Proteomics and Drug Discovery

The process of drug discovery within the modern scientific context is quite complex, integrating many disciplines, including structural biology, metabolomics, proteomics, and computer science, just to name a few. The process is generally quite tedious and expensive, given the sheer amount of possibilities of drug-to-target interactions *in-vivo*, and the necessity of successfully passing rigorous pharmacokinetic studies and toxicology assays prior to even being considered for clinical trials. Key components of the drug discovery process include target selection, lead identification, and preclinical and clinical candidate selection.

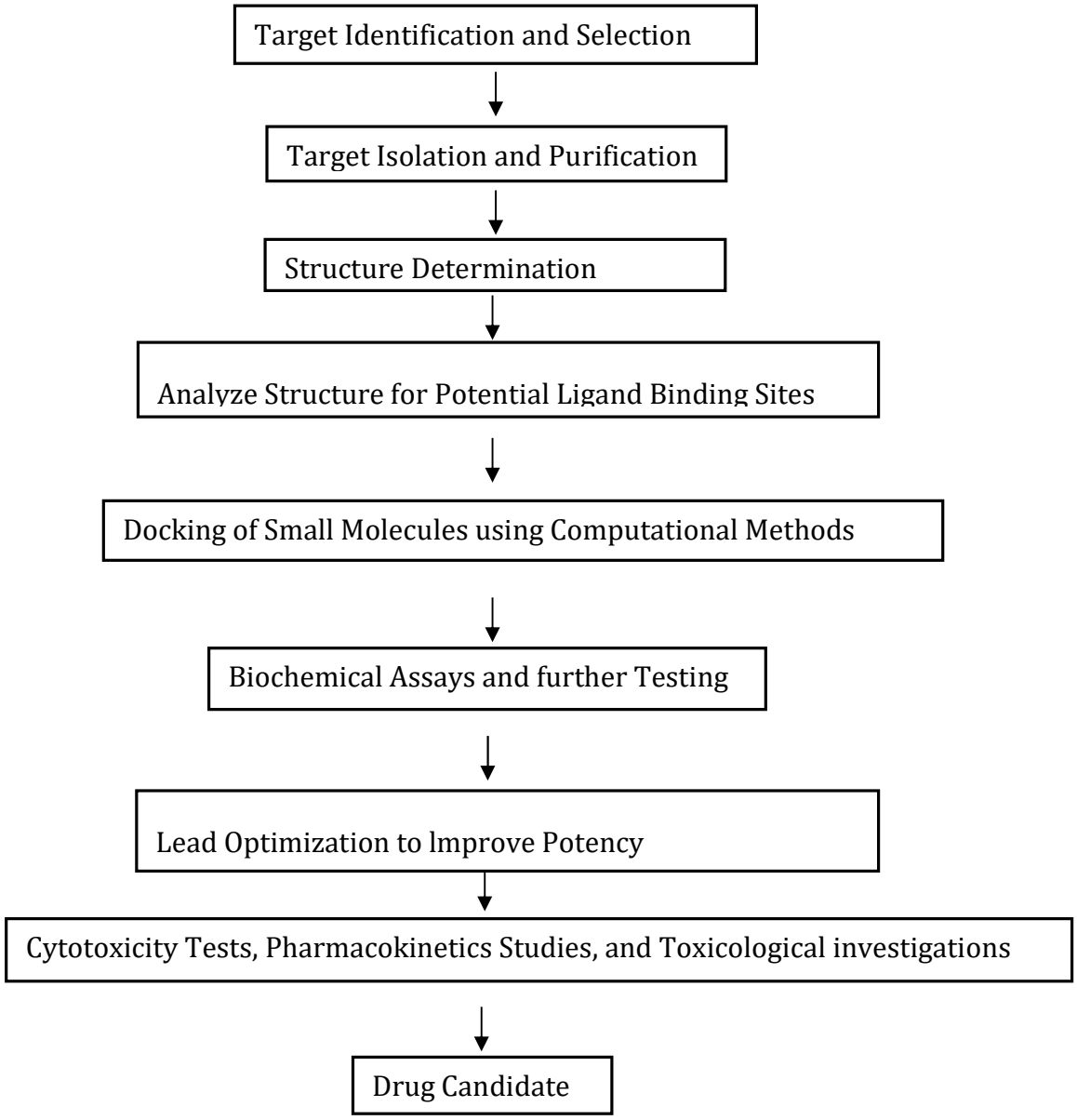
The recent boom of the proteomics field, or the analysis of the ever dynamic organismal proteome, has brought many advances with respect to the very nature of how the current drug discovery process is undertaken. The potential the field of proteomics brings in for identifying proteins involved in disease pathogenesis and physiological pathway reconstruction facilitates the ever increasing discovery of new, novel drug targets, their respective modes of action mechanistically, and their biological toxicology.

