Building the power house: recent advances in mitochondrial studies through proteomics and systems biology

Thuy D. Vo and Bernhard O. Palsson

Department of Bioengineering, University of California, San Diego, La Jolla, California

Submitted 19 April 2006; accepted in final form 28 July 2006

Vo TD, Palsson BO. Building the power house: recent advances in mitochondrial studies through proteomics and systems biology. Am J Physiol Cell Physiol 292: C164–C177, 2007. First published August 2, 2006; doi:10.1152/ajpcell.00193.2006.—The emerging field of systems biology seeks to develop novel approaches to integrate heterogeneous data sources for effective analysis of complex living systems. Systemic studies of mitochondria have generated a large number of proteomic data sets in numerous species, including yeast, plant, mouse, rat, and human. Beyond component identification, mitochondrial proteomics is recognized as a powerful tool for diagnosing and characterizing complex diseases associated with these organelles. Various proteomic techniques for isolation and purification of proteins have been developed; each tailored to preserve protein properties relevant to study of a particular disease type. Examples of such techniques include immunocapture, which minimizes loss of post-translational modification, 4-iodobutyrylphenylphosphonium labeling, which quantifies protein redox states, and surface-enhanced laser desorption ionization-time-of-flight mass spectrometry, which allows sequence-specific binding. With the rapidly increasing number of discovered molecular components, computational models are also being developed to facilitate the organization and analysis of such data. Computational models of mitochondria have been accomplished with top-down and bottom-up approaches and have been steadily improved in size and scope. Results from top-down methods tend to be more qualitative but are unbiased by prior knowledge about the system. Bottom-up methods often require the incorporation of a large amount of existing data but provide more rigorous and quantitative information, which can be used as hypotheses for subsequent experimental studies. Successes and limitations of the studies reviewed here provide opportunities and challenges that must be addressed to facilitate the application of systems biology to larger systems.

constraint-based modeling; kinetics-based modeling; data integration; standards; bioinformatics

ADVANCES IN BIOINFORMATICS and high-throughput technologies have rapidly expanded the availability of information at all levels of biological investigation. This information has greatly enhanced our knowledge of the molecular makeup, localization, and interactions among cellular components. The need for organizing and understanding these parts catalogs has led to the emergence of numerous biological databases and has also given birth to the field of systems biology. The launching of systems biology as a discipline represents a collaborative effort among scientists of different research areas to unite biological data, physical principles, and mathematical tools to unravel the complexity of living systems. Successes in this field depend on the interplay of four processes: 1) identifying key biological components, 2) reconstructing networks of interactions among these components, 3) quantitatively analyzing these networks, and 4) generating testable hypotheses for model validation and further experimental investigations. Together, these four steps form a cycle (Fig. 1) in which every iteration provides better understanding of the biological system than the sum of results individually collected from each process.

Component identification is an ongoing process in biological study and has accumulated data at many different levels of detail and scale. Early studies painstakingly identified and characterized enzymes and enzymatic activity individually (57, 115), whereas more recent studies have aimed at the organelle and cellular levels. Although high-throughput technologies have not been able to achieve the complete biochemical characterization of individual enzymes, they are excellent tools for identification of unknown components for further investigation. Large data sets of mitochondrial genomics and proteomics are currently available, and metabolomics is likely to appear on the horizon. Furthermore, current advances of proteomic techniques do not limit their use to component cataloging; they are also utilized in interaction discovery and functional-state characterization. For example, techniques discussed in this review, such as immunocapturing, can detect the assembly and interaction among proteins of each respiratory chain complex (RCC). Mass spectrometry-based technologies, including liquid chromatography (LC) followed by tandem mass spectrometry (LC-MS/MS), matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS), and surface-enhanced laser desorption ionization-time-of-flight mass spectrometry (SELDI-TOF MS), can reveal functional states of the cell based on identities of uncovered proteins. Knowledge about such interactions and functional states paves the way to better understanding of cellular physiology. The innovative high-throughput techniques, in some sense, can be considered disruptive technologies, as they force scientists to transform molecular biology to a systemic and quantitative discipline (65). As shown in Fig. 1, component identification is just a first step; for these “omics” data sets to provide tangible progress, methods must be established to analyze and interpret them, thereby completing the systems biology cycle. The use of computational tools is no doubt essential for the integration and analysis of omics and legacy data. Here we review successes and challenges resulting from recent experimental and computational investigations of the mitochondrion. First, we discuss why it is possible and useful to consider the mitochondrion a system. Second, we review results from proteomic studies, the most popular mitochondrial omics data type, in studying complex diseases associated with this
organelle. Finally, we summarize results from computational studies that integrate high-throughput and legacy data to improve the understanding of mitochondrial biology and mitochondrial diseases.

MITOCHONDRIUM AS A SYSTEM

Mitochondria are semiautonomous organelles. Due to their endosymbiotic origin (80, 81), the organelles retain their genome, bilayer membrane, transcriptional and translational capabilities, high metabolic and signaling activities, and division ability independent from the cell (34). The protein make up of these organelles contains components at all levels of the central dogma of molecular biology. For these reasons, mitochondria are natural candidates for systemic studies. With the mitochondrial genome encoding only a small fraction of mitochondrial proteins, proteomic technology becomes key in the identification of components and functions of this organelle. In addition, because of its manageable number of proteins, the mitochondrial proteome is an attractive target for initial application of new proteomic techniques. Furthermore, as components of the mitochondrial proteome encompass a wide range of hydrophobicity, molecular weight, isoelectric values, and copy numbers, this proteome serves as an ideal benchmark (85). To date, mitochondrial proteomes of many species have been reported, including human (36, 105, 121), rat (75), mouse (26), fruit fly (3), yeast (98), Neurospora crassa (111), rice (43), Arabidopsis thaliana (16), pea (120), and soybean (45). Successes in comprehensiveness and technology advancement in these studies have been reviewed extensively (25, 27, 37, 106); here we focus on the application and analysis of proteomic data in elucidating mitochondrial functions and mitochondrion-associated diseases.

