FEATURES

Inside a living cell

David S. Goodsell

A CLEAR PICTURE of the interior of a living cell that shows the average distribution of molecules at the proper scale, the proper concentration and with no missing parts, seems to me to be central to the understanding of the workings of life. However, this type of picture is virtually absent from the popular and technical literature. The reason for the paucity of comprehensive pictures is simple: there is no single experimental method to determine the information needed for their construction. Electron microscopy gives a view that is too coarse: subcellular structure is studied, but individual molecules are not seen. X-ray crystallography and classical biochemistry, at the other extreme, are too fine: individual molecules are studied in great detail, but information on their cellular environment is lost in their purification. The intermediate level – the molecular structure of cells – must be synthesized from information from these two extremes, fitting many individual molecular puzzle pieces together to form a realistic overall view.

In the drawings presented here, I have attempted to combine molecular composition data with structural results for a well studied organism, *Escherichia coli*. The three square illustrations show three different portions of a typical *E. coli* cell, magnified one million times. At this scale, with 1 nm enlarged to 1 mm, atoms are about the size of a grain of salt, ATP and chlorophyll are about the size of a rice grain and macromolecules fit easily into your hand. In these three illustrations, all small molecules – water, cofactors, biosynthetic intermediates, etc. – have been omitted, to clarify the distribution of macromolecules. All molecules, including these small molecules, are included in a 10 million times magnification, shown bounded by a circle. Imagine a similar dense packing of water and small molecules filling the interstices of the three square illustrations.

D. S. Goodsell is at the Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, CA 92037, USA.

Assumptions and calculations

Three books served as the major references for the macromolecular composition of *E. coli*[1–3]. Of particular use are Table I in the chapter by Schaechter and Neidhardt[1], which lists the molecular composition of a typical *E. coli* cell, and Fig. 2 in the chapter by Woldringh and Nanninga[2], which is a simpler version of the figures here.

The typical cell is an *E. coli* strain B/r in balanced growth at 37°C in glucose minimal medium. The cell is assumed to be 70% water, yielding a volume of 0.88 μm^3. This volume corresponds to a cell about 2.95 μm long and 0.64 μm wide. The volume of the envelope is 0.14 μm^3, assuming a width of 7.5 nm for the two membranes and a width of 10 nm for the periplasmic space. The nuclear material occupies another 0.14 μm^3, leaving 0.6 μm^3 of cytoplasm.

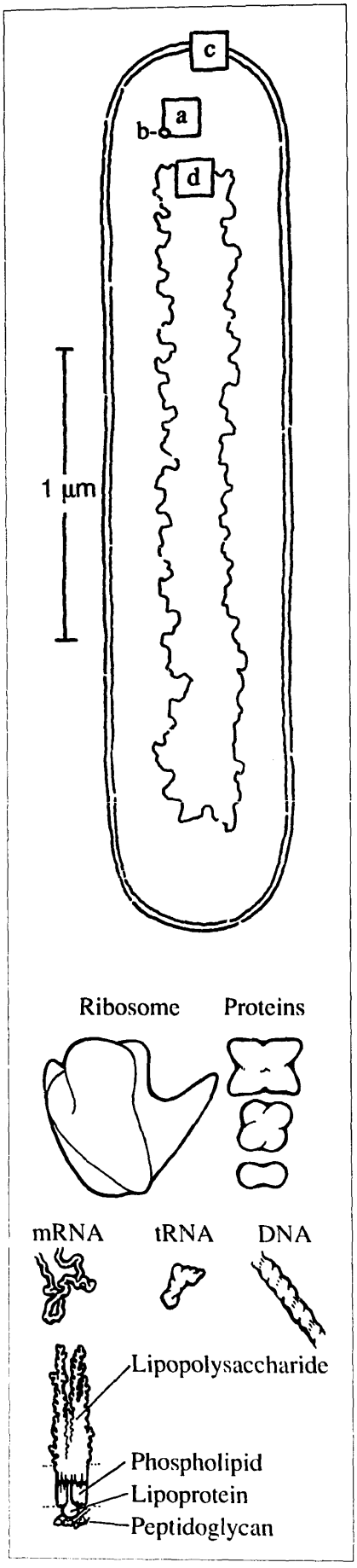
The partitioning of protein between the different portions of the cell is the major difficulty in the synthesis of the illustrations; other components such as lipid and nucleic acid may be read directly from Table I in the chapter by Schaechter and Neidhardt[1]. Protein comprises 55% of the dry weight of the cell (all percentage values will be percentage of the dry weight of the cell). The amount of protein in the outer membrane is calculated from Table I in the chapter by Nikaido and Vaara[1], assuming cylindrical proteins and a density of 1.33 g cm^{-3}, yielding a value of 6% of dry weight. Widely differing values of the percentage of protein in the inner membrane have been reported, ranging from 70:30 protein:lipid to 50:50. I used an intermediate value of 60:40 protein:lipid, yielding a value of 10% of dry weight. The concentration of protein in the nuclear region is also not well defined; Woldringh and Nanninga[2] use a value of 20 mg ml^{-1}, comprising about 1% of dry weight. This leaves about 38% of the dry weight of protein in soluble form. The protein components of ribosomes comprise 11% of this, leaving 27% soluble proteins. With a periplasmic volume of 0.057 μm^3 and a cytoplasmic volume of 0.6 μm^3, and assuming that the concentration of protein is the same in both compartments, 2% is in the periplasm and the remaining 25% is in the cytoplasm.

