and this is reflected in the characteristic shape of the resulting graph. The most useful descriptors were found to be a 2D structural fingerprint of the molecule's side chains, and two novel topomeric fields (shape descriptors) known as steric CoMFA (comparative molecular field analysis) fields and H-bonding CoMFA fields[2,3]. Random numbers were used as an example of an invalid descriptor, and gave diffuse plots with no distribution enhancement[2].

Diversity and similarity

In lead discovery it is most efficient to test compounds whose neighbourhood regions do not overlap. However, lead exploration requires the testing of compounds in the same neighbourhood region as the lead. The application of ChemSpace to this problem was addressed in a trial carried out by Tripos and Bristol-Myers Squibb[4]. The structures of four known drugs in the 'sartan' class of angiotensin II (A-II) receptor-antagonists were used as the 'hits'. The aim was to see whether structures identified as their close neighbours were at least as likely to be active as those selected by chemists using traditional concepts of similarity. An A-II-targeted virtual library of 2.6 billion compounds was set up and searched using similarity criteria based on Tripos's topomeric shape descriptor. The 63 closest compounds selected by ChemSpace using similarity criteria were synthesized, as were a further 362 using conventional selection criteria. All were tested for inhibition of A-II receptor binding. Seven compounds were found to be highly active; all were from the group of 63 selected by ChemSpace ($p<0.001$).[4]

According to Tripos, the technology has already successfully achieved results in ongoing projects. For example, a drug discovery collaboration with Arena Pharmaceuticals (San Diego, CA, USA) for novel drugs targeting G-protein-coupled receptors, moved from inception to lead series in under three months and then from lead optimization to *in vivo* biology in less than a further five months.

The future

'The topomer approach seems an excellent way to estimate the similarities of molecules in a biological setting,' says Robert Glen, Professor of Molecular Sciences Informatics at the University of Cambridge, UK. 'In library design and lead follow-up this is undoubtedly an extremely useful idea generator. The most urgent direction for future research is in generating better descriptors, in particular for ADME/Tox.'

Tripos says that the method has been extended to visualize the distribution of compounds in these thousand-dimensional descriptor spaces and to analyze HTS data. It is also being applied in preliminary studies of molecular toxicology and differential protein expression. Continued development of better and more specific descriptors is an ongoing area of research.

References

1 Cramer, R.D. *et al.* [Tripos] (2001) Method for selecting an optimally diverse library of small molecules based on validated molecular structure descriptors. US6185506

2 Patterson, D.E. *et al.* (1996) Neighborhood behaviour: a useful concept for validation of 'molecular diversity' descriptors. *J. Med. Chem.* 39, 3049–3059

3 Cramer, R.D. *et al.* (1996) Biosisosterism as a molecular diversity descriptor: steric fields of single 'topomeric' conformers. *J. Med. Chem.* 39, 3060–3069

4 Cramer, R.D. *et al.* (1999) Prospective identification of biologically active structures by topomer shape similarity searching. *J. Med. Chem.* 42, 3919–3933

Reversing the paradigm for HT protein identification and validation

Rebecca N. Lawrence, Supplements Editor

A new high-throughput technology that promises to rapidly identify and validate new protein drug candidates and targets has been awarded broad patent protection[1]. The technology, developed by Cytos Biotechnology AG (Zurich, Switzerland), is an integrated system for the functional expression of all proteins from a given cell or tissue and subsequent HTS of these proteins for a desired function.

The recent production of the draft sequences of the human genome has led to an explosion in the number of genes discovered. However, the identification of novel functions of genes and proteins has been unable to keep pace due to a lack of appropriate screening technologies. Rapid production of the corresponding purified proteins and effective vectors for expressing these genes in cells or animal models is also necessary for the successful exploitation of the therapeutic potential of the human genome.

Current approaches to identifying gene function start from the gene sequence and, through an elaborate process of *in vitro* and *in vivo* assays, eventually lead to elucidation of the gene function, but this process can sometimes take up to two years. The new technology developed by Cytos works by reversing this paradigm. Claudine Blaser,

