RIBONUCLEIC ACID BIOLOGY AND IMAGES

HELENA JAMBOR

Scientific Coordinator, CRTD, TU Dresden. Postdoctoral Research at the MPI-CBG, Dresden (2010–2015). Predoctoral researcher at the European Molecular Biology Laboratory, Heidelberg (2004–2008). Studies of Cell Biology and Biopsychology at the Freie Universität Berlin and Robinson College, Cambridge University (1999–2004). – Address: Zentrum für Regenerative Therapien TU Dresden, Exzellenzcluster an der TU Dresden, Fetscherstraße 105, 01307 Dresden. E-mail: helena.jambor@tu-dresden.de.

Everything as planned, nothing as planned summarizes my stay at the Wissenschaftskolleg in Berlin. To work on hidden messages in ribonucleic acids, I used my precious three months to learn the necessary computational skills. But I also allowed myself to drift off course and to trace the history of visualizations of ribonucleic acid data from the beginning of this field of research until today.

Everything as Planned

Ribonucleic acids are a main component of all cells on this planet, be it of plant, bacteria, or human origin. Besides water, cells are composed of DNA, the genetic material; lipids, which form a really thin, semi-permeable barrier surrounding the cellular content; and proteins, which perform most tasks. Around 20 % of the dry weight of cells is ribonucleic acids. This high abundance alone indicates that ribonucleic acids have critical importance.

The most famous task of ribonucleic acids is to transport genetic information from the DNA to the protein translation machinery, a process termed gene expression. In a way,

the ribonucleic acids thus work as a mobile and transient form of genetic information. In eukaryotic cells, instead of being translated into protein right away and thereby fulfilling their biochemical task, ribonucleic acids undergo an incredible number of regulatory steps, each of them slowing down the process in which a DNA is made into protein. In a circular turn of events, ribonucleic acids are not only the template for protein synthesis, but also actively control this process: specific classes of ribonucleic acids – ribosomal ribonucleic acids, transfer ribonucleic acids, and other small ribonucleic acids – are involved in all key steps and are found at the active center of the molecular machines that complete translation. The ribonucleic acids control the rate and efficiency of translation and can block it entirely. Moreover, ribonucleic acids not only control their expression into proteins, but also control which part of the DNA is even made into ribonucleic acid, how fast, and which parts are silenced for entire lifetimes. In other words, the key processes that sustain life would not function without the participation of ribonucleic acids.

Ribonucleic acids, however, cannot achieve any of these functions alone. Instead, they must functionally interact, work together with other ribonucleic acids, lipids, DNA, or proteins. Interestingly, we do not understand very much about how these interactions occur and how molecules in cells recognize the particular ribonucleic acids that they should interact with. This recognition is determined by the specific sequence of the four individual nucleotides adenine, uracil, guanine, and cytosine along the ribonucleic acids chain (*primary sequence*) or by the two- and three-dimensional shape into which the ribonucleic acid molecule can twist itself. Such secondary and tertiary structures are of course predefined by the arrangement of nucleotides along the ribonucleic acid, which can be up to several thousand nucleotides long. Secondary and tertiary structures are influenced by the chemical composition of its surroundings – such as the molecular density (*crowding*), the electrical charge of macromolecules, and the number of ions. To understand how specific ribonucleic acids work, we therefore must get a comprehensive understanding of their location, their cellular context, and how they associate with interaction partners.

As an example, I work with ribonucleic acids that, upon being formed in fruit fly ovary cells, exit the nucleus and, instead of being translated into protein immediately, are transported through the large cytoplasm of these cells to a specific destination within the cell. This "localization" process is essential for determining the embryonic axis, and when ribonucleic acid transport is disturbed, the fly embryo can develop without a head or abdomen. But in other cells the location of ribonucleic acids is also critical, such as in neuronal cells or in epithelial cells that form barriers lining the body surfaces (e.g. lungs,

and intestine). In the oocyte, ribonucleic acids must traverse up to 500 micrometers, which sounds rather little, but is very far for a tiny molecule. To cross this distance with diffusion alone would take two weeks, too long for cells that can divide faster than this. Mechanisms to achieve faster active transport evolved to overcome the limitations of relatively slow diffusive transport; these active transport mechanisms rely on interactions with proteins. To now search for hidden motifs that are required for this specific molecular interaction, we must know a few things: 1. Which ribonucleic acids accumulate at specific locations within the oocyte? 2. When do we observe them there? 3. What is the primary sequence when they localize? 4. What proteins bind to them? and 5. What kind of sequence motifs are we looking for? With this knowledge, we are then able to computationally search and find commonalities and motifs in the ribonucleic acids. In my previous work, I collected all the necessary data: we know the thousands of ribonucleic acids that can localize and the 591 ribonucleic acids that do so at a specific time; we know exactly the primary sequence of the ribonucleic acids accumulating and what proteins they must interact with. And for several ribonucleic acids, we have narrowed down the region where the motif must be hidden.

