sponsors development of the information technology infrastructure necessary for a National Virtual Observatory (NVO). There is a close cooperation with the particle physics community through the Grid Physics Network (GriPhyN). NASA supports astronomy mission archives and discipline data centers while developing a roadmap for their federation.

Impressively, these projects are all cooperating, and are working toward a future Global Virtual Observatory to benefit the international astronomical community and the public alike. There are similar efforts under way in other areas of science as well. The Virtual Observatory has had and will have significant interactions with other science communities, both learning from some and providing a model for others.

References

1. The Hubble Space Telescope, www.stsci.edu.
2. Chandra X-Ray Observatory Center, http://chandra.harvard.edu.
3. The Sloan Digital Sky Survey website, www.sdss.org.
4. The Two Micron All Sky Survey, www.ipac.caltech.edu/2mass.
5. Digitized Palomar Observatory Sky Survey, www.astro.caltech.edu/~george/dposs.
6. SIMBAD Astronomical Database, http://simbad.u-strasbg.fr.
7. NASA/IPAC Extragalactic Database, http://nedwww.ipac.caltech.edu.
8. *Virtual Observatories of the Future*, American Society for Physics Conference Series, vol. 25, R. J. Brunner, S. G. Djorgovski, A. S. Szalay, Eds., (The Astronomical Society of the Pacific, San Francisco, 2001).
9. X. Fan *et al.*, e-Print available at http://xxx.lanl.gov/abs/astro-ph/0108063.
10. Vizier Service, http://vizier.u-strasbg.fr/viz-bin/VizieR/
11. The Semantic Web, www.w3.org/2001/sw/; Web Services, www.w3.org/TR/wsdl.
12. I. Foster, C. Kesselman, Eds., *The Grid: Blueprint for a New Computing Infrastructure* (Kaufmann, San Francisco, 1998).
13. Public access Web site to the SDSS data, http://skyserver.sdss.org/

VIEWPOINT

Pathway Databases: A Case Study in Computational Symbolic Theories

Peter D. Karp

A pathway database (DB) is a DB that describes biochemical pathways, reactions, and enzymes. The EcoCyc pathway DB (see http://ecocyc.org) describes the metabolic, transport, and genetic-regulatory networks of *Escherichia coli*. EcoCyc is an example of a computational symbolic theory, which is a DB that structures a scientific theory within a formal ontology so that it is available for computational analysis. It is argued that by encoding scientific theories in symbolic form, we open new realms of analysis and understanding for theories that would otherwise be too large and complex for scientists to reason with effectively.

What happens when a scientific theory is too large to be grasped by a single mind? Decades of experimentation by molecular biologists to characterize the molecular components of single cells, combined with recent advances in genomics, have thrust biology into the position where the theoretical understanding of a system such as the biochemical network of *E. coli* is too large for a single scientist to grasp. This situation has a number of disturbing consequences: It becomes extremely difficult to determine whether such theories are internally consistent or are consistent with external data, to refine theories that are inconsistent, or to understand all of the implications of such large theories. As more details of such a complex system are elucidated experimentally, it is not so clear that our understanding of the system as a whole increases if the new understanding cannot be integrated with the larger theory it pertains to in a coherent fashion.

In this article I argue that as scientific theories reach a certain complexity, it becomes essential to encode those theories in a symbolic form within a computer database (DB). I describe pathway DBs as a case study in encoding scientific theories in computers. Although the scientific community clearly accepts the need to encode the ever-expanding quantity of scientific data within DBs, DBs of scientific theories, such as a theory describing the transcriptional regulation of *E. coli* genes, are much rarer. By data I mean measurements made from individual experiments; by theory I mean relationships inferred from the interpretation and synthesis of many experimental results. The biological sciences are particularly well suited to the DB approach because many theories in biology have a qualitative nature; they describe semantic relationships between systems with many different molecular components, and the causal relationships between these components have been measured in a qualitative rather than a quantitative fashion. The DB approach is probably less appropriate for quantitative theories that are best described by systems of differential equations, or other types of mathematical models in analytical form.

