**ABSTRACT**

**Motivation:** Metabolic networks combine metabolism and regulation. These complex networks are difficult to understand and visualize due to the amount and diverse types of information that need to be represented. For example, pathway information gives indications of interactions. Experimental data, such as transcriptomics, proteomics and metabolomics data, give snapshots of the system state. Stereoscopic virtual environments provide a true three-dimensional representation of metabolic networks, which can be intuitively manipulated, and may help to manage the data complexity.

**Results:** MetNet3D, a 3D virtual reality system, allows a user to explore gene expression and metabolic pathway data simultaneously. Normalized gene expression data are processed in R and visualized as a 3D plot. Users can find a particular gene of interest or a cluster of genes that behave similarly and see how these genes function in metabolic networks from MetNetDB, a database of Arabidopsis metabolic genes that behave similarly and see how these genes function in metabolism that is integrated more effectively by abstracting key relationships.

**Availability:** MetNet3D and some sample datasets are available at http://www.vrac.iastate.edu/research/sites/metnet/Download.htm

**Contact:** julied@iastate.edu

**Supplementary information:** Color snapshots and movies are available at http://www.vrac.iastate.edu/research/sites/metnet/Bioinformatics/SupplementaryInformation.htm

**INTRODUCTION**

MetNet3D integrates interactions between large-scale metabolic networks and gene expression data in a 3D, stereoscopic and fully immersive virtual reality (VR) system. VR has been shown to be a powerful tool to visualize, explore and manipulate complex data from a variety of disciplines (Cruz-Neira et al., 1993a). VR may be particularly well suited for metabolic networks due to the complexity of structures that represent metabolic processes. These structures and their relationships are not well known to scientists, and are difficult to investigate and understand on a large scale through traditional biological research and analysis methods. VR offers a unique perspective to scientists by immersing them in their data and giving them an intuitive control for navigation and exploration of the metabolic network. In such an environment, a scientist can observe internal complexities, follow a particular pathway and gain better understanding on the different relationships among all the elements of the networks. Furthermore, VR enables a group of scientists to simultaneously investigate the network by stepping into the projection system, providing an excellent “interactive whiteboard” to validate ideas, communicate results and discover new concepts.

While VR is being successfully applied to scientific visualization (e.g. medicine and architecture), there has been emerging research in the VR community to study the benefits of VR in abstract information visualization. Many studies have shown that users learn how to manipulate and understand data more quickly in VR environments than in conventional 3D desktop visualizations. For example, in a study conducted by Nelson et al. (1999) a group of experienced statisticians were asked to analyze data using their conventional software and also using a VR environment. The group in the VR environment was almost twice as effective (in terms of time, fewer errors and extracting more information) at finding patterns and relationships than the group using the conventional tools. Ware’s studies (Ware et al., 1993; Ware and Franck, 1994) quantified the benefits of using VR for complex graph visualization. Ware found that for any given error rate, graphs viewed in stereoscopic VR can be three times as large, in terms of the number of nodes, as graphs projected onto a 2D plane. Maletic et al. showed that software visualization in VR makes it easier to understand programs in development, maintenance and reengineering stages than UML (Unified Modeling Language) diagrams (Maletic et al., 2001).

Most of the current methods for visualizing biological networks use 2D graph models to represent pathways (Karp and Paley, 1994; Becker and Rojas, 2001; Dickerson et al., 2001; Schreiber, 2002, 2003). These 2D graph-based models of metabolic networks are
overloaded since edges and nodes must represent multiple concepts such as chemical reactions, rates, cell compartment identification and lab test results. 3D graph visualization can create a novel integrated information workspace for the study of metabolic networks. The extra dimension gives greater flexibility for placing nodes and edges, and edge crossings can always be avoided. Graph layout algorithms have been successfully applied to 3D space to use more dimensions for pathway data display. Brandes et al. (2003) focus on comparing pathways across species. Each pathway is drawn in one layer; then the layers are ordered in the third dimension such that the most similar pathways are adjacent. Rojdestvenski (2003) statically models metabolic networks as graphs in 3D space using Virtual Reality Modeling Language (VRML). Previous work by Dickerson et al. (2003b) displays 3D networks and allows the user to interactively visualize reactions of interest within the entire metabolic network in stereoscopic VR.

Work has also been done to combine gene expression and pathways within the metabolic networks, mostly by superimposing gene expression data onto static KEGG and AraCyC pathway diagrams (Wolf et al., 2000). In FCModeler (Dickerson et al., 2003a), a program for the dynamic display modeling of regulatory and metabolic networks, color animation is used to highlight the gene expression values in the display.

