Visualizing the Origins of Life: Molecular Animation for Scientific Research and Education

Janet Iwasa
Massachusetts General Hospital / Harvard University
jiwasa@gmail.com

Abstract

How did life evolve on Earth, and what form did it take? Biochemists are attempting to address these questions using molecular clues found in our own cells. All life on the planet, from bacteria to humans, share a system of storing genetic information in the form of DNA and/or RNA molecules. Very early in Earth's history, the first living cell is likely to have evolved this system from simple chemical compounds found in the early terrestrial environment. This project seeks to use scientifically accurate molecular animations to illustrate the leading theories on how cellular life arose on the early Earth, and the experiments that are currently being done in biochemistry laboratories to test these theories. Molecular visualizations of the origins of life, created in close collaboration with researchers and science educators, will be used to communicate cutting-edge science research to the public and will also be used for presentations within the scientific community.

Keywords

Molecular Visualization, Biochemistry Research, Scientific Communication, Origins of Life, 3D Animation.

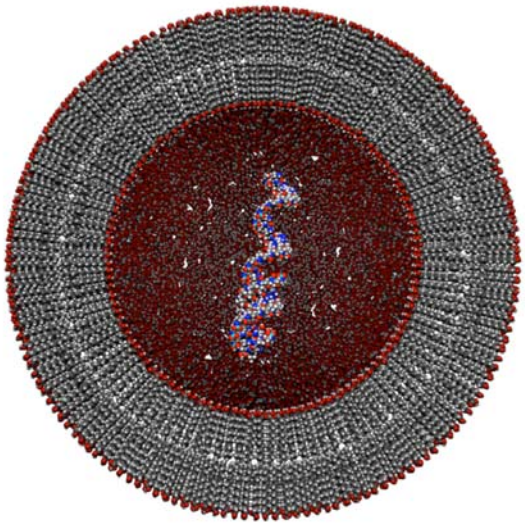

Figure 1. A model of a protocell showing RNA (center) encapsulated in a vesicle made up of fatty acids (hemisphere).

1. Introduction

How did simple chemical compounds form primitive, single-celled organisms, or protocells, that were destined to give rise to all known life on Earth? This is a question that has intrigued generations of scientists and non-scientists alike and spurred scientific research in diverse fields including chemistry, biology, and astronomy. Biochemists hypothesize about what molecules may have comprised the first living cell and how these molecules catalyzed reactions essential for life, such as replication. Ultimately, biochemists seek to build a model protocell that is able to undergo Darwinian evolution.

While the origins of life is a topic that is of broad interest to the public, visualizing the chemical and biochemical reactions behind the hypotheses poses a challenge. As a biochemist, I seek to explain how biological processes occur at a molecular level by elucidating how macromolecules, such as proteins and nucleic acids, behave in time and space. Too often, I have seen scientists resort to using blockish and simplistic representations of molecular interactions in their publications and presentations, only giving a crude estimation of conformational changes, cellular compartmentalization, and the like. These static illustrations not only lack structural accuracy and detail, but also do not allow for appreciation of their dynamic nature because they lack time resolution.

One method we can use to take on the challenge of illustrating complex and dynamic molecular processes is to create molecular animations. The use of movies and animations in scientific presentations is growing, and most journals now accept supplemental data in movie formats. More important, perhaps, is the role that animations can play in formal and informal science education. A single animation can provide a plethora of information about the shapes and scales of molecules as well as the roles they play within a larger biological system.

The goal of this 2-year project is to make scientific research and concepts on the chemical origins of life more accessible to the public at large and to the scientific community through the creation of dynamic, three-dimensional molecular visualizations. Animations will be designed with the oversight and input of a biochemistry laboratory actively researching the origins of life (the laboratory of Jack W. Szostak, MGH/Harvard University). Molecular visualizations of Szostak lab research will be created during close collaborations with Szostak lab members, and will be used by group members to communicate their research to diverse audiences. Portions of these animations will then be integrated into a multimedia exhibit at a popular science museum (Museum of Science, Boston), which will include live presentations, interactive touch-screen kiosks, and an online virtual exhibit. The planned exhibit will focus on theories of how life is thought to have evolved on Earth and how scientists are testing the feasibility of these theories in laboratories.

