molecular pharmaceutics

articles

Prediction of Drug Degradants Using DELPHI: An Expert System for Focusing Knowledge

David L. Pole, Howard Y. Ando,* and Sean T. Murphy

Research Formulations Department and Chemistry Department, Pfizer Global Research and Development, 2800 Plymouth Road, Ann Arbor, Michigan 48105

Received October 2, 2006; Revised Manuscript Received January 30, 2007; Accepted April 18, 2007

Abstract: DELPHI is an expert system that has been developed to predict possible degradants of pharmaceutical compounds under stress testing conditions. It has been programmed with the objective of finding relevant degradation pathways, identifying degradant structures, and providing tools to the analytical chemist to assist in degradation identification. The system makes degradant predictions based on the chemical structure of the drug molecule and precedent from a broad survey of the literature. A description of DELPHI's treatment of molecular perception is described as are many features of the heuristic degradation rules it uses to capture and apply chemical degradation knowledge. DELPHI's utility for capturing institutional knowledge is discussed in relation to an analysis of degradation prediction results for 250 molecules of diverse chemical structure collected over 5 years of use. As such, it provides a reliable, convenient, and rapid tool for evaluating potential pathways of chemical instability of pharmaceuticals.

Keywords: Expert system; drug degradation; stability; stress testing; prediction; institutional knowledge

Introduction

Pharmaceutical stress testing is an expensive and time-consuming part of drug development[1,2] that has not greatly benefited from the vast accumulation of public and corporate knowledge.[3,4] Currently, analytical methods for degradants are empirically developed based on stress-testing rather than on leveraging accumulated knowledge. Similarly, the identification of new degradants in a product is more reactive than proactive. Late identification of problems can often lead to production delays and possibly expensive reformulations. The ability to rapidly predict the potential degradants of a new chemical entity can create value in a number of ways.

For the new bench scientist, a degradant expert system could serve as a means for training and a benchmarking of personal expertise. For the institution or corporation, a degradant expert system can serve as a consistent means of focusing knowledge for problem-solving, proactive analytical method development, and directed-experimental evaluations.[5,6] Public knowledge encompasses the extensive litera-

* Address all correspondence to this author. Mailing address: 2800 Plymouth Rd., Ann Arbor, MI 48105. Phone: (734) 622-1278. Fax: (734) 622-3609. E-mail: Howard.Ando@pfizer.com.

(1) Alsante, K. M.; Friedmann, R. C.; Hatajik, T. D.; Lohr, L. L.; Sharp, T. R.; Snyder, K. D.; Szczesny, E. J. Degradation and impurity analysis for pharmaceutical drug candidates. In *Handbook of Modern Pharmaceutical Analysis*; Ahuja, S., Scypinski, S., Eds.; Separation Science and Technology, Vol. 3; Elsevier: San Diego, CA, 2001; pp 85−172.

(2) Alsante, K. M.; Martin, L.; Baertschi, S. W. A stress testing benchmarking study. *Pharm. Technol.* **2003**, *27* (2), 60, 62, 64, 66−68, 70, 72.

(3) Drucker, P. F. *Post-Capitalist Society*; Harper Business: New York, 1993.

(4) Choo, C. W., Bontis, N., Eds. *The Strategic Management of Intellectual Capital and Organizational Knowledge*; Oxford University Press: New York, 2002; p 772.

(5) Reid, D. L.; Calvitt, C. J.; Zell, M. T.; Miller, K. G.; Kingsmill, C. A. Early Prediction of Pharmaceutical Oxidation Pathways by Computational Chemistry and Forced Degradation. *Pharm. Res.* **2004**, *21* (9), 1708−1717.

(6) Freed, A. L.; Kale, U.; Ando, H.; Rossi, D. T.; Kingsmill, C. A. Improving the detection of degradants and impurities in pharmaceutical drug products by applying mass spectral and chromatographic searching. *J. Pharm. Biomed. Anal.* **2004**, *35* (4), 727−738.

10.1021/mp060103+ CCC: $37.00
Published on Web 06/28/2007

ture and monographs in the physical organic chemistry of chemical reactions and their mechanisms. Institutional knowledge draws on the extensive studies an organization has carried out on the chemical stability of the active pharmaceutical ingredient and on formulated products. Normally, this information is highly perishable due to the high attrition rate of pharmaceutical candidates. However, knowledge management of this intangible asset can create value if captured in a way that can be easily applied to future situations.[7] DELPHI, degradant expert leading to pharmaceutical insight, is an expert system that was designed to utilize both public and corporate chemical reaction and degradant information to predict potential pharmaceutically relevant degradants.

Experimental Section

Knowledge Sources for Drug Degradation Prediction. While it is clear that the ability to predict possible drug degradants in a routine fashion can have a significant impact on degradant identification, the scope and complexity of the problem balloons as the diversity in chemical structure, reactivity, stability conditions, and the physical state or purity of the stressed material are considered. In order to capture knowledge into an expert system, it is necessary to pull in data from as many places as possible and to bind that knowledge together into a consistent, rational framework.

There are no universal sources of knowledge or many effective tools available for the *in silico* prediction of drug degradants from molecular structure. Several computational systems are available that have been applied to the problem of drug degradation,[8−10] although none are specifically designed to address the unique reactivity space defined by pharmaceutical sciences problems. DELPHI has been constructed through an examination of published works including books, reviews, and journal articles in the field of pharmaceutical stability and degradant identification. While literature precedent can be a powerful source of information, it is frequently laborious, incomplete, and sometimes unreliable. For instance, many articles detailing pharmaceutical stability identify only decrease in parent and do not characterize the chemical structure of the degradants. That being said, the rules generated from well-studied examples can be applied readily to novel chemical structures. In order to do this, the rules must be interpreted in a general fashion so that they can be applied as broadly as possible.

DELPHI also benefits from our internal experience with drug degradation. Internal expertise in a large corporation is more abundant in research reports, technical memos, and various forms of internal communication. However, the most valuable source of internal expertise is the human expert. Experienced analysts and formulators have frequently accumulated good insight into problem structures and sources of instability. Internal expertise, however, is not well organized, is frequently anecdotal, and can conflict.

Even under the best conditions, the large number of relevant variables makes extrapolation of rules to related structures difficult. In addition, current literature is quite sparse on molecules that would represent the vast array of new synthetic templates coming from modern medicinal chemistry efforts.

