A Computerized Metabolic Map. 2. Relational Structure, Extended Modeling, and a Graphical Interface

Raymond S. Ochs,* Amer Qureschi, Adam Sycz, and James Vorbach

Departments of Pharmaceutical Sciences and Math and Computer Sciences, St. John's University, 8000 Utopia Parkway, Jamaica, Queens, New York 11439

Received September 12, 1995[⊗]

We have developed a computer representation of biochemical metabolism which allows stepwise selection of possible metabolic routes, retrieval of relevant associated information, and data entry to expand the underlying system. The system is based upon a data model that forges links between chemical substances, reactions, enzymes, and associated information. At the most fundamental level, sequential links can be viewed between metabolic reactions that occur in cells. In addition, variation in species as well as subcellular locale for reactions are represented. Our system allows a particularly relaxed definition of both *reaction* and *enzyme*. Thus, reactions include binding proteins (such as hormone receptors) that are neither catalytic nor involve transport steps. Enzymes include transport steps (which are formally translocation catalysts), and the model extends to representing enzymes as substrates of other enzymes. Thus, steps in hormone signaling such as hormone-receptor binding and noncatalytic binding of intracellular signal molecules can be incorporated into the same structure and output to the same sequence as steps involving enzymatic transformation or transport steps.

INTRODUCTION

The chemical reactions which occur within cells are recorded in metabolic charts, which schematize individual reactions and overall connected sequences (pathways). Such representations are skeletons of richer information sets which include chemical information on the compounds, the enzymes, localization, and the species which contain the reactions. The data itself currently exists in disparate articles, reviews, and textbook summaries, from which it is difficult to rapidly obtain an authoritative view of a particular pathway segment.

A fully developed computerized metabolic map will allow a user to view a metabolic segment from any point in the known universe of biochemical reactions, construct a pathway, and perform a wide array of analyses on that segment. This will supersede the use of paper representations and allow access to a wide array of related information, both electronic and paper-based.

Early approaches to applying computers to the understanding of enzyme linked reactions centered upon small, fixed datasets. Usually, the analysis centered upon solutions of simultaneous equations to determine rates of material flow given kinetic constants and concentrations of the compounds.[1] A variant of this approach[2] focuses on the cofactors ATP and ADP and their distribution in Mg^{2+} complexes at various values of pH. This implicitly recognizes the parallel connections of cofactors to metabolic pathways; still, this concept and other key notions such as noncatalytic binding proteins have not been incorporated into any existing computerized pathways.

Our first approach to developing a view of metabolic information was to construct a system for viewing linked lists of reactions, with representation of specific features of metabolic control.[3] These included the concept of near-equilibrium versus nonequilibrium reactions, the representation of separate species by applying a selection mask, and the separation of reaction components between pathway substrates on the one hand and cofactors (cosubstrates and coproducts) on the other. For the later, we specifically enforced the distinction between globally available coreactants (the cofactors, which connect reactions in parallel) and tightly bound prosthetic groups, which remain bound to enzymes and serve as part of the catalytic mechanism, but do not contribute to pathway flow. We also incorporated the concept of transporters, which can be viewed as enzymes that catalyze a cell space translocation, but not a chemical transformation. Subsequently, we reported our efforts to construct a GUI for viewing the metabolic map[4] and our first work toward systematizing the underlying representation itself, employing the notion of entity-relationship diagrams and an underlying relational database structure.[5] We report here our latest work which incorporates and extends these concepts.

OVERVIEW OF THE SYSTEM

The characteristics of our current metabolic map are twofold: a viewer for the metabolic information that separates the user from database concerns and a utility to maintain and update the database. The system is based upon a conceptual model, which allows the semantics of the information to be represented independently of the actual implementation. To implement our system, we used a Microsoft ACCESS, a relational database management program, and Microsoft VISUAL BASIC, an event-driven language to provide the user interface.

Two views into the system are available to the user: the Pathway View and the Data Entry View (Figure 1). The pathway view allows the creation, display, and storage of metabolic submaps of connected reactions, in addition to stored ancillary information. The data entry view allows additions and corrections to the database. This separation helps prevent data alteration when the intent is to explore

[⊗] Abstract published in *Advance ACS Abstracts*, February 15, 1996.

0095-2338/96/1636-0594$12.00/0

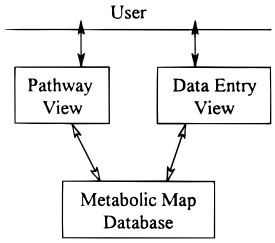

Figure 1. System overview. Schematic of interaction of the user with the system.

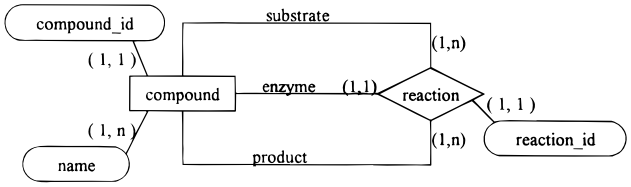

Figure 2. Basic ER model. The essential entity-relationship is displayed. The rectangle indicates entities, the diamond, relationships. The numbers in parentheses indicate the minimum and maximum number of participants in each relationship.

pathways and allows optimization of each of these tasks suited to their distinct needs. To assist the development of these views and ensure an accurate model of the underlying sense of the system, we used two conceptual modeling approaches: an extended entity-relationship (ER) information model[6] and an event-state model.[7] Information models define the information (entities and relationships) to be represented and define integrity constraints such as unique keys and the cardinality of relationships. Event-state models are useful in defining control and are used here to model the user's interaction with system.

