

XXX

Software for Teaching Structure-Hydrophobicity Relationships

Received for publication, August 4, 2004

Brian White[‡], Ilana Perna, and Reino Carlsson^d*From the Biology Department, University of Massachusetts, Boston, Boston, Massachusetts 02125*

We have developed a lecture and laboratory curriculum for introducing beginning undergraduate Biology students to chemical structure-function relationships. The laboratory portion of this curriculum employs cheminformatics software that provides instant feedback to help students develop an understanding of the relationship between structure and hydrophobicity. To evaluate the effectiveness of this curriculum, we measured students' understanding using an open-ended problem-based survey. Student responses to this survey improved significantly following the activities we describe, suggesting that they are effective teaching tools. This curriculum also provides a foundation for students' future structure-function studies in chemistry, biochemistry, and molecular biology.

Keywords: Laboratory exercises, cheminformatics, hydrophobicity, molecular properties.

Chemical structure-function relationships are a major topic in many biology and chemistry courses beginning in high school and continuing through graduate school. Their goal is for students to learn how to “read chemical structures”: to be able to predict or explain a molecule's chemical properties in terms of its structural elements. In General Biology I at UMass Boston, the first-semester introductory biology course for Biology majors, our goals for this part of the course are relatively modest. In order to understand modern molecular biology, students must have a basic understanding of protein structure and function. This requires that students be able to predict the noncovalent bonding capabilities and relative hydrophobicities of simple biological molecules based on their structures. In addition to explaining the behavior of proteins and small molecules, learning how to “read chemical structures” in this way serves as a foundation for the more sophisticated structure-function analyses they will be expected to carry out in organic chemistry, biochemistry, and molecular biology courses.

In General Biology I, the four key concepts that comprise this basic understanding of structure-hydrophobicity relationships are:

- Concept 1: Polar (-OH, -NH₂, -C = O, etc.) or charged (-O⁻ or -N⁺) groups are hydrophilic.
- Concept 2: Charged groups are more hydrophilic than uncharged groups.
- Concept 3: Non-polar groups (-CH₂-, -CH₃, -SH, etc.) are hydrophobic.
- Concept 4: On a per-atom basis, hydrophilic groups contribute more to the hydrophilicity of a molecule than hydrophobic groups contribute to its hydrophobicity.

Studies of students' learning have highlighted some of the challenges involved in teaching this type of material and have identified successful teaching strategies. Interpreting chemical structures requires an understanding of chemical bonding as well as facility with the different representations of molecular structure used by chemists. Previous studies have highlighted the many difficulties students have with these activities; these include misconceptions about bonding, inability to translate between different molecular representations, and confusion about intra- versus intermolecular forces [1–3]. Exercises involving the generation and exploration of different chemical representations have been shown to be effective in overcoming many of these [1, 3]. An important component of these exercises is rapid feedback, which has been shown to increase the effectiveness of instruction in general [4] as well as instruction in science problem solving in particular [5]. We have combined these two approaches into a laboratory exercise that uses cheminformatics software to help students learn how to interpret chemical structures in terms of their relative hydrophobicities. Our preliminary evaluation, using a problem-based survey, shows significant learning outcomes from this curriculum and suggest that it is a productive way to communicate a basic understanding of structure-hydrophobicity relationships.

THE CURRICULUM

General Biology I at UMass Boston is the first-semester introductory course for Biology majors; it covers genetics, chemistry, biochemistry, molecular biology, and cancer. The students attend three 50-min lectures and one 3-h laboratory per week. There are ~200 students in General Biology I each semester. All students attend the same lecture section; there are 10 laboratory sections taught by

[‡] To whom correspondence should be addressed: Biology Department, University of Massachusetts, Boston, 100 Morrissey Boulevard, Boston, MA 02125. E-mail: brian.white@umb.edu.

Fn1

graduate teaching assistants (TAs).¹ Many of these laboratory sessions involve problem-solving or computer-simulation exercises [6–8]. The course instructor (B.W.) gives all the lectures, writes the laboratory manual, and provides the TAs with detailed lesson plans for each laboratory session.