The available drug degradation literature must also be considered in light of fundamental organic chemistry; the combination can prove very powerful since the defined rules of organic chemistry help put the mottled empirical experience into a consistent, mechanistically accurate continuum. Organic chemistry can also facilitate interpolation and frequently extrapolation of single observations into well-founded heuristic principles. For instance, if the literature describes a reaction of a nucleophile, surely stronger nucleophiles will also have the potential to do the same sort of chemistry, all things being equal. Definition of ring structures, aromaticity, functional groups, and reaction types are organized best following the established rules of organic chemistry. Among others, reactivity of functional groups, linear free-energy relationships, regioselectivity, and frontier molecular orbital theory provide a sound theoretical architecture for building the chemical intelligence of DELPHI within which the less structured knowledge of drug degradation can be housed and manipulated.

There are several major sources of information for the development of the expert system. An excellent resource for many aspects of drug degradation and stress testing methods is the recent book by Baertschi[11] including a comprehensive chapter by Baertschi and Alsante on the Chemistry of Drug Degradation.[12] A particularly good, if somewhat dated, resource is Connors, Amidon, and Stella's book *Chemical Stability of Pharmaceuticals*,[13] which contains dozens of monographs on the degradation of pharmaceuticals. For issues relating to fundamental organic chemistry, March's

(7) Sveiby, K. E. *The New Organizational Wealth: Managing & Measuring Knowledge-Based Assets*; Berrett-Koehler: San Francisco, 1997.

(8) Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J.; Fleischer, J. M.; Gothe, S. A.; Helson, H. E.; Paderes, G. D.; Sinclair, S. CAMEO: a program for the logical prediction of the products of organic reactions. *Pure Appl. Chem.* **1990**, *62* (10), 1921−1932.

(9) Hoellering, R.; Gasteiger, J.; Steinhauer, L.; Schulz, K. P.; Herwig, A. Simulation of Organic Reactions: From the Degradation of Chemicals to Combinatorial Synthesis. *J. Chem. Inf. Comput. Sci.* **2000**, *40* (2), 482−494.

(10) Ihlenfeldt, W. D.; Gasteiger, J. Computer-Assisted Planning of Organic Syntheses: The Second Generation of Programs. *Angew. Chem., Int. Ed. Engl.* **1996**, *34* (2324), 2613−2633.

(11) Baertschi, S. W., Ed. *Pharmaceutical Stress Testing: Predicting Drug Degradation*, 1st ed.; Taylor & Francis: Boca Raton, 2005.

(12) Baertschi, S. W.; Alsante, K. M. Stress testing: the chemistry of drug degradation. In *Pharmaceutical Stress Testing Predicting Drug Degradation*; Baertschi, S. W., Ed.; Taylor & Francis: Boca Raton, 2005; p 51−140.

(13) Connors, K. A.; Amidon, G. L.; Stella, V. J. *Chemical Stability of Pharmaceuticals. A Handbook for Pharmacists*, 2nd ed.; J. Wiley & Sons: New York, 1986.

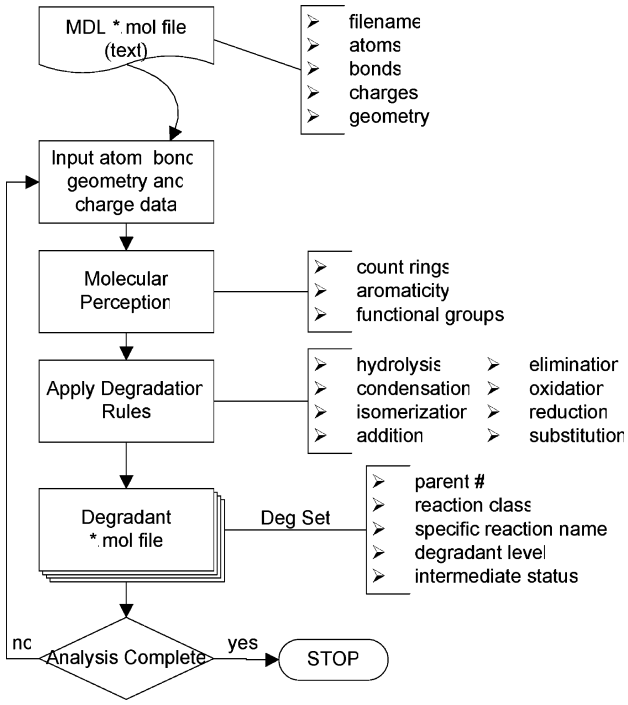


Figure 1. Schematic of the overall logic flow of DELPHI's degradation prediction. Input of structural information and the perception of relevant structural features of the molecule are followed by an application of heuristic degradation rules to generate a set of potential degradation products.

book[14] is a comprehensive source of information and references. For photodegradation information, Greenhill's reviews[15,16] have been a particularly useful resource. These main sources have been combined with information from the broader chemical and pharmaceutical literature in order to arrive at a fairly extensive set of rules.

Organization of Knowledge. The "chemical brain" of DELPHI is programmed in Amzi! Prolog, a programming language well suited to expert systems.[17] The Prolog code contains the chemical heuristic rules that we have developed to recognize molecular features (rings, aromaticity, functional groups) and rules of chemical reactivity. Molecular information (atoms and bond connectivity) is input from a source *.mol file. The molecular features are perceived in the artificial intelligence of the system, and the degradation rules are applied. Each degradant in turn is subjected to a similar analysis to create a degradation set. The program's overall logic flow is illustrated in Figure 1.

DELPHI's prolog engine uses the MDL molfile format[18] as a means to input and manipulate chemical structure information (atom, bond, geometry, and charge). Other structural formats are converted to MDL molfile format prior to being interpreted by Prolog, which translates enough relevant information to create a prolog representation of the structure. The prolog engine interprets the rings, aromaticity, and functional groups present in the molecule. Molecular perception is performed through atom-by-atom mapping cross-referenced against a catalogue of heuristic rules. A set of heuristic degradation rules is also applied to identify degradation pathways including perception of the molecular features of a potential reactant, examination of reaction conditions, and classification of reactive atoms, and, finally, the bond orders between reactive atoms are incremented or decremented according to a reaction code. The results of degradation rules are collected as a degradant set. Identified degradants are subjected to the same analysis until a predefined condition is reached (i.e., number of degradants).

Molecular Perception with Prolog. Prolog is a rule-based language and, as such, is well suited for encoding chemical definitions, principles, and rules. The Prolog routines perform all operations requiring chemical intelligence including handling the four main components in degradation prediction: ring perception, functional group analysis, reactivity analysis, and product analysis.

In structure analysis, the parent molecule is first analyzed for completeness and hydrogen atoms are added if necessary. The molecule is then analyzed for rings—first by finding the smallest set of smallest rings[19−23] and then by identifying aromatic rings and bicyclic systems. The parent molecule is also examined for double bonds that are capable of isomerization, and the initial geometry of these bonds in the parent structure is recorded.