THE INFORMATION MODEL

The extended ER model[6] allows a facile conceptualization of an application's semantics and a straightforward mapping to an underlying database. Entities (rectangle symbols) correspond to the meaningful units in an application, e.g., compounds; relationships (diamond symbols) identify connections between entities, e.g., reactions; attributes (oval symbols) represent descriptive properties (e.g., name). Figure 2 shows a simplified ER model of our system. Two central concepts of the database are represented in this figure: *compound* and *reaction*. The entity *compound* is related to itself through the relationship *reaction*.

Compounds, in the role of substrate, are transformed into compounds in the role of product, in reactions catalyzed by compounds in the role of enzyme. This reflexive relationship, i.e., compounds related to compounds, allows a wide range of metabolic processes to be easily supported, such as enzymes serving in the role of substrates of other enzymes ("enzyme-as-substrate"). This basic model identifies the semantic unit for the construction of pathway segments (submaps): one or more substrates are converted into one or more products, catalyzed by an enzyme.

Constraints such as "one or more..." are expressed by annotating the edges. The convention adopted (min, max) denotes the minimum and maximum number of participants in the relationship (equivalently, the minimum and maximum number of property values for each entity). This nomenclature allows us to specify key and cardinality constraints

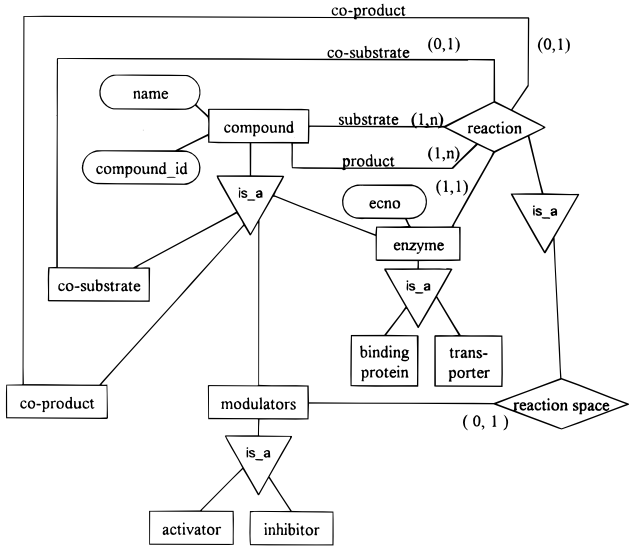

Figure 3. System ER model. This is a more complete entity-relationship diagram of the system. Ovals represent attributes, and the triangular IS_A symbols indicate a subclassification (hierarchy). The rectangle, diamond, and parenthetic number symbols are described in the legend to Figure 2.

for relationships. For example, each reaction must have a least one substrate, product, and enzyme compound. Each reaction can have n ($n > 0$) substrates and products but at most one enzyme. As an example of property values, each compound has just one compound_id, which uniquely identifies the compound.

The detailed model (Figure 3) provides a more complete description of the stored information. The extended ER model represents classification hierarchies using the IS_A symbol. Enzymes, coproducts, cosubstrates, and modulators are modeled as subclassifications of compounds. As subclassifications they inherit the properties of the "parent" compound but have in addition their own specific properties. For example *enzyme* inherits the name and compound id properties from compound but has in addition the enzyme commission number (ecno) property. *Binding protein* and *transporter* inherit name and compound_id from compound and the ecno property from *enzyme*. This scheme provides clarity in definition of relationships and allows efficient mapping to tables in a relational database. Note, for readability, most descriptive properties are omitted from the figure; a complete listing is provided in Table 1.

Enzymes acting in the role of substrates ("enzyme-as-substrate") are represented in our model and explicitly supported in our system. This important characteristic allows us to represent a much wider class of reactions within cells, including regulatory networks not previously possible. As an example, protein kinase A catalyzes the phosphorylation of a number of proteins, including the enzyme pyruvate kinase as illustrated in Figure 4. The modified protein in this case has a covalently attached phosphoryl group, indicated by the notation [pyruvate kinase]_P. The unmodified form of pyruvate kinase is more enzymatically active. This enzyme catalyzes the conversion of phosphoenolpyruvate to pyruvate. It is important to consider that the pyruvate kinase under discussion is a specific isozyme form; not all pyruvate kinase enzymes can be phosphorylated. This is represented in our database as a species-specific property.

Table 1. Properties of the Database System

Sub_Prod_Reaction	**Reaction_Space**
Substrate	Reaction_id
Product	Organism
Reaction_id	Organ
	Cell
Alternate_Compound_Name	Cell_Space
Compound_id	After_Cell_Space
Compound_name	Reversibility
Sort_name	Eq_Constant
	Eq_Constant_Lit_Ref
Comp_Coproduct	Constraint
Compound_id	Constraint_Lit_Ref
Reaction_id	Note
	Note_Lit_Ref
Comp_Cosubstrate	
Compound_id	**Enzyme**
Reaction_id	Enzyme_id
	Enzyme_Com_No
Enzyme_Reaction	
Enzyme_id	**Modulators**
Reaction_id	Compound_id
	Type
	Constant
	Comments
	Reference

Combining the binding protein notion with enzyme-as-substrate expands the possible number of processes that can be modeled even further. Figure 5 is a summary of recent research[8] on the molecular action of insulin signaling in cells. The insulin receptor has a dual role. On the one hand, it is a binding protein, specific for its ligand, insulin. This binding event is localized to the extracellular space (a species-type property). The bound complex Insulin:Insulin-Receptor has catalytic activity, representing the second role for the insulin receptor. As indicated in Figure 5, the substrate for the reaction catalyzed by this enzyme is the protein IRS-1; the product, the phosphorylated protein, IRS-1_P. The enzymatic event is localized to the cytosol, indicating the importance of representing cell space in the model. Further examples of the combination of binding protein and enzyme-as-substrate events are displayed for the subsequent steps in the insulin pathway of Figure 5. It is evident that the combination of the two notions of binding protein and enzyme-as-substrate allows a formal representation of an intricate sequence of events.

