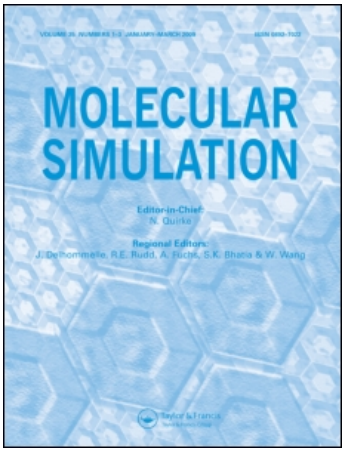
This article was downloaded by:
On: *14 January 2011*
Access details: *Access Details: Free Access*
Publisher *Taylor & Francis*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Biological network analyses: computational genomics and systems approaches

S. P. Walton[a]; Z. Li[a]; C. Chan[a]
[a] Cellular and Biomolecular Laboratory, Department of Chemical Engineering and Materials Science, Michigan State University, East Lansing, MI, USA

To cite this Article Walton, S. P. , Li, Z. and Chan, C.(2006) 'Biological network analyses: computational genomics and systems approaches', Molecular Simulation, 32: 3, 203 — 209
To link to this Article: DOI: 10.1080/08927020600647052
URL: http://dx.doi.org/10.1080/08927020600647052

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Molecular Simulation, Vol. 32, Nos. 3–4, 15 March–April 2006, 203–209

Taylor & Francis
Taylor & Francis Group

Biological network analyses: computational genomics and systems approaches

S.P. WALTON†·*, Z. LI‡ and C. CHAN¶

Cellular and Biomolecular Laboratory, Department of Chemical Engineering and Materials Science, Michigan State University, 3249 Engineering Building, East Lansing, MI 48824-1226, USA

(Received December 2005; in final form February 2006)

The complex responses of cells to stimuli are the aggregate of alterations at the genetic, protein, metabolic and cellular levels. The immense quantities of data now available from high-throughput genomic, proteomic and metabolomic sources require specialized analytical approaches. The integration of such data for the computational elucidation and analysis of cellular pathways and networks is an area of considerable current interest. As the quantity of available data continues to increase, strategies to extract the useful information from the data will continue to provide answers to important biological questions. Chemical engineers are actively involved in this field that has taken on the global title of "Systems Biology". These research groups have made considerable impact from both the computational and experimental standpoints. This commentary describes some of the recent contributions of chemical engineering researchers to the computational analysis of high-throughput, multi-source data as described in recent publications and presentations from the 2005 AIChE Annual Meeting.

Keywords: Systems biology; Network inference; Multi-source data; Signaling networks; Integrated networks

Downloaded At: 18:13 14 January 2011

1. Introduction

New insights into biological processes arise from both experimental and analytical innovations. Of particular recent interest is the generation and analysis of high-throughput data. This data is taken at many levels, e.g. the messenger RNA (mRNA) level (genomics), the protein level (proteomics) and the metabolite level (metabolomics). In each case, the goal of these "-omics" technologies is the simultaneous and quantitative measurement of all species in the class. For instance, genomics technology developments have led to oligonucleotide microarrays that can simultaneously generate over 500,000 data points from which the expression levels of over 47,000 unique mRNAs can be determined [1]. Concomitantly, advances in high-throughput data generation technologies have spurred new thinking and new approaches to analyzing such data to ensure that maximal information is extracted. This information potentially includes points of signaling and metabolic network connectivity that are too subtle to be detected in the examination of low-throughput (single gene or a few genes at a time) experiments. Mathematical tools are required that can integrate vast mRNA, protein, metabolite and pathway data. This is one powerful way that new analytical strategies are contributing to the understanding of biological systems.

By accounting for all of the available data simultaneously, statistical inferences of interactions that exist within a network can be made that would not be identified otherwise. A modeling framework offers the ability to predict cellular function and regulation and describes key aspects of network functionality, such as robustness, from the network structure. Numerous mathematical models, from abstract discrete Boolean networks [2] to detailed biological mechanistic-based models [3], have been applied to represent biological networks. Once the appropriate model framework is identified, manipulation of the network becomes a more straightforward task, with alterations to network components guided by model predictions.

