

## THE COMPLETENESS OF MOLECULAR BIOLOGY

HAROLD J. MOROWITZ

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

*Isr J Med Sci* 20: 750-753, 1984**Key words:** mycoplasma genetics; genome size; minimal cell

In the late 1950s a group associated with my laboratory began to focus on the question: What is the smallest autonomous self-replicating entity? The adjective autonomous excluded the viruses and *Chlamydia*, since the thrust of the question would not have been met if we were forced to study the complexity of the larger host cell. The first search took place by mail with a series of letters to microbiologists around the world asking for the organism of choice. The reason for our quest was to find a paradigm organism to synthesize all the newly emerging knowledge in the field of molecular biology. Just as the hydrogen atom, the smallest and simplest member of the periodic table, had served to sharpen many of the fundamental questions of spectroscopy and quantum mechanics, so, we reasoned, would a minimum biological system play an analogous role. The smallest system was being sought because it would probably be the simplest. With the emergence of molecular biology, the object of the search shifted from the smallest organism to that with the smallest genome.

Both from the criteria of size and genome size, it rapidly became clear that *Mycoplasma*, then known as pleuropneumonia-like organisms, PPLO, were the cells of choice. The reports on morphology and mode of replication were sufficiently confusing that we asked the naive question: Were we indeed dealing with a cellular system? The criterion of a cell in this sense is a closed vesicle of low conductivity and low dielectric constant material surrounding a highly conducting high dielectric constant core and immersed in a conducting matrix. Dielectric dispersion experiments had been the ideal method of getting this kind of informa-

tion ever since such work in the 1920s established the existence and nature of the plasma membrane of human erythrocytes. Extensive dielectric measurements were made on *Mycoplasma gallisepticum* A5969. They showed the existence of normal, although very small, cells surrounded by a membrane of ordinary dielectric properties.

Once we were convinced of the existence of conventional cells, the next problem was to characterize them and develop techniques of determining the genome size. At first the focus was on the amount of DNA per cell, but it was subsequently necessary to develop more sophisticated methods. Many of the early experiments on the sizing of cells by filtration were motivated by the relationship between cell size and atomicity, an important early consideration and one that still has fundamental relevance in the theory of cell function. Erwin Schrödinger in his very perceptive book, "What Is Life?," published 40 years ago, raised the question of how cells function in the presence of thermal noise. That question is still relevant, and the smaller the cell, the fewer the number of copies of each element of cellular hardware, the more difficult it becomes to understand life in terms of statistical physics.

In the early days of PPLO studies, reproductive bodies were reported with diameters in the range of 0.1 to 0.3 microns. At present, we would regard the upper end of that range as the lower limit to the size of viable cells. To grasp the atomicity problem, note that the number of atoms in a cell of radius  $r$  is given by:

$$n = \frac{4}{3} \pi r^3 \frac{\rho N}{M}$$

where  $N$  is Avogadro's number,  $\rho$  the density, and  $M$  the average atomic weight, a number approximately 6 or 7. Thus we find:

	Cell radius in microns	Atoms/cell
	1.0	$4.4 \times 10^{11}$
	0.5	$5.5 \times 10^{10}$
Smallest <i>Mycoplasma</i>	0.3	$1.2 \times 10^{10}$
	0.15	$1.5 \times 10^9$
	0.10	$4.4 \times 10^8$
	0.05	$5.5 \times 10^7$

These simple calculations and others of a similar nature convinced us that we were near a theoretical limit of cell size, since the number of atoms was beginning to limit the hardware available to carry out the various functions of the cell that were becoming known as the findings of molecular biology began to unfold. For example, for the nonaqueous portion of a 0.3-micron-diameter cell:

$$1.5 \times 10^8 \text{ atoms} \cong 1.5 \times 10^7 \text{ monomers} \cong 3 \times 10^4 \text{ polymers.}$$

Atomicity can be approached in other ways; for example, a 0.3-micron-diameter cell is 2,000 atomic diameters across and contains, at pH 7, an average of less than one hydrogen ion.

As work proceeded, along with the contemporaneous development of the dogma of molecular biology, we focused more directly on the genome size. A long, laborious series of investigations finally led to the elegant study by Hans Bode using *M. hominis* and the Kleinschmidt technique to dissect out and photograph the entire intact genome by electron microscopy. This led to a genome size of  $5.1 \times 10^8$  Da. That number and its subsequent confirmation and generalization to other species is still perhaps the most distinctive way for grouping the genus *Mycoplasma* and stressing the importance of this taxon to molecular biology as the minimum living system.

If one looks at the genome sizes of the eubacteria and archaeobacteria, they range from a minimum of  $10^9$  up to several times  $10^9$ . Among the wall-less cells, *Acholeplasma*, *Spiroplasma* and *Thermoplasma* are clustered around  $10^9$ . There is then a

gap, as all the *Mycoplasma* cluster around  $5 \times 10^8$ . This result seems most important, and every effort should be made to more precisely determine the genome sizes of the *Mollicutes*.

Daltons of DNA convert to bits of information, which for the *Mycoplasma* calculates out to be about  $1.6 \times 10^5$  bits. Alternatively, this much DNA will code for about 600 proteins—which suggests that the logic of life can be written in 600 steps. Completely understanding the operations of a prokaryotic cell is a visualizable concept, one that is within the range of the possible.

To make the preceding more tangible, I would like to very roughly outline a research plan to explore the logic of life. First, it is necessary to work with an organism that has a small-sized genome and grows in a completely defined medium.