As a point of notation, the Enzyme Commission divides enzymes into six major categories, dependent upon reaction type,[9] which accounts for the first digit (main category) of ecno. We have extended this nomenclature to provide an ecno to transporters (main category 7) and binding proteins (main category 8).

Cofactors may occur in pairs, as they remove or add a piece of the pathway substrate, leading to a pathway product and a modified cofactor. The cofactor reacting with the substrate is termed the cosubstrate; the cofactor formed is called the coproduct. Net flow through the reaction under consideration requires a continuous supply of cosubstrate. This means that separate (parallel) reactions are required to regenerate the cosubstrate. Note that from the standpoint of these other reactions, the roles of cosubstrate and coproducts are reversed. This metabolic interplay underscores the need for strict definition of cofactors and their reflexive quality. Examples of cofactor pairs in our database are ATP/ADP and NAD/NADH. A cofactor that does not appear as a pair is inorganic phosphate (Pi). The cardinality constraint (0,*n*) in Figure 3 on each cofactor link with reaction means a reaction may have either zero or many cofactors.

As is evident in Figure 3, modulators are compounds linked to reaction space. The reaction space relationship

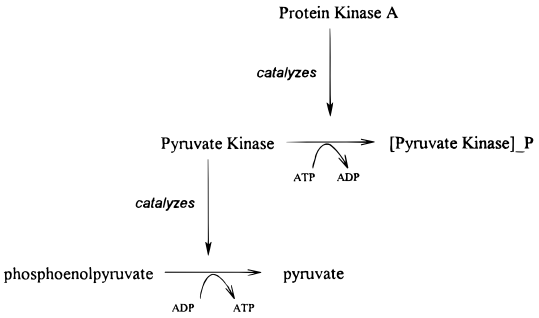

Figure 4. Enzyme-as-substrate. This concept is illustrated here with the example of pyruvate kinase. This enzyme is both a catalyst, in the conversion of phosphoenolpyruvate to pyruvate as well as a substrate of the enzyme protein kinase A. In the latter role, pyruvate kinase is converted to the phosphorylated form, [pyruvate kinase]_P, which is inactive.

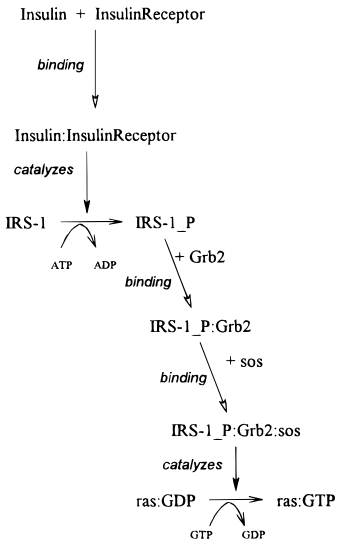

Figure 5. Binding protein and enzyme-as-substrate example. The insulin signaling system is shown as an example of the combination of the notion of binding proteins with that of enzyme-as-substrate. Bound complexes are indicated as individual compounds separated by colons.

links reaction to information at a lower, more detailed level. Modulators include compounds that alter enzymes in particular species, experimental data that vary between laboratories, and modulation that occurs as a result of distinct developmental states of an organism.

INFORMATION MODEL IMPLEMENTATION

In order to provide a higher degree of data independence and to minimize redundant data (and the concomitant data integrity enforcement problems), we choose an object-oriented design approach implemented in a relational database management system. Our key concepts, compounds and reactions, are uniquely identified by persistent object identifiers, *compound_id* and *reaction_id*, respectively. This makes reaction identification independent of ecno (enzyme commission numbers) which may be subject to change in the future. Separation of ecno from reaction has been previously employed by others, e.g., ref 10. The information model shown in Figure 3 was implemented as a relational database using Microsoft ACCESS. This software allows relatively simple interaction with the user, supports relational query language, SQL, and provides data integrity enforcement. The mapping from the ER formalism

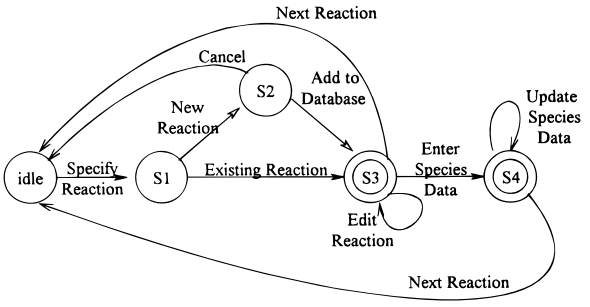

Figure 6. Data entry process. This event-state diagram portrays the sequence of user interaction with modification of the database. Circles represent events; double circles represent termination points.

