
Jonathan H. Chen and Pierre Baldi*  
Institute for Genomics and Bioinformatics and Department of Computer Science, School of Information and Computer Sciences, University of California, Irvine, Irvine, CA 92697; *pbalidi@ics.uci.edu

Cognitive theory and pedagogical experience have long indicated that practice problems with appropriate performance feedback are vital for students to master problem-solving skills (1). Paper homework assignments have traditionally been the means to challenge student knowledge and understanding, but manually grading paper homework for perhaps hundreds of students in a large university classroom is an unwieldy and labor-intensive task. Even if graded feedback is provided, corrections and useful comments commonly take up to a week after assignment submission, which limits the value of the feedback. Online problems sets and electronic tutorial systems can address many of these issues by offering automated grading capabilities, saving instructor’s time and providing immediate feedback to the student on the correctness of their answers to tighten the feedback loop. In fact, several online learning systems do exist to help address these needs in the chemistry classroom, including: ACEOrganic (2), LON-CAPA (3), MCWeb (4), OWL (5), WE_LEARN (6), and WebAssign (7).

While the existing systems already provide significant benefits over traditional instruction models, they are necessarily limited in scope by the quantity of problems they include. The problem sets available in the above systems for the most part simply transplant textbook problems into an electronic format. In particular, they consist of pre-constructed problems authored by human experts. Fixed problem sets inherently limit the “replay value” of such systems, as there is little reason to return to problems previously completed once the answers have been seen and memorized. A more valuable model would involve a system that can dynamically generate similar but non-identical problems on demand. Note that we are specifically interested in randomly generated problems that conceivably never existed before, as opposed to problems randomly selected from a pre-existing set. At least two of the systems mentioned above, OWL and MCWeb, support some dynamic generation of random problems, but this is only for general chemistry where problems are relatively straightforward to model in terms of simple mathematical formulas. In comparison, organic chemistry deals less with numerical values and algebraic equations and more with chemical structure and reactivity.

Especially for organic chemistry, problems composed for existing learning systems have been relatively constrained to closed-ended designs such as multiple-choice and fill-in-the-blank. This limitation in the variety and flexibility of problems ultimately restricts the creativity of the student, as well as the amount of meaningful feedback the system can give to the student beyond terse “correct–incorrect” responses. For open-ended problems, such as multi-step organic synthesis design, it can be very discouraging when a system prevents a student from proceeding as soon as he deviates even slightly from the intended solution. Particularly when it takes several steps to demonstrate alternative solutions, students should instead be free to experiment with different steps and combinations to find a solution.

Taking the idea of learning through experimentation further is the concept of inquiry-based learning (8). In this model, students gain mastery of a subject when they can move beyond passively absorbing information and instead explore the boundaries of their knowledge by actively asking questions. Having a qualified instructor answer questions and work through problems could be considered the “gold standard” for inquiry-based learning, but the combination of class discussion and instructor’s office hours still suffers scalability and accessibility issues similar to paper homework.

To support randomly generated problems and inquiry-based learning for organic chemistry, a tutorial system must be capable in its core to take an arbitrary reactant–reagent combination and reliably predict the major reaction product. No simple algebraic formula or computational construct can provide such power. The closest systems to embody such predictive power are the CAMEO (9) and EROS (10) systems, but these were built on previous-generation technologies and have largely fallen out of support. Since then, few projects, for example, ROBIA (11) and SOPHIA (12), have approached the computational reaction-prediction problem, and none have been applied towards modern chemical education.

Aims for a New System

Here we describe Synthesis Explorer, an online tutorial system for organic chemistry designed to enable learning in ways previously unrealized within existing models. The system focuses on some of the most important and challenging subjects in organic chemistry, including reaction product prediction, multi-step synthesis design, and reaction mechanism proposal. The system does not rely on a fixed collection of pre-authored problems, but instead relies on a collection of chemical reagent models with inherent predictive power. As a result, the system can dynamically generate and validate new problems at will, allowing students to freely explore novel reaction combinations, and respond to student inquiries related to reaction and mechanism predictions.

System Design and Description

The core features of the system are all related to the reactivity of organic molecules and thus content is organized around a collection of chemical reagent models with the built-in ability to predict the course of chemical reactions. This predictive power derives from an underlying expert system developed with the OEChem toolkit from OpenEye Scientific Software (13) and based on reaction transformation rules written in the SMIRKS language (14).

