Cell, Vol. 92, 291–294, February 6, 1998, Copyright ©1998 by Cell Press

The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists

Overview

Bruce Alberts
President, National Academy of Sciences
2101 Constitution Avenue NW
Washington, D.C. 20418
Professor, Department of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143

Introduction

We have always underestimated cells. Undoubtedly we still do today. But at least we are no longer as naive as we were when I was a graduate student in the 1960s. Then, most of us viewed cells as containing a giant set of second-order reactions: molecules A and B were thought to diffuse freely, randomly colliding with each other to produce molecule AB—and likewise for the many other molecules that interact with each other inside a cell. This seemed reasonable because, as we had learned from studying physical chemistry, motions at the scale of molecules are incredibly rapid. Consider an enzyme, for example. If its substrate molecule is present at a concentration of 0.5 mM, which is only one substrate molecule for every 10^5 water molecules, the enzyme's active site will randomly collide with about 500,000 molecules of substrate per second. And a typical globular protein will be spinning to and fro, turning about various axes at rates corresponding to a million rotations per second.

But, as it turns out, we can walk and we can talk because the chemistry that makes life possible is much more elaborate and sophisticated than anything we students had ever considered. Proteins make up most of the dry mass of a cell. But instead of a cell dominated by randomly colliding individual protein molecules, we now know that nearly every major process in a cell is carried out by assemblies of 10 or more protein molecules. And, as it carries out its biological functions, each of these protein assemblies interacts with several other large complexes of proteins. Indeed, the entire cell can be viewed as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines.

Consider, as an example, the cell cycle–dependent degradation of specific proteins that helps to drive a cell through mitosis. First a large complex of about 10 proteins, the anaphase-promoting complex (APC), selects out a specific protein for polyubiquitination (King et al., 1996; Zachariae et al., 1996); this protein is then targeted to the proteasome's 19S cap complex formed from about 20 different subunits; and the cap complex then transfers the targeted protein into the barrel of the large 20S proteasome itself, where it is finally converted to small peptides (Baumeister et al., 1998 [this issue]).

Ordered Movements Drive Protein Machines

Why do we call the large protein assemblies that underlie cell function protein *machines*? Precisely because, like the machines invented by humans to deal efficiently with the macroscopic world, these protein assemblies contain highly coordinated moving parts. Within each protein assembly, intermolecular collisions are not only restricted to a small set of possibilities, but reaction C depends on reaction B, which in turn depends on reaction A—just as it would in a machine of our common experience (Alberts, 1984).

Underlying this highly organized activity are ordered conformational changes in one or more proteins driven by nucleoside triphosphate hydrolysis (or by other sources of energy, such as an ion gradient). Because the conformational changes driven in this way dissipate free energy, they generally proceed only in one direction.

An earlier brief review emphasized how the directionality imparted by nucleoside triphosphate hydrolyses allows allosteric proteins to function in three different ways: as motor proteins that move in a polarized fashion along a filament or a nucleic acid strand; as proofreading devices or "clocks" that increase the fidelity of biological reactions by screening out poorly matched partners; and as assembly factors that catalyze the formation of protein complexes and are then recycled. (See figure 1 in Alberts and Miake-Lye, 1992.)

Since the time of that review, the number of protein assemblies that are recognized to employ such devices has substantially increased. In particular, the nearly ubiquitous use of energy-driven conformational changes to promote the local assembly of protein complexes, thereby creating a high degree of order in the cell, has become universally recognized. A simple generic diagram of such a process is presented in Figure 1.

We have also come to realize that protein assemblies can be enormously complex. Consider for example the spliceosome. Composed of 5 small nuclear RNAs (snRNAs) and more than 50 proteins, this machine is thought to catalyze an ordered sequence of more than 10 RNA rearrangements as it removes an intron from an RNA transcript. As cogently described in this issue of *Cell* by Staley and Guthrie (1998), these steps involve at least eight RNA-dependent ATPase proteins and one GTPase, each of which is presumed to drive an ordered conformational change in the spliceosome and/or in its bound RNA molecule. As the example of the spliceosome should make clear, the cartoons thus far used to depict protein machines (e.g., Figure 1) vastly underestimate the sophistication of many of these remarkable devices.

