

# A New Biology for a New Century

Carl R. Woese\*

*Department of Microbiology, University of Illinois, Urbana, Illinois 61801*

<b>INTRODUCTION</b> .....	<b>173</b>
<b>THE MOLECULAR ERA IN THE BIGGER PICTURE</b> .....	<b>173</b>
<b>Reductionism versus Reductionism</b> .....	<b>174</b>
<b>Synthesis</b> .....	<b>175</b>
<b>TOWARDS A NEW REPRESENTATION OF BIOLOGY</b> .....	<b>175</b>
<b>CHANGING THE OVERVIEW</b> .....	<b>176</b>
<b>SOME PERTINENT HISTORY</b> .....	<b>176</b>
<b>THE PANDORA'S BOX OF MICROBIOLOGY</b> .....	<b>177</b>
<b>The Dismantling of Bacteriology and a Deconstruction of the Prokaryote</b> .....	<b>177</b>
<b>Other Guesswork Solutions?</b> .....	<b>179</b>
<b>CELLULAR EVOLUTION: THE BUMPY ROAD TO WHO KNOWS WHERE</b> .....	<b>179</b>
<b>THE DYNAMICS OF CELLULAR EVOLUTION</b> .....	<b>180</b>
<b>The Key to Understanding the Character of HGT</b> .....	<b>181</b>
<b>From There to Here</b> .....	<b>182</b>
<b>An Interesting, if Not Relevant, Aside</b> .....	<b>183</b>
<b>When Is a Tree Not a Tree?</b> .....	<b>183</b>
<b>ONE LAST LOOK</b> .....	<b>184</b>
<b>REFERENCES</b> .....	<b>185</b>

## INTRODUCTION

Science is an endless search for truth. Any representation of reality we develop can be only partial. There is no finality, sometimes no single best representation. There is only deeper understanding, more revealing and enveloping representations. Scientific advance, then, is a succession of newer representations superseding older ones, either because an older one has run its course and is no longer a reliable guide for a field or because the newer one is more powerful, encompassing, and productive than its predecessor(s).

Science is impelled by two main factors, technological advance and a guiding vision (overview). A properly balanced relationship between the two is key to the successful development of a science: without the proper technological advances the road ahead is blocked. Without a guiding vision there is no road ahead; the science becomes an engineering discipline, concerned with temporal practical problems. In its heyday the representation that came to dominate and define 20th century biology, molecular biology, was a rich and inspiring blend of the two. By the end of the 20th century, however, the molecular vision of biology had in essence been realized; what it could see of the master plan of the living world had been seen, leaving only the details to be filled in. How else could one rationalize the strange claim by some of the world's leading molecular biologists (among others) that the human genome (a medically inspired problem) is the "Holy Grail" of biology? What a stunning example of a biology that operates from an engineering perspective, a biology that has no genuine guiding vision!

Look back a hundred years. Didn't a similar sense of a science coming to completion pervade physics at the 19th century's end—the big problems were all solved; from here on out it was just a matter of working out the details? Deja vu! Biology today is no more fully understood in principle than physics was a century or so ago. In both cases the guiding vision has (or had) reached its end, and in both, a new, deeper, more invigorating representation of reality is (or was) called for.

A society that permits biology to become an engineering discipline, that allows that science to slip into the role of changing the living world without trying to understand it, is a danger to itself. Modern society knows that it desperately needs to learn how to live in harmony with the biosphere. Today more than ever we are in need of a science of biology that helps us to do this, shows the way. An engineering biology might still show us how to get there; it just doesn't know where "there" is.

## THE MOLECULAR ERA IN THE BIGGER PICTURE

If the dominant molecular representation of biology is to be displaced by something deeper, something more comprehensive and inspiring, we need first to step back, define molecular biology, and place the molecular era into proper historical perspective.

Despite the fact that historians may well declare the 20th to be "the great century" in biology (24), it was in the 19th century that biology really came of age; consolidating itself, ridding itself of much of its ancient burden of mystical claptrap, and defining the great biological problems: Pasteur had banished spontaneous generation for good. He, along with Koch, Haeckel, Cohn, Beijerinck, and others, had shown the living world to comprise far more than plants and animals. Darwin had demystified evolution and recast it scientifically. The cell had

\* Mailing address: Department of Microbiology, University of Illinois, 601 S. Goodwin, Urbana, IL 61801. Phone: (217) 333-9369. Fax: (217) 244-6697. E-mail: carl@ninja.life.uiuc.edu.

emerged as the basic unit of biology. The gene had begun to take form (in the mind's eye). Embryology, given an experimental dimension, was to become an ever deepening and fascinating puzzle. Add to this biology's perennial concern with the nature and significance of biological form, and you had a science well worthy of the name (40). The proof of this, if such were needed, was that the more "mature" sciences, first chemistry and then physics, began to treat biology as worthy of interest in its own right—as a source of interesting problems, not just interesting products (41).

The great problems of 19th century biology were of two conceptually quite different types, and this difference would be greatly enhanced in the climate within which 19th and (especially) 20th century biology developed. On the one hand were the "encapsulatable" problems, those of the gene and the cell. Understanding here lay very much in the parts. On the other hand were the holistic problems, evolution and the genesis and nature of biological form (organization), where the parts don't give a real sense of the whole.

The climate just referred to, of course, was the colorless, reductionist world of 19th century classical physics, which by that time had strongly affected the outlook of western society in general. The living world did not exist in any fundamental sense for classical physics (53): reality lay only in atoms, their interactions, and certain forces that acted at a distance. The living world, in all its complexity and beauty, was merely a secondary, highly derived and complicated manifestation of atomic reality and, like everything else in our direct experience, could (in principle) be completely explained (away) in terms of the ever-jostling sea of tiny atomic particles (53). The intuitive disparity between atomic reality and the "biological reality" inherent in direct experience became the dialectic that underlay the development of 20th century biology.

Given the technological flow of society and science, it was just a matter of time before 19th century physics (in the guise of molecular biology and its adjunct, biochemistry) would enter biology's world (41). But molecular biology would prove a mixed blessing. On the positive side, those problems (or portions thereof) that were amenable to a reductionist approach would benefit from the fresh, no-nonsense outlook and experimental power of molecular biology. In addition, biology as a whole would benefit from the physicist's general *modus operandi*, i.e., from the well-honed understanding of what science is and how it should be done: the crisp framing of problems, the clear understanding of what is and what isn't established truth, the importance of hypothesis testing, and the physicist's disinterested approach in general. On the negative side, biology's holistic problems, which were not commensurate with the new molecular perspective, would remain relatively or completely undeveloped. The result was a distorted growth of biology in the 20th century. The most pernicious aspect of the new molecular biology was its reductionist perspective, which came to permeate biology, completely changing its concept of living systems and leading then to a change in society's concept thereof.

### Reductionism versus Reductionism

We cannot proceed further without clarifying and discussing what is meant by reductionism. The stakes here are high be-

cause the concept is deeply woven into the fabric of modern biology, and biology today has hit the wall of biocomplexity, reductionism's nemesis. Thus, a topic that previously had been left for the philosophers and scientific dilettantes has suddenly become a very real and global issue for the practicing biologist. "Reductionism" is a confused and cathected issue at the moment, in large measure because biologists use the term in two senses, usually without distinguishing them. This we now have to do. We need to distinguish what can be called "empirical reductionism" from "fundamentalist reductionism." Empirical reductionism is in essence methodological; it is simply a mode of analysis, the dissection of a biological entity or system into its constituent parts in order better to understand it. Empirical reductionism makes no assumptions about the fundamental nature, an ultimate understanding, of living things. Fundamentalist reductionism (the reductionism of 19th century classical physics), on the other hand, is in essence metaphysical. It is *ipso facto* a statement about the nature of the world: living systems (like all else) can be completely understood in terms of the properties of their constituent parts. This is a view that flies in the face of what classically trained biologists tended to take for granted, the notion of emergent properties. Whereas emergence seems to be required to explain numerous biological phenomena, fundamentalist reductionism flatly denies its existence: in all cases the whole is no more than the sum of its parts. Thus, biology of the 20th century was in the strange position of having to contort itself to conform to a world view (fundamentalist reductionism) that 20th century physics was simultaneously in the process of rejecting. In a metaphysical sense, molecular biology was outdated from the onset! What makes this curious period in biology's history doubly bizarre is that a fundamentalist reductionist perspective wasn't even needed in the first place in order to study biology on the molecular level. The simple empirical reductionist outlook would have done just fine, and technology was moving us in that direction anyway! It will be interesting to see what history has to say about the biology of the 20th century.

It is instructive to catalog some of the changes that fundamental reductionism wrought in our perception and practice of biology. Chief among these is that the biologist's sense of what is important and what is fundamental was retooled to conform to the classical physicist's perception thereof. From this followed changes in the biologist's concept of organism, in his or her view of what constitutes an explanation, in what constitutes a "comprehensive" understanding of biology, in what biology's relationship to the other sciences is, in what biology can tell us about the nature of reality, in what biology's role in the society is, and in what biology's future course will be. These in turn produced changes in how biological knowledge is organized—the structure of academic curricula, the nature and purview of biological disciplines and text books, the priorities of biological funding agencies—and an overall change in the perception of biology by the society itself. All has by now been set in stone. It is impossible to discuss modern biology without the cacophony of materialistic reductionism throughout.