Once the ring size has been identified, DELPHI attempts to match the rings to several known aromatic ring types identified in Table 1. In practice, key atoms in a chemical reaction can be compared to lists of atoms in aromatic rings

(14) March, J. *Advanced Organic Chemistry. Reactions, Mechanisms, and Structure*, 4th ed.; J. Wiley & Sons: New York, 1992.
(15) Greenhill, J. V. Is the photodecomposition of drugs predictable? Photostability of Drugs and Drug Formulations. *Int. Meet. Photostab. Drugs, 1st, Oslo, June 1995* **1995**, 83−110.
(16) Greenhill, J. V.; McLelland, M. A. Photodecomposition of drugs. *Prog. Med. Chem.* **1990**, *27*, 51−121.
(17) 2003. Amzi! Prolog + Logic Server Eclipse IDE. 7.0 ed., Lebanon, OH: Amzi! Inc.
(18) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of several chemical structure file formats used by computer programs developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32* (3), 244−255.
(19) Baumer, L.; Sala, G.; Sello, G. Ring perception in organic structures: a new algorithm for finding SSSR (smallest set of smallest rings). *Comput. Chem. (Oxford, U.K.) 15* (4), 293−299.
(20) Downs, G. M.; Gillet, V. J.; Holliday, J. D.; Lynch, M. F. Theoretical aspects of ring perception and development of the extended set of smallest rings concept. *J. Chem. Inf. Comput. Sci.* **1989**, *29* (3), 187−206.
(21) Balducci, R.; Pearlman, R. S. Efficient exact solution of the ring perception problem. *J. Chem. Inf. Comput. Sci.* **1994**, *34* (4), 822−831.
(22) Figueras, J. Ring Perception Using Breadth-First Search. *J. Chem. Inf. Comput. Sci.* **1996**, *36* (5), 986−991.
(23) Gasteiger, J.; Jochum, C. An algorithm for the perception of synthetically important rings. *J. Chem. Inf. Comput. Sci.* **1979**, *19* (1), 43−48.

Table 1. Catalogue of Aromatic Ring and Functional Group Features Perceived by DELPHI[a]

	Aromatic Rings	
benzene	furan	1,2,3-triazole
naphthalene	thiophene	1,3,5-triazole
anthracene	thiophene oxide	1,3,4-triazole
penanthrene	imidazole	1,2,3,4-tetrazole
pyridine	oxazole	1,2,3,5-tetrazole
pyrazine	thiazole	1,2-pyridone
pyrimidine	1,3,4-thiadiazole	1,4-pyridone
pyridazine	pyrazole	1,2-pyridazinone
pyrrole		
	Acyclic Functional Groups	
primary alcohol (Nu)	vinyl	thiol ether (Lg)
secondary alcohol (Nu)	alkene	amino-thiol ether
tertiary alcohol (Nu)	imine (CC) (MA)	sulfone (MA)
phenol (Nu)	hydrazone (CC) (MA)	sulfoxide (MA)
nitroso	hydrazine (Nu)	sulfonate ester
1,1-diol	amidine (MA)	sulfonic acid (MA)
transesterification intermediate	diene	sulfamic acid
ketone (CC) (MA)	alkyne	sulfamic acid ester
aldehyde (CC) (MA)	cyano (MA)	sulfonamide (MA)
formyl	allene	sulfonimide
thiol ester (MA) (CG)	ketene	peroxide
thiol acid (MA) (CG)	ether	hydrogen peroxide
carboxylic acid (MA) (CG)	fluoride (MA) (Lg)	phosphate ester
anhydride (MA) (CG)	chloride (MA) (Lg) (AA)	phosphonic acid
ester (MA) (CG) (Lg)	bromide (MA) (Lg) (AA)	phosphonic ester (MA)
aromatic ester (CG)	iodide (MA) (Lg) (AA)	phosphate amide
imide (MA) (CG) (Lg)	tertiary amine (Nu)	phosphamide
urea (Nu) (CG)	secondary amine (Nu)	phosphoramide
carbonate (CG)	primary amine (Nu)	aromatic amine
carbamate (CG) (Lg)	hydroxylamine (Nu)	TriHet
amide (Nu) (MA) (CG)	1° hydroxylamine (Nu)	DiHet
formamide	2° hydroxylamine (Nu)	imidate
oxime (CC) (MA)	thiol (Nu) (Lg)	
	Charged Functional Groups	
carboxylate anion	nitro	nitrone
ammonium ion (MA) (Lg)	amine oxide (MA) (Lg)	episulfonium ion (AA)
carbocation	aromatic amine oxide	aziridinium ion (AA)
	Cyclic Functional Groups	
hemiacetal (Nu) (Lg)	thiaza-ketal	epoxide (AA)
ketal (Lg)	beta-lactone (CG)	aziridine (AA)
acetal (Lg)	beta-lactam (CG)	episulfide (AA)
ortho ester	gamma-lactone	1,3-dicarbonyl tautomer
thiaza-acetal	gamma-lactam	uracil fragment (MA) (CG)
	Radical Functional Groups	
benzyl radical	phenoxyl radical	formyl radical
alkoxyl radical	alkoxyl radical	carboxyl radical
cyclopropyl radical	cyclopropyl radical	sulfonyl radical

[a] Functional groups are further grouped by reactivity including Nu = nucleophiles; CC = condensable carbonyls; MA = Michael acceptors activating groups; CG = condensation group (through intramolecular condensation with nucleophile); Lg = leaving group; AA = alkylating agents.

and reactivity can be adjusted accordingly. Aromatic rings are recognized as having either alternating single and double bonds in a six-membered ring or a five-membered ring with a heteroatom and alternating single and double bonds. Both aromatic and heteroaromatic rings included in Table 1 are perceived by DELPHI.

Functional groups are analyzed in three stages. First, acyclic and uncharged groups are found. Then, cyclic functional groups, such as epoxides and lactam rings, are found. Finally, charged groups, such as ammonium and phosphate ions, and odd-electron species, such as free radicals, are identified. Secondary functional groups can also

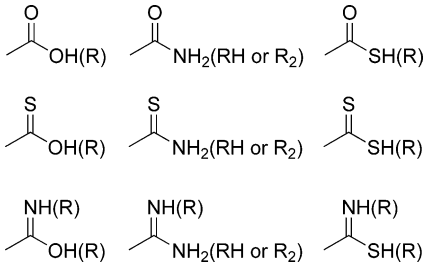


Figure 2. The generic functional group DiHet flags molecular reactivity for a variety of complex functional groups.

be defined that are built on combinations or special types of simple functional groups. The Prolog code for functional groups is composed of definitions in terms of necessary bonds and atoms and a specific arrangement. Table 1 contains a list of functional groups perceived by DELPHI.