*Corresponding author. Email: spwalton@msu.edu
†Tel.: +1-517-432-8733. Fax: +1-517-432-1105.
‡Email: lizheng1@egr.msu.edu.
¶Email: krischan@egr.msu.edu.

Molecular Simulation
ISSN 0892-7022 print/ISSN 1029-0435 online © 2006 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/08927020600647052

In general, the objectives of modeling approaches for metabolic and biological systems can be categorized into four groups: (i) feature selection, (ii) prediction, (iii) optimization and (iv) network inference. (i) Feature selection is used to build a classification model that evaluates patterns in the data and extracts "characteristic modes or features" within the data. These features are used to classify the data into groups, e.g. genes of similar regulation and function or similar cellular state or biological phenotype. The model, once trained, can then be used to assign the classification for sets of unknown data. (ii) Prediction is employed to determine, for instance, the cellular response as a function of the environmental variables and their genetic and metabolic profiles. Therefore, as these variables change, the model can predict how the cells will respond to the changes. (iii) Optimization involves finding values of the variables that optimize (maximize or minimize) an objective function. Variables may include concentrations of intracellular and extracellular species whose alteration results in a unique cellular behavior, while the objective function may be a particular cellular response, such as growth rate or specific production of a cellular component, for which an optimal behavior is desired. (iv) Network inference refers to the construction of network connectivity from high-throughput data. Inferring network interactions takes advantage of the available prior knowledge of the cause–effect relationships among the variables and, from this, proposes previously unidentified connections that exist. Knowing the full extent of the causal relationships among system components will help properly identify the variables that should be modulated (and how they should be modulated) to elicit a specific cellular or tissue response. A more detailed knowledge and understanding of the metabolic and genetic regulatory networks and their interactions facilitates optimization and prediction of cellular function for a variety of situations and conditions. As highlighted in this brief review focusing on the authors whose work was presented in the "Computational and Functional Genomics" sessions at the 2005 Annual AIChE meeting [4,5], a number of chemical engineering research groups are making valuable contributions to the computational methods required for network inference and analysis.

Downloaded At: 18:13 14 January 2011

2. Sources of data for computational cellular pathway analyses

Clearly experimental investigations provide the data to be analyzed by computational strategies, but experiments also provide critical constraints of known information that can improve the utility and accuracy of computational methods. Recently developed high-throughput data generation strategies provide the basis for the application of systems approaches to biological problems. High-throughput biological information can now be obtained at some degree for all biological molecules. In addition, databases of compiled information from both high-throughput experiments and literature data mining can provide sample sizes sufficient for the extraction of new pathway information with statistical significance that would be lacking from small-scale experiments alone.

2.1 Gene expression data

Oligonucleotide (short single-stranded DNA molecules) and cloned DNA (cDNA) microarrays are well established for the measurement of genomics data such as mRNA levels, genomic DNA resequencing, single nucleotide polymorphism (SNP) detection, etc. (recently reviewed in [6]). Microarray analyses depend on two unique properties of nucleic acids: (i) specific complementarity by Watson–Crick base-pairing (A:T and G:C) and (ii) mRNA amplification by reverse transcription-polymerase chain reaction (RT-PCR). Complementarity allows simultaneous separation of all targets in the sample. Amplification provides sufficient material for detection. For these reasons, genomics technologies have far outpaced high-throughput technologies for other cellular molecules, which do not possess these properties. Hence, systems biology approaches are still primarily based on genomics data.

With the quantity of data that is generated by microarray experiments, it is important to assess the value of each of the data points in providing classification of biological samples. Those genes that provide the greatest information in the expression dataset will be most useful for categorizing new samples in diagnostic applications. One approach to assessing the value of each data point is the application of decision trees [7]. In one study, an algorithm was trained with 63 samples containing representatives of each of four tumor cell types catalogued by pathology. It was determined that as few as three genes could perfectly categorize the samples into the four unique classes present in the training data. The method of gene selection proved reproducible with two genes being selected as most informative in 89 and 67% of cases, respectively. Each of these genes is known to be associated with the tumor cell types for which they proved most discriminatory. However, the genetic and phenotypic differences among different cancers are often dramatic. It is not expected that more subtle phenotypic differences could be accurately classified with as few informative genes, but it is expected that the described approach would minimize the number of genes required for any such discrimination, guiding more thorough analyses of the most relevant genes.

