TECHSIGHTING
PROTEIN INTERACTION

Complex Grabbing

Much has been learned from experiments in which the main goal is to find the identities of proteins that interact with each other in complexes. Unexpected relations can point researchers down entirely new avenues of fertile experimentation. For example, knowing that your favorite protein binds to RNA polymerase II would likely lead you into the transcription field. After many celebrated successes, large-scale fishing expeditions to look for molecular binding partners are well accepted today. What tools can you use for these kinds of experiments?

Most scientists are familiar with the yeast two-hybrid screen for protein-protein interactions. In this assay, two plasmids that contain fusion proteins are introduced into yeast. One of these plasmids is designed to contain a library of unknown genes, and the other contains a gene of interest. The fusions are created in such a way that interactions between the partners can be detected in vivo in a selection assay. For example, an essential gene's transcription can be controlled by a promoter that is responsive to the fusion pair. Although highly successful, the two-hybrid assay is not without problems. First, because a transcription readout is essential, complexes that block transcription or do not enter the nucleus efficiently will never be recovered. Second, large overexpression of the partners may be lethal for some constructs. Third, even after such a screen, much work is still needed to sort through promising candidate partners to weed out the true polypeptides that interact in vivo.

New methods to supplant the yeast two-hybrid screen may now be brewing at the intersection of the field of genomics and small-scale protein sequencing by mass spectrometry. Given the explosion of genomic data, such as the full sequencing of the yeast genome, can one purify complexes in vitro and obtain the identity of interacting proteins by direct sequencing? The simple answer is yes, but isolating enough peptide at a sufficiently high purity is the challenging part. Even using high-affinity reagents like antibodies will often yield persistent contaminants.

A group from Germany (*1*) has devised a clever method for isolating extremely pure complexes of proteins from cell cultures that can be used for protein sequencing. Their procedure is called tandem affinity purification (TAP). The key concept behind TAP is to attach multiple affinity tags to a protein, which can then be used in its isolation.

They first chose to test the affinity of commonly used tags fused to the yeast SmX4p protein. This protein is known to associate with the U6 small nuclear RNA, a molecule that can be analyzed quickly by primer extension. They used the following tags: FLAG, two immunoglobulin G (IgG) binding units, protein A (ProtA) of *Staphylococcus aureus*, the Strep tag, the His tag, the calmodulin binding peptide (CBP), and the chitin binding domain (CBD). None of the tags impaired protein function, but only the ProtA and CBP tags provided efficient recovery of greater than 50% of the SmX4p protein from crude extracts. Next, these two tags were combined to make a high-affinity fusion peptide that could be added to any protein of interest.

In brief, they designed a special carboxyl-terminus fusion peptide (TAP tag) that contained the following components (in order): NH_2, ProtA, a spacer region, a cleavage site for the tobacco etch virus protease, a spacer region, CMB, and COOH. The idea was to express fusions of this peptide and several genes of interest and express them in the yeast *Saccharomyces cerevisiae*. Using immobilized IgG, which binds tightly to the ProtA tag, they performed the first round of isolation from crude extracts. Next, they used the TEV protease to cleave the complexes back into solution. By adding calmodulin and Ca^{2+} to the solution, they recaptured the complexes to a matrix and eluted the resulting polypeptide mixture with EGTA. After SDS–polyacrylamide gel electrophoresis isolation, the individual peptides were isolated and sequenced.

They performed several experiments to test the system outlined above. For one, they made a fusion of the TAP tag to the U1 small nuclear ribonucleoprotein (snRNP). Using the TAP protocol, they purified a complex of 11 proteins from starting material obtained from only 2 liters of yeast cell culture. The identity of all 11 proteins was found by mass spectrometry. They showed that known U1 snRNP subunits were purified in the TAP method, and they identified a new member of the complex, called Snu30p. By tagging the Snu30p protein gene, they definitively showed that it bound to the U1 snRNP subunit.

The TAP method is a powerful tool for identifying new members of protein complexes. The high affinity of the TAP tag means that very pure samples can be obtained from relatively small amounts of starting material. This method, coupled with the massive amount of sequence data produced from the Human Genome Project, will allow mass spectrometry to be used as a rapid method of identifying protein partners.

