Affinity Selection from Biological Libraries

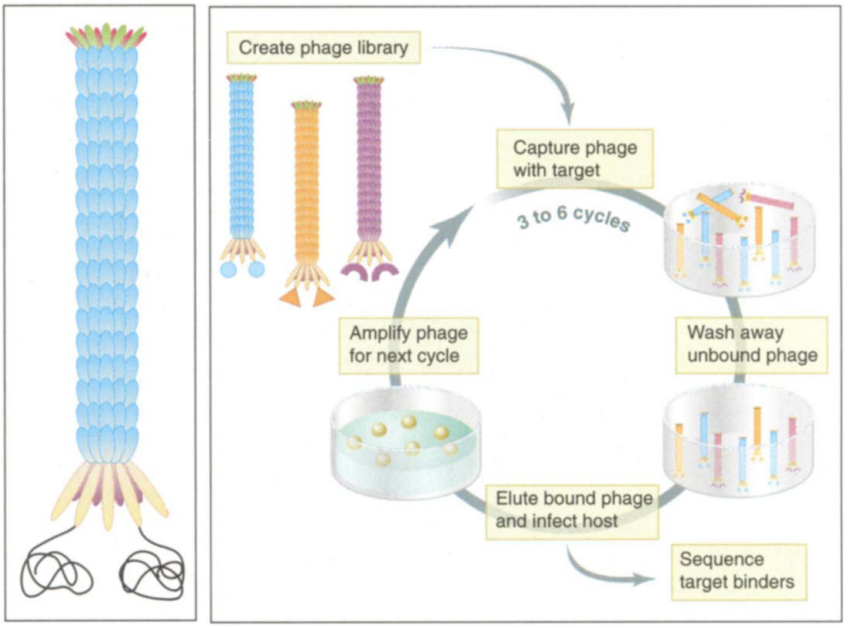
James F. Smothers, Steven Henikoff, Paul Carter

The electronic strands of the World Wide Web would tie us in knots were it not for clever search engines capable of selecting (and sometimes ranking the relevance of) documents from among billions of possibilities. Biologists practice a similar selection process when searching for interacting molecules. In this technique, called phage display, a pure protein or more complex target is used to select interacting partners from billions of particles that differ molecularly on their surfaces. The targets used in this procedure are bacterial viruses known as bacteriophage (or phage) that have been modified to carry potential interacting partner proteins or small peptides fused to exterior coat proteins. Just as search engines rank documents in order of relevance to the query, the physical selection of the phage display method identifies the best target binders.

Phage such as M13 have been widely used since the 1970s as "genetic warehouses" for cloning and sequencing of DNA fragments. In the mid-1980s, George Smith converted phage into "libraries" of genetic information whose encoded polypeptides could be displayed on the surface of the phage and then functionally interrogated (see figure). Smith first demonstrated this technique by genetically inserting a fragment of the bacterial enzyme Eco RI into the pIII coat protein of M13. He then used an antibody to the enzyme to enrich (or preferentially select) for the enzyme-displaying phage so that it was 1000 times more abundant than wild-type phage (*1*). Moreover, these mutated phage were still infectious and capable of passing along their genetic information to their progeny. Soon thereafter, Smith and colleagues constructed the first library of phage in which pIII was fused to random peptide sequences (*2*). An antibody was again used as the interacting target protein, but this time it selected from ~10^7 competing peptides, each of which could potentially form a specific complex with the antibody. After successive rounds of enrichment, several peptides representing the antibody's interaction site (epitope) were recovered. Jim Wells and colleagues advanced the technology by displaying two different human growth hormone variants on phage and selecting for the variant with higher binding affinity for the hormone receptor (*3*).

Four key strengths of phage display technology were demonstrated in these earliest investigations (*1*–*3*). (i) The peptide or protein displayed is physically linked to the genetic information that encodes it by virtue of its linkage to the viral coat encasing the genome. (ii) Successive rounds of selection—so called "biopanning" (see figure, above right)—hone in on target-specific protein interactions in the competitive presence of other potential partners. (iii) Phage libraries are plastic in design, allowing the display of either small random peptides or lengthy naturally occurring proteins with complex structures. (iv) Identification of the most efficacious complexes is possible on the basis of selecting interactors with the highest relative binding affinity.