1359-6446/01/$ – see front matter PII: S1471-1931(01)00157-4

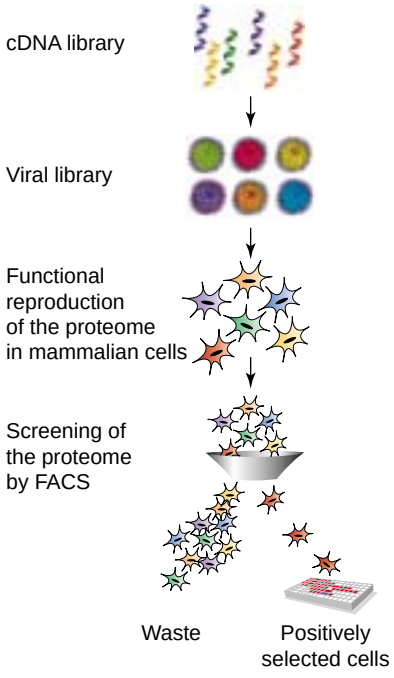

Figure 1. The one gene per cell approach. A culture of suspended single cells is infected with the alphaviral expression library, resulting in high-level expression of only one gene per cell. Single cells expressing a desired phenotype that result from the exogenous gene are sorted in a fluorescence activated cell sorter (FACS). This system can enable cell subset analysis according to a set of up to seven specified cell parameters at any time.

Director of Business Development at Cytos, says 'The desired function is chosen first, then all proteins from a given tissue are reproduced and screened for the desired function. From there, we have immediate access to the gene sequence.' The new technology can be used to identify, characterize and isolate nucleic acids encoding proteins of a desired property (including binding specificity, enzymatic activity, induction of proliferation, differentiation, apoptosis and others) and then provide rapid expression of such proteins or glycoproteins in mammalian cells.

How the technology works

The tissue of interest, whether it be diseased or healthy tissue, is chosen first and then a cDNA library is isolated from the tissue. Blaser highlights the significance of getting this step right: 'You can only pull out relevant proteins at the end if you start with a good cDNA library, so we take a lot of care in making sure we do this step as accurately as possible.' The cDNA originating from the tissue is then transformed into an infectious viral library through the use of alpha viruses so that each virus carries a single gene from the original RNA. These viruses are then used to infect mammalian cells such that each virus infects a single cell. 'Hence, every cell produces one protein originating from the cDNA library and one type of infectious virus, and this is very important,' says Blaser. The mammalian cell culture expressing the entire proteome of the chosen tissue can then be screened for specific protein functions.

One gene per cell approach

One use for the technology is for HTS using FACS (fluorescence-activated cell sorting), an approach that is most appropriate for e.g. finding natural receptors for known protein ligands or small organic molecules. The ligand is labelled with a fluorescent marker and then added to the cell pool. Those cells that then bind to the ligand can be identified and isolated from the pool. Blaser points out that: 'Because the cell also produces infectious virus particles, the virus can be easily isolated and amplified. This means that the viral RNA can be isolated and the gene identified very quickly.' See Fig. 1.

This method is also appropriate for several other types of studies such as in screening for activation of a specific signalling pathway or induction of apoptosis. Furthermore, this method can be performed in high-throughput. 'This technology can screen 10^7 cells per hour, which means that only a short time period is required to screen the whole proteome,' says Blaser.

One gene per well approach

Another use for the technology is through screening one gene per well in a 96-well plate. This method is most appropriate when screening for secreted molecules. Mammalian cells are seeded into the wells and then each well is infected with one virus which amplifies and infects all cells. Proteins that are secreted into the supernatant can then be assayed for their activity on specific target cells. Activity is measured by induction of proliferation, apoptosis, chemotaxis or expression of specific marker proteins. Again, because virus particles are also produced, it is a simple process to isolate the virus and then get back to the gene. See Fig. 2.

One gene per plaque approach

The third use for this technology is through using a one gene per plaque approach, which is suitable when screening for secreted, membrane-bound or intracellular proteins. The viruses used here are lytic and can cause cell death within a couple of days. Cells are seeded on a petri dish and, following infection with the virus, will form plaques after 2–3 days. The cell layer is overlaid with agarose followed by a nitrocellulose membrane, which attracts the proteins produced by the cells. The nitrocellulose layer can then be incubated with specific substrates to identify e.g. new enzymes. This approach can also be used to find new secreted ligands for known receptors.

Advantages of the technique

This new technique offers the advantages of one-step direct cloning of full-length cDNA and enables rapid expression of the protein of interest without subcloning or establishment of a production cell line. Blaser says: 'You end up being able to rapidly identify the virus in each case, which leads you straight back to the gene sequence. You also already have a viral expression vector and a lot of protein produced in a small-scale cell culture with which to do further studies. Furthermore, as you define at the beginning of the study what particular activity you are screening for, you already know the context of the gene sequence you find at the end of the process. You therefore do not have any genes in your sequence that do not play a role in that desired function.'

