

An Interactive Practical at the Interface of Web-based and Conventional Publishing

[Teresa K. Attwood](#)

School of Biological Sciences, The University of Manchester, UK

Background

When I was a student twenty years ago, I remember being introduced to CAL as part of a chemistry course. I recall being led down to the murky bowels of the department and being ushered into a dull room full of huge VDUs. Having followed the instructions to logon to these prehistoric machines, the class was left alone for 90 minutes in order to complete a computer-based 'project'. Not knowing quite what to expect, this turned out to be little more than a tedious multiple choice exercise, which felt totally divorced from the active learning opportunities provided by tutorials and practicals. To me at least, the entire experience seemed not only odd, but somehow rather pointless.

Today, things are a little different. In the current technological era, far from being confined to special rooms in university basements, computers are fast becoming common pieces of laboratory equipment. It is now almost taken for granted that biologists should have access to up-to-date databases, via advanced interrogation software, and that they should have developed skills to interpret diverse database search outputs. Yet, in spite of living in a highly-advanced, computer-literate age, in which many young people have access to computers from their own homes, persuading biology students not to be afraid of computers can still be surprisingly difficult.

In recent years, the World Wide Web has provided new opportunities for innovation in teaching. Web-based approaches allow students to gather information from different corners of the globe, literally at the click of a mouse button. This process attracts mounting interest when different web pages offer added extras, such as animations, or tools with which to commune with data interactively. A great advantage of learning on the web is that, depending on the design of the teaching material, students may be guided as much, or as little, as a particular course demands. Thus material may be used simply to supplement lecture courses, or may be completely self-contained. In either scenario, armed only with a URL, study may continue away from formal classes. Students may thus explore independently, in a self-paced setting, and compare notes when once again back in the laboratory.

A skills gap

Until recently, traditional biology teaching has not needed to address computational biology. The problem is that biological sciences are now suddenly feeling the impact of the genome projects, and are consequently being flooded with vast quantities of data. This *embarras de richesses* has spawned the field of bioinformatics, which is principally concerned with the acquisition, archiving, analysis and interpretation of these vast quantities of data. This has involved establishing and maintaining databases; designing powerful interrogation and analysis software;

and creating tools to help interpret results of analyses in biologically meaningful ways. To achieve this, and to do it well, requires the input of scientists skilled in relevant areas of both biology *and* computing.

Owing to the fecundity of the genome projects, bioinformatics is a fast growing discipline. Unfortunately, however, the rate of emergence of the field has created a skills gap: while there is an urgent and widespread need for trained bioinformaticians, there are virtually no courses available through which to provide the much-needed training. Postgraduate courses are still relatively rare, and mainstream undergraduate courses almost non-existent. In the meantime, there is a pressing need to make today's trained biologists more computer literate, a necessary process if bioinformatics is to make a significant impact in solving current post-genome problems.

Embracing the web

Despite the computer age, teaching computer applications to biologists can still be remarkably difficult, as many students (and their teachers) are 'silicophobic' - they may be used to handling systems *in vitro* or *in vivo*, but not yet *in silico*. Nevertheless, the emergence of the web, coupled with advances in web technology, has provided interesting new opportunities for teaching.

There are several advantages to web-based approaches. For example:

- 'point and click' interfaces are easy to use, even for the most computer-wary student;
- data from different sources can more easily be brought together, so teachers need not be expert in all areas, but may direct students to more appropriate expertise on the Internet;
- the use of interactive programs can render data more tangible and visually appealing, and hopefully, therefore, easier to understand and remember;
- the most up-to-date developments in the field can be accessed and exploited; and
- armed with a URL, study may continue away from formal classes, allowing students to work at their own pace, whether at home or on computers elsewhere on the College campus.

However, there are also several disadvantages to web-based teaching:

- Internet traffic can bring work to a stand-still, especially during "rush" hours;
- students can get lost in hyperspace, especially if navigation aids are not made obvious;
- students often don't bother to read web pages, so even if instructions are available, they will probably never be read;
- 'point and click' interfaces are easy to use - so easy, in fact, that clicking from beginning to end of a document designed to occupy a class for 2 hours can (and will) be achieved by some students in a matter of minutes;
- armed with a URL, study may continue away from formal classes - the ideal excuse for some students never to turn up at the laboratory; and
- maintaining web pages is the proverbial uphill struggle - sometimes it seems that no sooner has a page been cast in HTML than half its URLs are already broken!