Given the ubiquity of protein machines in biology, we should be seriously attempting a comparative analysis of all of the known machines, with the aim of classifying them into types and deriving some general principles for future analyses. Some of the methodologies that have been derived by the engineers who analyze the machines of our common experience are likely to be relevant. For example, modern machines comprised of subsystems from different "domains" (i.e., mechanical, electrical, fluid, thermal) are often analyzed by an energy-based approach. Here a mathematical description of the machine is achieved by considering certain scalar functions that represent the system energy (i.e., kinetic

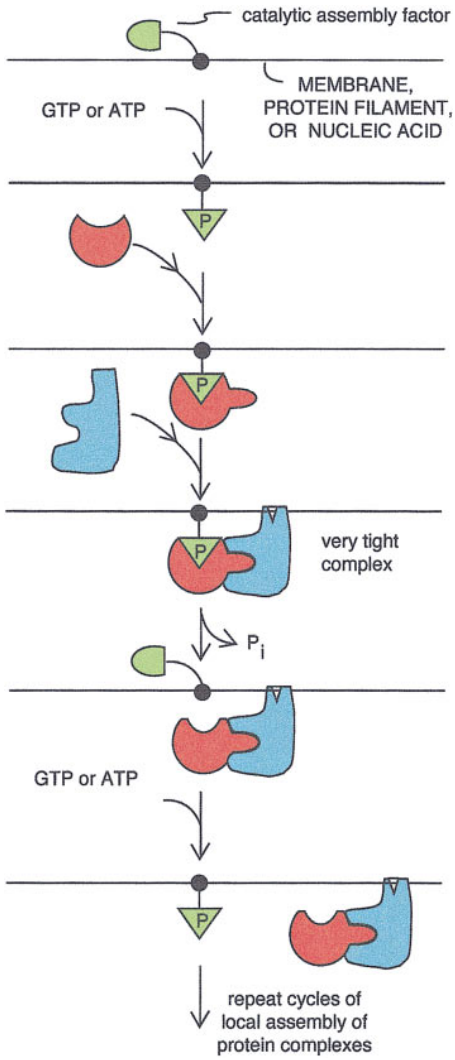

Figure 1. How the Energy Derived from Nucleoside Triphosphate Hydrolysis Makes Possible the Localized Assembly of Protein Complexes

In this schematic, the protein serving as a catalytic assembly factor either exchanges GDP for GTP, or is phosphorylated by a protein kinase using ATP. In either case, the added phosphate (P) activates this protein (green) to bind the red protein, which induces a conformational change that causes the blue protein also to bind. As indicated, this generates a very tight complex, in which each of the three proteins stabilizes the others in the complex. Loss of the indicated phosphate by hydrolysis then provides the energy needed to release the green protein, allowing it to be reused repeatedly as a local factor for assembling the other two proteins.

and potential energy) and the work done by external forces. The laws of nature are then enforced by application of first principles to arrive at the so-called equations of motion (Meirovitch, 1970; Ogata, 1992).

At the heart of such methods is the simplification and idealization of a real world machine as a composition of discrete elements. Engineers recognize certain fundamental behaviors in nature and then create an idealized element to represent each of those behaviors. Most simply, they classify elements as those that store kinetic energy, those that store potential energy, and those that dissipate energy. Any particular part of a machine might be modeled as consisting of one or more of these basic constituent elements. It seems reasonable to expect that different, but analogous approaches could profitably be applied to the protein machines that underlie the workings of all living things.

Should We Expect a Protein Machine to Be Well Engineered?

It is not hard to see why protein machines are advantageous to cells. A mere glance at the collection of articles in this issue of *Cell* should suffice to prove the point. Compare for example the speed and elegance of the machine that simultaneously replicates both strands of the DNA double helix (Baker and Bell, 1998 [this issue]) with what could be achieved if each of the individual components (DNA polymerase, DNA helicase, DNA primase, sliding clamp) acted instead in an uncoordinated manner.

But the devil is in the details. What, for example, has been the advantage to the higher eukaryotic cell of adding additional polypeptide chains to the DNA replication apparatus, while retaining the same basic functions as found in the bacterium *E. coli* and its viruses (Stillman, 1994)? And to what extent has the design of present-day protein machines been constrained by the long evolutionary pathway through which the function evolved, rather than being optimally engineered for the function at hand?

