FUNCTIONAL GENOMICS

Expression Cloning in the Test Tube

Randall W. King,* Kevin D. Lustig, P. Todd Stukenberg, Thomas J. McGarry, Marc W. Kirschner

TECH VIEW

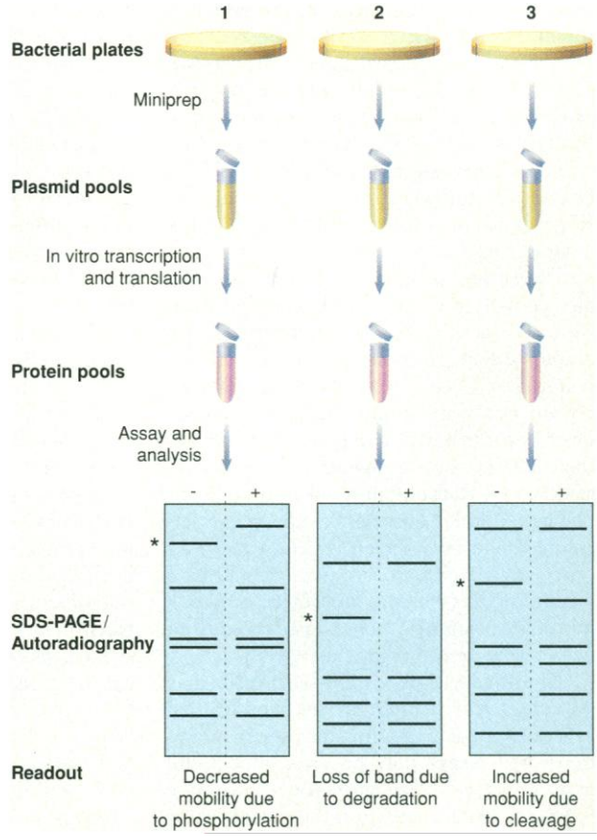
Traditional biochemical techniques such as protein purification are powerful tools for identifying proteins on the basis of their function. Yet, isolating the genes that encode these proteins is often cumbersome, requiring multiple steps of purification, peptide sequencing, and subsequent isolation of complementary DNA (cDNA). An alternative approach to functional gene identification, which perhaps could be called "reverse" biochemistry, is expression cloning, in which cDNAs are first translated in prokaryotic or eukaryotic cells and then assayed for a specific biologic activity. This approach circumvents the requirement for protein purification, yet it retains the use of a functional assay as the basis of cDNA identification. Cell-based expression cloning strategies, which use mammalian cell lines or *Xenopus laevis* oocytes to express encoded proteins, have identified genes encoding membrane receptors, transmembrane channels, and secreted growth factors, but they have not been widely used to identify intracellular activities. Here we discuss a new cell-free expression cloning approach (*1*) that substantially expands the range of biochemical assays that can be used to identify a cDNA on the basis of its function.

The cloning strategy is straightforward [see figure and (*2*, *3*) for details]. Rather than expressing cDNAs in intact cells, where the encoded proteins are not freely accessible, small plasmid pools (~50 to 100 clones per pool) are instead expressed in vitro in a single step by means of a coupled transcription-translation system (the Promega TNT system). Although this system is optimized to express single proteins, we found, surprisingly, that about 30 proteins can be expressed in a single reaction. Each protein pool is then screened for the presence of a desired biochemical activity, and positive pools are progressively subdivided until a single cDNA encoding the active protein is isolated. In principle, in vitro expression cloning (IVEC) can be used to rapidly identify cDNAs on the basis of almost any biochemical property for which an assay can be developed.

Many cellular processes are regulated by posttranslational protein modification, yet there are no general techniques that can identify the substrates of modifying enzymes on the basis of this property. By screening small pools of radiolabeled proteins, we have been able to systematically identify substrates of enzymes such as kinases and proteases. In a typical assay, a kinase or protease activity, either present in a crude extract or in purified form, is incubated with a radiolabeled protein pool, which is then resolved by SDS–polyacrylamide gel electrophoresis (SDS-PAGE; see figure). Because posttranslational modifications such as phosphorylation or proteolysis often change the electrophoretic mobility of a protein, pools containing a candidate substrate are identified by scoring for proteins whose mobility is altered by incubation with a given enzyme.