Biology's march into reductionism began in earnest with the "rediscovery" of the gene in the early 20th century. And the molecular dissection of the cell, which had begun with physiology being redefined (in part) on the level of enzymology, really took off with the advent of (molecular) genetics. The

problem of biological specificity soon became seen as the problem of specificity in molecular recognition, as manifested by enzymes and by antibodies. Familiar lock-in-key and hand-in-glove metaphors became the way to think about it. The whole problem of molecular specificity was raised to another level by Pauling and Delbruck when, in 1940, they proposed that not only was the notion of complementary molecular recognition useful in explaining enzyme and antibody specificity, but it could be used to conceptualize gene replication and gene expression as well (32). Molecular “templating,” tight apposition of molecular contours, seemed to be the *modus operandi* of biology, the basis of life.

The pinnacle of fundamentalist reductionism in biology was reached with the Watson-Crick structure of DNA. This structure, which clearly revealed the mechanism of gene replication, was hailed by molecular biologists as fundamentally solving the problem of the gene—a conclusion reified by the fact that once the Watson-Crick structure became known, most or all of the molecular biology coterie originally involved with the problem effectively packed their intellectual bags and moved on to “other great problems in Biology” (47). It is most interesting that molecular biologists declared the problem of the gene to be solved before the mechanism of translation (the core of gene expression) was at all understood—which, of course, was anathema to the classical biologist, who understood the gene to be defined by the genotype-phenotype relationship, by gene expression as well as gene replication. (I shall examine the implications of this signal, defining point in molecular biology’s history further below.)

### Synthesis

I think the 20th century molecular era will come to be seen as a necessary and unavoidable transition stage in the overall course of biology: necessary because only by adopting a heavily reductionist orientation and the technology of classical physics could certain biological problems be brought to fruition and transitional because a biology viewed through the eyes of fundamentalist reductionism is an incomplete biology. Knowing the parts of isolated entities is not enough. A musical metaphor expresses it best: molecular biology could read notes in the score, but it couldn’t hear the music.

The molecular cup is now empty. The time has come to replace the purely reductionist “eyes-down” molecular perspective with a new and genuinely holistic, “eyes-up,” view of the living world, one whose primary focus is on evolution, emergence, and biology’s innate complexity. (Note that this does not mean that the problems worked on in any new representation of biology will not be addressed by customary molecular methodology; it is just that they will no longer be defined from molecular biology’s procrustean reductionist perspective.)

I am obviously painting 20th century biology in too black-and-white a way. This is for didactic reasons. Of course trends don’t suddenly begin or suddenly end, and of course they don’t follow in mutually exclusive succession. However, it is often useful to portray them thusly because the trends themselves then stand out more clearly and the transitions between them are easier to recognize. For this reason I have ignored the progress that was made in evolution, morphology, and mor-

phogenesis during the 20th century. Yes, 20th century biologists did work to some extent on the holistic side of biology from the molecular perspective. Yet it is one thing to work on problems that are central to the governing paradigm, but quite another to work on those (such as molecular evolution) that are peripheral to it. In the former case, one’s work is swept into the mainstream, incorporated into the ruling world view, and vigorously developed. In the latter, the work more or less lies there as does rubble at a construction site, put up with but not appreciated, and hence underfunded and poorly developed. A future biology cannot be built within the conceptual superstructure of the past. The old superstructure has to be replaced by a new one before the holistic problems of biology can emerge as biology’s new mainstream and define its future goals (27).

### TOWARDS A NEW REPRESENTATION OF BIOLOGY

Nearly 40 years ago the physicist-philosopher David Bohm exposed the fundamental flaw in the mechanistic reductionist perspective (5): “It does seem odd . . . that just when physics is . . . moving away from mechanism, biology and psychology are moving closer to it. If the trend continues . . . scientists will be regarding living and intelligent beings as mechanical, while they suppose that inanimate matter is too complex and subtle to fit into the limited categories of mechanism.”

Bohm was warning us well before the fact that an engineering (mechanistic, reductionist) understanding of biology does not work because it is misleading and fails to capture biology’s essence. As is typical of prophecy, Bohm’s words went unheeded. Hopefully this time around there are ears to listen.

It has been known for some time that classical physics can deal with (formulate) only the more “linear” aspects of the world; true complexity, the vast “nonlinear” world that physics now recognizes to exist, is beyond the purview of classical physics (33). Thus, molecular biology, with its fundamentalist reductionistic mechanistic perspective, was faced with a difficult if not impossible task in developing a comprehensive understanding of biology. Not seeing the forest for the trees (and not caring what a tree was in any case), molecular biology took the only approach open to it: it clear-cut the forest. In other words, it dispensed with all those aspects of biology that it could not comprehend or effectively deal with (19). Molecular biology’s success over the last century has come solely from looking at certain ones of the problems biology poses (the gene and the nature of the cell) and looking at them from a purely reductionist point of view. It has produced an astounding harvest. The other problems, evolution and the nature of biological form, molecular biology chose to ignore, either failing outright to recognize them or dismissing them as inconsequential, as historical accidents, fundamentally inexplicable and irrelevant to our understanding of biology. Now, this should be cause for pause. Any educated layman knows that evolution is what distinguishes the living world from the inanimate. If one’s representation of reality takes evolution to be irrelevant to understanding biology, then it is one’s representation, not evolution, whose relevance should be questioned!

## CHANGING THE OVERVIEW

Let's stop looking at the organism purely as a molecular machine. The machine metaphor certainly provides insights, but these come at the price of overlooking much of what biology is. Machines are not made of parts that continually turn over, renew. The organism is. Machines are stable and accurate because they are designed and built to be so. The stability of an organism lies in resilience, the homeostatic capacity to reestablish itself. While a machine is a mere collection of parts, some sort of "sense of the whole" inheres in the organism, a quality that becomes particularly apparent in phenomena such as regeneration in amphibians and certain invertebrates and in the homeorhesis exhibited by developing embryos.

If they are not machines, then what are organisms? A metaphor far more to my liking is this. Imagine a child playing in a woodland stream, poking a stick into an eddy in the flowing current, thereby disrupting it. But the eddy quickly reforms. The child disperses it again. Again it reforms, and the fascinating game goes on. There you have it! Organisms are resilient patterns in a turbulent flow—patterns in an energy flow. A simple flow metaphor, of course, fails to capture much of what the organism is. None of our representations of organism capture it in its entirety. But the flow metaphor does begin to show us the organism's (and biology's) essence. And it is becoming increasingly clear that to understand living systems in any deep sense, we must come to see them not materialistically, as machines, but as (stable) complex, dynamic organization.

Twenty-first century biology will concern itself with the great "nonreductionist" 19th century biological problems that molecular biology left untouched. All of these problems are different aspects of one of the great problems in all of science, namely, the nature of (complex) organization. Evolution represents its dynamic, generative aspect; morphology and morphogenesis represent its emergent, material aspect. One can already see the problem of the evolution of cellular organization coming to the fore. And because of both its pressing practical and its fundamental nature, the problem of the basic structure of the biosphere is doing so as well.

My own career is one of the links between biology's reductionist molecular past and its holistic future. Thus, what follows will be autobiographically tinged.

## SOME PERTINENT HISTORY

I received my doctorate in biophysics from Yale University in the spring of 1953, just in time to celebrate the greatest achievement of the molecular era, the solving of the double-stranded structure of DNA (52). This one discovery, more than any other, exemplified the difference between the molecular perspective and that of the classical biologist. Here is where the battle between the two perspectives came to a head. As we have seen above, classical biologists effectively allowed biology itself to define the term "fundamental." Molecularists, on the other hand, imposed a reductionist definition of "fundamental," one that reflected their metaphysics. The process of gene replication was fundamental in molecularist eyes because it had a simple, reductionist, templating explanation, the mutual recognition of nucleotides according to the Watson-Crick pair-

ing rules ( $A \cdot T$  and  $G \cdot C$ ). Up to this point, classical biologists had no problem.

The clash came, however, over the issue of gene expression. Classical biologists naturally considered that process fundamental too. But for the molecularist, gene expression would be a fundamental biological process only if it too could be explained in simple molecular terms—for example, as the result of specific recognition of amino acids by corresponding oligonucleotides—and it was indeed along such lines that molecularists first sought to explain gene expression.

Enter the "era of the genetic code," when theoreticians and experimentalists alike were racing to see who would be first to "crack the code of life" (16, 22, 24, 30, 43). As we all know, once cracked, that code did not lead to a fundamental explanation of gene expression (translation). The code seemed to be merely an arbitrary correspondence table between the amino acids and corresponding trinucleotides. There seemed to be no simple physical-chemical interactions underlying the mechanism of gene expression (or that suggested the mode of its evolution). Could it be just another one of evolution's many "historical accidents"? Could there be nothing fundamental about it? That's how the molecularists saw it: outside of its structure, the only fundamental aspect of "the gene" was its mode of replication. Needless to say, classically trained biologists did not see it this way: in that translation (the heart of gene expression) was not yet understood, "the problem of the gene" could not possibly be completely (not to mention fundamentally) solved. No other single issue has exposed the difference between the molecular and classical perspectives more clearly than this one. Should the problem of translation be treated as just another (idiosyncratic) molecular mechanism (as it now is), or is that problem central, and thus fundamental, to the nature of the cell. As we shall see, biology today continues to live with this unresolved problem.

