology paradigm, where a protein structure has been determined to understand how it performs its known biological function at the molecular level. The purpose of genomics is to understand biology: not simply to identify component and develop the experimental and computational methods, but also to take advantage of as much sequence information as possible and to catalogue all the genes together with the information on their functions; to understand how the components come and work together to comprise functioning cells and to make up the physiology of an organism.

Similarities between unknown polypeptides and known proteins revealed only at the level

of high resolution molecular structures might suggest a biological function.

Until now, more than 30 prokaryotic and several eukaryotic genomes, including yeast, worm, fly, human, have been solved. The sequence data will have the greatest impact on molecular medicine, as they will allow to better formulate the diagnosis of a disease. Functional genomics is the next step in this biological revolution. It has evolved from a surrealistic or at least futuristic concept in the 1980s to an accepted part of science at the beginning of the new millennium. It is not simply the association of a function to the identified genes but the organization and control of genetic pathways that come together to make up the physiology of an organism. Therefore, we need to find the roles and principles governing the mechanisms of biological activity. However, the genome industry is already in full swing. As a commercial activity, it will stimulate profits from the genome well before any drugs, diagnostics or technical advances of any kind have ascended from the nucleotide sequence.

Unfortunately, the billions of bases of DNA sequence do not tell us what all the genes do, how the cells work, how the cells form organisms, what goes wrong in the course of a disease, how we age or how to develop a drug. This is where functional genomics comes into play.

Expression array and proteomic technologies will give us the ability to determine when a cell uses particular genes and when it does not. Classical metabolism told us how a cell lives whereas proteomics is necessary to tell how a cell dies. Proteomics can be used to correlate gene expression data to cell metabolism and the organism phenotype. These potential applications make proteomics useful for studying plant physiological mechanisms, and also for providing clues on proteins of unknown function. However, this approach alone may not provide insight into the mechanisms that establish protein expression patterns. Since the important regulatory proteins such as transcription factors and signalling proteins are usually not visualized on two-dimensional gels due to their low abundance. Parallel studies of proteomes and transcriptomes should not only allow for the to understanding the relationship between mRNA and protein levels. It should also respond to the questions posed by large scale proteomic studies about the genes/proteins involved in the regulation of genome expression, from transcription to post-transcriptional processes.

As in other fields before, biology will experience an increased use of systems mathematics and computer simulations. A new mathematical biology is emerging. Building on experimental data on organism development it uses the powerful computational methods to explore the properties of real gene networks. Will it ever reach a level of sophistication in mathematical modelling and simulation similar to other fields? The complexity of living systems and their continuous change through evolution makes many sceptical about the success of such endeavours.

The main method of analysis in molecular biology has been the cartoon representation in different pathways. However, for their full understanding, numbers need to be attached to the arrows, and equations should be related to the numbers. What about the entropic factors, which are of paramount importance for their understanding? How do we deal with water in these calculations. Can we even calculate the enthalpy barriers to individual reaction steps with useful accuracy? Can we foresee the effects of amino acid substitutions at the active site?

Over the past 20 years, it has become clear that a variety of RNA molecules have important or essential biological functions in cells, beyond the well-established roles of ribosomal, transfer and messenger RNAs in protein biosynthesis. In RNA, sequence conservation among functional homologs is usually

limited to short segments, making homology search more difficult than the search for proteins. Our understanding of RNA is currently limited by the lack of structural data.

It is not yet clear how many structural RNAs are expressed in different cell types, what biochemical pathways they participate in and what proteins they bind. Structural genomics of RNA (ribonomics) will be most interesting when integrated with experimental and computational methods for identifying novel RNA genes and determining their biological relevance.

For the futuredevelopment of biology, integrative analysis of the function of multiple gene products has become a critical issue. Such approach will rely on bioinformatics and methods for system analysis. In the future, thebiological sciences will be increasingly focused an the systems properties of cellular and tissue functions.

Where are we now in understanding evolution?

Now it is not the end. It is not even the beginning of the end. But it is, perhaps, the end of thebeginning.