Bioinformatics for Traumatic Brain Injury: Proteomic Data Mining

Su-Shing Chen[1,6], William E. Haskins[1,2,4], Andrew K. Ottens[1,2,4], Ronald L. Hayes[2,3,4], Nancy Denslow[1,5], and Kevin K.W. Wang[1,2,3,4]

[1] Center of Neuroproteomics and Biomarkers Research,
University of Florida, Gainesville, FL 32610
[2] Center for Traumatic Brain Injury Studies
University of Florida, Gainesville, FL 32610
[3] Department of Psychiatry,
University of Florida, Gainesville, FL 32610
[4] Department of Neuroscience,
University of Florida, Gainesville, FL 32610
[5] Interdisciplinary Center of Biomedical Research,
University of Florida, Gainesville, FL 32610
[6] Computer and Information Science Engineering,
University of Florida, Gainesville, FL 32610

Summary. The importance of neuroproteomic studies is that they will help elucidate the currently poorly understood biochemical mechanisms or pathways underlying various psychiatric, neurological and neurodegenerative diseases. In this paper, we focus on traumatic brain injury (TBI), a neurological disorder currently with no FDA approved therapeutic treatment. This paper describes data mining strategies for proteomic analysis in traumatic brain injury research so that the diagnosis and treatment of TBI can be developed. We should note that brain imaging provides only coarse resolutions and proteomic analysis yields much finer resolutions to these two problems. Our data mining approach is not only at the collected data level, but rather an integrated scheme of animal modeling, instrumentation and data analysis.

1 Introduction

With the complete mapping of the human genome, we are now armed with a finite number of possible human gene products (human proteome). There are approximately 30,000 to 40,000 hypothetical protein products transcribable from the human genome [2, 25, 26, 31, 32, 60, 61]. The study of the proteome is also aided by recent advances of protein separation, identification and quantification technologies not available even 3-5 years ago. Yet, the proteome is still extremely complex because by definition, proteome is organ-, cell type-, cell state- and time-specific. Proteins are also subjected to various

posttranslational modifications. In addition, cellular proteins are almost constantly subjected to various forms of posttranslational modifications (PTM), including phosphorylation/dephosphorylation by different kinases and phosphatases, proteolysis or processing acetylation, glycosylation and crosslinking by transglutaminases or protein conjugation to small protein tags such as ubiquitin or SUMO [35, 58, 59]. It has been proposed that a more feasible approach is to focus on a subproteome, such as that of single tissue or a subcellular organelle [33]. On the other hand, we proposed that focusing on the study of the proteome of the central and peripheral nervous systems (CNS and PNS) maybe more manageable and productive [17]. We further submitted that although the applications of proteomic technologies to nervous system disorders (e.g. neural injury, neurodegeneration, substance abuse and drug addiction) is still in its infancy, the potential insights one would gain from such endeavors are tremendous. The importance of neuroproteomic studies is that they will help elucidate the currently poorly understood biochemical mechanisms or pathways underlying various psychiatric, neurological and neurodegenerative diseases. The example we will focus on here is traumatic brain injury (TBI), a neurological disorder currently with no FDA approved therapeutic treatment. In general, proteomic studies of TBI create a huge amount of data and the bioinformatic challenge is two-fold: (i) to organize and archive such data into a useful and retrievable database format and (ii) to data-mine such database in order to extract the most useful information that can be used to advance our understanding of the protein pathways relevant to TBI.

This paper reports the bioinformatics component of the TBI research at the Center of Neuroproteomics and Biomarkers Research and Center for Traumatic Brain Injury Studiesat the University of Florida. In particular, we describe data mining strategies for proteomic analysis in TBI research so that the diagnosis and treatment of TBI can be developed. We should note that brain imaging provides only coarse resolutions and proteomic analysis yields much finer resolutions to these two problems. Our data mining approach is not only at the collected data level, but rather an integrated scheme of animal modeling, instrumentation and data analysis. Thus computing infrastructure is essential at all the protein separation, protein identification/quantification and bioinformatics levels.

The organization of the paper is as follows. In Section 2, we describe traumatic brain injury (TBI). Section 3 considers animal models, while Section 4 deals with the source of biological materials. Sections 5 and 6 address samples collection and pooling. Proteomic analysis is overviewed in Section 7. In Section 8, we present bioinformatics for TBI proteomics. Finally in conclusion we consider our future work and the prospect of systems biology in TBI research.

2 Traumatic Brain Injury (TBI)

Traumatic brain injury or traumatic head injury is characterized as a direct physical impact or trauma to the head, causing brain injury [17]. Annually there are 2 million traumatic brain injury (TBI) cases in the U.S. alone. They result in 500,000 hospitalizations, 100,000 deaths, 70,000-90,000 people with long-term disabilities and 2,000 survive in permanent vegetative state. Medical costs of TBI are estimated to be over $48 billion annually in U.S. The cause of TBI can be broken down into the following catalogues: motor vehicle accidents (50%), falls (21%), assault & violence (12%), sports & recreation (10%) and all others (7%) Importantly, 30-40% of all battlefield injuries have a head injury component.

Due to intensive research in both clinical setting and experimental animal models of TBI, there is now a general understanding of the pathology of TBI. It all starts with the impact zone, where there is mechanical compression-induced direct tissue injury and often associated with hemorrhage. Significant amount of cell death will occur very rapidly in this zone. More distal to the injury zone, due to the impact of the force, contusion injury also result, long fiber tracts (axons) are especially at risk to this type of injury. Usually after the first phase of cell injury/cell occurs, there is also the secondary injury which is believed to be mediated by neurotoxic glutamate release (neurotoxicity). Other significant alterations include inflammation responses by microglia cells, astroglia activation and proliferation and stem cells differentiation. Over time, if the TBI patient survives, these events lead to long-lasting brain tissue remodeling. Therefore, the spatial and temporal levels of biochemical and proteomic changes of TBI can be investigated.

3 Animal models of TBI

Over the past decades, basic science researchers have developed several animal models for TBI [19, 54]. There are several well characterized models of TBI, including controlled cortical impact (CCI) with compressed gas control, fluid percussion model that transduce a contusion force due to the movement of fluid in the chamber, and the vertical weight drop model with which a weight is dropped from a certain height within a hollow chamber for guidance. Thus it creates an acceleration force which direct on the top of the skull (either unilateral or bilateral injury [19]). In our work, we employ the rat CCI model of TBI. We have argued that the use of proteomic will greatly facilitate the biochemical mechanisms underlying the various phases of TBI pathology [17].

4 Source of Biological Materials

Proteomic studies for traumatic brain injury can be generally categorized into human studies, animal and cell culture-based studies. For the purposes

of this review, cell culture-based studies will not be discussed further. When comparing human vs. animal studies, there are pro and con in each scenarios. Regarding the sample types that can be exploited for proteomic analysis, they will include brain tissues, cerebrospinal fluid, blood (serum and plasma). For human TBI studies, samples that are the easiest to obtain would be blood samples (which are further fractionated into plasma or serum). Interestingly, there is increasing interests now focus on using cerebrospinal fluid (CSF) as its status will reflect the status of the central nervous system itself. Following severe traumatic brain injury, spinal shun or spinal tap are routinely done thus obtaining the CSF is not an issue. One of the major challenges of using clinical samples-based proteomic studies is that it is extremely difficult to control individual (biological) and environmental variables

(I) Brain Samples

Human brain materials from TBI would inevitably come from deceased TBI patients. These brain samples will be subjected to postmortem artifacts, compounded by various and significant time delay before samples can be obtained.

The biggest advantage of animal neuroproteomic studies over human counterparts is the ability to obtain brain tissues in a controlled laboratory environment. Furthermore, it is possible to harvest samples from defined anatomic regions. For example, for traumatic brain injury studies, we often focus on cortical and hippocampal samples. This is important as different brain regions might be selectively more vulnerable to traumatic or ischemic insults.

(II) CSF

CSF can be collected from the cisterna magna from lab animals, such as rats and mice. CSF contains rich brain proteome information that is relevant to disease diagnosis [16]. However, only about 50-100 ul can be withdrawn from a rat and 25-30 ul from a mouse. Care must also be taken not to contaminate samples with blood due to puncture. While more than one CSF draw might be possible, in our laboratory, we generally withdraw only one CSF sample followed by sacrifice. In the case of human TBI, CSF can also be collected routinely from ventriculotomy or from spinal tap.