The theory of the *E. coli* metabolic network is an example of a theory whose size and complexity are too large for a mind to grasp. The metabolic network is essentially a chemical processing factory within each *E. coli* cell that enables the organism to convert small molecule chemicals that it finds in its environment into the building blocks of its own structures, and to extract energy from those chemicals. The *E. coli* metabolic network, illustrated in Fig. 1, involves 791 chemical compounds organized into 744 enzyme-catalyzed biochemical reactions (*1*). On average, each compound is involved in 2.1 reactions. I posit that the majority of scientists cannot grasp every intricate detail of this complex network. Omission of even a single step from the network can be fatal for the cell.

One might argue that the biomedical literature is one embodiment of the theory of the *E. coli* metabolic network, and that as the biomedical literature enters electronic form, we need not be concerned with the size and complexity of biology theories. Although efforts to bring the biomedical literature online are tremendously useful, there are serious limitations to what they will achieve: We cannot compute effectively with theories within the biomedical literature. Natural-language texts still remain largely opaque to computers, despite many advances in natural-language processing. For example, one relatively simple question we might wish to ask of the *E. coli* metabolic network is how many of its reaction steps are catalyzed by multiple enzymes, meaning they have backup systems, and therefore would targeting a drug toward one of the enzymes catalyzing those steps be ineffective? Answering this question by using a pathway DB such as the EcoCyc pathway DB is trivial, but answering this question by processing the biomedical literature with a computer program would earn the programmer a Ph.D. in computer science.

Pathway Databases

A pathway is a linked set of biochemical reactions—linked in the sense that the product of one reaction is a reactant of, or an enzyme that catalyzes, a subsequent reaction. A pathway DB is a bioinformatics DB that describes biochemical pathways and their component reactions,

Bioinformatics Research Group, SRI International, EK223, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA. E-mail: pkarp@ai.sri.com

enzymes, and substrates. Most pathway DBs created to date describe metabolic pathways, but pathway DBs containing signaling and genetic-regulatory pathways are now beginning to appear. Most pathway DBs contain computable descriptions of pathways structured by using a formal ontology, as opposed to textual or semi-structured descriptions of pathways. A pathway/genome DB (PGDB) integrates pathway information with information about the complete genome of an organism. A PGDB is one type of model-organism DB (MOD), although most MODs do not contain pathway information.

Historically, pathway DBs arose at the intersection of genomics, biochemistry, DBs, and artificial intelligence. Genome DBs aim to catalog the molecular parts of an organism whose genome has been sequenced. They generally focus more on describing the genome map and sequence of the organism than they do on describing the functions of each gene in a structured computable fashion (functions are described using short English phrases). Biochemists have had a long-standing effort to catalog the catalytic activities of known enzymes, in printed form.

The EcoCyc project (*2, 3*) began in 1992 with the goals of integrating within a single DB the then incomplete genome map of *E. coli*, with detailed descriptions of the enzymes and pathways of *E. coli* metabolism. Thus, EcoCyc is a PGDB. By 1996, EcoCyc contained more than 100 pathways and 520 enzymes, entered by Riley's group at the Marine Biological Laboratory. Enzymes were connected to their genes, and to the genome sequence, when known. The DB contained detailed descriptions of the reaction catalyzed by each enzyme, the range of substrates the enzyme would accept, the chemicals known to activate or inhibit the enzyme, and its subunit structure. The DB also described each small molecule enzyme substrate. EcoCyc pathways include those for biosynthesis of cellular building blocks such as amino acids and cell wall components, those for catabolism of the different carbon sources that *E. coli* can utilize, and those for extracting energy from chemical compounds.

