Computer Scientist in Profile: **Yang Zhang**

Contributor: Amarda Shehu[1,2] [amarda@gmu.edu]
Dept. of [1]Computer Science, [2]Bioengineering
George Mason University, Fairfax, VA 22030
Tel: 703-993-4135 Fax: 703-993-1710

"Yes, I did have a band in college, and I still love music," says Yang, who now has his own highly-successful computational biology laboratory at University of Michigan, Ann Arbor. While most of us are quick to associate his name with top protein structure prediction servers, such as I-TASSER and QUARK, not many of us know that Yang's first passion was music. "When I was young," he recalls, "my dream was to roam around the world with my guitar." His dream did not last long, as he made a conscious decision to pursue a graduate degree after completing his undergraduate studies in physics in China. His Ph.D. thesis was on the interaction of elementary particles, such as quarks (that explains a lot). He was awarded the prestigious Humboldt fellowship, which allowed him to study quarks at the Free University in Berlin for two years.

Yang's research interests took a sudden, unplanned shift to biology when he read an article by Zhongcan Ouyang on the shape of membrane vesicles in 1999. He recalls being fascinated by the fact that the predicted results on the shape of membrane vescicles could be directly confirmed and viewed in the wet laboratory. That was something he found he had been missing in his previous studies. No one has ever seen quarks, and their existence can only be indirectly validated through high-energy particle collisions. The need to connect computation with experiment ultimately drove Yang to Ouyangs lab at the Chinese Academy of Science, where he spent the two next years on studying the elasticity of RNA and DNA molecules.

In 2001, Yang joined Jeffrey Skolnick's laboratory in the University at Buffalo via a recommendation by Ulrich Hansmann. It was in Skolnick's lab where Yang started to learn how to fold proteins in silico. He recalls his time in the lab as the "Golden Age" of his research, as he could concentrate fully on science without worrying on having to secure funding or preparing lectures and other teaching materials. In Skolnick's lab, Yang developed a number of computer algorithms, including TM-score, TM-align, and SPICKER, which are still widely used in the community for comparing and analyzing protein structures. The most recognized accomplishment of his time in the Skolnick's lab is perhaps his TASSER method, which in essence allows assembling new protein structures from segments cut from known structures of other proteins. Using TASSER, Yang built the first genome-wide structure database of G protein-coupled receptors (GPCRs) in the human. This was a marked accomplishment, as GPCRs are now widely considered to be the most important and prevalent drug targets.

Yang continued his work on protein structure prediction and folding in Skolnick's lab till 2005. He then moved to the University of Kansas as an assistant professor. He and his team in Kansas continued Yang's journey on protein structure prediction and folding, as Yang fully recognized that his work was not done. "After more than forty years of effort," Yang says, "we still have not solved the problem of protein folding." In Kansas, Yang extended and improved TASSER to I-TASSER by iterative structure assembly simulations. He shared I-TASSER with the community through a web interface, and this resulted in I-TASSER establishing itself as one of the most widely used online structure prediction services. Since its development, I-TASSER has been consistently ranked as the best server for structure prediction in the community-wide "Critical Assessment of protein Structure Prediction" (CASP) experiment since 2006. The server has attracted so far more than $50,000$ registered users, with hundreds of jobs waiting on the queue on any single day. I recall having sent many of my undergraduate and graduate students over the years to Yang's I-TASSER server to complete their homeworks and

their understanding of protein structure prediction. The server capabilities combine an intuitive and easy-to-use interface with serious algorithmic power and rigorous analysis.

"I-TASSER starts with a technique called threading, which requires the availability of homologous proteins," explains Yang. After I-TASSER, Yang wanted to do more and move to the *ab-initio* folding territory. To build a protein structure from scratch, Yang and his colleague, Dong Xu, developed a new algorithm based on "continuous fragment assembly,"' which Yang named QUARK. There seems to be a deeper story behind the naming than the subject of Yang's Ph.D. thesis. Yang explains, "In particle physics, hadrons, such as protons and neutrons, which account for the majority of the mass of all materials, are an assembly of quarks." In Yang's view, all protein molecules are an ordered reassembly of atomic building blocks (backbone fragments and side-chains), which is exactly the principle that QUARK follows in assembling structure models of novel protein sequences. QUARK made a debut worthy of its name. As soon as it was introduced to the community, QUARK stood out as the top *ab-initio* folding algorithm in the 9th and 10th CASP experiments.

In 2009, Yang moved his lab to the University of Michigan in Ann Arbor and joined the Department of Computational Medicine and Bioinformatics founded by Gil Omenn and Brian Athey. "I love what I am doing here and enjoy strong support from the department and colleagues," he says. He notes that one of the benefits of his new work environment is the ability to always find computational and experimental collaborators drawn to the same scientific quests as him. That has helped him pursue many projects and expand them from the silica to the wet laboratory.

"I am excited in particular by two major puzzles that we are now trying to solve in my lab: (1) What we can tell on a proteins role in cell when we are given the protein structure (mostly by computational prediction)? (2) How can we do the reverse of protein folding, i.e., design new protein sequences when given target structures?" Yang got started on the first puzzle by developing COFACTOR and COACH with his colleagues Ambrish Roy and Jianyi Yang. The programs detect drug- and ligand-binding partners from predicted structure models of proteins. The algorithms currently ranked at the top in the community-wide protein function annotation experiments (including CASP and CAMEO). To address the second second puzzle, Yang and his colleague, David Shultis, built up a new wet laboratory to crystallize proteins Yang designed in silico. They enjoyed a recent success in redesigning the BIR3 domain of the functional X-linked inhibitor of apoptosis, and the Phox membrane scaffolding domain; with the latter recently deposited in the Protein Data Bank.

"I consider it my duty to better serve the scientific community. If I do not share my inventions, they are essentially useless."

Despite the rapid string of successes, Yang does not forget what he considers his basic duty as a scientist. He and his team maintain a comprehensive set of web-based services for a variety of projects, ranging from protein structure prediction, protein function annotation, protein-protein interaction, and protein-ligand docking and drug screening. "The maintenance of multiple high-quality service systems can be time-consuming but worthwhile", Yang adds, "as one of our major goals when designing new algorithms is ultimately to better serve the scientific community through them."

"Going back to my first passion," he says, "I still play guitar and drum. I play drum at least 20 minutes every day, which helps me refresh my brain for a while from the crowding world of protein folding." He also routinely listens to Jazz. "But playing guitar or listening to Jazz are more enjoyable at night or during weekends, when it is quieter," he adds. Nurturing his first passion seems to be working for Yang, and, on that note, we have come to a natural conclusion.

Yang on one of his drumming sessions.