GuestEditorial

BY PETER LYSTER, PhD, PROGRAM DIRECTOR IN THE CENTER FOR BIOINFORMATICS AND COMPUTATIONAL BIOLOGY, NIGMS

Update on Biomedical Computation at NIH

As a program manager in biomedical computing and computational biology at the National Institutes of Health, I field many questions, particularly from new investigators. They ask questions like: Where do I find out about research funding? How do I navigate all the information? Whom do I contact? I want to take this opportunity to share a few insights.

NIH does not have a top-down approach for biomedical computing and computational biology, but it does have a highly coordinated community. Individual institutes and centers develop initiatives or are assigned incoming applications for funding in biomedical research through the Center for Scientific Review Division of Receipt and Referral (CSR DRR).

If anything comes close to centralizing biomedical computing and computational biology at NIH it's the CSR DRR—a hub that literally sorts out applications and assigns them to the most appropriate institute or center as well as the study section. The people who do this are science administrators who use their expert knowledge and excellent judgment to identify the right home for each application. When applicants tell me that they're going to request assignment to a certain program director, institute or study section, I tell them: "If you're not sure what you're doing, don't get tangled up in all that—let the experts at CSR DRR handle it so you can concentrate on the science."

All study sections at CSR can potentially review applications for research funding that involve some computing. However, there are nine study sections that review applications with a significant amount of biomedical computing. These include mainline modeling and analysis (MABS), data and analysis (BDMA), health informatics (BCHI), neurotechnology (NT), genomics and computational biology (GCAT), macromolecular structure and function (MSFD), biostatistics (BMRD), biomedical imaging (BMIT), and microscopy (MI). These nine study sections really point to the importance of computing in biomedical research and that these research areas merit special focus.

The glue that holds a lot of this together is BISTI, the trans-NIH Biomedical Information Science and Technology Initiative (BISTI) consortium. BISTI, for example, coordinates a number of initiatives in biomedical computing and computational biology. It is also the administrative center for the National Centers for Biomedical Computing, which are part of the NIH Roadmap for Medical Research. This program and its affiliated collaborations have funded more than $150 million in research in the past five years, and the effort will continue through 2015.

There are plenty of other opportunities for research funding across a range of size and complexity, and you can find them all listed on the BISTI Web site. Last year, BISTI reissued four broad-based program announcements to support "innovations in biomedical computing." They cover a range of areas, from the development of enabling technologies and non-hypothesis-based research to specific research relating to the needs of a disease or research area of interest to a specific IC. Of course investigators can also use the regular investigator-initiated R01 mechanism for requests for funding that have substantial components of computing.

BISTI and other related programs across the institutes and centers play an important role in providing both contacts and coordinating initiatives—and this creates a lot of communication within the NIH community. When I receive an application that I think may be more appropriate for another institute, I will use the BISTI Web site to find the right contact and then discuss the best home for review.

continued on page 29

NIH does not have a top-down approach for biomedical computing and computational biology, but it does have a highly coordinated community. Individual institutes and centers develop initiatives or are assigned incoming applications for funding in biomedical research through the Center for Scientific Review Division of Receipt and Referral (CSR DRR).

CHANGES IN THE NIH GRANT APPLICATION AND REVIEW PROCESS:

Want to see what's going on lately in the effort to enhance peer review? Go to the NIH site http://enhancing-peer-review.nih.gov/.
Since January 25, 2010, all applications are submitted on new forms with shorter page limits. The new page limits (http://enhancing-peer-review.nih.gov/ page_limits.html) include a 12-page Research Strategy for most applications.

```
float temp2;
{
  float multiplier = temp1;
  multiplier *= multiplier;
  multiplier *= multiplier;
  temp2 = multiplier;
  multiplier *= multipier;
  temp2 *= multiplier;
}
```

We are using only four multiplications to calculate a 12th power, which is much faster than the `pow()` function. Similarly, we can calculate the 6th power with three multiplications. But we can do even better by combining both of them into a single evaluation:

```
float temp2;
float temp3;
{
  float multiplier = temp1;
  multiplier *= multiplier;
  temp3 = multiplier;
  multiplier *= multiplier;
  temp2 = multiplier;
  temp3 *= multiplier;
  multiplier *= multipier;
  temp2 *= multiplier;
}
```

We are now calculating both powers at once with only five multiplications!