We applied this approach to identify proteins that become phosphorylated during mitosis (*3*). Pools of radiolabeled proteins were incubated with either *Xenopus* interphase or mitotic extracts, and candidate substrates were identified by screening for proteins whose electrophoretic mobility was reduced by treatment with the mitotic extract (panel 1 of figure). In a complementary screen, a monoclonal antibody that recognizes mitotic phosphoproteins was used to identify substrates that could be immunoprecipitated only after treatment with mitotic extract. After screening 12,000 proteins, we identified 20 distinct mitosis-specific phosphoproteins, including five that share no sequence similarity with known proteins.

We developed a similar functional screen to identify proteins that are degraded by the ubiquitin system as cells exit mitosis (*2*, *4*). Candidate substrates were identified by screening pools for proteins that were de-



The strategy of in vitro expression cloning. An unamplified cDNA expression library is plated at a density of approximately 100 clones per bacterial plate. Pooled plasmid DNA is obtained by scraping colonies from each plate and performing a small-scale plasmid purification. Each plasmid pool is then transcribed and translated in vitro with a commercially available system, such as the TNT system from Promega. The resulting protein pool is then assayed for the presence of an activity. In the illustrated experiment, a radioactive amino acid is included in the translation system to specifically label the pool of proteins. Incubation of a pool with a modifying enzyme (lanes labeled +) such as a protease or kinase can result in a change in mobility of a substrate (bands marked with asterisk). Pool 1 contains a protein whose mobility is reduced following treatment with a kinase; Pool 2 contains a protein that is degraded following treatment with an extract containing an activated proteolytic system; Pool 3 contains a protein that is specifically cleaved following treatment with a protease, decreasing its apparent molecular mass. Once a pool containing a candidate activity is identified, the original cDNA pool is subdivided and retested until the single cDNA encoding the protein of interest is isolated.

R. W. King, P. T. Stukenberg, T. J. McGarry, and M. W. Kirschner are in the Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA. K. D. Lustig is at Tularik Inc., Department of Biology, South San Francisco, CA 94080, USA. E-mail: lustig@tularik.com

*To whom correspondence should be addressed. E-mail: king@bcmp.med.harvard.edu

graded following incubation in a *Xenopus* mitotic egg extract (panel 2 of the figure). In addition to identifying several forms of cyclin B, as expected, this screen has identified a novel protein that shares little sequence similarity with cyclin B, with the exception of a small motif that is required for ubiquitin-dependent degradation.

Proteolysis of a different sort regulates the process of apoptosis, or programmed cell death. To screen for substrates that are specifically cleaved during apoptosis, pools of labeled proteins were incubated with extracts derived from apoptotic cells, and assayed for the disappearance of a labeled band or the appearance of a proteolytic fragment (*2*, *5*) (panel 3 of figure). In a similar screen for caspase substrates, purified proteases have been used in lieu of apoptotic extracts (*6*). These approaches have identified several candidate substrates whose roles in executing apoptosis are now being studied.

Although IVEC has successfully identified proteins that are posttranslationally modified, its potential applications are much broader, for it can be employed in screens that detect physical interactions between pool proteins and a wide variety of "bait" molecules. For example, to screen for transcription factors capable of binding specific DNA elements, unlabeled protein pools were incubated with a labeled DNA fragment and then assayed for protein-DNA binding by use of a gel mobility shift assay (*2*, *7*). Although this method has only been used to identify proteins that bind to "naked" DNA, in principle the bait could be a preassembled complex of DNA bound to known transcription factors. RNA-binding proteins could be isolated in a similar fashion.

Direct protein-protein interactions can also be used as a basis for cDNA identification. To identify binding partners, for example, a bait protein could be covalently coupled to a bead, mixed with radiolabeled protein pools, and interacting proteins reisolated by centrifugation. An important advantage of IVEC over conventional two-hybrid screening is that protein pools can be modified with purified enzymes or cellular extracts before the affinity screen is conducted. Another advantage is that the bait could be a large complex (such as a multimeric enzyme), a protein machine (such as a ribosome), or even a virus or cell organelle.

By using an antibody as a bait, IVEC may provide a useful alternative to traditional λ-gt11 expression cloning. Immunoprecipitation-based assays may be especially useful when an antibody recognizes only the native form of a protein, or when the protein must be modified to be recognized. For example, a screen could be designed to identify tyrosine-phosphorylated proteins recognized by anti-phosphotyrosine antibodies.

