

The Pathway Tools cellular overview diagram and Omics Viewer

Suzanne M. Paley* and Peter D. Karp*

Bioinformatics Research Group, SRI International, EK207, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA

Received February 8, 2006; Revised March 6, 2006; Accepted May 14, 2006

ABSTRACT

The Pathway Tools cellular overview diagram is a visual representation of the biochemical network of an organism. The overview is automatically created from a Pathway/Genome Database describing that organism. The cellular overview includes metabolic, transport and signaling pathways, and other membrane and periplasmic proteins. Pathway Tools supports interrogation and exploration of cellular biochemical networks through the overview diagram. Furthermore, a software component called the Omics Viewer provides visual analysis of whole-organism datasets using the overview diagram as an organizing framework. For example, gene expression and metabolomics measurements, alone or in combination, can be painted onto the overview, as can computed whole-organism datasets, such as predicted reaction-flux values. The cellular overview and Omics Viewer provide a mechanism whereby biologists can apply the pattern-recognition capabilities of the human visual system to analyze large-scale datasets in a biologically meaningful context. SRI's BioCyc.org website provides overview diagrams for more than 200 organisms. This article describes enhancements to the overview made since a 1999 publication, including the automatic layout capability, expansion of the cellular machinery that it includes, new semantic zooming and poster-generating capabilities, and extension of the Omics Viewer to support painting of metabolites, animations and zooming to individual pathway diagrams.

INTRODUCTION

Whole-organism analyses, whether computational or experimental, require whole-organism visualization tools. The human visual system has powerful pattern-recognition capabilities, but to harness that power, visual information must

be presented within a meaningful organization. Our Pathway Tools software (1) produces a pathway-based visualization of cellular biochemical networks, called the cellular overview diagram. The software supports interrogation and exploration of the network. A component called the Omics Viewer provides visual analysis of whole-organism measurements using the biochemical network as an organizing framework. For example, gene expression and metabolomics measurements, alone or in combination, can be painted onto the cellular overview diagram, as could computed whole-organism datasets, such as predicted reaction-flux values, or predicted essential genes.

Pathway Tools can automatically generate a cellular overview diagram for an organism from a Pathway/Genome Database (PGDB) describing the genome and biochemical networks of the organism. A PGDB can in turn be automatically generated from the annotated genome sequence of an organism. Thus, at the push of a few buttons, a scientist can generate from an annotated genome sequence a powerful device for interpreting system-biology studies of that organism. And as a PGDB is updated to reflect evolving scientific knowledge of the organism, its overview can be automatically updated to reflect those updates. SRI's BioCyc.org website provides overview diagrams for more than 200 organisms (2).

The cellular overview includes metabolic, transport and signaling pathways, and other membrane and periplasmic proteins, and can be displayed at user-selected magnification. The Omics Viewer can produce animated visualizations of whole-organism datasets, and it allows the user to inspect individual pathways at high resolution.

The overview and Omics Viewer have been under development since 1997 (3)—our group was the first to create a cellular pathway diagram that could be used to interpret omics data. This article reports on a number of recent enhancements to the software since 1999 (3) including automated layout capabilities, addition of periplasmic and membrane reactions and proteins, new zooming and web navigation capabilities, a novel poster-generating facility, and an extension of the Omics Viewer to support visualization of metabolite profiling data, customizable color schemes, animations for time course data and the superposition of high-throughput data onto individual pathway diagrams.

*To whom correspondence should be addressed. Tel: +1 650 859 5904; Fax: +1 650 859 3735; Email: paley@ai.sri.com

*Correspondence may also be addressed to Peter D. Karp. Tel: +1 650 859 4358; Fax: +1 650 859 3735; Email: pkarp@ai.sri.com

ORGANIZATION OF THE CELLULAR OVERVIEW DIAGRAM

The cellular overview diagram is a graph that depicts many aspects of the cellular biochemical network, including metabolic and transport reactions and pathways, bacterial signaling pathways, and periplasmic and membrane proteins. A sample cellular overview diagram is shown in Figure 1. Visual elements of the diagram encode the following aspects of the biochemical network.

Nodes in the overview represent biochemical species including small-molecule metabolites and, in some cases, proteins. The shape of a node indicates the type of chemical compound (e.g. triangles depict amino acids, squares depict carbohydrates and diamonds depict proteins). Nodes are filled to indicate phosphorylation of the molecule.

Lines between nodes represent the biochemical or transport reactions that occur in the cell. Transport reactions are shown as lines that span a membrane. Directionality of transport is indicated by arrowheads. Most pathways flow in the downward direction in the diagram (although reaction directionality is not shown explicitly, display of reversibility information where available will be supported in an upcoming version of the software).