(III) Blood Samples (Serum and Plasma)

In both human and animal traumatic brain injury studies, blood can be routinely collected and usually further processed into either serum or plasma fractions before subjecting to proteomic analysis. Like CSF, most proteomic researchers believe there is significant proteomic information in the blood that would reflect the status of the brain, particularly after TBI with possible blood-brain barrier compromise [53, 57].

5 Samples Collection and Processing Consistency

It needs to be emphasized here that for proteomic to be consistent and reproducible, one needs to take extra attentions to ensure the variables can be kept to minimal. All sample collection procedure should be discussed and finalized and the operators made familiar with the procedures. Some practice runs are highly desirable. For human studies, detailed record keeping is extremely important for future analysis or trouble-shooting purposes. For human studies, for example, CSF or blood samples should be taken at consistent intervals and ideally, food consumption might significantly affect blood proteomic profile. For animal studies which are conducted in controlled environment, it should be possible to keep brained and befouled sample collection time and routine as standardized as possible. Also, for animal studies, the animal subjects should be tagged and observed carefully and regularly; with any out-of-the norm observations recorded. They might become very helpful in enhancing proteomic analysis. Both tissue and biofluid samples, once obtained and processed, should be snap frozen and store at -85C until use.

6 Sample Pooling Considerations

There is also an important decision to be made before the proteomic analysis, i.e. whether to pool samples for analysis or analyze individual samples. Pooling samples significantly reduce minor individual variability and reduce the amount of workload. Yet, at the same time, its disadvantage is that it might miss certain proteomic changes that are present in only a subset of samples. On the other hand, analysis of individual samples has the advantage of being an exhaustive analysis of individual proteomic profile but it can be highly time-consuming and cost-prohibiting. If the protein amount in the samples are limiting factor, it would be useful to pool samples. Additionally, if there is a biochemical marker that correlates with TBI (such as alphaII-spectrin breakdown products), it can be used as positive controls for quality assurance and might even be used to guide inclusion criteria for sample pooling [51, 52]. It is also possible to incorporate both pooling and individual proteomic analysis in the same studies. For example, for pilot studies or initial proteomic profiling of TBI, pooled samples can be used while the final detailed analysis can be done with individual samples.

7 Proteomic Analysis Overview

Regardless whether we are dealing with human or animal samples or whether they are tissue lysate or biofluid (CSF, serum or plasma). The strategy we developed can be organized into three interacting scientific disciplines or phases:

protein separation, protein identification and quantification and bioinformatics analysis. By design, any proteomic center should spend two-thirds of its scientific and financial resources to establish robust readily usable proteomic platforms. However, it is equally important for the center to develop new or improve existing neuroproteomics technologies on all fronts.

7.1 Protein Separation Methods

In TBI neuroproteomic studies, we are less interested in descriptive and exhaustive characterization of the whole neuroproteomic, but rather we will focus on protein level or posttranslational changes that occur in TBI. With this in mind, it is important to devise methods in comparing and contrasting the two proteomic data sets: "control" versus "TBI". In order to productively identify all the proteins in a specific system of interest (subproteome) or a subset of proteins that are differentially expressed in TBI, it is essential that complex protein mixtures (such as brain sample or biofluid) be first subjected to multi-dimensional protein separation. Since proteins differ in size, hydrophobicity, surface charges, abundance and other properties, to date there is no single protein separation method that can satisfactorily resolve all proteins in a proteome.

Currently, there are two main stream protein separation methods used for proteomic analysis: (i) 2D-gel isoelectrofocusing/electrophoresis and (ii) multi-dimensional liquid chromatography.

(i) 2-dimensional gel electrophoresis approach

Two-Dimensional gel electrophoresis (2D-gel) is the most established protein separation method for the analysis of a proteome or subproteome [7]. It is achieved by subjecting protein mixtures to two protein separation methods under denaturing condition, in the presence of 6-8 M urea and cationic detergent such as SDS. Traditionally, proteins are first separated based on their PI value with a tube gel (polyacrylamide) by isoelectrofocusing with the aid of mobile ampholytes with different PI values. After IEF, the tube gel is placed atop a polyacrylamide gradient gel within which the SDS-bound proteins are separated by size. Due to poor gel-consistency, the IEF step (the first dimension) is most variable; however, a recent breakthrough in IEF technology utilizing immobilized pH gradient strips (IPG) for 2D-gel analysis provides improved reproducibility [6, 24, 31, 36]. Another disadvantage with 2D-gels is the inevitable gel-to-gel variability in exact location and patterns of protein spots. This proves problematic when comparing two samples directly (such as control vs. substance abuse brain). The recent advance of 2D-differential-in-gel-electrophoresis (2D-DIGE) has resolved this [49, 64]. Two protein mixtures are labeled with the fluorescent cyanine dye pairs Cy3/Cy5 that match in molecular weight and charge but matched have distinct excitation and emission wavelengths. These advantages are incorporated into our approach.

They include in particular the high resolving power for complex mixtures of proteins, and the capability of resolving post-translationally modified proteins, including acetylation, phosphorylation, and glycosylation and protein crosslinking [35, 58]. It is possible to annotate each protein of a proteome by PI and molecular weight values as X-Y coordinates to form a 2D protein map of which there is already a wealth of 2D-brain protein coordinates in publicly accessible and searchable databases [3, 20, 21, 42, 44]. There are however, several persistent weaknesses of 2D-gels. Proteins of extreme PI or minute quantity and proteins that are either very small or very large may be missed. Also, integral membrane proteins of which many are CNS disorder drug targets (membrane-bound receptors or neurotransmitter transporters) are lost due to their extreme hydrophobicity.

Regarding protein separation, there are also research in the direction of microfluidic 2D- protein separation with miniaturized IEF and electrophoresis. This approach is the advantage of reducing waste and sample usage without compromising detection sensitivity [18, 56].

(ii) 2-dimensional liquid chromatography approach

Alternative protein separation methods are needed to overcome some of the shortcoming of 2D-gels. Recently, there is significant movement toward multidimensional liquid chromatography methods to resolve complex protein mixtures [50]. The general idea draws on classic chromatographic principles including size chromatography (SEC) (gel filtration), ionic interaction (strong cation exchange (SCX) and strong anion exchange (SAX), hydrophobic interaction (C4- or phenyl-agarose chromatography), and isoelectrofocusing chromatography. One can envision combining multiple chromatographic approaches in series to achieve multidimensional separations. When selecting chromatographic separation methods, considerations must also been given to take advantage of the size, pI and hydrophobicity differences of the proteins of interests. IN addition, when dealing with membrane-bound proteins, the chromatographic method must be compatible with the use of proper neutral detergent (such as Triton X-100 or CHAPS). Importantly, minute proteins can be further concentrated to enhance their detectability. One weakness of this approach is that even with 2D LC separation, it is often not possible to separate all proteins individually. This problem will be addressed under the "Protein identification and Quantification" section. In summary, when compared to the 2D-gel electrophoresis method, the tandem liquid chromatography method described here is more compatible with membrane-bound proteins as well as can enrich proteins in minute quantity.

7.2 Protein Identification and Quantification Methods

The approach we are taking represents an effort to apply systematically the most contemporary proteomics approaches to identify and develop clinically

useful biomarkers for brain injury from traumatic causes, disease or drugs. Classical methods of protein identification involving protein separation by gels or liquid chromatography coupled to mass spectrometry to provide a potent and novel methodological array never applied systematically before to the detection of biomarkers of CNS injury, either alone or in combination. This integrated strategy makes possible both "targeted" analyses of known potential biomarker candidates as well as "untargeted" searches for novel proteins and protein fragments that could prove even more useful. Each of these technologies has advantages and disadvantages that together are complementary to each other. Thus, multiple proteomic strategies optimize opportunities for successful brain injury proteomic studies. Lastly, protein identification research also benefited from improved bioinformatics tools for protein database searching [9]. Thus, importantly, research designs must incorporate appropriate bioinformatics support.