In 1998 we integrated the complete genome sequence of *E. coli* determined by the Blattner laboratory into EcoCyc (*4*), and we expanded the EcoCyc collaboration to include additional biologists curating information on other aspects of the *E. coli* biochemical machinery: the transporters that move small molecules from the outside of the cell to the inside, and the mechanisms by which *E. coli* gene expression is controlled at the transcriptional level. Version 5.6 of EcoCyc released in June 2001 describes 162 *E. coli* transporters and 629 transcription

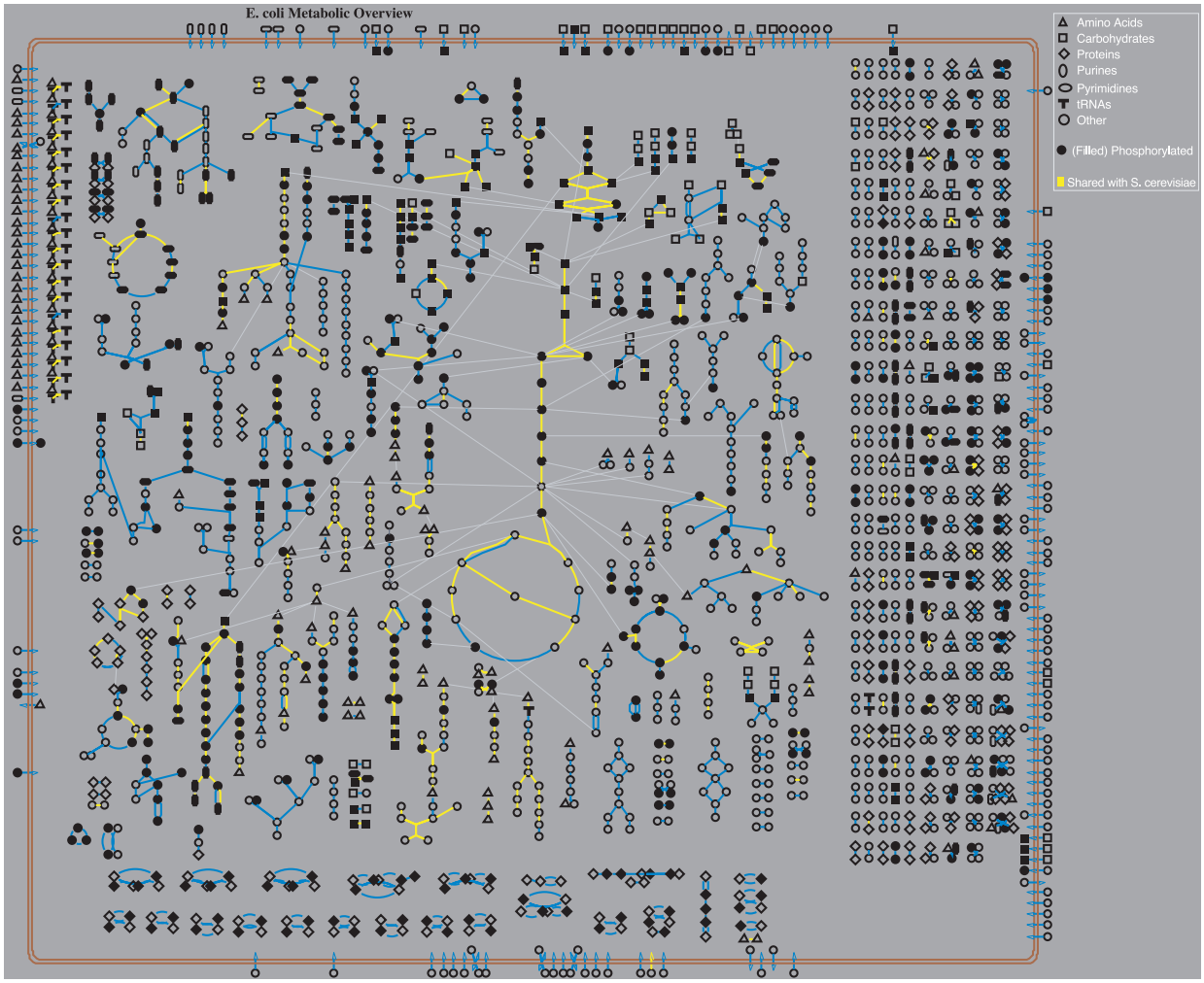

Fig. 1. An overview of the full known metabolic map of *E. coli*. Each blue and yellow line in this diagram represents a single enzyme-catalyzed reaction; each node represents a single metabolite. Glycolysis is in the middle of the diagram, biosynthetic pathways are to its left, catabolic pathways are to its right, and reactions that have not been assigned to a pathway are grouped along the far right-hand side. The shape of each metabolite encodes its chemical class: for example, amino acids are shown as triangles, and shaded nodes indicate phosphorylated compounds. This diagram was produced by combining automated graph layout algorithms with some manual positioning of regions of this graph. The reactions highlighted in yellow show the results of a species comparison between *E. coli* and the metabolic network predicted for the yeast *S. cerevisiae* from its genome by the PathoLogic program. Reaction lines in blue indicate reactions found in *E. coli* only; reaction lines in yellow indicate reactions found in both *E. coli* and *S. cerevisiae*. The species comparison does not generally include the transport reactions shown in the cellular membrane that surrounds the diagram.

units within *E. coli*, and a total of 165 *E. coli* metabolic pathways.

Other pathway DBs include KEGG (*5*), WIT (*6*), MetaCyc (*2*), the Connections Map DB (see www.stke.org), and UM-BBD (*7*); for a review see (*8*).

In parallel with the development of EcoCyc, SRI developed a software environment called the Pathway Tools (*9*) that supports query, analysis, and visualization operations for PGDBs. The Pathway Tools allows users to query EcoCyc DB by a variety of criteria including name matching and classification hierarchies (such as a taxonomy of metabolic pathways). Query results are displayed by using a library of visualization tools that automatically generate drawings of metabolic pathways, reactions, chemical compounds, chromosomes, and transcription units. A visualization tool called the Overview diagram depicts the full metabolic map of *E. coli* and its transporters. This tool is a powerful device for understanding global properties of the *E. coli* metabolic network. For example, the user can ask the software to highlight all occurrences of a given metabolite to understand all the pathways that can operate on it. The user also can highlight all metabolic reactions that are activated or inhibited by a specific metabolite at the substrate level, or whose transcription is controlled by a given transcription factor, to understand the regulation of the metabolic network. We have also developed a technique for painting gene expression data sets onto the Overview, thus providing an organizing framework to aid the interpretation of these complex data sets in a pathway context (see http://ecocyc.org:1555/expression.html).