The cellular overview diagram does not show all possible connections between pairs of metabolites in an organism. To do so would result in a diagram that was largely unreadable. Rather, the overview diagram shows all pathways defined in an organism's PGDB. When multiple pathways have been

combined within a single superpathway, the superpathway is shown rather than the individual pathways. To the extent that individual metabolites and reactions appear in more than one pathway, they will be duplicated in the overview diagram.

Pathways with related biological functions are grouped together in the diagram, with shaded background boxes delineating these groupings. The MetaCyc pathway ontology defines these groupings; examples for ontology classes include pathways of cofactor biosynthesis and pathways of carbohydrate utilization. The overall organization of the diagram positions biosynthetic pathways on the left, degradation pathways on the right and energy metabolism pathways in the middle. Any metabolic reactions in the PGDB that are not members of pathways appear to the right of the degradation pathways, laid out in tabular fashion. For PGDBs such as EcoCyc that contain signal transduction pathways (these are not inferred computationally, but can be created by curators), these pathways appear at the bottom.

Surrounding these pathways, we draw a border representing the cell membrane. For Gram-negative bacteria, we draw an inner and outer membrane with a periplasmic space in between, whereas for other organisms we draw only a single membrane. This representation is suitable for most bacteria—the cell architecture of more complex organisms is not currently represented in the cellular overview. Transport reactions are shown crossing the appropriate membrane(s), and any reactions that take place in the periplasmic