(i) Mass spectrometry approach

(a) *MALDI-TOF* (matrix-assisted laser desorption ionization mass spectrometry) - time-of-flight (TOF) approach: the most classical method for protein identification in a given protein mixture is to perform 2D gel electrophoresis followed by in-gel digestion of gel band(s) of interest followed by identification of proteins by mass spectrometry. The 2D-gel method has been improved by the use of immobilized pH gradient strips for the first dimension and the ability to label protein samples from control and experimental tissues with Cy dyes (Cy3 and Cy5) that form co-migrating labeled samples that are compared in the same gel. Differentially expressed proteins are easily found, cut from the gel, digested in the gel spot by trypsin, and then identified by MALDI-TOF [5]. It is important to understand that MALDI identifies peptides based on accurate determination of peptide masses since each amino acid has a unique mass and thus any given peptide which is composed of a unique combination of sequence will have a unique mass. However, this method of protein identification is not infallible. Although rare, peptides can have identical amino acid composition with which the order of these amino acid residues could be different. Thus, it is common practice that in order to positively identify the presence of a specific protein, at least two peptide fragments from the protein must be independently identified based on their mass. In addition, any posttranslational modifications when occurs at significantly high tachometric ratio, will make this type of mass prediction extremely difficult. This method is useful for distinguishing proteins that are either up-regulated or down-regulated due to injury, but it is also sub-optimal for finding small peptides from basic, very acidic, or hydrophobic proteins. Complementary to this method are direct mass spectrometry procedures that capture the entire range of proteins and peptides, but may not distinguish proteins that are post-translationally modified, also the maximal protein size is limited to

about 25,000 to 30,000 Delton. This approach is taken advantaged of by a modified MALDI approach called SELDI (invented by Ciphergen) which combines a protein separation phase with the MALDI using an affinity matrix based "Protein Chips" [65].

(b) *LC-MS/MS approach.* There are several 1D- and 2D-chromatography techniques [1] that can substitute for the 2D-gels that give reasonable resolution and include proteins that could be missed by the gel methods. These chromatography techniques can now be coupled to protein fragmentation (trypsinization) and reverse-phase chromatographic peptide separation, which is then coupled in-line with mass spectrometers. The main advantage of the in-line techniques is better recovery of peptides and thus, greater sensitivity. It is now possible to identify proteins that are present in tissues at the pM range. High-powered mass spectrometers including the quadrupole ion-trap (LCQ-Deca), the quadrupole time-of-flight (QSTAR), and the FT-ICR (Bruker BioApex 4.7) mass spectrometers can be used for identifying proteins. These methods work extremely well, especially when coupled with database searching and bioinformatics. Importantly, some of these MS can be configured to become tandem MS. The advantage of tandem MS (MS/MS) is that it can provide peptide sequence information while single MS can also provide peptide mass (see above) [29]. Briefly, in MS/MS, when peptides are ionized at the ion source in the first mass analyzer, selected peptide ions were further ionized in the collision cell. Due to the high energy of ionization inside the collision cells, peptides are actually fragmented randomly along the peptide backbone. Depending on whether the fragmentation site is at the N-terminal or the C-terminal, for each residue site, pair of a- b- and y-daughter ions will be generated. The exact mass of all the b-and y-daughter ions are then determined in the second mass analyzer. Thus, by analyzing this mass information using now available bioinformatic software, the sequence of peptide of interest can be reconstructed without ambiguity.

(ii) Protein and peptide quantification by MS

There are now no less than half a dozen MS-based protein/peptide quantification methods, which are reviewed recently [17]. In this section, we will focus on two most validated quantification methods that are applicable to TBI proteomics.

(a) *ICAT*: A direct chromatographic approach to evaluate differential expression is the use of isotope-coded affinity tags (ICAT) [28]. These tags can be used to label the protein samples on cysteine residues that are then compared mixed together following digested by trypsin. Fragments that are labeled by the tags can then be selectively isolated and analyzed by mass spectrometry. Differential expression is determined by relative peak heights of the two samples, and MS/MS sequencing and database searching directly identify the differentially expressed proteins [50, 66]. This method

is very powerful and quick. We already have experience with this approach and it works well.

(b) *AQUA*: Another innovative method to quantify differential expression is through the use of *A*bsolute *QUA*ntitation (AQUA) probes [22], which involves creating synthetic peptides containing heavy isotopes that can be spiked into the trypsin digest to act as exogenous calibrants for quantitation. For this method to work one must first identify by other means the protein that is differentially expressed, for example 2D-gel electrophoresis coupled to mass spectrometry or ICAT. The calibrant peptide is then synthesized and used within tryptic digests. This is a much quicker way of evaluating the effectiveness of a biomarkers and validating differential expression than waiting a specific antibody to be developed. Our preliminary experience with this method suggests that it will be a powerful way to proceed.

(iii) Antibody panel /array approach

Protein identification is also assisted by the availability of various platforms of antibody arrays or panels (Zyomyx protein Biochips, BD Powerblot and BD antibody arrays] [27, 34, 41, 48]. These methods all rely on antibody-based capturing of protein of interest. The quantification of the captured protein is either achieved by pre-labeling (including differential labeling) of protein with fluorsencent dye (dye-labeling detection), such as BD antibody arrays, similar to the gene chip mRNA quantification method. Alternatively quantitative detection with a second primary antibody specific to the same protein antigen (sandwich detection), similar to the sandwich enzyme-linked immunoabsorbant assay (sandwich ELISA) method (such as the Zyomyx protein chips). Thirdly, the BD Powerblot, as a variant, is in fact a high-throughput western blotting (immunoblotting) system with two distinct protein samples differentially subjected to a set of 5 blots. Each blot has 39 usable lanes with the use of a manifold system. Each lane is developed with 5-6 different fluorophores-linked monoclonal antibodies (toward antigen with non-overlapping molecular weight) Thus with this method, the samples will be probed with a total of 1,000 monoclonal antibodies. We have actually conducted several Powerblot experiments with animal TBI studies.

The major advantage of the antibody panel or array approach is that proteins of interest can be readily identified since all antibodies used have known antigens and their positional assignment on the antibody chip or panel is known Also, quantification is already built-into this antibody-based approach, without any additional effort. On the other hand, the major disadvantage of this approach is that it is practically impossible to be exhaustive as one would only have high fidelity antibodies to a subset of proteins. Furthermore, if antibodies are collected from many different sources it will likely results in uneven detection sensitivity. As in other immunoassay methods (Western blotting, immunostaining or ELISA). It is a given that antibody based method will likely

detect specially bound protein as well as non-specifically bound proteins or other substances. This will likely give rise to high background or false positive reactions or both. The authors believe that despite its shortcoming, the antibody-based protein identification approach is a perfect complement to the MS-based approach discussed above.

8 TBI Proteomic Bioinformatics

The current advance in databases and web portals has a natural convergence for knowledge and data sharing among local and remote scientists in any NIH domain. Large databases will be networked, while web portals will "federate and access" large databases. Such efforts need to develop for the neuroproteomic domain. Neuroscience has one of the most complex information structures - concepts, data types and algorithms - among scientific disciplines. Its richness in organisms, species, cells, genes, proteins and their signal transduction pathways provides many challenging issues for biological sciences, computational sciences, and information technology. The advances in neuroscience need urgently developing portal services to access databases for analyzing and managing information: sequences, structures, and functions arising from genes, proteins, and peptides (e.g. protein segments and biomarkers) [9].