More recently, the Pathway Tools has been generalized so that it can manage PGDBs for multiple organisms simultaneously. In the course of 1 week, a user of the PathoLogic component of the Pathway Tools (see below) can create a new PGDB for a sequenced microorganism. Once created, users can refine the PGDB by using a suite of interactive editors within the Pathway Tools, and can publish the PGDB on the Web. We have thus created a reusable "generic model-organism DB" toolkit that is now being used to create PGDBs as resources for the scientific communities that research many microorganisms. The SRI Web site at http://ecocyc.org contains PGDBs for eight microorganisms that were created by using the Pathway Tools.

Computing with Biological Theories in Pathway Databases

The artificial intelligence subfield of knowledge representation is concerned with devising symbolic encodings of complex collections of information in a manner that supports inference (reasoning) processes across that information. Key strategies in knowledge representation include (i) devising an ontology (DB schema) that captures important semantic distinctions in an accurate fashion, and that precisely defines the meaning of different DB fields, (ii) rigorously following the definitions in that ontology to encode a theory within the DB, and (iii) extending the ontology when new domain concepts are found to fall outside the scope of the ontology. The EcoCyc ontology (*10*) contains about 1000 classes that encode key concepts in biochemistry and molecular biology, and more than 200 *slots* that define properties of and relationships among those classes.

The EcoCyc DB is structured according to this ontology and consists of an interconnected web of frames (objects) stored in a frame knowledge representation system (similar to an object-oriented DB). Each frame represents a distinct biological object (such as a gene or a protein), and the labeled connections between those frames represent distinct semantic relationships among the objects, such as the relationship of a gene to its protein product, or the relationship of a protein to a reaction that it catalyzes.

Reasoning across such a DB is accomplished through computations that traverse this network, and represents a distinct nonnumerical style of computing that, in our experience, most scientists are not familiar with. It is symbolic computing that allows us to exploit the power of computational qualitative theories. The following sections explore global computations that can be applied to pathway DBs.

