

BY JOY P. KU, PhD

Stop Wheel Reinvention, Share Your Simulations!

Simbios has built a new publication repository that links publications to the research data and software behind them. The goal: to encourage and facilitate replication of published results and to foster use of what has already been accomplished rather than leaving others to reinvent the wheel. The repository is built upon Simtk.org—Simbios’ web-based infrastructure that provides open access to simulation software tools and models—making it easy to use and accessible to all.

“The publication repository is more than just the collection of data, models, and software used in the publications,” says **Jeanette Schmidt, PhD**, Executive Director of Simbios. “It provides the means for others to reproduce and build upon the results of your publication.”

GIVING YOUR PUBLISHED RESEARCH A FUTURE

Historically, when researchers have come across papers describing potentially useful software or data, their chances of actually getting their hands on that software or data were hit or miss. The student who did the research might have moved on, or the software developer might want to clean up the code first and take months (or longer) to do so. The Simbios publication repository for physics-based simulations of biological structures addresses this problem by providing a simple way to share and access the software, data, and other materials that support a particular research paper. It means that all the hard work behind the paper—the hours of coding, the repetitive experiments to get useable data—is captured and can easily enable future research.

That’s what motivated **May Liu, PhD**, a recent graduate from Stanford

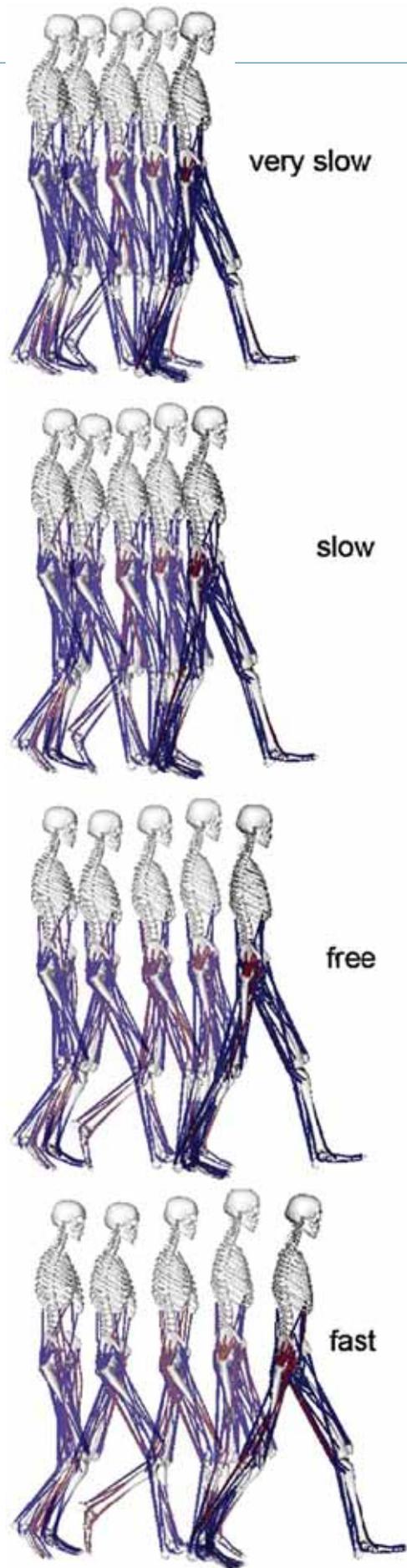
University’s mechanical engineering department, to create a publication project on Simtk.org. She is sharing 32 walking simulations used to analyze how muscle functions change with walking speed in children. It’s the largest number of simulations ever included in a muscle-driven simulation study, yet Liu sees it as just the beginning of further research rather than the end of the line. “The simulations themselves could become the starting point for a number of other studies,” Liu says. “There’s no reason why people should have to recreate simulations that already exist.”

Dahlia Weiss, a doctoral student in structural biology and chemistry at Stanford University, has a similar perspective. She established a publication project for her article comparing her *Climber* software tool against four other tools for interpolating between two molecular structures as one morphs into the other. *Climber*, based on a non-linear interpolation method, turned out to be very good at producing intermediate structures for very large, complicated changes.

“Knowing that we have a really good tool and not making it publicly available just seems really pointless,” says Weiss. She thinks *Climber* would be useful wherever high fidelity intermediate structures are required, not just for looking at structural movement.

REPLICATING RESEARCH GOES BEYOND SOFTWARE SHARING

But the Simtk.org publication repository is not just about software sharing. It supports and encourages sharing anything needed to replicate research results. For example, Stanford University researchers **Yuan Yao, PhD**, a post-doctoral fellow in the math department, and



On Simtk.org, May Liu created a publication project to share the three-dimensional simulation results from her latest publication analyzing eight subjects walking at four speeds (very slow, slow, free, and fast). Shown here are still images from simulations of a representative subject. The goal of the publication projects is to encourage and facilitate replication of published results. Courtesy of May Liu. Reprinted with permission from Liu, MQ, et al., Muscle contributions to support and progression over a range of walking speeds, Journal of Biomechanics (2008) 41:3243–3252.