In this bioinformatics component, two interlinked mandates are: (i) to build a local user-friendly proteomic databases, and (ii), to develop interoperable proteomic tools and architecture for multiple data integration and to integrate user and public domain-based databases. Data analysis applications should be interoperable with database operations and portal access. The TBI proteomics core technologies will provide an integrative approach to genomic and proteomic information by developing a common portal architecture, the TBI proteomics portal, at the University of Florida for data archiving and retrieval among core researchers and end users, and data linking and sharing to national and international neuroproteomic websites (e.g., Human Proteome Organization (HUPO, USA) [30] and Human Brain Proteome Project (Germany) [47]). (iii) Lastly, bioinformatics tools and software are also needed to enhance our ability to mine data, as well as to study protein-protein interaction, protein pathways and networks and complex post-translational modification such as (protein phosphorylation, processing, crosslinking and conjugation). This will help us develop knowledge bases about neuroproteomic functions and signal transduction pathways in terms of dynamic objects and processes [45, 46, 55, 62]. In addition, clinical information should be integrated with genomic and proteomic databases. The following diagram depicts the neuroproteomics bioinformatics core:

The three major functions of the bioinformatics component of TBI proteomic research can be further explained as follows:

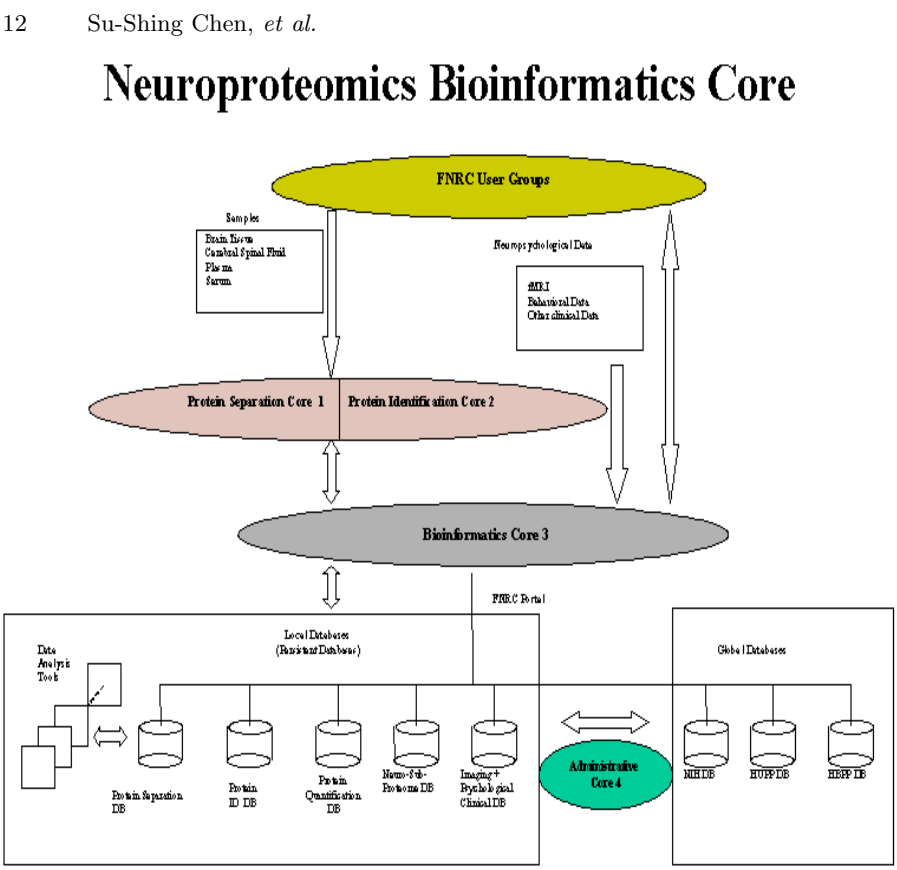


Fig. 1. Neuroproteomics Bioinformatics Core

(i) Permanence

Permanence is defined here as developing local databases for proteomics separation and identification, and link with national and international data sources. Local databases will include chromatograms, mass spectra, gel images, peptide and protein sequences, and fMRI images for control and diseased samples. Data modeling and semantics will be developed by proteomics and computer scientists together so that semantic equivalence of search attributes and semantic associations can be established.

Our Bioinformatics Core is in the process of combining different data semantics and knowledge trees in separate genomic, proteomic, and clinical databases. Our main contribution will be the development of data modeling and semantics by proteomics and computer scientists together so that semantic equivalence of search attributes and semantic associations can be established. A key requirement is the development of semantics (or ontology) of biological information, which are then captured in two components - semantic indexing and meta-information - of the intelligent search engines. A recent book

of S. Chen [15] has described these two important methods. Semantic indexing extends the existing full-text, hypertext, and database indexing schemes to include semantics or ontology of information content. Meta-information (or metadata) means "information about information" concerning content, context, and archival description. Both semantic indexing and meta-information are necessary to the semantic equivalence in intelligent search engines. Meta-information contains information about not only individual objects but also whole data collections and even resources. It provides collection-wide semantics to organize a widely distributed collection of information resources better. Furthermore both semantic indexing and meta-information complement each other, reduce the complexity of neural taxonomy and classification, and correlate semantically the proteomic types and phenotypes (e.g. behavior in drug abuse) at various (subcellular, cellular, and tissue or fluid) levels of neural activities. Dissemination to national and international data sources (e.g. HUPO-USA and HBPP-Germany) will be consistently maintained through our intelligent search engines.

(ii) Interoperability

Interoperability is defined here as integrating existing data analysis tools with local databases. A proteomic problem-solving environment will be established to provide users with rapid access to TBI neuroproteome center databases and analysis tools. This will include existing tools for proteomics research and drug abuse research. The range of these tools is very broad, from peptide sequencing and protein identification to image processing for fMRI images and data analysis for neuropsychological tests and diagnosis.

A critical component of our Bioinformatics Core will be distributed search at widely distributed resources of data analysis and multiple levels of proteomic clinical and behavior information. The distributed collections of heterogeneous information resources will be large-scale. The intelligent search engines are beyond the capability of current web search engines and protocols. A distributed information retrieval system, Emerge, has implemented some aspects of semantic indexing and meta-information of NIH's PubMed and Entrez databases, in a collaboration with NCSA of UIUC. The TBI neuroproteome center distributed information retrieval component is a set of search engines extending Emerge. Such an intelligent search engine should allow nomenclature, syntactic, and semantic differences in queries, data, and meta-information. It should permit type, format, representation, and model differences as well in databases. In our TBI neuroproteome research, we have to compare information among proteomic and clinical data, such as chromatograms, mass spectra, gel images, peptide and protein sequences, and fMRI images. This intelligent search engine must go to different databases to retrieve various data of potentially different types, formats, representations, and models. In an asynchronous way, data are compared to an abstract and conceptual schema for neuroscience domains. The object-oriented data modeling helps us to establish these mappings between the abstract and conceptual

schema and different database schemas. Due to the diverse nature of neuroscience information, we will need a set of interoperable search engines to guide users finding information of various domains, formats, types, and levels of granularity (e.g. peptide, protein, cell, and system levels). Since some abstract and conceptual schema has been developed for neuroscience domains, we will need a set of interoperable search engines for a wider set of analysis tools and databases.

Interoperability with analysis tools will be an important component. The starting point will provide a point and click interface for rapid access of neuroscience databases and analysis tools. The interoperability of databases and analysis tools will establish a proteomic problem-solving environment. Thus users of the problem-solving environment will also be factored into the interoperability. Whatever users need - small vs. large data sets, interactive vs. batch computation - will require design and implementation of data and event services. For the current research, we intend to develop a neuroproteomic workbench to gather a collection of data analysis tools for neuroproteomics as well as TBI neuroproteomic data sets (see data samples below) :

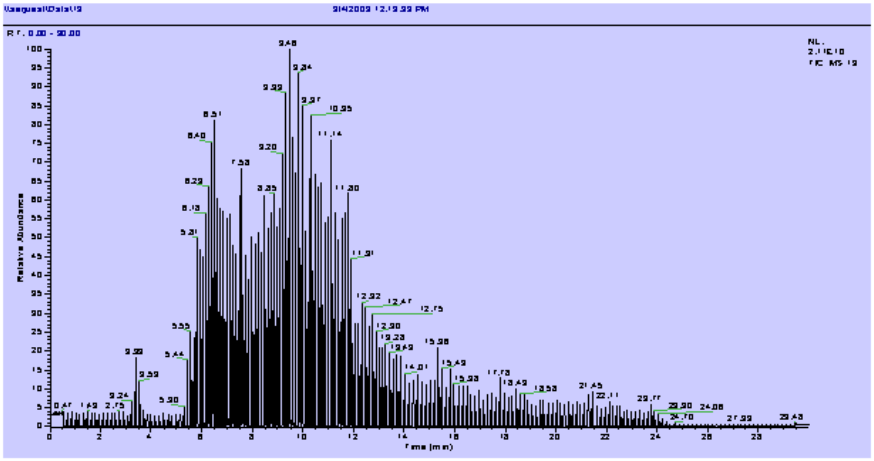
(1) Peptide sequencing and protein identification by MALDI-TOF-MS and capillary LC-MS/MS [43, 63].
(2) Protein peak patterns and single protein retention time from 1D or 2D-ion exchange or size exclusion chromatograms.
(3) Protein database searching algorithms such as SEQUEST [67].

