CBE—Life Sciences Education
Vol. 6, 65–73, Spring 2007

Article

Exploring DNA Structure with Cn3D

Sandra G. Porter,* Joseph Day,† Richard E. McCarty,‡ Allen Shearn,‡ Richard Shingles,‡§ Linnea Fletcher,‖ Stephanie Murphy,¶ and Rebecca Pearlman‡

*Geospiza, Inc., Seattle, WA 98107; †Northwest Lipid Metabolism and Diabetes Research Labs, University of Washington, Seattle, WA 98109; ‡Department of Biology and §Center for Educational Resources, The Johns Hopkins University, Baltimore, MD 21218; ‖Austin Community College, Austin, TX 78702; and ¶Bellevue Community College, North Campus, W225, Bellevue, WA 98007

Submitted March 22, 2006; Revised October 4, 2006; Accepted October 5, 2006
Monitoring Editor: A. Malcolm Campbell

Researchers in the field of bioinformatics have developed a number of analytical programs and databases that are increasingly important for advancing biological research. Because bioinformatics programs are used to analyze, visualize, and/or compare biological data, it is likely that the use of these programs will have a positive impact on biology education. Over the past years, we have been working to help biology instructors introduce bioinformatics activities into their curricula by providing them with instructional materials that use bioinformatics programs and databases as educational tools. In this study, we measured the impact of a set of these materials on student learning. The activities in these materials asked students to use the molecular structure visualization program Cn3D to locate, identify, or analyze diverse features in DNA structures. Both the experimental groups of college and high school students showed significant increases in learning relative to control groups. Further, learning gains by the college students were correlated with the number of activities assigned. We conclude that working with Cn3D was important for improving student understanding of DNA structure. This study is one example of how a bioinformatics program for visualization can be used to support student learning.

INTRODUCTION

The human genome project and the concomitant development of high-throughput technologies such as DNA sequencing, microarray analysis, mass spectrometry, various types of genotyping, and new tools for rapidly solving molecular structures have generated unprecedented quantities of biological data. This phenomenon has implications for both biology students and future biologists. First, new career pathways have developed because of the need to address the multiple challenges of collecting, storing, organizing, and analyzing large volumes of data, let alone interpreting the results. Second, the accumulation of data in national repositories makes it likely that some biological questions can be answered by searching through databases to find the relevant information and by working with the retrieved data. Third, the existence and availability of these data make it possible to design activities around selected data sets and support diverse learning goals.

Schools have responded to these challenges by developing education programs in computational biology and bioinformatics and by adding bioinformatics activities to existing biology courses (Altman, 1998; Bednarski *et al.*, 2005; Honts, 2003; Pevzner, 2004; Porter and Smith, 2000). Some bioinformatics degree programs emphasize computer science and algorithm development, with a course or two in molecular biology. Others focus on understanding the mathematics and algorithms inside of bioinformatics applications. Both kinds of educational programs treat bioinformatics as a separate and autonomous field that requires proficiency in both programming and biology. Some, however, have questioned whether mastery of both biology and computer science is possible, let alone worthwhile (Neeper, 2005).

Geospiza's education program began from a different perspective. We reasoned that students could benefit, in a gen-

DOI: 10.1187/cbe.06–03–0155
Address correspondence to: Sandra Porter (sandy@geospiza.com).
Potential conflict of interest: S.G.P. is also the author of the *Exploring DNA Structure* CD-ROM described in the article.

eral way, from activities that combine the goal of teaching bioinformatics skills with the goal of learning biology. Meeting these goals could help students better understand fundamental concepts in molecular biology, while developing the skills for modern research and familiarity with bioinformatics resources. With these thoughts in mind, we created several, freely available, instructional activities that use this approach (http://www.geospiza.com/education/materials.html).

In this study, we examined the effectiveness of one set of instructional materials that we have since packaged as a commercial product. The activities contained in the materials guide students in using Cn3D (Wang *et al.*, 2000), a program designed for visualizing and working with molecular structures, to learn about DNA. Through these activities, students view and interact with different portions of DNA structures in diverse ways. They find and identify the major and minor grooves, compare the relative orientation of the two strands, locate the components of the nucleotides, and examine the sugar–phosphate backbone. Students also compare covalent bonds with hydrogen bonds, identify bases that form complementary pairs, count the number of hydrogen bonds in each type of pair, and look at a variety of double-stranded structures to explore the positions of the two strands relative to each other.

We chose to work with the Cn3D program, in these activities, for a number of reasons. From a practical standpoint, the program is easy to use, freely available, and runs on many different versions of common computer platforms (Windows [Microsoft, Redmond, WA], Mac OS X [Apple, Cupertino, CA], and UNIX [The Open Group, San Francisco, CA]). Further, because Cn3D is maintained by the National Center for Biotechnology Information (NCBI) and is widely used by biomedical researchers, the National Institutes of Health funds ongoing support and development for Cn3D, which makes Cn3D a stable resource. From a pedagogical standpoint, we hypothesized that having students use Cn3D to visualize elements of DNA structure would improve learning by making concepts about structure and function less abstract. This idea is supported by other studies that have found links between student learning and visualization (Stith, 2004; Henry, 2004; Linn *et al.*, 2006; Noland and Juhn, 2001)