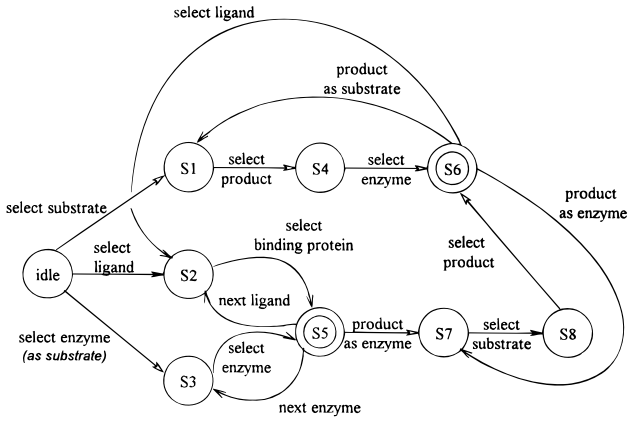

Figure 7. Navigation process. This event-state diagram portrays the user construction of pathway segments. Symbols are those described in Figure 5.

to database tables within ACCESS is straightforward. There is a roughly 1:1 mapping from entity and relationship symbols to relations. Property symbols in the ER diagram map to columns in the tables.

THE EVENT-STATE MODEL

The event-state model is a schematic indicating how the user interacts with the system. Figure 6 shows the data entry process for database modification, and Figure 7 shows the process of constructing metabolic pathways from the database.

An event-state model consists of nodes (states) connected by arrows (events). In our application, states identify control points in the process where the user is presented a choice to determine the next step (state). The idle state represents the system entry point where the choice between views is made. If data entry is selected, the process in Figure 6 applies. From the idle state, the user specifies a reaction (metabolic step). This is done by selecting substrate, product, and enzyme in any sequence. The system then determines if the reaction is already in the database. Either state S2 or S3 is entered depending on the result of this check. In S2 the user decides whether to add this reaction to the database or return to the idle state. In S3 the user can modify the selected reaction, enter additional experimental data, or exit the system. The double circles at S3 and S4 identify these states as terminal points. However for ease of use, the user is given the option to reset the system at any point in data entry process (omitted from Figures 6 and 7).

The pathway view (Figure 7) provides a choice between three selections: (a) by substrate, (b) by ligand, and (c) by enzyme. (a) The common pathway selection is the route from idle to S1, S4, and S6. This builds a metabolic sequence by sequential database queries and user-imposed restrictions at each step. Selecting S2 or S3 instead determines the reaction product in a different way. These choices break with the traditional substrate−product−substrate... submap and allow a choice of reaction by ligand or enzyme. (b) Choosing ligand (S2) restricts the next choice to a list of binding proteins. Examples include albumin, which binds fatty acids, and hormone receptors, which bind hormones. The latter example leads to the interesting possibility of considering the product as an enzyme, thus progressing from S5 to S7 and S8. From S8, it is possible to enter the upper limb of Figure 7 (S1 → S4 → S6) allowing re-entry to traditional metabolic routes. (c) S3 indicates direct selection of an enzyme which is a substrate of another enzyme (with progression to S5). From this point, progression through S5, S7, and S8, as before, allows construction of enzyme cascades. Alternatively, moving to S2 allows incorporation of binding proteins at this step. Overall, the submap produced can consist of virtually any combination of the three types of connections between compounds in the database.

IMPLEMENTATION OF DATA ENTRY

The opening screen of the data entry module is presented in Figure 8. Each of the large buttons (Enzymes, Substrate, Product) can be independently selected first to enable a search for existing entries in the database. Once it is determined that editing is required, the button EDIT enables this mode. At any point, a new compound entry can be entered into the database by simply typing the new the entry under substrate or product. Choosing one of the large buttons provides an alphabetical listing of compounds.

A significant problem of long standing is the occurrence of prefixes and suffixes in chemical compounds. These are important in correctly identifying the compound to the user but hinder alphabetical compound selection. A related difficulty is the variety of alternative names for the same compound. To resolve the first problem, compounds are now displayed in alphabetical order based only upon commonly recognized portions of the compound name. Symbols denoting stereochemistry or atomic positioning of groups are not recognized in the ordering. Nonetheless, the display provides the complete name. For example L-(−)-glucose and D-(+)-glucose would be adjacent listings. For the second problem, we now represent all compounds by a unique internal reference key (the object identifier, *compound_id*), which may be assigned to any number of chemical synonyms.

Several features are present in the editor that facilitate the modification of existing information. For example, all reactions are considered as reversible, so that selection of substrate or product produces a list that includes either one, whatever its appearance in the underlying database. New enzymes can be entered, and the ecno (enzyme commission number) field checked for consistency in format. Whenever an entry is made that does not exist in the database, a window is presented to inform the user of this situation, enabling a choice of starting over or adding the new entry. The ENTER button is selected when all modifications of the displayed reaction are complete and terminates the editing mode.