Since all reactivity in organic chemistry derives from interactions of functional groups, their accurate perception and interpretation will impact all subsequent areas of chemical prediction. Functional group perception involves defining and naming each functional group as a separate species. This strategy is superior to a structural fragment approach for our purposes since we use the functional groups to define regions of specific reactivity.

There is an expectation that reactive functional groups would be detected by DELPHI for nearly any new molecule input into the system. As such, the definition of rings and functional groups is broad and includes several more generic functional groups that would be expected to be reactive. For example, the generic functional groups DiHet and TriHet represent several possible combinations of reactive nitrogen, oxygen, or sulfur atoms without exhaustively specifying names for each functional group. The group DiHet, for instance, represents 15 combinations of OH, OR, SH, SR, and N(H/R) (see Figure 2). The functional group TriHet represents 75 such combinations.

Degradant Reactivity. Once functional groups have been identified, the reaction modules look for reactive combinations of these groups. The reaction modules are divided into the following sections: hydrolysis, condensation, isomerization, addition, elimination, oxidation, reduction, and substitution. Reactions are found by examining the functional groups found using the heuristic database of reaction rules.

When considering the problem of prediction of pharmaceutical degradation, the challenge quickly becomes overwhelming. The scope of the problem expands rapidly as the diversity in chemical structure, reactivity, stability conditions, and the physical state or purity of the stressed material is considered. It is impossible in an *in silico* setting to be able to anticipate those conditions to which a drug will be subjected since this is not standardized. DELPHI attempts to predict all products that might be reasonably expected given the breadth of pharmaceutically relevant degradation conditions. DELPHI uses expert system logic to predict possible degradants given the combination of molecular reactivity and the broad definition of pharmaceutically relevant reaction conditions. The resulting set of predicted degradation products thus represents the union of the reactivity rules captured in the expert system with the particular structural features of the molecule.

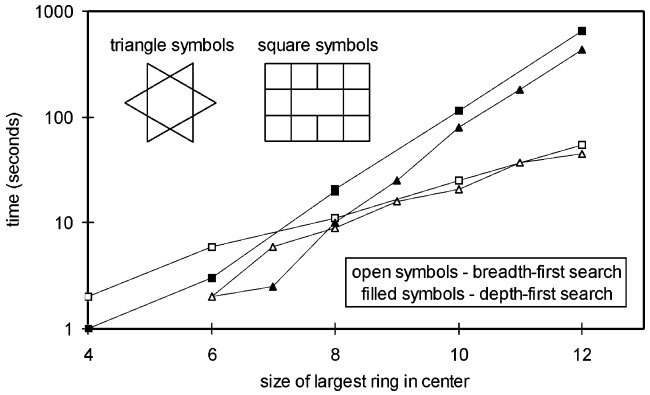


Figure 3. Comparison of depth-first and breadth-first ring-finding algorithms. Two series of test ring systems were used, one with triangles on the edges of a ring (△ symbols on graph) and another with squares on the edges (□ symbols on graph). The six-membered ring example for each series is shown. Note the *y*-axis is on a logarithmic scale.

Results

Ring Perception. An efficient algorithm has been implemented to find the smallest set of smallest rings (SSSR),[19,20] the size of which can be calculated from the number of atoms and bonds using Euler's formula. Two algorithms are implemented in the expert system, a depth-first search using the searching capabilities of Prolog and a breadth-first search from Balducci and Pearlman.[21] The depth-first search goes through the molecule atom-by-atom and checks if it is part of a ring of a specified size. In contrast, the breadth-first search checks all atoms simultaneously for inclusion in rings of a specified size. Because the depth-first search algorithm uses the built-in searching capabilities of Prolog, the algorithm is quite efficient for smaller size rings (<eight-membered rings). On the other hand, the breadth-first search, which is more difficult to implement in Prolog, was designed to be efficient for all size rings, and our implementation is faster than the depth-first algorithm for large-size rings (see Figure 3 for a comparison). A hybrid of these two search algorithms is typically used by DELPHI as the default option. In the hybrid algorithm, the depth-first search is initiated, but if all of the rings are not found after it has searched for eight-membered rings, the program uses the breadth-first routine to find the rest.

Functional Group Bundling. Functional group bundling is used to bin functional groups in terms of reactivity. In addition to their specific, individual reactivity, many functional groups play roles as nucleophiles, electrophiles, leaving groups, or electron donating or withdrawing groups. Rules of physical organic chemistry can be applied to make accurate predictions about the relative reactivity of functional groups in these roles. DELPHI contains subgroups of the

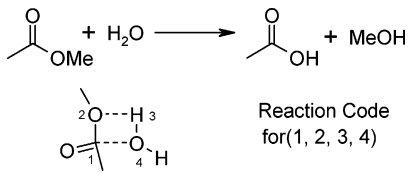


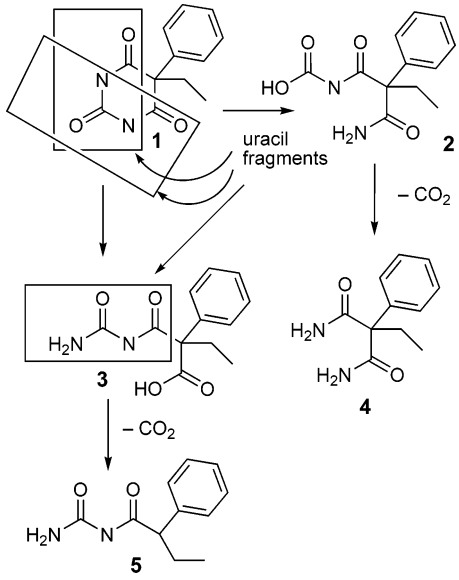
Figure 4. An illustration of an application of the Hendrickson line notation to identify a chemical reaction center.

functional groups described in Table 1 that are organized into lists of leaving groups, Michael acceptors, reactive carbonyl groups, nucleophiles, and alkylating agents among others.