Xuhui Huang, PhD, a research associate in the bioengineering department, and their colleagues developed Mapper, a tool that improves detection of low-density states within a massive amount of data. After creating a project for Mapper on Simtk.org, they submitted a paper showing how Mapper could be used to identify intermediate stable states during the RNA hairpin folding process, a difficult task when those states represent only two to three percent of the whole data set. In order for someone else to replicate that research, Yao and Huang posted (on Simtk.org) not only the Mapper software, but also the project's input data and instructions about how to use that data with Mapper.

"To reproduce the results from a paper is not an easy task," Huang says. "You need all the components together—the data, the program, your parameters, instructions—so that people can easily reproduce the results. Simtk.org provides such a platform, especially with this publication mode."

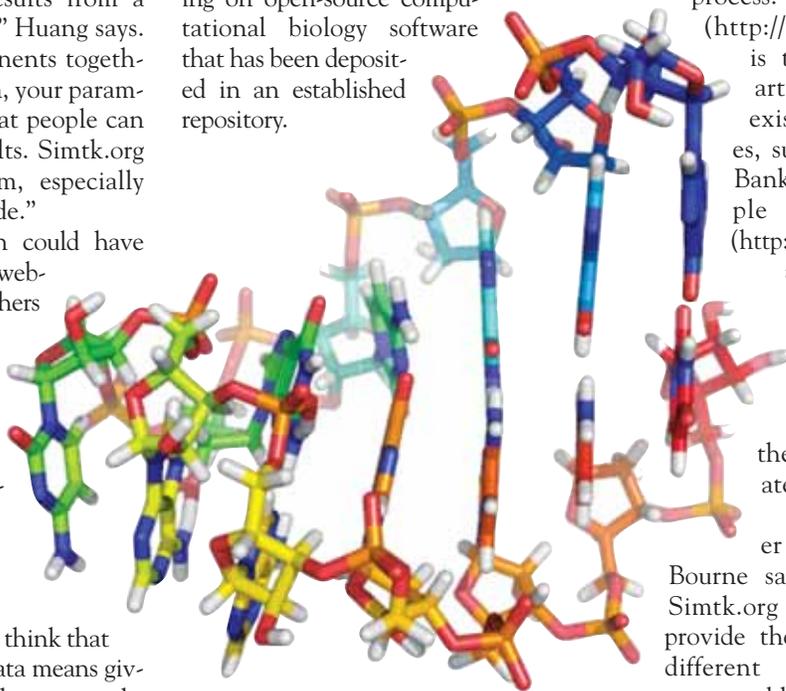
While the information could have been posted on his own website, Yao says that researchers from other fields would not think to look there. For an interdisciplinary field, a common platform like the Simtk.org publication repository is particularly valuable.

REWARDS FOR SHARING

While some researchers think that sharing their software or data means giving up their competitive advantage, others believe that it is a great way to build a successful career. "Careers often come from the application of software to make new discoveries in the life sciences," says **Philip Bourne, PhD**, a professor in pharmacology at the University of California at San Diego, and founding editor-in-

chief of the journal *PLoS Computational Biology*. "So by making the software available, researchers open up that possibility to benefit from what other people do with the software as well."

Bourne acknowledges that the current system does not always reward the work involved in preparing and supporting open-source software: answering questions from software users and providing documentation, examples, and tutorials. All of that effort takes time that could be spent doing research that would generate more publications—the metric by which academics are primarily judged. To address this concern, Bourne says, *PLoS* is considering having a special section that only publishes articles reporting on open-source computational biology software that has been deposited in an established repository.



Molecular dynamics were used to simulate the folding of the RNA hairpin structure (above), generating hundreds of thousands of molecular structures. Mapper was then used to sort through all that data to identify the relatively infrequent intermediate states that occur during the folding process. Courtesy of Yuan Yao.

A FIRST STEP TOWARD THE PUBLICATION OF THE FUTURE

Bourne sees the Simbios publication repository as a very positive step: "It actually speaks to the dream that I have." He envisions all aspects of research being accessible, with the paper being an access point to the experiment. From the paper, a researcher could retrieve and manipulate the associated data, and possibly discover new links and relationships via the data and tools—not just the paper citations—enhancing the research process.

Bourne observes that there are an increasing number of efforts to capture this whole research work flow process. The goal of his BioLit (<http://biolit.ucsd.edu>) project is to connect open access articles with information in existing biological databases, such as the Protein Data Bank (PDB). Another example is the Insight Journal (<http://www.insight-journal.org>), an open access on-line publication focused on medical image processing and visualization where authors are encouraged to provide the data and software associated with their papers.

"Most people get together because of content," Bourne says. Efforts such as the Simtk.org publication repository provide the infrastructure to share different types of information, enabling a dialog between the people who are using and developing the content.

"My sense is that in the next ten years, scientific discourse is going to change very dramatically as a result of these kinds of things." Bourne says. □

DETAILS

Learn more about the projects mentioned in this article.

Walking simulations: <http://simtk.org/home/mspeedwalksims>

Mapper: <http://simtk.org/home/mapper>

Climber: <http://simtk.org/home/climber>



Simbios (<http://simbios.stanford.edu>) is a National Center for Biomedical Computing located at Stanford University.