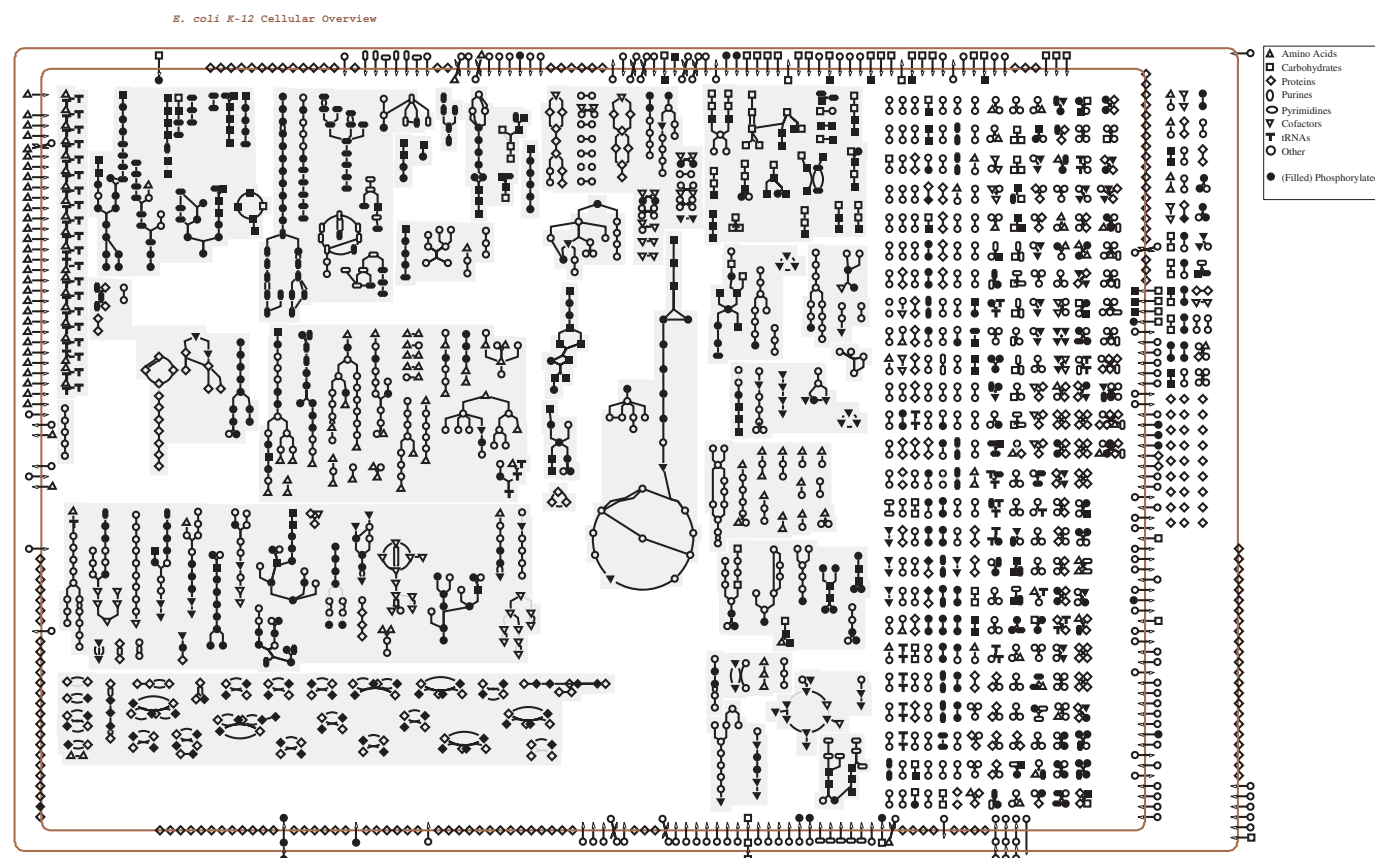


Figure 1. The cellular overview diagram for *Escherichia coli* K-12, from the EcoCyc database.

or extracellular space are drawn there. In addition, any proteins that are not enzymes but whose location has been assigned to a membrane or extra-cytoplasmic space will be drawn in the correct location (the overview does not explicitly draw cytoplasmic proteins, as there are too many of these to make their display useful).

The cellular overview diagram is generated entirely automatically from the data in a PGDB. Automated generation is a major advantage over previous versions of the Pathway Tools software (in which overview generation was partially automated but required significant manual oversight) and over other software systems, as it frees the curator from mundane and time-consuming layout tasks. It also means that as data are curated—pathways and reactions added or deleted, protein cellular locations updated, and so on—the curator does not need to worry about how this might affect the layout. With one simple command, the overview can be regenerated to include all the latest data updates. Automated layout is further essential for generating organism-specific overviews for each member of the BioCyc collection of more than 200 PGDBs; that large number of DBs precludes manual creation of separate overviews for each.

QUERYING THE OVERVIEW

The Overview provides querying capabilities to allow the user to explore the cellular networks encoded by this diagram. (The Pathway Tools software can run in either of two configurations: as a standalone desktop application, or as a web server. Although the cellular overview diagram is available to users under both configurations, their query capabilities differ somewhat, with more operations available to users of the desktop application.)

Mousing over part of the diagram generates a description of which pathway, compound and so on the mouse pointer is over. Clicking on an object produces a menu of links to pages showing details of the selected pathway, reaction or metabolite. In addition, the magnification of the entire diagram can be altered. Clicking on a pathway pops up a magnified view of the pathway, showing all the individual metabolite and gene names. This magnified view, which is a new feature of the software, allows users to quickly orient themselves and focus in on pathways of interest. It also greatly enhances the utility of the Omics Viewer, as described below.

The overview diagram is a particularly valuable tool for understanding and interrogating cellular networks. One of the requirements for effective pathway visualization systems identified by Saraiya *et al.* is to be able to show multiple pathways simultaneously, with interconnections between them (4). (We note that the evaluation performed by Saraiya and colleagues contains many errors of omission in failing to note where Pathway Tools (they use the term EcoCyc) satisfies many requirements listed in Tables 2–4 in their study.) In our cellular overview diagram, users can click on any metabolite and ask to see connections drawn between that node and every other place where that metabolite appears. They can do this for more than one metabolite, or for all the metabolites in a specified pathway. Thus, the user can quickly focus in on all connections of interest, while avoiding the visual clutter that would result if the overview always showed

all possible pathway interconnections. In addition, users can query and selectively highlight different elements of the overview in different colors. Users can search for a particular metabolite, gene, protein, reaction or pathway and highlight all the places where that object appears (we highlight reaction lines as a surrogate for the enzymes and genes that catalyze them).

Aside from metabolite connections, many other types of relationships can be explored. Users can search for and highlight, for example, all proteins with a particular cellular location, all reactions whose enzymes are activated or inhibited by a particular compound, all genes whose expression is regulated by a particular transcription factor, all reactions with multiple isozymes or any one of a number of other pre-defined queries. They can also highlight a list of genes, reactions and so on imported from a user-supplied file. A new zooming capability allows the diagram to be shown at different levels of magnification, which can make highlights easier to see. Figure 2 shows a section of a zoomed overview diagram, including selected metabolite connections and highlights.