The chemistry section of General Biology I consists of five lectures on atomic structure, covalent bonds, molecular structures, and noncovalent interactions. These concepts are reinforced by a set of ungraded practice problems and two 3-h laboratory sessions. All relevant materials are available for download (intro.bio.umb.edu/downloads/). In the first chemistry laboratory, the Chemical Structure Lab, students explore covalent bonding and several different molecular representations (paper and pencil, molecular visualization, and physical models). In the second chemistry laboratory, the Chemical Properties Lab, students explore structure-hydrophobicity relationships of molecules using cheminformatics software incorporated into a web page (www.securebio.umb.edu/daycgi/JMEclogp.pl). This software predicts the hydrophobicity of molecules designed by the students. The hydrophobicity of each molecule is expressed in terms of the logarithm of its octanol:water partition ratio: $\log P_{O/W}$ or simply $\log P$. The relationship between hydrophobicity and $\log P$ is straightforward: hydrophobic molecules will partition predominantly into the octanol phase, resulting in positive $\log P$ values, while hydrophilic molecules will partition predominantly in the water phase, resulting in negative $\log P$ values. The software analyzes each molecule and predicts its $\log P$ based on the particular features of its structure. The web page interface we have developed provides the students with the opportunity to use this tool to explore structure-hydrophobicity relationships with structures they have designed. In this way, each student can follow her own path to a clearer understanding of this material.

The Chemical Properties Lab begins with a short tutorial on how to use the structure drawing program and the $\log P$ calculator. Students then work in groups of three to solve a series of problems using the software. These ask the students to:

1. Look at structures, predict their relative hydrophobicities, and then check these predictions using the $\log P$ calculator.
2. Design structures with particular $\log P$ values. The first of these are simple: “Make a molecule more hydrophobic than molecule X”; later ones are more challenging: “Make a molecule with a $\log P$ value in between that of molecules X and Y.”
3. Design molecules to explore structure/hydrophobicity relationships. In these problems, students are asked to create a series of related molecules (R-CH₃, R-OH, R-SH, and R-NH₂; then the isomeric ether *versus* corresponding alcohol) and calculate their differing $\log P$ values. They are then asked to explain the relative $\log P$ values of these related

molecules in terms of the properties of the relevant atoms and functional groups.

4. Add a hydrophilic group, first uncharged (-OH or -NH₂), then charged (-O⁻ or -NH₃⁺), to a “core” molecule of their choice and then determine the number of hydrophobic groups (-CH₂- or -CH₃) they must add to the core in order to balance each hydrophilic group’s contribution.

As they work, the students receive immediate feedback from the software. This allows them to continually assess and revise their understandings of the relationship between molecular structure and chemical properties. During this process, students design their own molecules to answer the questions we have provided. The freedom afforded by the software allows each student to construct her own understanding using structures that she finds meaningful.

Following the computer exercises, the students participate in a brief interactive “wet” hydrophobicity demonstration as a practical example of what they have learned so far. Students are given the structures of four dyes: azulene, β -carotene, fast green FCF, and cresyl violet. They are then asked to predict, based on the structures and their understanding of hydrophobicity, where each dye will partition in a test tube of hexane and water. Two of the dyes have conspicuously hydrophobic structures and behavior; the other two are conspicuously hydrophilic. When each dye is added to a test tube containing hexane and water, its behavior visually confirms the students’ interpretation of its structure. As a final summary, the TA leads students in a discussion of the structure and noncovalent bonding properties of two other sample molecules.

AQ: A

EVALUATION OF LEARNING OUTCOMES

At the most basic level this curriculum is successful, as shown by the high scores students earn on their work in this laboratory: out of a possible 30 points, the average was 25.3 ($n = 205$). In addition, the server logs from groups performing this laboratory showed that most groups created a series of structures that followed the intended course of the laboratory (data not shown). Interestingly, these logs show that different groups of students solved the problems using different strategies. Table I shows one example of this; it presents the time sequence of structures devised by two different groups as they solved the first part of Problem 4.


