TECHSIGHTING
SOFTWARE

Life Sciences Data Analysis

The Life Science Workbench is a software suite from MDL Information Systems that helps biologists create and track protocols, capture experimental results, analyze data, integrate results, and provide extensive functions for screening data. The Data Analysis Toolbox (DAT), reviewed here, is an add-in to Microsoft Excel that performs the curve-fitting and data analysis functions of the suite. The program contains more than 40 models of physical processes including common linear, nonlinear, logarithmic, power, and polynomial families of curves. The advantage of add-in programs such as this one is convenience, since the functions of Excel are known to most users. DAT was installed within seconds on a 166-MHz Pentium and integrated seamlessly with Excel (requires version 7.0 or newer for Office 95). The software places an "Analyze" feature, containing six functions, within the "Data" group on the tool bar.

Data Analysis Toolbox—Life Science Workbench

version 1.01

MDL Information Systems, Inc., 14600 Catalina Street, San Leandro, CA 94577, USA. Phone: (510) 352-3024; Fax: (510) 895-6092; Web site: www.mdli.com

Users may select data for an analysis set by defining criteria (filters), such as data position, numerical size of the data set, lack-of-fit, and data distribution. The position and spread of the data can be filtered by defining a set point, upper and lower bounds, or difference from an estimated response. DAT reacts when these limits are exceeded, but unfortunately, not in a very useful fashion. No warning box appears on the screen. The calculations proceed unhindered, and the analyst must search for the output tab that includes a long row of ones and zeros to view disqualified data. The immediate appearance of a warning box stating the problem would be helpful.

A Wizard assists in graph construction. In the first step, a data group is chosen from categories that include enzyme kinetics, ligand-receptor binding, exponential, linear, logarithmic, and polynomial. Users next choose a specific model within a data group. For example, models available from the Michaelis-Menten subgroup include steady-state, random two-reactant system, or substrate plus inhibitor. One can display the equation of the final, selected model and set a weighting option for the data.

Subsequent steps define set-up parameters for input data and data selection, which, as in other Excel routines, may involve clicking cells or typing into a dialog box. Step four permits selection of the desired statistics: parameter estimates, standard error, confidence values, and analysis of variance tables. Finally, graph type is specified. An experienced user can complete the process in less than a minute and use Excel's functions to format output. After the graphics are completed, a "Back Calculate" feature generates revised output values for all input values, even those out of the original range. The "Preferences" box sets the value for convergence, number format, and confidence level.

A registration booklet and a user's guide are provided with DAT. Although the latter contains sections that are not completely clear, the examples are useful. The only noteworthy drawbacks to DAT are an unwieldy visual statistical output (stretched out over many columns rather than in a compact table) and the lack of P values for the F-test. The P value is the statistician's prized comparative metric, so this omission will not only be a problem to the novice user, who may not know how to calculate it, but a nuisance to the more experienced analyst, who comes to expect it.

—JOHN WASS

Abbott Laboratories, 200 Abbott Park Road, Abbott Park, IL 60045, USA. E-mail: john.wass@add.ssw.abbott.com

TECHSIGHTING
NET TIP

Digital Security, Part I

By now, you have probably bought into the idea of using the Internet for e-mail, database searching, and surfing the World Wide Web. You may have even purchased items online and may subscribe to a journal or two in digital form. But is the Internet secure? If you don't know about security technology, the answer is no. If you read a little further, it could easily be yes.

All items sent by e-mail without encryption are publicly accessible documents. Systems administrators can easily view all your outgoing and incoming mail without any prior approval. In fact, this information can be used against you in a court of law should you be subpoenaed. As seen in the recent Microsoft trial, Bill Gates' "private" e-mail messages are providing useful fodder for the prosecution. If you want to keep your e-mail messages from prying eyes, you must encrypt the messages before sending them. The message recipient must then be able to decrypt the message to restore it to readable form.