Coding Molecular Reactivity. A special Prolog module, the reactor module, processes the reaction codes and generates the product structure from the reactant and any other reactant molecules (water, oxygen, etc.). If a new reaction is found experimentally, a Prolog reaction code is created to specify the bonds to be made and broken to form the degradation product. The reaction code is based on a scheme developed by Hendrickson[24] for the classification of organic reactions. Hendrickson has devised a nomenclature that condenses a great variety of reaction mechanisms into a simple line notation. Reaction mechanisms are represented by mapping the atoms in the reaction center onto planar geometric polygons, typically of 3−6 vertices, with defined bond making and bond forming functions. An adaptation of this convenient nomenclature is illustrated in Figure 4. This nomenclature readily lends itself to implementation into an expert system. The top line indicates the reaction scheme for ester hydrolysis. The bottom line contains the simplified representation of the reaction center as a square shape with forming and breaking bonds. The simple Hendrickson line notation for the reaction is described using the atom numbers to represent the atoms in the reaction center.

The result of execution of the reaction code is the generation of a new molecular structure representing the degradant product. The new product is compared to a list of previously found products to prevent duplication. The expert system uses an atom-by-atom mapping routine in order to determine if a new degradant is the same as a degradant found previously. The atom-by-atom mapping determines whether two molecules have the same molecular structure. We call two structures with the same molecular structure isomorphic (same shape or morphology). Isomorphism will occur when the same atoms are bonded in exactly the same way in the two molecules. In addition, the new product is then analyzed to make sure it does not violate any chemical rules, such as Bredt's rule[25] or the restriction that double bonds in small rings must be *cis*. This element of quality control helps to ensure that the end user is presented only with chemically reasonable structures.

Scheme 1. Degradation Scheme for Phenobarbital as Described in Ref 13



Drug Degradation Predictions

The pharmaceutical degradation of phenobarbital has been described in detail.[13] This simple molecule will provide a good demonstration of DELPHI's degradant prediction abilities. Phenobarbital is known to undergo degradation in the presence of moisture. The known chemistry related in ref 13 is illustrated in Scheme 1. DELPHI prediction of degradation pathways for this molecule is represented in Scheme 2. The predicted degradation scheme matches that known from the literature quite well. The key degradants (species **1**−**5**) are identified and their relations to each other specified.

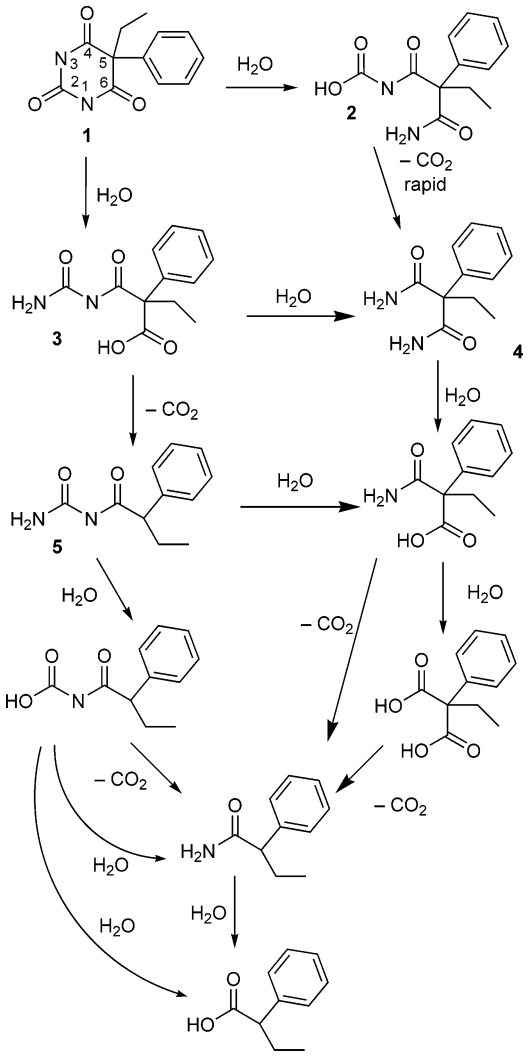
Phenobarbital is a symmetrical molecule with the complex heterocyclic six-membered ring characteristic of barbituric acid derivatives. DELPHI perceives this ring as well as the phenyl ring, which is also identified as having aromatic character. The barbituric acid functionality of phenobarbital is an instructive example of the complex functionality present in modern pharmaceuticals. Fragments of the barbituric acid could be perceived as amine, carbonyl, amide, urea, and imide functional groups; however, the definition of functional groups is always made based on the basis of chemical reactivity. DELPHI perceives the boxed substructures of phenobarbital as uracil derivatives and assigns reactivity from this definition. Although a uracil fragment is not a classical organic functional group definition, it is practical in this case to more specifically identify the degradation chemistry that occurs with this molecule.

Once functional groups have been identified, the reaction modules map perceived functionality against a set of heuristic rules describing known reactivity of a great number of reactive functional groups. The reaction modules are divided into hydrolysis, condensation, isomerization, addition, elimination, oxidation, reduction, and substitution sections. Reactions are defined around the expected chemistry for the particular functional groups identified in the molecule.

(24) Hendrickson, J. B. Comprehensive System for Classification and Nomenclature of Organic Reactions. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (5), 852−860.

(25) Bredt, J.; Thouet, H.; Schmit, J. *Justus Liebigs Ann. Chem.* **1924**, *437*, 1.

Scheme 2. Degradation Scheme for Phenobarbital as Predicted by DELPHI[a]



[a] The compound numbering in this scheme corresponds with that in Scheme 1. Additional degradation pathways are predicted.

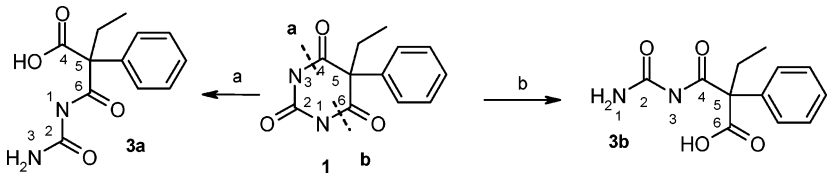
DELPHI provides additional information beyond that available from ref 13. For instance, there are several pathways to get several of the key compounds, a fact that is not obvious from Scheme 1, yet can provide insight to the scientist attempting to gain better understanding of a molecule's reactivity. The fact that DELPHI predicts not only the end product degradant but the intermediates that might be involved in a given pathway can allow an analyst to carry out focused stability studies under various conditions. Milder conditions should be represented by degradants early in the pathway while harsher conditions will correlate with later degradants. Such proactive knowledge can help to avoid surprises and can even lead to strategies to eliminate the possibility of a particular degradation reaction.