The work presented in this paper builds on these techniques to produce 3D network representations to be explored in VR. One critical aspect is to automatically generate and lay out those networks based on user’s actions and information requests. The algorithms are described in detail later in this paper.

### SYSTEMS AND METHODS

#### MetNet3D implementation

MetNet3D is written in C++ and Java. It uses three application programming interfaces (APIs), all of which are open-source: VRJuggler (http://www.vrjuggler.org/) is a flexible development platform for VR applications (Bierbaum et al., 2001); OpenSG (http://www.opensg.org/) is an OpenGL based API for graphics scene construction and rendering (Reiners et al., 2002); and R (http://www.r-project.org/) is a statistical application for online clustering and other data processing (Ihaka and Gentleman, 1996). This work uses the C6, a six-wall surround-screen projection-based virtual environment based on the CAVE system (Cruz-Neira et al., 1993b) equipped with stereo glasses, a six degree-of-freedom head tracker and joystick (http://www.vrac.iastate.edu/about/labs/c6/). The user has a stereoscopic view of the geometric scene representing the pathway information and gene expression data. The head tracker enables the stereo rendering of the scene based on the position and orientation of the viewer’s head. The joystick is used to navigate through the metabolic network and to choose nodes from a pathway network or genes from the 3D plot.

MetNet3D can run on a wide range of virtual reality and visualization platforms because it is built on VRJuggler. It provides abstraction from the underlying display and interaction platform, therefore allowing applications to run in a range of immersive environments such as the C6, head-mounted displays, workbenches and others, as well as in conventional desktop environments. Figure 1 shows users interacting with the same application in different platforms.

#### Visual metaphors for metabolic networks

Figure 2 shows the data flow diagram in MetNet3D. The XML file representing the metabolic network is loaded and converted into a graph model (A detailed description of the XML file format is available at http://www.vrac.iastate.edu/research/sites/metnet/xmlFileDescription.htm). The graph model is then converted into a scene composed of 3D geometric glyphs. The colors and shapes of the glyphs provide information about the nodes and edges within the graph, such as node type and edge type. The geometric glyphs in MetNet3D include spheres, cubes, cylinders and cones. A sphere or a cube represents a node. A cylinder with a cone on top forms an arrow for...
The immersive system to the desktop. Tweek in MetNet3D is that the application can be transparently moved from and pathway names. One of the main advantages of using VR Juggler and loaded metabolic network, such as EC numbers, gene family information interface lists textual annotations for the nodes and edges in the currently used in a Tablet PC in an immersive environment. In MetNet3D, the Tweek Navigating through the entire network gives the user a good understanding of Global layout of the entire pathway network ‘hclust’ package (Ihaka and Gentleman, 1996) so that genes with similar conditions. All genes are clustered using clustering algorithms from the R connected line in 3D, showing the expression levels of a gene under different in expression values. The series of spheres connected by cylinders form a reflected by the accumulation of its corresponding RNA. Values are indicated by color and height. The cylinder connecting two spheres shows the change in expression values. The series of spheres connected by cylinders form a connected line in 3D, showing the expression levels of a gene under different conditions. All genes are clustered using clustering algorithms from the R ‘hclust’ package (Ihaka and Gentleman, 1996) so that genes with similar expression patterns are close to each other in the plot.

User interface for detailed text information
One of the challenges is that nodes and edges in metabolic networks contain more information than those revealed by the visual metaphors mentioned in Tables 1 and 2 of Supplementary information, e.g. the subcellular compartment information for nodes and strength information for edges. Some of this information may not have an easy visual representation or it may clutter the user’s field of view. Furthermore, during the exploration of the network, scientists may want to know exact component names and numerical values associated with them. A floating text panel for this kind of information may block users’ sight of the network structure. MetNet3D resolves this problem by using a portable tablet PC to present textual information. A Java Graphical User Interface (GUI) that can be operated from a Tablet PC is connected to MetNet3D through a wireless network. This GUI, called Tweek, is part of the VR Juggler suite that enables compatibility with traditional keyboard and mouse interactions (Hartling et al., 2002). Figure 1a shows Tweek being used in a Tablet PC in an immersive environment. In MetNet3D, the Tweek interface lists textual annotations for the nodes and edges in the currently loaded metabolic network, such as EC numbers, gene family information and pathway names. One of the main advantages of using VR Juggler and Tweek in MetNet3D is that the application can be transparently moved from the immersive system to the desktop.