MITOCHONDRIAL PROTEOME AS A DIAGNOSTIC TOOL FOR MITOCHONDRIAL DISEASES

Mitochondria are present in almost all human cell types, with a few exceptions such as red blood cells. Mitochondrial proteomes from various human cells, including cardiomyocytes (36, 121), hepatocytes (82), and T leukemia cells (105), have been identified. The organelle once known only as the “power house” of the cell now emerges as a target for treating a variety of human disorders (131, 132). Although mitochondrial defects are implicated in a wide range of diseases, the more widely studied defects can be loosely classified into three categories: respiratory chain defects, neurodegenerative diseases, and cancer. This classification is used to provide a rough structure to our discussion and is not meant to be mutually exclusive in the underlying disease etiology.

Respiratory Chain Defects: Disease Diagnosis and Characterization by Immunologic Assays

Defects in proteins of the respiratory chain lead to a multitude of mitochondrial disorders, of which the most prominent...
are Leigh’s syndrome (complex I–IV), MELAS (complex I), and Leber’s hereditary optic neuropathy (LHON; complexes I, III, and V). Given the high number of subunits in each complex, it is not surprising that disorders involving the respiratory chain are reported to be more commonly a result of altered assembly than catalytic site dysfunction (42, 70). Therefore, methods for diagnosis and evaluation of such diseases should be capable of detecting misassembly, as well as reduced activity due to catalytic site mutations. In contrast to large-scale proteomic studies seeking to identify the entire set of subunits, monoclonal antibody-based focused proteomics (85) aims to identify specific proteins of interest with the maximal protein integrity. Instead of two-dimensional electrophoresis, which denatures proteins, monoclonal antibodies capture the protein complexes while preserving post-translational modifications and minimizing protein degradation (110). Immunocapture of complexes is also simpler, less expensive, and requires less sample than enzymatic assays. An obvious drawback, however, is that immunocapture can only be done for proteins for which antibodies are obtainable. A list of 35 commercially available monoclonal antibodies (MitoSciences, Eugene, OR) to mitochondrial proteins and their applications has been published (17). This immunologic approach can be applied at various stages of disease characterization, including detection of affected complexes (17, 70, 71), profiling of assembly patterns (84, 124), and measurement of complex activity (62). The following three studies are examples at each of these three stages.

The five RCC have been obtained by immunocapture and resolved by SDS-PAGE (85). Hanson et al. (42) proposed using this combined method to quickly diagnose affected complexes of the different respiratory chain defects. In particular, the cells were simultaneously probed with monoclonal antibodies against the RCC (distinguished by green fluorescent labels) and against porin as a control (distinguished by red fluorescent labels). In normal samples, equal expression of the RCC and porin, i.e., equal signals from the green and red fluorescent signals, makes the cells appear yellow. However, when a complex is low or missing, the dominant signal from porin antibodies gives the cell a red-orange color. Affected proteins, therefore, can be identified simply by viewing the sample with fluorescence microscopy. If antibodies recognizing proteins in the denatured state are also available, Western blots, where specific detection of the antibody is obtained with chemiluminescent substrates, can serve as an alternative (17). In addition, Western blotting is more amenable to protein quantification and allows one to confirm the molecular weights of target proteins.

Triepels et al. (124) examined 11 different patients whose complex I deficiency had been found with enzyme activity measurements. Mitochondria were isolated from fibroblasts of each patient and treated with a mixture of six different antibodies. Western blot signals, normalized to that of control fibroblasts, confirmed that the levels of complex I were diminished in most of the patient cell lines compared with the control. The steady-state levels of fully assembled complex I were also found to depend on expression levels of all the subunits being examined. When any subunit was mutated, the levels of the assembled complexes were reduced. The study also suggested which of the patients most likely had a mutated assembly factor. Besides diagnostic purposes, this immunologic assay also revealed phenotypes for mutations that had previously only been genetically defined. In particular, the mutation T423M of complex I was found to lead to a 25% reduction of catalytic function rather than a misassembly of the complex.

Complexes determined to be deficient in a particular patient by immunologic tests can be subjected to further enzymatic activity quantification. Using spectrophotometry, Kramer et al. (62) developed an automated assay for measuring concentration and enzyme activity of a full panel of respiratory chain enzymes. Enzyme activity was measured via absorbance change produced by a specific electron acceptor or donor in each complex. This assay was recommended for use in conjunction with Western blot analysis, as the spectrophotometric method can detect RCC deficiency in a protein not probed, or alterations in post-translational modification, whereas Western blot shows the physical amount of protein subunits. In particular, Western blot analysis explained the RCC deficiency in a patient who tested normal by the enzymatic activity assay. This combined spectrophotometry-Western blot assay provides rapid, precise results from a crude tissue sample without the need for isolation of mitochondria from tissue homogenate. The authors reported a 91% agreement with the established diagnostic method through a blinded comparison of 10 patients with known diagnoses and 1 healthy control. The applicability of spectrophotometry for an enzymatic assay, however, ultimately depends on the suitability of absorbance spectra of the enzymes.

Neurodegenerative Diseases: Oxidative Damage Identified as the Common Cause

Dysfunction in the electron transport chain and the consequent oxidative damage are also associated with many neurodegenerative diseases, including Alzheimer’s disease (AD), Parkinson’s disease (PD), and amyotrophic lateral sclerosis (ALS) (18). However, as the etiologies of these diseases are still poorly characterized, studies of these disorders tend to be exploratory, rather than focusing on the mechanism and extent of involvement of the respiratory chains on disease pathogenesis and progression. Besides sequencing efforts to discover associated genetic mutations, proteomic techniques have been applied to identify new and differentially expressed proteins in cell line models of these diseases. In particular, isotope-code affinity tags (ICAT) and LC-MS/MS have been employed to analyze mitochondrial proteins isolated from primary neuron cultures (76), dissected substantia nigra from Parkinsonian mice (52), and a mouse embryonic spinal cord-neuroblastoma hybrid cell line (33). Here we briefly describe methods and results obtained from these three studies.