The challenge in the drug discovery process is to find the exact causes of an underlying disease and find a way to negate them or bring them to normal levels. A mechanistic understanding of the nature of the disease in question is essential if we are to elucidate any target-specific remedy for it. While the causes of many documented clinical problems vary greatly in their nature and origin, in some cases, the cause is found at the protein level, involving protein function, protein regulation, or protein-protein interactions. One example of such a disorder would be alkaptonuria, characterized by a defect in the gene coding for the enzyme homogentisic acid oxidase inhibiting the metabolism of homogentisic acid to maleylacetoacetic acid, within the phenylalanine degradation pathway. While the underlying cause of this inborn disease is due to a single gene genetic defect, the clinical manifestations, which include excretion of black urine, are a function of the built up of homogentisic acid resulting from a defective [protein] enzyme.

Recent advances in applied genomics helped in the target identification process, since it allowed for high throughput screening of expressed genes. However, studies have shown that there is a poor correlation between the regulation of transcripts and actual protein quantities. The reasons for this are that genome analysis does not account for post-translational processes such as protein modifications and protein degradation. Therefore, the methods employed in the drug-discovery process started to shift from genomics to proteomics. Analysis of the dynamic organismal proteome, as opposed to the static genome, will certainly bring a much more accurate approach to identifying not only applicable biomarkers that will aid in diagnosis, but also effective remedies for diseases of varying origins.

The field of proteomics faces some daunting challenges, in comparison to genomics, for several reasons. First, protein science lacks an analogue of the polymerase chain reaction (PCR), which can generate many copies of a single, native molecule *in vivo* (nucleic acids in the case of PCR). However, several recent approaches have been applied in an effort to ameliorate this quandary. Methods of chemical synthesis exist, being limited by yield, particularly when it comes to synthesizing lengthy peptides. *In- vivo* expression synthesis methods exist as well, however, this approach cannot be applied to producing proteins which may alter normal cellular function. Also, cell-free synthesis ribosome kits can also be employed for accurate and rapid protein synthesis, though the intrinsic presence of ribosome inactivating enzymes contributes to the instability of these systems. Second, in contrast to DNA, protein levels vary significantly

depending on cell type and environment. Third, protein abundance is not directly correlated to protein activity. Protein activity is often determined by post-transcriptional modifications such as phosphorylation. Protein activity, not protein abundance, is of interest in the drug discovery process. Finally, proteins form many interactions with other proteins or small molecules. Elucidation of these interactions would greatly speed up the drug discovery process. One way this is currently being done is through ligand bound x-ray crystallographic studies.



The ideal proteomics technique suited for drug discovery would have the following features: it should be able to separate membrane proteins and detect low abundance proteins, two abilities not quite yet realized, yet required in current separations and analytical techniques. Furthermore, it should be able to identify protein activity independent of protein abundance. It also should reveal protein-protein and protein-small-molecule interactions. This method should

also be implemented easily, be automatable, and perform at high-throughput speed. Proteomics researchers are addressing these issues, and new methods are being developed.

Virtual drug libraries are being developed, both in the public and private sectors. These databases contain potential drug compounds; these compounds may or may not exist outside of a computer database, and new compounds developed through various methods of synthesis are continually added. Methods of modifying existing database entries to create new isomers and derivatives are also used, to more adequately cover a range of potential drug compounds. Docking and scoring are implemented using known and hypothetical drug targets on a protein, coupled with the databases of virtual chemical compounds. In docking, various computational methods are used to position a chemical properly within a protein binding site. Genetic algorithms and Monte Carlo methods are two popular algorithms for evolving an optimum binding position. This process screens for chemicals that are potential drugs, which initially are termed as hits. After docking, scoring is carried out using mathematical models. These models determine the chemical binding strength and energy state of the drug-protein complex. Those hits with high ranking scores are suqsequently subjected to in-vivo tests; hits with positive scores in both areas are then known to be leads.