Several features of contemporary phage display libraries distinguish them from their predecessors (*4*). Phage libraries now contain significantly larger sequence complexity than their historical counterparts: up to 10^{11} unique sequences as compared to ~10^7. Greater library complexity translates into higher probability that the most specific or highest affinity interactions will be resolved from amongst the total available interactions. This probability is greatly improved if relatively high stringency target-phage capture methods are applied during the selection process to avoid enrichment of less-specific interactors, which also increase in phage populations of greater diversity. Current researchers can design libraries on the basis of length and physical presentation of fusion peptides and proteins, which also improves their chances of finding selective interactions. Proteins can also be fused to either amino- or carboxy-termini by choosing judiciously among the five phage coat proteins [see figure, above left]. Phage biology has also been exploited to express these fusion proteins from small DNA structures, called phagemids, that are packaged into mature fusion-displaying phage with the aid of a "helper" phage. This process alleviates the need to genetically engineer the phage genome directly while allowing control over fusion protein expression levels. Consequently, modern phage libraries display fusion proteins either monovalently or multivalently to identify high- or low-affinity interactors, respectively.



Phage architecture and phage display. **(Left)** A protein (black coil) is shown fused on its carboxy-terminal end to the phage minor coat protein pIII (yellow) of a filamentous phage. Peptides and proteins have also been displayed by genetic fusion to the other minor coat proteins [pVI (purple), pVII (red), and pIX (green)] or to the major coat protein [pVIII (blue)]. **(Right)** Key steps in the selection of interactors from phage display libraries.

Today's varied and complex phage libraries often yield vast data sets of target-binding sequences, imposing demands on efficient evaluation of experimental results. Modern amino acid–alignment and comparison programs quickly and reliably interpret consensus and diversity of phage sequences (*5*). Public

J.F.S. and P.C. are in the Department of Antibody Technologies, Amgen Inc., Seattle, WA 98101, USA. E-mail: SmotherJ@amgen.com and CarterP@amgen.com. S.H. is at the Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. E-mail: steveh@fhcrc.org

VERSATILITY OF PHAGE DISPLAY LIBRARIES

Library	Common Applictions
Random peptides	Map protein-peptide interactions
	Identify specific binding reagents Diagnostics Receptor antagonists or agonists Enzyme inhibitors Mimics of epitopes and natural ligands (mimitopes)
	Determine protease substrate specificity
	Create vaccine
	Identify cell- or tissue-specific markers
Antibody fragments	Identify high-affinity specific binding reagents Human therapeutics and diagnostics Receptor antagonists or agonists Enzyme inhibitors
	Identify cell- or tissue-specific markers
cDNA libraries	Identify naturally occurring protein complexes
	Determine enzyme-substrate specificity
Protein fragments or variants	Map protein-protein interactions and epitopes
	Change protein function Enhance binding affinity Alter binding specificity Alter substrate specificity or enzyme kinetics

databases can also be searched using target-derived phage sequences to discover natural protein partners, to identify overlapping research areas, and to reveal new clinical arenas wherein drugs might be useful.

Antibodies are the most rapidly expanding class of drugs for the treatment of human disease (*6*). The display of functional antibody fragments on phage (*7*) and the subsequent antigen-driven selection from phage libraries (*8*) has provided a new way to generate antibodies without the use of rodents. The development of highly diverse (>10^9 unique sequences) human antibody phage libraries in the late 1990s has provided a powerful technology for the development of human antibodies for therapeutic use. Indeed, at least seven phage-derived human antibodies have progressed to clinical trials (www.cambridgeantibody.com).

Phage display has also been used to enhance the clinical potential of antibodies. For example, antibodies to the HIV glycoprotein gp120 and the tumor marker HER2/neu have been "affinity matured" via phage display, resulting in variant antibodies with 420- and 1230-fold increases in binding affinity, respectively (*9*, *10*). In both cases, human antibodies selected from phage display were mutated at sites of potential antigen contact to create "evolved libraries" that were then further enriched to identify enhanced affinity variants.

Molecules displayed on phage libraries are by no means limited to peptides and antibodies (see table, above). High-complexity cDNA phage display libraries have identified naturally occurring protein-protein complexes, similar to other two-hybrid methods (*11*). Libraries of random protein fragments have been created to map natural sites of protein-protein interaction, develop reagents for immunizing animals, and make vaccines (*12*). Site-directed mutagenesis and other techniques have been used to create libraries of either enzymes or substrates with altered specificity or kinetics (*13*).