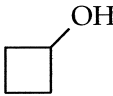

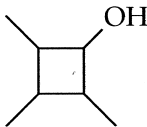
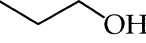
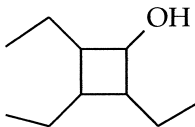
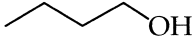
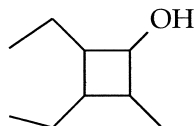
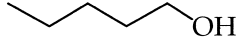
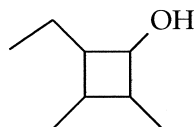
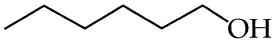
T1

The sequences in Table I show that, although both groups followed different paths, they both solved the problem correctly. Problem 4 asked the students to add a hydrophilic group (in this case, -OH) to a core structure and then determine how many hydrophobic carbons would be needed to balance the -OH’s contribution and restore the original $\log P$ value. Group 1 started with cyclobutane. They added an -OH and then found that three carbons were not sufficient, so they added three more, overshooting the desired $\log P$ value. They then quickly removed the extra carbons one-by-one until they had the desired $\log P$. Group 2 began with a linear structure, ethane, and worked more slowly, adding carbons one-by-one until they had the desired $\log P$. Although the two groups followed different paths, they found the same answer:

¹ The abbreviations used are: TA, teaching assistant; JME, Java Molecular Editor.

TABLE I



Group 1			Group 2		
Time (m:s)	Structure	logP (comment)	Time (m:s)	Structure	logP (comment)
0:00		2.236 (core)	0:00	$\text{H}_3\text{C}-\text{CH}_3$	1.752 (core)
0:39		0.149 (-OH added)	2:08		-0.235 (-OH added)
1:27		1.706 (+3 C's; not enough C's added)	4:59		0.294 (+1 C; not enough)
1:44		3.293 (+6 C's; too many C's added)	5:31		0.823 (+2 C's; not enough)
2:11		2.764 (+5 C's; still too many)	6:05		1.352 (+3 C's; not enough)
2:29		2.235 (+4 C's; OK)	6:35		1.881 (+4 C's; OK)

approximately four carbons are required to overcome the hydrophilic contribution of one -OH group (Concept 4). Other groups showed a variety of different strategies for solving each of the problems in the laboratory.

In the Fall of 2003, each of the 205 students in General Biology I completed an open-ended survey that compared their understanding of structure-hydrophobicity relationships before and after instruction; the survey questions are shown in Appendix A. This understanding was measured both in terms of their ability to solve a structure-hydrophobicity problem correctly as well as their use of appropriate chemical terminology. Briefly, in Part A, students were given two related uncharged molecules that had not been explored in lecture or laboratory and asked to circle the more hydrophobic of the two. Part B asked them to draw a molecule with a logP value in between those of the two molecules given in Part A. Part C asked them to explain, in a sentence or two, why the molecule in Part B had an intermediate logP value. Responses to all three parts were scored as correct or incorrect. In addition, responses to Part C were categorized by the type(s) of terminology used in their explanation, whether the terminology was used correctly or not. The types of terminology included:

- Blank: no response at all
- Not Relevant: an irrelevant response
- Atoms: mentioned specific atoms by element name ("oxygen", "C atom," etc.)
- Charge: mentioned charge ("+", "-", "ion," "charge")
- H-bonds: mentioned hydrogen bonds

All surveys were scored independently by two investigators; inter-rater reliability was determined using Cohen's Kappa [9]. In all cases, Kappa was greater than the usually accepted value of 0.7.

The survey was administered using a modified pre/post instruction protocol that we have used in other studies [8]. Students (169 total: 82% of the 205 students in the class) returned the pre-instruction survey immediately after the first three chemistry lectures; these lectures covered atomic structure and covalent bonding. Following lectures on noncovalent bonding, roughly half of the students ($n = 80$) completed the post-survey at the beginning of the Chemical Properties Lab. This group had only been exposed to the lectures; it is called the "Lecture Only" group. The other half of the class ($n = 89$) completed the post-survey at the end of the Chemical Properties Lab. This group had experienced both lecture and laboratory; it is called the "Lecture and Lab" group. Individual laboratory sections were randomly assigned to either group such that each TA taught one "Lecture Only" and one "Lecture and Lab" section. This protocol is diagrammed in Fig. 1.

