Proteomic tools in discovery-driven science

Eric J. Hayduk, Leila H. Choe and Kelvin H. Lee*

Department of Chemical Engineering, Cornell University, Ithaca, NY 14853-5201, USA

Proteomics, the study of all proteins expressed from a genome, has focused traditionally on the use of two-dimensional gel electrophoresis to analyse the protein expression profile of biological samples (i.e. cells, tissues and fluids). A strength of this strategy is the ability to visualize and compare the expression patterns of a large number of proteins in parallel, making it an outstanding discovery tool. When coupled to mass spectrometry or amino acid sequencing, specific changes in protein fingerprints can be characterized and searched against database sequences. In addition to this approach, a variety of chip-based protein characterization tools have recently been developed. Other technologies such as yeast two-hybrid systems and surface display also exist. These technologies are also excellent tools for protein analysis, that should increase throughput and reliability of proteomic discovery.

PROTEOMICS is an approach to life sciences research that employs a variety of technologies to characterize, study and understand the protein complement of a biological sample (i.e. a proteome)[1]. The simultaneous analysis of all expressed proteins from a biological system poses significant technological challenges because proteins exhibit a great diversity of physicochemical properties. Traditionally, proteomics involved separating and visualizing proteins using two-dimensional gel electrophoresis (2DE). A comparison of gel images from different phenotypes or samples of interest identifies proteins that are uniquely expressed, up- or down-regulated in a particular physiological state. Proteins can then be excised from the gel, proteolytically digested and characterized using mass spectrometry (MS) coupled with database searches[1,2].

Although technological advances have helped proteomics mature significantly in past years, progress is still limited by a relative lack of technology for quantitative, sensitive and rapid analyses of protein mixtures. Among the emerging technologies showing promise are devices constructed using photolithographic printing and robotics techniques from the microelectronics industry. These technologies allow for the development of silicon chip-based proteomics. Several groups have used this technology to spot small quantities of protein solution onto silicon wafers or other substrates at high spatial density[3]. These protein chips function as valuable screening tools capable of binding a single protein from complex biological samples and are able to measure antibody–antigen specificity among other screens. Chip-based technologies can also be combined with mass spectrometry to enhance separation and quantification power. Other advances such as peptide surface display and yeast forward and reverse hybrid systems have made important contributions to our understanding of various proteomes.

The field of proteomics, which combines all of these technologies, and many more, may currently be viewed as providing discovery-based tools for life sciences research. These tools, which permit one to visualize and screen a large number of proteins provide excellent methods to discover new interactions and to elucidate new regulatory information at the protein level. The discoveries are useful in the generation of new hypotheses about the underlying biochemistry. This review is an attempt to cover some, but certainly not all, of the recent advances in proteomics and to review key principles of more mature technologies.

Two-dimensional gel electrophoresis and mass spectrometry

2DE is a multi-step separation technique in which the soluble protein content of a biological sample is separated first by isoelectric point (pI) and then by molecular weight in a polymer matrix[4]. Because the pI and the size of proteins are independent characteristics, one achieves excellent resolution of complex protein mixtures (Figure 1). The main steps in this procedure are sample preparation, isoelectric focusing (IEF), sodium dodecyl sulphate (SDS) polyacrylamide gel electrophoresis (PAGE), staining and detection, analysis, and spot characterization. This article is not intended to provide an extensive review of all six categories, but will discuss briefly the basic concepts and recent advances.

Sample preparation is a necessary step that allows one to denature and solubilize proteins. However, no one sample preparation technique will effectively permit the analysis of all proteins in a mixture, due to the physicochemical diversity of proteins and the need to use non-ionic or zwitterionic detergents. Several groups have recently suggested new protocols to prepare various protein mixtures[5]. Soluble proteins are subsequently applied to a low