2.2 Protein expression and interaction data

Typical proteomic experimental systems take both array-based and non-array-based forms. In non-array-based proteomics techniques, separation and detection are performed in distinct operations. Two principal methods for proteomic separation are two-dimensional electrophoresis

(2DE) [8], where proteins are separated by their isoelectric point (p*I*), in essence molecular charge, and their size (molecular weight) and liquid chromatography (LC) [9], but others have also been used [10]. Following separation, proteins are identified by western blotting or single or tandem mass spectrometry (MS or MS/MS) [9]. Unfortunately, MS does not innately provide quantitative information about peptide concentrations in a sample [11]. Recent efforts to establish absolute and relative quantitation for MS samples, e.g. using peptide standards and stable isotope labeling techniques such as ICAT and iTRAQ, are beginning to improve the utility of MS as a quantitative measure of peptide concentrations [12].

Arrays for protein detection use the same principle of simultaneous separation as nucleic acid arrays [13]. Protein arrays can be separated into two classes, (i) those designed to identify interactions of a specific protein with other molecules or (ii) those designed to measure many protein levels in a sample. The first class of arrays includes DNA-binding protein arrays, small molecule arrays for drug targets, lectin arrays for glycoprotein binding and protein arrays for protein–protein interactions [13–24]. The second class of arrays typically utilizes antibody–protein binding for specific protein separation. Antibody arrays for classes of proteins, including cytokines, have been successfully applied [25,26]. However, any approach based on antibodies will suffer from product variability, affinity and kinetics limitations, and possible protein denaturation [27]. Aptamers, RNA molecules that can specifically bind proteins like antibodies, have also been developed for protein measurements on both solid substrates and confined microspheres [28–34]. Current aptamer array technologies, though, are in their infancy relative to DNA microarrays and require specialized instruments for use, detection and analysis [31,32,34]. Better solutions for array-based proteomics are an area of considerable current interest and investigation.

In addition to protein expression data, protein–protein interaction data is critical in identifying connections among cellular pathways that might not be evident in other ways. One technique that has proven successful for assessing protein–protein interactions is termed "the yeast two-hybrid assay". The technique is predicated on the fact that transcription factors have distinct domains for DNA binding and transcriptional activation (the original demonstration of the method utilized a factor, GAL4, specific for activation of galactose metabolism genes in yeast [35]). To test whether two proteins, A and B, interact each of these proteins is conjugated to one domain of the transcription factor. If the A and B conjugates interact, then their attached transcription factor domains will be brought into proximity resulting in the upregulation of expression from the reporter gene. Though arguments can arise as to the biological relevance of some of the measured interactions, the technique provides a robust and flexible means of identifying those proteins that show stable intracellular interactions and may, therefore, be linked in some fashion in a cellular network.

2.3 Metabolomic data

Metabolic data often reflects the aggregate activity of the cellular network and can be a valuable means of comparing the global effects of different stimuli on cells. High-throughput metabolic data or metabolomic data, is typically acquired through combined chromatographic separation (either GC or LC) and MS [36]. One powerful approach was recently described that identified over 5000 metabolites in a 20 μl sample using primarily automated sample handling [9]. As with MS-based proteomics, inter- and intra-sample comparisons of metabolite concentrations are very difficult, requiring addition of complex internal standards. Nonetheless, the presence or absence of a particular metabolite of interest in a sample can be readily established, at least for a single instrument set up and sample handling protocol [37], and its concentration quantified relative to a standard or by more traditional analytical methods if the importance of the molecule dictates such scrutiny.