The integration of databases with proteomic computational algorithms will be based on the object-oriented data modeling and data semantics discussed earlier. The ODMG compliant data analysis and databases are highly relevant to the Common Component Architecture [8]. In high throughput computing, in terms of parallel or multi-threaded objects, components (data and algorithms alike) may be distributed over a wide area grid of resources and distributed services.

(iii) Data Mining

Our neuroproteomic initiative has placed significant effort in new data mining and analysis tools for differential protein expression, protein network and modification analysis and validation. A unique data-mining workbench will be created to explore protein network and pathways underlying the pathobiology of TBI from a neuroproteomic perspective. Novel data-mining tools will include a differential analysis tool for research on proteins and protein fragments involved in TBI and construction of cognitive maps [4, 40, 68, 69], a graphical network method to represent knowledge and information. Furthermore, the cognitive maps will be used for TBI-induced Differential Neuroproteome Validation and possible brain injury diagnosis and severity monitoring. These data mining steps are described in the following:

Capillary LC separation of tryptic peptides

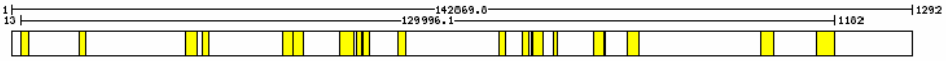


Representative CLC chromatogram of tryptic peptides from the unlabeled injured (I) sample in gel slice 19.

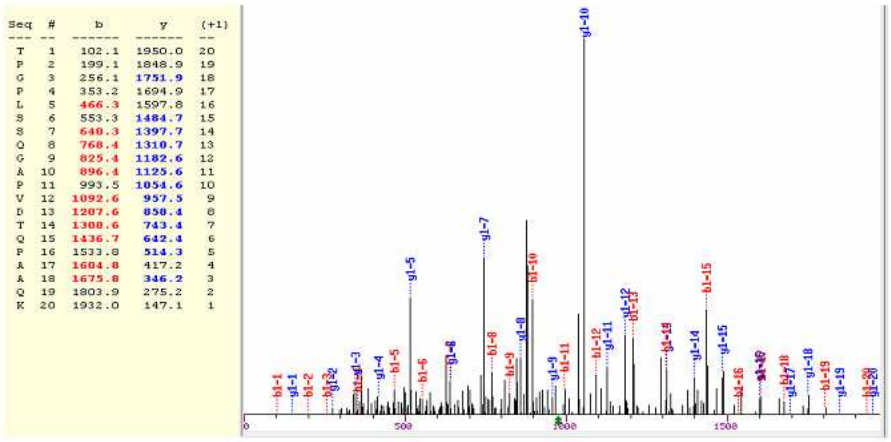
Tryptic Peptides Observed in the 1st- ranked protein (Synaptojanin) in Gel Slice 22: Injured-Cy3 (I) HC

MAFSKGFRIY HKLDPPPFSL IVETRHKEEC LMFESGAVAV LSSAEKEAIK GTYAKVLDAY GLLGVLRLNL GDTMLHYLVL
VTGCMSVGKI QESEVFRVTS TEFISLRVDA SDEDRISEVR KVLNSGNFYF AWSASGVSLD LSLNAHRSMQ EHTTDNRFFW
NQSLHLHLKH YGVNCDDWLL RLMCGGVEIR TIYAAHKQAK ACLISRLSCE RAGTRFNVRG TNDDGHVANF VETEQVIYLD
DCVSSFIQIR GSVPLFWEQP GLQVGSHRVR MSRGFEANAP AFDRHFRTLK DLYGKQIVVN LLGSKEGEHM LSKAFQSHLK
ASEHASDIHM VSFDYHQMVK GGKAEKLHSV LKPQVQKFLD YGFFYFDGSA VQRCQSGTVR TNCLDCLDRT NSVQAFLGLE
MLAKQLEALG LAEKPQLVTR FQEVFRSMWS VNGDSISKIY AGTGALEGKA KLKDGARSVT RTIQNNFFDS SKQEAIDVLL
LGNTLNSDLA DKARALLTTG SLRVSEQTLQ SASSKVLKNM CENFYKYSKP KKIRVCVGTW NVNGGKQFRS IAFKNQTLTD
WLLDAPKLAG IQEFQDKRSK PTDIFAIDFE EMVELNAGNI VNASTTNQKL WAVELQKTIS RDNKYVLLAS EQLVGVCLFV
FIRPQHAPFI RDVAVDTVKT GMGGATGNKG AVAIRMLFHT TSLCFVCSHF AAGQSQVKER NEDFVEIARK LSFPMGRMLF
SHDYVFWCGD FNYRIDLPNE EVKELIRQQN WDSLIAGDQL INQKNAGQIF RGFLEGKVTF APTYKYDLFS EDYDTSEKCR
TPAWTDRVLW RRRKWPFDRS AEDLDLLNAS FQDESKILYT WTPGTLLHYG RAELKTSDHR PVVALIDIDI FEVEAEERQK
IYKEVIAVQG PPDGTVLVSI KSSAQENTFF DDALIDELLQ QFAHFGEVIL IRFVEDKMWV TFLEGSSALN VLSLNGKELL
NRTITITLKS PDWIKTLEEE MSLEKISVTL PSSTSSTLLG EDAEVSADED MEGDWDDYSA EVEELLPQHL QPSSSSGLGT
SPSSSPRTSP CQSPTAPEYS APSLPIRPSR APSVTPGPLS SQGAPVDTQP AAQKESSQTI EPKRPPPPRP VAPPARPAPP
QRPPPPSGAR SPAPARKEFG APKSPGTARK DMGRNQFSI QAGDAGFGFS GTGAARPTIP ARAGVISAPQ SQARVSAGRL
TPESQSKPLE TSKGPAVLPE PLKPQAAFPP QPSLPTPAQK LQDPLVPIAA PMPPSIPQSN LETPPLPPPR SRSSQSLPSD
SSPQLQQEQP TG

Mass (mono): 142869.8 **Identifier:** gi|1166575 **Database:** C:/Xcalibur/database/rat.fasta
Protein Coverage: 262/1292 = **20.3%** by amino acid count, 28224.7/142869.8 = **19.8%** by mass



1st-ranked Tryptic Peptide MS-MS analysis in Gel Slice 22 (Synaptojanin): Injured-Cy3 (I) HC



Seq	#	b	y	(+1)
T	1	102.1	1950.0	20
P	2	199.1	1848.9	19
G	3	256.1	1751.9	18
P	4	353.2	1694.9	17
L	5	466.3	1597.8	16
S	6	553.3	1484.7	15
S	7	640.3	1397.7	14
Q	8	768.4	1310.7	13
G	9	825.4	1182.6	12
A	10	896.4	1125.6	11
P	11	993.5	1054.6	10
V	12	1092.6	957.5	9
D	13	1207.6	858.4	8
T	14	1308.6	743.4	7
Q	15	1436.7	642.4	6
P	16	1533.8	514.3	5
A	17	1604.8	417.2	4
A	18	1675.8	346.2	3
Q	19	1803.9	275.2	2
K	20	1932.0	147.1	1