2. Scientific Background

One of the key tenets of molecular biology, known as the "central dogma" describes the process by which genetic information is stored and utilized by cells. Long strands of DNA are housed in the cell nucleus and are faithfully copied when the cell divides. Specific segments of DNA (genes) are used as templates from

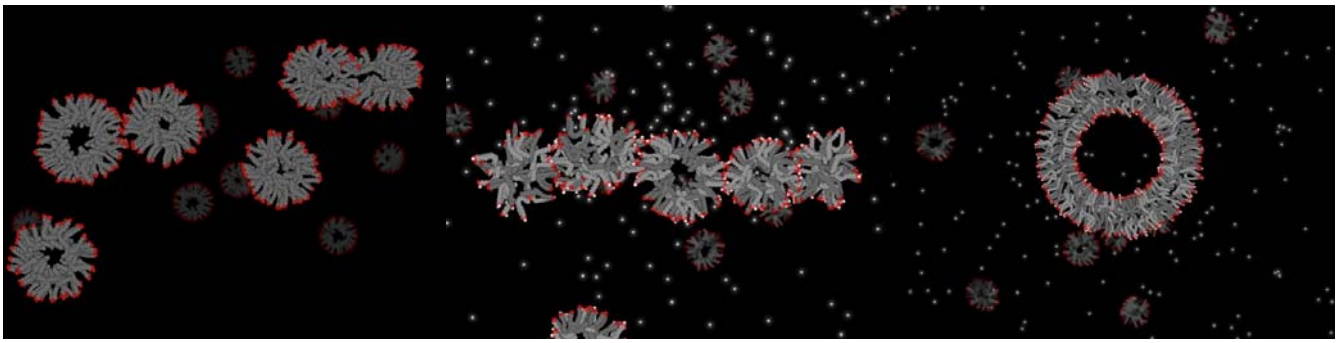

Figure 2. Still frames from an animation depicting vesicle formation from fatty acid micelles.

which RNA molecules (messenger RNAs) are copied. These RNA molecules direct the synthesis of proteins whose order of amino acids is derived from the sequence of nucleotide bases in their matching mRNAs. Proteins are the workhorse of the cell and have highly specialized roles, ranging from enyzmatic catalysis to the maintenance of cellular architecture. The central dogma is often simplified diagrammatically as: DNA → RNA → protein; this process is repeated millions of times in every cell of every living organism on Earth.

How did cells evolve this strategy of encoding genetic information? The "RNA world" hypothesis claims that ancient organisms actually used RNA to carry out basic biological functions and only later evolved to use DNA and proteins (Orgel 2004). This hypothesis was supported by the discovery of ribozymes, RNA molecules that are able to catalyze chemical reactions in an enzymatic manner (Kruger, Grabowski et al. 1982). Ribosomes are, in fact, ribozymes, and the finding that protein synthesis is carried out primarily by RNAs gave credence to the RNA world hypothesis (Moore and Steitz 2002). As both a storehouse of genetic information and a catalyst of metabolic reactions, RNA may have been the critical component that enabled the evolution of life. Some of these early ribozymes must have acted as RNA replicases, capable of using neighboring RNA molecules as a template to create additional replicases. In a prebiotic pool lacking compartmentalization, however, there would be no selective advantage for an improved replicase; a better replicase will simply copy other replicases more efficiently, whether they are closely related or not. Compartmentalization would allow better replicases to preferentially grow in number and evolve (Szostak, Bartel et al. 2001). Compartmental boundaries were likely to have been formed by lipid bilayers that assembled spontaneously into vesicles in the prebiotic environment (Hargreaves, Mulvihill et al. 1977). There is evidence that show that minerals, such as the clay montmorillonite, are capable of catalyzing both the polymerization of RNA as well as the conversion of lipid micelles into vesicles (Hanczyc and Szostak 2004). Reactive substrates such as montmorillonite may well have given rise to simple protocells that were the precursors of all life on Earth (Figure 1).