2.4 Molecular interaction databases

Database information collected from many sources is used to ensure that the overarching biology is accounted for in computational results. The Kyoto Encyclopedia of Genes and Genomes (KEGG) database is one of the most highly utilized [38] and has features akin to those available from a number of databases that are currently freely available online. Included in the KEGG database are the proteins and small molecules involved in major metabolic and signaling cellular pathways. In addition, sequence, structure, known functions and localization information are provided for proteins, when available. Connections among various cellular components have been built from the available literature and are continually updated. One commercial database has combined manual and automated data collection to yield a curated database that is proposed to reflect biologically relevant pathway information more accurately than databases that rely solely on automated text searching of the literature [39].

3. Construction and analysis of cellular networks

3.1 Signaling networks

Cells respond to environmental and physiological changes through an extensive transcriptional regulatory network, which concludes with transcription factors acting on their target genes. Protein receptors on the outer surface of cellular membranes bind to specific ligands present in the extracellular space and initiate a cascade of events leading to the activation of transcription factors. These transcription factors bind to the promoter regions of specific genes to either positively or negatively regulate expression of those genes. The activities of transcription factors depend both on the quantity of the transcription factor proteins present as well as their state of activation.

Downloaded At: 18:13 14 January 2011

High-throughput technologies can measure the mRNA expression levels of transcription factors. However, genome-wide measurement of transcription factor activities remains a challenge that is being addressed computationally from a number of unique perspectives.

Typically, a gene is regulated by several different transcription factors, and the specific role of each transcription factor in regulating each gene is not easily ascertained. Until high-throughput protein analysis approaches can accomplish this formidable task, computational strategies are better suited to elucidating these relationships. One tool for accomplishing this task is promoter analysis and interaction network tool (PAINT) [40]. Integrating promoter database information and genomic data, PAINT returns a network identifying the "transcriptional regulatory elements" (TREs) for each gene. TREs provide insight into the coordinated regulation of each gene through multiple transcription factor activities. It was determined from two case studies that the majority of genes have between 8 and 12 TREs, with the majority of TREs regulating fewer than 10 genes. This indicates that multiple transcription factors can potentially play redundant, synergistic or antagonistic roles in affecting gene expression, perhaps providing finer control of the expression of important genes than would be achieved with a single TRE. The success of PAINT also demonstrates the value of integrating multi-source data for improved understanding of cellular function.

Another approach to understanding transcription factor activities and regulation, network component analysis (NCA), demonstrates the integration of known pathway information with statistical analysis to improve the description of the underlying biology [41]. NCA predicts the function of each transcription factor in the regulation of the signaling network through a "control strength". The method requires no assumptions about the underlying data structure, unlike independent component analysis (ICA) and principal component analysis (PCA), and incorporates known pathway information in the form of statistical constraints. If the data satisfy three criteria (see online appendices of [41] for further information), a qualitatively unique network structure can be obtained. NCA utilizes available network information to reconstruct unknown connections that must exist to result in the measured output signal dynamics. The technique accurately reconstructed known spectral data in a model case and, in a biological test, identified transcription factor activities during a carbon source transition in *E. coli* despite little change in the mRNA expression levels for the transcription factor genes. Expression levels of 100 genes were used to determine the activities of 16 transcription factors that control those genes. Subsequent experimental results validated the predicted transcription factor activities.

NCA succeeds in part due to the existing wealth of information on transcription factor–gene connectivity. As more such information is generated and validated, NCA could be applied for the whole-cell determination of transcription factor activities, which would permit the formulation of global transcriptional regulatory networks. One limitation of the initial formulation of NCA was its inability to incorporate experimental constraints, such as gene knockouts, into the mathematical formulation. However, the method has recently been generalized to make it feasible to include such information [42].

3.2 Integrated metabolic and signaling networks

The interest of chemical engineering groups in the analysis of genomics data grew from strengths in metabolic engineering. Numerous mathematical approaches were developed to analyze the systematic properties and behavior of complex metabolic networks and the pathways of which they are comprised [43]. A metabolic pathway consists of enzymatic reactions that convert reactants (substrates) to products. Each reaction starts with a substrate(s) and terminates with a product(s). A substrate can participate in any number of reactions or pathways. A metabolic network is a linked set of complex interconnected pathways. Existing literature knowledge is relied upon heavily to reconstruct or build metabolic networks (e.g. KEGG). These databases combine genome sequence data with information on biological function to reconstruct metabolic pathways.