Visual metaphors for gene expression data
MetNet3D represents microarray expression data as a 3D plot, as shown in Figure 3. Each sphere represents an expression level of a given gene as reflected by the accumulation of its corresponding RNA. Values are indicated by color and height. The cylinder connecting two spheres shows the change in expression values. The series of spheres connected by cylinders form a connected line in 3D, showing the expression levels of a gene under different conditions. All genes are clustered using clustering algorithms from the R ‘hclust’ package (Ihaka and Gentleman, 1996) so that genes with similar expression patterns are close to each other in the plot.

Global layout of the entire pathway network
Navigating through the entire network gives the user a good understanding of the data structure and serves as the starting point of the exploration. The layout method for the entire network in MetNet3D is a modified 3D Graph-Embedder a directed edge. Tables 1 and 2 in the Supplementary information give shape and color encodings used in this study.

User interface for detailed text information
One of the challenges is that nodes and edges in metabolic networks contain more information than those revealed by the visual metaphors mentioned in Tables 1 and 2 of Supplementary information, e.g. the subcellular compartment information for nodes and strength information for edges. Some of this information may not have an easy visual representation or it may clutter the user’s field of view. Furthermore, during the exploration of the network, scientists may want to know exact component names and numerical values associated with them. A floating text panel for this kind of information may block users’ sight of the network structure. MetNet3D resolves this problem by using a portable tablet PC to present textual information. A Java Graphical User Interface (GUI) that can be operated from a Tablet PC is connected to MetNet3D through a wireless network. This GUI, called Tweek, is part of the VR Juggler suite that enables compatibility with traditional keyboard and mouse interactions (Hartling et al., 2002). Figure 1a shows Tweek being used in a Tablet PC in an immersive environment. In MetNet3D, the Tweek interface lists textual annotations for the nodes and edges in the currently loaded metabolic network, such as EC numbers, gene family information and pathway names. One of the main advantages of using VR Juggler and Tweek in MetNet3D is that the application can be transparently moved from the immersive system to the desktop.

Visual metaphors for gene expression data
MetNet3D represents microarray expression data as a 3D plot, as shown in Figure 3. Each sphere represents an expression level of a given gene as reflected by the accumulation of its corresponding RNA. Values are indicated by color and height. The cylinder connecting two spheres shows the change in expression values. The series of spheres connected by cylinders form a connected line in 3D, showing the expression levels of a gene under different conditions. All genes are clustered using clustering algorithms from the R ‘hclust’ package (Ihaka and Gentleman, 1996) so that genes with similar expression patterns are close to each other in the plot.

Global layout of the entire pathway network
Navigating through the entire network gives the user a good understanding of the data structure and serves as the starting point of the exploration. The layout method for the entire network in MetNet3D is a modified 3D Graph-Embedder

User interface for detailed text information
One of the challenges is that nodes and edges in metabolic networks contain more information than those revealed by the visual metaphors mentioned in Tables 1 and 2 of Supplementary information, e.g. the subcellular compartment information for nodes and strength information for edges. Some of this information may not have an easy visual representation or it may clutter the user’s field of view. Furthermore, during the exploration of the network, scientists may want to know exact component names and numerical values associated with them. A floating text panel for this kind of information may block users’ sight of the network structure. MetNet3D resolves this problem by using a portable tablet PC to present textual information. A Java Graphical User Interface (GUI) that can be operated from a Tablet PC is connected to MetNet3D through a wireless network. This GUI, called Tweek, is part of the VR Juggler suite that enables compatibility with traditional keyboard and mouse interactions (Hartling et al., 2002). Figure 1a shows Tweek being used in a Tablet PC in an immersive environment. In MetNet3D, the Tweek interface lists textual annotations for the nodes and edges in the currently loaded metabolic network, such as EC numbers, gene family information and pathway names. One of the main advantages of using VR Juggler and Tweek in MetNet3D is that the application can be transparently moved from the immersive system to the desktop.

Visual metaphors for gene expression data
MetNet3D represents microarray expression data as a 3D plot, as shown in Figure 3. Each sphere represents an expression level of a given gene as reflected by the accumulation of its corresponding RNA. Values are indicated by color and height. The cylinder connecting two spheres shows the change in expression values. The series of spheres connected by cylinders form a connected line in 3D, showing the expression levels of a gene under different conditions. All genes are clustered using clustering algorithms from the R ‘hclust’ package (Ihaka and Gentleman, 1996) so that genes with similar expression patterns are close to each other in the plot.