Encryption of e-mail can be done either by symmetric key or asymmetric key encoding. In the first case, the sender and recipient both use the same key to scramble and unscramble the message. By a simple analogy, say the message is simply the number "7." If the key is the product of the original message content and the number "3," then the scrambled message is "21." If the cipher text "21" is sent to someone who has the key, then the original message "7" can be decrypted. Symmetric key cryptography can be used for complex documents, but the sender and recipient must first exchange the shared key. The problem of exchanging the key safely makes symmetric cryptography difficult to use on the Web.

Asymmetric cryptography (also called public key cryptography) is the system that you will most likely use to send secure e-mail. It is, for example, the technology behind Pretty Good Privacy (PGP), a commonly used commercial product for securing files and e-mail. Public key encryption works by taking advantage of special mathematical formulas in which the forward and reverse encoding can be performed by different keys. One key is called the private key and is used only by the owner of the key. Private keys are never distributed. The other key, called the public key, can be used by anyone. In fact, the public key is usually published in a Web directory for anyone to obtain and use.

Asymmetric keys have many uses, from sending secure documents by e-mail between two users to enabling entire companies to interact with remote sales operations. As an example, say that person A wants to send a secure document to person B. First, both A and B obtain a pair of keys, one private and one public. Each could then publish their public keys or exchange them via the Web. Person A uses B's public key to encrypt the document and then sends the cipher text to B. When B receives the message, she uses her private key to decrypt it. Because B is the only one who has her private key, the document cannot be read by anyone else during Web transit. If the document were to be intercepted, only a scrambled message could be viewed.

Even this scheme is too simple for most situations that arise on the Web where A

may not even know B. Also, because Web identitites do not include faces or voices, how would A know that B's public key is really from B? An imposter can set up a Web site and distribute a false B public key. In practice, the public key is usually turned over to a third party, called a certificate authority, before it is distributed. The certificate authority must authenticate the identity of the user behind the public key. The resulting key is modified and is called a digital certificate. For the highest security, this could mean that the user must present the key to a registered professional in person for notarization.

In next month's column, we will explain how to acquire and use a digital certificate with today's browsers.

—ROBERT SIKORSKI AND RICHARD PETERS

TECHSIGHTING
CLONING

Cloning 3.0?

If cloning technologies were software releases, you could say that two major versions have been created to date. Version 1.0 would probably include the basic techniques of cloning: using plasmids, restriction enzymes, and so on. Intermediate releases, say 1.1 to 1.5 came about with the improvement of vectors to detect inserts and the conversion of lambda clones. Version 2.0 came into being when the polymerase chain reaction (PCR) was developed in the late 1980s. PCR technology provided methods to better refine the DNA molecules to be inserted. It allowed novel molecules to be created in vitro that would be difficult or impossible to produce any other way. What, then, can we expect to find in the next "release" of cloning technology? A group led by Steve Elledge at Baylor College of Medicine (*1*) may have just answered this question and, in the process, ushered in a novel set of molecular technologies for the postgenome era.

One of the major tasks for a molecular biologist is moving known pieces of DNA from one cloning vector to another to create various fusions and clones. In the process, a researcher may use PCR to design joining segments, internal mutations, and other elements. All of these version 2.0 tasks rely on careful, custom construction by the genetic engineer. Many common DNA constructs (reporter fusions, GST fusions, *myc*-tag fusions) could be made more easily if a standard set of plasmids and manipulations were available. However, moving passenger DNA between template plasmids by restriction digests requires detailed knowledge of the passenger DNA, which is not always available. Also, as whole-genome research (for example, the Human Genome Project) progresses, cloning techniques to manipulate entire genomes become necessary.

The method proposed by Q. Liu *et al.* (*1*), the Univector Plasmid-fusion System (UPS), generates a new approach to cloning. Basically, they devised a highly scalable system in which site-specific recombination is used to rapidly generate new fusion proteins. Instead of using restriction enzymes and DNA ligase to cut and paste DNA molecules, they used Cre recombinase, a protein that can catalyze the joining of two sequences that each have a defined 34–base pair (bp) sequence called loxP.