The quantitative analysis of metabolic networks and other metabolic engineering techniques have traditionally been applied to optimize cellular function primarily for the purpose of enhancing the production of specialty chemicals from prokaryotic cells, yeasts, fungi and to some limited extent, mammalian cells [44,45]. However, prior to the application of metabolic engineering techniques, the network map of the metabolic pathways for the cell or tissue of interest must be constructed. Mapping all the destinations for a particular enzymatic substrate *a priori* would facilitate the building of metabolic networks.

Systems approaches to analyze metabolic network information quantitatively are gaining in importance. Information on how cells are regulated and behave under a given environment is contained within the gene expression and metabolic profile data, and thus could be readily captured by an appropriate model. The information contained within the data includes, the linear and nonlinear relationships between the external and internal variables. Systems techniques can help extract this information from the data and facilitate the reconstruction of metabolic networks and the coupling of gene regulatory and signaling pathways to the metabolic networks.

Once relationships among the cellular networks are identified, they can be used to instruct how to modulate tunable variables (gene expression levels, protein activities and extracellular stimuli) to obtain a desired level of cellular function or response. Knowledge of the metabolic network and the interactions among the various networks (gene, signaling, etc.) facilitates optimization and prediction of cellular processes. Recent work demonstrates two approaches to integration of signaling and metabolic

Downloaded At: 18:13 14 January 2011

networks: use of genomic data for reconstruction of metabolic pathways and simultaneous kinetic modeling of signaling and metabolic reactions.

Metabolic pathways were successfully reconstructed using a genetic algorithm/partial least squares (GA/PLS) approach that integrated metabolic response data, metabolic pathway information and genomic data [20]. The goal of the study was to identify those genes whose expression level changes predicted most closely metabolic data obtained from cells exposed to fatty acids and hormone-supplemented culture media. The sets of relevant genes selected by GA/PLS were found to depend on the initial conditions that were randomly chosen by the algorithm. GA is a simple and robust algorithm that is able to find solutions rapidly for difficult high-dimensional problems. GA is an efficient search algorithm when the search space is large, complex, poorly understood or when the domain knowledge is scarce. A disadvantage of GA is that it cannot guarantee an optimum or global solution. Therefore, to address this limitation, the GA algorithm was run many times and the frequency with which each gene was selected was counted to predict a metabolic or phenotypic response in the PLS model [46]. A cut-off value of 55% was chosen to identify the genes selected with a probability higher than random chance. Examining the biological function of the selected genes, the urea cycle, tricarboxylic acid (TCA) cycle and metabolic pathways for triglyceride synthesis were reconstituted. However, this approach utilized existing network information to connect the relevant pathway components. Approaches that generate a network without prior connectivity knowledge were thus considered.

It was demonstrated that a Bayesian framework could reconstruct metabolic networks from metabolic data [47], and this approach was extended to the reconstruction of the signaling networks involved in an observed metabolic response [48]. This method requires a large amount of data for the network reconstruction. Therefore, an alternative approach was developed that identified the active pathways without interaction measurements and with a limited amount of data, namely, by integrating gene expression and metabolic profiles. The approach, termed Three-stage Integrative Pathway Search (TIPS©), first couples GA/PLS and constrained ICA to identify those genes that were statistically most likely to be involved in the process of interest, in this case cell death as characterized by LDH release. Bayesian network analysis was then applied to the selected genes to reconstruct network connectivity. An advantage of this approach is the coupling of multi-source data, which may capture more of the cellular state than with gene expression data alone. This is done at the expense of a reduced sample size, which can be overcome, if necessary, using techniques such as interpolation and resampling [49]. Perturbation analysis identified genes in the reconstruction network whose expression levels were uniquely important in mediating the biological response. Experimental inhibition of these genes validated their connection to the cytotoxic response, as predicted by the modeling approach.