Computing global properties of a pathway DB. By integrating the fragments of biological theories that are scattered through the biomedical literature, we can discover global properties of those theories that were heretofore elusive. Ouzounis and Karp wrote a set of programs that computed statistics on relationships in the DB that showed how very simplistic is the classical notion of one gene, one enzyme, and one reaction (*1*). The program found that *E. coli* contains 100 enzymes that catalyze more than one biochemical reaction, 68 cases where the same reaction is catalyzed by more than one enzyme, and 99 cases where one reaction is used in multiple *E. coli* pathways.

More recently, Karp and Collado studied the global properties of the *E. coli* genetic network stored within EcoCyc, which is the most detailed model of the genetic network of any organism. The model describes the control of 630 *E. coli* transcription units (containing 27% of all *E. coli* genes) by 97 transcription-factor proteins. Figure 2 shows a visualization of the *E. coli* genetic network defined by EcoCyc. A number of interesting properties are present in this network. A significant portion of the network is unconnected, that is, 50 of the 85 transcription factors do not regulate other transcription factors—they regulate other genes that do not encode transcription factors, and, in some cases, they regulate themselves. The network is not very deep—it has a maximum depth of three nodes. Only two transcription factors (CRP and Fnr) directly control more than two other transcription factors. Aside from autoregulation (when a transcription factor directly controls its own expression), there are only two feedback loops in this graph (between MarR and MarA, and between GutR and GutM). Negative autoregulation is the dominant form of feedback.

Jeong *et al.* studied the topology of the protein interaction network for the yeast *Saccharomyces cerevisiae*, and found that the network topology is heterogeneous and scale-free, meaning that there are relatively few highly connected nodes, and that the probability of finding a network node (protein) with many connections (interactions) follows a power law (*11*). They also found that deletion of proteins with high numbers of interactions was more likely to be lethal to the organism. In earlier work, they found that metabolic networks are also scale-free (*12*).

Pathway Prediction from Sequenced Genomes

Another form of inference with a Pathway DB occurred in 1995 when Karp, Ouzounis, and Paley demonstrated that the EcoCyc DB could be used to predict the metabolic network of an organism from its genome (*13*). The PathoLogic program that they developed takes two inputs: an annotated genome sequence that includes the locations and predicted functions of genes within the genome, and a reference pathway DB. The output produced by the program is a new PGDB that includes a set of pathways predicted to be present in the organism by PathoLogic. PathoLogic uses SRI's MetaCyc pathway DB (*2*) as the reference DB; MetaCyc contains 450 pathways from many different organisms. PathoLogic matches enzymes in the annotated genomes against enzymes in the MetaCyc DB and computes a score for the presence of different pathways on the basis of the number of matching enzymes, and their positions within the pathway.

Prediction of Pathway Flux Rates

Schilling and Palsson have devised a numerical method to predict the reaction flux rates for the entire metabolic network of an organism (*14*). Given experimental measurements of the mass composition of metabolites within the cell, and a list of all metabolic reactions known to occur in the cell (such as those provided by EcoCyc), an optimization procedure is used to calculate the equilibrium rates at which substrates are processed by each metabolic reaction. The flux rates predicted by their technique have been verified experimentally. An extension of this computational technique was used to predict the lethality of *E. coli* deletion mutants; it correctly predicted the growth potential of mutant strains in 86% of the genes examined (*15*).

Consistency of Metabolic Network with Cellular Growth-Media Requirements

The growth of *E. coli* can be supported by a number of alternative chemically characterized growth media, such as a combination of glucose, ammonia, and minerals. The *E. coli* metabolic network is able to synthesize all of the compounds essential for its growth (such as the amino acids and nucleoside triphosphates) from these simple precursors. Therefore, it should be possible to verify the completeness and correctness of the theory of the *E. coli* metabolic network embodied in the EcoCyc DB by computationally propagating the known chemical nutrients of *E. coli* through the EcoCyc network, and determining whether the resulting chemical products included all of the compounds known to be essential for growth.

