MULTIDIMENSIONAL EXPLORATION INTO BIOCHEMICAL PATHWAYS

JOHANN GASTEIGER*1,2, MARTIN REITZ1 AND OLIVER SACHER2

1Computer-Chemie-Centrum and Institute of Organic Chemistry, University of Erlangen-Nuremberg, Naegelsbachstraße 25, 91052 Erlangen, Germany
2Molecular Networks GmbH, Naegelsbachstraße 25, 91052 Erlangen, Germany
E-Mail: *gasteiger@chemie.uni-erlangen.de

Received: 17th December 2004 / Published: 22nd July 2005

ABSTRACT
The famous Biochemical Pathways wall chart has been converted into a reaction database. The web based retrieval system C@ROL has been interfaced to this BioPath database providing a wide variety of search methods for chemical structures, enzymes, and reactions that can allow one to explore the endogenous metabolism of different species. The database has been made accessible on the internet at: http://www2.chemie.uni-erlangen.de/services/biopath/index.html and http://www.mol-net.de/databases/biopath.html.

It is shown how the information in this database can be used to explore enzyme inhibitors as transition state mimics. Furthermore, it is shown how the classification of biochemical reactions based on physicochemical effects at the reaction site, corresponds with the classification of enzymes by the EC code.

INTRODUCTION
Massive efforts have gone into the deciphering of the human genome. With this goal achieved an important milestone has been reached. However, it was clear from the very beginning that this can only be the start of our understanding of the processes that keep us alive and which might go wrong thus causing diseases.
Thus, attention has shifted to proteomics, the study of the proteins that regulate these processes. Furthermore, increasing research is devoted to metabolomics, the study of the compounds that are produced in living organisms and the way in which they are produced.

Even before the advent of genomics a large body of evidence had been accumulated on the compounds and reactions occurring in our bodies, the biochemical reactions. The essential results of all this work have been compiled quite beautifully in the famous Biochemical Pathways wall chart, produced by G. Michal and coworkers, and initially distributed by Boehringer Mannheim, and now by Roche [1]. These biochemical pathways are outlined in greater detail in an Atlas on Biochemical Pathways [2].

The task now, is to bring these two avenues of investigation, genomics and proteomics on one hand, and biochemical pathways on the other, closer together to link these sources of information. On one side, with genomics and proteomics we have huge amounts of data that have to be processed by bioinformatics methods. And on the side of biochemical pathways we have quite detailed information on small molecules and their interconversions that ask for a more thorough analysis. This is where chemoinformatics can come in, a field with a long history of developing methods for the representation, manipulation and analysis of chemical structures and reactions, which has matured to a discipline providing powerful methods [3,4].

**BIOCHEMICAL PATHWAY DATABASES**

The Biochemical Pathways wall chart contains a cornucopia of information in a highly condensed manner. This makes it difficult to locate individual compounds, reactions, or enzymes. Figure 1 shows a small selection of this wall chart which by the very fact of its outline of reaction arrows emphasizes the problem: we have a highly connected network of reactions that emerge and zero into compounds that sometimes participate in quite as many reactions.

In essence, biochemical pathways form a high-dimensional space that had to be projected into two dimensions to produce the poster. The task is then to exploit the full high-dimensionality of biochemical pathways and allow it to be explored. This is where chemoinformatic methods can come in to allow searches for structures, substructures, reactions and enzymes. To achieve this task, the information on the Biochemical Pathways wall chart had to be stored in a database of structures and reactions.
Figure 1. A cut out from the Biochemical Pathways wall chart.

This is exactly what we did:

- metabolites, coenzymes, and regulators were stored by their name and as connection tables;
- enzymes were stored by name and EC code;
- chemical reactions were indicated by their starting materials and products as well as the enzymes and coenzymes involved. Furthermore, it was indicated whether a reaction occurs as a general pathway, only in animals, in higher plants, or in unicellular organisms. Above all, the bonds broken and made in a reaction were marked.

The representation of chemical structures on the atomic level indicating all atoms and bonds in a molecule in the form of a connection table, allows one to use the full arsenal of chemical database searching methods such as full structure and substructure searching. At the outset of our work no database on biochemical reactions containing structures in the form of connection tables was available.
In the meantime several such databases have appeared such as KEGG [3] and the collection of databases on the BioCyc [4] web page. However, none of these databases have marked the reaction centre, an essential feature for correct reaction searching, as we will see later. This is a unique feature of our BioPath database.