These are only some of the numerous possible advantages and disadvantages. Thus, while it is easy to put lecture notes and diagrams on the Internet, designing *effective* web-based teaching material is rather less trivial. Obviously, a good tutorial should aim both to capitalise on the powers and to avoid the pitfalls, but this is not as easy as it sounds, because no matter how well you think you've thought it through, at least one student will always achieve the unexpected!

In spite of the potential problems, the web has been embraced with enthusiasm as a teaching medium, and courses and primers in all aspects of science have proliferated. However, some courseware is still little more advanced than the CAL exercises of the late '70s, involving screen by screen presentation of static information, with embedded questions that allow progress to the next page when answered correctly. Other more advanced courses require registration (and usually fees), but offer greater interaction (e.g. via timetabled virtual meetings in multi-user dungeons) and often certification that the course has been attended or passed¹.

An interactive bioinformatics web practical

Conventional teaching of bioinformatics has tended to rely on the use of commercial command-line-driven sequence analysis packages². Such programs can be difficult and tedious to use for all but the computer enthusiast, and are opaque and off-putting for bench biologists. In 1995, in an attempt to tackle some of these problems, I developed a short bioinformatics lecture course, which I supplemented with an interactive web-based practical (*BioActivity*) (see Figure 1), giving students first-hand experience of sequence analysis on the Internet^{3,4}. By contrast with commercial packages, *BioActivity* provides an interface to analysis resources around the world, so students are not confined to a particular software environment or to a limited commercial perspective.

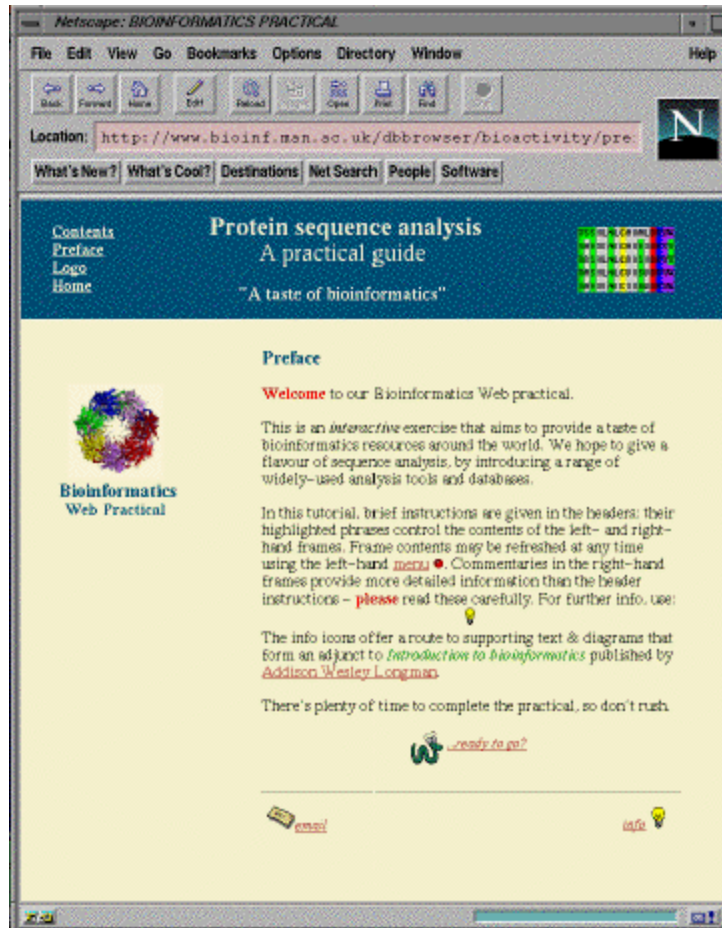


Figure 1. *BioActivity* home page. Each practical page contains a header, beneath which the screen is divided into left- and right-hand frames. As shown, the right-hand frame always provides a detailed commentary, which gives the rationale for the relevant part of the practical; this is augmented with pictures, to help illustrate more explicitly what is meant in the text. Beneath each commentary is also an 'info icon', which provides further explanations, additional pages of background information, and so on.