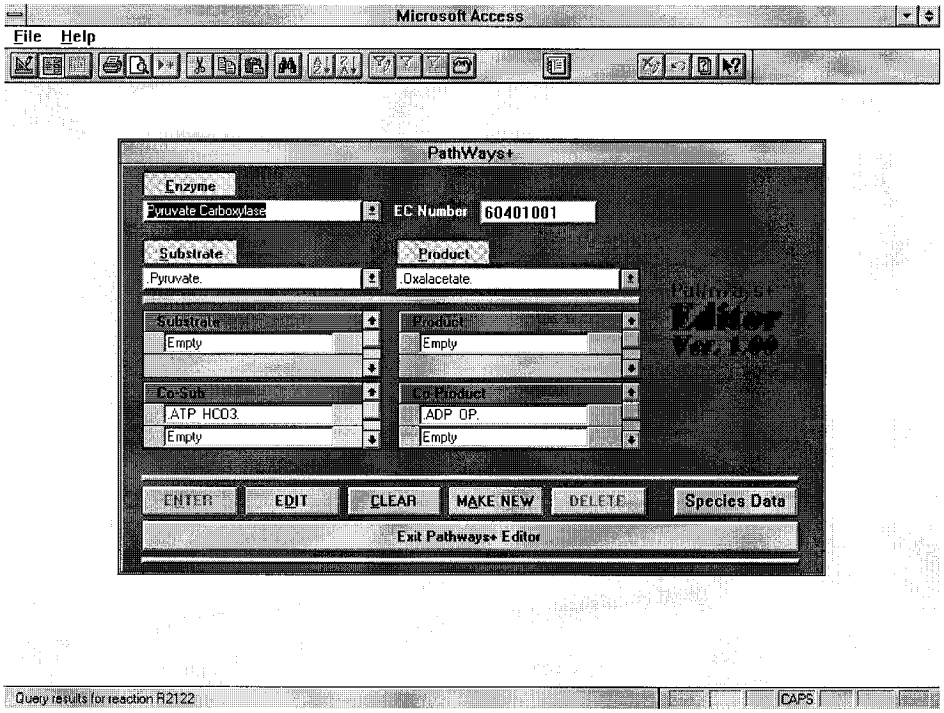

Figure 8. Pathways editor, main screen. Screen display shows the opening and main screen of the editor, with functions apparent on buttons for selection of substrate, product, or enzyme and their modification.

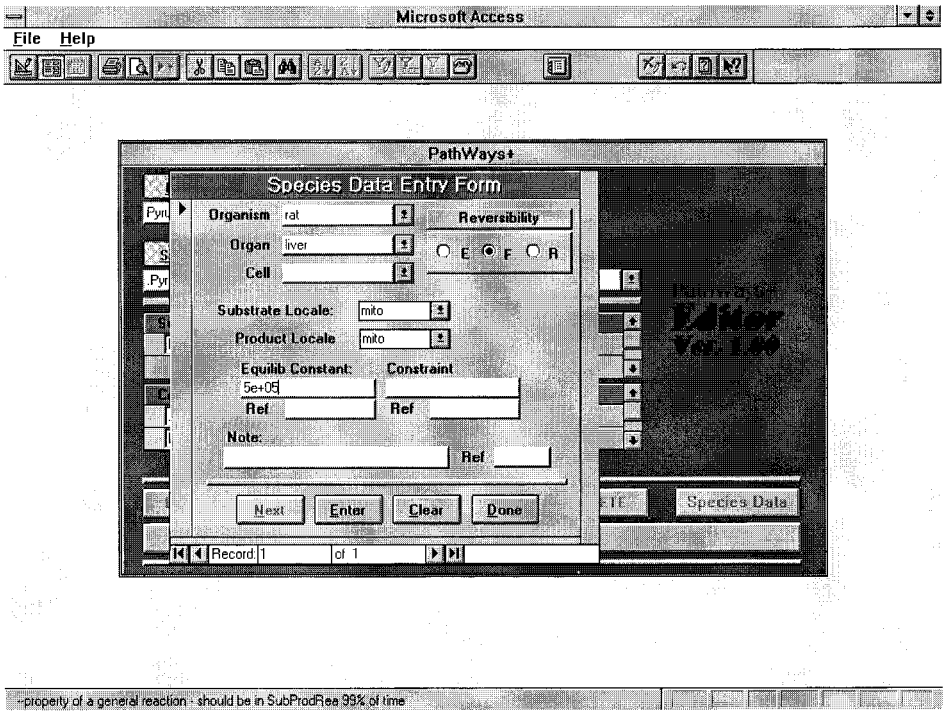

Figure 9. Pathways editor, species screen. Display shows subform of editor in which species specific information can be entered.

A second window, hidden from initial view, is opened by the Species Data button. Figure 9 shows this new entry form. The connection between the two forms coincides with the reaction−reaction space connection in the information model. The species form allows entry of information that is species-specific as well as that which formally is not, such as values for equilibrium constants. Equilibrium data, however, are experimental values which do vary between laboratories reporting them. Thus, different values for an equilibrium constant can be collected for the same reaction. The species fields are linked to reaction_id field directly and include information that specifies the portion of the cell in which the reaction substrates and products exist separately, information on whether the reaction is metabolically reversible or not, and a constraint field, useful for circumscribing the reaction under other situations such as developmental state of the organism. Any number of species entries can be called up for a given reaction and each are stored related to that reaction entry.

IMPLEMENTATION OF PATHWAY VIEWER

The pathway view implemented under Visual Basic accesses the database maintained by the data entry system.