MATERIALS AND METHODS

Cn3D is an open-source program that can be freely downloaded from the NCBI website (http://www.ncbi.nih.gov/Structure). DNA structures from both NMR and x-ray crystallography experiments were obtained from the NCBI's Molecular Modeling Database (http://www.ncbi.nlm.nih.gov/Structure). The following structures were used in this study: 1AIO, 1G4Q, 1K5F, 1MXJ, 1QKG, 1CX3, 1HX4, 1K8P, 1N1N, 365D, 1DCW, 1ILC, 1K9L, 1NAB, 384D, 1EJS, 1JPQ, 1KFI, 1NGU, 452D, 1FHZ, 1JUC, 1MNV, 1NZM, 459D, 1FJ5, 1JUO, 1MP7, 1OVF, 1FYY, 1K2Z, 1MV7, 1P20, 127D, 1FJ5, 1K61, 1R4R, 365D, 1C2X, 1NAJ, 1ER1, 1NXR, 1GX0, 1QTQ, 1IZH, 1LAI, 1MSW, 1AWC, 1FMS, 1LEJ, 1VZK, 3CRO, 1BOS, 1FTD, 1LO1, 1WOU, 459D, 1CIT, 1HCR, 1M6F, 261D, 6BNA, 1MK6, 298D, H9T, 1COW, 1I3J, 1MM8, 2RAM, 1DSZ, 1JO1, 1MXJ, 303D, 1DVL, 1JTO, 1PUF, 334D, 1EEL, 1K2Z, 1R0O, 360D, 1AWC, 1I3J, 1K61, 1R0O, 3CRO, 1K2Z, and 1R4R. Several of these structures were used in a modified form that included annotations to either highlight or hide different portions of the structure. To make these materials more convenient for student use, we packaged the structure files (both annotated and nonannotated) and Cn3D together on a CD with an electronic textbook, laboratory activities, and animated tutorials. The resulting CD-ROM, *Exploring DNA Structure*, is a commercial product that is compatible with Microsoft Windows (95/98/Me/NT/2000/2003/XP) and Mac OS X and can be obtained from www.ExploringDNAStructure.com (Porter, 2005). Some of the features on the CD (e.g., autorun menus, the Windows and Mac versions of Cn3D, and Adobe Acrobat Reader [Adobe, San Jose, CA]) are not likely to function on UNIX systems; however, if students download the UNIX version of Cn3D from the NCBI, obtain a UNIX version of Adobe Acrobat, and use Firefox as a Web browser, they will be able to view all of the Web-based tutorials, use all of the structures, and have access to the manual and all of the activities.

Development of Instructional Materials

As described above, the *Exploring DNA Structure* CD-ROM developed in this project contained the following elements: a manual with worksheets designed to guide student thinking, several DNA structures, animated tutorials (written in JavaScript) to provide assistance in using Cn3D, and the Cn3D program. Example worksheets and tutorials can be seen at the Geospiza website (http://www.geospiza.com/education/products/eds_cd.html). NMR structures were included in addition to structures derived through x-ray crystallography so that students could view hydrogens, which are not revealed in crystal structures. All of the activities on the CD (see Box 1) involve active learning and include opportunities for inquiry-driven work as recommended by National Science Education Standards (National Research Council, 1996). Throughout this series, students uncover features of DNA structure by working with the structures and making observations. Many of the activities have students make predictions about the structures, record their predictions, and analyze the structures to find the answers.

Box 1. Activities on the exploring DNA structure CD

1. Identify the major and minor grooves
2. Investigate interactions between DNA and other molecules
3. Compare the orientation of the two DNA strands
4. The building blocks of DNA
5. The DNA backbone
6. How are two DNA strands held together?
7. Transmitting genetic information to the next generation
8. Base-pairing in double-stranded DNA
9. Investigate other DNA structures
10. Further investigation using the Internet

A representative activity is the investigation of complementary base pairs, entitled "How Are Two DNA Strands Held Together?" In this activity, students determine which bases form complementary pairs and count the number of hydrogen bonds in each pair. They begin by opening a DNA structure file in Cn3D and clicking a letter in the DNA sequence to highlight a base in the three-dimensional structure. They identify the complementary base by using Cn3D to find and highlight the nearest base on the opposite strand. Last, they use a space-fill rendering style to view and count the number of hydrogen bonds between the two bases (Figure 1).

Study Design

Because the central dogma of biology is fundamental, some aspects of DNA structure are taught in nearly every introductory biology course. Thus, we were able to test the effectiveness of the materials at multiple sites and with students at different educational levels. Also, because we had originally developed materials that covered the structures and functions of both DNA and RNA, we were able to divide the materials, create two CDs with comparable sets of

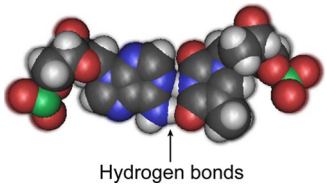

Figure 1. Hydrogen bonds between adenine and thymine. Laboratory activities were designed to guide students through investigating diverse features of DNA structures. This image from Cn3D shows a view that students would obtain during an investigation of base pairing. Students would identify each nucleotide, determine which elements share hydrogens, and count the number of hydrogen bonds in different nucleotide pairs.

materials (text, tutorials, worksheets, Cn3D, and structures), and provide students with the RNA CD as a negative control. Although some of the RNA materials did introduce similar information to that found on the DNA CD, such as the structure of RNA molecules and the difference between the nucleotides in RNA and DNA, many of the activities on the RNA CD were focused on the steps in gene expression and interpreting the genetic code. Specifically, the materials on the RNA CD emphasized comparing the nucleotide and sugar structures in DNA and RNA, identifying different types of RNA, examining the base pairing between the tRNA and the codon, and looking at tRNA structure. The assignments from the RNA CD, however, did not include the activities in which RNA and DNA structures were compared.