They first intoduced their gene into the Univector (pUNI) in the form loxP-GENE-Kanr-Ori. The gene is placed in-frame with the open reading frame of loxP. Next, they made a collection of pHOST vectors that had a DNA sequence in the form Promoter-Tag-loxP-Ampr-Ori. They then mixed a pUNI clone in vitro with a pHOST plasmid and Cre protein. The result was a highly efficient, 20-min recombination reaction that produced a fusion plasmid joined at the loxP sites. The reading frame of the loxP site on pHOST, pUNI, and the Tag sequences are the same, so that fusion of pUNI and pHOST results in the creation of both a promoter-gene fusion and a protein-protein fusion. Furthermore, through genetic trickery, they devised a selection process through which only fusion plasmids grow upon transformation. The fusion reaction is so efficient that they can transfer a whole library in pUNI into a different vector without loss of representation.

Thus, with a simple, systematic, and easily reproduced in vitro reaction, they were able to generate a large collection of gene fusions with little knowledge of the chosen gene's sequence. Once a gene is inserted properly into pUNI, there is no more need to worry about its sequence, its reading frame, or the recipient vector. There is also no need for more fragment isolations or ligations; only a 20-min reaction is required.

This technique now allows the systematic manipulation of large gene sets or even complete sets of full-length complementary DNAs from a particular species (so-called "Unigene" sets). This capability will become even more important as we progress into the postgenome era of proteomics. The paper also contains a must-read collection of additional molecular biology tips and tricks that no graduate student should be without.

—ROBERT SIKORSKI AND RICHARD PETERS

References

1. Q. Liu, Z. M. Li, D. Leibham, D. Cortez, S. J. Elledge, *Curr. Biol.* **8**, 1300 (1998).

TECHSIGHTING
GENE SEQUENCING

Size Matters

Small changes in DNA sequences can have profound effects. Take a small 3–base pair (bp) deletion in the CTFR gene that leads to cystic fibrosis or a 1-bp substitution in the RAS oncogene that sends a defective signal in cancer cells. In fact, single nucleotide polymorphisms (SNPs) are now being used as genomic signposts in many applications in human genetics. Clearly, new methods to detect changes at defined genome regions are of interest to many scientists today.

Researchers have now devised a clever way to analyze small fragments of defined DNA by mass spectroscopy. The technique of electrospray ionization mass spectrometry has been used in the past to measure the molecular mass of small (less than 20 bp in length) oligonucleotides. The problem arises because most DNA specimens are likely to be derived from the polymerase chain reaction (PCR), and each oligonucleotide used in a PCR reaction can easily exceed 20 bp. PCR fragments are just too long for mass analysis by this technology. A technique to generate very small targets of a defined location in a genome is needed.

A group from Johns Hopkins (*1*) has come up with a new method called SOMA (short oligonucleotide mass analysis). The trick was to add a type IIS endonuclease cleavage site (Bpm I) into their PCR primers. Type IIS cut DNA adjacent to their recognition sequence. By appropriately spacing two new type IIS sites, a very small fragment can be digested out of any DNA target. As a test case, the authors were able to generate 7-bp sequences, which contained changes in a single base pair, from the human APC gene. The resultant small piece of DNA could then be interrogated unequivocally by electrospray ionization mass spectrometry.

—ROBERT SIKORSKI AND RICHARD PETERS

References

1. S. J. Laken *et al.*, *Nature Biotechnol.* **16**, 1352 (1998).

Tech.Sight is published in the third issue of each month. Contributing editors: Robert Sikorski and Richard Peters, Medsite Communications, Boston, MA, and Kevin Ahern, Department of Biochemistry and Biophysics, Oregon State University, Corvallis, OR. Send your comments by e-mail to techsight@aaas.org.