Fan layout of reactions of interest
The global layout reveals the overall structure of the metabolic network to users. Users may also be interested in a more detailed view of a specific metabolite. MetNet3D can bring them Reactions of interest (ROI) focusing on a user-selected metabolite. A ROI is defined as the set of all metabolic reactions that a metabolite takes part in. The metabolite is called a focus node. Users can interactively choose focus nodes by using the joystick or through the Tweek GUI.

A layout algorithm, called a fan layout, is proposed to draw ROIs. In the fan layout, the reactions are around the focus node (Fig. 4), like leaves of a fan. Each reaction is drawn following the convention that all substrates are on one side of the reaction node, all products on the other side and all enzymes are above the reaction node. One uniform property of metabolic reactions, i.e. the hierarchical structure among enzymes (genes, RNA, polypeptides and protein complex) forming a quasi-tree rooted at the reaction node, helps to

Visual metaphors for gene expression data
MetNet3D represents microarray expression data as a 3D plot, as shown in Figure 3. Each sphere represents an expression level of a given gene as reflected by the accumulation of its corresponding RNA. Values are indicated by color and height. The cylinder connecting two spheres shows the change in expression values. The series of spheres connected by cylinders form a connected line in 3D, showing the expression levels of a gene under different conditions. All genes are clustered using clustering algorithms from the R ‘hclust’ package (Ihaka and Gentleman, 1996) so that genes with similar expression patterns are close to each other in the plot.

Fan layout of reactions of interest
The global layout reveals the overall structure of the metabolic network to users. Users may also be interested in a more detailed view of a specific metabolite. MetNet3D can bring them Reactions of interest (ROI) focusing on a user-selected metabolite. A ROI is defined as the set of all metabolic reactions that a metabolite takes part in. The metabolite is called a focus node. Users can interactively choose focus nodes by using the joystick or through the Tweek GUI.

A layout algorithm, called a fan layout, is proposed to draw ROIs. In the fan layout, the reactions are around the focus node (Fig. 4), like leaves of a fan. Each reaction is drawn following the convention that all substrates are on one side of the reaction node, all products on the other side and all enzymes are above the reaction node. One uniform property of metabolic reactions, i.e. the hierarchical structure among enzymes (genes, RNA, polypeptides and protein complex) forming a quasi-tree rooted at the reaction node, helps to
A key role of genes in metabolic networks is to encode proteins that catalyze reactions. In the graph model, the edges go out from the genes to the reactions that their encoded proteins catalyze. This property makes the radial layout (Eades, 1992) applicable. The radial layout algorithm arranges all the children nodes on a circle around the parent node. In MetNet3D, the radial layout algorithm arranges the reaction nodes on a circle around the gene (Fig. 5). It shows which reactions a gene product catalyzes. The radius of the circle is the minimum distance that guarantees that there is no overlap between any two reactions. The detailed algorithm is given in the Supplementary information.

Drawing radial layouts centered on different genes side-by-side helps to compare these reactions and visualize any differential regulation of the catalysis. Rotating the radial layout along the center of the circle, such that at each time one reaction is horizontal, gives the viewer a better view of each reaction.

Radial layout of reactions of interest

While navigating through the profiles, the user finds two genes, ‘AT5G35170’ and ‘AT3G60510’, with similar expression levels and from the entire metabolic network. Both metabolic networks and gene expression data contain the same gene names. However, not all genes in the expression data link to the networks since even in model organisms such as Arabidopsis the function of most genes is still unknown.

Choosing a metabolite from the metabolic network. The scene switches to a fan layout of the ROI focusing on the node. Choosing a node other than the current focus node generates another ROI. A series of ROIs let the user navigate through a pathway.

Choosing a gene from the network view and the gene expression data are given below.

- Choosing a cluster of genes to display their expression data. Due to the huge number of genes in large-scale expression studies (e.g. 22 000 for Affymetrix GeneChips), it is impractical to display the expression values for all genes simultaneously. MetNet3D displays the expression profiles of gene clusters (the left part in Fig. 3) and the expression profiles of genes of the current selected cluster (the right part in Fig. 3).

- Choosing a gene from gene expression profiles. The user can either look at a radial layout of the ROI focusing on the selected gene or watch the animation created by pulling out the reactions that the gene participates in from the entire metabolic network. Both metabolic networks and gene expression data contain the same gene names. However, not all genes in the expression data link to the networks since even in model organisms such as Arabidopsis the function of most genes is still unknown.