Pre- and posttests were used to measure student learning. Each test had 14 multiple-choice questions that addressed the same topics and mapped to specific activities on the CD (see the sample test in the online Supplemental Material). In our initial study, students were randomly assigned either of the two tests (dna1 or dna2) as pre- and posttest pairs in order to compare the difficulty of the two versions. Students took the pre- and posttests online by selecting a link from their CD and logging in to an online testing system that assigned an ID number and stored a time stamp and results for each test in a MySQL database (MySQL AB, Uppsala, Sweden). To compare the results from different tests and to determine whether the mean values were significantly different, we converted the number of correct answers to percentages and used these values to calculate test statistics. To protect student privacy, we did not store any information that might be used to identify individual students. We were aware that college students might work together while taking tests. Therefore, we did not offer students any incentives that were tied to test performance.

Two other concerns were identified that we sought to address. These were the possibility that information learned from the pretest might carry over to the posttest and that the two tests were of comparable difficulty. To minimize carryover from one test to the next, we considered both the subject matter included in the tests and the time between tests. In terms of subject matter, the two forms of the test differed from each other in the following ways: the questions were presented in a different order; sometimes the questions used different images, such as an image of a DNA molecule with a drug bound to the minor groove instead of an image with a protein bound to the major groove; and in different versions of the test students were asked to identify different parts of the same object. For example, one test might ask a student to identify a cytosine in a DNA structure, whereas the other version might ask a student to identify a guanine. We were only partially successful in controlling the time between tests, as we will discuss in the *Results*; therefore, some of our work took place in retrospect through limiting our analyses of the college student data to pre- and posttest pairs that had been taken at least 5 d apart. This was not an issue with the high school students because their teacher was able to control when they took the tests.

STUDY GROUPS

Student Interns

A small pilot study was carried out to provide a formative evaluation with seven student interns who assisted on the project. Two of the interns were college students, two were high school juniors, and three had just completed ninth grade. All of the interns had taken ninth-grade biology, and one college student was a biology major. The interns took a pretest, worked through the materials, completed worksheets with minimal instruction, and then took a posttest at least 1 d after completing the materials.

Community College Students

Students in the biotechnology programs at Austin Community College (ACC) and Oklahoma City Community College (OKCC) participated in the initial study along with students in the general biology course at Seattle Central Community College (SCCC). These students were given CDs with either RNA or DNA materials. These students normally learn about DNA structure through lectures, reading, and laboratory activities with plastic models.

Lynnwood High School

Lynnwood High School (LHS) is located just north of Seattle and had an enrollment of 1407 students during the test period. Because we worked with a vocational biotechnology class, and not an upper-level honors class, the demographic makeup in our test class reflected the school as a whole. Forty-four percent of the students enrolled at LHS were ethnic minorities, with the following distribution: 65.8% Caucasian, 8.8% Hispanic, 5.1% African or African American, 19.3% Asian or Pacific Islander, and 1.0% American Indian. Approximately one-third (33.5%) of the students at LHS qualified for free or reduced lunch, and 10.8% of the students were categorized as transitional bilingual (Washington State Office of Superintendent of Public Instruction, 2006). At the end of the fall quarter of 2004, 134 students took pretests. Students who stayed in the class during the winter quarter were randomly given either DNA CDs or RNA CDs to use for 2 wk and then were asked to take posttests. In the fall quarter section of this course, these students had completed assigned readings and worked on laboratory activities related to DNA structure, such as digesting DNA with restriction enzymes and separating fragments by agarose electrophoresis.

The Johns Hopkins University

Our largest and longest study was carried out in collaboration with the biology faculty from the Johns Hopkins University (JHU). Normal classroom activities in the JHU biology program involve lectures and assigned readings. During the fall semesters of 2004 and 2005, approximately 500 General Biology students participated. The students in these classes were expected to show a demographic distribution similar to that found in a 2002 survey. In 2002, many students taking General Biology were freshmen (54%) and sophomores (27%), with a few juniors (12%) and very few seniors (<1%). The 2002 class was also ethnically diverse, with 36% of the students describing themselves as Caucasian, 13% African American, 15% Hispanic, and 36% Asian. A subset of JHU students, identified as Biology Workshop students, also participated in the study. Biology Workshop students are those who took Advanced Placement (AP) Biology in high school. Because these students are considered to have completed the equivalent of a college biology course, they have fewer requirements and only attend a subset of the General Biology activities and a weekly research seminar.

Our research protocol differed slightly during the 2 yr of the study. In year 1, students were assigned to four random groups and took one of four combinations of pretests and posttests. Half of the students used our control CD (with RNA materials), and the other half used the DNA CD. Our ability to measure the effect of the CDs

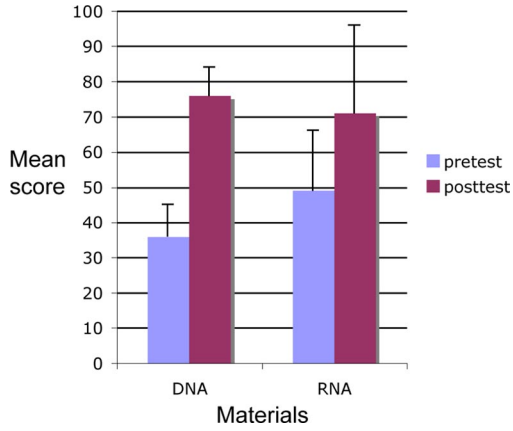

Figure 2. Comparison of DNA and RNA pre- and posttest results from the pilot study. Pre- and posttests were used to measure learning gains by student interns after working with either DNA or RNA materials. Error bars, SD.

was somewhat restricted during the first year, because the fact that some students had the DNA CD and some did not limited our ability to assign a large number of activities. During the second year of the study (2005), a greater number of activities were assigned from the CD, all students used the DNA CD, and all students completed the same pair of pre- and posttests. Based on student feedback, minor changes were made to the CD menu between 2004 and 2005 to improve navigation.