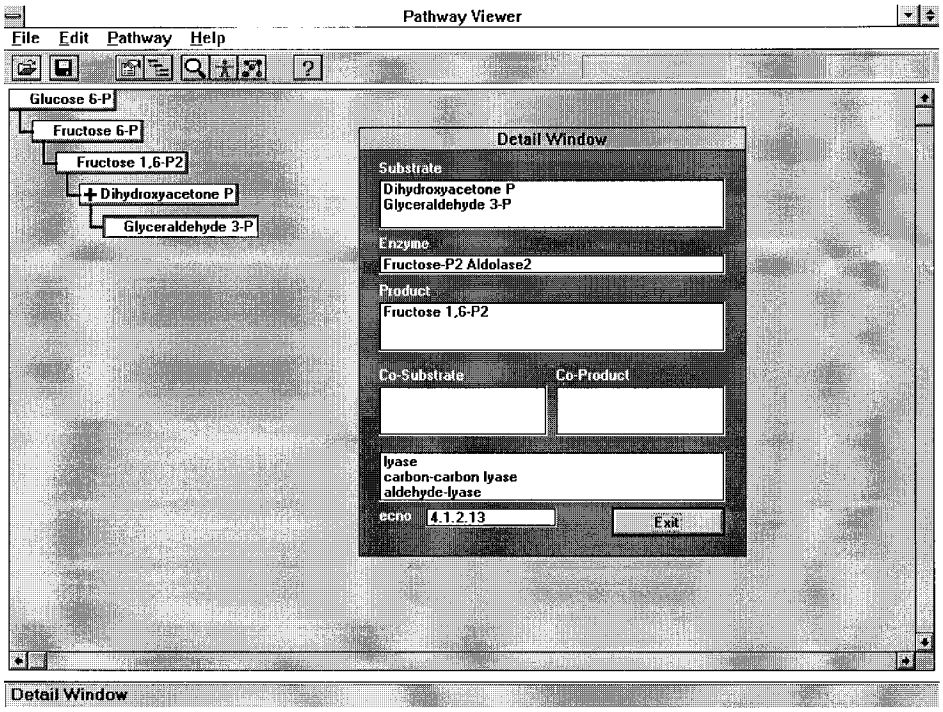

Figure 10. Pathway viewer display of a simple submap. A metabolic sequence that was selected sequentially is shown, along with a detail window for the reaction leading to glyceraldehyde 3-phosphate.

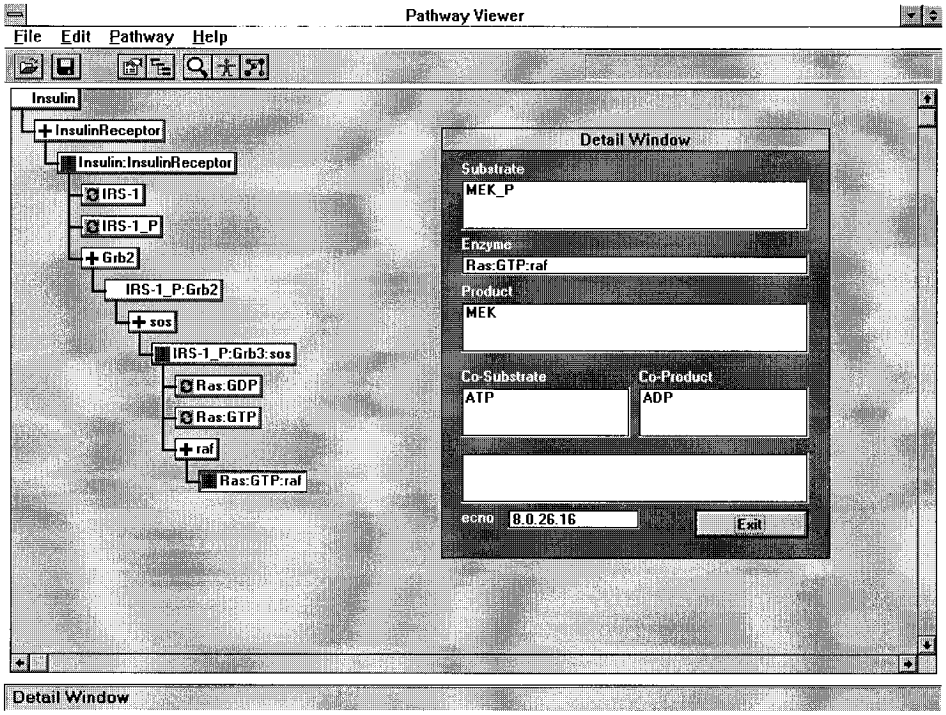

Figure 11. Pathway viewer display of a complex submap. This more complicated metabolic sequence is an implementation of the binding protein and enzyme-as-substrate concepts. The information displayed includes the reactions sketched in Figure 5.

Selection of substrates is made from a list sorted as described in the previous section, that is, by excluding chemical prefix and infix notations, but displayed in native form. Once a substrate is selected, a choice of the next step is presented in a second list box. Finally, an enzyme is selected, and a node is populated on the screen. Branches can be introduced by selecting the add pathway button, and trips can be saved to disk and used as the basis of later pathways. An example is shown in Figure 10. The "+" sign indicates more than one pathway substrate or product is present. Note that the detail box (in this example, aldolase) includes the ecno; just above this is a box containing a translation of the ecno which provides a formal naming of the enzyme.

A very similar display can accommodate more complex pathways, including binding proteins and enzyme-as-substrate. Implementation of these features is indicated in Figure 11, in which a separate icon is displayed to indicate that this selection mode has been chosen. The map displayed includes all of the steps schematized in Figure 5. Enzymes are selected from lists in a very similar way to substrates (not shown); they are made distinct to the user by the use of different colors.

Access to information about a single reaction is available by choosing a metabolite and then selecting the detail window button. This displays the information of the underlying database, including not only the complete reaction record but also the species information that exists. Chemical structures can be displayed by selecting the structure button. The chemical structures themselves were entered into a separate program (ChemBase, from Molecular Designs, Ltd.) which facilitates graphical data entry and is linked to the data entry system via the compound_id field.

The flexible tree metaphor of the display takes advantage of the extensibility of Visual Basic. Many software tools called Visual Basic Extensions (VBXs) are available which supplement its functionality. In our case, the TreeBuilder VBX (from Visualsoft) greatly reduced the complexity of implementing the pathway graphics. The presentation of submaps using our system is simple yet shows at a glance the hierarchical relationships.

