Unindexed Late Abstracts

The Environmental Molecular Sciences Laboratory:
Application to Biology and Biological Grand Challenges

J. W. Rogers
EMSL Director, Pacific Northwest National Laboratory, (509) 376-5328, jw.rogers@pnl.gov,
Richland, WA

The William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) is a national scientific user facility located at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. The EMSL was developed and is operated for the DOE as a multiprogram national user facility for molecular studies focused on solving the major environmental and biological problems facing DOE and the Nation. The EMSL http://www.emsl.pnl.gov/ has signature characteristics which include: integration of theory, modeling, and simulation, with experiment; multidisciplinary teams and collaborative modes of operation to solve major scientific problems; teams responsible for development of extraordinary tools and methodologies; scientists who design experimental strategies and operate state-of-the-art instruments; integrated operation and coordinated execution; education and training in the use of sophisticated instrumentation / computation / systems and approaches; a cyber infrastructure that facilitates productive remote interactions; and the charter to deliver capability in a transparent manner and to facilitate user outreach. The EMSL also boasts unparalleled resources and infrastructure in high-performance computing and informatics (e.g., Linux cluster supercomputer which supports computational biology and bioinformatics), nuclear magnetic resonance spectroscopy (e.g., 12 NMR spectrometers (300-900 MHz) and one pulsed EPR spectrometer), multimodal optical spectroscopies and imaging technologies (lasers; magnetic resonance; atomic force, near-field optical, and other scanning microscopies; electron microscopies), as well as advanced mass spectrometers for global proteomics (four Fourier transform ion cyclotron resonance mass spectrometers, Sciex QSTARR quadrupole time-of-flight mass spectrometer, Five Finnigan LCQ ion trap spectrometers, and a Finnigan TSQ 7000 triple quadrupole spectrometer), all of which are applicable to biological research. Supporting these technologies are world-class staff at PNNL, including biologists, chemists, physicists and software engineers, as well as instrument designers and builders.

The EMSL is now seeking concept papers for a scientific grand challenge in the area of membrane biology. Successful proposals will pose scientific questions that cannot be readily addressed without access to the full range of scientific instrumentation, computational resources, and research teams located at the EMSL for substantial periods of time. They are expected to attract and involve some of the best research scientists in the area chosen for study. Research areas of interest include biological membrane processes in cells (e.g., energy transduction, photosynthesis, signal transduction, dynamics of membrane proteins, and regulation of conformation states of proteins). Understanding membrane processes requires a systems-level analysis of fundamental cellular processes, the characterization of which is particularly well suited to the capabilities of the EMSL. The EMSL requests submission of a concept paper outlining the grand challenge goal, approach, technical requirements, and expected scientific and technical outcomes. Interested parties should become familiar with the EMSL facility and capabilities (http://www.emsl.pnl.gov/) and contact the EMSL Director for further information. An external review committee for the biology grand challenge will select concepts for further development. EMSL will host a workshop(s) for the successful concept team(s) to develop the scientific plan, strategy, and resource requirements for the grand challenge, which will be used in the final selection process. The EMSL staff will assist successful teams in implementing their program.

In Search of Complexity: Bioinformatics and Molecular Tools in the Search for Protein Recoding

Norma Wills, Barry Moore, Andy Hammer, Mike Howard, Clark Henderson, Jason Simmons, John Atkins & Ray Gesteland
Department of Human Genetics, University of Utah, Salt Lake City, Utah, USA

Estimates of the complexity of the human genome are converging to somewhere around 30,000 genes. This is only 5 times more than the number of genes in the simplest eukaryote – the single celled yeast, and less that twice the number of genes found in the worm. How do we account for the much greater complexity of the human? We must look to the proteins. It is well known that the number of unique proteins expressed by the genome is much greater than the number of genes. A number of mechanisms add to the complexity of the proteome. Some of these mechanisms such as alternative splicing and post-translational modification are well understood. Other sources of complexity are just beginning to be understood such as, programmed ribosomal frameshifting, transcriptional slippage, RNA editing and ribosomal bypassing. In the field of proteomics the study of peptides generated from full length proteins by protease cleavage remains the standard protocol. Peptides are readily removed from 2-D gels which have traditionally been proteomic's separation tool of choice, and they are much more efficiently separated by 2-D liquid chromatography which has been increasingly popular in recent years as a tool for proteome separation. However with all the benefits of peptide proteomics, valuable information is lost when full length proteins aren't considered.

Separation of very complex mixtures of full length proteins into fractions for analysis by mass spectrometry is exceptionally challenging. Abundant proteins are routinely found throughout the separated fractions, masking the presence of less abundant proteins, or those for which ionization is difficult. Tagging full length proteins and expressing them under their native promoter, or with over expression, is an alternative for preparing high quality protein samples for mass spec analysis. However, tagging, purifying and mass analyzing all known and predicted proteins from a genome would be monumental task. Here bioinformatics can inform our decisions about which proteins or predicted proteins are most likely to show non-standard decoding.

Here we discuss work that involves extensive searching of the human genome for telltale signs of recoding. We use Perl and Java (along with the biological counterparts Bioperl, and Biojava) to search predicted gene regions for frameshifting motifs, transcriptional slippage sites, RNA secondary structure, unusual ORF architecture, and cross species conservation to find potential targets cases of recoding. We use β-galactosidase and dual luciferase fusions to assay for frameshifting, readthrough, and transcriptional slippage. We employ affinity-tagging protein purification to isolate recoded protein products, and electrospray ionization mass spectrometry to identify the full length protein mass. This coupled with site specific mutagenesis allows us to tease apart the molecular details of known and predicted translational recoding events. As we

learn more about these details of recoding we will be better able to apply that knowledge to de novo prediction of recoding sites in raw genomic sequence.