System content is organized into reaction categories corresponding to chapters from undergraduate organic chemistry textbooks (Table 1). The Bruice (15), Loudon (16), and Smith...
(17) textbooks were used as models for chapter organization. This textbook organization is provided primarily for convenience—reference to a textbook is not necessary in order to use the system. Every “chapter” of material is shared by essentially every undergraduate textbook. The textbooks simply order and filter the complete list of possible chapters. Students and instructors using other textbooks can simply ignore the chapter numbers and instead attend only to the descriptive labels. Since the system can generate problems for chapters in any order and combination, it is flexible enough to fit virtually any textbook and lesson plan.

**Synthesis Design Workspace**

The major functionalities of Synthesis Explorer revolve around multi-step synthesis design problems and begin with the student selecting one or more categories or chapters of content they wish to review through the Web interface. From the selected categories, the system presents a pool of available reagents and starting materials as well as a target synthetic product (Figure 1). The student may then interactively select any combination of these reactants and reagents and the system will predict what (intermediate) products result. Note that this product prediction is calculated by the system dynamically and is not based on pre-coded reaction examples. These intermediate products can then be carried over to further reactions to build increasingly varied and complex molecules. The student’s goal is to reconstruct the target product using these tools.

**Dynamically Generated and Customizable Problems**

With its inherent predictive power, the system has the unique ability to offer randomly generated organic chemistry problems in addition to typical pre-constructed problems. It does so by taking the collection of reactants and reagents that

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Alkenes</td>
</tr>
<tr>
<td>9.04</td>
<td>Substitution Reactions of Alkyl Halides</td>
</tr>
<tr>
<td>9.05</td>
<td>Elimination Reactions of Alkyl Halides</td>
</tr>
<tr>
<td>10</td>
<td>Alcohols and Epoxides</td>
</tr>
<tr>
<td>11.04</td>
<td>Epoxides and Organometallic Compounds</td>
</tr>
<tr>
<td>11.05</td>
<td>Oxidation of Alcohols and Alkenes</td>
</tr>
<tr>
<td>14</td>
<td>Alkynes</td>
</tr>
<tr>
<td>15</td>
<td>Dieno, Conjugation, Diels–Alder</td>
</tr>
<tr>
<td>16</td>
<td>Electrophilic Aromatic Substitution</td>
</tr>
<tr>
<td>17</td>
<td>Allylic and Benzyl Reactivity</td>
</tr>
<tr>
<td>17.02</td>
<td>Alkanes, Radical Reactions</td>
</tr>
<tr>
<td>18</td>
<td>Transition Metal (Pd) Catalysis</td>
</tr>
<tr>
<td>18.04</td>
<td>SnAr and Benzene Reactions</td>
</tr>
<tr>
<td>19</td>
<td>Aldehydes and Ketones</td>
</tr>
<tr>
<td>20.1</td>
<td>Redox of Alcohols and Carboxyls</td>
</tr>
<tr>
<td>21</td>
<td>Carboxylic Acid Derivatives</td>
</tr>
<tr>
<td>22</td>
<td>Enolate Chemistry</td>
</tr>
<tr>
<td>22</td>
<td>Acetoacetic and Malonic Ester Synthesis</td>
</tr>
<tr>
<td>22.04</td>
<td>Aldol Chemistry and Michael Addition</td>
</tr>
<tr>
<td>22.05</td>
<td>Claisen Condensations</td>
</tr>
<tr>
<td>22.08</td>
<td>Organometallic Addition, Conjugate Addition</td>
</tr>
<tr>
<td>23</td>
<td>Amines</td>
</tr>
<tr>
<td>23.1</td>
<td>Aminediazonium Reactions</td>
</tr>
<tr>
<td>24</td>
<td>Naphthalene and Heteroaromatic EAS</td>
</tr>
<tr>
<td>24.05</td>
<td>Pyridine Derivatives</td>
</tr>
<tr>
<td>25</td>
<td>Pericyclic Reactions</td>
</tr>
<tr>
<td>26.04</td>
<td>Amino Acid Synthesis</td>
</tr>
<tr>
<td>26.07</td>
<td>Peptide Synthesis</td>
</tr>
<tr>
<td>27</td>
<td>Carbohydrates</td>
</tr>
</tbody>
</table>

Note that problems can be dynamically generated from these categories in any order and combination, thus the system is not intimately tied to any particular textbook or lesson plan. Please refer to the system Web site for the most current list.