Another biological response, biomass accumulation subsequent to extracellular stimulation, was simulated using a multi-scale kinetic model based on flux balance analysis [50]. In this model, an integrated signaling and metabolic network was built leading to the accumulation of biomass. Estimates of reaction rate constants for the various processes that occur (i.e. receptor–ligand association, phosphorylation, ATP hydrolysis, transcriptional activation/repression, protein synthesis, etc.) were obtained from the literature [51]. The pseudo-steady-state assumption was then used for "fast reactions" (processes/reactions occurring on the order of seconds) while "slow reactions" (processes/reactions occurring on the order of minutes or hours) were modeled as occurring after a time delay of appropriate length in discretized time units. Biomass accumulation was represented by a reaction in which the combination of a number of components in a specified stoichiometry yielded a "molecule" of biomass. Alterations in the initial signaling events were shown to propagate through the network and alter the rates of biomass accumulation and carbon depletion. This technique provides a unified approach to modeling a unified signaling and metabolic network by treating all pathways as chemical reactions of known kinetics.

4. Conclusions

Biological systems are complex and integrated. Classical experimental techniques have yielded considerable information about the key participants in important signaling and metabolic pathways. The advent of high-throughput technologies has altered the manner in which data are collected and subsequently analyzed. Newly developed analytical techniques take advantage of the available data as a whole and integrate as many data sources as possible to characterize the biological situation. Much recent work has focused on establishing functional pathways and regulatory roles for the various constituents of the cellular milieu. With validated statistical models in hand, system modifications, such as gene knockouts or overexpressions, can be tested *in silico* without the cost and time expense of laboratory experiments. In this way, prime targets for intervention, whether they be genetic, protein activation or metabolic, can potentially be identified more directly. As the quantity and quality of data increases, the value of these approaches will be further demonstrated.

Acknowledgements

The authors gratefully acknowledge support from the National Science Foundation (BES 0222747, BES 0331297, and BES 0425821), the Environmental

Downloaded At: 18:13 14 January 2011

Protection Agency, the Whitaker Foundation, and Michigan State University. Z. L. is supported in part by a deVlieg fellowship for Computational Engineering at Michigan State University.