Tryptic peptide ID: TPGPLSSQGAPVDTQPAAQK

Fig. 2. Observed Data of Tryptic Peptides

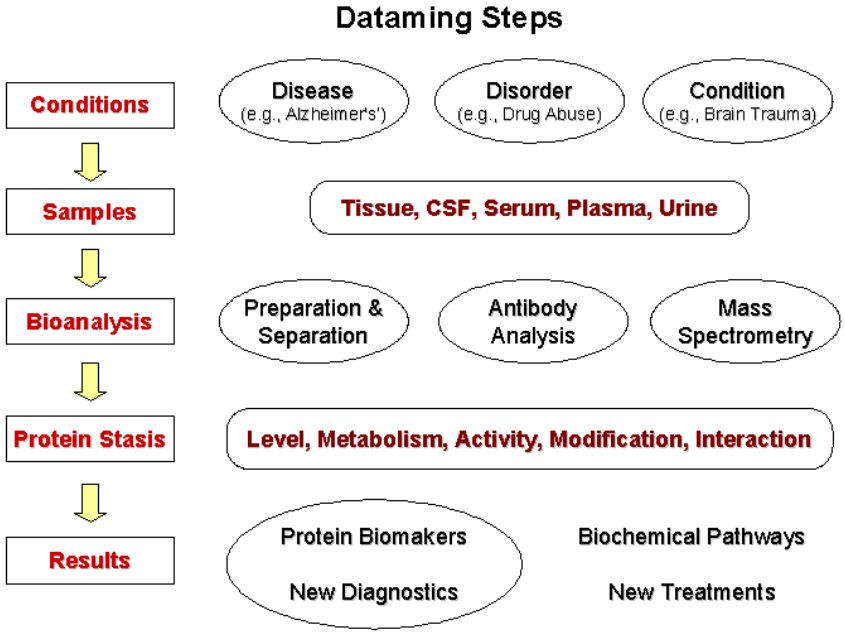


Fig. 3. Data Mining Steps

a) Creating Cognitive Maps for TBI-induced differential proteome

New data mining tools for TBI-induced differential proteome analysis and validation are being developed at our center. There are three major zones of neuroproteomics information (i) pathophysiological stasis (including TBI, other CNS injuries, such as ischemic stroke, aging, environmental toxin or substance abuse-induced brain injury, neurodegenerative diseases such as Alzheimer's disease or Parkinsonism), (ii) neuroproteome stasis (such as differential protein expression, protein synthesis and metabolism, alternative mRNA splicing and RNA editing, protein-protein interaction, enzymatic activity or protein functions) post-translational modifications (such as protein crosslinking, acetylation, glycosylation protein proteolysis and processing, phosphorylation) and protein-protein interaction networks and signal transduction pathways and (iii) sources of neuroproteomic data (brain tissue from different areas or anatomical regions of the brain, such as hippocampus), biological fluids such as the cerebrospinal fluid (CSF), blood samples (including plasma and serum) where brain proteins stasis might be reflected upon via diffusion-based equilibrium or blood brain barrier compromise (e.g. from brain to CSF to blood).

Collection of data from these three components will enable the construction of multiple cognitive maps [4, 40, 68, 69]. For instance, cognitive maps can be constructed for the TBI-induced differential proteome in the following figure. Automated reasoning and knowledge discovery algorithms on the cognitive maps [10, 11, 12, 13, 14, 15, 39] will distill the information and present the knowledge gained from a systems biology perspective. Thus, cognitive maps will enable the brain trauma researchers to gain a greater understanding of the entire TBI-induced differential neuroproteome and hopefully the mechanistic protein-pathways of TBI.

b) Using Cognitive Maps for TBI-induced Differential Neuroproteome Validation

A statistical analysis tool is also being developed for TBI-induced differential neuroproteome validation and possible TBI protein-pathways elucidation. For example, up- or down-regulation of multiple proteins and protein fragments in control and injured samples will be quantified by ICAT, AQUA, or ELISA to validate differential TBI neuroproteome. Linear discriminant analysis (LDA) will be used to calculate the probability of a correct diagnosis given the number of injury-specific biomarkers measured the number of samples, etc. Thus, statistical analysis tools are expected to provide an important component for all the neuroproteomics research conducted at our neuroproteomic center. These statistical analysis data will be fed into the cognitive maps to reach decision on

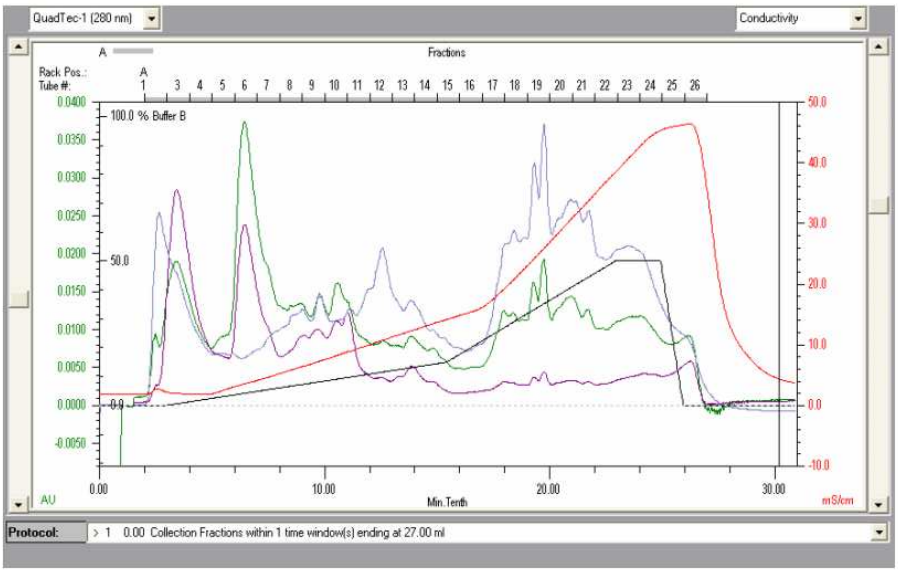


Fig. 4. Input Data for Neuroproteomics Cognitive Maps

diagnosis, monitoring and treatment. We have both statistical/probabilistic and fuzzy reasoning capabilities in our cognitive maps [40, 68, 69].

Cognitive maps are directed graphs representing relations (by links) among concepts/attributes (by nodes). Cognitive maps include several knowledge representation schemes. Semantic networks or frames form a special class of cognitive maps. Inference networks and causal networks form other classes of cognitive maps. In cognitive maps, link weights may be assigned to relations representing their compatibility degrees, and node values may be assigned to concepts and attributes representing relevance factors. A hierarchical cognitive map consists of several cognitive maps, each of which represents gene network interaction or metabolic pathway. The knowledge bases of hierarchical cognitive maps will effectively capture the complex behavior of biological systems. A hierarchical cognitive map is alternatively represented as a large cognitive map combining several individual ones in the following diagram.

Cognitive maps can extend to probabilistic, or fuzzy cognitive maps, and further to neural network learning maps. These numerically enabled cognitive maps can be interfaced with other numerical simulation packages in biology.

Now we briefly describe the relaxation computation in a cognitive map. Let Σ be a collection of biological objects $\{x_1,...,x_n\}$ (e.g., gene sequences, protein structures, metabolites, genotypes, and phenotypes), and let Λ be a collection of labels $\{\lambda_1,...,\lambda_m\}$ with any mathematical structure (e.g., concentrations and intensities). The labeling problem is to find a consistent labeling of biological objects in Σ by Λ, given a set of relations among objects and

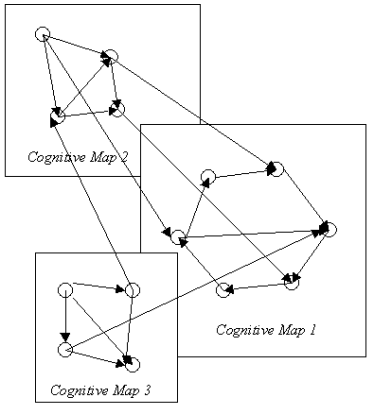


Fig. 5. Hierarchical Cognitive Maps

a set of constraints among objects and their labels. For each x_i, let Λ_i be a subset of Λ that is compatible with x_i. For any pair $\{x_i,x_j\}$ of objects (i, j distinct), let Λ_{ij} be a subset of compatible pairs of labels in $\Lambda_i \times \Lambda_j$. A labeling $L = \{L_1,...,L_n\}$ is an assignment of a set of labels Λ_i in Λ to each x_i. L is consistent if for each i, j and all λ in Λ_i, $(\{\lambda\} \times \Lambda_j)$ intersects with Λ_{ij}. L is unambiguous if it is consistent and assigns only a single label to each object. The semantic labeling of cognitive maps determines the results of TBI-induced differential neuroproteome validation. The semantic labeling is to assign a measure $m_i(\lambda)$ to the statement "λ is the correct label of x_i". An arbitrary labeling of a neuroproteome may not be consistent and unambiguous, because the constraint satisfaction is required among either objects or a combination of new input evidences. The interaction with external users and systems is through a query system. At the initial stage, the $m_i(\lambda)$ is either estimated by the user or is provided by another cognitive map or simulation tool. Now the initial measures go through a constraint satisfaction checking by the label relaxation, which iterates the process until the convergence to final measures is reached. The final measures are sent back to the query subsystem for either clinical decision or further data analysis.