*For correspondence. (e-mail: KHL9@cornell.edu)

concentration polyacrylamide gel matrix containing a pH gradient, wherein proteins are separated according to their pI in the presence of an electric field. Traditionally, IEF has been performed in hand-cast carrier ampholyte IEF gels that form a pH gradient using a mixture of ampholytic species in free solution[4]. The use of carrier ampholytes poses several problems such as pH gradient drift, which can lead to proteins migrating out of the gel at either end and gaps in the middle, and limited reproducibility[6]. To overcome these issues, immobilized pH gradient (IPG) strips have been developed in which the pH gradient-forming groups are co-polymerized in the polyacrylamide IEF gel[7]. The immobilized pH gradient affords better reproducibility, longer focusing times, and an improved capacity for higher voltages. IPG gels can be hand-cast, but are also commercially available in different lengths and pH gradients through Bio-Rad (Hercules, CA, USA) and Amersham–Pharmacia (Piscataway, NJ, USA). To further simplify isoelectric focusing, new instruments that combine IPG strip rehydration and IEF in one unit have been developed[8]. These instruments also afford a wider range of voltage programming capabilities and better temperature control.

After IEF, proteins typically undergo an additional preparation for SDS PAGE. They are first treated with a reducing agent such as dithiothreitol, to break disulfide bonds, and then an alkylating agent such as iodoacetamide, to protect the exposed sulphydryl groups[9]. These steps are often combined with an equilibration step that resolubilizes the focused proteins for further separation. SDS PAGE remains generally faithful to classic Laemmli electrophoresis, wherein proteins are loaded onto polyacrylamide gels with the anionic detergent, SDS, which uniformly binds to the proteins[10]. When an electric field is applied, these SDS–protein complexes migrate through the gel towards the anode and the gel matrix acts as a molecular sieve through which smaller proteins travel faster than larger ones. Pre-cast gels in different pore size distributions (%T, %C) and gradients are commercially available, as well as equipment to hand-cast such gels.

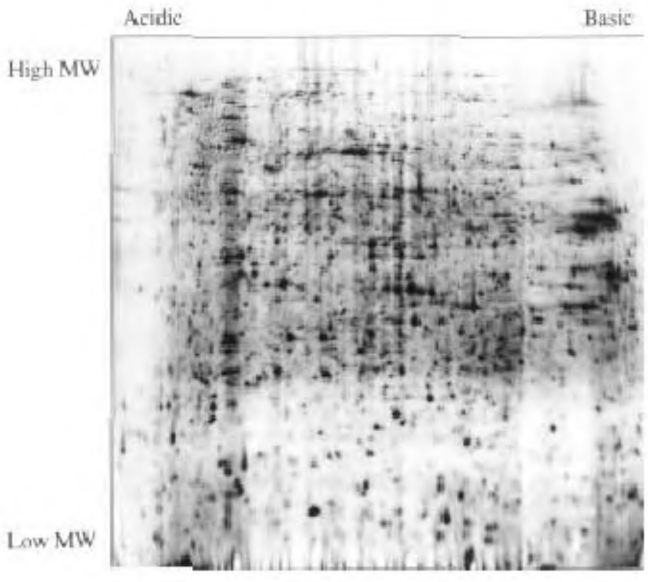

Figure 1. Silver-stain 2DE showing proteins from a microbial mixed culture separated on 3–10 nonlinear IPG strips and 12% T SDS PAGE gel.

After SDS PAGE, the proteins must be visualized to enable image acquisition and analysis. There are several in-gel staining options available and the specific choice is dictated by the needs of the investigator. Silver stain[11], Coomassie brilliant blue (CBB)[12], and colloidal blue[13] are at least three methods by which gel images can be visualized without special equipment. Among these choices, silver stain is the most sensitive for detection of low abundance proteins; however, silver staining is a lengthy process and silver-stained spots are often difficult to quantify. CBB is a relatively fast and simple detection method; however, it has lower sensitivity and often requires background destaining. Colloidal blue is a clear background stain that combines the simplicity of CBB and the sensitivity of some silver stains; however, it is not as fast as CBB and not as sensitive as an ammoniacal silver stain. Gel images resulting from these three stains can be captured by a laser densitometer to be analysed at a later date. Other choices for detection methods include pre-gel or post-run labelling with fluorescent dyes or pre-run labelling with a radioisotope (e.g. S^{35} or P^{32}). Visualization with these methods requires additional equipment such as a phosphorimager and/or fluorescence imager.