The cellular overview diagram can also be used for comparative analyses of the complete metabolic networks of two or more organisms. Given the display of an overview for one organism, the software can highlight all reactions that are either shared with, or not shared with, other combinations of organisms for which PGDBs are available. For example, given the *Bacillus subtilis* overview, a user interested in antimicrobial drug discovery in Gram-positive bacteria might request a highlighting of reactions shared between *B. subtilis*, *Bacillus anthracis* and *Streptococcus pneumoniae*. The user could request highlighting of reactions present in all these organisms, or reactions present in *B. subtilis* and at least one of the other organisms selected, or choose to look at only those reactions present in all these organisms. This comparison is not performed at the sequence level, since the question of whether the organisms share common enzymatic activities is orthogonal to whether the enzymes that catalyze those activities share sequence similarity. Rather, two organisms are considered to share a reaction if the PGDBs for the organisms both specify that some enzyme catalyzes that reaction.

A new feature of the Cellular Overview, which will be available in our summer 2006 release for the desktop software only, is a semantic zooming capability. As the user zooms in on a region of the overview diagram, first reaction arrowheads become visible, then pathway labels and metabolite names, and finally, at the highest semantic zoom level, enzyme names, and gene names become visible.

A new command provides the capability to print out the overview diagram at its maximum zoom level as a poster-sized metabolic chart (formatted as a postscript file). The ability for the owner of a PGDB to automatically generate a high quality wall chart of their organism's metabolism at the push of a button is unique to Pathway Tools. Previous metabolic charts, such as the Boehringer-Mannheim chart (5), are mosaics of pathway information from many organisms, and do not precisely describe the metabolism of any one organism. In contrast, metabolic charts of Pathway Tools are organism specific, and may be generated from any of the 200 PGDBs in the BioCyc collection, or from PGDBs created by other groups. Examples of metabolic

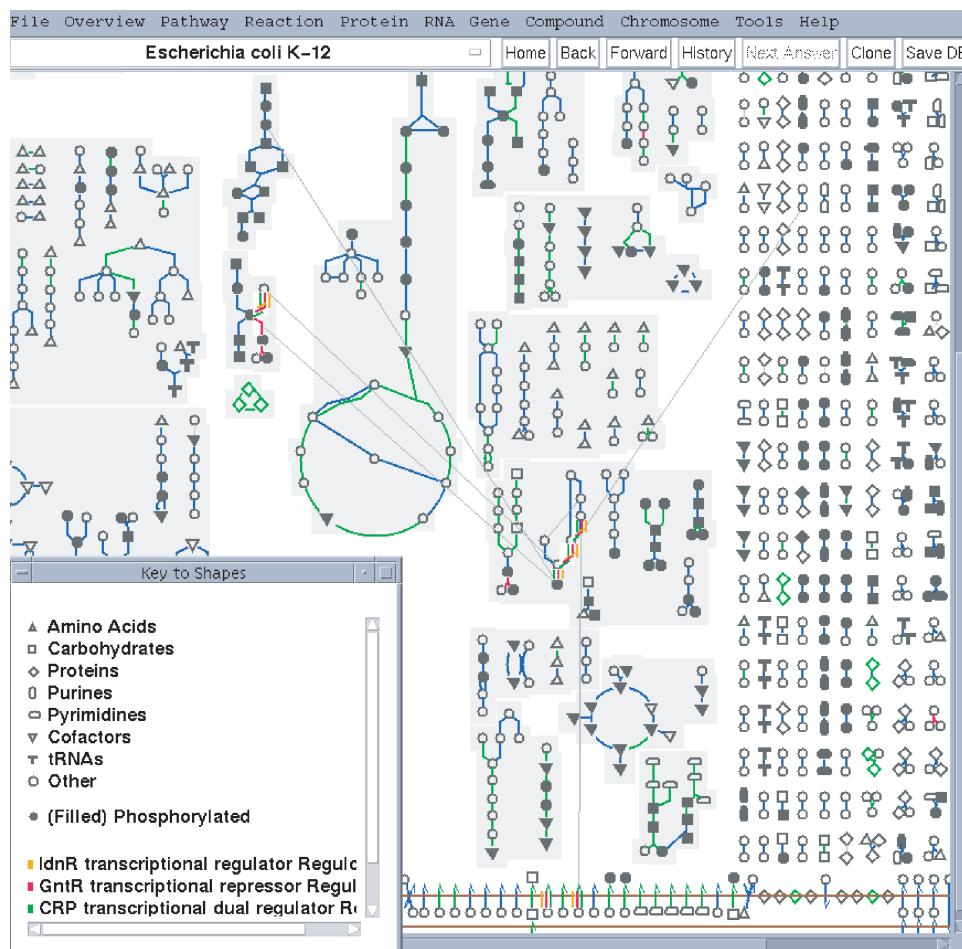


Figure 2. A portion of the cellular overview diagram for *E.coli*. The pathway in the center of the image is ketogluconate metabolism. The diagram shows all connections from metabolites in this pathway to other pathways or reactions. This pathway is regulated by three transcription factors—the regulon for each of these transcription factors is highlighted in a different color, as shown in the key.

charts for *Caulobacter crescentus* and *Mycobacterium tuberculosis* are included with this paper as Supplementary Data.