The genetic code became for me the looking glass through which I entered the world of real biology. Like many molecularists of the day, I was taken by the code, and at first I emulated their cryptographic approach to the problem (55). But that approach didn't have a biological "feel" to it. Wasn't it wrong to consider the codon assignments in cryptographic isolation? Weren't they just a superficial but important manifestation of something deeper and more interesting, i.e., how translation evolved? Here was the real problem of the gene, how the genotype-phenotype relationship had come to be. Translation, far from being just another relatively uninteresting study in biological idiosyncrasy, actually represented one of a new class of deep evolutionary questions, all of which had to be formulated and addressed on the molecular level.

Universal evolutionary problems of this kind can be approached only in the context of a universal phylogenetic framework, and in the mid-1960s, when I set out to study the evolution of translation, no such framework existed. Animal and plant phylogenies were reasonably fleshed out, but the huge and overwhelming bacterial world was effectively virgin phylogenetic territory. A massive job lay ahead merely to establish a framework within which to begin operating.

Fortunately, the technology for tackling the job had recently been developed by the one individual who, more than any other, had made 20th century biology technologically possible: Fred Sanger. In the mid-1960s, on his way to developing DNA

sequencing technology (and a second Nobel Prize), Sanger had come up with a method for partially characterizing RNA sequences, a two-dimensional paper electrophoretic method called oligonucleotide cataloging (38). Here was just what was needed, and it had come along at just the right time. While protein sequences were starting to be used to infer phylogenetic relationships, it was already evident that no known single protein sequence had the phylogenetic “reach” required to infer a universal tree (1, 2). However, one particular type of RNA might have. That was rRNA. rRNA molecules are relatively large, universal in distribution, and constant in function. Importantly, their sequences are highly conserved overall (13, 59), and, as central components of a complex and essential cellular mechanism, rRNAs arguably would be less subject to the vagaries of reticulate evolution than would other cellular components (13). If the universal tree could be inferred at all from one single molecular type, then Sanger’s oligonucleotide mapping method applied to rRNA was the way to go about it!

As the research program I had set in motion unfolded and the universal phylogenetic jigsaw puzzle began to assemble itself in its hit-or-miss way, the majority of taxa previously proposed (above the level of genus) were swept away (15, 59). In 1976 to 1978 the *Archaea* surfaced (65), with the first methanogen showing in June of 1976 (in a collaboration with Ralph Wolfe and his lab) (14); in May 1977 the first extreme halophiles appeared (28), to be followed at the end of that year by *Thermoplasma* and *Sulfolobus* (done with some initial help from Tom Langworthy) (14, 31).

### THE PANDORA’S BOX OF MICROBIOLOGY

In bringing to light the large-scale evolutionary order of life, our studies also made it apparent what a scientific mess 20th century microbiology was in. The discipline had languished too long: it had no concept of itself, was pulling itself hither and yon, and seemed headed for the pit of anonymity. Since the beginning of the century, microbiologists had wrestled with the problem of the natural (phylogenetic) relationships among the bacteria, which held the key to establishing bacteriology as an organismal discipline (as zoology and botany already were). Through no fault of their own, microbiologists had failed to create the needed phylogenetic framework, thus preventing bacteriology from developing into a real organismal discipline. The discipline lacked a meaningful concept of the organisms it studied (45), and there was no contemporary awareness of the serious effect this was having, not merely on the development of bacteriology but on the course of all of biology. (A bacteriology that was a full-fledged organismal discipline would have ameliorated the crippling procrustean reductionism of the molecular paradigm.)

Twentieth century bacteriology was a prime example of a science not seeking to define itself, letting itself instead be defined by external influences. The discipline had never sought to frame the overarching questions that synthesize and define a field. Quite the contrary: when such questions happened to come along, microbiologists either shied away from them or papered them over with guesswork. There was one occasion (perhaps the only one) on which the “lack of a concept of a bacterium” was recognized and denounced as the “abiding scandal” of bacteriology (45). But, rather than use this insight

to begin a much-needed dialog within the field, the authors concocted a guesswork solution to settle the matter then and there, thereby removing the question/problem from the arena of discourse. Enter the infamous “procaryote.” Not only did this bit of thimblery appear to settle the immediate issue (see below), but it forever changed the course of microbiology. In retrospect the “procaryote” episode (see discussion below) was microbiology’s historical nadir. For the sake of trying to understand what microbiology (bacteriology) is today and where it is (should be) going, we need to go into this strange juncture in the field’s course in some detail. I have come to see the whole unfortunate episode and its outcome as the product of the clash between the classical (home-grown) perception of biology and the fundamentalist reductionism introduced by molecular biology. Bacteriology was effectively shattered by this encounter and did (does) not have the “self-awareness” to pull itself back together—although there is now hope.

### The Dismantling of Bacteriology and a Deconstruction of the Procaryote

One thing that makes this juncture so interesting and important is that it may well have represented a genuine fork in the road for 20th century biology, and the “road not taken” might have led (as mentioned above) to a more inclusive, a more “biological” kind of biology than the harsh molecular reductionist regimen that was actually followed—though we shall never know. The critical period is the decade surrounding 1960. Microbiology’s search for a natural classification of bacteria, the key to bacteriology as an organismal discipline, had clearly reached an impasse; classical approaches to a natural bacterial taxonomy could not crack the problem. Some leading microbiologists had thrown up their hands about a natural classification, their frustration rising to the level of toying with the defeatist notion that bacterial phylogenies are inherently unknowable (44, 50).

This attitude was worlds apart from the one prevailing in the molecular arena. Here technology had come to the point (with Sanger’s development of protein sequencing in the early 1950s [36, 37]) where comparative sequence analysis seemed to offer taxonomy a bright new future, a fact that had not been lost on the molecularist Francis Crick (11): “Biologists should realize that before long we shall have a subject which might be called ‘protein taxonomy’—the study of amino acid sequences of proteins of an organism and the comparison of them between species. It can be argued that these sequences are the most delicate expression possible of the phenotype of an organism and that vast amounts of evolutionary information may be hidden away within them.”

The implications of this for bacteriology were far reaching—a whole new approach to the stalled problem of the natural relationships suddenly became possible. But microbiology was no longer willing to fight the battle. All it now wanted was to leave the past and defeat behind and recast the field in a new, more productive (reductionist) way. Microbiologists were of no mind to hear, much less embrace, Crick’s prescient proclamation.

The crisis came for microbiology in 1962, when the term (and concept) “procaryote” slithered onto the scene (45). The procaryote was invoked in order once and for all to overcome

(actually, obscure) the impasse over bacterial phylogenetic relationships and to provide microbiology with its long-needed “concept of a bacterium” (45). All bacteria, it was asserted, are procaryotes. In other words all shared a basic “procaryotic” organization and, therefore, had come ultimately from a common procaryotic ancestor (45, 46). The fact that all bacteria were of a kind (phylogenetically and structurally) would then serve as the basis for developing the long-sought “concept of a bacterium” in a new and different way, namely, from knowing in detail how procaryotes differed (in structure-function ways) from eucaryotes (45). This meant that the concept of a bacterium could be gained without having to know the natural relationships among bacteria. Consequently, the question of their relationships could be finally dispensed with, or so it seemed.

The official history that accompanied the reintroduction of the “procaryote” was that the “procaryote-eucaryote” dichotomy was actually not new. It was a prescient insight on the part of the protozoologist Edouard Chatton in the 1930s (8, 45). The reasoning was simple: just as nucleated cells represented a monolithic grouping structurally and phylogenetically, nonnucleated cells (bacteria) must also. That surely was simple—a bit too simple. But it made for a neat and appealing dichotomy—so neat and appealing that mid-century microbiologists saw no need to test the monophyletic nature of procaryotes experimentally. Knowing the properties of one or a few representative procaryotes would suffice.

If it wasn't clear at the time, it is more than clear today that this “procaryote” prescription for gaining the critical “concept of a bacterium” doesn't work. Regardless of the fact that there have never been any facts to support the monophyly of the bacteria, a concept of a group of organisms cannot be gained simply by knowing differences between that group and some other (unrelated) organismal group; it requires knowing both differences and similarities within the group. Why, as scientists, biologists then and now (21, 29) accepted the procaryote-eucaryote argument at face value is a mystery. What made this concept so attractive that microbiologists unquestioningly bought it? How firmly did their predecessors believe in the monophyly of the procaryote, and what were their feelings about this one-size-fits-all organization for bacterial cells? What did Chatton actually say about these matters? The history that answers these questions is nothing short of scientifically disconcerting.

Microbiologists had long been aware that the bacteria, which had no visible nucleus and did not undergo mitosis, were very different from the nucleated forms (9, 10). Bacteria were traditionally viewed as more primitive than their nucleated counterparts and as their likely progenitors (4, 9). However, the monophyly of the bacteria was by no means taken for granted. The following quote, from the protozoologist Copeland, speaks the conventional wisdom up to his time (10): “The most profound of all distinctions among organisms is that which separates those without nuclei from those which possess them. The former are the bacteria and bluegreen algae . . . Whether or not life originated more than once, it is certain that life possessing nuclei came into existence once only, by evolution from nonnucleate life.”

The significance of this quote lies as much in what was not said as in what was.

