CHEMICAL & BIOMOLECULAR ENGINEERING

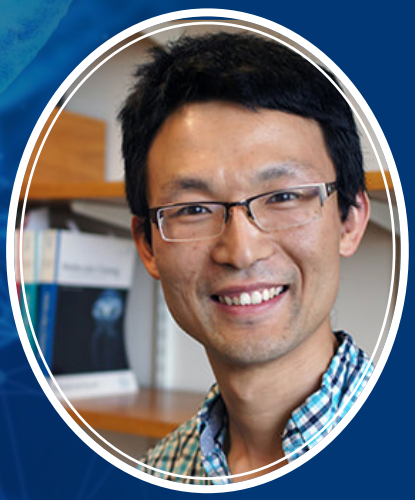
Center for Biomanufacturing Science & Technology

CBST SEMINAR

Wednesday, October 30, 2019

366 Colburn Lab

10:00 a.m.



Kang Zhou

National University of Singapore

Dr. Kang Zhou is an Assistant Professor of Chemical and Biomolecular Engineering, National University of Singapore (NUS). He obtained his Bachelor's degree from Tianjin University in 2007 and his PhD degree from Singapore-MIT Alliance in 2012. He subsequently did his postdoctoral research work at MIT and then joined NUS as Assistant Professor in 2015. He joined DiSTAP as a PI when the Singapore-MIT Alliance for Research and Technology program was launched in 2018. Besides the tool development topic mentioned in this presentation, his current research interests also cover the following topics: (1) engineering microbes to utilize new substrates as carbon source, (2) constructing new pathways in microbes to synthesize plant natural products, and (3) using compartments to enhance metabolite detection and synthesis.

"New Bioinformatic and Molecular Biology Tools for Metabolic Engineering"

Metabolic engineering uses and manipulates cellular metabolism to convert cheap substrates into more valuable products. Most of the involved biochemical transformations are catalyzed by enzymes. The flux from substrate to product can be limited by activity of one enzyme, especially when its coding gene is from another species. The flux can also be limited by imbalanced activities of multiple enzymes involved in the process, which has various, partially understood causes, such as accumulation of toxic metabolic intermediates, over-drainage of cellular energy and building blocks, and activation of cellular defence mechanisms. In this presentation, I will share with the audience our efforts in developing new tools for improving protein activity and for balancing metabolic pathway.

In the first study, we developed a computational tool to improve protein activity in collaboration with Dr. Xiaonan Wang's group at NUS. We trained a mathematical model by using machine learning algorithms to predict solubility of a protein from its amino acid sequence. Protein solubility instead of activity was used as the model output because (1) protein solubility and activity are correlated to certain extent, and (2) solubility data from different classes of proteins can be pooled together, making data collection much easier. With the

UNIVERSITY OF DELAWARE

https://sites.udel.edu/cbst

model, we were able to optimize protein's sequence *in silico* to maximize its solubility. We have experimentally tested some optimized protein sequences and found that a good fraction of them indeed resulted in more soluble proteins. For example, solubility of tyrosine ammonia lyase (TAL) was increased from 0.40 to 0.85 (if 15% of TAL forms inclusion bodies, its solubility is defined as 0.85). Subsequent *in vitro* TAL activity assay revealed a more than four-fold improvement in TAL activity when the original TAL sequence was replaced by the optimized one. The final *in vivo* test also showed that using the optimized TAL sequence can boost *Escherichia coli*'s ability of producing coumaric acid from glucose.

In the second study, we improved molecular biology tools for expressing multiple genes. Since there is currently no theory or computational tool to effectively guide balancing activity of multiple genes in metabolic engineering, one still needs to screen many combinations of gene expression levels to identify the best condition. To facilitate this process, we developed a new DNA assembly standard (GT Standard [GTS]), which allowed rapid construction of plasmids from reusable, standard DNA parts in an almost scarless way. With GTS, we can define coding genes and regulatory elements (promoters, RBS, etc.) as standard parts, and can easily arrange them in various ways to express a set of coding genes at different levels. As a demonstration, we constructed 72 *E. coli* strains for production of coumaric acid, and found that these strains resulted in a wide range of product titer (1-250 mg/L). GTS has also been used to facilitate the experimental validations in the first study –tags were added to proteins to increase their solubility by using reusable DNA parts.

Related publications:

X. Han, X. Wang*, K. Zhou*, Develop machine learning based regression predictive models for engineering protein solubility, *Bioinformatics*, 2019

X. Han, W. Ning, X. Ma, X. Wang*, K. Zhou*, Use a protein solubility-predicting model to guide protein engineering, *Manuscript under preparation*

X. Ma, H. Liang, X. Cui, Y. Liu, H. Lu, W. Ning, N. Poon, B. Ho, K. Zhou*, A standard for near-scarless plasmid construction using reusable DNA parts, *Nat. Comms, 2019 (In principle accepted)*