Evaluation of docked and scored complexes are then made, selecting an arbitrary number of top hits to be further screened manually. The first two steps are done entirely in silico; however, the best complexes now need to be examined using software visualization, often in three-dimensional setups. This allows scientists to ensure that the determined docking orientation looks acceptable, and that the scoring 1s correct based on known interaction energies such as hydrogen bonds and ionic interactions.

The compounds that make it through docking, scoring, and evaluation become drug leads, and are then passed on to undergo drug testing techniques by scientists in a wet lab, to ensure that only compounds with effects relatively unique to the target system and safe to the rest of organism are considered. However, the drug company has already saved much time and money up to this point by having computers do chemical screening, rather than human scientists.

Biomarker Discovery and Drug Development: A Proteomics Approach

A new era of proteomics has dawned owing to the completion and annotation of the human genome and new refinements in the techniques to study proteins on the large scale. Researchers all over the world are applying proteomics to gain a better understanding of disease pathogenesis, to discover new and reliable biomarkers for early detection of diseases and to accelerate drug development.

The dynamic nature of the proteome of cells (diseased) provides ample information for studying a disease at protein level, but to gather all the information from a cell requires implementation of multiple strategies and technologies. Two traditionally used techniques in proteomics are two dimensional polyacrylamide gel electrophoresis (2-DE) and mass spectrometry (MS). Many improvements have been made in these techniques to make them more effective and informative. Moreover, other advanced non-gel-based techniques like protein chip technology, phage display, activity based assays, two hybrid assays, isotope coded affinity tagging are being used in disease proteomics.

Some recently developed strategies for classical proteome analysis are isotope coded affinity tagging (ICAT) and multidimensional protein identification technique (MudPIT). ICAT categorize and quantify all the proteins in the proteome. It is a very sensitive technique with high throughput. MudPIT is another method for classical proteome study. It is the liquid

chromatography/mass spectrometry (LC/MS) method which can be directly applied on crude samples. Functional analysis of proteome involves protein arrays, phage display methods and two- hybrid systems. These techniques quantify proteins and also help in determining protein-protein interactions. Some of them form bridge between classical and functional approach, like activity based probes which analyze the active protein component of proteome.

Proteomics is being extensively used to study molecular basis of various diseases and development of novel drugs with better understanding of targets. Substantial interest has been generated in identifying disease biomarkers. It is a molecule that indicates changes in the physiology of a cell under diseased state and hence can be used as a diagnostic tool, therapy guidance and prognosis monitoring of diseases. Cancer biomarkers are a good example which is not only help in diagnosis of disease but also helps in determining different stages of cancer, studying therapy response and verification of clinical end points. Impressive data can be collected by comparative analysis of disease tissue and its normal counterpart to identify protein with aberrant expression. The sera of the patients can also be screened for auto-antibodies against tumor antigens. Immune response against tumor cells in patients with cancer is being increasingly reported resulting in production of auto-antibodies against various intracellular and surface antigens. Identification of these antigens might help in cancer screening, diagnosis and immunotherapy against the disease. Lung cancer has been extensively studied using proteomic approach. Cytokeratin isoforms were reported to correlate with patient survival and oncoprotein 18 over expression was associated with poor differentiation status in lung carcinoma. Another type of cancer that has been studied is bladder tumor, including transitional cell carcinoma, squamous cell carcinoma, and adenocarcinoma. Researchers have also concentrated on identification of biomarkers of breast cancer, ovarian cancer and colon cancer. A new technology, differential in-gel electrophoresis (DIGE), combined with LC/MS, has been claimed to be a powerful proteomic procedure for the molecular characterization of tumor development and for the detection of tumor- specific biomarkers in esophageal scans cell cancer. Moreover, proteome analysis has also been reported to provide insights into cardiovascular diseases, inflammatory and immune diseases like rheumatoid arthritis and hepatitis.

Proteomics play an important role in drug development. All major pharmaceutical companies are implementing proteomic programs. As majority of drugs act by targeting proteins or they are protein themselves, proteomics along with bioinformatics can meet the needs of pharmaceutical industry in identifying new targets to understand insight of drug action. Bioinformatics is a tool that interprets biological information using computer-aided data. It offers algorithms for gene and protein identification, structure and function relationship predictions and functional interactions among proteins. Information technology enables mining of DNA and protein sequence databases for similarities, screening of active compounds in silico by virtual screening and docking of analyses. Informatics enables easy optimization of leads in drug design, and the selection of pre-clinical candidates. Growth in computing power and systematic databases has made such automated analyses possible.