As you can imagine, the matter of the organization of bacterial cells was so ill-defined in those early days that there was little point in making specific suggestions about the subject, especially in proposing that all bacteria had essentially the same basic organization. In 1949 Pringsheim, a prominent bacteriologist and contemporary of Chatton, thoroughly reviewed the literature regarding the relationship of the blue-green algae (*Myxophyceae*) to the bacteria and concluded that although the bulk of the latter (the eubacteria) were not related to the cyanobacteria (*Myxophyceae*), the myxobacteria, which microscopically appeared to be apochlorotic cyanobacteria, might well be (34). What comes through in reading the earlier literature is that throughout the first half of the 20th century microbiologists strongly distinguished the monolithic nucleated forms from the (nonnucleated) bacteria, but the matter of bacterial relationships, be they phylogenetic or organizational, was far below the factual horizon. It was no more feasible to draw conclusions and generalizations about bacterial cellular organization than it was to draw conclusions about their phylogenetic relationships.

Chatton himself seems to have been one of the few ever to use the terms eucaryote and procaryote. (Stanier and van Niel apparently did not use them [or the procaryotic concept]) prior to their 1962 publication [44]. The historian Jan Sapp informs me (personal communication) that the terms procaryote and eucaryote probably first appeared in print in a 1925 article by Chatton (7). Here Chatton's use of “procaryote” is confined to two figures, as a label for the bacteria. The term does not appear in the text. Moreover, Chatton appears not to have used “procaryote” to connote common structural organization or a common ancestry for bacteria, but rather to suggest that schizomycetes (procaryotes) preceded nucleated cells (eucaryotes) in evolutionary sequence and somehow gave rise to them; this is implied in Chatton's Fig. 2 by the positioning of the procaryote grouping immediately beneath the root of the eucaryote phylogenetic tree (7).

So, what are we now to conclude about the “procaryote episode”? The meaning of the term procaryote that appeared in 1962 seems to have no historical justification. In 1962 the term meant that all bacteria shared the “distinctive structural properties associated with the procaryotic cell . . .,” which allowed us “therefore [to] safely infer a common origin for the whole group in the remote evolutionary past. . .” (46). Chatton, on the other hand, appears to have used the term simply to imply that eucaryotes somehow arose from procaryotes. The use of these terms in 1962 then becomes an example of “name expropriation”: a term used in a past scientific context being applied at a later time to a new context in order to give the latter historical justification, the illusion of “tried and true.” Needless to say, the term receives a conceptual makeover in the process.

This entire strange period in microbiology's history can be rationalized as an attempt to bury the old microbiology (along with its past failures) in order to remake the field along more progressive (read reductionist) molecular lines. Unfortunately, the process left microbiologists knowing less about what bacteriology is than before, and the field became the technological playground for other biological disciplines and for medical and related practical concerns.

Things might have been very different had microbiologists

been willing to tolerate a gap in their knowledge (regarding the natural relationships among bacteria) for a decade or so longer, and there was no good reason why this shouldn't have been, given the advances in comparative protein sequence analysis then well under way (as mentioned above). The question of the phylogenetic relationships and the cellular organization of bacteria should have remained as active and alive as the questions of the origin of the chloroplasts and mitochondria were at the time (39). Some microbiologists did indeed feel this way. In his above-mentioned 1949 analysis of the relationship of blue-green algae to bacteria, Pringsheim had ended by basically throwing in the towel. But the scientist in him had added, "Modern methods of extracting specific proteins and other compounds of high molecular weight may eventually afford the clue to the problem [of the natural relationships among bacteria]" (34).

What are we now to do? Obviously, it is not scientifically appropriate (one might even say ethical) to teach the prokaryote concept any more. At the same time, given the ingrained nature of the term prokaryote, it is not useful (not to mention feasible) suddenly to discard it. The way out of this conundrum may be to redefine the term once again. Let prokaryote now mean only cells that are noneucaryotic, with no monophyly implied. It is important that the next generation of biologists understand this and understand why the term's previous connotations are invalid. In this way "prokaryote" can still be used as long as conveniently needed, but it will now imply nothing about relationships or structure (or even evolutionary relationship to eucaryotes). If this strange "prokaryote" period in microbiology's development needs an epitaph that speaks to the future, then the following words from the great physicist Erwin Schrödinger would seem appropriate (42): "In an honest search for knowledge you quite often have to abide by ignorance for an indefinite period. . . . The steadfastness in standing up to [this requirement], nay in appreciating it as a stimulus and a signpost to further quest, is a natural and indispensable disposition in the mind of a scientist."

#### Other Guesswork Solutions?

The prokaryote episode makes one leery: are there other guesswork explanations woven into biology's conventional wisdom that also mask important unanswered questions and so impede progress? We should look particularly at evolution, where conjecture is necessarily the mainstay of defining and understanding issues. Remember, it is not guesswork per se that is anathema; it is guesswork, conjecture, and the like that masquerade as problem-solving, interest-ending fact and so violate scientific norms.

One needs look no further than the "doctrine of common descent" to find a candidate; common descent is something that essentially all modern biologists have taken for granted. Where did this doctrine come from? Why, Darwin, of course: didn't he say that all life stems from a single primordial form? Indeed he did. But look at the context and way in which Darwin addresses the issue in *Origin of Species*. Herein we read (12): ". . . [we may infer] that all the organic beings which have ever lived on this earth may be descended from some one primordial form. But this inference is chiefly grounded on analogy and it is immaterial whether or not it be accepted. No

doubt it is possible, as Mr. G. H. Lewes has urged, that at the first commencement of life many different forms were evolved; but if so we may conclude that only a very few have left modified descendants."

That doesn't sound like doctrine to me! Darwin was merely speculating about ultimate origins—a great gap in our knowledge and something to be defined and resolved when the time came. For Darwin, common descent was an open question, an invitation to discussion. What elevated common descent to doctrinal status almost certainly was the much later discovery of the universality of biochemistry, which was seemingly impossible to explain otherwise (49). But that was before horizontal gene transfer (HGT), which could offer an alternative explanation for the universality of biochemistry, was recognized as a major part of the evolutionary dynamic.

In questioning the doctrine of common descent, one necessarily questions the universal phylogenetic tree. That compelling tree image resides deep in our representation of biology. But the tree is no more than a graphical device; it is not some a priori form that nature imposes upon the evolutionary process. It is not a matter of whether your data are consistent with a tree, but whether tree topology is a useful way to represent your data. Ordinarily it is, of course, but the universal tree is no ordinary tree, and its root no ordinary root (61). Under conditions of extreme HGT, there is no (organismal) "tree." Evolution is basically reticulate.

By now the lesson is obvious: hold classical evolutionary concepts up to the light of reason and modern evidence before weaving an evolutionary tapestry around them. Most of them will turn out to be fluid conjectures that 19th century biologists used to stimulate their thinking, but conjectures that have now, with repetition over time, become chiseled in stone: modern concepts of cellular evolution are effectively petrified versions of 19th century speculations. Evolutionary study today is on a fresh, new molecular footing. This is no time to be shackling our thinking with a collection of refurbished antiques, ideas that automatically make us think in a 19th century mind-set about problems that above all require open minds. I don't feel it helps us to debate these antiquated notions (in modern dress) in the present context.

#### CELLULAR EVOLUTION: THE BUMPY ROAD TO WHO KNOWS WHERE

Approaching evolution of the cell with a clean slate requires establishing a perspective, a framework, and ground rules—not simply for this one problem but for biology in general. Let us begin by recalling David Bohm's prescient quote above and try to imagine a biology released from the intellectual shackles of mechanism, reductionism, and determinism.

A heavy price was paid for molecular biology's obsession with metaphysical reductionism. It stripped the organism from its environment; separated it from its history, from the evolutionary flow; and shredded it into parts to the extent that a sense of the whole—the whole cell, the whole multicellular organism, the biosphere—was effectively gone. Darwin saw biology as a "tangled bank" (12), with all its aspects interconnected. Our task now is to resynthesize biology; put the organism back into its environment; connect it again to its evolutionary past; and let us feel that complex flow that is organism,

evolution, and environment united. The time has come for biology to enter the nonlinear world.

From a theoretical point of view, one thing can be said about evolution with fair assurance: it is a complex, dynamic process. But it is only now, in the context of computer algorithms, fractals, and chaos mathematics, that we are beginning to get a useful feeling for what that means (33, 51), and it means that evolution is a bumpy road to who knows where. "Bumpy" implies that evolution, as a complex dynamic process, will encounter critical points in its course, junctures that result in phase transitions (drastic changes in the character of the system as a whole) (19, 26, 33, 51). "Who knows where" implies that the outcomes of these transitions, saltations, are not predictable a priori. Biologists now need to reformulate their view of evolution to study it in complex dynamic-systems terms.

When one starts looking for major evolutionary saltations, they are not all that hard to identify (48). It is immediately apparent that one of them is the development of language(s). Human language is a development that has set *Homo sapiens* worlds apart from its otherwise very close primate relatives, adding new dimensions to the phase space within which human evolution occurs. Another good critical-point candidate is the advent of (eucaryotic) multicellularity. Here too the saltation is accompanied by a qualitatively new world of possibilities.