Phenobarbital hydrolyzes by breaking either the 1−2 bond (i.e., **1** → **2**) or the 1−6 bond (i.e., **1** → **3**) as illustrated in Scheme 3. In order to predict the hydrolysis primary pathways (**1** → **2** and **1** → **3**), DELPHI recognizes the reactivity of the barbituric acid, introduces the water molecule into its knowledge base, identifies the bonds on phenobarbital and water that will need to break or form, and then deals with the product(s) of the reaction. In the example in Scheme 3, identical degradation products from two different pathways and with two different atom numberings are possible and thus new degradants are mapped, atom-by-atom, with the entire list of previous degradants to ensure that duplications are eliminated.

Reactive Intermediates. Degradant **2** is a substituted carbamic acid. Carbamic acids are quite unstable and cannot generally be isolated. It is common that a molecule formed in a chemical reaction is much more reactive than its parent. Such molecules are termed reactive intermediates. Intermediates are defined by DELPHI as species that are sufficiently unstable chemically that they may not be isolable. Such species may be ions, radicals, or other reactive intermediates such as enols, carbamic acid derivatives, and α-haloamines. Identification of a molecule as a reactive intermediate is based on the structure of the molecule. Certain structural features are responsible for the high reactivity. For example, the enol **6** can be distinguished as a reactive intermediate by the presence of a hydroxyl group bound to a nonaromatic alkene (Scheme 4). The presence of a single functional group characteristic for high reactivity renders the entire molecule an intermediate.

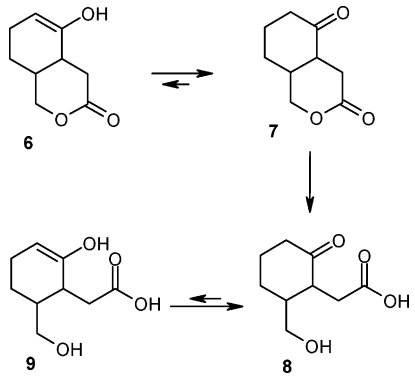
DELPHI must be able to distinguish intermediates from stable degradants for two reasons. First, the expert system must be able to tell the end user that a particular degradant should not be expected in an HPLC trace or some other analytical method. Second, the reactive intermediate will undergo reaction at the highly reactive site much faster than it will react at any of the other functional groups that the molecule possesses. For instance the enol **6** will tautomerize to the keto **7**; hydrolysis of the lactone of **6** will not occur in competition with tautomerization (Scheme 4). Lactone hydrolysis is possible, but only from the keto tautomer to give **8**. In this way, unreasonable products are not found in the degradation pathway. For instance, it would be unreason-

Scheme 3. The Effect of the Two Symmetric Uracil Fragments of Phenobarbital on the Generation of Identical Degradants[a]



[a] Degradants **3a** and **3b** are identical, but the atom numbering is different.

Scheme 4. Behavior of Reactive Intermediates[a]



[a] Formation of **9** by hydrolysis of **6** is not predicted because tautomerization of the enol should occur much faster than lactone hydrolysis.

able for the product **9** to be predicted to be formed directly from hydrolysis of **6**.

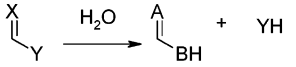
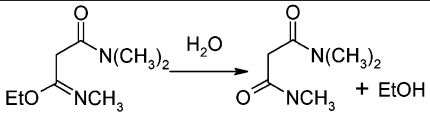
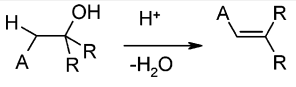
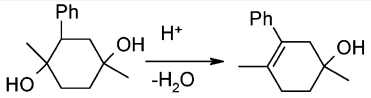
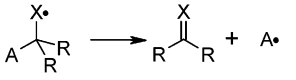
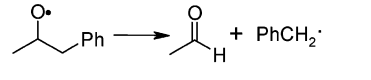
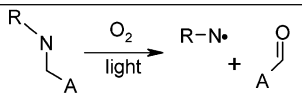
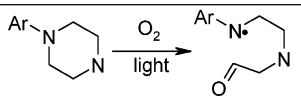
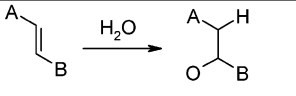
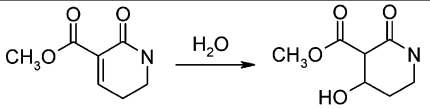
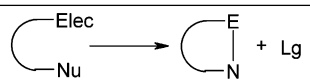
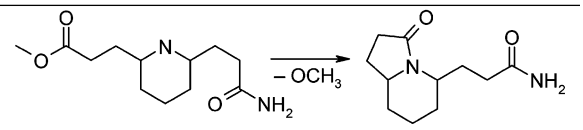
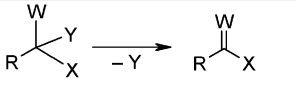
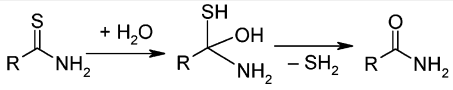
In order to put the higher reactivity of the enol tautomer into context, DELPHI treats the reactions of intermediates separately from the reactions of more stable degradants. For instance, degradant **2** undergoes rapid decarboxylation of the carbamic acid to give **4**. Hydrolysis of the amide or hydrolyses of the imide functional group do not occur until *after* decarboxylation.

Decarboxylation of the carbamic acid intermediate **2** differs from decarboxylation of degradant **3**, which is designated as a stable degradant. Decarboxylation of **3** is also predicted by the expert system (to give **5**), but the reaction would occur in conjunction with any other chemistry that might be possible for the molecule (i.e., hydrolysis of the uracil fragment). Distinguishing intermediates from stable degradants gives us two levels of chemical reactivity that can be dealt with independently. As a result, the quality of the expert system analysis is greatly improved.

Contexting Degradation Reactivity. The treatment of reactive intermediates is not the only example of contexted reactivity used in DELPHI to introduce a higher quality degradation prediction. The great diversity of chemical structure space provides challenges to the prediction of degradation reactivity. When reactivity is distilled down to a simple transformation of functional groups, the context of this chemistry in the larger reactivity of the molecule cannot be lost. For instance, hydrolysis of both amides and esters is well-characterized in the pharmaceutical literature. However, in a molecule with both an ester and an amide, it is certain that hydrolysis of the more labile ester will dominate the chemistry. To make a successful prediction of reactivity, DELPHI must put the individual reactivity of the functional groups into the larger context of the reactivity of the whole molecule.