RESULTS AND DISCUSSION

Results with Student Interns

The pilot study with student interns showed significant increases in test scores after students worked with Cn3D and either the DNA materials or the RNA materials (Figure 2). Students who worked with the DNA materials had significantly greater scores on the DNA posttests ($p < 0.01$). Significantly greater scores were seen on RNA posttests as well, although the change was not as great ($0.05 > p > 0.025$). In both cases, the p values were calculated using a one-tailed, paired t test from the percent of correct answers. On average, the students answered 35% of the questions correctly on the DNA pretest. Because the RNA test covered topics that were not part of the DNA materials, such as identifying the cellular locations and molecules involved in translation, information about the genetic code, and identifying structural elements in a tRNA, we expected students to score at a similar level on the RNA pretest. This assumption proved to be incorrect because the average RNA pretest score showed students answering close to 50% of the questions correctly. Because the DNA materials were used first, it is likely that these students performed better on the RNA pretest because of the information they learned while working with the DNA structures. In fact, 4 of the 16 questions on the RNA test did ask about concepts that partially overlapped the DNA structure materials. These questions asked about the nucleotides and sugar that would be found in RNA, identifying the 2′ carbon, and choosing the correct sequence and orientation of an anticodon, when given the sequence of a codon. Classroom learning was not a contributing factor in the pilot study, because these activities took place during the summer when the students were out of school.

Comparison of the Two Versions of the DNA Test

Pretest results from the two versions of the DNA test (dna1 and dna2) were compared to determine whether the tests were equivalent in terms of difficulty and, at the same time, to learn something about the background knowledge of the different test groups. Figure 3 shows the mean scores from pretests taken by students at JHU (n = 91, 31, and 29, respectively, for dna1, dna2, and the Biology Workshop students), ACC (dna1, n = 5; dna2, n = 9), OKCC (dna1, n = 6), SCCC (dna1, n = 21), and LHS (dna1, n = 84; dna2, n = 50). We expected that the high school students would have the lowest scores for a variety of reasons, and this was confirmed by the data. This result was not surprising for a number of reasons. These students were younger (15–17 yr old) and less educated, and despite some exposure to DNA

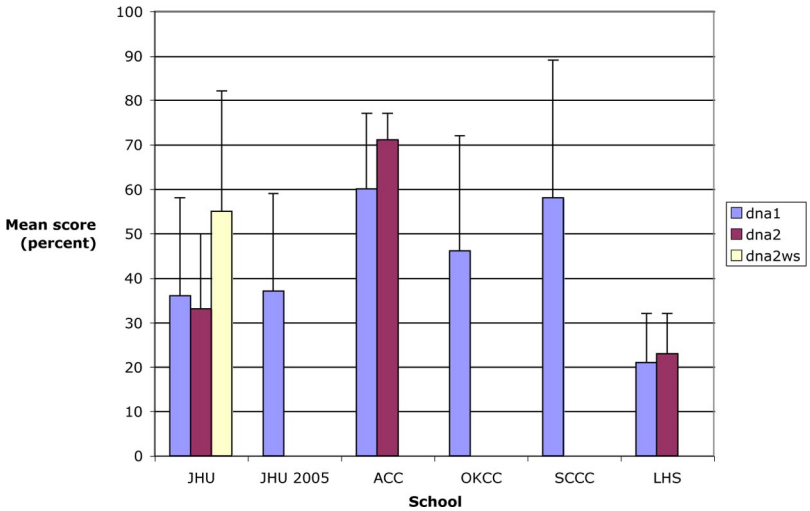

Figure 3. DNA pretest results from different study groups. The mean scores on two versions of the DNA pretest were compared for students from JHU, ACC, OKCC, SCCC, LHS, and JHU Biology Workshop students (dna2ws). Error bars, SD.

in terms of performing restriction digests and carrying out agarose gel electrophoresis, these students had not explored the subject material in depth. The community college students obtained some of the highest scores on the pretests. In the case of ACC and OKCC, we expected the scores to be higher because those students were in biotechnology programs and had more experience in working with DNA. In the case of SCCC, we anticipated that the JHU students and the SCCC students would have similar scores on the pretests because they were all enrolled in General Biology. In contrast, the SCCC scores were higher than those seen at JHU. One explanation for the higher scores in the SCCC sample could be the timing of the pretest. The SCCC students took the pretests later in the quarter than did the students at JHU, making it likely that similar topics had been covered in class. It is also possible, given the demographics of community college students (median age = 26 at SCCC in 2005) that some students had previous coursework in biology.