ICAT enables quantitative proteomic analysis of two different pools of proteins based on differential isotopic tagging of each pool (39, 41, 116). The ICAT reagent (Applied Biosystems, Foster City, CA) labels proteins by specifically reacting with thiol groups on cysteine residues. Each sample pool is treated with an isotopically different ICAT reagent, which is either light, through the use of 1H or 13C linker, or heavy, through the use of 2H or 15C linker. The combined sample is proteolytically digested with trypsin. Avidin affinity chromatography is used to capture and isolate peptides containing labeled cysteines, and the resulting fractions are analyzed with

AJP-Cell Physiol • VOL 292 • JANUARY 2007 • www.ajpcell.org

Downloaded from http://ajpcell.physiology.org/ by 10.220.33.3 on June 11, 2017
LC-MS/MS. LC-MS separates and determines the differential expression of each tagged peptide pair from the signal intensity ratios of isotopically light- and heavy-labeled forms. Peptides exhibiting significant differential expression can be further sequenced with the tandem MS. ICAT is particularly useful for samples <100 μg (76) and allows for simultaneous quantification and identification of proteins in one analysis. The method, however, is applicable only to cysteine-containing proteins, which make up about 30% of the total proteome.

AD is the most common neurodegenerative disease, affecting ~10% of Americans >65 yr old and 50% of those >85 yr old (30). Causes for the late-onset, sporadic AD are largely unknown. Numerous causal connections, however, have linked mitochondrial dysfunction with sporadic AD: somatic mitochondrial DNA (mtDNA) mutations, generation of reactive oxygen species (ROS), cytochrome c oxidase deficiency, misregulation of redox-active heavy metals (Fe and Cu), and impaired energy metabolism resulting from these causes (19, 102). Studies of AD have often centered on amyloid-β peptide, a protein frequently found in brains of patients with AD. To investigate the peptide’s functional effects on neurotoxicity, Lovell et al. (76) applied the ICAT labeling technique and LC-MS/MS to quantitatively evaluate changes in mitochondrial proteins of primary neuron cultures exposed to amyloid-β. ICAT analysis of mitochondria from control and amyloid-β-treated cultures identified 45 nonredundant mitochondrial-specific proteins. Ten proteins showed significantly altered expression (P < 0.05), and nine demonstrated a trend toward significance (P < 0.09) in amyloid-β-treated cultures. Among the enzymes with increased expression were those playing essential roles in energy metabolism, such as pyruvate kinase, glyceraldehyde 3-phosphate dehydrogenase, and creatine kinase. Such increased expression was possible because of compensatory response of amyloid-β-treated cells to the declining mitochondrial membrane potentials induced by oxidative stress. Other proteins exhibiting elevated levels were guanine nucleotide-binding protein-β (modulator in signaling pathways), coifin (control of polymerization/denpolymerization), Na+/K+-transporting ATPase (ion balance), voltage-dependent ion channels (VDAC1 and VDAC3), cyclophilin A (interacting with apoptosis-inducing factors), dihydropyrimidinase (axonal outgrowth), and 60S ribosomal protein L4 (protein synthesis). Although the authors could not explain the statistically significant decrease of ATP synthase γ-chain, they concluded that the increased production of the other proteins in primary cortical neurons undergoing apoptosis was perhaps the cell’s effort to maintain ATP production.

PD is the second most common neurodegenerative disorder after AD (52). Pathological characteristics of PD include the progressive and selective loss of nigrostriatal dopaminergic neurons and deposits of the protein α-synuclein called Lewy bodies. The mechanisms underlying PD development and Lewy body formation are not fully characterized, although increasing evidence suggests that mitochondrial dysfunction, oxidative damage, excitotoxicity, and inflammation are contributing factors. Jin et al. (52) employed ICAT in combination with LC-MS/MS to study mitochondrion-enriched fractions isolated from the substantia nigra in mice treated chronically with N-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). The compound MPTP produces pathological and biochemical changes similar to the well-established changes in PD patients (64). The authors identified 318 proteins in mitochondrion-enriched fractions isolated from mouse substantia nigra. Of these proteins, >100 exhibited substantial differences in relative abundance between MPTP-treated and control mice. This is a much higher number of proteins compared with the 45 proteins identified by Lovell et al. (76) (by the same method), because these authors combined results from triplicate experiments using pooled samples of 15 MPTP-treated mice. The authors argued that combining results from multiple experiments was necessary, as current LC-MS/MS technology achieves only 30% reproducibility with moderately complex, identical samples. Besides complex I, which has been shown to be inhibited by MPTP (114), the levels of several subunits of complexes III and V were also shown to decrease after MPTP treatment. This study concludes that MPTP-induced mitochondrial dysfunction leads to an increase in ROS production and a decrease in ATP production.

ALS is a fatal neurodegenerative disease characterized by progressive loss of voluntary motor neurons. Similar to PD, the majority of ALS patients (90%) have the sporadic form of the disease, while only 10% have the familial form (fALS). Approximately 20% of fALS are caused by mutations in superoxide dismutase SOD1, an enzyme that normally removes superoxide radicals by converting them to molecular oxygen to mitigate oxidative damage to the cell. Studies have shown that clinical progression and pathological alteration of motor neurons in sporadic ALS are very similar to those from both fALS patients carrying SOD1 mutations and transgenic mice expressing mutant G93A-SOD1 (33, 38). It is further thought that fALS caused by SOD1 mutations is a result of a toxic gain of function, rather than SOD1 deficiency, because some of these mutations do not seem to affect normal SOD1 activity and SOD1-null mice do not develop the disease (79). Fukada et al. (33) used two independent approaches, two-dimensional gel electrophoresis followed by MALDI-MS/MS and SDS-PAGE followed by LC-MS/MS, to characterize changes in the mitochondrial proteome in the presence of the mutant G93A-SOD1. The chosen model system was a mouse embryonic spinal cord-neuroblastoma hybrid cell line (NSC34) exhibiting motor neuron phenotype. The authors searched for mitochondrial proteins that were altered by G93A-SOD1 under the hypothesis that mitochondrial dysfunction and apoptosis activation play a role in motor neuron death in fALS. Combining results from both approaches, 470 unique proteins were identified in the mitochondrial fraction, of which 75 were newly discovered proteins that had been reported previously only at the cDNA level. Two-dimensional gel electrophoresis identified 40 proteins displaying differential expression, 24 of which were results of different post-translational modifications caused by the G93A-SOD1 mutation. In particular, the authors found five distinct isoforms of VDAC2 and decreased abundance of the channel in G93A-SOD1 cells compared with wild-type cells. Modifications of VDAC2 and apoptotic activation were suggested to be a result of SOD1-mediated toxicity, although possible mechanisms were not discussed. Five chaperones and heat shock proteins were also found to have decreased abundance in G93A-SOD1 cells. The authors suggested that the reduced activity of these proteins somehow facilitated the entry of mutant SOD1 into the mitochondria, thereby inducing mitochondrial abnormality. Wild-type SOD1 was not found in mitochondrial matrix (74, 92). Intriguingly, although fALS has...
never been regarded as a metabolic disease, half of the differentially expressed proteins were those involved in various metabolic pathways, including subunits of complexes I, II, IV, and V. Although the authors did not speculate about the involvement of these metabolic enzymes, it is possible that as the motor neurons deteriorate, the cell reduces its metabolic demand, thereby reducing the requirement for these proteins.