Next comes the evolution of the eucaryotic cell itself. While biologists have traditionally seen this as a step (saltation) beyond the stage of bacterial cells, I do not. The idea that eucaryotic cell structure is the product of symbioses among bacteria, and so represents a higher stage than that of the bacterial cell, goes back a good century and a half, but there has been no effort to seriously rethink the matter in the light of modern biological knowledge. Nowhere in thinking about a symbiotic origin of the eucaryotic cell has consideration been given to the fact that the process as envisioned would involve radical change in the designs of the cells involved. You can't just tear cell designs apart and willy-nilly construct a new type of design from the parts. The cells we know are not just loosely coupled arrangements of quasi-independent modules. They are highly, intricately, and precisely integrated networks of entities and interactions. Any dismantling of a cell design would not reverse the evolution that brought it into existence; that is not possible. To think that a new cell design can be created more or less haphazardly from chunks of other modern cell designs is just another fallacy born of a mechanistic, reductionist view of the organism.

But what about the mitochondrion; isn't that a direct counterexample of what has just been said? No, it is not. Evolving the mitochondrion through (endo)symbiosis is fundamentally different from evolving the eucaryotic cell in this way. Whereas the latter process would involve a disruptive dismantling of the preexisting eucaryotic cellular design, acquisition of a mitochondrion does not significantly perturb the eucaryotic cell's basic organization, which is in essence the same with or without the mitochondrion's presence. I take it as a general rule in biology that the more complex, integrated, and specific a cell design becomes, the more intolerant of change that design is. For modern cells, the changes possible in their designs (other than degeneration) are all of a trivial, but not necessarily unimportant, nature. (Granted, the organization of the mitochondrial endosymbiont is radically changed during its evolution,

but that change is a degeneration to a far simpler "cell-like" design, and the mitochondrial design could never evolve back to the level of complexity that its free-living [bacterial] ancestor had.)

In the remote evolutionary past lies the RNA world (18) or, as I call it, the era of nucleic acid life (57), an evolutionary stage whose existence is here taken for granted. The transition that gave rise to this era must have been one of the great evolutionary saltations, as was the transition(s) from that era ultimately to the world of the (proteinaceous) cells as we know them. Somewhere along the line there had to have occurred a saltation that we could call the "coding threshold," where the capacity to represent nucleic acid sequence symbolically in terms of a (colinear) amino acid sequence developed, a development that would generate a truly enormous new, totally unique evolutionary phase space.

What, if anything, do these examples of presumed evolutionary critical points have in common? How might they have come about? All of them, of course, involve the emergence of higher levels of organization, which bring with them qualitatively new properties, properties that are describable in reductionist terms but that are neither predictable nor fully explainable therein. A common thread that links language and multicellularity is communication (interaction at a distance). In each case a complex, sophisticated network of interactions forms the medium within which the new level of organization (entities) comes into existence (3). The advent of translation can be seen similarly (3). Translationally produced proteins, multicellular organisms, and social structures are each the result of, emerge from, fields of interaction when the latter attain a certain degree of complexity and specificity. In the first case, we speak of tRNA "adaptors"; in the second, of morphogenetic fields (17); in the third, of language. Communication, networking, and discrimination are all buried deep in the evolutionary dynamic.

Cells today are complex enough, especially compared to their presumed rudimentary RNA world ancestors, that a number of major critical points may well have occurred in the passage from the presumed simple, primitive aboriginal cellular designs to modern ones. In the existing molecular sequence data, there exists evidence, I claim, for at least one such previously unrecognized critical point. The case rests upon a phenomenon called canonical pattern, which can be seen in sequence comparisons (59, 66): for nearly all of the proteins involved in transcription and translation, the archaeal and bacterial versions of each, although clearly homologous, are remarkably dissimilar. The divergence between the two types borders on the qualitative—far in excess of the degree of divergence seen within either of the two bacterial domains—despite the fact that each domain has had over three billion years of evolution during which its individual lineages have diverged from one another (61). What could this difference in "genre" (between domains) signify other than a period of drastic evolutionary change? Canonical pattern is the molecular "fossil remains" of an evolutionary saltation (59, 66).

## THE DYNAMICS OF CELLULAR EVOLUTION

Two factors strongly influence cellular evolution: HGT and the constraints imposed upon the evolving cell by a primitive

evolving translation apparatus. The quality of HGT is mainly determined by (i) a given gene's functional significance, (ii) the nature of the organismal community within which the recipient organism finds itself (which determines the spectrum of alien genes to which it is exposed), and (iii) the overall organization (design) of the recipient cell. The genes in a genome thus fall into fairly discrete categories depending upon these HGT characteristics. (I will leave out of consideration those genetic elements that have no functional significance to the cell.) One category could be called "cosmopolitan genes." These would be specialty genes, genes that come and go as environmental circumstances change. Cosmopolitan genes are special life style genes; they allow adaptation to unusual environments. Examples are genes conferring antibiotic or heavy metal resistance or any niches that have unusual physical properties, energy sources, and so on. I would venture that some cosmopolitan genes will turn out to be more characteristic of particular environments than they are of particular organismal lineages.

Then there are the genes whose functions are central to general cellular metabolism and so are crucial for the cell's existence under any (natural) condition. For the majority of the main metabolic pathways, alternatives appear to exist, i.e., different enzymes catalyzing the same reaction, different pathways from one compound to another, etc. Under these circumstances one might expect to, and does, find that given enzymes in pathways can be replaced (via HGT) by functional equivalents or alternatives—so long as there is functional continuity throughout the process.

Finally there are the genes that define the organizational fabric of the cell, those that give the cell its basic character. By and large genes of this type are highly and idiosyncratically woven into the cellular fabric. The more integrated a component is, the less likely it becomes that there exists an alien equivalent that fits its design specifications well enough to displace it successfully. Therefore, many genes in this category tend to be fixed in the cellular genome and collectively give the organismal lineage its stable genealogical trace.

### The Key to Understanding the Character of HGT

One cellular system in particular is especially informative regarding HGT, and that is the translation apparatus. The translational componentry is a mix of molecules, some of which are highly refractory to horizontal gene displacement and others of which are relatively susceptible. rRNAs, ribosomal proteins, and elongation factors, for example, are refractory; instances of a ribosomal protein gene transferring from one bacterial taxon to another exist, but these occur rarely. However, no instances in which an archaeal ribosomal protein gene has moved into a bacterial genome (or vice versa) are known. On the other hand, the aminoacyl-tRNA synthetases show quite a few examples of HGT, and a significant (and striking) fraction of these involve transfers from archaea to bacteria (66). This situation raises two important questions: one concerning the ways in which and extent to which HGT has influenced the phylogenies of the aminoacyl-tRNA synthetases and the other concerning why the aminoacyl-tRNA synthetases should be so susceptible to HGT when the ribosomal proteins and elongation factors are not.

With regard to the first question, the phylogenetic pattern for each of the aminoacyl-tRNA synthetases shows evidence for at least some HGT. Despite this, one can see remnants of the universal phylogenetic tree branching—the basic canonical pattern and even further detail—in roughly two-thirds of them. However, in their deviations from the established branching pattern, no two synthetase patterns agree; each exception is idiosyncratic (66). (The reader who wishes further detail should spend some time with reference 66.) The only reasonable interpretation of these data is that the canonical pattern is aboriginal and that it has persisted despite erosion by HGT (66).

Why the aminoacyl-tRNA synthetases are relatively susceptible to HGT (while other members of the translational componentry are not) is possibly the most revealing question of all when it comes to cellular evolution. The answer is simple and lies in the connectedness of the componentry (61). Obviously the ribosomal proteins and the elongation factors interact with the ribosome as a whole in a spectrum of complex, intricate, and well-defined ways, and their interconnections tend to be idiosyncratic, more or less different in different taxa. The aminoacyl-tRNA synthetases are not like this (66). These enzymes interact minimally with others of the cellular componentry. They are functionally self-defining, and their interactions are confined in each case to a small subset of the tRNAs. The aminoacyl-tRNA synthetases are, in effect, modular elements, woven only superficially into the cellular fabric. The universality of the tRNA charging function and the near constancy in shape of tRNA molecules ensure that the aminoacyl-tRNA synthetases that work in one organismal setting will probably work fairly well in many others. Hence, the (modular) aminoacyl-tRNA synthetases are excellent candidates for horizontal gene exchange. The difference in HGT profiles between the aminoacyl-tRNA synthetases and other members of the translational componentry demonstrates that cell design (the manner and extent to which components are fitted into the cellular matrix) is the primary factor in determining the ways in which and the degree to which cellular componentry is subject to HGT.

The lesson is simple and clear: altering cellular design alters the ways in which HGT affects the cell. Yet progressive changes in the design of cells are the essence of cellular evolution. In all likelihood primitive cells were loosely connected conglomerates, in which the connections among the parts were relatively few in number and imprecise in specification, and primitive cellular organization was likely minimal and largely horizontal in nature (60, 64). In other words, the primitive cell is a loose confederation of a relatively small number of rather simple modules. For cells of this type, most if not all cellular componentry would be open to HGT, making the combinatorics of gene transfer far and away the major factor in early cellular evolution.

Is there any support for this notion of primitive cellular organization? Yes, although none of it is direct. I would claim that the notion is inherent in the way the translation mechanism must have evolved. Consider the following argument: modern translation mechanisms are complex, tightly and precisely coupled aggregates of many components (on the order of 100). It is not reasonable to expect early primitive versions of this mechanism to have had anywhere near that complexity,

anywhere near the functional precision of the modern mechanism (60, 64). A sufficiently imprecise translation mechanism would strongly limit the general types of proteins that could evolve (64). Errors in amino acid-codon matching and in reading frame maintenance would prevent the evolution of the long protein chains we see today. But it is precisely these large proteins that are crucial to basic cellular functions today. It is today's small proteins, the ribosomal proteins, the cytochromes, etc., proteins that are ubiquitous in cellular functioning, that probably most resemble primitive proteins.