Figure 1. Multi-step Synthesis Explorer screenshot from the chapter on electrophilic aromatic substitution. The student is presented with a target product molecule (top-right) to derive a synthesis pathway for, as well as several control buttons and a context-sensitive help box. Completing the synthesis involves selecting the proper sequence of reactants and reagents from the available, scrollable, pools (top-left and top-middle, respectively). As the student selects different reactant–reagent combinations as possible steps in the synthesis pathway, these combinations are presented in the “Pathway” workspace area (bottom) along with each intermediate product the system predicts. Once the target product itself is borne out of one of these reactions, the system can validate and record that the student was able to solve the problem.
will be available to the student and then applying them in a random sequence until it yields a reasonable target product that the student can retro-synthetically reconstruct using those starting materials. This provides a unique opportunity to tailor problems specifically to fit a student’s needs and interests. In particular, the student can select any combination of chapters and the maximum number of steps of a synthesis problem, that is, the primary measure of its difficulty. The system will generate a customized problem fitting those specifications on demand. Currently the system leaves these customization choices to the student, but conceivably the system could automatically assess the student’s competence and dynamically adapt the content and complexity of the generated problems to this competence in appropriately challenging ways (18, 19).

**Free Form Exploration**

When working with Synthesis Explorer, students are ultimately expected to combine reactants and reagents in a particular sequence towards a target synthesis product, but the system does not constrain the student to this purpose. The system provides the unique opportunity for the student to deviate from the intended “correct” pathway and to instead learn through experimentation. Students are allowed to combine arbitrary reactant and reagent combinations to observe the predicted products and any warning or hint messages that accompany them. These messages are generated to remind students of common pitfalls that are often overlooked such as when carbocation rearrangements or allylic resonance occurs in the course of a reaction. Not only can students *select* any of the available reactants listed by the system, they can even *input* their own novel reactant structures through a chemical sketcher interface (JME Editor, courtesy of Peter Ertl of Novartis). For the kinds of structures that would be found in an undergraduate organic chemistry text, the system can still make reasonable predictions for the reactivity of novel compounds. As a result, common questions a student would ask an instructor can instead be answered by the system providing a basis for inquiry-based learning (e.g., see Figure 2).

**Mechanism Explorer and Reaction Details**

When the reaction product and any accompanying cautionary messages are insufficient to help a student understand a reaction, they can simply view the system’s expected solution for the synthesis or, even better, they can follow a link to the Mechanism Explorer interface (Figure 3). The Mechanism Explorer interface is a powerful new addition to the system that allows students to interactively work through the proposal of a complete curved-arrow mechanism diagram for any reaction, taking advantage of the capabilities of the MarvinSketch applet (20). Full mechanism solutions are available for most reactions in the form of a dynamically generated reaction detail information page (Figure 4). These pages expose the internal logic used by the system to predict the course of a reaction, enabling the student to ask not only what the final product of a reaction is, but also how the reaction proceeded. Detailed mechanisms are depicted via arrow-pushing diagrams showing the elementary sub-steps and intermediate products of the overall reaction. As with the synthesis problems, these diagrams are not based on a pre-constructed set of mechanism diagrams. It is rather the underlying expert system that can predict reaction mechanisms on demand for arbitrary reactant–reagent combinations.
Classroom Trials

Subject Group Description

To assess the efficacy of Synthesis Explorer, it has been made available to several undergraduate classes taught by the chemistry department at the University of California, Irvine. These classes, each taken by hundreds of students, represent the second of a three quarter undergraduate organic chemistry curriculum where students learn and use many chemical reactions and synthesis strategies. Subjects covered include reactions of alkanes (radical chemistry), alkyl halides, alcohols, epoxides, organometallics, benzene derivatives, and carboxylic acid derivatives. Students were offered an incentive of a ~2% boost to their final grade for completing multi-step synthesis problems generated by the system (~5 per week). None of these classes included written homework assignments, though non-graded paper problem sets were available for students to practice.