3. Specific Project Goals

3.1 Creating Animations for Scientists

In collaboration with the Szostak lab, I will create animations of the primary components of early cells, simulating RNA replication by a ribozyme and the process of vesicular growth and division in a prebiotic environment. Visualizing these events is fundamental to the understanding of the current hypotheses on the origins of life. Subsequent animations will demonstrate how these components might work together in a simple cell to create an interdependent and evolving system. The evolution of a cell, which takes place by subtle and gradual changes at the molecular level, is an ideal candidate for visualization by animation.

Animations that are currently being created delve into the question of how vesicles grow and divide, a question actively being pursued by members of the Szostak lab. Membranes pose a challenge to visualize because of their dynamic nature; while the larger vesicular structure may remain stable and apparently unchanging for weeks or months, there is a constant and fast exchange of lipids and other molecules occurring at the sub-microscopic level at the membrane surface. Vesicle formation from fatty acid micelles can be clearly illustrated by molecular animations (Figure 2). Additionally, the dynamic processes of vesicle growth by micelle incorporation and vesicle division by extrusion through porous material are prime targets for illustration by animation.

How might RNA molecules act as both templates and enzymes that catalyze RNA replication? Ribozymes form complex secondary and tertiary structures that are uniquely suited to catalyze specific chemical reactions. For this animation, I will first show an unfolded, nascent replicase RNA molecule to give the viewer a general idea of its structure and topology. Next the RNA will slowly begin to fold, giving rise to a secondary and final, folded tertiary structure. The concept of how an unstructured macromolecule consistently folds to form a very specific 3-D structure is important to grasp in order to understand how proteins and ribozymes operate. Even small subtle mistakes made during secondary or tertiary structure formation will likely cause the RNA to be a non-functioning ribozyme. An animation can clearly illustrate the importance of accurate folding, and can also show how a mutation at the level of a single base may cause drastic changes in secondary and tertiary structure. This visualization also provides the background necessary for understanding how a ribozyme may evolve to form a more efficient replicase.

In the second part of this ribozyme visualization, I will show the newly folded replicase associating with a template RNA strand. This template strand will be essentially identical to the unfolded nascent replicase RNA introduced at the beginning of the animation. A close-up shot will show how the replicase could theoretically utilize the energy of nucleotide hydrolysis to polymerize a new RNA molecule using the unfolded RNA as a template. The RNA template and newly synthesized strand can then fold into replicases themselves after polymerization is complete.

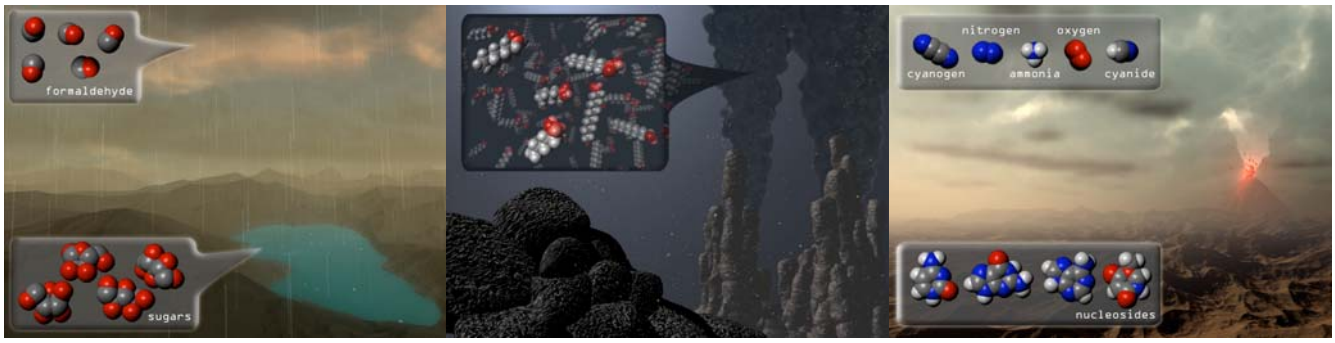

Figure 3. A triptych of scenes of early Earth, showing where some of the key molecules involved in protocell formation may have originated.