—ROBERT SIKORSKI AND RICHARD PETERS

References
1. G. Rigaut *et al. Nature Biotechnol.*, in press.

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

TECHSIGHTING
NET TIP

Lab Search Engine

Every lab amasses its own private collections of research protocols, documents that serve as the essential blueprints for day-to-day operation. These little tips and tricks are the stuff that makes that cell transformation work at high frequencies or those frog eggs grow two heads in an assay. Labs that focus on common protocols for all users can turn their enterprise into a well-oiled machine. New students, postdocs, or technicians can easily hit the ground running with the right collection of protocols from day one.

In the past, protocols were passed around in multiply-photocopied documents, which became illegible over time. Although the scribbling in the side margins added some information, a paper-based system is not the best choice. Paper protocols are useful, but they have serious problems with versioning (maintaining and navigating multiple versions of a given document) and searching. A paper protocol is difficult to maintain on a regular basis without creating a versioning nightmare. For example, suppose your group has decided that the Sorvall centrifugation of your new cell line should be done for twice as long as in the previous protocol for an assay. You would need to correct this one line of the protocol in every copy of that protocol that exists in the lab. If the protocol is not in a central location, you would create multiple versions of the protocol. Clearly, you could number the versions, but you would still have to notify everyone to change their versions. Also, aside from a simple index in a binder, it would be difficult to search for specific protocols.

The answer to these problems will be found when you go digital. Below we outline a relatively simple and inexpensive solution for creating a lab protocol knowledge base that uses commercial software.

First, you will need to start with a computer running Windows NT. The workstation version of NT is a bit easier to set up and maintain, but the server is preferable because it provides additional features you will need as you grow. Next, add the NT4 option pack that will include Microsoft Index Server and the Web server Microsoft Internet Information Server (IIS).

Next, configure IIS to run as a Web server on your local network so that each lab computer can access a common Web site. IIS is probably one of the easiest servers to run and maintain, so once it is up and running, you should not have to touch its settings again. Before you begin, you should have IP addresses and an IP scheme mapped out that will work with your current local network. With IIS, you can then set up directory security so users are restricted to certain directories. This may be important if you add any sensitive data to your intranet.

With a working Web server, you can now turn your attention to converting your protocols to digital form. You have two choices here: convert your protocols to HTML or leave them in a common file format, like MS Word. As you will see, Microsoft Index Server will search through HTML documents of a wide collection of file formats. Many tools exist, such as Allaire's HomeSite or Microsoft's FrontPage, that will guide you through the process of adding HTML tags to protocols. Alternatively, you could write your protocols in a Microsoft Office application, such as Word, and use the built-in automatic HTML converter.

Finally, install and configure Microsoft Index Server. This free tool works under the IIS Web Server to gather, index, and search an entire intranet site. Currently, Index Server will catalog all Microsoft Office documents (including PowerPoint presentations), HTML, and common word processing formats like WordPerfect. Index Server is essentially maintenance free once installed. When you add a new document to your lab intranet directory, Index Server will automatically index the file in its database. This means that, within a few seconds, the new document will be available to any searches on your site.

Searches of Index Server can be made through boiler plate scripts that come with the software. The searching language of Index Server is quite robust and will allow you to, for example, find all documents created within certain time ranges, do natural language queries, and perform complex Boolean queries. The results of your search will be displayed in a format similar to Web search engines, like HotBot or Lycos. The abstracting engine of Index Server will automatically create abstracts of your documents, making visual scanning of your knowledge base rapid. The title of your document in the results page will be hyperlinked to the document itself on the server. If the document is in HTML form, it will be displayed in the browser after you click on its link. If it is in another format, like Word, the click will launch Word on your PC and load it with the chosen document. Also, if you have security set up on directories, queries will only return documents for which a user has permission to access. This neat feature means that you can send search queries down your entire intranet server tree without worrying that sensitive material will be delivered to unintended users.