At least for protein synthesis on the ribosome, the evolutionary history—dating back to an "RNA world"—is thought to have played a predominant role (Green and Noller, 1997; Wilson and Noller, 1998 [this issue]). And when one examines the other protein assemblies known to operate in cells—such as the various complexes of RNA polymerase and its sets of accessory factors that catalyze transcription in eukaryotes—one is sometimes reminded of the many irrational complexities of a Rube Goldberg cartoon (Tjian, 1996; Greenblatt, 1997; Kadonaga, 1998 [this issue]). But perhaps this is only because we still understand so little of what the cell needs to accomplish with each of its various protein assemblies.

About ten years ago, I was struck by the speed and the elegance of the protein machine that replicates DNA (Alberts, 1987) in comparison to what I viewed then as a slow and ponderous ribosome. This led to a speculation: those present-day reactions that evolved early in the history of life on the earth (like protein synthesis) should have originated in a cell dominated by RNA catalysis; these reactions might therefore remain relatively inefficient, due to constraints traceable to their evolutionary history. In contrast, those present-day reactions that evolved later (like DNA replication), in a cell dominated by protein catalysis, could be expected to be much more efficient (Alberts, 1986). The complexity of the spliceosome might support this view, if one assumes that RNA splicing was a very early event that predated the existence of cells rich in proteins. However, the argument has certainly been weakened by the unexpected complexity of DNA transcription processes in eukaryotes, which I would have predicted to mimic DNA replication in their elegance and their simplicity.

Answers to puzzling questions like these will require that we acquire a much more complete understanding of the many protein assemblies that carry out the important functions of the cell.

How Should We Educate the Next Generation of Molecular Biologists?

This brings me to the central point of this introduction. A careful reading of this volume should convince everyone of at least two things: first, that we have made incredible progress in deciphering what we know today about protein assemblies; and second, that we still have an enormous amount more to learn. Thus, for example, our current drawings of the structure of the nuclear pore complex seem reminiscent of the sketches of houses that are drawn by young children, and they probably bear a similar relation to the real thing. Determining the structure of this fascinating cellular component, approximately 25 times larger than a ribosome (Ohno et al., 1998 [this issue]), remains a daunting challenge that will probably require methodologies not yet developed. And what new techniques will allow us to follow the kinetics and structure of each of the intermediates involved in the many fascinating transport reactions that occur deep within the lipid bilayer membrane? (See Matlack et al., 1998 [this issue]).

Even when we know the detailed structure of a protein assembly at an atomic level, as we do for the chaparonin GroEL-GroES, much will remain to be studied. As the article by Bukau and Horwich (1998) makes clear, any real understanding of the function of a protein machine will require not only its resting structure in atomic detail, but also a knowledge of the kinetics and energetics of each of its reaction intermediates. New techniques will need to be developed to facilitate such research. But, as always in biology, it will be crucial to define the key parameters that need to be determined, since much more can be measured than should be measured. Outstanding prototype investigations that are clearly explained and reexplained in review articles and textbooks can help both to shape this exciting new field and to recruit young scientists to it.

Many of my generation fear that the molecular biology revolution that we have just been through has made biological research look deceptively easy. Perhaps as a result, we generally find that even our most talented graduate students lack the background in the physical sciences that they are likely to need to decipher the detailed chemistry of protein machines. These individuals have taken the standard undergraduate courses in physics and chemistry, but they have generally not seen these subjects as central to carrying out research in molecular biology. True, in an era dominated by gene cloning, many of today's most distinguished scientists have been enormously productive without any quantitative skills. But the students of today will carry out most of their research in a post-genome-sequencing era, when most of the advances in molecular biology will come from successfully dissecting complicated in vitro systems composed of pure components (e.g., proteins, nucleic acids, and/or membranes). Here a deep understanding of the key constraints on the system posed by thermodynamic and kinetic factors, as well as an ability to use new developments in chemistry and physics as appropriate tools, will often be vital for success.

From my point of view, the education that we are offering today to young biologists in our colleges and universities is seriously in need of a major rethinking. The good news is that biology has become an increasingly popular major for our undergraduates, and there is no reason why we cannot excite all of them about science—whether budding researchers, premedical students, or those aiming for other professions. But the bad news is that far too many of our introductory courses are tedious surveys of an entire field—as if, for example, one could hope to gain any real understanding of all of biology in a single year. And in an era where there is a uniform push for exposing K-12 students to "science as inquiry," as emphasized in the National Science Education Standards (National Research Council, 1996, 1997), it remains hard to find any evidence of inquiry in most of our introductory college science laboratories.