Other advantages include the fact that the products can be expressed in mammalian cells ensuring the production of correctly folded and glycosylated proteins, and the ability to rapidly purify the material without explicit knowledge of the nature of the protein. 'This whole process can significantly speed up the drug discovery process. Instead of taking sometimes up to 2 years to assign the function to a given gene sequence, this technology means that it takes only about 3 months to screen for proteins displaying a desired function', she adds.

Current and future uses

Cytos is currently using the technology in-house for the discovery of novel drug candidates in four therapeutic areas (inflammation, allergy, cancer and Alzheimer's disease). 'We have already identified some candidates in inflammation and cancer. Although it is hard to predict, we are hoping to have some of these candidates in preclinical development in the next year', says Blaser.

The company is also using the technology to offer a research service to large pharmaceutical companies. 'We prefer to do the research ourselves rather then licence out our technology as we can then provide the best possible service to our clients and remain independent to use the strategic advantage of our technology for our in-house product development,' she adds.

Cytos is now trying to develop the technology to further improve the screening assays and to increase the throughput by using greater automation. 'As the screening assays require enough cells or supernatant to be able to do the screening, increases in throughput are likely to be through increased automation rather than through moving into more miniaturized formats', says Blaser. 'Our other main aim is to use the technological advantage we now have to identify our own protein molecules and develop our own novel drugs.'

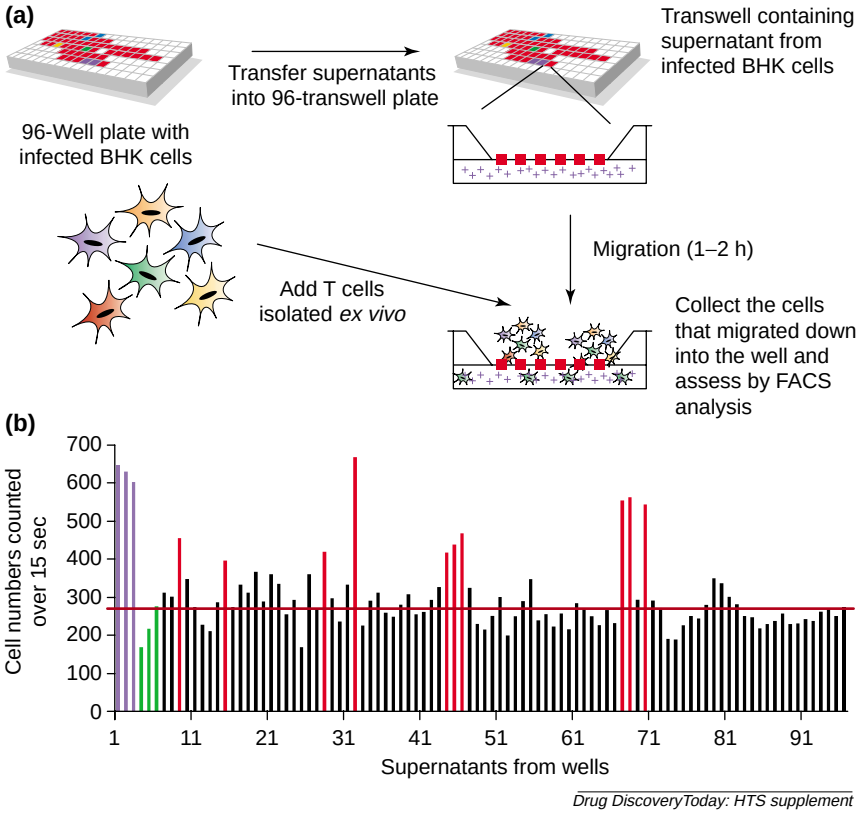

Figure 2. The one gene per well approach. (a) An alphavirus expression library was constructed from spleen mRNA, and ten viral particles per well were incubated on a confluent BHK cell layer in 96-well plates. Supernatants containing secreted proteins encoded by the recombinant viruses were harvested and used to measure their ability to induce migration of *ex vivo*-activated T cells. Migratory T cells were assessed in a 96-well-based transmigration system, and viruses from wells containing supernatants with migration-inducing proteins were subcloned, re-tested and plaque purified. (b) Cells were incubated with supernatants containing MCP-1 as a positive control (purple) and with supernatants from cells infected with wild-type virus as a negative control (green). Bars above the red line indicate supernatants from wells that have induced true migration. Abbreviation: FACS, fluorescence activated cell sorter.

Reference

1 Renner, W.A. *et al.* [Cytos Biotechnology] Expression cloning processes for the discovery, characterization, and isolation of genes encoding polypeptides with a predetermined property. US6197502