JHU students split into two defined groups, consistent with their educational background: those currently enrolled in their first college-level biology course (General Biology 1) and those enrolled in their second college-level biology course (Biology Workshop I). The Biology Workshop students (identified as dna2ws in Figure 3), scored higher on the pretest, with a mean of 55%, comparable to the pretest scores from the community college students. The JHU General Biology students (2004) had mean scores of 36 and 33%. These scores were comparable to the average value of our student interns' scores in the pilot study and were consistent with scores from the JHU General Biology students in 2005 (mean = 37%).

Next, we asked whether the two versions of the DNA test were of equal difficulty by comparing the results from groups that took different versions of the test. We used a one-tailed nonpooled t test to determine the p values from the percent of correct answers. No significant differences were seen between the two versions of the test for the LHS students, ACC students, or the JHU 2004 students. All p values were >0.10, allowing us to accept the null hypothesis and conclude that the two tests are equivalent in terms of difficulty.

Time between Tests

Before addressing further questions, we looked at the amount of time that passed between the pre- and posttests for different groups of students (Figure 4). The high school students took tests more than a month apart, in class, on dates that were determined by their instructor. In 2004, all of the college students were able to take tests at any time because they could use the links on the CDs to access the pre- and posttests.

For this portion of the study and because the number of community college students was small, we combined the community college students into one group (CC) and compared their test-taking patterns with the students from JHU in 2004. Although some (n = 10) JHU students took both the pre- and posttests within 30 min, most waited at least 5 d before taking the posttests. In contrast, most of the community college students completed both tests within 30 min or less, making it unlikely that they could have worked through any of activities between the two tests. On the basis of these results, we omitted the community college data from further analyses. Although our results from the community college group were uninformative in terms of evaluating the effectiveness of the materials, these data did provide an important insight that informed our testing procedure during the second year of our study. In 2005, we removed the test links from the CD and instead utilized the Blackboard system at JHU to make test links available at specified times. This system provided better control over testing times and allowed us to enforce a minimum time between tests (55–80 d).

Did the DNA CD Materials Enhance Student Learning?

Once we established that the two versions of the DNA test were comparable in terms of difficulty and identified our study groups, we asked whether the DNA CD contributed to student learning by comparing scores on pre- and posttests for both the LHS students and the JHU students.

Lynnwood High School

We compared student performance on the DNA pre- and posttests from students who had either DNA CDs or the

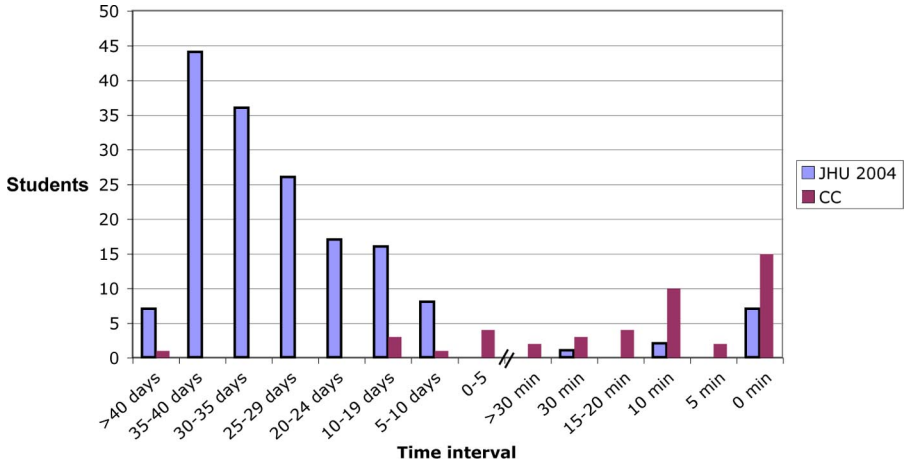

Figure 4. Time between pre- and posttests. The amount of time between taking the pretest and taking the posttest for the JHU students and the CC students, combined together in one group. The y-axis shows the number of students taking pre- and posttests at different time intervals.

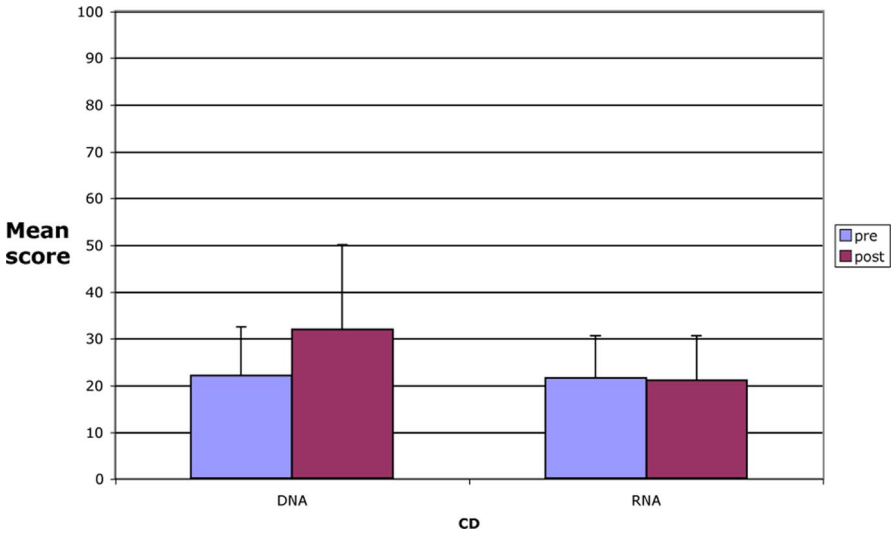

Figure 5. DNA test results from LHS students. The mean scores are shown for the high school students (LHS) on the DNA tests, who had different kinds of CDs. Students with the RNA CDs comprised the control group. Error bars, SD.

control CDs (RNA; Figure 5). The students with the DNA CDs (n = 73, pretest; n = 37, posttest) showed a significant increase in scores on the DNA posttest ($p = < 0.005$, using a nonpooled, single-tailed *t* test). In contrast, students with the control CDs (n = 61, pretest; n = 26, posttest) obtained similar scores on both tests (pre- and posttest scores, 22 and 21%, respectively) and did not show a significant increase.