**Cancer: Biomarker Discovery and Clinical Application of Mitochondrial Proteomics**

Genetic alterations in the mitochondrial genome (mtDNA) have been suggested to be the cause of numerous types of cancer (51, 54). In fact, the frequency of mtDNA mutations in cancer cells was reported to be 10 times higher than that of nuclear DNA mutations (127). The involvement of mitochondria in energy metabolism, apoptosis, and cellular oxidative stress responses further implicates the organelle’s role in the development of malignant cells. Mitochondrial proteomes naturally become an attractive target for biomarker identification and monitoring disease progression.

Using the lipophilic cation 4-iodobutyltriphenylphosphonium (IBTP), Lin et al. (72) developed a method to chemically label and measure redox changes in mitochondrial thiols. Driven by the large membrane potential across the inner mitochondrial membrane, IBTP compounds accumulate in the mitochondrial matrix, where they oxidize exposed thiol groups to form a stable thioester. Conversely, thiols already oxidized as a result of cellular oxidative stress are unable to react with IBTP, thereby reducing the amount of label observed. Using this technique, the group showed that free thiol groups on complexes I, II, and IV were vulnerable to oxidative stresses further implicating the organelle’s role in the development of malignant cells. Mitochondrial proteomes naturally become an attractive target for biomarker identification and monitoring disease progression.

**BUILDING THE POWER HOUSE: FROM PROTEOMIC TO SYSTEMS BIOLOGY**

The unifying theme across the proteomic studies described here is the search for a physiological context, mostly in terms of protein expression levels, of the cellular response to a particular disorder or disturbance. This theme reflects the transition of biology from component identification to functional state characterization. In fact, though it is useful to know which mutation causes a particular disease, it is much more beneficial, in terms of therapeutic development, to understand downstream effects of the mutated enzyme’s actions. For example, there are approximately 100 SOD1 mutations attributable to fALS (79), but a cure for the disease awaits discoveries about consequences of the mutations on the cellular and organism systems as a whole. In understanding cellular physiology, it is necessary to conduct experiments and analysis on the cellular scale. Proteomic studies, as well as genomic and microarray experiments, have begun to meet the need for cellular-scale investigation. Cellular-scale analysis can be broken into two components: data integration and quantitative analysis. These two components, together with experimental investigations, complete the cycle that defines systems biology.

**Data Integration: Databases and Network Reconstruction**

Numerous databases have been created to centralize, organize, and facilitate the accessibility of various mitochondrial omic data sets (Table 1). Of these databases, MitoP2 (5, 97) is perhaps the most comprehensive in providing data relevant to mitochondrial proteomic studies. Released in 2004, MitoP2 replaced an earlier version named MITOP. The database combines information regarding the genetic, functional, and pathogenetic aspects of nuclear-encoded mitochondrial proteins. The current version contains data for human, mouse, yeast, *N. crassa*, and *A. thaliana*. Besides data for proteins known to be mitochondrial, referred to as the reference set by the authors, MitoP2 also provides information about putative mitochondrial proteins identified by homology search tools. Each protein entry is annotated with function, chromosomal localization, subcellular localization, homologs and associated confidence values, gene ontology (GO) number, applicable Online Mendelian Inheritance in Man (OMIM; 40), literature references, and cross-references to external databases. Users can search for...
proteins by species, functional category, disease, homology, subcellular localization, or scores from homology calculations.

Network abstractions can be developed to summarize available data and interaction among known components. Networks of proteins and metabolites are the most common and can now be reconstructed at the genome scale (28, 104). Reconstructed networks serve as a highly organized repository for results of a large number of experiments, from which testable hypotheses can be drawn (Fig. 2). Interrogating the properties of these networks allows one to evaluate their accuracy and functions. In particular, the biochemical pathways describing the metabolism of a mitochondrion have been constructed into a network that precisely preserves the interconnectivity of genes, proteins, and metabolites in this organelle (128). This reconstruction integrated data from the mitochondrial genome (4), the heart mitochondrial proteome (87, 121), and the annotated human genome (63, 126), together with a large volume of literature on mitochondrial metabolism. The resulting network, with 189 biochemical reactions and 230 metabolites, described energy metabolism, ROS detoxification, and heme synthesis, as well as nitrogen and lipid metabolism. Every metabolite and reaction in the network was localized to one of three cellular

Table 1. Databases dedicated to mitochondria

<table>
<thead>
<tr>
<th>Database</th>
<th>Website</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>mtDB (49)</td>
<td><a href="http://www.genpat.uu.se/mtDB">www.genpat.uu.se/mtDB</a></td>
<td>mtDNA sequences and polymorphic sites</td>
</tr>
<tr>
<td>GOBASE (113)</td>
<td>megasun.bch.umontreal.ca/gobase/</td>
<td>Genetic maps of completely sequenced mitochondrial genomes and RNA secondary structures of mitochondria-encoded molecules</td>
</tr>
<tr>
<td>Mitomap (15)</td>
<td><a href="http://www.mitomap.org">www.mitomap.org</a></td>
<td>Polymorphisms and mutations of human mtDNA</td>
</tr>
<tr>
<td>MitoProteome (22)</td>
<td><a href="http://www.mitoproteome.org">www.mitoproteome.org</a></td>
<td>Fully annotated human mitochondrial sequences</td>
</tr>
<tr>
<td>MitoDat (67)</td>
<td><a href="http://www.ecb.ncifcrf.gov/mitoDat">www.ecb.ncifcrf.gov/mitoDat</a></td>
<td>Enzymes and genes involved in mitochondrial biogenesis and function</td>
</tr>
<tr>
<td>Human Mitochondrial Protein Database</td>
<td>bioinfo.nist.gov:8080/examples/servlets</td>
<td>Comprehensive data on mitochondrial and human nuclear encoded proteins involved in mitochondrial biogenesis and function</td>
</tr>
<tr>
<td>MitoP2 (5, 97)</td>
<td>ihg.gsf.de/mitop2</td>
<td>Comprehensive list of mitochondrial proteins of several species including human. Links and search tools for datasets relevant to study of mitochondrial proteomes</td>
</tr>
</tbody>
</table>

URLs of each database are provided, however, due to the fluidity nature of domain names, the address can be quickly outdated, and it is advisable that users search for the databases using search engines such as Google.