Inside the practical, as shown in Figure 2, each page carries brief header instructions, coupled with a more detailed commentary and embedded diagrams for those wishing to learn more (I borrowed this idea from the Rupert Bear Annuals I remember reading as a child - it was always exciting to unfold the story quickly by reading the couplets beneath the pictures, but it was ultimately more satisfying to discover the full tale from the free-form prose at the foot of the page. I doubt, however, that any user to date has made the Rupert connection!).



Figure 2a.



Figure 2b.

Figure 2. (a) Page from within *BioActivity*, showing the list of practical contents, together with header instructions and practical commentary. The consequence of clicking on header instructions is to change the information appearing in the left- and right-hand frames, as seen in (b). The commentary can always be restored using the menu in the header (left-hand side).

In addition to the instructions and commentaries, further information and suggestions for additional reading are provided in associated 'info loops', which are accessed via hyperlinked icons in the bottom right-hand corner of each page (see Figure 3). Much of this background material forms the basis of the accompanying lecture course; although not essential for the practical, students are advised to read it, knowing that most will probably ignore it until examination revision is upon them. Nevertheless, to get the most from the practical, the commentaries urge students to read as much as possible.

Protein sequence analysis
A practical guide
"A taste of bioinformatics"

A metaphor for biology
In sequence analysis, orthology vs paralogy complicates the marriage of biological information with sequence data: from a database search, it may be unclear how much functional information can be correctly transferred to a query from a matched homologue (thus errors have been incorporated and propagated through databases via automatically-derived annotations). The process is further complicated as similarity is confined only to part of an alignment, as with protein autonomous folding units that may be used as building blocks to form functions on the parent protein, through multiple combinations of via combinations of different modules to form mosaics, as shown that nature behaves like a tinker, using a patchwork of existing components to create a new, visible whole.

We can appreciate these difficulties by considering a rather fanciful biology encapsulated by **Rube Goldberg machines**. These are machines whose construction is so bizarre that an unknown 5% could not be deduce the machine was known. Biology could be considered the ultimate in which spare parts were grabbed during evolution's tinkering process (or adapted from their original functions), and frozen by chance as within a coherent, surviving organism. Echoes of those earlier functions but the remnants may be useless in surmising the new functions. The ramifications for sequence analysis – understanding the biology is to correctly infer relationships between components within complex systems.

Self opening umbrella
(RUBE GOLDBERG™ and © property of Rube Goldberg, Inc. Distributed by United Media)

Figure 3. Typical page from within an 'info loop', showing background material with embedded diagrams

Having progressed to the interactive aspects of *BioActivity*, results of searches are sent to additional left- and right-hand output pages, as shown in Figure 4. This means that the central commentary always remains visible, while allowing results from different database searches to be viewed simultaneously in separate windows (this facilitates both cutting and pasting of information between different pages, and comparison of different search outputs). At the end, references are given to the databases and search tools used, and an extensive glossary of terms is provided. A feedback form allows students to convey comments or problems (anonymously if they wish), so that any obvious difficulties can be quickly remedied.