Students at JHU

The pre- and posttest results from four groups of JHU students are shown in Figure 6. All four groups of JHU students showed a significant increase in mean scores between the pre- and posttests (Figure 5). The students, in 2004, with the RNA CDs (n = 74) represented the control group. Their increase in learning between the pre- and posttests ($p < 0.005$) showed that the standard mixture of lectures and reading assignments used in the General Biology course resulted in significant learning gains that could not be attributed to using the DNA CD.

To assess the contribution to learning from the DNA CD, we used a nonpooled, single-tailed *t* test to assess the significance of the differences between the mean scores (percent correct) on the posttests from the groups that worked with the DNA CD and the mean posttest scores obtained by the RNA control group. We found that the scores for the JHU General Biology students (2004, n = 48) were significantly higher for the students who had the DNA CDs than the scores from the control group ($p < 0.05$). The observation that these materials had a significant impact on learning was confirmed in 2005, with a second, larger, group of students (n = 129) who showed greater learning gains ($p < 0.005$). It is likely that the increased scores in 2005, over 2004, resulted from increased use of the materials. Because JHU adopted the CD for the General Biology class, the instructors had greater latitude and assigned an increased number of activities from the CD, because all students now had a copy of the same (DNA) version. A simplified menu on the DNA CD may have influenced test scores by making it easier to navigate through the CD and complete the materials.

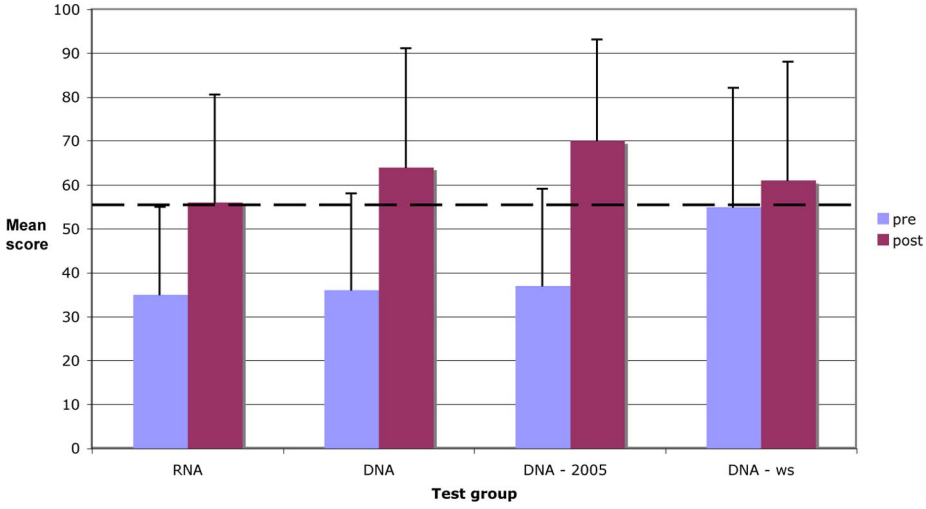

Figure 6. DNA test results from JHU students. The mean scores (percent correct answers) are shown for the DNA pre- and posttests from the JHU students with different kinds of CDs. The mean posttest score for the control group (56%) is indicated by the dashed line. Error bars, SD.

Although the Biology Workshop students (n = 29, labeled DNA - ws) did perform better on the posttest than the RNA control group, the difference between their posttest scores and those of the control group were not significant. In some respects the posttest scores from these students are surprising. Because this group of students completed AP Biology in high school, they enter the General Biology course with a greater amount of prior knowledge than most students, as evidenced by their performance on the pretest (mean = 55 vs. mean = 37 for other biology students). These characteristics might lead us to predict that these students would attain better scores on the posttest. One explanation for the difference might be that these students studied less because they entered the course knowing a large amount of the material. A more likely explanation might be the different course requirements for these students. Because Biology Workshop students are assumed to have completed a college-level course (AP Biology), they are exempt from participating in many of the General Biology activities. Their lower scores on the posttest may reflect their decreased involvement in the course.

Examination of Test Results by Item

We next looked at the percent of students who answered each question correctly, in order to determine which topics mapped to student learning and whether or not these improvements correlated with specific assignments. The LHS students, who used the DNA materials, scored better on 10 of 14 test items (Figure 7). One test item, however, identifying either the major or minor groove, showed an unusual pattern. Not only did an unusually high number (65%) of the LHS students answer this question correctly on the pretest, the percentage of correct answers did not change between the pretest and the posttest. Conversations with the LHS teacher indicate that this item may have been used to demonstrate how to take the test. That the fraction of LHS students answering this question correctly is unusually high is supported by the contrast with the JHU students; only 43% of the JHU 2004 students answered this question correctly on the pretest.