![Network reconstruction as applied to the glycolytic pathway. A reconstruction of a biochemical pathway starts with identification of enzymes, metabolites, and directionality of each reaction. Left: list of reactions participating in the glycolytic pathway. Each reaction is labeled with an abbreviation of the corresponding enzyme name. Each reaction is written to balance elements and charges of reactants on both sides. Net reaction shows a gain of 2 protons, which can possibly contribute to the proton gradient across the mitochondrial membrane. Right: mathematical representation (S matrix) of the list of reactions. Each element $S_{ij}$ denotes participation of metabolite $i$ in reaction $j$. Reconstruction also directly integrates associations of locus, gene, protein, and reaction for each enzyme when possible. Examples are shown for the enzyme phosphofructose kinase (PFK). PFK consists of 4 subunits and has tissue variations (muscle, liver, platelets) as well as transcript variations (locus 5211 has 2 alternative splice, 5211.1 and 5211.2). Muscle and liver isoforms are homotetramers of the M (PFKM) and L (PFKL) subunits, respectively; platelets contain various heterotetramer combinations such as 3 PFKL and 1 PFKM or 2 PFKL and 2 PFKM (29).](image-url)
compartments: mitochondrial, cytosolic, or extracellular. Metabolites were also characterized by molecular formulas and predominant charge forms determined at pH 7.2. Such procedures for manually reconstructing and curating metabolic networks, i.e., using a bottom-up approach (44, 89), have been detailed in the literature (103). This integration serves as a curated database that can assist in the evaluation of genomic annotation and protein expression studies (88).

The top-down network reconstruction approach, i.e., based on high-throughput data, has also been applied to mitochondrial proteomic data sets. In one of the most comprehensive mitochondrial proteomic studies, Mootha et al. (83) identified 591 mitochondrial proteins using mitochondria samples from mouse brain, heart, kidney, and liver. Combining this data set with mRNA expression, the authors hierarchically clustered 386 genes into subnetwork modules. Each of these gene modules was characterized by tightly correlated gene expression across the four tissues, suggesting shared transcriptional regulatory mechanisms as well as cellular functions among module members. A number of these modules had clear functional associations, such as oxidative phosphorylation, branched-chain amino acid metabolism, steroidogenesis, and heme biosynthesis. The modules were further extended to mitochondrial neighborhoods, which grouped genes that were functionally coordinated with mitochondrial processes. These modules represented a first step toward a systematic, functional characterization of mitochondrial and mitochondrion-associated genes. Advantages of this top-down network reconstruction are that it is relatively quick, automated, and not subject to bias from prior information about the system. One disadvantage, however, is that the method is solely dependent on the often noisy high-throughput data sets (112) and the parameters used for clustering.

Quantitative Analysis: Model Validation and Functional State Determination

A large number of mitochondrial models describing a wide range of mitochondrial functions, including energy transduction (10–12, 50, 61, 69, 100), Ca²⁺ signaling (31, 86, 134), and respiratory regulation (59, 60, 133), can be found in the literature. The increasing accessibility of computational power and availability of data have made the building of larger and more comprehensive models possible. Analysis of model predictions, either validation or falsification, drives the iterative model-building process toward a fuller and more accurate representation of the corresponding biological system. The following are results from some of the recent models of mitochondria.

Kinetics-based modeling. A biological process can be quantitatively described by relating the reaction rate to the change in concentrations of involved metabolites (Fig. 3). The reaction

\[
\begin{align*}
\text{CS:} & \quad \text{acccoa} + \text{h2o} + \text{oa} \rightarrow \text{cit} + \text{coa} + \text{h} \\
\text{ACO:} & \quad \text{cit} \leftarrow \text{icct} \\
\text{ICDH:} & \quad \text{icct} + \text{nad} \rightarrow \text{akg} + \text{coa} + \text{nad} \\
\text{AKGD:} & \quad \text{akg} + \text{coa} + \text{nad} \rightarrow \text{co2} + \text{nad} + \text{succoa}
\end{align*}
\]
rate is primarily dependent on the kinetics of the corresponding enzyme, which, in turn, depends on the enzymatic mechanism, regulatory state, and quantity of available enzymes. A kinetics-based description is perhaps the most straightforward, although not always simple or feasible, approach to quantitatively model biological systems. Kinetic models of mitochondria have been consistently improved in scale and sophistication. Recently, Beard (8) expanded the kinetic modeling approach with thermodynamic principles to study the mitochondrial respiratory system. He stressed that consistency and accuracy of the mathematical description are vital in developing models of biophysical processes that span multiple cellular scales. This thermodynamically consistent model applied differential equations to describe the proton gradient, substrate transport, and ATP synthesis and used distinct state variables for mitochondrial membrane potentials. The basic components included in the model were reactions at complexes I, III, IV, and V, adenine nucleotide translocase, phosphate transport, and cation fluxes. The model accounted for three separate compartments: mitochondrial matrix, intermembrane space, and external space. The overall flux balance was governed by a system of 17 differential equations and 35 parameters. The parameters were categorized into three classes: 16 adjustable parameters, the values of which were determined by fitting model simulations to published data; 17 fixed parameters, the values of which were established in the literature; and 2 free parameters, the values of which were set arbitrarily large. The author cleverly used 9 independent data curves (14) to estimate values for the 16 adjustable parameters. A unique advance made by this model is that the proton gradient and the membrane potential (2 separate components of the electrochemical gradient) are treated as independent variables. Changes in proton concentration were expressed as a function of fluxes by dehydrogenase enzymes, respiratory complexes (I, III, IV, and V), ion transport, and membrane leak. The membrane potential was also considered to be a function of these components but was adjusted by the inner membrane capacitance. Separate treatments of these two components allowed the author to discover effects of phosphate control on the mitochondrial membrane potential. In fact, results from this study strongly suggested that phosphate modulates the activity of complex III in a concentration-dependent manner. This valuable hypothesis can be readily tested in vitro with isolated mitochondria. This study highlights the power of kinetic models in uncovering control points in a system and in driving testable hypotheses that can be used for further experimental study.