An average polypeptide has a molecular mass of 40 kDa (Ref. 1), so 25% of the 2.8×10^{-13} g of dry weight corresponds to about 1 000 000 individual polypeptide chains in the cytoplasm. However, most proteins do not exist as monomers. Using the concentrations of 25 soluble proteins reported by Albe *et al.*[4], we obtain an average oligomerization state of about 4, so there are about 250 000 protein entities in the cytoplasm.

Every attempt has been made to use the results of structural studies to illustrate each individual molecule. The structure of molecules studied by X-ray crystallography are taken from the Brookhaven Protein Data Bank. Results from electron microscopy were used for larger molecules, such as ribosomes and membrane proteins. In some cases, such as the flagellar motor complex, only rough guesses of size are available. The cell biology text by Alberts *et al.*[5] was used as a frequent source for references for these structures. Molecules are drawn in the simplest possible form, preserving only the gross shape and solvent-excluding volume.

Inside an *E. coli* cell

Figure 1a is a 100 nm window centered in the cytoplasm. The volume of cytoplasm in a typical *E. coli* cell is sufficient to fill about 600 cubes $(100\ nm)^3$. Each cube would contain a diverse collection of molecules. For the synthesis of protein, each cube contains an average of 30 ribosomes, over 100 protein factors, 30 amino acyl-tRNA synthetases, 340 tRNA molecules, 2–3 mRNA molecules (each about 10 box widths in length) and 6 RNA polymerase molecules. On this basis, about