Phage display libraries have even been injected into animals to explore interactions in vivo (*14*). After injection, an organ or tissue of interest is removed and any phage found within it are amplified for the next cycle of biopanning. The selection process is repeated until an enriched pool of organ- or tissue-specific phage is identified. Amidst a contemporary debate of the method's practical utility, Ghosh and colleagues provided a compelling proof-of-concept demonstration (*15*). Using a mosquito host, these researchers identified peptides that selectively bind to the insect's midgut and, after exposure to gametocyte-infected blood, largely block malarial infection. Genetically engineered versions of these mosquitoes express these small peptides and show the same resistance to infection (*16*), providing new hope in the battle against a disease for which effective vaccines do not exist.

Recently, phage libraries have been used to rapidly map the key amino acid residues that contribute to the energetics of a protein-protein interaction (*17*). In this "shotgun scanning" strategy, phage libraries that contain alanine substitutions of residues in a displayed protein of interest monitor the location of residues where the least amount of variation is allowed to achieve binding, thus identifying the residues that contribute most to target binding. This technique has been shown to confirm the findings of a traditional analysis of mutant protein binding energetics.

The history of phage display illustrates an advantage in screening libraries with maximal diversity, thereby increasing the probability that interactions reflecting the most desired characteristics will be found, e.g., highest affinity binders. Modern phage libraries are constructed with vast complexities of ≤10^{11} sequences, and ≤10^{13} phage are commonly screened in each experiment. Further increases in library size impose practical difficulties on their efficient construction, propagation, and sampling. As a phage library increases in complexity, fewer replicates of each clone are sampled, which presumably decreases the probability of retrieving any given clone. An attractive alternative to creating ever larger libraries is to introduce additional diversity between rounds of selection via site-directed mutagenesis or DNA shuffling, thus evolving binders that are increasingly tailored to the target as selection progresses. This is analogous to how the mammalian immune system responds to antigen recognition, perpetuating an evolved subset of diverse antibodies with varied affinity for the foreign target. Other recently developed selection technologies, such as ribosome and messenger RNA display, enable the use of libraries of far greater diversity (10^{14}) than their phage counterparts and may be advantageous where even greater complexity is desired (*18*).

The diverse application, popularity, and success of phage display is largely due to the method's flexiblity and its practical uses, which remain largely under control of the investigator. Recently, phage display has even been used to identify peptides that selectively bind inorganic materials used to make semiconductors (*19*). Investigators plan to fuse these peptides to other peptides that specifically bind organic molecules, creating hybrid products for use in "nano-scale" semiconductors that could effectively bridge the gap between biological materials and inorganic substrates. The comparison between phage display and a search engine may, thus, be more than a convenient metaphor; it may become a literal link between biological and physical technology.

References

1. G. P. Smith, *Science* **228**, 1315 (1985).
2. J. K. Scott, G. P. Smith, *Science* **249**, 386 (1990).
3. S. Bass, R. Greene, J. A. Wells, *Proteins* **8**, 309 (1990).
4. C. G. Adda *et al.*, *Comb. Chem. High Throughput Screen.* **5**, 1 (2002).
5. J. F. Smothers, S. Henikoff, *Comb. Chem. High Throughput Screen.* **4**, 585 (2001).
6. M. P. Quan, P. Carter, in *Anti-IgE and Allergic Disease*, P. M. Jardieu, R. B. Fick Jr., Eds. (Marcel Dekker, New York, 2002), vol. 164, chap. 20, pp. 427–469.
7. J. McCafferty, A. D. Griffiths, G. Winter, D. J. Chiswell, *Nature* **348**, 552 (1990).
8. T. Clackson, H. R. Hoogenboom, A. D. Griffiths, G. Winter, *Nature* **352**, 624 (1991).
9. W. P. Yang *et al.*, *J. Mol. Biol.* **254**, 392 (1995).
10. R. Schier *et al.*, *J. Mol. Biol.* **263**, 551 (1996).
11. R. Crameri, G. Achatz, M. Weichel, C. Rhyner, *Methods Mol. Biol.* **185**, 461 (2002).
12. K. Manoutcharian *et al.*, *Curr. Pharmacol. Biotechnol.* **2**, 217 (2001).
13. R. H. Hoess, *Chem. Rev.* **101**, 3205 (2001).
14. M. Trepel, W. Arap, R. Pasqualini, *Curr. Opin. Chem. Biol.* **6**, 399 (2002).
15. A. K. Ghosh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 13278 (2001).
16. J. Ito *et al.*, *Nature* **417**, 452 (2002).
17. G. A. Weiss *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8950 (2000).
18. P. Amstutz, P. Forrer, C. Zahnd, A. Pluckthun, *Curr. Opin. Biotechnol.* **12**, 400 (2001).
19. S. R. Whaley *et al.*, *Nature* **405**, 665 (2000).