A qualitative simulation of the *E. coli* metabolic network is obtained—qualitative in the sense that the approach is not attempting to predict the quantities of chemical products that *E. coli* metabolism produces over time, but to predict *if* certain chemical products can be produced from a given set of precursors. This qualitative simulation was obtained by using a computational device called a production system, which is the basis of many expert systems. A production system consists of a set of rules (corresponding to reactions) and a set of propositions currently listed as true in the *working memory* of the production system (which correspond to metabolites). Each rule is of the form $A \wedge B \rightarrow C \wedge D$, meaning that if A and B are present in working memory, then add the presence of C and D to working memory. A production-rule *inference engine* repeatedly searches for a rule for which all of the propositions on its left side are present in working memory, and fires the rule by adding the propositions on its right side to working memory.

Every metabolic reaction of the form $A + B = C$ was computationally translated into a production rule of the form $A \wedge B \rightarrow C$. And the working memory of the production system was initialized to contain a proposition for each known growth nutrient of *E. coli*. The initial outcome of the resulting qualitative simulation (*16*) was that known *E. coli* growth media could *not* produce the compounds essential for *E. coli* growth. Examination of these results revealed (i) the existence of bugs in the EcoCyc DB (since corrected); (ii) the existence of metabolic intermediates required for the production of some essential compounds, but for which the metabolic synthesis route is unknown; and (iii) that we had neglected to include certain important precursors in our simulation, such as some proteins that are metabolic substrates (such as thioredoxin and acyl carrier protein). Once these discrepancies were resolved (*16*), it could be verified that all 41 essential compounds could be produced from the M63 minimal growth medium.

Discussion

Pathway DBs have several purposes. They are encyclopedic references for pathway information that can be queried by scientists who want to search out specific facts, or search for patterns. They can also be queried by computer programs that perform global analyses and pattern searches. The symbolic nature of pathway DBs means that many types of programs and inference procedures can be written to compute with this information. Although one might argue that computational theories have existed for many years in the form of Fortran programs that model physical systems, the procedural nature of a Fortran program means that it can be used in only one way, namely, to be executed. The knowledge representation community has long recognized the flexibility that results from symbolic representations (*17*).

Pathway DBs are a mechanism for easing the cognitive overload produced by genome data: An analysis of the pathway content of a microbial genome reduces the complexity of that genome by allowing the scientist to think in terms of hundreds of pathways rather than in terms of thousands of gene products. Similarly,