Also, a primitive translation apparatus could have produced a type of protein that would be undesirable today but may have been far from that in the past (see below). A sufficiently imprecise translation mechanism could produce "statistical proteins," proteins whose sequences are only approximate translations of their respective genes (54). While any individual protein of this kind is only a highly imprecise translation of the underlying gene, a consensus sequence for the various imprecise translations of that gene would closely approximate an exact translation of it.

Constraints on the length and general composition of primitive proteins will affect all aspects of the primitive cell, not only individual specific functions but also the cell's overall character (64). The aboriginal processes of DNA replication and transcription could not be as complex and, so, as precise as are their modern equivalents because both of these mechanisms today are dependent upon large proteins (60, 63). Imprecise primitive genome replication implies that primitive genomes could comprise relatively few (unique) genes (64). This in turn argues for simplicity of primitive cell designs and a general looseness and imprecision in those designs (64).

To summarize: a primitive, loosely connected, and highly modular primitive cellular organization would be subject to rampant HGT. For such cells, many highly novel functions could be readily introduced without disrupting the loose, ill-defined, and permissive cellular organizations to their breaking points. By the same token, existing componentry could be relatively easily lost or displaced by something only roughly equivalent (in shape or function). The primitive counterparts of some of today's specific enzymes may have been only reaction class specific. Cellular entities of this kind would not have stable genealogical records; this had to be a period of ephemeral organismal genealogies. The world of primitive cells feels like a vast sea, or field, of cosmopolitan genes flowing into and out of the evolving cellular (and other) entities. Because of the high levels of HGT, evolution at this stage would in essence be communal, not individual (63). The community of primitive evolving biological entities as a whole as well as the surrounding field of cosmopolitan genes participates in a collective reticulate evolution.

Although we can infer essentially nothing about the hypothetical primitive entities under discussion, it is nevertheless worthwhile to consider their possible relationships to one another. Were they communal relationships only in an abstract sense, a virtual community defined only by gene transfers, or did they form actual physically structured groupings, perhaps resembling modern bacterial consortia but even more diverse in makeup and modes of interaction? Some time ago I said that what I now call the pre-Darwinian era (63) "may be more a world of semiautonomous subcellular entities that somehow

group to give 'loose' (ill-defined) cellular forms" (58). The panoply of interactions that such an image evokes (interactions that go far beyond HGT alone) is strongly suggestive of physical communal organization, one not only of "cells" but of a spectrum of biological entities, many of them not self-replicating in their own right (and not all on paths to become "modern" cells).

### From There to Here

Now, how are such primitive cellular entities, these loose confederations of simple modular elements, turned into the much more complex cells of today? Talking specifics is not feasible at this point, but the general character of the evolutionary course is self-evident. The overall thrust of early evolution is toward greater organization, complexity that leads to finer discrimination, to increased coordination, and to biological specificity in general. Key to this transition is an increase in the connectivity of the parts, leading to a more complex and integrated network of interactions.

As its connectivity increases, a complex dynamic system tends to encounter critical points, points where the system undergoes phase transitions, in which its overall nature changes dramatically (26). I do not think that biologists can avoid the conclusion that during the evolution of (modern) cellular organization, such phase transitions have occurred. In particular, I assert that it was one such transition that took the cell out of its initial primitive state in which HGT dominated the evolutionary dynamic (and evolving cells had no stable genealogical records and evolution was communal) to a more advanced (modern) form (where vertical inheritance came to dominate and stable organismal lineages could exist). The obvious choice of a name for this particular evolutionary juncture would be Darwinian threshold or Darwinian transition, for it would be only after such a saltation had occurred that we could meaningfully speak of species and of lineages as we know them (63).

Three questions are central to understanding cellular evolution: (i) when (under what circumstances) did the evolution of (proteinaceous) cells begin, (ii) how was the incredible novelty needed to create these first proteinaceous cells generated, and (iii) did all extant cellular life ultimately arise from one or from more than one common ancestor? The second of these questions, how the overwhelming amount of novelty needed to bring modern cells into existence was generated, is the central and most challenging question of the three. This is a kind of novelty that we would not encounter in the modern biological era, and it had to have been generated in a kind of way that we have yet to fathom.

Arguably there has to have been a very definite (and so recognizable) stage at which the evolution of modern cells began. The transition was too drastic, too profound not to have somehow left its mark. It seems highly likely that the stage in question was the onset of translation, the emergence of the capacity to represent nucleic acid sequence (colinearly) in an amino acid language (as mentioned above). Hence, the onset of cellular evolution is likely to have occurred in an RNA world context. Over the last several decades biologists have become increasingly aware that translation is defined by its RNA componentry, and so the idea that the aboriginal mechanism was

an RNA-based device has become increasingly attractive (6, 56, 62). (Note that I take the RNA world [era of nucleic acid life] to be a period before proteins were translationally produced, when, regardless of whatever else existed, nucleic acids capable of complementary [templating] replication existed and were the drivers of evolution. In this view, peptides could have existed, although, by definition, they could not have been generated by a translation process [57].)

I take the RNA world to have been typically biological in two ways: (i) organized entities that were probably encapsulated and analogous to modern cells existed, but these were entities whose organizations centered about their nucleic acid componentry, not protein, and (ii) considerable diversity had evolved among these (nucleic acid-based) entities. One can't say whether the initial products of translation were largely of no value to the existing nucleic acid-based entities or for unknown reasons played significant roles right from the start. Considerations possibly bearing on this include whether the genetic code reflected preexisting specific interactions between nucleic acid and amino acids or proteins; what the relationship, if any, between translationally produced peptides and any preexisting nontranslationally produced peptides was; and what the general nature of the role(s) played by the first translationally produced peptides was.

One thing, at least, seems likely: horizontal gene flow, which probably predated the first translation system, was essential to evolving the protein-based cellular organization from its onset. It is also likely that the genetic code has remained in effect universal because it is the lingua franca of genetic commerce (61, 63). It is even reasonable to see the code originating as a lingua franca, being the product of, and belonging to, the community from the start.

The creation of the enormous amount and degree of novelty needed to bring forth modern cells is by no means a matter of waving the usual wand of variation and selection. What was there, what proteins were there to vary in the beginning? Did all proteins evolve from one aboriginal protein to begin with? Hardly likely! (Evolution's rule, to which there are, fortunately, a few exceptions, is that "you can't get there from here.") Our experience with variation and selection in the modern context does not begin to prepare us for understanding what happened when cellular evolution was in its very early, rough-and-tumble phase(s) of spewing forth novelty.

It is useful to try to envision the problem in phase space terms, in terms of an evolutionary process that wanders through an enormous space of possible protein sequences. (The nature of such an evolutionary phase space is a deep philosophical issue, but that shouldn't prevent our using the concept superficially, as an aid to discussion.) In a phase space framework the problem of generating the novelty needed to evolve the cell becomes one of finding an optimal strategy for searching the phase space, a strategy that pulls as much novelty out of it as possible. Novelty, of course, doesn't exist in a vacuum; it has to have selective value in some environment. For this reason I see no way out of the conclusion that cellular evolution began in a highly multiplex fashion, from many initial independent ancestral starting points, not just a single one. Such a strategy automatically optimizes both the amount and diversity of novelty generated because it generates a great variety of selective contexts. Needless to say, such a multiplex

evolutionary strategy requires that the various evolving "foci" spreading into the phase space be linked, that the novelty generated "over here" can end up (and be tested) "over there": another argument for HGT.

### An Interesting, if Not Relevant, Aside

In the context of search strategies for probing evolutionary phase space, let us revisit the matter (discussed above) of statistical proteins (54). In a modern biological context one is hard pressed to think of a *raison d'être* for proteins that are imprecise translations of a gene—but not in a primitive context. Statistical proteins form the basis of a powerful strategy for searching protein phase space, finding novel proteins. Today, evolution explores the space of possible protein sequences in effect by mapping one point in nucleic acid space to a corresponding point in protein space (ignoring codon degeneracy) and testing the match by selection. This is like shooting a rifle bullet at a target. The probability of hitting the target would be much higher, however, were a shotgun used instead. Statistical proteins in effect allow evolution to use a "shotgun" strategy: only one of the many sequences (pellets) that make up a statistical protein need "hit the target" for the underlying gene to have selective value. At that point the evolving cell can employ a "variation on existing themes" (local variation/selection) approach to optimizing the underlying gene sequence ("center" it on the target)—and as a by-product of the strategy the surrounding phase space is automatically explored for proteins of similar sequence, which means that evolving one protein function could end up evolving a family of related functions (or even parts of an enzymatic pathway). (Put another way, statistical proteins provide a strategy for shrinking the effective phase space, from an enormous collection of points to a smaller [but still large] collection of "locales" [of related points].)

Perhaps you have noticed that in a formal sense the antibody system is equivalent to a statistical protein(s), but in the former case, sequence diversity is generated at the genetic, not the translational, level. Nevertheless, the antibody system stands as a concrete example of a kind of role statistical proteins could have played early on in finding novelty—a role that could well have been essential at the inception of cellular evolution.