Many complex cellular processes can now be reconstituted in vitro, extending the capabilities of IVEC beyond simple interaction-based screens. The process of secretion can be modeled in vitro by supplementing the translation system with purified microsomes. A screen for secreted proteins could therefore be developed, either by using a protease protection assay, or by scoring for changes in electrophoretic mobility that occur as a consequence of the microsome-dependent cleavage of a signal sequence. The incorporation of an epitope or chemical tag in the expressed proteins could also enable the design of assays that identify proteins on the basis of localization to other cellular structures. Pools of biotinylated proteins could be incubated with complex cellular extracts, and incorporation into nuclei or centrosomes could be monitored with streptavidin-coupled fluorophores and microscopy.

Enzyme-activity–based screens are another unexplored but potentially exciting application of IVEC. The assays would be similar to those used in standard biochemical approaches, with the exception that protein pools rather than chromatographic fractions would be tested for activity. For example, protease cDNAs could be cloned by testing unlabeled protein pools for the ability to cleave a known substrate. This approach may also expedite the identification of new "extremophilic" enzymes that operate at extremes of temperature, pH, or ionic strength.

Enzyme activity screens could also be used to identify mutant enzymes with novel properties. To identify mutant enzymes that are insensitive to an inhibitor, small pools of mutagenized proteases could be constructed and screened for activity in the presence of the inhibitor. This approach may facilitate the identification of mutations that lead to drug resistance, or, using sequential rounds of mutagenesis and screening, enable investigators to evolve enzymes with activity toward novel substrates.

Although IVEC dramatically expands the types of assays that can be used as a basis for expression cloning, it has several limitations. First, IVEC is unlikely to succeed when multiple proteins are required for activity, as the likelihood of expressing multiple subunits in the same pool is low. If the putative binding partners are known, it may be possible to supply them as recombinant proteins or fractionated extracts. Second, for activity-based approaches to be successful, the expression system must be devoid of the activity being assayed. We have only used reticulocyte lysates to express cDNA pools; alternative expression systems, such as wheat-germ lysates, may have different background activities. Finally, the concentration of expressed proteins is low relative to the protein concentration of the lysate. Although protein labeling facilitates the identification of substrates, the detection of enzymatic or binding activities may require the expression of greater amounts of protein. The use of smaller cDNA pools increases the abundance of each protein, but also makes the screening process more time-consuming. Alternatively, it may be possible to isolate and concentrate the expressed proteins, either by constructing epitope-tagged cDNA libraries or by incorporating biotinylated amino acids to specifically tag newly expressed proteins.

IVEC is feasible in small laboratories because DNA preparation and protein expression technologies have become rapid and reproducible. The initiation of saturating, genome-wide screening, however, will depend on automating the process. Fortunately, the same robotic systems that have been developed for genome sequencing and high-throughput drug screening should be easily adaptable to the steps of cDNA pool preparation and in vitro translation, and, in many cases, to the biochemical assay itself. In the not-so-distant future, the availability of completely normalized libraries, in which each cDNA found within the genome is represented once, will provide the ultimate resource for IVEC. This is currently feasible in *Saccharomyces cerevisiae*, in which the sequence of each of its ~6200 potential genes is known. These cDNAs could be expressed and tested individually, or pooled together to expedite assays.

IVEC is a hybrid technique, containing elements of both genetics and biochemistry, that has enabled us to ask new questions that could not be addressed with either approach alone. Whereas cell-based expression cloning technologies have excelled in the identification of secreted growth factors and their receptors, IVEC now makes it possible to identify proteins that are active in the cytoplasm and nucleus. Within the coming decade, the human genome will be completely sequenced. Expression cloning techniques, using both cell-based and in vitro expression systems, will be important tools for characterizing the tens of thousands of genes whose sequences will be available, yet whose functions remain obscure.

References

1. R. W. King, K. D. Lustig, P. T. Stukenberg, M. W. Kirschner, U.S. Patent 5,654,150 (1997).
2. K. D. Lustig *et al.*, *Methods Enzymol.* **283**, 83 (1997).
3. P. T. Stukenberg *et al.*, *Curr. Biol.* **7**, 338 (1997).
4. T. J. McGarry and M. W. Kirschner, unpublished data.
5. V. Cryns and J. Yuan, personal communication.
6. D. Xue and H. R. Horvitz, personal communication.
7. P. Mead and L. Zon, personal communication.

TechWire Forum:
www.sciencemag.org/dmail.cgi?53281