

## Focus for John Yates III, 2004 ASMS Biemann Medal Awardee



It is a pleasure to introduce in this issue a focus section consisting of five articles honoring the achievements of John R Yates III, Cell Biology Professor at the Scripps Research Institute California. Over the past decade, John's research has been crucial in making the enormous leap from genomics to proteomics, particularly in relation to human disease. Specifically, his development of the now benchmark technique affectionately known as MudPIT (Multidimensional Protein Identification Technology) has revolutionized proteomic analysis of biological assemblies in complex mixtures. MudPIT is widely recognized as one of the most important of the emerging MS systems, enabling detailed analyses of the expression and function of the proteome and its response to disease.

Among his many accolades, John has numerous publications in *Science*, *Nature* and *Cell*. In 2002, he led a collaboration of scientists from the USA and the UK to determine the proteome of the most deadly form of the malaria pathogen, identifying more than 2400 proteins. Not only did he uncover those genes expressed at different stages of the life cycle of *P. falciparum*, he also determined which proteins are inside the cell and which are membrane associated (*Nature*, October 3, 2002). This crucial information led directly to the iden-

tification of candidate antigens in the design and development of potential vaccines. In other ground-breaking research, John has employed subtractive proteomics to more than quadruple the number of known lamina proteins, and went on to identify 23 new nuclear membrane proteins which map to chromosomal regions linked to human dystrophies (*Science*, September 5, 2003).

John has inherited a flair for invention from his one-time PhD supervisor and long-term mentor, D. F. Hunt, whose article, "Electron Transfer Dissociation of Peptide Anions" is published in the June issue. Where established manufacturers have failed to meet his exacting requirements, John has gone on to develop his own innovative technology. By applying his imagination and intellect to fundamental biological questions, John has produced reams of new information and a remarkable variety of potential candidates for the future treatment of human diseases.

The theme of human disease and its elucidation by MS techniques such as MudPIT is continued in the first of our dedicated articles. A. Emili and his colleagues present an analysis of complex tissue expression patterns in normal and subcellular organelle fractions of normal and diseased heart tissue. Employing an ingenious adaptation of subcellular fractionation to isolate

Published online June 23, 2005

discrete organellar fractions from heart tissue, this group has highlighted, and successfully overcome, many of the challenges presented by this fibrous tissue. In addition, the technical and computational solutions offered in this work make it an invaluable resource for researchers involved in, or thinking of becoming involved in, the MudPIT approach.

In a similar vein, the second of our dedicated papers takes us a step closer to the goal of early disease diagnosis by identifying protein species in biofluids that are symptomatic of disease states. The wealth of histopathological information contained in sera can often be masked by contamination from proteins that are abundant but not related to disease. T. D. Veenstra and T. P. Conrads et al. present an ultrafiltration methodology for the enrichment of low molecular weight (LMW) serum proteome, to concomitantly identify and quantify putative biomarkers in a single, high-throughput procedure. Combining MudPIT with differential labelling to examine trypsin-mediated incorporation of  $^{18}\text{O}$  in the LMW serum proteome of xenografted tumor-bearing mice, this group has distinguished over 6000 proteins, the largest number of proteins identified in serum to date. Almost 30% of the proteins identified by multiple fully tryptic peptides displayed altered abundance in mice xenografted with Lewis lung cancer. Of these, over 400 were quantitatively determined as having disrupted abundance in relation to Lewis lung carcinoma. The methodology presented by this group has wide ranging implications for the quantitative analysis of protein species indicative of human disease.

The use of protein biomarkers to detect the earliest stages of disease by monitoring changes in protein abundance is a theme picked up in the third of our articles. J. Gao and collaborators present a set of recommendations developed for the peptide hits technique (PHT), a promising procedure for semi-quantitative proteome screening. By refining and optimising the necessary bioinformatics and statistical analyses for a well defined sample, this group have opened up a simple and rapid whole proteome screening technique for the future analyses of unknown biomarker and/or target discovery samples.

With the completion of each new genome comes the further expansion of potential peptide sequences, and the associated increase of ambiguity in peptide/protein identification. In the fourth of our papers, R. D. Smith and his co-workers present a theoretical study addressing the statistical challenges involved in unequivocal identification of proteins from a progressively more

complex pool. *In silico* tryptic digests of all predicted proteins from the genomes of three organisms of varying complexity are evaluated by amalgamating accurate mass with calculated LC normalized elution time (NET). Their results show that where accurate mass alone may leave room for doubt, in combination with NET, a unique peptide fingerprint can be generated to support peptide identification in the context of high-throughput measurements. Importantly, this work demonstrates the potential for employing NET in addition to accurate mass to evaluate proteomes of high complexity whereby the majority of proteins should be uniquely identifiable.

A large part of the remarkable progress witnessed in MS in recent years is associated with the concurrent increase in computer power and the development of creative software to manage the staggering quantity of information produced. A ubiquitous challenge in any high-throughput approach such as shotgun proteomics is that of spectral redundancy. This issue is addressed by W. H. McDonald and colleagues in the last of our dedicated articles. Here a novel algorithm, MS2Grouper, is presented to elucidate structure in proteomic data sets. Significant improvements to current strategies are described to identify MS/MS spectra which result from a single parent peptide. By autonomously detecting repetition of spectra, assessing related spectral groups and then replacing these groups with representative synthetic spectra, MS2Grouper cuts both database search times and spectral noise without diminishing the number of peptides identified. Further, the effect of false positive or negative results is largely overcome by adopting a paraclique-based criterion for assessing analogous spectra, which guards against erroneous similarity links. Importantly, this approach can be adapted to run on arbitrarily large libraries of spectra. By presenting data acquired from biologically relevant samples, McDonald demonstrates the considerable benefits MS2Grouper offers proteomics.

John R. Yates III was awarded the Pehr Edman Award in 1998 and the American Society for Mass Spectrometry Research award in 1996. In his meteoric, but still youthful career, John's work has become a benchmark of excellence and in recognition of this he truly is a very worthy recipient of the 2004 Biemann Medal.

**M. G. McCammon**  
Guest Editor

**C. V. Robinson**  
Associate Editor