Fig. 2 shows the results of scoring the surveys of all four groups; the bars in this figure are shaded to match the corresponding bars in Fig. 1. Our survey administration protocol allows several relevant comparisons between these groups. First, because the two pre-instruction groups ("Pre-Lecture" and "Pre-Lecture and Lab") are random samples of the student population taken at the same time, their scores should not differ

F1

F2

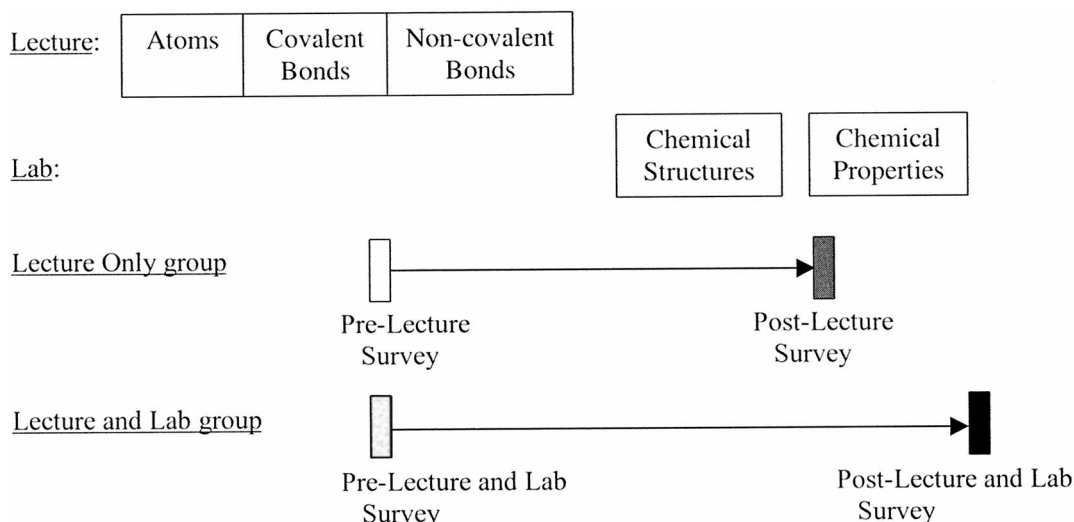


FIG. 1. Survey administration timeline.

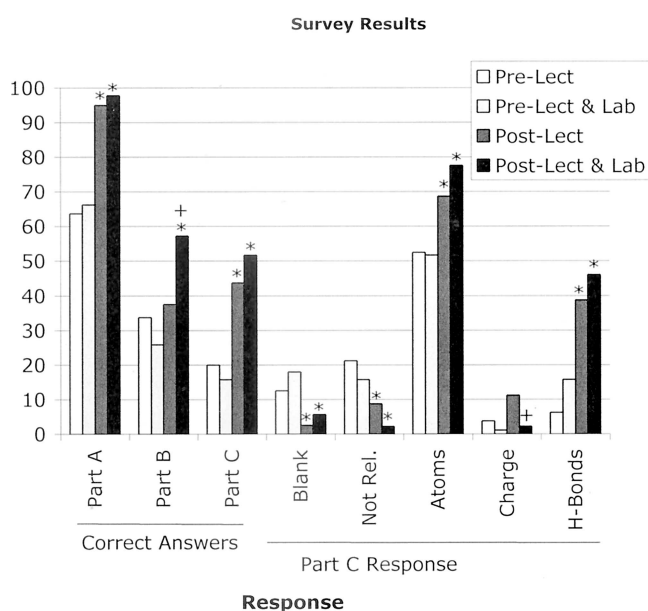


FIG. 2. Results of survey scoring. An asterisk indicates significant ($p < 0.05$) difference from corresponding pre-instruction group; “+” indicates significant ($p < 0.05$) difference from other post-instruction group.

significantly. This was found to be true, using a χ^2 test, for all categories shown.

Differences between the “Pre-Lecture” and “Post-Lecture” groups show the effect of the lectures by themselves (McNemar’s test for significance; $p < 0.05$ indicated by an asterisk). Based on this, the percentage of students who were able to recognize the more hydrophobic of the two structures (Part A) and create correct explanations for the structures they generated in Part B increased following the lecture presentation. In terms of response categories, “Blank” and “Not Relevant” responses to Part C dropped significantly and mention of particular atoms in the molecule and hydrogen bonds increased significantly. The laboratory activities has no further significant effect on these response categories.