In addition to being a tool to enable members of the public to better understand current research, molecular animations also allow scientists to think more critically or differently about their own models and theories. Envisioning molecules as participants in a dynamic and complex process by manipulating them in three dimensions often raise new questions on how molecular events occur and contributes to the progress of scientific research.

3.2 Creating Animations for Public Outreach

The Boston Museum of Science (http://www.mos.org/) is one of the most popular science museums in the nation, attracting over 1.5 million visitors a year. Featuring hundreds of interactive, hands-on exhibits that encourage active thinking, the Museum of Science effectively communicates complex scientific ideas to diverse audiences.

Animations for the Museum of Science will focus on current hypotheses on the chemical origins of life and experiments that are being conducted to test the feasibility of these hypotheses. One such animation will describe how simple organic molecules were formed on the early abiotic Earth (Figure 3). The atmosphere before the appearance of photosynthetic bacteria was low in oxygen, allowing for the stable formation of molecules, including nucleotides and fatty acids, from energy sources such as ultraviolet light, lightning and meteorite collisions. Animations at atomic-level resolution could show, as examples, the formation of a simple nucleic acid or amino acid, using methane or carbon dioxide as a carbon source. These newly made molecules collect in pools of water, the "primordial soup." An accompanying animation can recreate the famous 1953 Miller-Urey experiment in which Stanley Miller and Harold Urey introduced sparks of electricity in a chamber containing methane, hydrogen gas, ammonia and water, forming a number of biologically important molecules over the course of a week (Miller and Urey 1959). As new theories emerged on the gaseous composition of the atmosphere of early earth, different mixtures of gases and sources of energy have been tested with similar results, demonstrating the robustness of this hypothesis.

Next, the formation of more complex structures from simple molecules will be illustrated. Minerals, such as montmorillonite, catalyze the formation of vesicles from micelles (Figure 2), and simultaneously catalyze the polymerization of RNA strands from nucleotides, sometimes resulting in one or more RNA strands that are enclosed in a lipid bilayer. Extremely rarely, an RNA strand is produced that possesses catalytic activity. Critical for evolution, as mentioned in the previous section, is the ability for RNAs to replicate themselves. Vesicles containing self-replicating RNAs fuse with other vesicles containing the raw materials, nucleotides and fatty acids, necessary for creating more replicases contained in more vesicles. In the last section, simple protocells will evolve more complex functions. The replicase machinery is prone to errors at some small rate, creating a diverse population of RNAs. Protocells with a more efficient replicase will grow faster, and thus outnumber those with poorer replicases. Some protocells may evolve RNAs with other activities, such as the ability to import nucleotides and energy sources, giving those protocells a selective advantage.

Additional animations will focus on concepts regarding how life may evolve on other planets. Is it likely that aliens would also be carbon-based? An animation focusing on this question could demonstrate how carbon is able to combine with itself and other elements to create a huge number of complex, stable compounds, a characteristic that makes it very possible that alien life would be based on carbon (Goldsmith and Owen 2002). Similarly, another animation could consider the qualities of water that make it ideal for the evolution of life compared to other solvents commonly found in interstellar matter, such as ammonia.

These animations, each several minutes in length, will be able to be viewed alone or in combination to create a longer narrative, and will include voice-overs and text to further describe context. Visualizations of the chemical origins of life will be incorporated into a variety of exhibition spaces at the Museum of Science, including the Current Science and Technology Center, web-based virtual exhibits, and touch-screen monitors on the museum floor.

4. Methods

4.1 Importing Molecular Structures

The atomic structures of numerous macromolecules, including proteins and nucleic acids, have been solved and are freely available for download from publicly accessible databases. The Protein Data Bank (References) is the central repository for macromolecular structural data, and currently holds over 30,000 molecular structures. PDB files contain the three-dimensional coordinates and identities of every atom in a given molecule. The molecule is then easily viewed and manipulated using one of a number of freely available PDB viewers such as Pymol (References) or UCSF Chimera (References).