Although this column is not a cookbook for setting up a digital protocol manual on your own private intranet, hopefully, it will serve as an introduction to the topic. With some knowledge of Windows NT and some time spent on the Microsoft Web site, this project is one that any lab can tackle with a modest amount of effort.

—ROBERT SIKORSKI AND RICHARD PETERS

TECHSIGHTING SOFTWARE

Pull-Down Stats

Prescient by several decades, T. S. Eliot mused, "Where is the knowledge we have lost in information?" Finding meaning in the huge amounts of data available in virtually all areas of science requires statistical tools.

Happily for scientific spreadsheet users, the Numerical Analysis Group (NAG) of Oxford, UK, has ported their standard statistical analysis subroutines (previously available only as C and FORTRAN code libraries) to the menu bar of the Microsoft Excel spreadsheet. Powerful and sophisticated statistical analyses are within the reach of a point and click, placing powerful statistical analyses into the hands of everyday scientists who need it, who have the computer power to use it, and who may not have the time or the expertise to work at the level of software code.

Statistical Add-Ins for Excel
Numerical Algorithms Group
Downers Grove, IL.
630-971-2337
www.nag.com
www.nag.co.uk

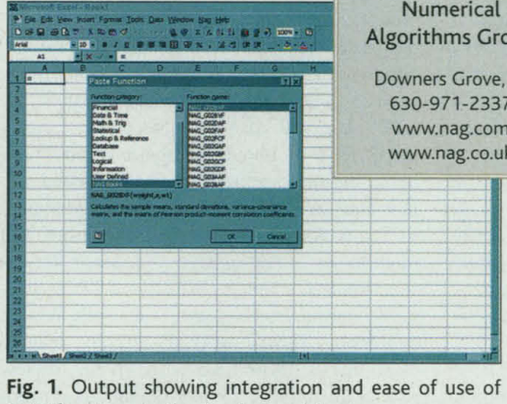

Fig. 1. Output showing integration and ease of use of NAG functions in the Excel spreadsheet.

After the NAG Statistical Add-In modules are installed, their use in statistical analysis is essentially transparent. As the name implies, they appear as an additional feature of the resident Excel spreadsheet program (Fig. 1). The NAG add-ins may be used within Excel 7.0 or Excel 97, running under Windows 95 or NT on a PC platform. There is no noticeable slowdown in the operation of the Excel program itself when the add-ins are present.

The add-in package consists of 51 routines in five categories: Statistical Modeling, Multivariate Methods, Analysis of Variance (ANOVA), Time Series Analysis, and Generalized Linear Models. On selecting the Excel Function Wizard, the user gains access to statistical routines, such as normal distribution; t distribution; χ^2 test; linear regression; correlation matrix; time series analysis; autocorrelation; seasonal autoregressive integrated moving average (ARIMA); ANOVA block, row, and factorial; confidence intervals; hierarchical clustering; and discriminants.

Operation is simple. The desired statistical function is chosen from a pull-down menu. Arguments are specified by designating the spreadsheet cells that contain the data. In this regard, it is just like using any of the other arithmetic functions that are part of Excel. Output, likewise, fills a designated set of cells. The use of these functions has been streamlined by requiring only problem-specific input data. Optional (default) arguments may be set as needed. Once one of the add-in functions is embedded in a spreadsheet problem, the output cells are updated, continuously and automatically, to reflect any changes in the input cells.

The software is available as a CD-ROM or by downloading a 4.5-Mb compressed .zip file from the NAG Web site. Installation proceeds easily with an Installation Wizard included in the setup file. Complete installation requires a software key, provided by NAG to registered users. Ease of use of the software is augmented by extensive online help files, which can be consulted while Excel is running. The online help is searchable by keywords as well as by index. In addition, NAG has provided an animated PowerPoint demonstration tutorial that shows how to access and use the add-in functions, together with an example spreadsheet.

The user-friendliness and the power of this software make it a sure winner. It is highly recommended for any spreadsheet user who employs statistics in everyday scientific work. The reasonable price makes it accessible to almost every level of user, including students. —EMILE M. BELLOTT

Pharm-Eco Laboratories, Lexington, MA 02421, USA.
E-mail: ebellott@pharmeco.com