THE OMICS VIEWER

One of the most exciting uses of the cellular overview diagram is as a vehicle for pathway-based visualization of high-throughput experimental data. The Omics Viewer is so named because it supports overview-based visualizations—both static and animated—of data from a variety of different kinds of ‘-omics’ experiments, such as gene expression data, metabolomics data, proteomics data or any other kind of experimental data in which numerical data values are assigned to many individual genes, proteins, reactions or metabolites. Being able to relate pathway networks to quantitative high-throughput data is another key requirement identified by Saraiya *et al.* for pathway visualization systems (4).

So that experimental data can be visualized using the Omics Viewer, each value in a user-supplied dataset is assigned a color. For example, in a very simple color scheme, a log ratio >2 might be assigned red, values between 2 and −2 might be assigned black, and values less than −2 might

be assigned green. In the web version, a few predefined color schemes are available, including a default scheme computed automatically based on the range of the provided data. On the desktop version, the color scheme is fully customizable. Each node or line in the cellular overview diagram is then drawn in the color corresponding to the value for the particular gene, protein, reaction or metabolite. The result is a color-coded metabolic chart that enables the user to instantly see which pathways or key metabolic or transport steps are turned on or off under particular sets of conditions. An example display can be seen at <http://biocyc.org/expr-examples/single-expt.html>.

Many experiments, such as time-series experiments, involve a series of data points per gene, metabolite, and so on. In these situations, the Omics Viewer offers the ability to view the output as an animation, with one frame per time point. The resulting animation can be viewed all together as a movie, or can be stopped and advanced, one time point at a time. An example animation display is available at <http://biocyc.org/expr-examples/animation.html>.

The Omics Viewer takes as input a simple tab-delimited file containing the omics dataset. The first column contains the name or ID of a gene, protein, reaction or metabolite.

The name can be any name or synonym that is uniquely associated with the entity in the PGDB. Following the first column can be any number of data columns. The user specifies whether the first column contains genes, metabolites, proteins, reactions or a mixture of types, and which subsequent data column or columns are to be used to generate the display—if multiple data columns are specified then an animated display is produced. The user indicates whether the numbers represent relative or absolute measurements, whether they should be taken from a single column or as the ratio of two different columns and whether the values are based on a log or a linear scale. For microarray experiments, the desktop version of the Omics Viewer can also accept as input the output file generated by SAM [Significance Analysis of Microarrays (6)], a plug-in for the Microsoft Excel spreadsheet program that combines the results of multiple repetitions of a single microarray experiment to produce lists of statistically significant positively and negatively regulated genes.

Most input files contain only a single type of data (i.e. genes or metabolites, but not both) but it is also possible to combine multiple types of data into a single file so that one display can be produced showing a combination of, for example, gene expression data and metabolite profiling data from the same set of experimental conditions. In this case, the metabolite profiling data would color the nodes in the diagram (representing individual metabolites), and the gene expression data would be used to color the edges between the nodes (representing reactions, enzymes and the genes that code for them). When combining different types of data in this fashion, keep in mind the caveat that the scales for the two types of data must be the same in order for the results to be meaningful. This limitation can be avoided in the desktop version, where it is possible to overlay the data from multiple files, specifying a different color scheme for each to reflect differences in scale.

If a reaction has multiple isozymes, then multiple genes or proteins will be associated with a single reaction line, each with its own associated data value. However, on the full Cellular Overview Diagram there is room to draw only a single color for each line, so only one of the data values can be chosen for display (we try to choose the value expected to be of most interest to users: the highest value for absolute measurements, or the one that shows the greatest deviation from 0 or 1 for relative measurements). We have recently addressed this issue in the desktop Omics Viewer by allowing users to zero in on pathways of interest by navigating from the Omics Viewer to an individual pathway display, where they can see all their omics data superimposed on a much larger picture of that pathway. In this individual pathway display, multiple isozymes with different data values cause the reaction arrows to consist of multiple parallel lines, with one color for each isozyme. When accessing the Omics Viewer over the web, we achieve the same effect, not by including omics data on regular pathway pages, but by showing it on magnified pathway images that pop up when the user clicks on a pathway in the cellular overview diagram—see Figure 3 for an example.