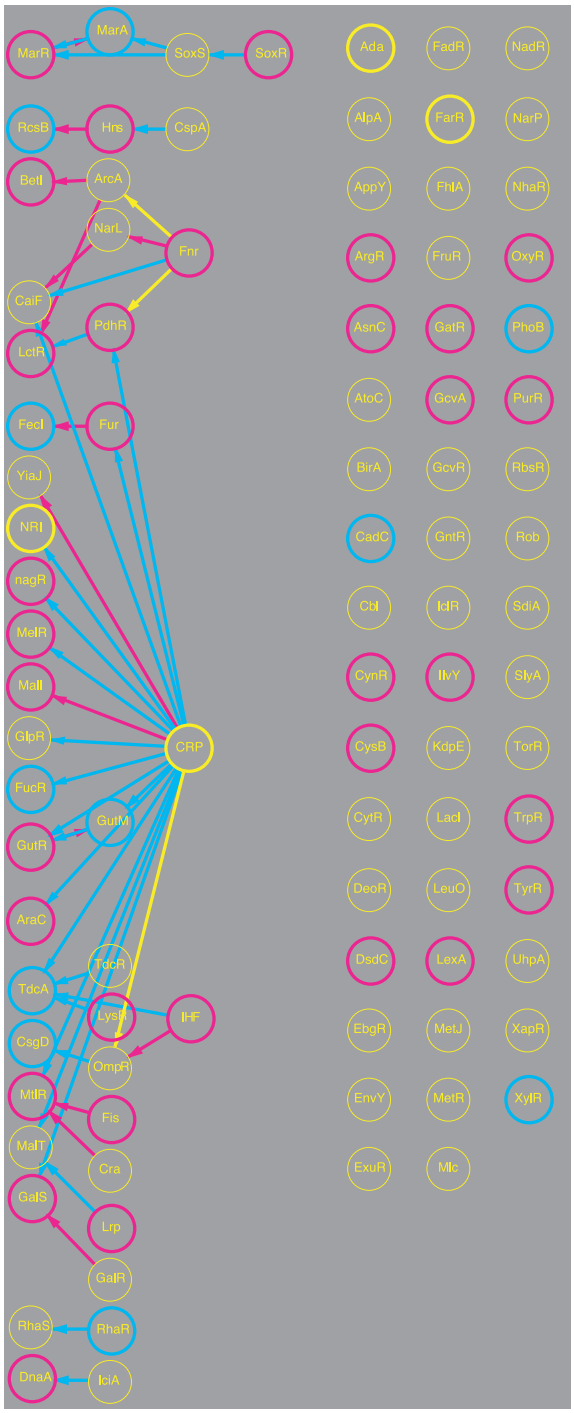

Fig. 2. A visualization of the known network of transcription factors in EcoCyc. Each circle represents a single transcription factor. A blue arrow from protein *A* to protein *B* indicates that *A* activates the transcription of *B*. Pink arrows indicate repression of transcription, and yellow arrows represent both positive and negative regulation of transcription. Circles with a blue outline represent positive self-regulation. Pink circles represent negative self-regulation. Yellow circles with a thick outline represent both positive and negative self-regulation. Circles with a thin yellow outline do not regulate their own transcription. Thus, for example, IHF represses transcription of its own genes and those of OmpR, but activates transcription of the genes for TdcA. The majority of transcription factors depicted here regulate the transcription of many other genes; this diagram shows only regulation of other transcription factors.

pathway DBs can impose an organizing framework on complex gene expression (or proteomics) data sets that facilitates their interpretation.

Future challenges for pathway DBs include modeling of large signaling networks in eukaryotic organisms; performing automated layout similar to that shown in Fig. 1 of the much larger pathway networks that exist in eukaryotic organisms, and supporting methods for user navigation through such a larger pathway network; defining standard ontologies for exchange of pathway data among different DBs and application programs; and creating new analysis algorithms for extracting new insights from pathway networks, such as to aid drug design by analyzing diseased human pathway networks, or predicting optimal drug targets for antimicrobial drug design.

One lesson for computer scientists provided by pathway DBs (and by other bioinformatics applications) concerns the importance of DB content to solving computational problems. Most computer scientists focus their attention on algorithms, thinking that the best way to solve a hard computational problem is through a better algorithm. However, for problems such as predicting the pathway complement of an organism from its genome, or predicting metabolic products that an organism can produce from a given growth medium, I know of no algorithms that can solve these problems without being coupled with an accurate and well-designed pathway DB.

By encoding scientific theories in a symbolic DB, scientists can more easily check those theories for internal consistency and for consistency with external data, can more easily refine theories that are found to violate external data, and can more easily assess the global properties of the system that such a theory describes. The genome revolution is increasing the need for pathway DBs in the biological sciences, and similar developments will occur in other sciences. However, effective implementation of this paradigm is hampered because most biologists (and most other scientists) receive essentially no education in DBs or knowledge representation. Although many scientists learn a computer programming language as part of their undergraduate education, introductory programming courses completely omit DB and knowledge representation concepts such as data models, ontologies, DB query languages, logical inference, DB design, and formal grammars—which explains why many biological DBs do not have a regular syntactic structure, much less a consistent or precisely defined semantics. As science enters the information age, it is crucial that the computer-science education that scientists receive covers symbolic computing as well as numerical computing.