- Choosing a metabolite from the metabolic network. The scene switches to a fan layout of the ROI focusing on the node. Choosing a node other than the current focus node generates another ROI. A series of ROIs let the user navigate through a pathway.

- Choosing a gene from the network view and the gene expression data are given below.

RESULTS AND DISCUSSION

The utility of the MetNet3D can be demonstrated by examples from the gene expression data, from an experiment in carbohydrate metabolism that used Affymetrix ATH1 GeneChip™ data from the Wurtele Lab (Li et al., 2004) and metabolic networks in Arabidopsis from the MetNet Database (Wurtele et al., 2003) and the AraCyc Database (Mueller et al., 2003).

Initially, MetNet3D displays the global layout of a metabolic network together with the expression profiles of genes and gene clusters. The network includes pieces from 200 different pathways such as chorismate biosynthesis, fatty acid elongation (saturated), glucose 1-phosphate metabolism, purine biosynthesis, isoleucine biosynthesis I and nucleotide metabolism. While navigating through the network, the user becomes interested in the metabolite ‘GMP’. Figure 4 shows a fan layout of reactions of interest focusing on ‘GMP’, in the nucleotide metabolism and purine biosynthesis pathways.

While navigating through the profiles, the user finds two genes, ‘AT5G35170’ and ‘AT3G60510’, with similar expression levels and
wants to know their functions in the metabolic network. Side-by-side, the radial layout compares the reactions that their encoded proteins have catalyzed (Fig. 5).

The user can also choose a gene from profiles and look at the animation by pulling out the relative reactions. The movie of an animation (http://www.vrac.iastate.edu/research/sites/metnet/paper.htm) shows how ‘AT2G22190’, the putative gene for trehalose-6-phosphate phosphatase, fits into the known pathways in MetNetDB. The animated portion pulled out of the main graph consists of all the reactions from the trehalose biosynthesis and trehalose degradation pathways that the gene AT2G22190 takes part in.

Integrating metabolic pathway and gene expression data in three-dimensions with stereoscopic VR enhances visualization, navigation and interaction capabilities. The ability to view multiple types of data at once leads to a better understanding of both the pathway and the gene expression data. Pulling out reactions of interest from a complex pathway network allows the users to see details of the reactions. Furthermore, the ability to use natural human responses to explore 3D environments significantly helps in exploring the complex metabolic networks. In a typical desktop setting, if the view presented to the user is confusing, the user must know the specific interactions to change the view point to be able to visually disambiguate the complexity due to overlapping portions of the graph. In the virtual environment this is done naturally, by simply walking around the graph or moving the head side to side as we do in the real world when something is obstructing our view.

Interaction makes the integration more effective as users do not need to search for a specific node in the pathway network when its expression levels in 3D plot interest them. Visualizing pathways and gene expression plots in immersive stereoscopic VR gives the users a more realistic feeling and more natural interaction. By displaying pathways and expression data around the user, the user can manipulate them in the same way as he manipulates real objects.

Stereoscopic VR have the potential to help people to understand the interrelationships within the complex metabolic networks and with the gene expression profiling data. Visualization and navigation in 3D space make it possible and easy to explore a much larger dataset than in 2D space. Consistent and rich visual metaphors in 3D space simplify the process of knowledge perceiving. Interaction and navigation give users more control of the learning procedure. Current users, some biologists and biology students, use MetNet3D to analyze new data from their experiments, after some training. A formal user study is being designed.

The integration of metabolic pathways and gene expression data in VR environments is a promising new method to explore metabolism. Immersive stereoscopic VR is a relatively new human computer interface compared with the traditional computer screen, mouse and keyboard. As such, application of immersive VR to bioinformatics raises new research opportunities, but also presents challenges and extends existing areas, such as pathway layout problem. New navigation and data display methods need to be investigated to fully realize the potential of VR for understanding complex data.

ACKNOWLEDGEMENTS

The authors are grateful to Jie Li and Pan Du for developing and providing the XML data files, and to Carol Foster and Ling Li for providing the microarray data and for helpful discussions. This work was performed at the Virtual Reality Application Center, Iowa State University. Funding for this project, and the Open Access Publication charges for this article, was provided by grants from the National Science Foundation in the Arabidopsis 2010 and Information Technology Research Program (Grant Number: 0219366) and by a research initiation grant from the Curver Foundation.

Conflict of Interest: none declared.

REFERENCES


Friedrich, C. and Hoole, M.E. (2001) Graph drawing in motion II. In 9th International Symposium on Graph Drawing, LNCS 2265, Vienna, Austria, pp. 220–231.