The differences between the “Post-Lecture” and “Post-

Lecture and Lab” groups reflect the incremental effect of laboratory following the lectures (using a χ^2 test for significance; $p < 0.05$ indicated by a “+”). Following the lectures, some students mentioned charge (Concept 2) in their answers, even though it is not relevant to this problem. The fraction of students giving this inappropriate explanation decreased significantly following the laboratory. Most significantly, students’ ability to draw chemically correct structures of intermediate hydrophobicity (Part B) increased only after students had performed the laboratory. These results suggest that, while the lectures provided the terminology, the laboratory was necessary for students to put these concepts into practice. These results are consistent with the laboratory’s emphasis on drawing structures and exploring their logP values and indicate that it provides an essential component of the curriculum. Because the students spend most of their laboratory time working through the computer exercises, it is likely that these make an important contribution to their learning.

CHEMINFORMATICS SOFTWARE FOR PREDICTING HYDROPHOBICITY

The web page used in the Chemical Properties Lab (www.securebio.umb.edu/daycgi/JMEclogp.pl) is based on two separate pieces of software. The first is the Java Molecular Editor (JME), a Java applet that allows the user to draw chemical structures; it is available at no charge from www.molinspiration.com/jme/getjme.html. Fig. 3 shows JME with a structure drawn. JME provides the user with several template structures and a variety of editing features indicated by the top row of buttons. It allows the user to generate structures that follow Lewis’ structure rules for covalent bonding and charge but does not prevent drawing unstable or chemically improbable structures. Once the student has drawn a structure, she clicks a button that submits the structure to the second piece of software, ClogP, that runs on our web server. ClogP has been developed for the prediction of $\log P_{OW}$ [10–12] and is used extensively in industry and academia. ClogP ana-

lyzes the structure by considering it as a set of chemical fragments, each of which makes a well-characterized contribution to the overall logP of the molecule. ClogP also includes several compensation factors to deal with more global structural features. It returns an approximate calculated logP value along with some information about how the value was calculated. Sample ClogP output for the structure from Fig. 3 is shown in Fig. 4; the number at the lower-right corner of the display indicates a logP of 1.414 for *p*-amino-toluene. ClogP is available for a yearly license fee; access to our ClogP site (www.securebio.umb.edu/daycgi/JMEclogp.pl) is free to all users. This site requires a browser that supports both java and javascript; it is compatible with Netscape 4.78 and above, Internet Explorer, and Safari. Although our server is capable of handling the needs of small classes in addition to those of our students, for large groups it may be necessary to set up an on-site hydrophobicity calculator. This calculator could use ClogP or one of the alternative cheminformatics programs described below.

There are two freely available logP calculators that could be used in a laboratory as we have described as lower-cost alternatives to ClogP. Both of these would require some setup for use in a teaching laboratory. The Environmental Protection Agency offers EPISUITE for free download (www.epa.gov/opptintr/exposure/docs/episuite.htm).

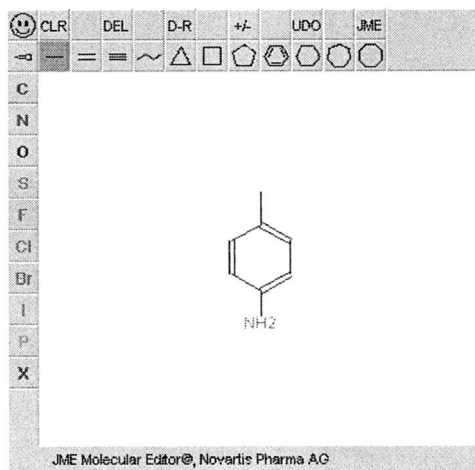


FIG. 3. JME user interface.

```
logPstar found: 1.39
-----
| SMILES: Cc1ccc(N)cc1 |
| ATOM #: 12.345.6.78. |
| ISOC-ID: Za.aaa...aa. |
| FRAG-ID: .....1.... |
| H-COUNT: 3..11..2.11. |
| RING 1: .a.aaa...aa. |
-----
| Class | Type | Log(P) Contribution Description | Comment | Value |
-----+-----+-----+-----+-----+
| Fragment | # 1 | Primary Amine [a] | Measured | -1.000 |
| Carbon | | 1 aliphatic isolating carbon | | 0.195 |
| Carbon | | 6 aromatic isolating carbons | | 0.780 |
| ExFragment|Hydrog| 7 hydrogens on isolating carbons | | 1.589 |
| Benzylbond|Simple| 1 benzyl bond to simple aromatic | | -0.150 |
-----+-----+-----+-----+
| RESULT | DB=17 | All fragments measured | CLOGP=4.01 | 1.414 |
-----+-----+-----+-----+-----+-----
```

FIG. 4. Sample ClogP output.

