Chemistry & Biology, Vol. 10, 783–786, September, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2003.09.003

Previews

Bacteriophage that Display Small Molecules

Technology has been developed to display small molecules on phage particles. This innovation enables the generation of libraries of phage-tagged compounds with novel properties that are well suited for in vivo assays.

The increasing importance of high-throughput screening (HTS) in drug discovery has intensified the demand for streamlined screening methods. For example, HTS protocols would be simplified, could be implemented more efficiently, and be less expensive to run if libraries were composed of distinguishable members, screenable as a mixed population. To this end, ingenious approaches have been adopted, such as designing libraries of compounds with distinct individual molecular weights and tagging the bead support on which compounds are synthesized with DNA labels [1, 2]. However compounds bearing unique identifiers still must be isolated in sufficient quantities for identification, a process that may be difficult for in vivo screens. Phage display is a robust method for generating diverse libraries of identifiable proteins and peptides tethered to the coat of either filamentous or icosahedral phage [3]. The major limitation of phage display is that, with the exception of selenopeptides [4], the libraries have been restricted to members composed of the 20 natural amino acids. In this issue of *Chemistry & Biology*, scientists from XenoPort present new methodology that combines the structural diversity of chemical libraries with phage display's physical combination phenotype and genotype, as illustrated in Figure 1 [5].

The authors show that chemical libraries can be efficiently attached to the protein coat of phage particles using hydrazone chemistry. The unique DNA sequence securely packed within the phage particles acts as an identification tag for the small molecule attached to its surface. The individual small molecules must be kept separate before attachment to the phage to ensure that unique combinations of compound and phage particle are produced. A "master" array of the DNA from all the phage in the library produces a key to indicate which small molecule is associated with which phage particle. Phage retrieved from a chemical screen are rapidly amplified in *E. coli*. Comparing the DNA sequence from the amplified phage and the master DNA array enables the easy identification of the small molecule selected in the HTS.

Displaying small molecules using the phage format has several advantages. First, phage are highly soluble in aqueous solution; concentrations of up to 10^{13} particles per milliliter can be achieved, and therefore screens can be carried out in homogeneous solution. Second, by regulating the reaction conditions, both the number of compounds bound per phage and the correlated avidity can be rationally controlled. Third, the infection yield of bacteria by phage is close to 100%, allowing the detection and the exponential amplification of each unique phage particle that is isolated in minute amounts. Finally, phage are stable, and the outer coat efficiently protects the DNA within the core of the particle from degradation and prevents the nucleic acid from interfering with the chemical screen.

The authors demonstrate the utility of their method in a series of experiments. Initially, a small amount of fluorescein-labeled T7 phage is added to two populations of undecorated phage at levels of 1 in 10^4 and 1 in 10^6, respectively. Each of the two mixed phage populations was then sorted with an anti-fluorescein antibody to select for phage that is labeled with the fluorescein tag. In each case, 90% of the phage recovered from antibody screening was fluorescein labeled, signifying a significant enrichment in both cases. An undecorated phage library was then "spiked" with phage labeled with one of four different fluorophores to a final ratio of 1 to 10^4 labeled to unlabeled phage. Sorting the pooled library against four different antibodies, each of which specifically binds one of the fluorophores, resulted in exclusive isolation of phage with the corresponding fluorophore's antigen attached. The results of these phage-panning experiments mirrored the selectivity of the antibodies used in the screen, showing that the display of the fluorophores on T7 phage does not perturb the interaction between the antigen and antibody.

The most important point demonstrated in the paper is that the stability, size, and shape of phage particles facilitate their use in cell-based assay and in vivo formats [6]. The XenoPort group makes use of their technique to select for compounds that assist the internalization of the whole phage particle by KB cells, a human nasopharyngeal carcinoma cell line that expresses the folate receptor at high levels. Previous observations suggested that molecules bound to the folic acid receptor are internalized via endocytosis, so it was believed phage displayed compounds that are tightly bound to the folic acid receptor would be more efficiently internalized [7]. A 980-member library of folate analogs attached to 10^{10} T7 phage particles was incubated with KB cells. After washing the cells to remove excess phage, the internalized phage particles were recovered. The recovered phage was amplified in *E. coli* and screened with radiolabeled nucleic acid probes representing the phage DNA. The DNA from 14 of the 20 clones hybridized strongly to phage clones that carried folate analogs with high affinity for the folate receptor. The recovered compounds were enriched by 10- to 30-fold compared to their starting concentration in the original library. Other compounds that were enriched by the screen were later shown to be false positives, a ubiquitous problem in HTS.