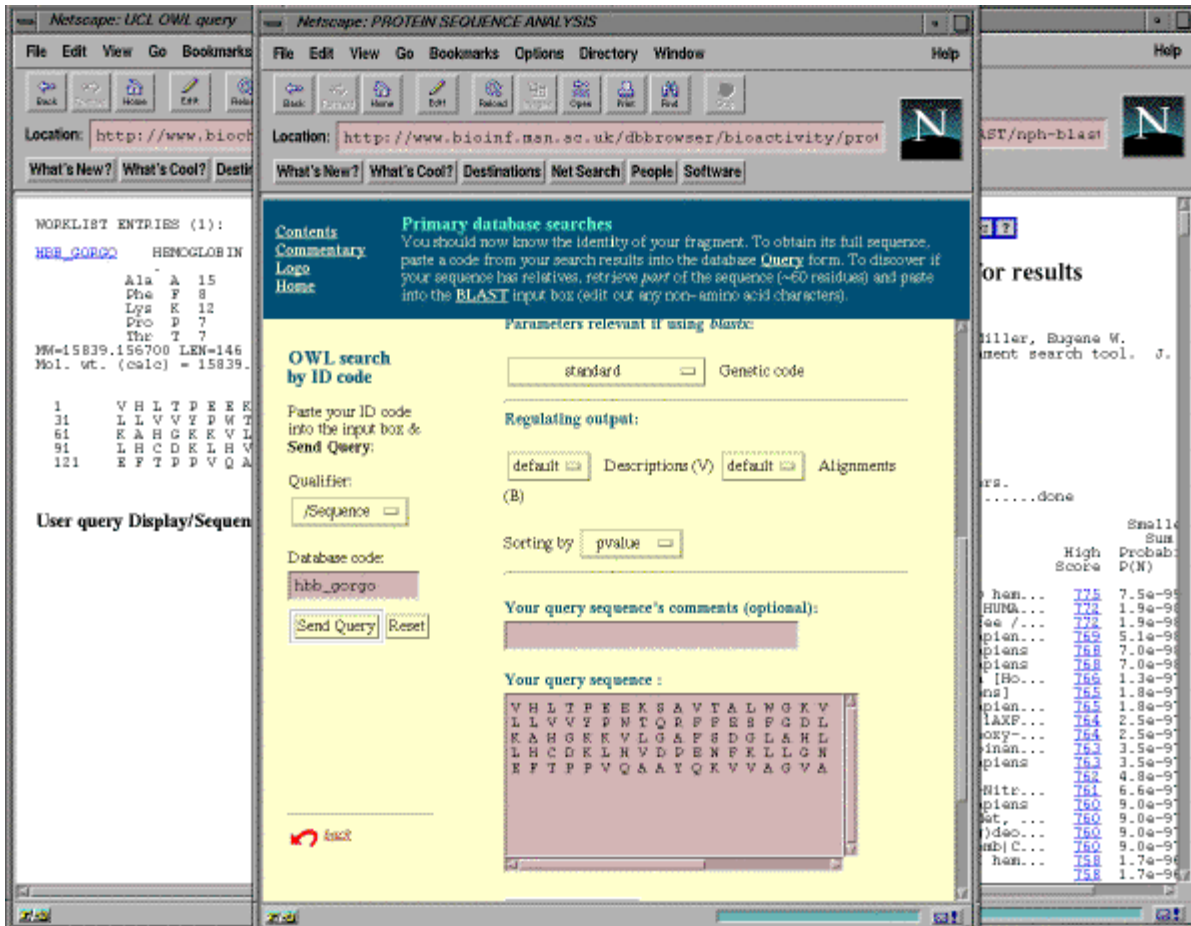


Figure 4. *BioActivity*'s first fully interactive page, illustrating the results of database searching having been sent to separate left- and right-hand output windows

Depending on how *BioActivity* is being used, students may tackle built-in exercise pages before progressing deeper into the practical (Figure 5); this ensures that the different stages have been understood, and encourages students to think about what they've done before moving on. Again, depending on context, at the end of the practical, students may be required to write a short report, detailing the main conclusions of the analysis, and placing them in a biological context (for the purposes of this exercise, they are warned not simply to submit pages of output without explanation, but rather to discuss the results in a reasoned and concise manner).