The relaxation scheme is mathematically described as follows. An initial assignment of measures $\{m_i(0)(\lambda)\}$ to $\{x_i\}$ is given at time 0. A relaxation operator R is defined to transform one set $\{m_i(k)(\lambda)\}$ of measures to another set $\{m_i(k+1)(\lambda)\}$. The limit $\{m_i^*(\lambda)\}$ of $\{m_i(k)(\lambda)\}$ gives the unambiguous labeling under compatibility constraints, as k approaches to infinity. In reality, we expect the limit to be attained after a finite number of iterations. In practice, the limit $\{m_i^*(\lambda)\}$ may not be unique (we are not always getting an unambiguous labeling). The multiple labelings are sent back to the users so that they can select an appropriate result for further analysis.

There are several ways to define the relaxation operator R. A relaxation operator R should produce $m_i(k+1)(\lambda)$ from the combination of

$m_i(k)(\lambda)$ and support $s_i(k)(\lambda)$ by some update equations, where $s_i(k)(\lambda) = \Sigma r_{ij}(\lambda s \lambda')m_j(k)(\lambda')$, where $r_{ij}(\lambda, \lambda')$ is the compatibility function of "label λ is assigned to x_i and label λ' is assigned to x_j", and j-indices are indices of all source nodes leading to the i-th node. A relaxation operator R is defined by the following update equations:

$$m_i(k+1)(\lambda) = min[1, max(0, m_i(k)(\lambda) + s_i(k)(\lambda))],$$

$$s_i(k)(\lambda) = \Sigma(r_{ij}(k)(\lambda, \lambda') + \Delta r_{ij}(k)(\lambda, \lambda'))m_j(k)(\lambda'),$$

$$\Delta r_{ij}(k+1)(\lambda, \lambda') = a_{ij}\Delta r_{ij}(k)(\lambda, \lambda') + b_{ij}m_i(k+1)(\lambda)m_j(k)(\lambda'),$$

where a_{ij} and b_{ij} are learning parameters. The first equation makes sure that $m_i(k+1)(\lambda)$ stays between 0 and 1. The second equation provides the network input to the (i, λ)-th node. The third equation includes the Hebbian learning rule.

9 Conclusion

In summary, proteomic studies of both human and rat traumatic brain injury, if approached systemically, is a very fruitful and powerful analytic technology. In order to obtain a comprehensive TBI neuroproteome data set, it is important to integrate multiple protein separation and protein identification technologies. Equally important is the optimization of individual protein separation identification methods. Bioinformatics platform then becomes the critical adhesive component by serving two purposes: (i) integrating all proteomic data sets and other relevant biological or clinical information, and (ii) inferring and elucidating the protein-based pathways and biochemical mechanisms underlying the pathobiology of TBI and identifying and validating biomarkers for the diagnosis and monitoring of TBI [23]. Ultimately, if we are to be successful in doing these, the TBI proteomic approach outlined here must be further integrated with genomic, cytomics as well as systems biology approaches [37, 38].

References

1. Adkins JN, Varnum SM, Auberry KJ, Moore RJ, Angell NH, Smith RD, Springer DL, Pounds JG. (2002) Toward a human blood serum proteome: analysis by multidimensional separation coupled with mass spectrometry. Mol Cell Proteomics. 1:947-55.
2. Aebersold, R.; Watts, J. D. (2002) The need for national centers for proteomics. *Nature Biotech.* 20 (7): 651-651.

3. Appel, R.D., Bairoch, A., and Hochstrasser, D.F. (1999) 2-D Databases on the World Wide Web in Methods in Molecular Biology, Vol 112: *2-D Proteome Analysis Protocols* (A.J. Link, Ed) Humana Press Inc, Totowa, NJ, pp 383-391.
4. Axelrod, R. (1976) Structure of Decision, Princeton University Press, 1976.
5. Bienvenut WV, Deon C, Pasquarello C, Campbell JM, Sanchez JC, Vestal ML, Hochstrasser DF. (2002) Matrix-assisted laser desorption/ionization-tandem mass spectrometry with high resolution and sensitivity for identification and characterization of proteins. Proteomics. 2002 Jul;2(7):868-76.
6. Bjellqvist, B.; Ek, K.; Righetti, P.G.; Gianazza, E.; Gorg, A.; Westermeier, R. (1982) Isoelectric focusing in immobilized pH gradients: Principle, methodology and some applications. *J. Biochem.Biophys Methods* 6: 317-339.
7. Boguslavsky, J. (2003) Resolving the proteome by relying on 2DE methods. Drug Discovery Develop. 6(7), 57-60.
8. Cattell R. G. G. (Ed.), (1996) The Object Database Standard: ODMG-93, Morgan Kaufmann, 1996.
9. Chakravarti, D.N., Chakrarti, B., Moutsatsos, I. (2003) Informatic tools for proteome profiling. High throughput proteomics; protein arrays, a supplement to BioTechnoques March 2003, 4-15.
10. Chen, (1988) Knowledge acquisition on neural networks, Uncertainty and Intelligent Systems, edited by B. Bouchon, L. Saitta and R. R. Yager, Lecture Notes in Computer Science, Springer-Verlag, Vol. 313, 1988, pp. 281-289.
11. Chen, S. (1986) Some extensions of probabilistic logic, Proc. AAAI Workshop on Uncertainty in Artificial Intelligence, Philadelphia, Penn., August 8-10, 1986, 43-48; An extended version appeared in Uncertainty in Artificial Intelligence, Vol. 2, edited by L. N. Kanal and J. F. Lemmer, North-Holland.
12. Chen, S. (1987), Automated reasoning on neural networks: A probabilistic approach, IEEE First International Conference on Neural Networks, San Diego CA, June 21-24, 1987.
13. Chen, S. (2000a) Knowledge discovery of gene functions and metabolic pathways, IEEE BioInformatic and Biomedical Engineering Conference, Washington DC, November 2000.
14. Chen, S. (2000b) Knowledge representation for systems biology, First International Conference on Systems Biology, Tokyo Japan, November 14-16, 2000, 107-112.
15. Chen, S. (1998) Digital Libraries: The Life Cycle of Information, BE Publisher, 1998.
16. Davidsson P, Westman-Brinkmalm A, Nilsson CL, Lindbjer M, Paulson L, Andreasen N, Sjogren M, Blennow K. (2002) Proteome analysis of cerebrospinal fluid proteins in Alzheimer patients. Neuroreport. 2002 Apr 16;13(5):611-5.
17. Denslow, N.D., Michel, M.E., Temple, M.D., Hsu, C., Saatman, K., and Hayes, R.L. (2003) Application of Proteomics Technology to the Field of Neurotrauma, J. Neurotrauma, 20: 2003, 401-407.