Woiwode et al. give a detailed account of each step in the screening process from phage display vectors to

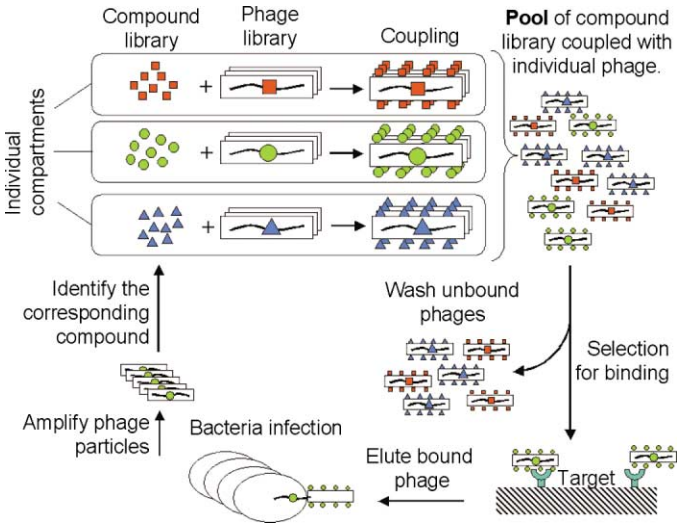

Figure 1. Diagram Illustrating the Formation, Screening, and Analysis of a Small Molecule Library Displayed on Phage

synthesis of the folic acid derivatives. The data analysis is straightforward and easy to follow. In summary, this report presents a compelling case for the utility of the methodology, especially for selecting molecules engaged in cell targeting, uptake, and translocation.

Frederic Fellouse and Kurt Deshayes
Department of Protein Engineering
Genentech Inc.
1 DNA Way
South San Francisco, California 94080

Selected Reading

1. Lenz, G.R., Nash, H.M., and Jindal, S. (2000). Drug Discov. Today *5*, 145–156.
2. Lam, K.S., Lebl, M., and Krchnak, V. (1997). Chem. Rev. *97*, 411–448.
3. Sidhu, S.S., Fairbrother, W.J., and Deshayes, K. (2003). Chembiochem *4*, 14–25.
4. Sandman, K.E., Benner, J.S., and Noren, C.J. (2000). J. Am. Chem. Soc. *122*, 960–961.
5. Woiwode, T.F., Haggerty, J.E., Katz, R., Gallop, M.A., Barrett, R.W., Dower, W.J., and Cwirla, S.E. (2003). Chem. Biol. *10*, this issue, 847–858.
6. Trepel, M., Arap, W., and Pasqualini, R. (2002). R. Curr. Opin. Chem. Biol. *6*, 399–404.
7. Westerhof, G.R., Schornagel, J.H., Kathmann, I., Jackman, A.L., Rosowsky, A., Forsch, R.A., Hynes, J.B., Boyle, F.T., Peters, G.J., and Pinedo, H.M. (1995). Mol. Pharmacol. *48*, 459–471.

Chemistry & Biology, Vol. 10, September, 2003, DOI 10.1016/j.chembiol.2003.09.006

Ontology Recapitulates Physiology

High-content information experiments in the post-genomic era hold the promise of deciphering age-old questions in biology and new ones in the biomedical arena. In response, researchers are devising computationally intensive and novel strategies to extract answers from multidimensional data sets.

The post-genomic era has brought with it a vast collection of data from disparate sources, raising new questions about how to interpret the information and derive something meaningful. First, the human genome sequence and high-density gene expression arrays came, followed by high-throughput bioassays, SNPs, proteome biochips, and, more recently, genome-wide gene knockdown screens in cells, the collective interpretation of which, in the absence of "A Beautiful Mind," is computationally challenging [1–6]. Needless to say, the influx of large-scale data sets has shifted the biomedical research focus toward challenges in computational science [7]. Extracting knowledge contained in the patterns of these experiments into a structured format useful to biologists and medical researchers may highlight an underlying "method to the madness" and could prove critical to an understanding of how cells work.

Attempts to systematically identify novel gene and/or drug function from genome-scale data have thus far relied on acts of heroism both at the bench and in front