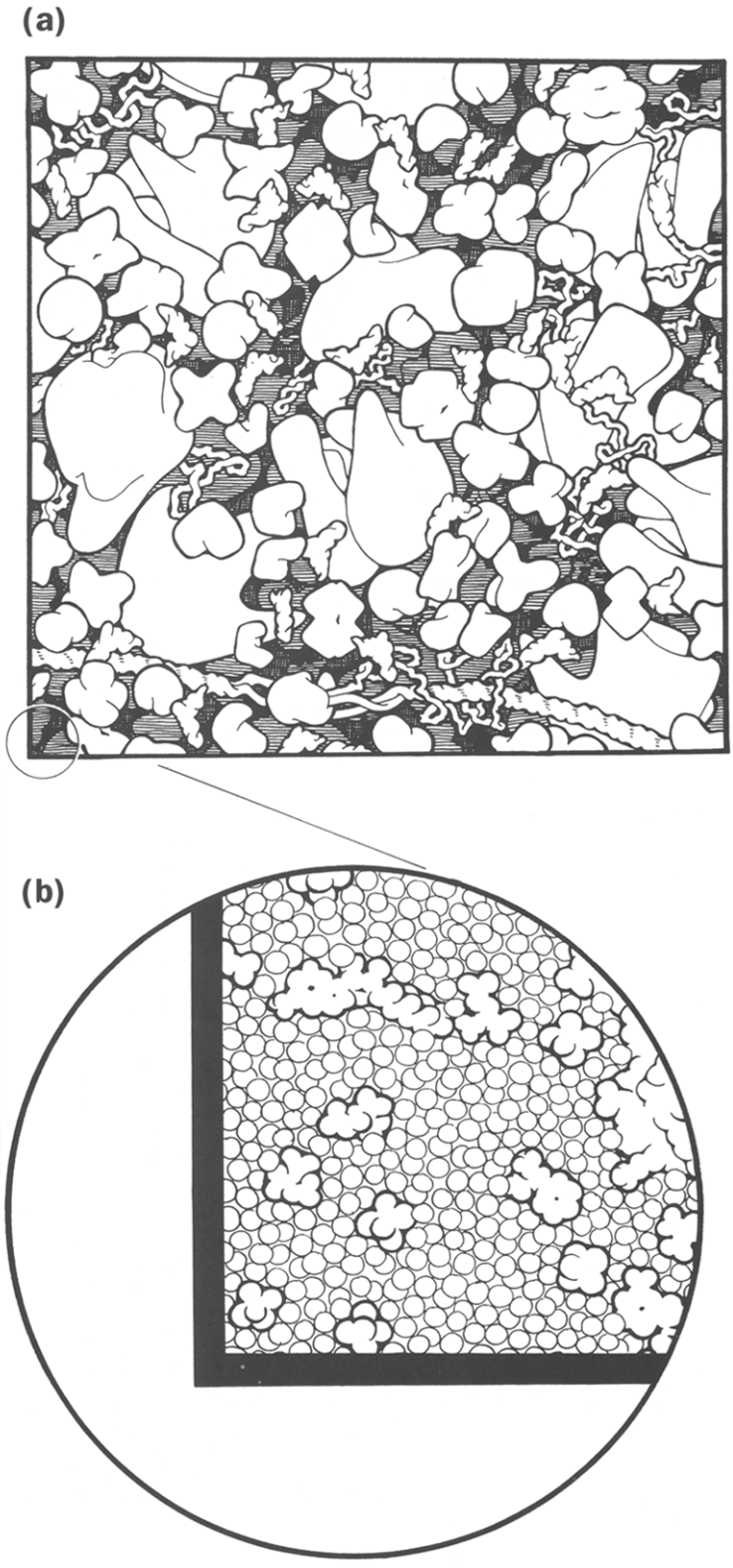
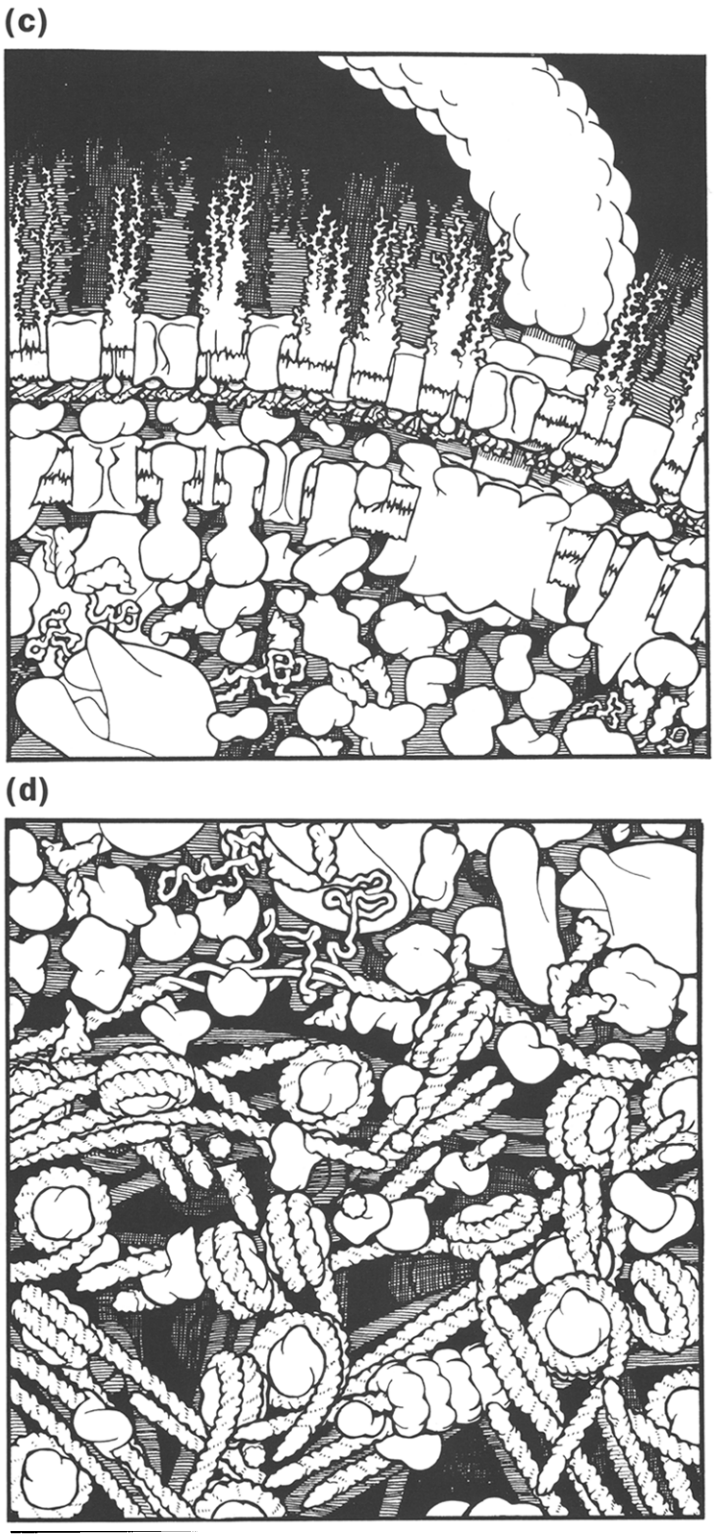