EPISUITE runs on Windows 98 and above and includes the KowWin program for calculating logP. KowWin uses an algorithm like that used by ClogP [13] and requires input in SMILES format. SMILES are text strings that express the structure of simple molecules [14]. JME can output molecules in SMILES format by clicking on the smile-face in the upper-left corner of its window; the resulting SMILES can then be copied and pasted into KowWin for analysis. EPISUITE also contains several useful programs for predicting other molecular properties of environmental interest. Another program, xlogp, calculates logP values using an algorithm that characterizes each atom in a given molecule into one of 90 different types based on the element, its hybridization, and its neighboring atoms. The individual atom values are then summed and combined with several more global correction factors to produce the overall logP value [15]. XlogP requires that molecules be submitted in SYBYL Mol2 format; the free-ware program OpenBabel (openbabel.sourceforge.net) can convert virtually all chemical file formats, including those produced by JME, into Mol2 files. The source code for xlogp is available for free download (mdl.ipc.pku.edu.cn/drug_design/work/xlogp.html) and can be compiled for many platforms.

Unfortunately, none of these programs, including ClogP, are capable of calculating the logP of charged structures (Concept 2). This is because the logP of charged structures is more complex to calculate in general and is highly dependent on pH [16, 17]. For exercises where Concept 2 is not important, ClogP, xlogP, and KowWin will be useful, as we have shown. In situations where it is necessary to calculate the correct logP values for charged molecules, one particularly outstanding package, Marvin (www.chemaxon.com/marvin/) is capable of calculating many properties of charged and uncharged molecules including logP, pK_a , polarizability, and refractivity. Marvin is available for a yearly license fee.

Because Concept 2 is an important part of our curricular goals, we have developed jlogP, a java applet that can calculate the approximate logP of both charged and uncharged molecules. JlogP sacrifices some of the accuracy of the commercial software programs while retaining their pedagogically relevant properties. The algorithm used by jlogP is based on the atom-additive method used in xlogp

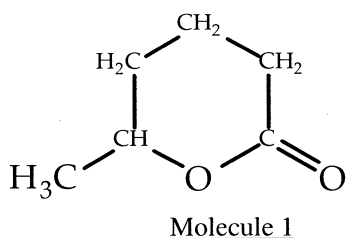
[15]. In addition to using most of the atom types from xlogp, jlogP includes several charged atom types that are absent from xlogp. The logP contributions of the charged atom types were estimated based on published values [10, 17, 18]. To simplify software development, jlogP does not implement any of the global correction factors used by xlogp and employs an extremely simplified charge algorithm that does not account for pK_a and pH. Although this means that the absolute logP values generated by jlogP do not necessarily correspond to actual partition coefficients, the relative logP values of different molecules reflect Concepts 1 through 4. Because all the laboratory exercises involve comparisons of logP values, this level of accuracy is sufficient for exercises like the Chemical Properties Lab. We have developed a web page (intro.bio.umb.edu/111-112/OLLM/111F98/jlogp/test.html) that uses JME and jlogP that is virtually identical to our ClogP-based web page. Although we have not formally evaluated jlogP with our students, because jlogP uses JME for drawing and produces logP output like that of ClogP, it is likely that jlogP will be similarly effective. Moreover, because jlogP can calculate the logP of charged molecules, we will be using jlogP for the Chemical Properties Lab in the future. JlogP is part of the General Biology I website and runs entirely on the client machine; no server-based hydrophobic calculator is required. It is compatible with Win98 and above as well as MacOSX. JlogP, and its source code is available for free download under the GPL license and thus can be modified or improved as desired.

