Thermo Fisher Award Recipient

Dr. Andrew Doxey, University of Waterloo, Waterloo, ON

My interests in bioinformatics and microbial genomics started in my undergraduate years (2003) in a summer research project with Drs. Brendan McConkey and Marilyn Griffith at the University of Waterloo. I was tasked with developing a method to predict proteins with ice-binding ("antifreeze") activity. This ultimately led to my Ph.D. work on the computational prediction of protein functions and evolutionary adaptations. I then completed an NSERC postdoctoral fellowship at Stanford University in the lab of Dr. Gill Bejerano, turning my attention to understanding non-coding regulatory sequences and how they work. Following my postdoc, I returned to the University of Waterloo in 2013 as an Assistant Professor to set up my laboratory. My lab's research merges my interests in protein biochemistry with computational genomics, and focuses on genomic data mining, unexplored molecular diversity, and protein function discovery. We have chosen microbial genomes and metagenomes as a central target for data-mining given the vast diversity of uncharacterized sequences (genes and genomes) that exist in the microbial world. We are particularly interested in the discovery of new protein domain families and domain combinations indicative of new biological functionality, and are currently focusing our efforts on bacterial toxins and degradative enzymes involved in bacterial biofilms and host tissue decomposition. My lab is supported by grants from NSERC, MITACs, and an Ontario Early Researcher Award

ThermoFisher SCIENTIFIC

Searching for biological novelty in a sea of genomes: detecting the unexpected

Andrew C. DOXEY, Department of Biology and Cheriton School of Computer Science, University of Waterloo, Waterloo, ON

There are now hundreds of thousands of microbial genomes and metagenomes accessible to any researcher with a computer and internet connection. Encoded within these datasets is a plethora of novel sequences, but interpreting this novelty and placing it within a biological context remains a significant challenge. In this talk, I will describe our work in protein bioinformatics, and the application of function prediction methods to discover new protein families, pathways, and biological functions within public databases. I will highlight three examples of 'unexpected' data-driven predictions made by my lab, and how these predictions have shed light on three areas within microbiology: microbial vitamin B12 production, the evolution of bacterial toxins, and the functional plasticity of bacterial flagella. Key to our approach is the use of multiple bioinformatic methods, including remote homology detection, structural modeling, phylogenetics, and genomic context analysis. By integrating these methods, it is possible to sensitively and accurately predict functions for a substantial fraction of genomic and metagenomic sequences, including many marked as "orphans" or "hypothetical genes" of unknown function. With the further development of these methods, we will increasingly be able to mine genomes and metagenomes for not only known biological functions of interest, but for entirely novel ones as well.