In 2004, the number of JHU students answering 9 of 14 questions correctly was greater when they had the DNA CD than when they had the control RNA CD. In 2005, this increased to 12 of 14 questions, with an increased number of students providing correct answers for 9 of 14 questions (Figure 7). In 2004, the assignments for the JHU students were aligned with the questions on the major and minor groove, the elements in DNA, and the parts of a nucleotide. All three of these questions showed an increase in correct scores over the fraction of correct scores from the control group. In the JHU 2005 sample, we found an increase in both the number of topics that were answered correctly and the number of students who answered the topics correctly, consistent with the greater number of assignments.

The only questions that did not show an increase in the number of correct answers were those in which students were asked to identify 5′ and 3′ carbons in a drawing of double-stranded DNA (Figure 7). We do not know, but suspect that this result might indicate a difficulty with the online test format (discussed below). This idea will be investigated in future work by comparing the written and online versions of the tests. If the written tests also show a difference for this item, we will add activities to the materials that are related to DNA replication in order to draw more attention to the chemical differences between the 5′ and 3′ ends.

In the case of the JHU students, our results may portray smaller learning gains than the true change. First, the learning gains may have appeared lower because our control group of students from JHU might have learned information from the RNA CD that favorably impacted test scores. This idea is supported by the observation that the control group, with the RNA CD, showed a significant improvement on the posttest. Of course, in this case, we cannot distinguish between test scores that improved as a result of the RNA

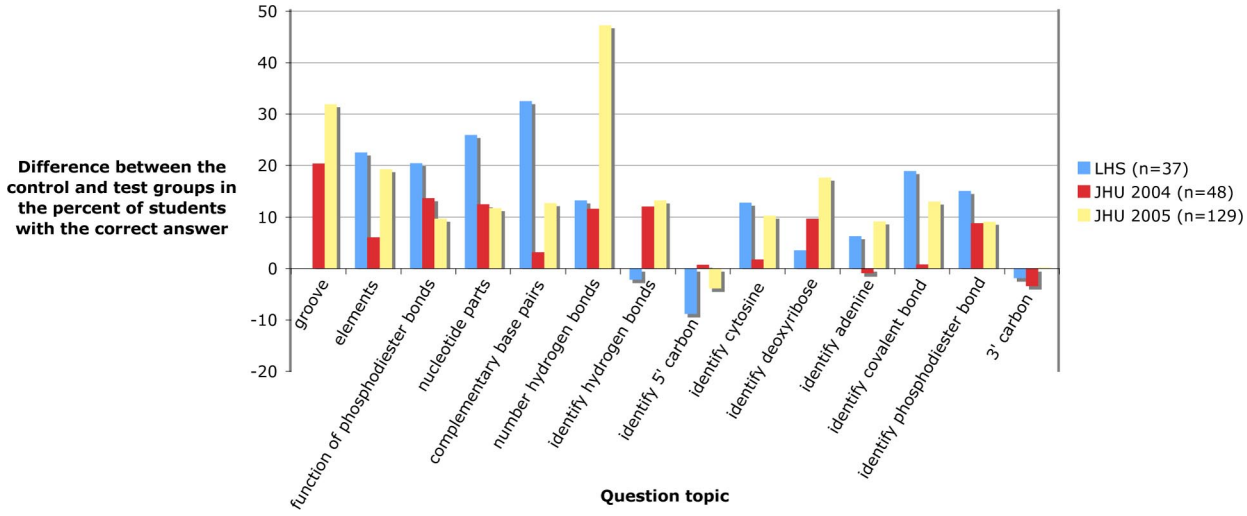

Figure 7. Correct test answers broken down by topic for different groups of students. The *y*-axis represents the change between the pre- and posttests in the fraction of high school (LHS) and college students (JHU 2004 and 2005) providing correct answers on different topics. A negative value indicates fewer correct answers on the posttest.

materials and test scores that were better as a result of taking the class. We can address this question more fully in future studies by having classes take pretests and posttests before later classes begin working with the materials.

An alternative explanation for potentially lower learning gains concerns the usability of the online test. We noticed that overall, the second half of the test showed lower scores. This can be seen in Figure 7; the items concerned begin toward the middle, at the question about identifying hydrogen bonds, and continue to the right side of the graph. It might be that these topics were more difficult or that students who failed to complete the activities had difficulty in translating the three-dimensional visual information from Cn3D to a two-dimensional drawing. We suspect that usability issues contributed to the lower scores for the following reason: All of the questions on this half of the graph were related to identifying specific items in a DNA image. Technical issues with the database and our test website caused this image to be displayed on a webpage that was separate from the rest of the test. Students who knew how to adjust the sizes of the Web browser windows could see the image and the test at the same time; however, for students who did not know how to adjust window sizes, it would make the second half of the test harder because students would have to remember the question, change windows to look at the DNA image, and remember the answer until they changed back to the test window and clicked the correct button on the screen. We have observed, through giving professional development workshops, that many college instructors do not know that it is possible to adjust window sizes and to view two windows on a screen at the same time. Therefore, we think this might be true of students as well. This issue will be addressed in future studies by comparing written and online tests and modifying the Web pages accordingly, to put the image on the same page as the questions.

CONCLUSION

We found that working with Cn3D and the related learning activities on the *Exploring DNA Structure* CD-ROM led to a significant improvement in the performance of both high school and college biology students on tests designed to measure their knowledge of DNA structure. Our results with control groups of students (RNA CDs) demonstrated that this increase in performance was greater than the increase that could be attributed to conventional teaching methods (lectures, assigned reading, and lab activities). Further, the learning gains exhibited by the college students (JHU) increased during the second year of the study, when a greater number of activities were assigned from the CD. Although some of this increase may have resulted from improvements in navigation related to changes in the CD menu, it is more likely that the increased performance was tied to an increased use of the materials.