Subsequent analyses of gel images can be performed with any one of a variety of software packages available such as Melanie 3 (GeneBio, Geneva, Switzerland), PDQuest (Bio-Rad, Hercules, CA, USA), ImageMaster (Amersham–Pharmacia, Piscataway, NJ, USA) or Investigator HT Analyzer (Genomic Solutions, Ann Arbor, MI, USA). These software packages allow for multiple-gel comparisons and analysis over several parameters, whereby one can detect quantitative and qualitative changes. It should be noted that subsequent to SDS PAGE, gels may also be blotted, wherein the proteins and their resulting pattern are transferred onto polyvinylidene-difluoride (PVDF) or nitrocellulose membranes. Blotted 2DE images can then be visualized with a total protein stain such as Ponceau S or Colloidal Gold. However, blotting is not an effective method of studying an entire proteome due to loss of proteins during transfer and lower staining sensitivities.

The full utility of the 2DE approach is only realized when protein spots of interest can be characterized and linked to genes of known function. There are several technologies available to characterize spots of interest, such as direct amino acid sequencing[14] and amino acid analysis[15], but recent advances in technology have brought mass spectrometry (MS) to the forefront of proteomic analysis. There are two main approaches to MS spot

characterization : peptide mass mapping and tandem mass spectrometry. In both of these experiments, protein spots of interest are first excised from a stained 2D gel or blot (either manually or with an automated spot-picker) and then digested with a proteolytic enzyme (e.g. trypsin) to produce a mixture of short peptides. These peptides are ionized and studied using MS. In peptide mass mapping, the masses of all peptides resulting from the proteolytic digestion are measured and matched against a database of proteins and their theoretical digest masses or against theoretical six-frame translations of nucleotide databases which are digested *in silico*. For this type of analysis, a matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometer is often used. In tandem MS, a particular peptide ion is pre-selected to undergo further fragmentation, producing peptide fragments that have been predominantly broken at amide bonds between amino acids. The spectrum of these fragments displays peaks that differ in mass to charge ratio by one amino acid, allowing for partial peptide sequencing, which can be matched against amino acid and six-frame translations of nucleic acid sequence databases. The most commonly used instruments for this type of analysis are triple quadrupole or ion-trap mass analysers fitted with an electrospray ionization (ESI) source. For a detailed discussion of MS, the reader is directed to recent reviews[16,17].

The 2DE approach to proteome analysis, while effective, does have limitations which include: limited dynamic range, limited solubility of hydrophobic proteins, relatively poor focusing of very basic and very acidic proteins, poor or no visualization of very low abundance proteins, and limited inter-laboratory reproducibility[18]. Thus, there is a desire to overcome these issues by implementing alternate technologies. For example, Link and co-workers[19] have developed a 2D capillary liquid chromatography system coupled with tandem MS to analyse global protein digests[19]. Pasa-Tolic and coworkers[20] are currently evaluating the use of single-dimension ultrahigh-resolution capillary liquid chromatography combined with Fourier transform ion cyclotron resonance (FTICR) MS to analyse global protein digests. Wang *et al.*[21] describe a microfluidic device that integrates an ESI MS with a protein digestion bed, a capillary electrophoresis channel and an injector on a monolithic substrate. They demonstrate that the device could rapidly digest, separate and identify proteins. The protein digestion bed incorporates trypsin immobilized on microbeads, which allows for faster digestion, eliminates autodigestion products that hinder sample identification and increases the protein processing capacity of the ESI MS step. However, to date, 2DE remains a popular and effective discovery tool because of its high resolving power and readily available resources. When coupled with the ability of MS to characterize proteins even at low femtomole levels, the analytical power of this technique is greatly extended.

Chip-based technology

A variety of chip-based technologies have been recently developed to study proteins. Typically, small spots of proteins are immobilized on silicon, glass or other substrates and used to simultaneously screen protein mixtures. With existing technology, it is possible to array an entire complement of proteins produced from a library of cDNA clones onto a surface and then use the array to search for small molecule-binding interactions, antibody specificity or identify unique proteins in biological samples using fluorescent probes[3].