CONCLUSION

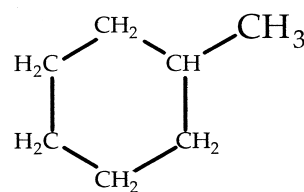
We have developed a lecture and laboratory curriculum that is effective at communicating core concepts of small molecule structure-hydrophobicity relationships to introductory-level college biology students. This curriculum allows exploration of these relationships in a student-centered manner. We have also developed jlogP, a freely available open-source java applet that calculates approximate logP values. The lecture notes, laboratory manual, software, and source code are all available for free download (intro.bio.umb.edu/downloads/). Our success with this method suggests that similar applications of cheminformatics may facilitate learning of other structure-function relationships in chemistry and biochemistry.

APPENDIX A: SURVEY INSTRUMENT

1. Consider the following two molecules:



Molecule 2



Molecule 3

- A) Circle the molecule above that is more hydrophilic (more able to dissolve in water).
- B) In the space between the two molecules above, draw a molecule that has intermediate hydrophilicity—that is, its hydrophilic character is in between that of Molecule 1 and Molecule 3. You need not show all H-atoms.
- C) Explain why the molecule you drew in Part B (Molecule 2) has a hydrophilicity between that of Molecule 1 and Molecule 3.

REFERENCES

- [1] A. G. Harrison, D. F. Treagust (2000) Learning about atoms, molecules, and chemical bonds: A case study of multiple-model use in grade 11 chemistry, *Sci. Educ.* **84**, 352–381.
- [2] R. B. Kozma (2003) The material features of multiple representations and their cognitive and social affordances for science understanding, *Learning and Instruction* **13**, 205–226.
- [3] H.-K. Wu, J. S. Krajcik, E. Soloway (2001) Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom, *J. Res. Sci. Teaching* **38**, 821–842.
- [4] B. F. Skinner (1968) *The Technology of Teaching*. Appleton Century Crofts, New York, NY.
- [5] R. Taconis, M. Ferguson-Hessler, H. Broekkamp (2001) Teaching science problem solving: An overview of experimental work, *J. Res. Sci. Teaching* **38**, 442–468.
- [6] B. White (1999) The red and white yeast lab: An introduction to science as a process, *Am. Biol. Teacher* **61**, 407–410.
- [7] B. White (1998) A curriculum for recitation sections in introductory biology, *J. College Sci. Teaching* **27**, 407–410.
- [8] B. White, S. Kim, K. Sherman, N. Weber (2002) Evaluation of molecular visualization software for teaching protein structure: Differing outcomes from lecture and lab, *Biochem. Mol. Biol. Educ.* **30**, 130–136.
- [9] J. Cohen (1960) A coefficient of reliability for nominal scales, *Educ. Psychol. Measurement* **20**, 37–46.
- [10] C. Hansch, A. Leo (1979) *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley Interscience, New York NY.
- [11] J. Chou, P. Jurs (1979) Computer-assisted computation of partition coefficients from molecular structures using fragment constants, *J. Chem. Info. Comp. Sci.* **19**, 172–178.
- [12] A. Leo, D. Hoekman (2000) Calculating log P(oct) with no missing fragments: The problem of estimating new interaction parameters, *Perspect. Drug Discov. Design* **18**, 19–38.
- [13] W. M. Meylan, P. H. Howard (1995) Atom/fragment contribution method for estimating octanol-water partition coefficients, *J. Pharm. Sci.* **84**, 83–92.
- [14] Weininger, D. (1988) SMILES 1: Introduction and encoding rules, *J. Chem. Info. Comp. Sci.* **28**, 31.
- [15] R. Wang, Y. Gao, L. Lai (2000) Calculating partition coefficient by atom-additive method, *Perspect. Drug Discov. Design* **19**, 47–66.
- [16] V. N. Viswanadhan, et al. (1989) Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, *J. Chem. Info. Comp. Sci.* **29**, 163–172.
- [17] J. Szegedi, F. Csizmadia (2004) Prediction of distribution coefficient using microconstants, in American Chemical Society National Meeting.
- [18] F. Csizmadia, et al. (1997) Prediction of distribution coefficient from structure. 1. Estimation method, *J. Pharm. Sci.* **86**, 865–871.

AQ: B

AQ: C
AQ: D

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

1

A—Please define FCF.

B—Please provide all authors for Ref. 16.

C—Is Ref. 17 an abstract presented at this meeting. Is it printed in a Proceedings of this meeting?
Please provide details of the publication of this reference.

D—Please provide all authors for Ref. 18.