Hydrolysis, oxidation, and even intramolecular reactivities can be ranked to refine degradation predictions. A selection of these contexted reactions is described in Table 2. According to the mechanism described by Greenhill,[15,16] the photooxidation of a benzylamine, for instance, will dominate the oxidation of purely aliphatic amines. Thus, for the molecule **10** in Scheme 5, DELPHI predicts the degradant

Table 2. Examples of Types of Reactivity Contexting Implemented within DELPHI[a]

Example Rankings	Generic Reaction	Contexted Example
Hydrolysis of Reactive Carboxyls ranked on relative reactivity to hydrolysis [β-lactone; thiol ester] > [β-lactam; ester] > [Amide; Amidine] > Urea		
Elimination of Tertiary Alcohol ranked on stability of resulting alkene [aryl; EWG subst.] > 2° alkene > 1° alkene > methyl		
β-Cleavage ranked on radical stability [Ph_2CH• ; $(EWG)_2CH$•] > [$PhCH_2$• ; $(EWG)CH_2$•] >[RO•;R_2N•; 3°alkyl] >[2°alkyl] >[1°alkyl] >CH_3•		
N-Dealkylation of Amine [R=Ar ; A = $(EWG)_2$] > [R=Ar ; A = EWG] ~ [R=3° alkyl ; A = $(EWG)_2$] > [R=Ar ; A = 3° alkyl] > [R=3° alkyl; A = 3° alkyl]		
Hydration of Alkene ranked on electrophilicity of double bond atom [$(strongEWG)_2C$=**C**] > [(strongEWG)(weakEWG)C=**C**]		
Condensation ranked on reactivity of Nucleophile / Electrophile Combination [β-lactam + amine] > [β-lactam + alcohol] ~ [Ester + amine] > [Ester + alcohol] ~ [Amide + amine]		
Acetal elimination ranked on leaving group strength. Leaving group SH > OH > NH.		

[a] EWG = electron withdrawing group; Lg = leaving group; Nu = nucleophile; Elec = electrophile.

Scheme 5. Illustration of the Contexting of the Photooxidation of Amines Showing Ranking Based on Relative Reactivity

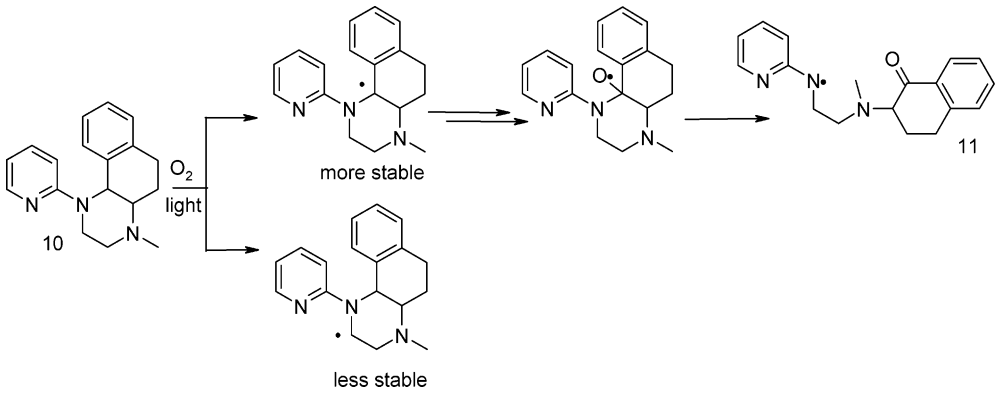


Table 3. Frequency of Occurrence of Reaction Classes in DELPHI Database

reaction class	frequency	number
oxidation	4535	23
radical	3583	25
hydrolysis	1799	82
reactive intermediate	1126	45
elimination	292	64
condensation	229	51
reduction	215	6
isomerization	197	18
addition	136	56
substitution	59	4
total	12171	374

Table 4. Analysis of the Frequency of Occurrence of Reactions Predicted by DELPHI

reaction name	reaction class	frequency
H-atom abstraction α to amine	oxidation	1214
benzylic hydroxylation	oxidation	930
alkoxyl radical - H abstraction	radical	676
N-oxide formation on aromatic N	oxidation	645
radical beta-cleavage of H atom	radical	632
benzylic hydroperoxylation	oxidation	626
radical beta-cleavage of aminyl	radical	546
nonaromatic amine → *N*-oxide	oxidation	474
radical beta-cleavage of Bz group	radical	347
amide hydrolysis	hydrolysis	314
free radical - H abstraction	radical	310
aminyl radical - H abstraction	radical	285
hydrated imine	intermediate	232
photodehalogenation of aromatic Cl	reduction	190
sulfonamide photocleavage	oxidation	153
beta-cleavage of aminoradical	radical	153
aldehyde to carboxylic acid	oxidation	152
2° amine hydrolysis near aromatic "N"	hydrolysis	148
imine hydrolysis	hydrolysis	146
carbamate hydrolysis	hydrolysis	126

11. The pyridyl-substituted nitrogen in the piperazine activates this amine toward photooxidation, and the greater stability of the benzyl-stabilized intermediates favors the product **11**.[15,16] The variety of possible degradation products from the one reaction type can be simplified to a single dominating degradant by building the reaction rule in the context of the greater reactivity of the molecule.

To achieve these results, DELPHI uses several ranked lists of functional groups and reactions. Nucleophiles, electrophiles, leaving groups, and other reactivities are collected and bundled into broad classes, thus, a great number of reactions provide both possible reactions and some ranking of reactivity.

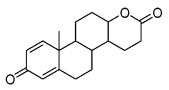
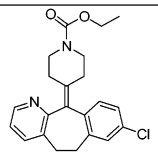
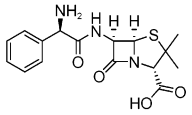
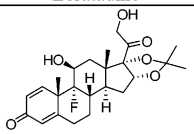
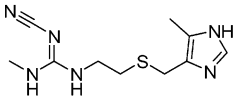
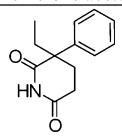
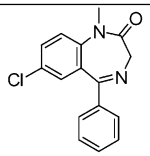
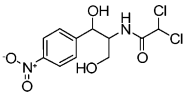
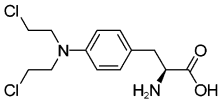
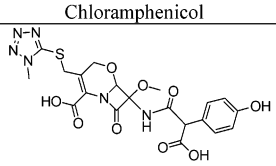
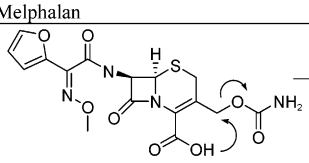
DELPHI's ability to put reactivity into context has some limitations but is nevertheless a powerful tool. The limitation in this contexting is that reactions that span diverse reactivities—oxidation and hydrolysis, for instance—cannot be ranked since the reactions conditions have no frame of reference in common. Thus it is impossible to say whether oxidation of an amine or hydrolysis of an amide will predominate given the completely in silico nature of the program.