Leuking and coworkers[22] developed a transfer stamp with blunt-end tips measuring 250 µm. Spotting robots are able to deposit 5 nl drops of protein solution onto small (25 mm × 75 mm) PVDF membranes. These arrays can be used to bind antibodies to specific spots with sensitivities of 10 fmol antigen/µl above background, as determined by using glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and a chemiluminescent screen. Others have developed high throughput screens by arraying recombinant antibodies themselves rather than antigens. De Wildt *et al.*[23] used robotic picking and high-density gridding of antibody-expressing bacteria to produce antibody arrays. Arrays are then probed for specific protein–protein interactions using a filter-based enzyme-linked immunosorbent assay (ELISA). Simultaneous screens of 18,342 antibody clones can be achieved using this system. This approach has high sensitivity resulting from immobilizing live bacteria rather than purified protein and avoids inherent background problems associated with screens of phage libraries[24].

High-density screens can also be designed to immobilize enzymes onto an array surface for use as screens against a variety of small molecules. Macbeath and Schreiber[25] have immobilized proteins onto a glass microscope slide by using high precision contact printing to deliver small protein samples to an aldehyde-coated glass surface. With this system, densities of 1600 spots/cm^2 with spot diameters of 150–200 µm can be achieved. Applications include identification of protein kinase substrates and probing protein–protein interactions.

In addition to these emerging technologies, Ciphergen Biosystems (Fremont, CA, USA) has a commercially available device that combines chip-based techniques with mass spectrometry. Surface-Enhanced Laser Desorption/Ionization (SELDI) technology used in the ProteinChip® System[26,27] permits the selective capture of proteins from biological samples. Protein mixtures are incubated with any number of chips which probe hydrophobic, electrostatic, coordinate covalent bonding or Lewis acid/base interactions. The surfaces of the chips are precoated with chromatographic affinity surfaces, or the user can covalently couple enzymes, antibodies, receptors, DNA or other small molecules, which are used to probe the

sample mixture. Surface-enhanced affinity capture, the most promising version of SELDI technology, uses probe surfaces to extract, structurally modify or amplify a particular protein. After the addition of a matrix solution to enhance laser energy transfer and sample ionization, samples are analysed by time of flight MS. Advantages of this approach include reduced amount of sample preparation prior to MS and the ability to capture trace amounts of protein directly from biological fluids. Probes also provide information on hydropathicity, total charge, isoelecteric point, phosphorylation, glycosylation and primary structure[27]. Recently, ProteinChip Arrays have been used to identify disease markers[28]. For example, prostate-specific membrane antigen, a 100 kDa transmembrane glycoprotein thought to indicate prostate cancer tumour progression, can be detected in blood sera and quantified based on normalized peaks via ProteinChip technology. ProteinChip assays for other cancers are also being developed[29].

Nelson and coworkers[30] have combined surface plasmon resonance-biomolecular interaction analysis (SPR-BIA) with MALDI-TOF analysis (BIA-MS analysis). SPR-BIA is a chip-based technique that uses an immobilized receptor to monitor biomolecular interactions. Since SPR detection is nondestructive, proteins may be further subjected to TOF analysis for identification. This approach has detection limits at or below 20 fmol and a mass range up to 150 kDa. Applications of this technology include the isolation and detection of tagged polypeptides retrieved from the expression system on a BIA chip and ligand fishing in which a known biomolecule is used to selectively bind molecules from solution.

Another important area of technology is the development of microfluidic devices for the analysis of proteins. Miniaturization reduces the amount of analyte needed and potentially provides low-cost, disposable chips for a variety of applications. Pinto *et al.*[31] report the use of a novel interface for microfabricated fluidic systems with ESI MS. This study involves construction of a transfer capillary, the use of continuous sample infusion via nonaqueous electrolytes and novel cationic coatings. Liu *et al.*[32] and Gottschlich *et al.*[33] report lab-on-a-chip techniques for separating and detecting protein mixtures. Gottschlich and coworkers[33] have integrated a microreactor, injector and electrophoretic separator and a second reactor for derivatization on a monolithic substrate followed by fluorescence detection. Liu *et al.*[32] integrate capillary electrophoresis, post-column labelling and fluorescence detection on a microfabricated system. Others at the University of Michigan have developed integrated microfluidic systems capable of accurately metering nanolitre drops (0.5–125 nl) in channels using a combination of hydrophobic surface modification and air pressure[34]. Additional information about miniaturization can be found in a recent review[35].

