# Towards a High-Level Programming Language for Standardizing and Automating Biology Protocols

Vaishnavi Ananthanarayanan and William Thies
Microsoft Research India
{t-vaisha, thies}@microsoft.com

For decades, biologists have relied on written descriptions of protocols to guide their experiments in the laboratory. However, due to recent technology trends, the practice of describing protocols with free-flowing English-language text is quickly becoming inadequate and obsolete. First, we are witnessing immense advances in laboratory automation systems. Microfluidic chips have been evolving at a pace faster than Moore's Law, with the number of valves per chip doubling every four months [1]. In order to leverage such technologies for biological experimentation, it will be necessary to express the protocols in a format that is not only comprehensible by humans, but also by machines. Second, the complexity of biology protocols is increasing dramatically. As we attempt to synthesize living systems as a composition of many parts, we will need to execute lengthy protocols with great precision. This will require a standard language for unambiguously describing the steps needed to synthesize a part, as well as for composing parts into a larger system.

We are developing a high-level programming language that enables standardization and automation of biology protocols. Our vision is to change the way that experimental methods are communicated: rather than publishing a written account of the protocols used, researchers will simply publish the code. As microfluidic devices mature, one could automatically replicate a colleague's experiment – or call it as a subroutine – by downloading the code to your own microfluidic chip. In the meantime, the code can be automatically converted to human-readable steps for manual execution in the laboratory. When written as a computer program, biology protocols can be parameterized to facilitate reuse in different contexts. They can also be mapped automatically to the setup of a given laboratory, taking into account the equipment and reagents that are available.

As a step towards this vision, we are defining and implementing the BioStream language for molecular biology protocols. In previous work, we demonstrated that an early version of BioStream could describe simple protocols and automatically execute them on different microfluidic devices [2]. Our current focus is expanding the language capabilities to encompass a broad and useful range of protocols. To date, we have expressed 15 protocols in the language, with protocols sourced from textbooks, classes, and published papers; our benchmark suite totals 2300 non-comment, non-blank lines of code. In the process of formalizing these protocols, we discovered and fixed several errors in the original descriptions. We have also implemented a C library that converts BioStream code to a human-readable format. Using this format, scientists at the Indian Institute of Science have successfully completed a DNA extraction procedure using our auto-generated protocol as the only reference. To the best of our knowledge, this represents the first time that a high-level programming language has been used to direct the actions of an experimentalist in a laboratory.

In the future, we look forward to collaborating with the synthetic biology community to explore ways of further standardizing their protocols, and to leverage a programmed description for maximum benefit to both manual and automated experiments.

**Acknowledgments.** We thank Utpal Nath, Mansi Gupta, Subhashini Muralidharan, and Sushmita Swaminathan for providing an experimental validation of BioStream, and Eric Klavins for helpful conversations.

## References

[1] J. W. Hong and S. R. Quake. Integrated Nanoliter Systems. *Nature BioTechnology*, 21(10), 2003.

[2] W. Thies, J. P. Urbanski, T. Thorsen, and S. Amarasinghe. Abstraction Layers for Scalable Microfluidic Biocomputing. *Natural Computing*, 2007.