There exists such a medium for *M. mycoides* subsp. *mycoides*, which supported a growth rate and yield similar to those obtainable in the best nondefined media. The major problem in developing defined media appears to involve lipid toxicity, and synthetic high polymer lipid adsorbers provide a possible solution to that problem. Given a mycoplasma and a defined medium, a metabolic map could be established. This is not a simple task, but it is a possible one. The genome could be sequenced. If one used a species of very low guanine plus cytosine (G+C) content (*M. mycoides* also fits this description), then it should be possible to uniquely read from the code to the amino acid sequence. A thorough study of the ultrastructure and the transport systems of the organism could also be undertaken. The next step is to assign coding space to structures and functions until the entire genome is assigned. What is being described is a very large amount of work, but involves no conceptual difficulties. For example, the Epstein-Barr virus of 172 kbp has been sequenced; the mycoplasmas we are discussing have a genome size somewhere around 700 kbp.

Since the amino acid sequence can be uniquely read out from the genome, it is not necessary to determine a complete amino acid sequence in order to map a protein. In general, the first seven or eight N terminal amino acids will locate the protein on the genome, and the remainder can be read out.

If all the preceding operations were carried out we would have: 1) a complete metabolic map indicating all the synthetic, degradative, and energy-yielding pathways; the enzymes can all be

located on the genome; the nonenzymatic control loops must be established, but here we have considerable guidance from general biochemistry; 2) a mapping of ribosomal RNA, ribosomal protein, transfer RNA and transfer enzyme sites on the genome; 3) a catalogue of transport proteins and their mapping on the genome; 4) chemomechanical systems involved in motility or cell division; and 5) genome regions for which functions must be sought.

At 600 steps, a computer model is feasible, and every experiment that can be carried out in the laboratory can also be carried out on the computer. The extent to which these match, measures the completeness of the paradigm of molecular biology. Matching the model to the cell will clearly generate more experiments, but the task is nonetheless finite. We can, as in any experimental domain, anticipate surprises, but the informational domain limits their magnitude.

To date, our theoretical efforts have been directed to the question of the smallest cell we can design using the conventional results of molecular biology. Such a hypothetical cell would require about  $2.5 \times 10^8$  Da to encode all of the necessary hardware. The gap between the theoretical minimum and the actual minimum is a factor of two.

The picture so far seems to understress those parts of the cellular hardware concerned with the actual process of cell division or separation of replicated DNA. The present view is consistent with a cell which is in the old biochemical nomenclature "a bag of enzymes." The existence of fibers, blebs and other internal structures, as well as the dense packing of the cytoplasm and the existence of mycoplasma motility, all suggest that we should view the cell as a more or less ordered liquid crystal performing a set of operations in which spatial orientation and enzymatic activity are coordinated. The characteristic shape of many *Mycoplasma* species in the absence of a rigid cell wall demands either the existence of a cytoskeleton or some other ordering structure. Shape maintenance may, of course, be an active metabolic process requiring the continuous expenditure of energy. In any case, for very small cells, the partitioning of materials at cell division is too important to be left to chance, and the requirement of some ordered protocol of cell division seems a necessity.

A word is in order about the place of mycoplasmas in the global evolutionary scheme. The mycoplasmas have been linked to the clostridia and the *Thermoplasma* to the archaebacteria on the basis of ribosomal RNA sequence. While I accept the heuristic value of this work, some caveats are in order. 1) The restriction enzyme fingerprints are not in themselves adequate to generate a reliable metric of taxonomic distance. Full sequence is necessary. 2) Taxonomic distance does not represent direction in time. Whether the mycoplasmas predate or are derived from bacteria cannot be determined solely from taxonomic trees. The preceding indicates the extent to which it is still debatable whether the mycoplasmas are descended from the bacteria or from prebacterial protokaryotes.

Let me note some properties we would be likely to ascribe to a primitive cell on an a priori basis. 1) It would be wall-less. 2) It would have a minimum genome. 3) It might use a nondegenerate code and could thus have an extreme base ratio—say 23% G+C. 4) It would have a rather simple—if inefficient—metabolism. If that description begins to sound familiar, then keep an open mind about the evolutionary position of the *Mollicutes*.

As theoretical biologists, let's now ask about the possibility of a photosynthetic mollicutes. It would need only a protein comparable to bacterial rhodopsin and a transmembrane ATP synthetase system and could presumably carry out direct photophosphorylation. Have present screening techniques ruled out such an organism? I doubt it, but I think if such a cell is found, it will illuminate some of these taxonomic issues.

Let me return to the mycoplasmas as the simplest living cells. Their existence with all the properties of life says that the "logic of life" is finite, relatively simple, and subject to full exploration. They allow us to put the doctrine of the completeness of molecular biology to the full test. Within 5 years or so we can determine whether the reductionist paradigm can explain cell reproduction and cell regulation. At this level of inquiry, the distinction between genetics and cell physiology disappears. I think that this doctrine of possible completeness is so important that its full significance has not yet been fully realized.

The chief factor in all of this is not the unusual

features of mycoplasmas, but the fact that they are so ordinary in spite of a genome of  $5 \times 10^8$  Da. The techniques are at hand. If we could agree on a strain of choice and launch an international effort to determine the necessary informa-

tion, we should be able to proceed quite rapidly. We could be asking about some of the most profound generalizations in the history of biology. There can hardly be a more worthwhile goal in our field of study.