Surface display screens

An alternate high-throughput screening approach, based on fluorescence-assisted cell sorting (FACS), is known as surface display. A library of genes encoding various proteins are inserted into expression vectors. Proteins are expressed as fusions to membrane-binding sequences and are displayed on the surface of phage, yeast or bacteria cells. The cells can then be probed with fluorescent ligands or other proteins and sorted by FACS. This approach is particularly useful when applied to protein evolution. Isolation of sorted cells followed by further growth and sorting allows selective enrichment of a property of interest. Applications include the analysis of gene regulation, the screening of small molecule libraries, the isolation of mutants with improved expression characteristics, the screening of enzyme libraries based on catalytic turnover and the screening of antibody fragments[36]. Daugherty *et al.*[36] have recently reviewed the basic approaches and issues associated with protein surface display technology. Recently, screens have been used to display functional T-cell receptor mutants on a yeast surface[37] and correlate secretion levels of single chain T-cell receptor with its surface display in *S. cerevisiae*[38]. Among other advantages, this screen provides the ability to select rare cell types[39,40] and provides a link to the gene producing the protein of interest.

Functional discovery

Proteomics also seeks to understand the interactions between proteins based on structure information and how these interactions help to form metabolic networks. Examples of the proteomic tools used to identify protein–protein interactions are yeast forward and reverse hybrid systems, developed almost ten years ago. It was shown that DNA binding and activating functions of yeast transcription factors were located on two different domains. By fusing an unrelated protein (protein A) to the activation domain and a second protein (protein B) to the binding domain, one is able to examine whether or not these proteins interact to restore transcriptional activity in yeast cells. Transcription of a reporter gene is used to identify yeast cells with restored activity. This system is known as the yeast two-hybrid system[41]. The power of the two-hybrid system is clearly demonstrated by publication of a complete protein–protein interaction map of *S. cerevisiae* by CuraGen Inc. (New Haven, CT, USA)[42]. A reverse two-hybrid system used to detect interaction defective alleles and alternative *n*-hybrid systems are discussed in a recent review[41].

Several groups have expanded upon two-hybrid technology to extend its capabilities. Beranger and coworkers[43], for example, have shown that fusing activating domains and DNA-binding domains to the C-terminus

of unknown proteins works as effectively as N-terminal fusions. This approach allows screens requiring a free N-terminus for protein–protein interactions to be probed. Serebriiskii and coworkers[44] have created a novel, two-hybrid dual bait system that identifies proteins interacting with one binding domain and proteins that do not react with a second binding domain.

Conclusion

Proteomics includes a diversity of tools, methods and devices. These include relatively more mature technologies such as MS and 2DE, and relatively newer technologies such as chip-based methods and technology to probe protein function. Clearly, no one technology can be singled out as a benchmark for proteomic study. Rather, it is a combination of the technologies described above (and others) that will eventually help elucidate a more complete picture of how proteome expression relates to any particular physiological state.

The desire to probe and study the proteome is certainly motivated in large part by the advancements made in high-throughput DNA sequencing and mRNA gene expression analysis. Indeed, if the information obtained at all three levels is integrated, it will provide a deeper understanding of gene expression (transcription and translation) and its relationship to phenotype. However, many challenges remain that extend beyond this discovery-based approach. For example, recent experiments and analyses suggest that there is not a simple relationship between mRNA and protein expression[45,46]. Thus, how does one correlate information at these two levels? Further, there is a need for better quantitative tools to monitor protein expression as well as a significant need to develop high throughput methods for the study of intracellular metabolites. Still, the existing methods provide a wealth of information about biological systems and have clearly demonstrated the ability to enrich our understanding of these systems.