Instead of focusing on the pH gradient and reactions of the respiratory chains, the model by Cortassa et al. (20) integrates these components with enzymes in the tricarboxylic acid (TCA) cycle and Ca$^{2+}$ dynamics (20). The major advance of this model over its predecessors (58, 77, 130) is the direct incorporation of Ca$^{2+}$ concentration into the mechanisms of α-ketoglutarate and isocitrate dehydrogenases and explicit examination of their influences on mitochondrial energy metabolism. This unified model was used to explore the coupling of energy demand in response to changes in substrate supply, inhibition of the respiratory chain, and Ca$^{2+}$ level in heart mitochondria. The 12 ordinary differential equations included in the model described the time derivatives of the mitochondrial membrane potential and concentrations of NAD, NADH, ADP, ATP, Ca$^{2+}$, and the TCA cycle intermediates. Specifically, the authors evaluated effects of substrate availability (through the acetyl-CoA level), phosphorylation potential, proton leak, and respiratory inhibition on mitochondrial energetics. A twofold increase in respiration was found when acetyl-CoA varied from 0 to 10 μM. Saturation of oxygen consumption was reached at 5 μM acetyl-CoA. A steep increase in respiration occurred in response to small changes in ADP concentration, indicating a shift from state 4 to state 3 respiration. Inhibition of the respiratory chain, simulated by decreasing the available electron carriers, led to a decrease in membrane potential accompanied by a concomitant decrease in oxygen consumption. The study also addressed the hypothesis that Ca$^{2+}$ influx acts as a control factor for increasing ATP production. Two opposing effects of Ca$^{2+}$ were discussed: an increase in mitochondrial Ca$^{2+}$ stimulated TCA dehydrogenase activities, but the influx of cations into the matrix decreased the inner mitochondrial membrane potential. As a result, positive responses in terms of oxygen consumption, proton gradient, and ATP production to increased Ca$^{2+}$ were observed only when the increase of proton pumping by activated dehydrogenases exceeded the depolarization due to Ca$^{2+}$ entry. In this respect, the model disagrees with observations from experiments with isolated porcine heart mitochondria (122) that high Ca$^{2+}$ concentration continued to produce a positive stimulation of Ca$^{2+}$ on oxygen consumption and ATP synthase activity. The authors suggested that a further incorporation of Ca$^{2+}$ in the ATPase kinetic expression would eliminate this inconsistency. This suggestion forms an interesting hypothesis of Ca$^{2+}$ influence on F$_{1}$F$_{0}$-ATPase activity, which can be verified with theoretical and experimental investigations. Particularly, it will be significant to see whether this positive effect is reproducible at the tissue and organ levels or is an artifact of in vitro experiments. Conclusions from this study were that mitochondrial energetics may be controlled through processes upstream (“push” condition) and downstream (“pull” condition) of NADH production. Stimulatory effects of Ca$^{2+}$ on oxygen uptake, membrane potential, and ATPase activity were higher, percentage-wise, in the push condition, but higher in magnitude in the pull condition. The authors acknowledged that the model could not provide an unequivocal estimate of the relative contributions of control points upstream and downstream of the TCA cycle and suggested that further experiments can help resolve this matter. However, as biological control is a mostly theoretical concept at the present, a precise definition in terms of measurable quantities is necessary before the concept is readily confirmed or refuted by experiments.

Besides scientific findings, the studies by Beard (8) and Cortassa et al. (20) also contributed significantly to the mitochondrial modeling community by publishing all governing equations and values and sources of parameters associated with their models. This contribution is arguably more important than the scientific findings themselves, as they facilitate the reproducibility of the work as well as follow-up studies. It is thus recognized that kinetics-driven models such as these, although useful, require a large number of system- and condition-specific parameters for which values are difficult and laborious to obtain. The lack of data for such parameters is a major hindrance to kinetics-based modeling, rendering it inaccessible for many systems of interest, particularly those not amenable to experimental studies.
Constraint-based modeling. An alternative, data-driven, constraints-based approach has been developed to partially overcome this difficulty (Fig. 3). This modeling approach seeks to narrow the range of possible phenotypes that can be displayed by a metabolic system by imposing physicochemical constraints, rather than precisely determining the exact behavior of the system (24, 96). The constraint-based method is typically used in combination with a reconstructed network representing the system of interest. Quantitatively, the reconstructed network can be represented by a stoichiometric matrix $S$ ($m \times n$), where $m$ is the number of metabolites and $n$ is the number of reactions (103). Each element $S_{ij}$ represents the coefficient of metabolite $i$ in reaction $j$, following the convention that $S_{ij}$ is positive if the metabolite is the product of the reaction and negative otherwise. A zero entry is used when the metabolite does not participate in the reaction. Reversible reactions can be written in either direction. Under the steady-state assumption, the rate of consumption of every metabolite equals its rate of production. This mass conservation relation translates to a system of ordinary differential equations

$$\frac{d[X]}{dt} = S \cdot v = 0 \quad (1)$$

where $X$ represents the vector of metabolite concentrations, $S$ is the stoichiometric matrix, and the variable $v$ represents a flux vector containing the steady-state reaction rate for every reaction in the network. Solutions for $v$ are systematically determined by the successive application of additional constraints, such as those representing directional (reversible vs. irreversible) and enzymatic capacity considerations. These constraints have the following form

$$\alpha_i \leq v_i \leq \beta_i \quad (2)$$

where $\alpha_i$ and $\beta_i$ represent the lower and upper bounds on the steady-state rate of each reaction. Energy-balance constraints have also been developed to disallow fluxes in thermodynamically infeasible internal reaction cycles (9, 94). These bilinear constraints introduce a second set of variables, stored in the vector $\Delta \mu$, which represents the change in chemical potential associated with reactions in the network:

$$K \cdot \Delta \mu = 0 \quad v \cdot \Delta \mu < 0 \quad (3)$$

The matrix $K$ (Eq. 2) stores the null space basis for a matrix $S'$, where $S'$ contains rows and columns in $S$ corresponding only to internal reactions. The resulting solution space often contains a range of possible values, rather than a unique number, for each reaction rate $v_i$ that satisfies the stated constraints.