The final important optimization is to translate all expressions at once as a single unit. The above example shows only the expression for the energy, but in OpenMM we need to calculate the derivative of the energy as well. The two expressions share many subexpressions. For example, the derivative includes $(\sigma/\varepsilon)^{11}$ and $(\sigma/\varepsilon)^5$, so by translating both expressions together, we can compute four different powers at the same time.

In practice, we find these techniques work extraordinarily well for generating optimized OpenCL code to evaluate mathematical expressions. Our preliminary benchmarks with OpenMM show that the automatically generated GPU kernels are only a few percent slower than hand-tuned versions. At the same time, the user gains enormous flexibility to select the precise interactions they want in their simulations. □

Guest Editorial

cont'd from page 1

Every application that gets exchanged like this goes through CSR DRR.

While there is no central institute or office for biomedical computing and computational biology at NIH, there is a very vibrant and organic entity.

Now you're probably wondering about the outcomes of all these activities. In the last six years, the four broad-based BISTI announcements funded a total of 297 research grants in the amount of $355 million. In addition, the Continued Development and Maintenance of Software announcement funded 106 research grants in the amount of $160 million. In that same period, 5560 unique grant applications were reviewed in the informatics study sections (MABS, BDMA, BCHI, NT, GCAT, MSFD, BMRD, BMIT, MI and Continued Development and Maintenance special study section), and of these 1330 (24 percent) were funded.

For early stage investigators who want to add to these numbers by submitting successful grant applications, I offer the following advice:

- Team up with experienced mentors who can help you through the science and logistics of the NIH process.
- Talk to NIH program staff about your ideas. You can identify the appropriate contacts from the BISTI funding page/funding contacts link, http://www.bisti.nih.gov/funding/index.asp.
- Visit the BISTI Web site, which offers many useful resources, including a list of ongoing government programs, initiatives and public-private partnerships dealing with multiscale modeling, ontologies and data management, mathematical biology, systems biology, and numerous other biomedical informatics or computational biology efforts.
- Whether a new or seasoned NIH investigator, always focus your applications on the science because, after all, biomedical and health-related research is the NIH mission. □

Seeing Science

cont'd from page 30

rational approach to predicting gene function in *Arabidopsis thaliana*, a plant widely studied by plant geneticists. Dubbed AraNet, the work was published in the February 2009 issue of *Nature Biotechnology*. Marcotte and Lee are currently using the same approach to study gene function in humans.

"The idea is that we're making functional links between genes based on their behavior in a lot of different assays," Rhee says, including microarray analyses, protein-protein interactions and inferences from animal orthologs culminating in 24 different data sets.

The researchers started by analyzing pairs of genes with known function in order to set a baseline score for inferring related function. They then looked at about 27000 *Arabidopsis* genes—most of which are uncharacterized—to identify possible gene-gene associations among them. "By then asking 'what are the functions of the neighboring genes?' we can try to infer the functions of the uncharacterized genes," Rhee says. When her team experimentally tested the predictions for three uncharacterized genes, two out of the three had functions that were predicted by the network.

Rhee is interested in using inferences from AraNet to narrow down the candidate genes involved in complex traits. Although she'll be doing this work in plants, Rhee says the approach will be applicable to all organisms. She's also curious about uncharacterized genes that are connected only to other uncharacterized genes. "Perhaps we can use the network to characterize some undiscovered processes."

Ideally, Rhee says, researchers will combine AraNet's predicted functions with their own knowhow to try to design the best sorts of experiments to conduct. It's like rational drug design, she says: "You're using all the available information to be as systematic as possible in designing your experiments. This is a good application of systems biology." □