Performance Results

A typical final score histogram for one of these classes, separated into 2 groups is shown in Figure 5. Students who completed at least 45 problems in Synthesis Explorer by the end of the course were classified as “participants” and offered the extra credit. All other students were classified as “non-participants.” The chart illustrates that the participants score distribution is distinctly up-shifted from the non-participants, in this case by 10.4% of their final grade (this is before the additional 2% of extra credit was offered). A similar pattern was observed for every class, with the score improvement ranging from 7% to 14%.

The above results offer a qualitative indication that the system can indeed improve student learning and performance. We realize that a causal relationship cannot be strictly established, in particular because in this experiment use of the system was voluntary and student motivation could be a confounding factor in the analysis. Other study designs were considered, such as comparing average scores against previous classes where the system was not available, or normalizing results against student scores in past courses. Unfortunately, all these designs can also be subject to criticism for some confounding factor, such as courses being taught by different professors, or examinations having different degrees of difficulty. The best experiment one could imagine would be to randomly select half of the students from a class and require them to use the system, while the remaining half is forbidden from using the system as a control group. However, this has the potential for creating an unfair learning environment for the control group since the control students are denied access to a resource that their peers can use. Ultimately, we felt that a fair learning environment where the system was open and optional to all students was more important than having a fully randomized control group for this analysis.

Summary and Discussion

The Synthesis Explorer system specializes in challenging problems of reactivity in organic chemistry, particularly reaction product prediction, mechanism prediction, and multi-step synthesis design. From our experience, the system’s support of a mix of random and pre-constructed problems works best with random problems for extensive practice and pre-constructed problems for targeted testing and evaluation. The system is built on an underlying expert system whose inherent predictive power enables a richer learning experience through experimentation and interactive dialogue with customized feedback in the form of both text and predicted chemical structures.

The majority of reagents covered in a second-year organic chemistry curriculum are already modeled within the system, and new reagents and reaction mechanisms are being added periodically. Mechanism prediction problems have recently been added alongside the existing synthesis design problems. Other features that could be pursued in the future include the capability to automatically assess the student’s abilities and dynamically tailor the problems to fit the student’s particular strengths and weaknesses.

The expert system upon which Synthesis Explorer is built has been presented in an educational setting, to facilitate the learning of chemistry in ways previously unrealized. But the same expert system can be used in other applications in chemical informatics and modeling. For example, computerized retrosynthesis decision support systems (21, 22) could be based on the same technology. In fact, the underlying expert system is already being used to help solve the very kinds of synthesis problems generated by Synthesis Explorer (23). As the content and robustness of chemical expert systems expand, these will become useful not only to undergraduate students, but also to professional chemists.

Distribution and Access

Presently, the system is hosted on computer servers at UC Irvine and is freely accessible via the Internet to any user or institution. Students may work anonymously on the system or “login” with a unique identifier (e.g., student ID number) to let the system record their progress automatically, while instructors may review the records of the students for whom they have the identifier. Walkthroughs are available on the Web site to guide users step-by-step through the basic functionalities. For more advanced instructor features, such as the ability to assign specific (vs random) problems with enforced due dates, instructors should contact the developers listed on the Web site to arrange an appropriate collaboration. Possible partnerships with com-
panies, such as textbook publishers, or other organizations are being explored to facilitate a more widespread distribution of the system and its integration into more formalized assignment and assessment programs. The latest information on how to access the system will always be available via the Web site and the respective help page (24).

Acknowledgments

Work supported by an NIH Biomedical Informatics Training grant (LM-07443-01) and NSF grants EIA-0321390 and 0513376 to PB. We acknowledge OpenEye Scientific Software, Peter Erlt of Novartis (JME Editor), and ChemAxon for academic software licenses. We thank Suzanne Blum, David Van Vranken, Zhibin Guan, Elizabeth Jarvo, Susan King, Larry Overman, Mare Taagepera, Chris Vanderwal, and Gregory Weiss, who taught the undergraduate chemistry classes, and all the participating students for their feedback. We acknowledge Peter Phung and Paul Rigor for contributing to software design and development. We thank James Nowick, Scott Rychnovsky, and Kenneth Shea for additional feedback and comments.

Literature Cited


Supporting JCE Online Material

Abstract and keywords

Full text (PDF)

Links to cited URLs and JCE articles

Supplement
Description of the reagent models, single-step reaction drills, and structure details
User feedback and survey statistics
Implementation details