1. Wasinger, V. C. *et al.*, *Electrophoresis*, 1995, **16**, 1090–1094.
2. Dutt, M. and Lee, K., *Curr. Opin. Biotechnol.*, 2000, **11**, 176–179.
3. Walter, G. *et al.*, *Curr. Opin. Microbiol.*, 2000, **3**, 298–302.
4. O'Farrell, P., *J. Biol. Chem.*, 1975, **250**, 4007–4021.
5. Rabilloud, T., *Electrophoresis*, 2000, **21**, 3329–3344.
6. Corbett, J. M. *et al.*, *ibid*, 1994, **15**, 1205–1211.
7. Bjellqvist, B. *et al.*, *ibid*, 1993, **14**, 1357–1365.
8. Choe, L. H. and Lee, K. H., *ibid*, 2000, **21**, 993–1000.
9. Lane, L. C., *Anal. Biochem.*, 1978, **86**, 655–664.
10. Laemmli, U. K., *Nature*, 1970, **227**, 680–685.
11. Merrill, C. M., *Science*, 1981, **211**, 1437–1438.
12. Weber, K. and Osborn, M., *J. Biol. Chem.*, 1969, **244**, 4406–4412.
13. Neuhoff, V. *et al.*, *Electrophoresis*, 1985, **6**, 427–448.
14. Aebersold, R. H. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 6970–6974.
15. Latter, G. I. *et al.*, *Clin. Chem.*, 1984, **30**, 1925–1932.
16. Andersen, J. S. and Mann, M., *FEBS Lett.*, 2000, **480**, 25–31.
17. Patterson, S. D., *Curr. Opin. Biotechnol.*, 2000, **11**, 413–418.
18. Smith, R. D., *Nature Biotechnol.*, 2000, **18**, 1041–1042.
19. Link, J. *et al.*, *ibid*, 1999, **17**, 676–682.
20. Pasa-Tolic, L. *et al.*, *J. Am. Chem. Soc.*, 1999, **121**, 7949–7950.
21. Wang, C. *et al.*, *Rapid Commun. Mass Spectrom.*, 2000, **14**, 1377–1383.
22. Leuking, A. *et al.*, *Anal. Biochem.*, 1999, **270**, 103–111.
23. De Wildt, R. M. T. *et al.*, *Nature Biotechnol.*, 2000, **18**, 989–994.
24. Irving, R. A. and Hudson, P. J., *ibid*, 2000, **18**, 932–933.
25. MacBeath, G. and Schreiber, S. L., *Science*, 2000, **289**, 1760–1763.
26. Senior, K., *Mol. Med. Today*, 1999, **5**, 326–327.
27. Merchant, M. and Weinberger, S. R., *Electrophoresis*, 2000, **21**, 1164–1177.
28. Xiao, Z. *et al.*, *Protein Exp. Purif.*, 2000, **19**, 12–21.
29. Paweletz, C. P. *et al.*, *Drug Dev. Res.*, 2000, **49**, 34–42.
30. Nelson, R. W. *et al.*, *Electrophoresis*, 2000, **21**, 1155–1163.
31. Pinto, D. *et al.*, *ibid*, 2000, **21**, 181–190.
32. Liu, Y. *et al.*, *Anal. Chem.*, 2000, **72**, 4608–4613.
33. Gottschlich, N. *et al.*, *J. Chromatogr. B*, 2000, **745**, 243–249.
34. Handique, K. *et al.*, *Anal. Chem.*, 2000, **72**, 4100–4109.
35. Sundberg, S. A., *Curr. Opin. Biotechnol.*, 2000, **11**, 47–53.
36. Daugherty, P. S., Iverson, B. L. and Georgiou, G., *J. Immunol. Methods*, 2000, **243**, 211–227.
37. Keike, M. C. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 5651–5656.
38. Shusta, E. V. *et al.*, *J. Mol. Biol.*, 1999, **292**, 949–956.
39. VanAntwerp, J. J. and Wittrup, K. D., *Biotechnol. Prog.*, 2000, **16**, 31–37.
40. Daugherty, P. S. *et al.*, *Protein Eng.*, 1999, **12**, 613–621.
41. Vidal, M. and Legrain, P., *Nucleic Acids Res.*, 1999, **27**, 919–929.
42. Uetz, P. *et al.*, *Nature*, 2000, **403**, 623–627.
43. Beranger, F. *et al.*, *Nucleic Acids Res.*, 1997, **25**, 2035–2036.
44. Serebriiskii, I. *et al.*, *J. Biol. Chem.*, 1999, **274**, 17080–17087.
45. Hatzimanikatis, V. and Lee, K., *Metal Eng.*, 1999, **1**, 275–281.
46. Gygi, S. *et al.*, *Mol. Cell. Biol.*, 1999, **19**, 1720–1730.

ACKNOWLEDGEMENTS. E.J.H. is supported by a Merck Graduate Research Fellowship. K.H.L. is a DuPont Young Professor. This work is supported in part by the National Science Foundation (BES-9874938) and the New York State Science and Technology Foundation.