References and Notes

1. C. Ouzounis, P. Karp, *Genome Res.* **10**, 568 (2000).
2. P. Karp *et al.*, *Nucleic Acids Res.* **28**, 56 (2000).
3. P. Karp, in *Nucleic Acid and Protein Databases and How To Use Them* (Academic Press, London, 1999), pp. 269–280..
4. F. Blattner *et al.*, *Science* **277**, 1453 (1997).
5. M. Kanehisa, S. Goto, *Nucleic Acids Res.* **28**, 27 (2000).
6. R. Overbeek *et al.*, *Nucleic Acids Res.* **28**, 123 (2000).
7. L. Ellis, C. Hershberger, L. Wackett, *Nucleic Acids Res.* **28**, 377 (2000).
8. P. D. Karp, *Trends Biochem. Sci.* **23**, 114 (1998).
9. P. Karp, M. Krummenacker, S. Paley, J. Wagg, *Trends Biotechnol.* **17**, 275 (1999).
10. P. D. Karp, *Bioinformatics* **16**, 269 (2000).
11. H. Jeong, S. P. Mason, A.-L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (2001).
12. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabasi, *Nature* **407**, 651 (2000).
13. P. Karp, C. Ouzounis, S. Paley, in *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology,* D. States, P. Agarwal, T. Gaasterland, L. Hunter, R. Smith, Eds. (American Association for Artificial Intelligence, Menlo Park, CA, 1996).
14. C. Schilling, B. Palsson, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4193 (1988).
15. J. Edwards, B. Palsson, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5528 (2000).
16. P. Romero, P. Karp, in *Proceedings of the Pacific Symposium on Biocomputing,* R. Altman, T. Klein, Eds. (World Scientific, Singapore, 2001), pp. 471–482.
17. T. Winograd, in *Representation and Understanding* (Academic Press, New York, 1975), pp. 185–210.
18. J. Collado-Vides, I. Paulsen, M. Riley, and M. Saier are collaborators on the EcoCyc project. J. Collado-Vides assisted in the analysis of the *E. coli* genetic network. S. Paley produced Fig. 2. C. Ouzounis, T. Garvey, A. Rzhetsky, and I. Sim provided valuable comments on this manuscript. The development of EcoCyc, MetaCyc, and the Pathway Tools has been funded by grant 1-R01-RR07861-01 from the Comparative Medicine Program of the NIH National Center for Research Resources.

VIEWPOINT

Limits on Silicon Nanoelectronics for Terascale Integration

James D. Meindl,* Qiang Chen, Jeffrey A. Davis

Throughout the past four decades, silicon semiconductor technology has advanced at exponential rates in both performance and productivity. Concerns have been raised, however, that the limits of silicon technology may soon be reached. Analysis of fundamental, material, device, circuit, and system limits reveals that silicon technology has an enormous remaining potential to achieve terascale integration (TSI) of more than 1 trillion transistors per chip. Such massive-scale integration is feasible assuming the development and economical mass production of double-gate metal-oxide-semiconductor field effect transistors with gate oxide thickness of about 1 nanometer, silicon channel thickness of about 3 nanometers, and channel length of about 10 nanometers. The development of interconnecting wires for these transistors presents a major challenge to the achievement of nanoelectronics for TSI.

Silicon technology has advanced at exponential rates in both performance and productivity throughout the past four decades. From 1960 to 2000, the energy transfer associated with a binary switching transition—the canonical digital computing operation—decreased by about five orders of magnitude and the number of transistors per chip increased by about nine orders of magnitude. Such exponential advances must eventually come to a halt imposed by a hierarchy of physical limits. The five levels of this hierarchy are defined as fundamental, material, device, circuit, and system (*1*). A coherent analysis of the key limits at each of these levels reveals that silicon technology has an enormous remaining potential to achieve TSI of more than 1 trillion transistors per chip, with critical device dimensions or channel lengths in the 10-nm range. This potential represents more than a three-decade increase in the number of transistors per chip and more than a one-decade reduction in minimum transistor feature size compared with the state of the art in 2001. Fundamental physical limits that are independent of the characteristics of any particular material, device structure, circuit configuration, or system architecture are virtually impenetrable barriers to future advances of TSI.

Binary switching transitions implemented with transistors are indispensable to performing computation in a digital system. The energy transfer per binary transition is a revealing metric for comparing the performance of

School of Electrical and Computer Engineering, Microelectronics Research Center, Georgia Institute of Technology, Atlanta, GA 30332–0269, USA.

*To whom correspondence should be addressed. E-mail: james.meindl@mirc.gatech.edu