A greater increase in test scores overall was seen with the college students than the high school students. This difference reflects multiple factors including maturity, cognitive development, and socioeconomic and motivational differences between students in 10th grade and students in college. In addition, the LHS students only had the CDs for 2 wk, whereas the JHU students had the CDs for an entire semester. When the test questions were analyzed by topic, we found that the high school students showed a greater gain on the first half of the test questions. Because the LHS students worked at their own pace and we do not know the number of assignments that were completed, it is likely and consistent with the results (Figure 7) that they performed better on questions related to the first sets of activities. This may also reflect a difference in background knowledge and course content, because the high school students had not taken chemistry and had covered DNA structure in less depth than the students who took college biology. Further studies are needed, however, in order to better assess the impact of using these materials on high school student learning.

Bioinformatics resources, in particular those used for visualizing macromolecules, such as Cn3D, hold great promise for enhancing biology education. The ability to visualize molecules or create animations to illustrate abstract concepts is an important development in computing technology and has been especially beneficial for promoting student learning in subjects such as cell biology (Stith, 2004) and chemistry (Henry, 2004). In a high school setting, visualization activities, when presented in an integrated framework along with inquiry-based learning and assessments, have been tied to learning improvements in physics, chemistry, earth sciences, and biology (Linn *et al.,* 2006). Visualization tools have also been used in other educational settings, with physicians using these types of programs to communicate health information to their patients (Noland and Juhn, 2001). We plan to expand our work with Cn3D as well, completing the materials that we developed for studying RNA and expanding materials that we have developed for structure comparisons. Although this study was largely confined to Cn3D activities, we have developed other materials that utilize multiple bioinformatics programs and databases. This study concentrated on Cn3D, in part, because it runs on a personal computer, and we were informed by high school teachers that they did not want to be constrained by needing access to the Internet (Porter *et al*., 2003). As Internet access becomes more prevalent in the classroom, we will have a greater ability to include other types of bioinformatics-based activities in our investigations of student learning.

ACKNOWLEDGMENTS

We thank Charlotte Mulvihill for reviewing the manuscript, Erin Dolan for her analysis of student feedback, Paul Thiessen and Eric Sayers for advice and guidance in using Cn3D, and our student interns, Jeremy Friend, Kate Anderson, Camille Beck, Andrea Porter-Smith, Sam Ryan, Marian Deuker, and Thea Chard, for all of their work in testing and reviewing materials. We also thank Doug Clarkson for his kind advice on performing the statistical analyses. This project was supported in part by a grant to Geospiza from the National Science Foundation's Course, Curriculum, and Laboratory Improvement Program (DUE-0127599) and a Visiting Scientist Fellowship to S.G.P. from the NCBI.

REFERENCES

Altman, R. B. (1998). A curriculum for bioinformatics: the time is ripe. Bioinformatics *14*, 549–550.

Bednarski, A., Elgin, S., and Pakrasi, H. (2005). An inquiry into protein structure and genetic disease: introducing undergraduates

to bioinformatics in a large introductory course. Cell Biol. Educ. *4*, 207–220.

Henry, C. (2004). Visualizations and Translations. Chem. Eng. News *82*, 46–48.

Honts, J. (2003). Evolving strategies for the incorporation of bioinformatics within the undergraduate cell biology curriculum. Cell Biol. Educ. *2*, 233–245.

Linn, M., Hee-Sun, L., Tinker, R., Husic, R., and Chui, J. (2006). Teaching and assessing knowledge integration in science. Science *25*, 1049–1050.

National Research Council (1996). National science education standards. Washington, DC: National Academy Press.

Neeper, S. (2005). Packing it all in: curricula for biomedical computing. Biomed. Comput. Rev. *1*, 7–13.

Noland, K., and Juhn, G. (2001). The Visual Health Experience: Tools, Technology, and Information for Providers and Their Patients. Adam, Inc. http://www.adam.com/VLWP.pdf (accessed 25 September 2006).

Pevzner, P. A. (2004). Educating biologists in the 21st century: bioinformatics scientists versus bioinformatics technicians. Bioinformatics *14*, 2159–2161.

Porter, S., Fletcher, L., Murphy, S., Pearlman, R., Staheli, J., and Johnson, E. (2003). Teaching Undergraduate Biology with Bioinformatics. BISTI Symposium on Digital Biology. http://www.bisti.nih.gov/2003meeting/Abstracts/II10.pdf (accessed 1 October 2006).

Porter, S. (2005). Exploring DNA Structure. Seattle: Geospiza. http://www.cafepress.com/geospiza.22638483 (accessed 25 September 2006).

Porter, S. G., and Smith, T. M. (2000). Bioinformatics in the biotechnology classroom. J. Indust. Microbiol. Biotechnol. *24*, 314–318.

Stith, B. J. (2004). Use of animation in teaching cell biology. Cell Biol. Educ. *3*, 181–188.

Washington State Office of Superintendent of Public Instruction (2006). Data Administration. http://www.k12.wa.us/DataAdmin (accessed 10 January 2006).

Wang, Y., Geer, L., Chappey, C., Kans, J., and Bryant, S. (2000). Cn3D: sequence and structure views for Entrez. Trends Biochem. Sci. *25*, 300–302.