At present, the BioPath database contains about 1500 structures and about 2200 reactions. Having a database with chemical structures in the form of connection tables allows one to enrich the database by data generated by chemoinformatic methods. Thus, we have sent all the molecules contained in the BioPath database through our 3D structure generator CORINA [5,6] and the program ROTATE [7,8] that generates multiple conformations. For each metabolite, coenzyme, and regulator, 3D molecular models for an ensemble of conformations have been stored in the BioPath database, allowing 3D substructure searching. Figure 2 shows a molecular model of coenzyme A as obtained by CORINA.

![Figure 2. The 3D structure of coenzyme A generated by CORINA.](image)

Furthermore, a variety of links have been added to the BioPath database, particularly those to other bioinformatics databases such as BRENDA [9] or those on the ExPASy server [10].
SEARCHING IN THE BioPath DATABASE

In order to exploit the rich information content in the BioPath database we have developed the web based C@AROL retrieval system [11], which allows searches either on chemical structure information or on chemical reaction information.

Chemical structures can be searched by:
- name or name fragments
- full structure
- substructure
- structure similarity
- 3D substructure

Enzymes can be searched by:
- name
- EC code

Chemical reactions can be searched by:
- The structure of the starting materials or products
- Enzyme name or EC code
- Chemical transformation

Clearly, chemical transformation searching is a very important feature because it allows one to search for the essential transformation of a substrate invoked by an enzyme. For such transformation searching, it is essential that the atoms and bonds directly participating in a reaction are marked. In reaction searching it does not suffice to specify the functional groups in the starting materials and in the interconversion of products in which one is interested. Rather, one has also to specify the mapping of the atoms and bonds in the functional groups. Thus, if one is interested in the oxidation of primary alcohols to aldehydes and only specifies that the starting materials should contain an alcohol group and the product should contain an aldehyde group one would obtain the reaction shown at the bottom of Fig. 3. However, in reality this reaction involves the phosphorylation of glyceraldehyde to glyceraldehyde-3-phosphate as a hit, for the starting material having an alcohol function and the product an aldehyde group.
Only if one is indicating the bonds broken and made in a reaction and requires the atoms of the bonds broken to be mapped onto the atoms of the bonds made in the reaction will it be realized that this reaction is a phosphorylation and not an oxidation of an alcohol to an aldehyde.

![Chemical structure diagram]

**Figure 3.** The phosphorylation of glyceraldehydes to glyceraldehyde-3-phosphate (bottom) obtained as a result of the substructure search shown on top.

We do not have sufficient space here to outline all the rich and diverse search possibilities in the BioPath database. A more extensive presentation has recently been published [12]. Only a few examples are presented here. The reader is encouraged to explore the BioPath database by her/himself as it has been made accessible together with the C@ROL retrieval system on the web at: [http://www2.chemie.uni-erlangen.de/services/biopath/index.html](http://www2.chemie.uni-erlangen.de/services/biopath/index.html) and [http://www.mol-net.de/databases/biopath.html](http://www.mol-net.de/databases/biopath.html)

**Structure and Substructure Searching**

The structure editor JME [13] has been integrated into C@ROL allowing easy graphical specification of structure or substructure queries. Figure 4 shows the query for a substructure search. Here the substructure tetrahydropyrane has been input. The search resulted in 83 hits for this example.
Figure 4. Input of a substructure search into C@ROL via JME.

Reaction Searching

Searching by enzyme

Inputting the EC number 3.1.3.3 provided two hits, one reaction being indicated in Fig. 5, the reaction catalysed by phosphoserine phosphatase.

Following the link in the ExPASy server contained in the BioPath database provides the image shown in Fig. 6.

Thus, you can see in which environment this reaction is embedded. Incidentally, this also shows one of the advantages of the BioPath database that clearly highlights this reaction, whereas it is hard to extract from the image in Fig. 6. The markings are ours and are not contained in the image on the ExPASy server.
**Figure 5.** The reaction catalysed by the enzyme phosphoserine phosphatase (EC 3.1.3.3).