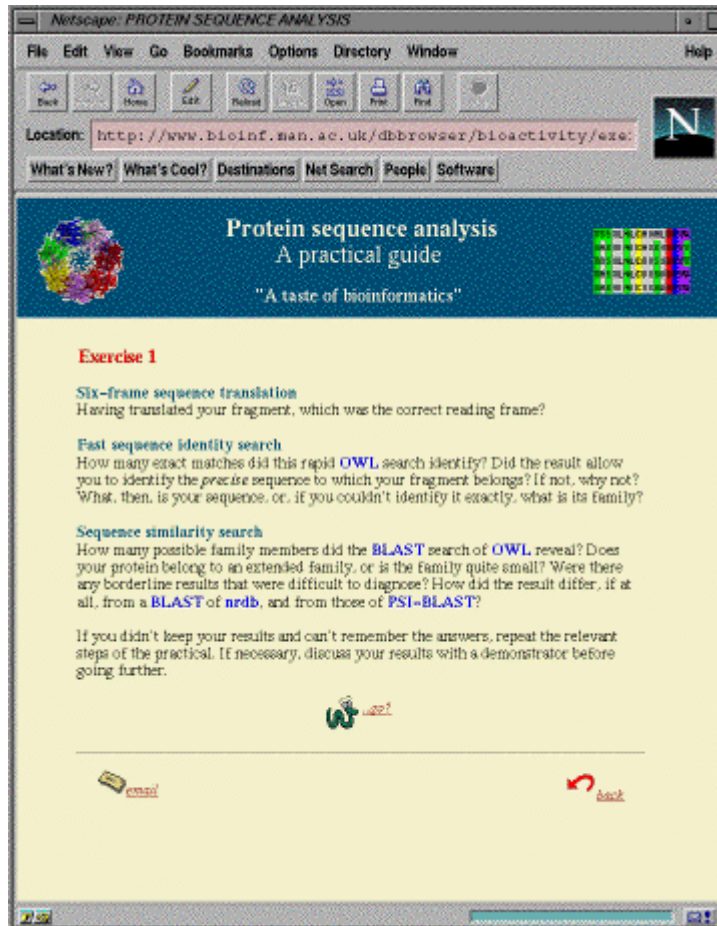


Figure 5. Example exercise page

The practical scenario

In terms of its content, the practical is built around the following scenario. A molecular biologist has sequenced a number of "unknown" 30-base fragments of DNA. The student is invited to select one of the fragments and, by means of a variety of real-time database searches, to discover the protein family to which it belongs and its biological function. If there is a known 3D structure, this is examined and its structural class uncovered. There are three main stages in the discovery strategy: (i) nucleic acid sequence analysis; (ii) protein sequence analysis; and (iii) protein structure analysis. The aim of database searching is to discover if a homologue, or homologous family, exists, and to reveal the presence of potential functional sites; interrogation of the structure classification resources then allows rationalisation of functionally important motifs in structural terms. Because the practical is 'live' on the Internet, rather than simply offering a collection of canned results, it offers the genuine experience of working with the web, which inevitably is sometimes good and sometimes bad.

Lessons learned

The practical runs adequately on PCs, but may respond differently to different browsers (*Netscape Navigator* is probably still the best option). *BioActivity* includes links to some Java software, allowing direct interaction with data downloaded from the Internet. However, in view of the possible problems associated with Java (speed, browser compatibility, available processing power, etc.), these are not core components.

A potential problem with PCs is the tendency of users to fill their screens completely with whatever application they're running. In the practical, a complication may arise when the main page 'speaks' to its left- and right-hand results windows, which will be obscured if the browser occupies the full screen. Of itself, the management of several windows is not a problem, but, surprisingly, it seems to demand a high level of organisation from the students.

BioActivity has been accessible via the web for several years and is gradually maturing. For example, I now try to provide alternative sites for some database searches, effectively offering 'rat runs' to avoid inevitable traffic jams. I have tried to improve navigation aids, so students don't get lost, but inevitably some still wander! Where possible, I have kept words to a minimum in the central pages, knowing that student attention span is inversely proportional to the length of the text. I have also provided baffles, in the form of short exercises, to delay the rampant 'point and clickers'. In every class, there is always a student who manages to finish in half an hour. In theory, forcing the student through the intervening exercises (the answers to which can be discussed, for example, with demonstrators) slows their progress and should make them reflect more carefully upon what they've done. Making a written report compulsory also helps to maintain their attention.