The software can now also generate a table of magnified views of all pathways that have one or more experimental values exceeding a specified threshold, with omics

data painted onto the pathway. An example is shown in Figure 4.

The Pathway Tools' capabilities for overlaying multiple kinds of data onto both the overview as a whole and onto individual pathways, for showing time course data as an animation, and for allowing users to specify their own color schemes are intended to maximize the opportunities for scientists to derive biologically meaningful insights from their data. Example uses of the Omics Viewer by life sciences researchers are as follows.

- The Conway laboratory investigated the relative significance of individual metabolic pathways for the colonization of *Escherichia coli* in the mouse intestine by analyzing time series and gene knockout gene expression data using the Omics Viewer (7).
- The Schoolnik laboratory investigated how the metabolism of *Vibrio cholerae* changes, and which metabolic pathways are selectively up- or down-regulated as the organism transitions from an aquatic environment to the mammalian gut, by analyzing gene expression experiments using the Omics Viewer.
- Researchers at the Max Planck Institute of Molecular Plant Physiology used the Omics Viewer with a combination of gene expression and metabolite profiling data to identify and analyze changes in the metabolism of *Arabidopsis thaliana* under conditions of sulfur starvation (8). The Supplementary Data attached to this study includes figures showing the overview diagram for *Arabidopsis* superimposed first with gene expression data and then with metabolite profiling data.
- Researchers at TAIR (The *Arabidopsis* Information Resource), who maintain and curate a PGDB for *A.thaliana*, use the Omics Viewer with metabolite profiling data to quickly identify pathways involving metabolites whose quantities change significantly during the course of an experiment. They also use the Omics Viewer to point out gaps in knowledge in their PGDB, e.g. if a metabolite profiling experiment identifies important metabolites that are missing corresponding pathways in the overview diagram, this information can suggest targets for future curation and research efforts.

RELATED WORK

The KEGG database (9), to which the BioCyc databases are often compared, provides a single overview map for all organisms in KEGG, which contains links to functionally related groups of reference pathways. That diagram is not queryable or interactive in any other way, and there is no way to find out what individual nodes or lines represent other than by navigating to an individual reference pathway map. Pathway Solutions, Inc. provides a web CGI tool for mapping omics data onto individual KEGG reference pathway maps (10) (but not onto the full overview diagram), accepting as inputs either gene names, EC numbers or metabolite IDs. This tool does not offer animations or customized color schemes.

Reactome (11) has produced an analog of our cellular overview diagram. Their diagram can be used for navigation and querying. It focuses primarily on human pathways, but can