**Figure 6.** The reaction of phosphoserine phosphatase (EC 3.1.3.3) depicted on the Biochemical Pathways wall chart.
Searching by transformation

Oxidations are one of the most important reactions in metabolism, on one hand providing energy, and, on the other hand, transforming organic compounds into more water-soluble species, an important step in the excretion of xenobiotics. Figure 7 shows the query for searching for oxidations of C-H bonds to C-OH bonds.

![Figure 7. Reaction centre search in C@ROL: search for reactions converting a C-H bond into a C-OH bond.](image)

This query provided 97 reactions as hit underlining the importance of this reaction type. Figure 8 shows one example from this hit list, the oxidation of leukotriene to 20-hydroxy-leukotriene.
**APPLICATIONS**

We consider the BioPath database as an important milestone in the understanding of the processes that keep living species alive. It is an information source at the interface of bioinformatics and chemoinformatics. Bioinformatics is concerned with the expression of enzymes by genes, whereas now we can use chemoinformatics methods to study the reactions that are controlled by the enzymes, on the very detailed level of individual atoms and bonds in the molecules. We are just at the very beginning of exploiting this rich source of information. Here, two examples are given of the use of the BioPath database to explore details of enzyme action.

**Inhibitors as Transition State Analogues**

Nearly 60 years ago Linus Pauling emphasized that enzymes are complementary in structure to the activated complexes of the reactions that they catalyse.

**Figure 8.** One of 97 hits of reactions of the query of Fig. 7 converting a C-H bond into a C-OH bond.
In other words, enzymes should bind the transition state of a reaction with far greater affinity than the substrate or the products. As a case in point we have analysed the reaction catalysed by AMP deaminase (EC 3.5.4.6) (Fig. 9) converting adenosine-monophosphate (AMP) into inosine-monophosphate (IMP).

The reaction quite certainly proceeds first through the addition of a water molecule to produce the intermediate, also shown in Fig. 9; the transition state of this reaction being close in geometry to this intermediate. Coformycin (Fig. 10) is an inhibitor of this enzyme.

Following Pauling one would conjecture that this inhibitor should be closer in geometry to the intermediate than to AMP or IMP. To test this hypothesis, the inhibitor coformycin was superimposed with AMP, IMP, and the intermediate, respectively. This was done with the aid of the program GAMMA (Genetic Algorithm for Multiple Molecule Alignment) [14]. Figures 11-13 show the three superimpositions.

As can be seen, the geometry of the inhibitor fits best with the geometry of the intermediate, in particular, the two OH-groups are almost perfectly aligned. Apparently, the orientation of this incoming OH-group - which one knows to be coordinated to a Zn$^{2+}$ ion - plays a crucial role in this reaction.
Figure 11. 3D superimposition of AMP with the inhibitor coformycin.

Figure 12. 3D superimposition of IMP with the inhibitor coformycin.

Figure 13. 3D superimposition of the AMP deaminase reaction intermediate with the inhibitor coformycin. The OH groups, which are crucial for proper binding, are very well aligned.
Various other enzyme-catalysed reactions have been studied along similar lines by comparison of the geometry of inhibitors of enzymes [15], with the transition states or intermediates of the reaction catalysed by these enzymes.

All these studies support this hypothesis that an inhibitor of an enzyme should be particularly similar in geometry to the transition state of the reaction catalysed by this enzyme.

**Classification of Enzymes vs. Classification of Reactions**

Enzymes are classified by the widely accepted EC code [16] that builds its classification on a variety of criteria such as reaction patterns, substrates, transferred groups, and acceptor groups. Thus, the EC classification is not quite coherent as, depending on the EC class, the emphasis shifts between different criteria.

Clearly, the most important action of an enzyme is the catalysis of a reaction, an event that breaks and makes bonds. The question therefore is can we build the classification of enzymes on considerations that only take into account the properties of the bonds directly involved in the reaction event? In other words, we want to classify enzyme-catalysed reactions and compare this classification with the EC classification system.

Biochemical reactions are governed by the same kind of physico-chemical effects as more traditional organic reactions, effects that are involved in the discussion of reaction mechanism such as charge distribution, inductive, resonance, or polarizability affect. Some time ago, we developed procedures that allow the calculation of these affects providing values that have shown their importance in modelling reaction mechanisms [17,18].