Using different molecular representations will allow the viewer to more easily transition between events on macro (a protocell) or micro (small molecule) scales. Molecular representations commonly used by biochemists include spacefilling (Figure 4a) and ribbon (Figure 4b). Spacefilling models are useful for showing atomic identities and the overall size and shape of a molecule, while ribbon models are ideal for

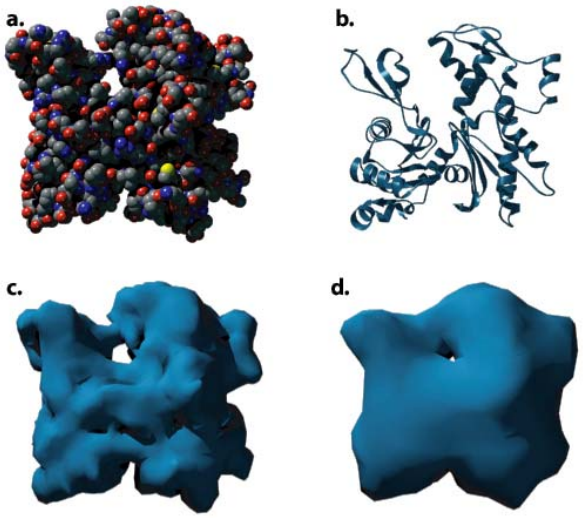

Figure 4. Different molecular representations that can be imported into a 3D application include (a) spacefilling, (b) ribbon, (c) rough surface, and (d) smooth surface. Images were created using UCSF Chimera.

illustrating its folded structure. UCSF Chimera (References), a free program used primarily as a PDB viewer, contains a feature that also allows users to convert a PDB structure into a surface with one of eight levels of resolution as well as space-filling and ribbon models (Figure 4). These molecular representations can then be exported from Chimera as a VRML file, which can be imported directly or indirectly into a 3-D graphics application. For importing structures into Maya (Autodesk), VRML files need to first be converted to .ma, .mb or .obj files using programs like Deep Exploration (Right Hemisphere). In general, high levels of resolution are useful for depicting close-up inter- and intra-molecular interactions and chemical reactions, while lower resolution molecular surfaces will for the incorporation of many molecules into an animation without overwhelming computing resources.

To create small molecules, such as fatty acids and single nucleotides, chemical modeling software such as Chem3D (Cambridgesoft) can be used to create a PDB file from which 3D surfaces will be constructed. For larger molecules that lack crystal structures, closely related molecular structures can be used, if available. Molecules can also be modeled from images collected from electron microscopy and from structural predictions. Larger structures, such as vesicles and cells, may be modeled based on light or electron microscopy data.

4.2 The Challenge of Animating Molecules

When creating molecular animations, it is often difficult to balance scientific accuracy, educational clarity and aesthetic judgment. Some of the most common issues I have faced center around the large numbers of dynamically moving molecules needed for a scene, and depicting how they move in space and time as accurately as possible.

One early challenge I faced in animating fatty acids was the sheer numbers that I needed in order to make a realistically large vesicle (Figure 2). In addition, each fatty acid needed to be moving in a random manner while moving laterally within the larger micelle or vesicle structure. I eventually chose to use particle sprites, each playing a looped movie of a single toon-shaded fatty acid, starting at a random frame (to give the illusion of randomness). The sprites can then be affected with turbulence fields and goals in order to add an extra layer of apparent randomness to the movement of the fatty acids. This method allowed me to create animations with large numbers of individually moving fatty acids whose movement and rotation could be easily adjusted. Importantly, large numbers of sprites are quick to playback and render.

