

PROJECT K: "THE COMPLETE SOLUTION OF E. COLI"*

F. H. C. CRICK†

It is convenient to consider future developments in molecular biology (in the widest sense) under three headings: (a) studies on cell components, (b) studies on unicellular organisms, and (c) studies on multicellular organisms. The latter, although of great importance, will not be dealt with here. The division between cell components (which may come from any sort of cell) and organisms is admittedly arbitrary and is only introduced here to make the discussion easier. In practice, most work on complete organisms is supplemented by studies on the components of that organism.

It is first necessary briefly to take stock of the present position. As far as classical biochemistry is concerned, many enzyme reactions are known, and for a minority of these the action of the pure enzyme is understood in outline. For no case have the details of the enzymatic action been firmly established in chemical terms. Within the field of molecular biology (in the narrow sense) we now understand in outline the synthesis of the nucleic acids and of proteins, their interrelation in the genetic code, and a little about their control mechanisms.

It seems likely that future progress will take place in several broad areas:

1. The more detailed test-tube study of the structure and chemical action of biological molecules (especially proteins). Typical of such studies will be the detailed action of enzymes (already getting very close with the solution by X-ray crystallography of the structure of several enzymes), the way proteins fold themselves up (a backward field), the radiation damage to molecules, especially to DNA, and many other topics. It is characteristic of these studies that they involve the application of complicated and advanced methods of physical chemistry to biological molecules, and often

* The idea arose in conversation with Dr. Sydney Brenner, who invented the title "Project K" and whom I have to thank for useful discussions on the topic. This short paper was originally circulated in a European Molecular Biology Organisation (EMBO) document [1] toward the end of 1967. It still seems to me to be an attractive scheme for people of the right temperament, and since EMBO is now unlikely to take it up I thought that it might be useful to give the idea wider publicity.

† MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, England.

require rather large amounts of (pure) material. We may also expect the chemical synthesis of model compounds to play an important part.

2. The filling-in of the broad outlines already established, for example, the biochemical mechanism of protein synthesis, the unwinding of DNA, the mechanism of genetic recombination (which is probably related to DNA repair mechanisms), and, on the classical biochemical side, the exploration of more metabolic pathways and especially their interrelationships.

3. Work on subjects of fundamental importance which are little studied at the moment, for example, the structure and function of cell membranes,¹ the mechanism of cell division, and the biochemistry of spore formation.

4. The study of control mechanisms at all levels, in particular the interrelation of the known mechanisms, leading to an appreciation of the economy and "design" of the cell.

5. The behaviour of natural cell populations and their population genetics, leading to the consideration of the evolution of the cell.

The above discussion is necessarily sketchy, but it clearly brings out three important points: (a) an enormous amount of work remains to be done without ever going to multicellular organisms; (b) important problems exist at all levels of complexity; and (c) there is likely to be an increasing demand for large amounts of pure cell components present in the cell in rather small amounts. For these reasons, it seems certain that in spite of the obvious opportunities awaiting the study of organisms having many cells, a major effort will almost certainly continue to be applied to single-cell organisms, in particular to bacteria.

The point of this paper is to argue that such work should be concentrated on one organism (probably *Escherichia coli*, K12) and that a case exists for centralizing many aspects of such work in a "central laboratory."

The major reasons for wanting to have the "complete solution" of a bacterial cell have been listed above. In addition, there is the intellectual satisfaction of having a single living cell "completely" explained. Of course, it is unlikely that the work will ever be pushed to the point that every possible detail about the cell is known. It does not seem very probable, for example, that all the various proteins of the cell (which may number several thousand) will all have their amino acid sequences and stereochemical structure determined. By "complete" one means complete in the intellectual sense, implying that nothing appears to remain which further experiment could not easily explain using well-established facts and ideas.

It is clear that if the cell is going to be considered as a well-integrated chemical factory, information from many different laboratories will have

¹ The understanding of cell membranes has advanced greatly since this was written.

to be pooled. This might argue for a central laboratory to act as a focus for such work, but there are additional reasons of a technical nature which make the case even stronger.

In the first place, the technique of studying the action of a gene by picking up specific mutants of it is likely to continue to be widely used. Now for a limited class of genes it is possible to devise special selective techniques, but this cannot be done for the majority. For many genes, however, it is possible to produce "conditional lethal" mutants. These are usually of two classes: (1) temperature-sensitive mutants, which grow at one temperature but not at another, and (2) suppressible mutants. Unfortunately, when producing these mutants one cannot usually obtain mutants of just the gene being considered, but mutants in many different genes. These must then be screened to obtain the class of mutants in which one is interested. The rest are usually discarded, which is clearly a wasteful process. It would be a great advantage if such mutants were characterized as far as possible so that they could be made available to other workers who might wish to study them in more detail. A central laboratory for producing mutants, and for receiving mutants from other laboratories, which would then be classified, stored, and made available to others would be an obvious help to everybody in the field.

Another reason for a central laboratory would be the production of cells on a large scale. Again, at the present time the tendency is for each laboratory to grow a large culture for one particular chemical component in the cell and to discard the rest. This is wasteful and will become more so as larger batches are needed in order to obtain sufficient supplies of the rarer molecules in the cell. Whereas some growing could be done commercially, this will not be enough in the long run, as large batches of special mutants will eventually be required for certain pieces of work.

All this suggests that there is a case for a central laboratory to coordinate and assist experimental work going on in many different places. It remains to discuss the choice of a suitable organism. There is, of course, no reason why eventually a central laboratory might not deal with several organisms, or alternatively that several such laboratories be started, each with its own special organism. However, in the first place it would seem sensible to start with one only.

The obvious requirements are: (1) the organism should be reasonably small to reduce the complexity of the problem; (2) it should be easy to handle and able to grow on a relatively simple, defined medium; and (3) the same strain should be used by most workers. Yeast is probably too big, and has the complication that many different yeasts are in use (e.g., "bakers' yeast" and "brewers' yeast"). It has the advantage, however, of easily forming stable diploids. The pleuropneumonia-like organisms (PPLO), though small, are difficult to handle and need a complicated

growth medium. The obvious choice is *Escherichia coli* (probably the K12 strain for genetic reasons), but *Bacillus subtilis* and possibly *Salmonella* would also have to be considered. The main point is that several reasonably satisfactory organisms are already well known. The final choice between them could be left open at this stage.

A central laboratory might well contain the following groups: (1) a genetic group for developing new and rapid methods of genetic mapping and screening; (2) a group to develop instruments for the automation of experiments; (3) a biochemical genetics group to produce, receive, classify, and supply mutants of all possible genes; (4) a fractionation group for developing more and better methods of fractionation; (5) a production group for supplying very large batches of partly fractionated material; and (6) a group to study control mechanisms and the general economy and design of the cell. To this could usefully be added various associate groups and visiting workers studying areas of growing interest (such as membranes or cell division), who would find the facilities provided by the rest of the laboratory an attraction and who in turn would point out the material most needed at any particular time.

Postscript: Since the above was written, some of the last-enumerated suggestions have been taken up, in particular (3) and (5). However, as far as I know, no one is at the moment trying to set up a central laboratory, and the idea of making our knowledge of *E. coli* (say) as complete as possible has not been announced by any group as their acknowledged target.

REFERENCE

1. Proposal for a European laboratory of molecular biology. Document CEBM 68/31E, European Conference of Molecular Biology, November 1967.

THAT LITTLE GAMETE

Of all the spermatozoan in the sea
You my stout lad swam free;
How fortunate am I that you were me.

GEORGE W. GILL