Data Mining. DELPHI has been used for the last 5+ years to make predictions of potential degradation products to aid early forced or purposeful degradation studies. Over this time, roughly 250 molecules have been analyzed by DELPHI and the results saved to a database. The analyses are typically performed at the discovery and preclinical stages, and the molecules analyzed represent real-world chemical structures from 10+ therapeutic areas. In this analysis of a wide set of molecules, it should be reiterated that the number of times reactivity is found is a measure of how many times the reactive functionality has been found—not necessarily the facility with which the reaction occurs.

The stored results offer an opportunity to mine useful trends from the collected dataset. For the sake of consistency, all data on degradants is limited to primary (direct products of the parent molecule) and secondary (products of the primary degradants) degradants. There are 374 unique reactions specified in the database, which, as detailed in Table 3, broadly span the range of reaction classes. When these 374 unique reactions are applied to the 250 molecules in the database, a total of 12171 individual instances are found.

This analysis is noteworthy in that oxidation is predicted very frequently even though the number of unique oxidation rules is less than one-third that of hydrolysis. Oxidation is a critically important mechanism for pharmaceutical stability

Table 5. Comparison of Experimental and Predicted Results for a Selection of Literature Reactivity

Compound	Experimental Reactivity	Compound	Experimental Reactivity
 Testolactone	✓ Lactone hydrolysis	 Loratadine	✓ Carbamate hydrolysis
 Ampicillin	✓ β-lactam hydrolysis ✓ Dimerization	 Triamcinolone acetonide	✓ Ketal hydrolysis
 Cimetidine	✓ Nitrile Hydrolysis	 Glutethimide	✓ Imine hydrolysis
 Diazepam	✓ Imine hydrolysis ✓ Amide hydrolysis	 Chloramphenicol	✓ Amide hydrolysis
 Melphalan	✓ Chloride hydrolysis	 Moxalactam	✓ Aliphatic decarboxylation
 Cefuroxime			✓ Carbamate hydrolysis × Lactone formation

✓ = Experimental reactivity predicted by DELPHI; × = Experimental reactivity not predicted by DELPHI

of both the active pharmaceutical ingredient (API) and formulated drug product.[26,27] This indicates that chemical functionalities prone to oxidation are ubiquitous in candidate molecules. The classification "radical" refers to subsequent chemistry of a free radical species. Because radical chemistry propagates in a chain, it is not surprising that the frequency of radical reactions is relatively high. Oxidation reactions lead to radicals, and therefore both must be considered when evaluating the extent of oxidation chemistry. Together, oxidation and radical reactions constitute 67% of all chemistry predicted across the test set of candidate molecules. By this analysis, one can anticipate that greater understanding of oxidation-based degradation chemistry has the ability to improve the accuracy of degradation prediction.

The results can be broken down more specifically by looking beyond the broad reaction classes and into specific degradation rules. Table 4 provides a tabulation of the 20 most frequently observed reaction mechanisms and represents a total of 68% of the total predicted reactions. The majority of this reactivity arises from a relatively small, ubiquitous chemical space. Alkyl tertiary amines are frequently employed in medicinal chemistry to impart ionization and solubility to a lead molecule in order to improve biopharmaceutical properties. These tertiary amines are known to be reactive through many pharmaceutically relevant pathways including *N*-oxide formation and H-atom abstraction from the weakened C−H bond in the position α to the amine. Aromatic amines (substituted pyridine, pyrimidine) as well as sp^2 nitrogens in five-membered heteroaromatic systems are prone to *N*-oxide formation and of widespread utility as H-bond acceptors in molecular interactions at enzymatic receptor sites. Benzylic C−H bonds are also very common in pharmaceuticals and have been shown to be more prone to oxidation because of weaker C−H bond strength.[28]

(26) Waterman, K. C.; Adami, R. C.; Alsante, K. M.; Hong, J.; Landis, M. S.; Lombardo, F.; Roberts, C. J. Stabilization of pharmaceuticals to oxidative degradation. *Pharm. Dev. Technol.* **2002**, *7* (1), 1−32.

(27) Hovorka, S. W.; Schoneich, C. Oxidative Degradation of Pharmaceuticals: Theory, Mechanisms and Inhibition. *J. Pharm. Sci.* **2001**, *90* (3), 253−269.

Fundamentally, DELPHI is a tool for knowledge capture and application to drug degradation problems for new molecules. As a test of the breadth of degradation rules, DELPHI was used against a selection of degradation chemistries described by Baertschi and Alsante[12] and evaluated for its ability to predict the primary degradation chemistry of a range of molecules. This recent review was written after the bulk of DELPHI's rules were generated, and thus this exercise serves as a reasonable benchmark for the extent of DELPHI's knowledge base. Ten molecules were selected to represent a variety of chemistries including testolactone, ampicillin, cimetidine, diazepam, melphalan, loratadine, triamcinolone acetonide, glutethimide, chloramphenicol, and moxalactam. DELPHI successfully predicted the primary reactivity of the ten compounds that comprised the test set (Table 5). On the other hand, the chemistry of cefuroxime is known to include lactone formulation. While lactone formation from the condensation of carboxylic acids and alcohols is included in DELPHI's rules, the intramolecular displacement of the carbamate moiety is not. This is a good example that illustrates the fact that DELPHI is a system that captures knowledge; rules that update the chemical database to include carboxylic acids and carbamates as nucleophiles and leaving groups, respectively, can easily be coded into this learning system.

Conclusion

The ability of DELPHI to predict the degradation of pharmaceuticals has been described. The reliability of these predictions has been determined through long-term use of the program in conjunction with degradant isolation studies. Prediction of pharmaceutical degradation is a challenging problem that truly tests the capabilities of an expert system to emulate the wisdom of an experienced chemist. DELPHI is frequently updated as additional reactivity is discovered or the diversity of chemical space available to the medicinal chemist is expanded.

Acknowledgment. The authors would like to thank several colleagues that have either contributed to the development and application of DELPHI or contributed significantly to DELPHI's knowledge base. We thank Dinos Santafianos, Karen Alsante, Carol Kingsmill, Tim Hurley, and Michael Lovdahl for their contributions to the knowledge content of DELPHI and John Harju, Wilson Rao, John Dela Cruz, and Asad Lateef for their many contributions to the realization of the information technology related to DELPHI.

MP060103+

(28) Lewin, J. L.; Cramer, C. J. Rapid Quantum Mechanical Models for the Computational Estimation of CH Bond Dissociation Energies as a Measure of Metabolic Stability. *Mol. Pharmaceutics* **2004**, *1*, 128−135.