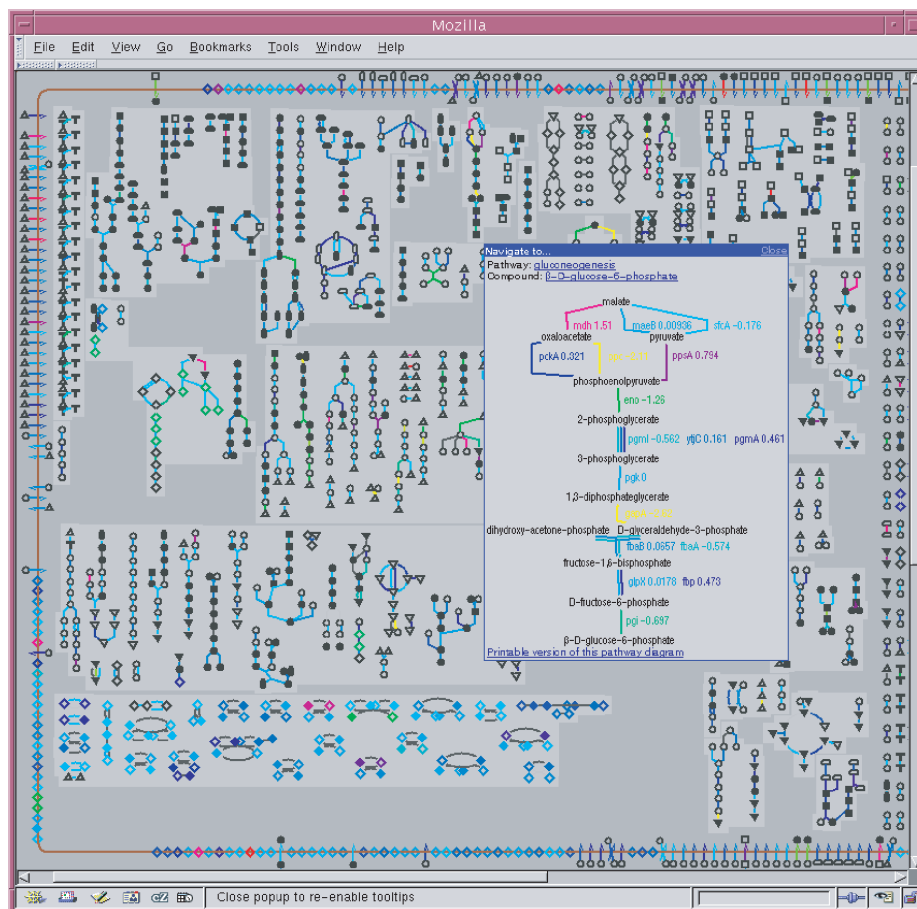


Figure 3. The cellular overview diagram for *E.coli* overlaid with data from a gene expression experiment, showing a magnified view of the gluconeogenesis pathway. Notice that in this magnified view, isozymes are distinguished as parallel lines within one reaction arrow, and gene names and their corresponding data values are drawn in the color for that data value.

show subsets applicable to 14 additional eukaryotic model organisms. The Reactome SkyPainter tool maps omics data onto the reaction arrows in the overview diagram, including generating animations for time series data. Because the Reactome overview contains only reaction arrows, and not individual metabolites, it is not suitable for displaying metabolomics data. It cannot show omics data on individual pathways, nor can it distinguish isozymes. Because all the overviews for the different model organisms are shown as subsets of the human overview, many pathways are missing from the overviews for other organisms. For example, although Reactome has an overview for *A.thaliana*, it includes only those pathways shared between humans and *Arabidopsis*, and does not contain any additional plant-specific pathways.

Other available packages allow mapping of experimental data onto pathway images in various forms. Examples are GenMapp (12), VitaPad (13) and ArrayXPath (14). These packages show only single pathways (their pathways are often defined somewhat differently), rather than a full cellular network.

These examples attest to general recognition of the need for a biochemical overview map, but because the diagrams associated with these other tools are laid out by hand, their customizability is limited. Our automated layout algorithms

allow us to generate a customized overview diagram for each organism, and they can quickly regenerate the diagram when new pathways are added or spurious ones deleted.

FUTURE WORK

We are considering future extensions to our software to expand the abilities of users to visualize omics data in the context of cellular networks, including (i) extending the cellular overview to include representations of different types of gene products, such as transcriptional regulators, so that these genes and proteins can also be visualized using the Omics viewer; (ii) adding eukaryotic subcellular compartments to the overview; (iii) creating generic non-organism-specific overviews (such as the one for bacteria, another for plants) containing all the relevant MetaCyc pathways and (iv) adding further query operations, such as query by Gene Ontology ID.

AVAILABILITY

The Cellular Overview Diagrams for more than 200 organisms are freely available to all at <http://BioCyc.org>. The web version supports navigation and most of the Omics Viewer functionality, as indicated above. Accessing these features of the website requires a modern javascript-enabled