DISCUSSION

We have constructed a system for representing connections between chemical compounds which exist within cells. Combining direct pathway links with the notion of enzymes catalyzing these links, properties of the enzymes, modifiers, species variations, reaction information, and new variations on the notion of enzyme itself is readily handled by a strongly structured database environment. The system allows complex models to be readily constructed with a simple interface. Currently, only those investigators immersed in a specialty area are well versed in particular metabolic segments and the intricacies of their related facts. With the advent of a system which enables storage and recall of key facts related to chemical compounds and their cellular context, understanding of metabolism can become far more widely appreciated.

Other investigators have produced software for metabolic modeling, but typically the underlying motivation is not directed to problems of metabolism, its interconnections, and regulation. These approaches can be grouped into four types: chemical database representation, correlation to nucleic acid sequence data, quantitative modeling, and bioengineering.

An example of specifically targeted chemical database is that of Caffrey et al.[11] and Caffrey and Hogan.[12] This is a collection of data on lipid phase transition temperatures. While clearly narrowly focused, the approach taken for data collection—requesting data on short paper forms from various labs, followed by central collation, data entry, and checking—may well be important for others (including ourselves) in future database population efforts.

Closer to our own approach is the work of Rouxel *et al.*[13] This system allows searches of metabolic information, but pathway construction is limited. Key features of metabolic structure, such as classification of cofactors as a separate parallel reactions, and the ability to represent multiple pathway intersections are absent. The guiding principle stated by the authors is that a computer database should reflect paper representations as given by wall charts; indeed, a major information source of this database is the Boehringer wall chart itself. As argued by Morrell,[14] however, a computer representation should be quite different. Instead of duplicating paper representations, computer versions should take advantage of the unique computer abilities, such as large scale sorting and associations. An earlier model[15] takes an approach similar to Rouxel *et al.*[13] Here, too, the key strength is that of a look up engine for ascertaining particular assorted chemical facts rather than a systems view of metabolic events. A chemical database available for viewing on the Internet[16] is essentially a set of reactions searchable by substrate and arranged in fixed pathways. This contrasts with our approach in which pathway segments are created by the user. In our approach, pathway segments can be stored in data files but are not a data attribute. Other database collections[17,18] also assign pathways as part of the data fields, presumably because it is a way of placing the enzyme in context. This reflects the intent of these efforts as a lookup engine to determine the identity of an enzyme. In one case[17] the data is exclusively from E. Coli; in the other, from C. Elegans. Links to nucleic acid sequences demonstrate a key interest of such collections, namely, an extension of the ability to identify a protein by its nucleic acid sequence and perhaps the function of such a protein. One of the problems in chemical representation, the various prefixes and infixes in chemical terminology, is evident in several of these databases as currently represented. Thus, our approach of separating out these characters from the search routines provides a small but very convenient advance in the first step of data query, namely, the selection of an appropriate substrate or enzyme.

Quantitative modeling approaches have a long history in chemical and biochemical literature. In the case of pathways, one of current interest is the computation of control strength,[19−22] a sensitivity analysis to determine relative importance of each step in controlling the rate of overall flow. New approaches to computer construction of general quantitative models have appeared, which allow direct construction of simple programs,[23] and even graphical manipulation of icons and their connections for automated construction of equations for numerical analysis.[24] What remains to be done is to join a modeling system for quantitative analysis with a database system. With the well represented data structure of our program, we suggest that future model systems can be easily applied to pathway segments, making the pathway segment itself as a variable in the analysis. Since database systems and quantitative modeling systems can be considered as separate entities, a variety of quantitative approaches can be applied. For example, Liao and Lightfoot[25] proposed an analysis which requires time-dependent measurement of pathway intermediates rather than relying upon rate constants assayed in vitro. It is possible to link this analysis to our database with relatively minor modifications.

The use of metabolic pathways as a tool for organic synthesis is the goal of much of metabolic engineering. This requires a knowledge both of pathway sequences as well as its regulation. As recently reviewed by Cameron and Tong,[26] this area is still in a nascent stage. Nonetheless, there is a clear overlap of interests between the bioengineers and those constructing metabolic pathway. Thus, database representations such as the one presented here will help advance such efforts.

Beyond the work summarized above, a number of data sets exist which are relevant to our efforts in more or less direct ways. Much of this can be described as “educational”, in as much as the aim is to provide a teaching tool for beginning students. Thus, there is an electronic textbook in

which glycolysis can be divined,[27] an online biology textbook,[28] and the beginnings of a medical biochemistry text online.[29] These also represent pathways as fixed segments. While they may be more convenient than textbook representations, they provide no deeper insight than the textbook into the pathways themselves.

Overall, the wide range of approaches to representing chemical information of metabolism almost seem to resist a true compilation and compendium. However, this is surely a critical goal. We suggest that an important method to reach that goal is to ensure that the representation of the information be strongly structured so that links between different data sets may be more easily drawn.

Future extensions already planned for our system include addition of other databases (presently, enzyme modulators and foods for which chemical composition is known), expansion of the species type information to include kinetic constants, searching of stored submaps, mapping of atoms within reactions for isotopic analysis, and quantitative modeling based upon the stored submaps. Beyond this, the inclusion of enzymatic mechanisms as a microcosm of metabolic pathways will employ a similar formalism and add extra information density to the metabolic map. Finally, we plan an ongoing effort of database population through a number of approaches, including collaborative efforts with others building similar databases as well as those involved with research involving particular aspects of metabolism. It is possible that an automated system for entry and verification through the Internet will eventually serve to enable continuous upgrading of the database.