Accurately depicting scale, both in space and time, can also be difficult. Atoms exist on an angstrom (1×10^{-10} meter) scale, proteins on a nanometer (1×10^{-9}) scale, and cells on a micron (1×10^{-6}) scale. The use of zoom in animation, while keeping some point of reference (usually a molecule in the center of the screen) is often an effective way to deal with the issue of scale, but can also be disorienting if used too frequently. Another option is to utilize a split-screen, with one side showing macroscopic events (such as a bubbling deep-sea vent) and the other showing sub-microscopic events (fatty acid formation) that is occurring within the context of the macroscopic scene.

A trickier issue is how to deal with molecular movements that occur on very different timescales. I have dealt with this issue in part by incorporating a small clock in the corner of the animation that slows down or speeds up to show the relative rates of molecular events. This is, however, not an ideal solution, as it draws the viewer's attention away from the main action of the scene.

One of the most difficult challenges facing molecular animators is how to accurately depict complexity and crowding in molecular systems. Living organisms, at a molecular level, are packed completely full of proteins and other molecules. This surprisingly high degree of crowdedness is shown beautifully and effectively in the artwork of David Goodsell (Goodsell 1993). While it would be technically very difficult, and perhaps undesirable to show this level of complexity in an animation, it is important to demonstrate that molecular reactions do not occur in a vacuum or in isolation, and that many reactions are occurring simultaneously within an organism. Dramatic lighting and creative shading and camera angles can then be used to effectively guide the eye of the viewer towards the molecule or reaction of interest within the context of a crowded and complex scene.

5. Conclusion

As beneficiaries of public funds, scientists have a responsibility to share scientific findings with the public and to educate them on the possible ramifications of their research. In biochemistry, this has been made difficult due to the complex and sub-microscopic nature of biochemical and cellular systems. Research in biochemistry has been progressing rapidly, and we have recently entered the age of genomics and proteomics. It is now more important than ever to supply the public with tools they can use to inform and educate themselves on the progress of research in science and medicine. Molecular animation will be a powerful tool that will enable diverse members of the public to better understand the complex and dynamic molecular processes that many in the scientific community are currently working to elucidate. Animations on the origins of life can help address areas of public misunderstanding, such as how evolution occurs at the level of a single molecule of nucleic acid -- a concept that is fundamental to understanding the origins of life and, in turn, the evolution of a species.

Acknowledgements

This project is supported by NSF grant no. 0610117 (Discovery Corps Fellowship). I thank Jack W. Szostak (MGH/Harvard University), Szostak laboratory members, and M.J. Morse (Museum of Science, Boston) for their collaboration, support, and guidance.

References

GOLDSMITH, D. and OWEN, T. (2002). The Search for Life in the Universe. Sausalito, CA, University Science Books.

GOODSELL, D. S. (1993). The Machinery of Life. New York, Springer-Verlag.

HANCZYC, M. M. and SZOSTAK, J. W. (2004). "Replicating vesicles as models of primitive cell growth and division." *Curr Opin Chem Biol 8*(6): 660-4.

HARGREAVES, W. R., MULVIHILL, S. J. and DEAMER, D. W. (1977). "Synthesis of phospholipids and membranes in prebiotic conditions." *Nature 266*(5597): 78-80.

KRUGER, K., GRABOWSKI, P. J., ZAUG, A. J., SANDS, J., GOTTSCHLING, D. E. and CECH, T. R. (1982). "Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena." *Cell 31*(1): 147-57.

MILLER, S. L. and UREY, H. C. (1959). "Organic compound synthesis on the primitive earth." *Science 130*(3370): 245-51.

MOORE, P. B. and STEITZ, T. A. (2002). "The involvement of RNA in ribosome function." *Nature 418*(6894): 229-35.

ORGEL, L. E. (2004). "Prebiotic chemistry and the origin of the RNA world." *Crit Rev Biochem Mol Biol 39*(2): 99-123.

Protein Data Bank. http://www.rcsb.org.

Pymol. http://pymol.sourceforge.net/.

UCSF Chimera. http://www.cgl.ucsf.edu/chimera/.

SZOSTAK, J. W., BARTEL, D. P. and LUISI, P. L. (2001). "Synthesizing life." *Nature 409*(6818): 387-90.