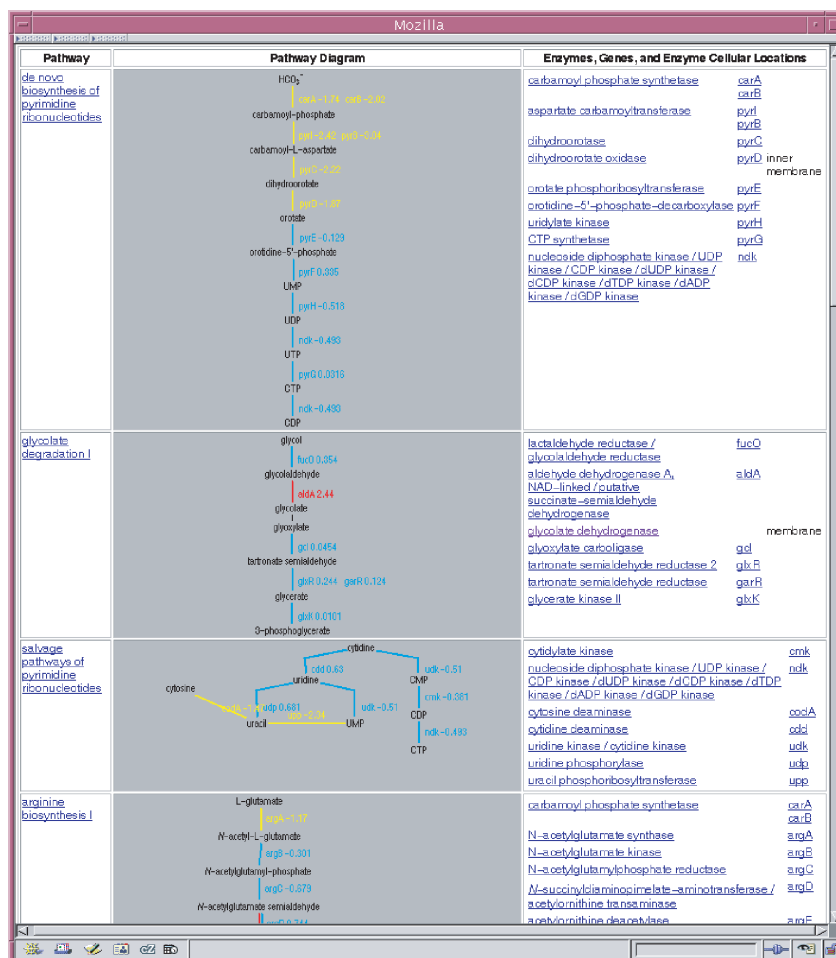


Figure 4. A portion of a table of *E. coli* pathways that are significantly up- or down-regulated as measured by a gene expression experiment. The pathway diagram shows all genes with an expression log ratio >1 in red, and all those with a ratio less than -1 in yellow. The column on the right lists each enzyme in the pathway, its associated genes and, where available, its cellular location.

web browser such as Firefox, Safari or Internet Explorer version 4.0 or later. Some additional PGDBs are hosted by other institutions on their own websites, such as TAIR and SGD, and the same functionality should be available there.

The locally installed software/database bundle for Pathway Tools and BioCyc, which supports the additional desktop operations described in this paper and allows users to generate PGDBs and overviews for their own organisms, is available for Sun workstations, and for PCs running either Windows or Linux. It is available free to academic or non-profit institutions for research use; a fee applies to other forms of use. See <http://biocyc.org/download.shtml> for more details and download instructions.

SUPPLEMENTARY DATA

Supplementary data are available at NAR online.

ACKNOWLEDGEMENTS

This work was supported by grants GM70065, GM75742 and RR07861 from the National Institutes of Health. The

contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health. Funding to pay the Open Access publication charges for this article was provided by NIH GM70065.

Conflict of interest statement. None declared.

REFERENCES

- Karp, P.D., Paley, S. and Romero, P. (2002) The Pathway Tools Software. *Bioinformatics*, **18**, S225–S232.
- Karp, P.D., Ouzounis, C.A., Moore-Kochlacs, C., Goldovsky, L., Kaipa, P., Ahren, D., Tsoka, S., Darzentas, N., Kunin, V. and Lopez-Bigas, N. (2005) Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res.*, **33**, 6083–6089.
- Karp, P., Krummenacker, M., Paley, S. and Wagg, J. (1999) Integrated pathway/genome databases and their role in drug discovery. *Trend. Biotechnol.*, **17**, 275–281.
- Saraiya, P., North, C. and Duca, K. (2005) Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Inform. Visualiz.*, **4**, 191–205.
- Michal, G. (1982) *Biochemical Pathways Wall Chart*. Boehringer Mannheim GmbH Biochemica.

6. Tusher, V.G., Tibshirani, R. and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci., USA*, **98**, 5116–5121.
7. Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S. *et al.* (2004) Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc. Natl Acad. Sci., USA*, **101**, 7427–7432.
8. Nikiforova, V.J., Kopka, J., Tolstikov, V., Fiehn, O., Hopkins, L., Hawkesford, M.J., Hesse, H. and Hoefgen, R. (2005) Systems rebalancing of metabolism in response to sulfur deprivation, as revealed by metabolome analysis of *Arabidopsis* plants. *Plant Physiol.*, **138**, 304–318.
9. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y. and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
10. Arakawa, K., Kono, N., Yamada, Y., Mori, H. and Tomita, M. (2005) KEGG-based pathway visualization tool for complex omics data. *In Silico Biol.*, **5**, 419–423.
11. Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L. *et al.* (2005) Reactome: a knowledge base of biological pathways. *Nucleic Acids Res.*, **33**, D428–D432.
12. Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. and Conklin, B.R. (2002) GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nature Genet.*, **31**, 19–20.
13. Holford, M., Li, N., Nadkarni, P. and Zhao, H. (2004) VitaPad: visualization tools for the analysis of pathway data. *Bioinformatics*, **15**, 1596–1602.
14. Chung, H.J., Kim, M., Park, C.H., Kim, J. and Kim, J.H. (2004) ArrayXPath: mapping and visualizing microarray gene-expression data with integrated biological pathway resources using Scalable Vector Graphics. *Nucleic Acids Res.*, **32**, W460–W464.