Equations 1 and 2 are most commonly applied to study metabolic states of cellular systems because of their simple mathematics and guaranteed solutions. Such simplification is based on two fundamental assumptions that must be discussed. First, it is assumed that all biochemical reactions and physical interactions can be written as equations with known participants. Second, the flux distribution $v$ calculated based on Eqs. 1 and 2 assumes a steady state where the change in concentration of every chemical species in the network is approximately zero. Rationales and ramifications associated with these two assumptions are as follows. The first assumption is straightforward for most biochemical reactions where reactants and products are well defined. However, participants and stoichiometry of reactions or interactions involving signaling molecules, activation and inactivation of enzymes, and voltage-gated responses are frequently ill defined, making it difficult to incorporate them into the stoichiometric matrix. There are two implications of the steady-state assumption: 1) there must be no internal buildup of metabolites in the cell, so that the mass conservation equation (Eq. 1) holds perfectly, and 2) observable phenotypes or biological phenomena of interest occur on a time scale longer than the rate at which metabolites are produced and consumed by reactions in the network. Consequently, Eq. 1 and the mentioned constraints apply only to a subset of biological systems and only at a time scale satisfying assumption 2. Specifically, these equations are most appropriately used to investigate time-invariant biological qualities, such as network topology (95) and metabolite pool identification (32), or to characterize gene essentiality (28) and end points of adaptive evolution (47). Successes in these studies have clearly demonstrated the strength of constraint-based modeling in studying such biological behaviors (7, 23, 47). In contrast, the steady-state assumption precludes the use of constraint-based methods to study concentration-dependent behaviors associated with transient or periodic dynamics observed in regulatory (activation, inhibition, feedback) or signaling responses by neuronal and muscular cells. For example, even when the components and interactions of the JAK-STAT signaling network are painstakingly identified, only topological characteristics can be satisfactorily analyzed (91). With respect to the mitochondrial systems, the constraint-based approach cannot be used to model oscillatory behaviors resulting from muscle excitation (6, 21) but is ideal for studying metabolic disturbance due to enzymatic defects (100). These limitations are associated only with the use of Eq. 1 and are not inherent in the constraint-based methodology. In other words, the constraint-based approach allows the use of Eq. 1, i.e., the omission of kinetic data, but it does not preclude the use of such data if they are available. In fact, if such data are available, they can be formulated into constraints to further resolve flux calculation. On the other hand, since the constraint-based approach was developed as a way to avoid the need for kinetic data and the nonlinearity present in kinetic rate laws, most researchers apply the constraint-based method only when they do not have kinetic data and/or want to simplify the calculation. These assumptions are thus not direct results from the constraint-based philosophy but have always been associated with it.

Methods of analysis under the constraint-based framework.

Under the constraint-based modeling framework, flux balance analysis is commonly applied to find a physiologically meaningful $v$ within the solution space (13, 96). Specifically, linear programming is employed to search for a flux distribution $v$ that maximizes a particular end product such as NADH production (Fig. 3). The previously described mitochondrial metabolic network (see Data Integration) was used to assess a broad range of the mitochondria’s physiological functions, such as ATP production, heme biosynthesis, and phospholipid biosynthesis (128). The linear programming formulation assumed that the flux distribution that most efficiently carried out a given mitochondrial function under the constrained cellular resources would most resemble the mitochondrial physiological state in the cell. Notably, the maximal ATP yield per
glucose of the reconstructed network was calculated to be 31.5. Although simple, the significance of this calculation is at least twofold. First, it resolves an inconsistency in a quantity that exists in the literature as both 36–38 (129) and 31 (109, 118). Second, it represents the power of this method to systematically account for both costs and gains of a cellular process. The value of 31.5 ATP per glucose molecule was found through discovery of a net difference of two protons per glucose molecule between the present calculation and those previously reported (109, 118) after all protons consumed in glycolysis, the malate-aspartate shuttle, and phosphate transport are accounted for. These two protons were gained through glycolysis when every reaction was elementally and charge balanced. Such a theoretical calculation does not account for the variable proton leak or electron loss in the respiratory chain that is inherent in all mitochondria and thus should be considered only as an upper-bound approximation.

Due to the common underdetermined nature of constraint-based models, the flux distribution calculated to maximize a particular objective function, ATP demand, for example, is not unique. Alternate pathways, termed alternate optima (78), produce the same optimal objective values. The flux distributions corresponding to these paths are found by setting the maximal objective value as an additional constraint and repeating the search for v. Enumeration of all possible alternate optima is computationally intractable for large networks, but is feasible for sufficiently small systems, including the present mitochondrial metabolic network. All possible pathways that can be used by the mitochondrial network to optimally satisfy each of the three metabolic functions were identified (128). When glucose, fatty acids, and glutamate were simultaneously available to produce ATP, only four optimal flux distributions were found. A much larger number of alternate optimal solutions were found for the maximal synthesis of heme (8,288 solutions) and phospholipids (21,863 solutions), respectively. The number of alternate optima with respect to an objective represents the network redundancy in carrying out the corresponding metabolic function.