In addition to these developments, I have provided soft and hard routes through the practical. In other words, the starting materials (the initial 'unknown' DNA fragments) are extremely easy to analyse - their identities are usually discovered in the first exercise. However, once the practical has been completed with easy sequences, a set of 'mystery' sequences is available, which are much less likely to yield quick answers - of course, as a stand-alone teaching tool on the web, anybody may use the practical with his/her own sequence(s). *BioActivity* is thus suitable for different levels of user, and is adaptable to different teaching contexts.

Assessment and evaluation

The practical requires students to tackle a range of new concepts, including databases and software tools, which they're meeting for the first time. They must become familiar with a variety of web sites, and not only understand the differences between different types of database, but also the differences between the databases themselves and their software interfaces. Students have to evaluate critically the results from searching the different resources, and present their work as a coherent study, written in the style of a short journal publication.

The practical doubtless has its faults, but it evolves with increasing use and user feedback. To date, responses to questionnaires have been very positive, with 90% ranking the practical as a very useful aid to learning. For the newcomer, it is clear that there is a learning curve to be negotiated, but the overall experience seems to be worthwhile, providing a good foundation for

future exploration of bioinformatics resources on the web. Students moving on to higher degrees have reported that their work has benefited from the practical experiences gained on the course.

Other applications

BioActivity is suitable for use in a variety of teaching contexts, with appropriate modification of the supporting lectures: thus, for example, it is currently an effective tool for undergraduate, MSc and PhD classes, and for staff training. Furthermore, its web-based nature means that the material can be used as a stand-alone resource within a completely virtual environment (e.g. it has provided a core practical component of the Birkbeck College Virtual Course in Protein Structure Analysis⁵, and will soon be integrated into Manchester University's distance learning course).

A novel twist?

An important recent extension to the practical has been a link to a newly-published textbook⁶. Or, perhaps more precisely, the book is the natural extension of the practical. In putting together the initial undergraduate course and preparing the background material for *BioActivity*, it became clear that there were no textbooks at a suitable level available to recommend as further reading. I therefore used the existing background material as the basis for an introductory book, which leads step by step to a worked example of the practical in the penultimate chapter. In the book there was therefore scope to develop the theories behind some of the practical concepts (which would have been tedious to spell out on the web), while the interactive practical gives life to the worked example in the book. In addition to allowing me to provide supplementary full colour illustrations (the book is 2-colour only), this arrangement allowed me to keep the theory within the practical relatively concise, and will hopefully encourage students to consult the textbook if further details are needed.

Conclusion

Computers are now commonplace in almost all areas of life, from work and commerce to the home. For current generations of students, exposure to computers both at school and at home is more or less taken for granted, and future generations are now playing the latest computer game before they are able to tie their shoelaces! Yet using computers effectively in teaching can still be a challenge. By integrating the solid security of a textbook with the virtual excitement of the web, in what I hope is a relevant practical approach, I have tried to avoid the sterility of my earliest CAL experience. I hope this will help to make bioinformatics a more accessible subject for biologists, and perhaps spur new generations of students on to MSc or PhD courses - one small step towards bridging the current skills gap.

References

1. Anderson, C. (1994) Cyberspace offers chance to do Virtually Real Science. *Science*, **264**, 900-901.
2. Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**(1), 387-395.

3. Attwood, T. K. (1997) BioActivity: An Interactive Bioinformatics Practical on the WWW. *Life Sciences Educational Computing*, **8**(3), 10-13.
4. Attwood, T. K. (1999) An interactive bioinformatics practical on the World Wide Web. In *Laboratory Work in Science Education: an Evaluation with Case Studies*. Ed. L. Jervis, SEED Publications, University of Plymouth.
5. Durham, A. (1995) Protein Knowledge at a Click. The Times Higher Education Supplement (February 10 1995) multimedia supplement p.v.
6. Attwood, T. K. and Parry-Smith, D. J. (1999) *Introduction to bioinformatics*. Addison Wesley Longman Ltd., Edinburgh Gate, Harlow, Essex, UK.

Teresa K. Attwood
School of Biological Sciences
The University of Manchester
Oxford Road
Manchester
M13 9PT
United Kingdom
attwood@bioinf.man.ac.uk
<http://www.bioinf.man.ac.uk/dbbrowser/bioactivity/prefacefrm.html>