In order to investigate how reaction classification corresponds with enzyme classification we have chosen reactions that are catalysed by hydrolases. These fall into the EC category 3.1.x.y. The question is then, how will the classification of reactions catalysed by enzymes match classes 3.1.1.y, 3.1.2.y, 3.1.3.y, 3.1.4.y, 3.1.5.y, and 3.1.6.y?

Each bond broken in the reaction with water was characterized by six physico-chemical values, calculated by simple empirical procedures: difference in total charges [19], difference in $\sigma$-electronegativities [19], difference in $\pi$-electronegativities [20], effective bond polarizabilities [21], and delocalization stabilization of a positive or negative charge. In effect, each reaction is then an event in a six-dimensional space spanned by the above six physico-chemical descriptors.
as coordinates, with each bond hydrolysed having a specific value for each of these six coordinates.

In order to visualize the distribution of each reaction in this six-dimensional space we projected this space into two dimensions by a Kohonen neural network [22,23]. The dataset contained 49 reactions which were mapped into a 8x8 Kohonen network. Figure 14 shows the map thus obtained.

**Figure 14.** Kohonen map showing the classification of reactions catalysed by enzymes of EC class 3.1.x.y (49 reactions in dataset).

As can be seen, reactions catalysed by enzymes of different subclasses of the EC code 3.1. are mapped, by and large, into different, well separated areas of the two dimensional map. There are only two squares (neurons) where reactions belonging to different subclasses of EC 3.1. are mapped. Reactions catalysed by enzymes of subclass 3.1.4. and one catalysed by an enzyme of subclass 3.1.5. are mapped in one square. However, it should be realized, that the enzyme EC 3.1.5.1, dGTPase, involves the hydrolysis of a phosphate much like the enzymes of EC 3.1.4.y. Thus, indeed these reactions have much in common, and somehow the EC classification artificially separates them. In the other case, reactions catalysed by enzymes of EC 3.1.2. are mapped, all invoking hydrolysis of thioesters, and one reaction catalysed by EC 3.1.6.2, steryl sulfatase.
Similar results have been obtained in investigations of other EC classes. This leads us to emphasize that the classification of reactions based on physico-chemical affects corresponds to a large extent to the EC classification system.

However, the classification based on reactions provides chemically sounder results and also shows finer details in the reactions catalysed by the enzymes.

**SUMMARY**

The BioPath database constitutes a rich source of information on the structures and reactions involved in the endogenous metabolism. The web based retrieval system C@ROL provides a host of search methods that allow the exploration of these biochemical pathways. We are just at the beginning of discovering new insights into the all-important reactions that keep living species alive. Two examples were given here. The three-dimensional structures of the intermediates in biochemical reactions provide insight into the geometric situations in the binding pocket of enzymes, information that can be used to search for inhibitors of enzymes. In the other application we have shown that the classification of enzyme catalysed reactions based on the physico-chemical affects operating on the bonds directly involved in the reactions corresponds by and large with the EC code classification system. However, it constitutes a chemically more meaningful classification and provides new insights into enzyme reactions.

**ACKNOWLEDGEMENTS**

We appreciate assistance in the construction of the BioPath database and the analysis of its contents through projects funded by the Bundesministerium fuer Bildung und Forschung (BMBF projects no. 08 C 5850 0, 08 C 5879, 031U112D, 031U212D, 031U112A, and 031U212A). We appreciate the initiation of the Biochemical Pathways project by Spektrum Akademischer Verlag and the collaboration with Prof. Guido Moerkotte and Dr Carl-Christian Kanne, University of Mannheim, Germany in establishing our data scheme. Dr Wolf-Dietrich Ihlenfeldt, at that time at the CCC, provided important contributions to this project, most notably the CACTVS system. Discussions with Dr Gerhard Michal were always stimulating. We are indebted to a number of students who carefully input the structures and reactions into
the BioPath database. The BFAM project, initiated by Prof. Hans-Werner Mewes allowed us to continue the work on the BioPath database.

REFERENCES


[6] CORINA can be tested online at http://www2.chemie.uni-erlangen.de/software/corina/free_struct.html and is available from Molecular Networks GmbH, Germany, info@mol-net.de, http://www.mol-net.de.