For most biological systems, there are an infinite number of theoretical steady-state v satisfy Eqs. 1 and 2. A distribution of possible values for each reaction rate (v_i) can be determined by sampling the entire null space of S. Such results allow one to evaluate how changes in steady-state rates of a subset of reactions systemically affect the rates of the remaining reactions in the network. A random walk algorithm (artificial centering hit-and-run) (55) has been used to evaluate steady-state flux distributions in the cardiac mitochondrion subjected to metabolic disturbances associated with diabetes (123). Effects of diabetes on mitochondrial metabolism were assessed by simultaneous application of constraints reflecting 1) an increased fatty acid uptake via the carnitine-palmitoyltransferase-I shuttle, 2) a decreased glucose consumption due to the reduced number of glucose transporters in cell membrane, and 3) an increased ketone body uptake due to its higher blood concentration (Fig. 4). In the normal condition, allowable flux values (flux values satisfying all imposed constraints) of each reaction formed a broad distribution. The peak of each distribution represented the most probable flux value of that reaction. However, the increased mitochondrial fatty acid uptake in diabetes led to much narrower distributions and hence smaller allowable flux ranges, for most reaction fluxes. In particular, the phospholipid synthesis activity increases sharply, but the range of feasible flux values for maintaining cellular steady state is very small (Fig. 4A). Increasing the ketone body and/or glucose uptake to normal physiological values led to a minimal increase in allowable range of reaction fluxes. Interestingly, the flux through the mitochondrial pyruvate dehydrogenase enzyme was significantly restricted by network stoichiometry when the fatty acid uptake was increased (Fig. 4B). Many studies have tried to identify factors that affect the inhibitory mechanism of pyruvate dehydrogenase under conditions such as diabetes (117, 119); this study showed that an increase in cellular fatty acid uptake flux forced a significantly lower flux through this enzyme as a direct consequence of overall network stoichiometry.

**FUTURE DIRECTIONS: BRIDGING EXPERIMENTAL AND COMPUTATIONAL STUDIES**

With the plentiful experimental data and computational models analyzing such data, what is needed to complete the
systems biology cycle (Fig. 1) is the generation of new hypotheses from such model and model-driven experimental studies) is needed to complete the systems biology cycle (Fig. 1). Since mathematical models provide the most precise and informative representation of current understanding about the corresponding biological system, careful analysis of model predictions is essential for designing subsequent experiments and driving forward the iterative model-building procedure that is at the heart of systems biology. Model-driven experimental studies are rare, but they are not completely absent from the literature. Examples of successful studies include the discovery of regulatory elements in yeast galactose utilization (48), regulatory interactions operating on *Escherichia coli* metabolism (23), and identification of regulatory interactions in the SOS pathway in *E. coli* (35). Nevertheless, in order for systems biology to significantly impact the way biological research is conducted in the future, such studies should become the norm, rather than the exception. Completing each systems biology cycle is not trivial, however, as it requires not only the participation of researchers from multiple disciplines but also the development of a standard for data representation and procedures. The need for such a standard is reflected in the paradox that while the field is clearly flooded with high-throughput and legacy data (1, 53, 56, 108), every modeler, including those whose work is reviewed here, is keenly aware of the absence of data needed to complete his study. In recognizing the needs, the National Institute of Standards and Technology (NIST), with the primary missions are to establish technological standards and promote innovative research, has supported workshops on identifying standards for model systems, methodologies, and data archival for the use of mitochondrial proteomics in heath care (101, 125). Such efforts, however, should not be limited to mitochondrial proteomics, as standards are vital for the future of systems biology. Responsibilities for realizing such standards should be taken up by both experimental and computational biologists. As an experimental biologist, such responsibilities may include the following.

**Standardization of criteria for model systems.** Such criteria should include, but not be limited to, organism choice, relevance of the chosen model systems, availability of samples, and ease of maintenance.

**Standardization of experimental protocols.** Although it may not be possible to specify every detail involved in an experimental procedure, information regarding upper and lower limits of sample types and sizes, reagent grades, duration of measurements, and temperature, among others, would help increase the reproducibility of each experiment and facilitate the development of new protocols.

**Standardization of data presentation.** The primary cause for the data availability paradox is the inconsistency in data presentation across different experimental studies, making it difficult or impossible to directly compare or combine them. A standard for reporting units is therefore absolutely necessary and would hopefully be a natural result of standardized experimental protocols.

**Making samples (cell lines, vector constructs, proteins) accessible to those who are interested.** Such availability would not only facilitate the verification and reproducibility of new findings but will also form new collaborations among researchers.

Responsibilities of computational biologists, on the other hand, may include several criteria.

**Outlining data specificity.** Currently, most available biological data are qualitatively described, and quantitative data are often reported in arbitrary units. In moving toward standardized data representation, modelers should provide guidelines on what is desirable in terms of units, data types (snap shot vs. time series, in vivo vs. in vitro), and number of replications necessary for statistical confidence.

**Providing testable hypotheses.** For modeling studies to drive the systems biology cycle forward, analysis of such models should not only confirm or falsify experimental interpretations but ought to lead to testable hypotheses. Such hypotheses should provide expected results when possible and should be realistic and practical within the constraints of experimental studies.

**Suggesting experimental design.** Computational analyses of experimental data or model prediction can identify which measurements are the most informative. Such results can greatly benefit the design of subsequent experiments.

**Making model and analysis programs available for others.** Many computational studies only briefly describe the underlying algorithms and do not publish associated computational programs, as it is not required by most peer-reviewed journals. However, for results to be reproducible, a documented source code of such programs should accompany every study. Standards for distributing models, such as systems biology markup language (SBML) (46) and minimum information requested in the annotation of biochemical models (MIRIAM) (68), could be adopted, and a centralized database could be created as a repository for these models.

Perhaps one way to accelerate the realization of such standards and to drive systems biology forward is to have experimental and computational researchers meet each other half way. It is probably no longer realistic for each group to stay within the boundary of their respective discipline. To transform biology into a more quantitative discipline, efforts should be contributed by both groups of researchers. Experimentalists should understand the basics of computer-aided analytical tools, and computational biologists should understand the systems beyond the mathematical representation by experiencing experimental biology firsthand at the bench. Through interdisciplinary training, perhaps the next generation of researchers will no longer have to identify themselves as either experimentalist or computational biologist but, rather, biologists of a new era, the systems biology era.

**ACKNOWLEDGMENTS**

The authors thank Andrew Joyce and Dr. John Mongan for critical reading of the manuscript.

**GRANTS**

This work was supported by University of California Systemwide Biotechnology Research and Education Program GREAT Training Grant 2005-246 to T. D. Vo.

**REFERENCES**