Are there any except theoretical considerations to suggest a multiplex evolution of proteinaceous cells? One thing we do know is that the possibility is no longer ruled out (as a doctrine of common descent would demand). I think the RNA world (although hypothetical) makes a compelling case for multiplexing: with the onset of translation in an RNA world setting, preexisting nucleic acid-based entities can (and will) become bedecked with proteins. In this way, many different entities in the RNA world could serve as starting points for the evolution of proteinaceous (RNP) entities, and HGT and other modes of communication would serve to unite them into coevolving communities. Multiplex origin of cellular organization is a question that definitely deserves serious consideration.

### When Is a Tree Not a Tree?

The universal phylogenetic tree converges to what is conventionally interpreted as a root, a common ancestor locus.

What is this convergence and how does it relate to the pre-Darwinian era in cellular evolution and the Darwinian threshold? Analyses of genomic data have shown that many cellular functions were probably well developed before the stage represented by this so-called root had been reached. Moreover, “prehistoric” gene families also existed, for example, the translation elongation factor family and the aminoacyl-tRNA synthetase families (66). While these examples involve universal gene families, nonuniversal “prehistoric” gene families were there as well, with the eucaryotic tubulin subunit family (23), the “archaeal histone” family (35), or various members of the families in the archaeal signature (20) being examples. It would appear that the three major cell designs may have each taken on some kind of characteristic primitive form well before the stage of the root of the universal tree was reached. What, then, is this “root”?

The root of the universal tree is an artifact resulting from forcing the evolutionary course into tree representation when that representation is inappropriate (60). In the pre-Darwinian era the evolutionary course cannot be represented by an organismal tree topology. It is only after a more advanced stage in cellular evolution has been reached that tree representation begins to become useful. That stage is the Darwinian threshold, the critical point before which HGT dominates the evolutionary dynamic and after which it does not—thus allowing stable organismal genealogies to emerge (63). Only then can living systems finally be conceptualized in discreet, idiosyncratic species terms. Note the phrase “begins to become” above: if only one of the major evolving cell designs were to cross its Darwinian threshold, tree representation would appear to be appropriate because that one lineage (only) would be distinguishable from all the rest, despite the fact that the others did not yet exist as discrete stable lineages, having not yet undergone Darwinian transitions of their own. What tree representation does at this stage is effectively to lump these others by exclusion into a common “negative branch,” which is how tree topology must represent an “A versus ~A” distinction. The result is an apparent bifurcation from an apparent root point (63).

There is no reason to expect the three primary cell designs all to have crossed their Darwinian thresholds simultaneously (58). Indeed, because each is a unique design, there is every reason to expect the opposite. The existing universal tree topology demands that the bacterial design be the first to reach its Darwinian threshold, leaving the archaeal and eucaryotic designs still in their pre-Darwinian, prehistoric, condition (along with any other prehistoric designs, cellular or otherwise, that might then have existed).

The question then becomes which of the two remaining cell designs was next to cross its threshold. Although customary phylogenetic analysis cannot provide the answer, I would argue for the archaeal design being the next one—simply on the grounds that of the many features that the archaea and eucaryotes specifically share, the archaeal versions tend to be the simpler and, so, possibly closer to older “less evolved” (more cosmopolitan) “ancestral” forms (63). However, even with the emergence of the archaeal design as a discrete lineage, the tree representation is only apparently applicable in full, for the eucaryotic design remains in its pre-Darwinian condition and its “lineage” is still defined by exclusion. Therefore, what ap-

pears in the tree representation to be a common ancestral trunk shared by the archaea and eucaryotes does not actually exist. The order in which the three cell designs crossed their respective Darwinian thresholds is, then, the bacterial first, the archaeal second, and finally the eucaryotic (an order that was first suggested by Otto Kandler [25]).

Keep in mind that crossing a Darwinian threshold does not mean that HGT is eliminated. It is postulated merely to diminish (dramatically) in scope and frequency (63, 66). The regression of HGT will continue as each primary lineage consolidates and begins to spawn its own major sublineages—until HGT and cellular organization reach the levels at which they exist today. This conjecture is testable in that it predicts HGT to occur in a more intense form while the major (early) sublines within a given domain are developing than later on, when these major sublines within each domain are in turn spawning their own daughter lineages (66). This effect should be detectable in a distribution of horizontal transfer events (for example, from the *Archaea* to the *Bacteria*) scaled according to taxonomic rank; the higher the taxonomic rank, the more likely the taxon is to have incorporated phylogenetically distant (alien) genes. The phylogenetic distribution of the several kinds of bacterial chlorophyll-based photosynthesis (which appears to have involved phylum rank HGT) seems to support this notion (59).

## ONE LAST LOOK

Enough time has elapsed that we can begin to look back at 20th century biology with some perspective and, so, see the molecular era for what it is in the larger picture. The 19th was biology’s defining century. There, for the first time, biology’s great problems lay scientifically outlined and assembled, with all of them effectively in early stages of development. Nineteenth century biology was a potpourri of problems in that some (like the natures of the gene and of the cell) cried out for dissection, analysis in terms of their parts, whereas others (such as evolution and morphogenesis and the significance of biological form in general) were holistic, metaphysically challenging, not fundamentally understandable as collections of parts.

The 19th century as a whole had a reductionistic world view, if for no other reason than because of the outlook of classical physics. Physics at that time saw a fundamentally reductionistic world, in which ultimate explanation lay completely in the properties and interactions of atoms: to know the positions and momenta of all of the fundamental particles at a given point in time was in principle to know their positions and momenta at any other point in time, past or future. Nothing added, nothing subtracted; just the endless deterministic jumble of bouncing atomic balls in a directionless time (33). Biologists of the 19th century were no exception to the reductionist zeitgeist, but theirs tended to be an empirical, analytical reductionism, not a metaphysical one: one would be hard put to explain evolution and the problem of biological form in reductionist terms alone.

Given the temper of the times, the entry of chemistry and physics into biology was inevitable. The technology that these sciences would introduce was not only welcome but very much needed. Also, biology was now well enough scientifically understood that it began to appeal to physicists. But the physics

and chemistry that entered biology (especially the former) was a Trojan horse, something that would ultimately conquer biology from within and remake it in its own image. Biology would be totally fissioned, and its holistic side would be quashed. Biology would quickly become a science of lesser importance, for it had nothing fundamental to tell us about the world. Physics provided the ultimate explanations. Biology, as no more than complicated chemistry, was at the end of the line, merely providing baroque ornamentation on the great edifice of understanding that was physics—the hierarchy physics→chemistry→biology is burned into the thinking of all scientists, a pecking order that has done much to foster in society the (mistaken) notion that biology is only an applied science.

In the last several decades we have seen the molecular reductionist reformulation of biology grind to a halt, its vision of the future spent, leaving us with only a gigantic whirring biotechnology machine. Biology today is little more than an engineering discipline. Thus, biology is at the point where it must choose between two paths: either continue on its current track, in which case it will become mired in the present, in application, or break free of reductionist hegemony, reintegrate itself, and press forward once more as a fundamental science. The latter course means an emphasis on holistic, “nonlinear,” emergent biology—with understanding evolution and the nature of biological form as the primary, defining goals of a new biology.

Society cannot tolerate a biology whose metaphysical base is outmoded and misleading: the society desperately needs to live in harmony with the rest of the living world, not with a biology that is a distorted and incomplete reflection of that world. Because it has been taught to accept the above hierarchy of the sciences, society today perceives biology as here to solve its problems, to change the living world. Society needs to appreciate that the real relationship between biology and the physical sciences is not hierarchical, but reciprocal: physics↔biology. Both physics and biology are primary windows on the world; they see the same gem but different facets thereof (and so inform one another). Knowing this, society will come to see that biology is here to understand the world, not primarily to change it. Biology's primary job is to teach us. In that realization lies our hope of learning to live in harmony with our planet.

#### ACKNOWLEDGMENTS

My work is supported by grants from the National Aeronautics and Space Administration and the Department of Energy.