18. Derra, S. (2003) Lab-on-a-chip technologies emerging from infancy. Drug Discovery Develop. 6(5), 2003, 40-45.
19. Finnie J. (2001) Animal models of traumatic brain injury: a review. Aust Vet J. 2001 Sep;79(9):628-33.
20. Fountoulakis, M.; Hardmaier, R.; Schuller, E.; Lubec, G. (2000) Differences in protein level between neonatal and adult brain. *Electrophoresis.* 21(3): 673-678.
21. Fountoulakis, M.; Schuller, E.; Hardmeier, R.; Berndt, P.; Lubec, G. (1999) Rat brain proteins: two-dimensional protein database and variations in the expression level. *Electrophoresis.* 20(18): 3572-3579.
22. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A. 100:6940-5.
23. Goldknopf, I., Park, H.R., Kuerer, H.M. (2003) Merging diagnostics with therapeutic proteomics. IVD Technology 9(1) 1-6,
24. Gorg, A.; Postel, W.; Gunther, S. (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 9: 531-546.
25. Grant SG, Blackstock WP. (2001) Proteomics in neuroscience: from protein to network. J Neurosci. 2001 Nov 1;21(21):8315-8.
26. Grant, S.G.N. and Husi H. (2001) Proteomics of multiprotein complexes: answering fundamental questions in neuroscience. Trends Biotechnol. 2001 Oct;19(10 Suppl):S49-54.
27. Graslund S, Falk R, Brundell E, Hoog C, Stahl S. (2002) A high-stringency proteomics concept aimed for generation of antibodies specific for cDNA-encoded proteins. Biotechnol Appl Biochem. 2002 Apr;35(Pt 2):75-82.
28. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 17:994-9.
29. Gygi, S.P. et al. (2000) Mass spectrometry and proteomics. Analytical techniques, 4: 489-494.
30. Hanash S. (2004) HUPO initiatives relevant to clinical proteomics. Mol Cell Proteomics. 2004 Jan 27
31. Hanash, S. M. (2003) Disease proteomics. *Nature* 422(6928):226-232.
32. Hochstrasser, D. F.; Sanchez, J. C.; Appel, R. D. (2002) Proteomics and its trends facing nature's complexity. *Proteomics.* 2(7): 807-812.
33. Husi,, H.; Grant, S. G. (2001) Proteomics in neuroscience: from protein to network. *J. Neurosci.* 21(21): 8315-8318.
34. James, P. (2002) Chips for proteomics; a new tool or just hype. High throughput proteomics; protein arrays, a supplement to BioTechniques Dec 2002, 14-23.
35. Janssen, D. (2003) Major approaches to identifying key PTMs *Genomics and Proteomics* 3(1): 38-41.
36. Jungblut, P.; Thiede, B.; Zimny-Arndt, U.; Muller, E.C.; Scheler, C.; Wittmann-Liebold, B. (1996) Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis.* 17: 839-847.
37. Kitano H. (2002a) Systems biology: a brief overview. Science. 2002 Mar 1;295(5560):1662-4. Review.

38. Kitano H. (2002b) Computational systems biology. Nature. 2002 Nov 14;420(6912):206-10. Review.
39. Kitano, H. (2000) Perspectives on systems biology, "New Generation Computing", Vol.18, No.3, Ohm-sha, Springer-Verlag, New York Inc., 2000.
40. Kosko,B. (1986) Fuzzy cognitive maps, Int. Journal of Man-Machine Studies, 24, 1986, pp. 65-75.
41. Kusnezow, W. and Hoheisel, J.D. (2002) Antibody microarrays: promises and problems. High throughput proteomics; protein arrays, a supplement to BioTechniques Dec 2002, 14-23.
42. Lemkin, P. F. (1997) Comparing Two-dimensional Electrophoretic Gel Images Across the Internet. *Electrophoresis*. 18: 2759-2773.
43. Lu B, Chen T. (2003) A suffix tree approach to the interpretation of tandem mass spectra: applications to peptides of non-specific digestion and post-translational modifications. Bioinformatics. 2003 Oct;19 Suppl 2:II113-II121.
44. Lubec, G.; Krapfenbauer, K.; Fountoulakis, M. (2003) Proteomics in brain research: potentials and limitations. Proteomics in brain research: potentials and limitations. *Prog Neurobiol.* 69(3): 193-211.
45. Mendes, P. (1993) and control of biochemical and other systems, Computer Applications Biosciences, 9, 1993, pp. 563-571.
46. Mendes, P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3, Trends in Biochemical Sciences, 22, 1997, pp. 361-363.
47. Meyer HE, Klose J, Hamacher M. (2003) HBPP and the pursuit of standardisation. Lancet Neurol. 2003 Nov;2(11):657-8.
48. Moody, M.D. (2001) Array-based ELISAs for high-throughput analysis of human cytokines. Biotechniques 31:186-194.
49. Patton, W. F. (2002) Detection technologies in proteome analysis. *J Chromatogr B Analyt Technol Biomed Life Sci.* 771(1-2): 3-31.
50. Peng, J., Elias,J.E., Thoreen, C.C., Licklider, L.J. Steven P. Gygi (2003) Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large scale protein analysis: The yeast proteome. J. Proteome research 2, 43-50.
51. Pike BR, Flint J, Dave, J.R., Lu, X.-C. M, Wang, K.K.W., Tortella, F.C., Hayes R.L. (2004) Accumulation of calpain and caspase-3 proteolytic fragments of brain-derived αII-spectrin in CSF after middle cerebral artery occlusion in rats. J. Cereb. Blood Flow Metab. 24(1): 98-106.
52. Pike, B. R.; Flint, J.; Dutta, S.; Johnson, E.; Wang, K. K. W.; Hayes, R. L, (2001) Accumulation of non-erythroid aII-spectrin and calpain-cleaved aII-spectrin breakdown products in cerebrospinal fluid after traumatic brain injury in rats. *J. Neurochemistry*, 78: 1297-1306.
53. Raabe, A., C. Grolms, and V. Seifert, (1999) Serum markers of brain damage and outcome prediction in patients after severe head injury. Br J Neurosurg, 13: 56-9.
54. Raghupathi R, Graham DI, McIntosh TK. (2000) Apoptosis after traumatic brain injury. J Neurotrauma. 2000 Oct;17(10):927-38. Review.
55. Reddy, V. N. Mavrovouniotis, M. L. and M. N. Liebman, (1993) Petri net representations in metabolic pathways, Proc. ISMB, 1993, pp. 328-336.

56. Reyes DR, Iossifidis D, Auroux PA, Manz A. (2002) Micro total analysis systems. 1. Introduction, theory, and technology. nal Chem. 2002 Jun 15;74(12):2623-36
57. Romner B, Ingebrigtsen T, Kongstad P, Borgesen SE. (2000) Traumatic brain damage: serum S-100 protein measurements related to neuroradiological findings. J Neurotrauma, 17(8): p. 641-7.
58. Schäfer, H.; Marcus, K.; Sickmann, A.; Herrmann, M.; Klose, J.; Meyer, H. E. (2003) Identification of phosphorylation and acetylation sites in alphaA-crystallin of the eye lens (mus musculus) after two-dimensional gel electrophoresis. *Anal Bioanal Chem.* 376(7): 966-972.
59. Schwartz, D. C.; Hochstrasser, M. (2003) A superfamily of protein tags: ubiquitin, SUMO and related modifiers. *Trends Biochem Sci.* 28(6):321-328.
60. Service, R. F. (2001) Gold rush - High-speed biologists search for gold in proteins. *Science.* 294(5549): 2074-2077.
61. Smith, R. D, (2000) Probing proteomes–seeing the whole picture? *Nature BioTech*, 18: 1041-1042.
62. Somogyi, R. and Sniegoski, C. A. (1996) Modeling the complexity of genetic networks: understanding multigenic and pleiotropic regulation, Complexity, 1, 1996, pp. 45-63.
63. Tabb, D.L., McDonald, W.H., and Yates, J.R., (2002) DTASelect and contrast: Tools for assembling and comparing protein identifications from shotgun proteomics, Journal of Proteome Research, 1 (2002) 21-26.
64. Unlu, M.; Morgan, M. E.; Minden, J. S. (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis.* 18(11): 2071-2077.
65. Wiesner A. (2004) n of Tumor Markers with ProteinChip(R) Technology. urr Pharm Biotechnol. 2004 Feb;5(1):45-67.
66. Yates, J.R. III, Carmack, E., Hays, L. Link, AJ., Eng, J.K. (1999) Automated Protein Identification using Microcolumn Liquid Chromatography-Tandem Mass Spectrometry. In: Methods in Molecular Biology, Vol 112: 2-D Proteome Analysis Protocols (A.J. Link, Ed) Human Press Inc, Totowa, NJ, pp 553-569..
67. Yates,J.R., Morgan,S.F., Gatlin,C.L., Griffin,P.R., and Eng,J.K., (1998) Method to compare collision-induced dissociation spectra of peptides: Potential for library searching and subtractive analysis, Analytical Chemistry, 70 (1998) 3557-3565.
68. Zhang, W.R., Chen, S., Wang, W. and King, R.S. (1992) A cognitive map based approach to the coordination of distributed cooperative agents, IEEE Trans. SMC, 22, 1992, pp. 103-114.
69. Zhang, W.R., Chen, S., and Bezdek, J. C. (1989) Pool2: A generic system for cognitive map development and decision analysis, IEEE Trans. SMC, 19, 1989, pp. 31-39.