Understanding basic motifs in ribonucleic acids is rewarding on several levels: given that ribonucleic acids have been around for four billion years and were likely very much involved in the emergence of the first life forms on this planet, it is exciting to think of such motifs as a possible starting point for cells altogether: some kind of motif must have allowed the first ribonucleic acids to interact with molecules of other kinds, to then form a somewhat more permanent assembly and allow formation of a protected environment in the turmoil of the early Earth. It is also fascinating that, as today ribonucleic acids have pretty much the same role in all organisms, their basic principles of functioning must have been around very early on – and therefore, motifs in ribonucleic acids must in principle be interchangeable across organisms, too. Thus, deciphering the rules of how, when, and for what purpose ribonucleic acids interact with other molecules is a universal question.

Nothing as Planned

Ribonucleic acids are key for all life forms, yet, they are not as popular and as widely known as other molecules. All work on the molecular scale is visible only with electron microscopy or by leaving a molecular trace, which is challenging to communicate and requires a visual code, and this is also true for the history of nucleic acids. For the longest

time, we did not know that ribonucleic acids even existed – nucleic acids were discovered in 1871 by Friedrich Miescher, but it took almost another century until it was discovered that there are two forms, one being ribonucleic and the other nucleic acid. Soon after it emerged that there were two classes, DNA was discovered to be the heritable material, and ribonucleic acids sank back into irrelevance. It wasn't until the 1960s that ribonucleic acids garnered attention again, when their role for protein translation was elucidated. But now, their role was entirely that of a helper in the cell – helping the proteins perform an important job and helping to express the DNA code into protein machinery. Only in recent years, with the discovery of ever more classes of ribonucleic acids and their respective roles as active regulators of cellular fate, are ribonucleic acids slowly gaining recognition.

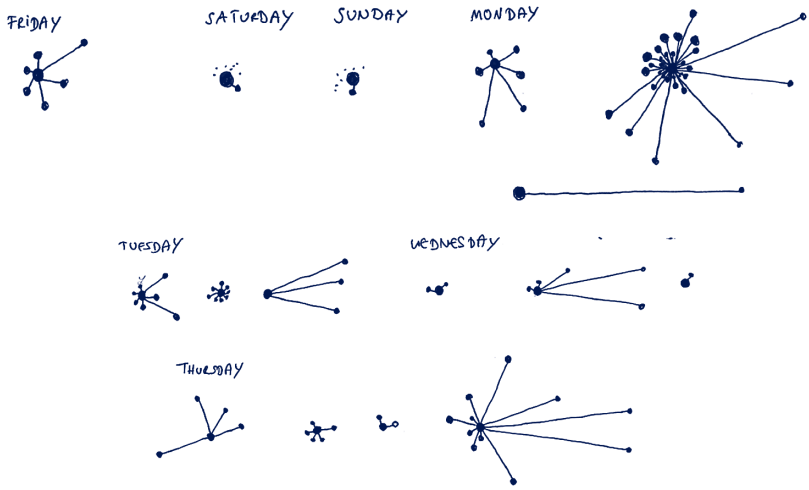
What were the factors that delayed the discovery and investigation of ribonucleic acids? For one thing, ribonucleic acids are much more fragile than DNA – and they are also much shorter, some so short that they were considered fragmented junk for the longest time. Ribonucleic acids are also very heterogeneous in their functions – from being an enzyme to encoding proteins, from being a sponge that collects cellular dirt, to acting as small signaling molecules and serving as a platform for complex cellular assemblies. And last, ribonucleic acids are not photogenic – they can constantly change their shape, length, and structure and can even adapt their structures through small fluctuations in local ion composition. As a result, the history of ribonucleic acid research is poor in images.

When reporting on the discovery of all nucleic acids, Miescher wrote a 30-page text with only tables as supporting evidence. Later, the first discovery and the description of the different ribonucleic acid forms were documented mostly in text, with mostly tables and one line chart showing adsorption spectra.

For cytoplasmic ribonucleic acids, the topic of my research, first descriptions come from the sea urchin egg. These observations were documented with text (in French) and tables. The authors observed that after fertilization of the egg, the pool of ribonucleic acids rapidly disappears. The author states that this rapid clearance is also the reason for the absence of a detailed graph, as the process was not observable. Cytoplasmic ribonucleic acids thus were not very photogenic from the very first time they were observed, which certainly is a difficult starting point for arguing for its cellular importance. Another report from 1949, providing exciting evidence of a dynamic presence of ribonucleic acids in the cytoplasm, was entirely devoid of any figures or tables. This work, despite being highly relevant today, lacks convincing images, which could be one of the reasons it has not been cited a single time in its almost 70 years of existence.

The first visualization of a cytoplasmic ribonucleic acid is a powerful image of a thin section through a cell in which the ribonucleic acids are visualized just as they are released into the cytoplasm. Since its publication in 1963, many such images have been published, always illustrating different aspects of cytoplasmic translocation of ribonucleic acids, and I added around 50,000 images to this growing collection. However, until today, we mainly see the cell in such images, and the ribonucleic acid is always visualized very indirectly. Until today we lack an iconographic representation of this central molecular class.

After spending my time with old and new images of ribonucleic acids, I then used my data to try out new computational tools to visualize large amounts of data. As I teach my students to start every visualization by hand, I also challenged myself to try new visualization strategies with pen and paper. One example is the sketch shown below – it summarizes what the Wissenschaftskolleg was for me – a chance to interact every day with many different people, from different fields, with different perspectives.



Conversation during the first week with colleagues (dots), mostly during meal times. Line length encodes conversation length. Dotted lines: footsteps outside my door. Note: the boat was Monday afternoon.