#### REFERENCES

- Ambler, R. P., M. Daniel, J. Hermoso, T. E. Meyer, R. G. Bartsch, and M. D. Kamen. 1979. Cytochrome  $c_2$  sequence variation among the recognised species of purple nonsulphur photosynthetic bacteria. *Nature* **278**:659–660.
- Ambler, R. P., T. E. Meyer, and M. D. Kamen, M. D. 1979. Anomalies in amino acid sequence of small cytochromes  $c$  and cytochromes  $c'$  from two species of purple photosynthetic bacteria. *Nature* **278**:661–662.
- Barbieri, M. 2003. The organic codes: an introduction to semantic biology. Cambridge University Press, Cambridge, United Kingdom.
- Beijerinck, M. W. 1905. *Versl. Kon. Akad. Wetensch. (Amsterdam)* **14**: 168–169.
- Bohm, D. 1969. Some remarks on the notion of order, p. 18–40. *In* C. H. Waddington (ed.), *Towards a theoretical biology*, vol. 2. Sketches. Edinburgh Press, Edinburgh, United Kingdom.
- Cech, T. 2000. The ribosome is a ribozyme. *Science* **289**:878.
- Chatton, E. 1925. *Pansporella perplexa*. Reflexions sur la biologie et la phylogenie des protozoaires. *Annales des Sciences Naturelles* 10e, Series vii, p. 1–84.
- Chatton, E. 1938. *Titres et travaux scientifiques (1906–1937)*. Sottano, Sète, France.
- Cohn, F. 1875. *Untersuchungen über Bacterien II. Beitr. Biol. Pfl.* 1 Heft **3**:141–208.
- Copeland, H. F. 1947. Progress report on basic classification. *Am. Nat.* **81**:340–350.
- Crick, F. H. C. 1958. The biological replication of macromolecules. *Symp. Soc. Exp. Biol.* **12**:138–163.
- Darwin, C. 1859. *On the origin of species*. [Reprint, Modern Library, New York, N.Y., 1998.]
- Fox, G. E., K. R. Pechman, and C. R. Woese. 1977. Comparative cataloging of 16S ribosomal RNA: molecular approach to procaryotic systematics. *Int. J. Syst. Bacteriol.* **27**:44–57.
- Fox, G. E., L. J. Magrum, W. E. Balch, R. S. Wolfe, and C. R. Woese. 1977. Classification of methanogenic bacteria by 16S ribosomal RNA characterization. *Proc. Natl. Acad. Sci. USA* **74**:4537–4541.
- Fox, G. E., E. Stackebrandt, R. B. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. Tanner, L. J. Magrum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen, and C. R. Woese. 1980. The phylogeny of prokaryotes. *Science* **209**:457–463.
- Gamow, G. 1954. Possible relations between deoxyribonucleic acid and protein structures. *Nature* **173**:318.
- Gilbert, S. G., J. M. Opitz, and R. A. Raff. 1996. Resynthesizing evolutionary and developmental biology. *Dev. Biol.* **173**:357–372.
- Gilbert, W. 1986. The RNA world. *Nature* **319**:618.
- Goodwin, B. C. 2000. The life of form. Emergent patterns of morphological transformation. *C. R. Acad. Sci. Ser. III* **323**:15–21.
- Graham, D. E., R. Overbeek, G. J. Olsen, and C. R. Woese. 2000. An archaeal genomic signature. *Proc. Natl. Acad. Sci. USA* **97**:3304–3308.
- Harold, F. M. 2001. *The way of the cell*. Oxford University Press, Oxford, United Kingdom.
- Hoagland, M. 1960. The relationship of nucleic acid and protein synthesis as revealed by studies in cell-free systems, p. 349–408. *In* E. Chargaff and J. Davidson (ed.), *The nucleic acids*, vol. III. Academic Press Inc., New York, N.Y.
- Jenkins, C., R. Samudrala, I. Anderson, B. P. Hedlund, G. Petroni, N. Michailova, N. Pinel, R. Overbeek, G. Rosati, and J. T. Staley. 2002. Genes for the cytoskeletal protein tubulin in the bacterial genus *Prostheco bacter*. *Proc. Natl. Acad. Sci. USA* **99**:17049–17054.
- Judson, H. F. 1996. *The eighth day of creation: makers of the revolution in biology*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.
- Kandler, O. 1994. Cell wall biochemistry in Archaea and its phylogenetic implications. *J. Biol. Phys.* **20**:165–169.
- Kauffman, S. 1995. *At home in the universe*. Oxford University Press, Oxford, United Kingdom.
- Kuhn, T. S. 1996. *The structure of scientific revolutions*. University of Chicago Press, Chicago, Ill.
- Magrum, L. J., K. R. Luehrsen, and C. R. Woese. 1978. Are the extreme halophiles actually “bacteria”? *J. Mol. Evol.* **11**:1–8.
- Murray, R. G. E. 1974. A place for bacteria in the living world, p. 4–9. *In* R. E. Buchanan and N. E. Gibbons (ed.), *Bergey's manual of determinative bacteriology*, 8th ed. Williams & Wilkins, Baltimore, Md.
- Nirenberg, M., P. Leder, M. Bernfield, R. Brimacombe, J. Trupin, F. Rottman, F., and C. O'Neal. 1965. RNA codewords and protein synthesis. VII. On the general nature of the RNA code. *Proc. Natl. Acad. Sci. USA* **53**: 1161–1168.
- Olsen, G. J., N. R. Pace, M. Nuell, B. P. Kaine, R. Gupta, and C. R. Woese. 1985. Sequence of the 16S rRNA gene from the thermoacidophilic archaeobacterium *Sulfolobus solfataricus* and its evolutionary implications. *J. Mol. Evol.* **22**:301–307.
- Pauling, L., and M. Delbruck. 1940. The nature of the intermolecular forces operative in biological processes. *Science* **92**:77–79.
- Prigogine, I. 1996. The end of certainty: time, chaos and the new laws of nature. Simon and Schuster, Inc. New York, N.Y.
- Pringsheim, E. G. 1949. The relationship between Bacteria and Myxophyceae. *Bacteriol. Rev.* **13**:47–98.
- Reeve, J. N., K. Sandman, and C. J. Daniels. 1997. Archaeal histones, nucleosomes, and transcription initiation. *Cell* **89**:999–1002.
- Sanger, F., and H. Tuppy. 1951. The amino-acid sequence in the phenylalanyl chain of insulin. *Biochem. J.* **49**:481–490.
- Sanger, F., and E. O. P. Thompson. 1953. The amino-acid sequence in the glycyl chain of insulin. *Biochem. J.* **53**:353–374.
- Sanger, F., G. G. Brownlee, and B. G. Barrell, B. G. 1965. A two-dimensional fractionation procedure for radioactive nucleotides. *J. Mol. Biol.* **13**:373–398.
- Sapp, J. 1994. *Evolution by association: a history of symbiosis*. Oxford University Press, Oxford, United Kingdom.
- Sapp, J. 2003. *Genesis: the evolution of biology*. Oxford University Press, Oxford, United Kingdom.
- Schrödinger, E. 1944. *What is life*. Cambridge University Press, Cambridge, United Kingdom.

42. **Schrödinger, E.** 1954. *Nature and the Greeks*. Cambridge University Press, Cambridge, United Kingdom.
43. **Soll, D., E. Ohtsuka, D. S. Jones, R. Lohrmann, H. Hayatsu, S. Nishimura, and H. G. Khorana.** 1965. Studies on polynucleotides. XLIX. Stimulation of the binding of aminoacyl-sRNAs to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. *Proc. Natl. Acad. Sci. USA* **54**:1378–1385.
44. **Stanier, R. Y., M. Doudoroff, and E. A. Adelberg.** 1957. *The microbial world*, 1st ed. Prentice-Hall Inc., Engelwood Cliffs, N.J.
45. **Stanier, R. Y., and C. B. van Niel.** 1962. The concept of a bacterium. *Arch. Mikrobiol.* **42**:17–35.
46. **Stanier, R. Y., M. Doudoroff, and E. A. Adelberg.** 1963. *The microbial world*, 2nd ed. Prentice-Hall Inc., Engelwood Cliffs, N.J.
47. **Stent, G. S.** 1968. That was the molecular biology that was. *Science* **160**:390–395.
48. **Szathmary, E., and J. Maynard Smith.** 1997. From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* **187**:555–571.
49. **van Niel, C. B.** 1949. The “Delft school” and the rise of general microbiology. *Bacteriol. Rev.* **13**:161–174.
50. **van Niel, C. B.** 1955. A century of progress in the natural sciences 1853–1953, p. 89–114. California Academy of Sciences, San Francisco.
51. **Waldrop, M. M.** 1992. Complexity: the emerging science at the edge of order and chaos. Simon and Schuster, New York, N.Y.
52. **Watson, J. D., and F. H. C. Crick.** 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature* **171**:964–966.
53. **Whitehead, A. N.** 1929. *Process and reality*. Collier-Macmillan Canada, Ltd., Toronto, Canada.
54. **Woese, C. R.** 1965. On the evolution of the genetic code. *Proc. Natl. Acad. Sci. USA* **54**:1546–1552.
55. **Woese, C. R.** 1967. *The genetic code: the molecular basis for genetic expression*. Harper & Row, New York, N.Y.
56. **Woese, C. R.** 1970. Molecular mechanics of translation: a reciprocating ratchet mechanism. *Nature* **226**:817–820.
57. **Woese, C. R.** 1972. The emergence of genetic organization, p. 301–341. *In* C. Ponnampereuma (ed.), *Exobiology*. North-Holland Publishing Co., Amsterdam, The Netherlands.
58. **Woese, C. R.** 1982. Archaeobacteria and cellular origins: an overview. *Zentbl. Bakteriol. Mikrobiol. Hyg. Abt. 1 Orig. C* **3**:1–17.
59. **Woese, C. R.** 1987. Bacterial evolution. *Microbiol. Rev.* **51**:221–271.
60. **Woese, C. R.** 1998. The universal ancestor. *Proc. Natl. Acad. Sci. USA* **95**:6854–6859.
61. **Woese, C. R.** 2000. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. USA* **97**:8392–8396.
62. **Woese, C. R.** 2001. Translation: in retrospect and prospect. *RNA* **7**:1055–1067.
63. **Woese, C. R.** 2002. On the evolution of cells. *Proc. Natl. Acad. Sci. USA* **99**:8742–8747.
64. **Woese, C. R., and G. E. Fox.** 1977. The concept of cellular evolution. *J. Mol. Evol.* **10**:1–6.
65. **Woese, C. R., and G. E. Fox.** 1977. The phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA* **74**:5088–5090.
66. **Woese, C. R., G. J. Olsen, M. Ibba, and D. Soll.** 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**:202–236.