References

[1] Affymetrix—GeneChip; Human Genome U133 Plus 2.0 Array, Available online at: http://www.affymetrix.com/support/technical/datasheets/human_datasheet.pdf (accessed 15 November 2005).
[2] S. Liang, S. Fuhrman, R. Somogyi. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pac. Symp. Biocomput.*, 18 (1998).
[3] T. Chen, H.L. He, G.M. Church. Modeling gene expression with differential equations. *Pac. Symp. Biocomput.*, 29 (1999).
[4] AIChE Annual Meeting Program—Session: #494—Computational and Functional Genomics (21010). Available online at: http://aiche.confex.com/aiche/2005/techprogram/S1342.htm (accessed 13 December 2005).
[5] AIChE Annual Meeting Program—Session: #208—Computational Genomics (21012). Available online at: http://aiche.confex.com/aiche/2005/techprogram/S1344.htm (accessed 13 December 2005).
[6] C. Kidgell, E.A. Winzeler. Elucidating genetic diversity with oligonucleotide arrays. *Chromosome Res.*, **13**, 225 (2005).
[7] I.P. Androulakis. Selecting maximally informative genes. *Comput. Chem. Eng.*, **29**, 535 (2005).
[8] A. Gorg, W. Weiss, M.J. Dunn. Current two-dimensional electrophoresis technology for proteomics. *Proteomics*, **4**, 3665 (2004).
[9] Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H.R. Udseth, R.D. Smith. Automated 20 kpsi RPLC-MS and MS/MS with chromatographic peak capacities of 1000–1500 and capabilities in proteomics and metabolomics. *Anal. Chem.*, **77**, 3090 (2005).
[10] L. Breci, E. Hattrup, M. Keeler, J. Letarte, R. Johnson, P.A. Haynes. Comprehensive proteomics in yeast using chromatographic fractionation, gas phase fractionation, protein gel electrophoresis, and isoelectric focusing. *Proteomics*, **5**, 2018 (2005).
[11] R. Aebersold, M. Mann. Mass spectrometry-based proteomics. *Nature*, **422**, 198 (2003).
[12] D.R. Graham, S.T. Elliott, J.E. Van Eyk. Broad-based proteomic strategies: a practical guide to proteomics and functional screening. *J. Physiol.*, **563**, 1 (2005).
[13] A. Lueking, D.J. Cahill, S. Mullner. Protein biochips: a new and versatile platform technology for molecular medicine. *Drug Discov. Today*, **10**, 789 (2005).
[14] K.T. Pilobello, L. Krishnamoorthy, D. Slawek, L.K. Mahal. Development of a lectin microarray for the rapid analysis of protein glycopatterns. *Chembiochem* (2005).
[15] G. MacBeath, S.L. Schreiber. Printing proteins as microarrays for high-throughput function determination. *Science*, **289**, 1760 (2000).
[16] S. Mukherjee, M.F. Berger, G. Jona, X.S. Wang, D. Muzzey, M. Snyder, R.A. Young, M.L. Bulyk. Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nat. Genet.*, **36**, 1331 (2004).
[17] T. Egener, E. Roulet, M. Zehnder, P. Bucher, N. Mermod. Proof of concept for microarray-based detection of DNA-binding oncogenes in cell extracts. *Nucleic Acids Res.*, **33**, e79 (2005).
[18] J. Kim, A.A. Bhinge, X.C. Morgan, V.R. Iyer. Mapping DNA–protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nat. Methods*, **2**, 47 (2005).
[19] O. Stoevesandt, M. Elbs, K. Kohler, A.C. Lellouch, R. Fischer, T. Andre, R. Brock. Peptide microarrays for the detection of molecular interactions in cellular signal transduction. *Proteomics*, **5**, 2010 (2005).
[20] K. Kim Guisbert, K. Duncan, H. Li, C. Guthrie. Functional specificity of shuttling hnRNPs revealed by genome-wide analysis of their RNA binding profiles. *RNA*, **11**, 383 (2005).
[21] Y.S. Choi, S.P. Pack, Y.J. Yoo. Development of a protein microarray using sequence-specific DNA binding domain on DNA chip surface. *Biochem. Biophys. Res. Commun.*, **329**, 1315 (2005).
[22] G. MacBeath, A.N. Koehler, S.L. Schreiber. Printing small molecules as microarrays and detecting protein–ligand interactions en masse. *J. Am. Chem. Soc.*, **121**, 7967 (1999).
[23] F.G. Kuruvilla, A.F. Shamji, S.M. Sternson, P.J. Hergenrother, S.L. Schreiber. Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature*, **416**, 653 (2002).
[24] J.L. Naffin, Y. Han, H.J. Olivos, M.M. Reddy, T. Sun, T. Kodadek. Immobilized peptides as high-affinity capture agents for self-associating proteins. *Chem. Biol.*, **10**, 251 (2003).