SUMMARY

We have designed a computerized metabolic for the exploration of biochemical pathways. The system is composed of two parts: an editor to build the database and a browser to view metabolic segments and related information. Chemical compounds are linked by enzymes, transporters, and binding proteins. Complex events which require recursive manipulation of relationships between compounds are represented. Thus, a ligand may bind a binding protein and the resulting complex become an enzyme. The system was developed with strong underlying database models and can serve as a platform for quantitative modeling in future versions.

ACKNOWLEDGMENT

This work was supported by a grant from the National Science Foundation, BIR 94-96289.

REFERENCES AND NOTES

(1) Garfinkel, D. Computer modeling of metabolic pathways. In *Metabolic Regulation*; Ochs, R. S., Hanson, R. W., Hall, J., Eds.; Elsevier: New York, 1985; pp 20−6.
(2) Alberty, R. A.; Goldberg, R. N. Standard thermodynamic formation properties for the adenosine 5′-triphosphate series. *Biochemistry* **1992**, *31*, 10610−10615.
(3) Ochs, R. S.; Conrow, K. A Computerized Metabolic Map. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 132−137.
(4) Ochs, R. S.; Conrow, K.; Venkatasubramaniam, S.; Grover, S. A Graphical Environment for a Computerized Metabolic Map. In *The Second International Conference on Bioinformatics, Supercomputing, and Complex Genome Analysis*; Lim, H., Fickett, J., Cantor, C., Robbins, R., Eds.; FSU Press: Singapore, 1993; pp 287−96.
(5) Ochs, R. S.; Khadkikar, S.; Ozsoyoslu, G. et al. A relational database model for metabolic information. In *Third International Conference on Bioinformatics*; Lim, H. A., Ed.; World Scientific Publishing: Singapore, 1995.
(6) Korth, H.; Silberschatz, A. *Database System Concepts*, 2nd ed.; McGraw-Hill: New York, 1991.
(7) Peters, L. *Advanced Structured Analysis and Design*; Prentice Hall: Englewood Cliffs, 1987.
(8) White, M. F.; Kahn, C. R. The insulin signaling system. *J. Biol. Chem.* **1994**, *269*, 1−4.
(9) International Union of Biochemistry. Nomenclature Committee. *Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry on the Nomenclature of Enzyme-Catalyzed Reactions*; Academic Press: San Diego, CA, 1992.
(10) Karp, P. D. A knowledge base of the chemical compounds of intermediary metabolism. *Comput. Appl. Biosci.* **1992**, *8*, 347−357.
(11) Caffrey, M.; Moynihan, D.; Hogan, J. A database of lipid phase transition temperatures and enthalpy changes. *Chem. Phys. Lipids* **1991**, *57*, 275−291.
(12) Caffrey, M.; Hogan, J. LIPIDAT: A database of lipid phase transition temperatures and enthalpy changes. DMPC data subset analysis. *Chem. Phys. Lipids* **1992**, *61*, 1−109.
(13) Rouxel, T.; Danchin, A.; Henaut, A. METALGEN.DB: metabolism linked to the genome of Escherichia coli, a graphics-oriented database. *CABIOS* **1993**, *9*, 315−324.
(14) Morrell, R.; Wasilauskas, B.; Winslow, R. Expert systems. *Am. J. Hosp. Pharm.* **1994**, *51*, 2022−2030.
(15) Barcza, S.; Kelly, L. A.; Lenz, C. D. Computerized retrieval of information on biosynthesis and metabolic pathways. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 243−251.
(16) http://www.mcs.anl.gov/home/towell/metabhome.html; 1995.
(17) http://www.ai.sri.com/ecocyc/metabolism.html; 1995.
(18) http://moulon.inra.fr/cgibin/nphacedb3.1/acedb/metabolisme; 1995.
(19) Salter, M.; Knowles, R. G.; Pogson, C. I. Quantitation of the importance of individual steps in the control of aromatic amino acid metabolism. *Biochem. J.* **1986**, *234*, 635−647.
(20) Kacsar, H.; Burns, J. A. The Control of Flux. In *Rate Control of Biological Processes*; Davies, D. D., Ed.; Cambridge University Press: Cambridge, 1973; pp 65−104.
(21) Niederberger, P.; Prasad, R.; Miozzari, G.; Kacser, H. A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast. *Biochem. J.* **1992**, *287*, 473−479.
(22) Groen, A. K.; Wanders, R. J. A.; Westerhoff, H. V.; Van Der Meer, R.; Tager, J. M. Quantification of the contribution of various steps to the control of mitochondrial respiration. *J. Biol. Chem.* **1982**, *257*, 2754−2757.
(23) Keen, R. E.; Spain, J. D. *Computer simulation in biology. A BASIC* introduction; Wiley: New York, 1991.
(24) Hannon, B.; Ruth, M. *Dynamic Modeling*; Springer-Verlag: New York, 1994.
(25) Liao, J. C.; Lightfoot, E. N. Characteristic reaction paths of biochemical reaction systems with time scale separation. *Biotech. Bioeng.* **1993**, *31*, 847−854.
(26) Cameron, D. C.; Tong, I. T. Cellular and metabolic engineering. *Appl. Biochem. Biotech.* **1993**, *38*, 105−140.
(27) http://bmbwww.leeds.ac.uk/designs/glyintro/home.htm; 1995.
(28) http://esgwww.mit.edu:8001/esgbio/7001main.html; 1995.
(29) http://ubu.hahnemann.edu/HemeIron/NetWelco.htm; 1995.

CI9501812