Figure 1

If we could magnify a cell one million times, making molecules the size of everyday objects, what would we see? Three portions of a typical *E. coli* cell are magnified one million times. A schematic of the cell at 50 000 times magnification shows the location and size of each 100 nm window with respect to the whole cell and the key identifies the macromolecular components. Although only three



examples are shown in the key, proteins come in many shapes and sizes. **(a)** The cytoplasm, showing all macromolecular components. **(b)** Close-up of one portion of the cytoplasm, showing all molecules, including water (circles), small molecules (dark outlines) and a small portion of a protein. **(c)** The cell wall, showing all macromolecular components. **(d)** The nuclear region, showing all macromolecular components.

36% of the dry weight of the cell is dedicated to protein synthesis, which is a prodigious allotment of resources. Another 330 protein molecules also fit into each 100 nm cube, including about 130 glycolytic enzyme molecules, 100 enzyme molecules from the citric acid cycle and a host of other anabolic and catabolic enzymes.

A ten-times enlargement of the corner, Fig. 1b, shows the concentration of small molecules in the cytoplasm. Assuming an average molecular mass of 200 Da, each 100 nm cube contains 30 000 small molecules, including precursors and cofactors. Also included are approximately 50 000 ions. In a cubic lattice, small molecules are about 3.2 nm apart, and ions about 2.7 nm. Note that water molecules adjacent to small molecules and proteins will be bound and not part of the bulk solvent phase.

The *E. coli* cell is bounded by a complex cell wall, cut in cross section in Fig. 1c. Our 100 nm cube isolates a square patch of the cell wall; there is enough cell wall for about 600 of these patches. Most apparent is one of the cell's flagella, rising out of the motor apparatus integrated into the cell wall. The outer membrane has two very different faces, with lipopolysaccharide molecules facing outward and phospholipids facing inward. In this 100 nm square patch of inner membrane, about 100 porin molecules span the lipid bilayer and 380 lipoprotein molecules interact with the layer of peptidoglycan lying below. The inner membrane, composed of a phospholipid bilayer, is packed with a diverse set of transport and energy-production proteins.

Figure 1d shows a 100 nm portion of the nuclear region. The nuclear region comprises about 140 cubes of this size. As expected, the region is dominated by DNA. With a single copy of the genome, about 100 box widths of DNA are packed into each cube. However, our typical cells are constantly replicating their DNA, so there is an average of 2.3 genomes per cell or over 200 box widths of DNA. The HU protein is shown at a level of 30 000 per cell, or about 200 in each 100 nm cube.

Dynamics

These illustrations are snapshots of an instant in time, as the densely packed molecules are actually in constant motion. To get an intuitive feel for these motions, let us compare the thermally driven motion of proteins to

motion in the macroscopic world. If an average 160 kDa protein were unhindered by surrounding molecules, it would travel at an average speed of about 500 cm s^{-1} at 300 K (calculated for an ideal gas). The molecule would move a distance comparable to its own size of 10 nm in about 2 ns. But the protein is surrounded by other molecules, which cause constant changes in direction and force the protein to perform a 'random walk' in space. In solution, the 160 kDa protein molecule requires almost 2 ms to traverse the 10 nm distance, almost a thousand times as long. In this 2 ms, the molecule samples a great deal of the space in between, instead of traveling a straight line.

Imagine a two-dimensional analogue in our macroscopic world. You enter an airport terminal and must reach a certain window. The distance is several meters, comparable to your own size. If the room is empty, the distance is traversed in a matter of seconds. But imagine that the room is extremely crowded, with people packed on all sides, trying to reach other windows. With all of the pushing and shoving in different directions, it takes you fifteen minutes to get across the room, a thousand times longer than the one or two seconds it would take if empty. You would bump into many different people and would definitely not travel in a direct line to your destination.