Proteomic profiling technologies are evolving in a way to emphasize the need of increased sensitivity and high throughput. No technology can provide all the necessary information, so concurrent refinement of a number of techniques will be required for generation and interpretation of data necessary for understanding of processes involved in cell function and regulation. Thus, proteomics, particularly applied to drug discovery and disease proteomics, is evolving toward an increasingly interdisciplinary hunt that combines aspects of biology,

chemistry, engineering and information science. Further improvements in these technologies will continue to drive the pursuit for better diagnostics and effective drug candidates.

APPLICATION OF PROTEOMICS IN BIOMARKER DISCOVERY

Biomarkers are usually disease-associated proteins that can be detected and quantitatively measured for disease diagnosis, staging, prognosis and treatment monitoring.

The development of a disease condition is a multi-step process involving different biological pathways. Many proteins are altered in expression levels and/or expression types such as modification during this process. These altered proteins can be detected in tissue, blood, urine or other body fluids and thus provide indicators for the disease. An ideal biomarker should have high specificity for a certain disease condition; this kind of biomarkers is rare, however. Most of biomarkers are those proteins expressed by many different types of diseases but with variant expression levels from type to type. Combining several unspecific biomarkers together may lead to a specific index for a particular disease. In this regard, proteomics offers suitable and powerful technique platforms for the biomarker discovery, characterization and evaluation because of its capacity of globally examining the protein expression profiles under given conditions. Actually, since they were introduced, proteomics approaches, especially 2- DE/MALDI-TOF and SELDI/ProteinChip, have been extensively used to identify biomarkers for various diseases.

Application of Proteomics Technologies in the Drug Development Process

Proteins are the main targets in drug discovery. Most large pharmaceutical companies now have a proteomics-oriented biotech or academic partner or have started their own proteomics division. Common applications of proteomics in the drug industry include target identification and validation, identification of efficacy and toxicity biomarkers from readily accessible biological fluids, and investigations into mechanisms of drug action or toxicity. Target identification and validation involves identifying proteins whose expression levels or activities change in disease states. These proteins may serve as potential therapeutic targets or may be used to classify patients for clinical trials. Proteomics technologies may also help identify protein–protein interactions that influence either the disease state or the proposed therapy. Efficacy biomarkers are used to assess whether target modulation has occurred. They are used for the characterization of disease models and to assess the effects and mechanism of action of lead candidates in animal models. Toxicity (safety) biomarkers are used to screen compounds in pre-clinical studies for target organ toxicities, as well as later in development during clinical trials. Complementary approaches such as metabolomics and genomics can be used in conjunction with proteomics throughout the drug development process to create more of a unified, systems biology approach. The traditional approach to drug discovery is based on generation of a hypothesis based on a biochemical and pharmacological approach to a disease. Targets are defined on the basis of this hypothesis and lead discovery is a matter of chance. The classical drug-discovery effort also involves isolating and characterizing natural products with some biological activity. These compounds are then ‘refined’ by redesigning their molecular structures to yield new entities with higher biological activity and lower toxicity/side effects. The main limitation of such a process is that the discovery of natural products with defined biological activity is essentially a hit-or-miss approach and therefore lacks a rational basis. Advances in drug discovery include introduction of combinatorial chemistry, which involves the use of high-throughput technologies for preparing a large number of compounds for use in screening of a

variety of biological targets. Genomics was used as a means to improve our understanding of disease with the hope that a comprehensive knowledge of an organism's genetic makeup would lead to more efficient drug discovery. Although useful, DNA sequence analysis alone does not lead efficiently to new target identification, since one cannot easily infer the functions of gene products (proteins) and protein pathways from DNA sequence. Drug discovery is a lengthy and expensive process with shortage of promising drug leads. Functional genomics and proteomics have provided a huge amount of new drug targets. The challenge now is to increase the efficiency of testing lead efficacy and toxicity. In practice, this is not easy because an infinite number of genes, proteins and other molecules interact with each other in signaling pathways to direct cell function. With advances in proteomic technologies, there is an increasing interest in the application of these to improve the drug-discovery process. Because most of the drugs act on proteins, it is important to focus drug-discovery efforts at this level.