

DARPA Open-Source Biology Letter

Rob Carlson (rcarlson@molsci.org)
Roger Brent (rbrent@molsci.org)
The Molecular Sciences Institute, Berkeley, CA
October, 2000

We are requesting money to begin the development of "Open Source Biology". We are working towards the day when well-characterized molecular components, and the know-how to use them to design and implement new biological systems, will be available to anyone who wishes. Like the software movement from which it takes its name, the Open Source Biology community will rely on individuals and small groups of people to take charge of (and receive credit for) maintaining and improving the common technology, open to all, usable by all, modifiable by all. We believe that this development will have a number of positive consequences and that it may decrease the probability of some negative ones.

In particular, we are requesting funds to begin to develop and maintain a body of publicly available technology, to foster a community of researchers who contribute to this open-source technology repository, and to publicize the concept and the actual workings of open-source biology through meetings and the web. Our near term goal is to generate a set of interoperable components sufficient to comprise a basic "kernel" or basic "OS" for phage, bacterial, viral, plant, and animal systems. Components of the core OS will include tissue specific gene regulatory elements, transcription regulatory proteins and sites whose activities are tunable and switchable by small molecule chemicals, site specific recombinases and sites, and protein domains that can be used to direct specific protein-protein interactions. Extensions of the OS will enable other kinds of input into the system, and may include the creation of tool kits of broadly applicable effector molecules (outputs). Would-be designers of biological systems will be able to effect the controlled expression of effector molecules such as proteins specific to their goals.

Although much of it is currently proprietary, there is a great and increasing amount of genomic information available to the public for free. That is not true for the components and knowledge needed to design new living organisms. Here, there are fewer interoperable components, and researchers that possess and can work with these components to achieve desired ends are confined to high-end academic labs and corporations. Overall, too many relevant developments are protected by patents or outright secrecy. The highest concentration of ability occurs in plant biology, where there are really only four corporations that control collections of reagents and patent rights general enough to allow construction of most desired transgenic plants (Monsanto, Aventis, Novartis, and Dupont/Pioneer Hi-bred). For example, Dupont owns the rights to the most widely used site specific recombination system, and other companies need to spend millions of dollars working around the existing patents, or forgo the advantages that site specific recombination brings.

To the extent that some of the methods are becoming widely disseminated, open source biology may be already becoming a reality. For example, considerable information is already available on how to manipulate and analyze DNA in the kitchen. A recent *Scientific American* Amateur Scientist column provided instructions for amplifying DNA through the polymerase chain reaction (PCR), and a previous column concerned analyzing DNA samples using homemade electrophoresis equipment. The PCR discussion was immediately picked up in a slashdot.org thread where participants provided tips for improving the yield of PCR. Detailed, technical information can be found in methods manuals, such as *Current Protocols in Molecular Biology*, which contain instructions on

how to perform almost every task needed to perform modern molecular biology, and which are available in most university libraries. More of this information is becoming available online. Many techniques that once required PhD-level knowledge and experience to execute correctly are now performed by undergraduates using kits containing labeled (or even color coded) bottles of reagents. DNA synthesis (and DNA synthesizers) are becoming faster, cheaper, and longer, and it is possible that in ten years specified large (>10kb) stretches of sequence will be generated by dedicated machines. Should this capability be realized, it will move from academic labs and large companies to smaller labs and businesses, perhaps even ultimately to the home garage and kitchen.

However, although methods are becoming disseminated, interoperable components and tool kits are not. We thus think it is time to get out ahead of the developments a little bit, to try to make them happen faster and with more sophistication. We would expect such a development to have a number of positive consequences. First, by laying out clear design goals, we can enlist the efforts of numerous participants who have an interest and a stake in the design and engineering of biological systems but who are not now able to assemble the sets of tools and expertise to allow them to participate. For example, we see every reason that labs of individual investigators and even entire departments in state schools of engineering and agriculture can and should take charge of developing components that work with the kernel that are relevant to their design problems or species. Second, we believe that rapid attainment of a publically available kernel will enable smaller players, for example philanthropic foundations or startup companies, to perform sophisticated manipulations in support of their own goals. Third, by greatly enlarging the community of people who have experience designing and building new biological systems, we can increase the talent pool, foster economic growth, and increase the number of citizens who have some sophistication on these issues and can participate in the political choices that increasing biological capability can bring.

Similarly, we would hope that development of a public-domain kernel could avert some negative consequences. These include, for example in agriculture, averting further consolidation of the existing oligopoly and the consequent higher prices and delays to innovation that will result from proprietary OSs. That is, while development of proprietary OSs may be beneficial for a very few individual corporations, the economy as a whole will be stunted by a lack of competition and diverse innovation. Negative consequences also include include averting the damage done by "bugs", due to the greater robustness, larger pool greater developer talent, and faster ability to fix errors that open source structure brings to bear. To make this example specific, we think it would be a shame if, in 2009, most of the wheat in this country was dependent on an operating system of the quality and stability of Windows '95. We also hope to reduce the sensitivity of engineered systems to deliberate acts of sabotage ("viruses") by ensuring that the ability to work around complex biological systems is widely distributed throughout a self-confident, self-aware community.

Open-source biology will aid in maintaining a technological edge through diversified research. Like other distributed systems, biological research and biological engineering efforts conducted in an open source manner will be robust and adaptive, providing for a more secure economy and country.