Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Editorial

Challenges and prospects of proteomics of non-model organisms

The elucidation of the double helical structure of DNA, and the mechanistic implications summarized in one sentence ("*It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.*") at the end of the seminal article published in 1953 by James D. Watson and Francis H. C. Crick [1], represented the birth of molecular biology. Less than 50 years later, the first draft of the entire human genome had been sequenced [2,3], heralding the new era of genomics. With an increasing list of genomes of organisms across all domains of life being routinely sequenced, it becomes clear that a large gap exists in predicting the phenotype from the genotype. In other words, understanding the genome alone does not allow obvious extrapolations into the complex biology that occur downstream from the gene. Proteins are the major workhorses of biological systems, the targets of natural selection, and their abundance is regulated by post-transcriptional and post-translational processes that cannot be detected, or even inferred, only by genomic and transcriptomic analyses. Gene regulation gives the cell control over structure and function, and is the basis for cellular differentiation, morphogenesis and the versatility and adaptability of any organism. Gene regulation may also serve as a substrate for evolutionary change, since control of the timing, location, and amount of gene expression can have a profound effect on the phenotype of a cell or multicellular organism, and phenotypic plasticity impacts on diversification and speciation [4].

At any time point in the natural history of a living system, the genotype–phenotype map is the outcome of very complex dynamics that include environmental effects. Bridging the gap between the genotype and the phenotype is synonymous with understanding these dynamics, and this goal demands a quantitative biology approach. The need for the complete picture engendered the development of new tools (most notably biological mass spectrometry [5]) and even a new field, proteomics [6,7]. Established in the 1990s, as a powerful analytical technique, proteomics has catalyzed an expansion of the scope of biological studies from the reductionist approach of conventional protein chemistry to the parallel and rapid proteome-wide measurements of hundreds to thousands of proteins. The subsequent development of quantitative MS techniques for the simultaneous study of proteins, ultimately the whole proteome, that are expressed in response to changes in gene activity, has represented a major advance in the field over the first decade of the XXI century [8].

Although significant advances in the comprehensive profiling of proteins have occurred in model organisms (*i.e.* those that, because of their small size and short generation times, facilitate experimental laboratory research), proteomics research in non-model organisms, albeit representing the vast majority of the biodiversity that exists on Earth [9], has not advanced at the same pace. There are several reasons that hamper the application of high-throughput proteomics to non-model, particularly to non-genome-sequenced, organisms. On the one hand, peptide sequencing *via* MS/MS requires a great deal of *de novo* interpretation of product ion mass spectra [10], involving the manual intervention of an expert mass spectrometrist, to subsequently carryout cross-species, homology-based database searches [11]. Automated *de novo* peptide sequencing remains a challenge, and even a simple MS/MS spectrum may require minutes for a trained expert to interpret. On the other hand, the application of proteomic approaches to non-model organisms whose genome is not available or ill-defined entails the principal challenge of making it necessary to infer protein functionality through the evolutionary conservation of gene sequences and protein domains between species. In addition, shotgun protocols do not work for non-genome-sequenced organisms, precluding the use of shotgun and targeted proteomics for the relative or absolute quantification of proteins. Notwithstanding these limitations, the proteomics of non-model organisms also presents its own strengths. Hence, proteogenomics [12] and proteotranscriptomics [13] offer the possibility of refining, even proof-checking, species-specific genome and transcriptome databases using proteomics data. Further opportunities of proteomics investigations of non-model organisms

http://dx.doi.org/10.1016/j.jprot.2014.04.034

can be guessed using a simile borrowed from the field of structural biology.

The first crystal structure of an integral membrane protein ("non-model protein") dates from 1985 [14]. By then, almost three decades after Kendrew, Phillips and others reported the first structure of a protein, myoglobin [15], protein crystallography had produced hundreds of crystal structures of soluble ("model") proteins. At present, the growth of crystals suitable for X-ray diffraction studies remains the major bottleneck of membrane protein structure determination. On the other hand, the advent of structural genomics (SG) initiatives, that seek to describe the 3-dimensional structure of every protein encoded by a given genome, has exponentially enhanced the elucidation of "model proteins". As opposed to traditional structural biology, the determination of a protein structure through a SG effort often comes before anything is known regarding the protein function. Conversely, non-SG structural biology approaches focusing on non-model membrane proteins whose significance is already appreciated, have immediate scientific impact by employing the "non-model" protein structure as a Rosetta Stone to interpret in structural terms a number of detailed biochemical and biophysical studies collected over years. Similarly, the combined application of multidisciplinary analytical methodologies to unravel the biology and natural history of non-model organisms facilitates the comprehensive integration of proteomics data and findings from diverse research areas across the biological system.