Most important for the future of our field, the departmental structures at most universities seem to have thus far prevented any major rethinking of what preparation in mathematics, what preparation in physics, and what preparation in chemistry is most appropriate for either the research biologists or the medical doctors who will be working 10 or 20 years from now. The result is a major mismatch between what today's students who are interested in biology should be learning and the actual course offerings that are available to them. It is largely for this reason, I believe, that so many talented young biologists feel that mathematics, chemistry, and physics are of minor importance to their careers.

It is my hope that some of the young scientists who read this issue of *Cell* will come to the realization that much of the great future in biology lies in gaining a detailed understanding of the inner workings of the cell's many marvelous protein machines. With this perspective, students may well be motivated to gain the background in the quantitative sciences that they will need to explore this subject successfully. But they will need the faculty in our colleges and universities to lead them.

Acknowledgments

I am indebted to Jonathan Alberts for his explanations of how engineers analyze machines, Mei Lie Wong for preparation of the figure, and Teresa Donovan for manuscript preparation.

References

Alberts, B.M. (1984). The DNA enzymology of protein machines. Cold Spring Harb. Symp. Quant. Biol. *49*, 1–12.

Alberts, B.M. (1986). The function of the hereditary materials: biological catalyses reflect the cell's evolutionary history. Amer. Zool. *26*, 781–796.

Alberts, B.M. (1987). Prokaryotic DNA replication mechanisms. Phil. Trans. R. Soc. Lond. B *317*, 395–420.

Alberts, B., and Miake-Lye, R. (1992). Unscrambling the puzzle of biological machines: the importance of the details. Cell *68*, 415–420.

Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. Cell, this issue, *92*, 367–380.

Baker, T.A., and Bell, S.P. (1998). Polymerases and the replisome: machines within machines. Cell, this issue, *92*, 295–305.

Bukau, B., and Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. Cell, this issue, *92*, 351–366.

Green, R., and Noller, H. (1997). Ribosomes and translation. Annu. Rev. Biochem. *66*, 679–716.

Greenblatt, J. (1997). RNA polymerase II holoenzyme and transcriptional regulation. Curr. Opin. Cell. Biol. *9*, 310–319.

Kadonaga, J.T. (1998). Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. Cell, this issue, *92*, 307–313.

King, R., Deshaies, R., Peters, J., and Kirschner, M. (1996). How proteolysis drives the cell cycle. Science *274*, 1652–1659.

Matlack, K.E.S., Mothes, W., and Rapoport, T.A. (1998). Protein translocation: tunnel vision. Cell, this issue, *92*, 381–390.

Meirovitch, L. (1970). Methods of Analytical Dynamics (New York: McGraw-Hill).

National Research Council (1996). The National Science Education Standards (Washington, D.C.: National Academy Press). <www.nap.edu/readingroom/books/nses/>

National Research Council (1997). Science Teaching Reconsidered (Washington, D.C.: National Academy Press). <www.nap.edu/readingroom/books/str/>

Ogata, K. (1992). System Dynamics, 2nd Ed. (Englewood Cliffs, NJ: Prentice-Hall).

Ohno, M., Fornerod, M., and Mattaj, I.W. (1998). Nucleocytoplasmic transport: the last 200 nanometers. Cell, this issue, *92*, 327–336.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell, this issue, *92*, 315–326.

Stillman, B. (1994). Smart machines at the DNA replication fork. Cell *78*, 725–728.

Tjian, R. (1996). The biochemistry of transcription in eukaryotes: a paradigm for multisubunit regulatory complexes. Philos. Trans. R. Soc. Lond. B *351*, 491–499.

Wilson, K.S., and Noller, H.F. (1998). Molecular movement inside the translational engine. Cell, this issue, *92*, 337–349.

Zachariae, W., Shin, T., Galova, M., Obermaier, B., and Nasmyth, K. (1996). Identification of subunits of the anaphase-promoting complex of Saccharomyces cerevisiae. Science *274*, 1201–1204.