[25] R.J. Feezor, H.V. Baker, W. Xiao, W.A. Lee, T.S. Huber, M. Mindrinos, R.A. Kim, L. Ruiz-Taylor, L.L. Moldawer, R.W. Davis, J.M. Seeger. Genomic and proteomic determinants of outcome in patients undergoing thoracoabdominal aortic aneurysm repair. *J. Immunol.*, **172**, 7103 (2004).
[26] A. Sreekumar, M.K. Nyati, S. Varambally, T.R. Barrette, D. Ghosh, T.S. Lawrence, A.M. Chinnaiyan. Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res.*, **61**, 7585 (2001).
[27] C.K. O'Sullivan. Aptasensors—the future of biosensing? *Anal. Bioanal. Chem.*, **372**, 44 (2002).
[28] J.R. Collett, E.J. Cho, J.F. Lee, M. Levy, A.J. Hood, C. Wan, A.D. Ellington. Functional RNA microarrays for high-throughput screening of antiprotein aptamers. *Anal. Biochem.*, **338**, 113 (2005).
[29] D. Smith, B.D. Collins, J. Heil, T.H. Koch. Sensitivity and specificity of photoaptamer probes. *Mol. Cell. Proteomics*, **2**, 11 (2003).
[30] T.G. McCauley, N. Hamaguchi, M. Stanton. Aptamer-based biosensor arrays for detection and quantification of biological macromolecules. *Anal. Biochem.*, **319**, 244 (2003).
[31] R. Kirby, E.J. Cho, B. Gehrke, T. Bayer, Y.S. Park, D.P. Neikirk, J.T. McDevitt, A.D. Ellington. Aptamer-based sensor arrays for the detection and quantitation of proteins. *Anal. Chem.*, **76**, 4066 (2004).
[32] Y. Liu, C. Lin, H. Li, H. Yan. Aptamer-directed self-assembly of protein arrays on a DNA nanostructure. *Angew. Chem. Int. Ed. Engl.*, **44**, 4333 (2005).
[33] T.H. Koch, D. Smith, E. Tabacman, D.A. Zichi. Kinetic analysis of site-specific photoaptamer-protein cross-linking. *J. Mol. Biol.*, **336**, 1159 (2004).
[34] M. Lee, D.R. Walt. A fiber-optic microarray biosensor using aptamers as receptors. *Anal. Biochem.*, **282**, 142 (2000).
[35] S. Fields, O. Song. A novel genetic system to detect protein–protein interactions. *Nature*, **340**, 245 (1989).
[36] S.J. Park, S.Y. Lee, J. Cho, T.Y. Kim, J.W. Lee, J.H. Park, M.J. Han. Global physiological understanding and metabolic engineering of microorganisms based on omics studies. *Appl. Microbiol. Biotechnol.*, **68**, 567 (2005).
[37] O. Fiehn. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155 (2002).
[38] KEGG: Kyoto Encyclopedia of Genes and Genomes. Available online at: http://www.genome.ad.jp/kegg/ (accessed 23 Nov 2005).
[39] Ingenuity Systems. Available online at: http://www.ingenuity.com/ (accessed 23 Nov 2005).
[40] R. Vadigepalli, P. Chakravarthula, D.E. Zak, J.S. Schwaber, G.E. Gonye. PAINT: a promoter analysis and interaction network generation tool for gene regulatory network identification. *Omics*, **7**, 235 (2003).
[41] J.C. Liao, R. Boscolo, Y.L. Yang, L.M. Tran, C. Sabatti, V.P. Roychowdhury. Network component analysis: reconstruction of regulatory signals in biological systems. *Proc. Natl. Acad. Sci. USA*, **100**, 15522 (2003).
[42] L.M. Tran, M.P. Brynildsen, K.C. Kao, J.K. Suen, J.C. Liao. gNCA: a framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation. *Metab. Eng.*, **7**, 128 (2005).
[43] J.E. Bailey. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnol. Prog.*, **14**, 8 (1998).
[44] J. Nielsen. Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol. Bioeng.*, **58**, 125 (1998).

Downloaded At: 18:13 14 January 2011

[45] G. Stephanopoulos. Metabolic engineering. *Biotechnol. Bioeng.*, **58**, 119 (1998).
[46] Z. Li, C. Chan. Integrating gene expression and metabolic profiles. *J. Biol. Chem.*, **279**, 27124 (2004).
[47] Z. Li, C. Chan. Inferring pathways and networks with a Bayesian framework. *FASEB J.*, **18**, 746 (2004).
[48] Z. Li, S. Srivastava, X. Yang, C. Chan. Inferring pathways that confer a cellular phenotype by integrating gene expression and metabolic profiles. Paper presented at the AIChE Annual Meeting, Cincinnati, OH (2005).
[49] J. Yu, V. Smith, P. Wang, A. Hartemink, E. Jarvis. Using bayesian network inference algorithms to recover molecular genetic regulatory networks. Paper presented at the International Conference on Systems Biology (2002).
[50] J.M. Lee, J. Papin. Dynamical analysis of an integrated signaling network at a genome-scale. Paper presented at the Annual AIChE Meeting, Cincinnati, OH (2005).
[51] E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nat. Biotechnol.*, **23**, 975 (2005).

Downloaded At: 18:13 14 January 2011