Acknowledgements

I wish to thank A. J. Olson for helpful discussions.

References

1 Neidhardt, F. C., ed. (1987) Escherichia coli *and* Salmonella typhimurium: *Cellular and Molecular Biology Vol. 1*, American Society for Microbiology
2 Nanninga, N., ed. (1985) *Molecular Cytology of* Escherichia coli, Academic Press
3 Ingraham, J. L., Maaloe, O. and Neidhardt, F. C. (1983) *Growth of the Bacterial Cell*, Sinauer Association
4 Albe, K. R., Butler, M. H. and Wright, B. E. (1990) *J. Theor. Biol.* 143, 163–195
5 Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. (1989) *Molecular Biology of the Cell*, Garland Publishing

LETTERS

The calcium-binding proteins

The February issue of *TIBS* featured troponin C (TN-C) on its cover and included inside an article by da Silva and Reinach on calcium-binding proteins (*TIBS* 16, 53–57, 1991). I do not wish to comment on the excellent description of crystal structures given there, but I do believe that the general science public should not be left with the impression that these structures can be used so directly to interpret the functional significance of the calcium-trigger proteins. The authors of this article appear to have missed entirely a huge literature on the properties of these proteins in solution using nuclear magnetic resonance (NMR) spectra. Apart from the work of my own group, I draw attention to the work of the groups of Forsen, Bax, Sykes, Carafoli and Hikichi, among others. This work shows that the calcium-trigger proteins in solution have the following properties.

First, in the calcium-free state at pH = 7.0. (1) The proteins do not have a fixed structure independent of temperature over the range 20–50°C. Unlike structures such as that of lysozyme, many are highly flexible. (2) There has been no evidence published for the long helix shown in your cover picture despite considerable efforts to find it. It is possible, some would say probable, that the protein is not present in this form at 37°C and that the long stretch between the two pairs of hands visits many states. (3) The hands are empty and in pairs, back to back with hydrogen bonds stretching between short β-strands, but there is also temperature dependence even in these regions. (4) The change to a fully denatured state is in fast exchange (>10^4 s^{-1}).

The conclusion is that these proteins are mobile to a greater or lesser degree, depending on the exact calcium-binding protein under discussion. Of course mobility is necessary if they act like a dimmer switch rather than an on/off switch. The results are in accord with thermodynamical studies by Privalov showing molar specific heat (ΔC_P) variations with temperature.

Second, the proteins bind protons in the hands below pH = 6.0 and protons affect their conformation. We do not know if (in part) the crystal structures have protonated 'unoccupied' hands. The crystal structures were carried out at pH<6.0. What bound state are they in?

Equally exacting NMR studies of solution structures have been made on calcium-bound proteins, which show the following. (1) The calcium binding is, to different degrees for different proteins, cooperative. However there are discernable states with different calcium loadings, i.e. either two or four in the case of TN-C. (2) The proteins are now much more rigid, but even now it is unlikely that there is a straight, long helix in solution. The diagrams shown have the beauty that crystals tend to exhibit but the truth may be much more ugly. Crystallization drives toward regular structure so as to give packing. (3) The structure changes in steps on binding calcium and it changes in subtle ways involving a (small) change in the strength of hydrogen bond network of the hands, as well as of their structure; a realignment of the short β-strands, a relative rotation and translational motion of the helices and some ill-defined rearrangements of the loops, which were not at all well defined in the calcium-free state. (4) The off-rate of Ca^{2+} is around 10^{-3} and Ca^{2+} exchange is in fast to intermediate exchange on the NMR time-scale at pH = 7.

So far nothing has been said about these proteins when bound in organized systems but we know that data will soon be available.

I believe that we should not, many would say must not, rely too heavily on crystallographic studies to reveal functional properties without the aid of spectroscopic studies especially by NMR. The NMR 'structures' can not be as good as crystal structures – there are intrinsic problems – but they reveal the nature of proteins in solution more clearly and the NMR method is more useful in describing changes of 'structure' and structural mobility. Every scientific method has weaknesses and we must not hide them from one another.

[For a recent review see Evans, J. S., Levine, B. A., Williams, R. J. P. and Wormald, M. R. (1988) *Calmodulin, Vol. 5* (Cohen, P. and Klee, C. B., eds), pp. 57–82, Elsevier.]

R. J. P. WILLIAMS

Inorganic Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QR, UK.