Historically, research communities have focused on model organisms to gain an insight into the general principles that underlie various disciplines, such as genetics, development and evolution. As a consequence, species that are not among the handful of exhaustively studied model organisms have been often ignored due to the lack of tractable genetics, and searchable genomic databases. However, the situation is changing, for the number of sequenced genomes of non-model organisms is accumulating at a faster than ever rate, with no signs of deceleration [16]. Incorporating proteomics analyses in ecology and population studies of non-model organisms, which are often characterized by their unique genomic and phenotypic characteristics, offers an exciting reverse genetic venue towards the study of adaptation, trait evolution, and species divergence, including ontogenetic changes [17] and the impact of non-genetic mechanisms in the evolution of phenotypic diversity within a lineage and the fixation of fitness-related traits [18]; and for addressing a wide range of old evolutionary questions from a new perspective [19]. During the past few years, environmental physiologists and ecotoxicologists have made great progress in applying proteomics to the study of how organisms respond to a changing environment and to pollution. The contributions to a recent symposium on "Comparative Proteomics of Environmental and Pollution Stress" [20] offer the novice a starting point for assessing the potential opportunities and challenges of proteomics to generate novel hypotheses about how organisms interact with their environment.

The current Special Issue of the Journal of Proteomics brings together a collection of proteomics studies carried out in such diverse groups of non-model organisms as plants (*Mangifera indica* and *Malus domestica* fruits; the arsenic hyperaccumulator *Pteris vittata*; the nuclear phosphoproteome of developing chickpea, *Cicer arietinum*, seedlings; mechanisms of environmental stress response of the important fiber crop, *Gossypium* spp. (cotton); and proteomics of seeds and needles of two Mediterranean tree species of agronomic interest, *Quercus ilex* and *Pinus radiata*); marine organisms (including population proteomics of the great scallop, *Pecten maximus*; lipid accumulation under nitrogen starvation in the domesticated oleaginous algae *Tisochrysis lutea*, a non-model organism of major interest for biomass feedstock, food and biofuel production; proteomics of dinoflagellates, producer of an essential component of the food chain in the marine ecosystem, and major causative species of various shellfish poisonings; osmoregulation in the Japanese eel, *Anguilla japonica*, during acclimatization from seawater to freshwater conditions; proteome variance within European whitefish, *Coregonus lavaretus*, populations adapted to different salinity environments; proteomic characterization of the hemolymph of *Octopus vulgaris* to understand the basis of octopus tolerance-resistance to the protozoan parasite *Aggregata octopiana*; proteomics investigation to disclose mechanisms underlying the adaptive response of different organisms to ecological and stress conditions); arthropods (structure and post-translational modifications of a major ampullate silk protein produced by Nephila spiders for use in the construction of the frame and radii of orb webs, and as a dragline to escape from predators; studies of the salivary proteomes of phytotoxic Greenbug, *Schizaphis graminum* Rondani, biotypes; first report on the proteome of the most important *Amblyomma* tick species for their relevance as vectors of zoonotic pathogens worldwide; elucidation of the unexplored biodiversity of ant venom peptidomes, and their application for chemotaxonomy); parasite helminth (elucidation of the surface proteome involved in the interaction of *Dicrocoelium dendriticum* with the host, representing potential targets for intervention against parasitic helminths); prokaryotes (host's environmental signals sensed by a pathogenic strain of *Bacillus anthracis* to regulate its expression of virulence-related genes); bats (proteomics investigation of the strategy adopted by torpid *Myotis ricketti* bats against brain dysfunction during hibernation); and snakes (venomics and functional characterization of the venom of the eastern coral snake, *Micrurus fulvius*, responsible for numerous snake bites in southern United States; overview of the venom proteomes of venomous snakes of Costa Rica, and preclinical antivenomics analysis of the homologous and paraspecific efficacy of a polyspecific antivenom; proteomics characterization of the ontogenetic and activity changes of the Chinese short-tailed pitviper, *Gloydius brevicaudus*, venom; and proteomics assessment of the stability of snake venoms stored for up to eight decades!).

Snake venomics and antivenomics form part of a biology-driven conceptual framework to unveil the genesis and natural history of venoms [13]. Identifying the molecular basis of adaptations in natural populations is an important yet largely unrealized goal in evolutionary biology. Such information is of broad significance because it addresses fundamental questions about the connection between genotype and phenotype for fitness related traits, such as venom, and more explicitly, the relative importance of structural