3
Harvesting Molecular Diversity – Biology's New Commodity

JEFFREY W. SMITH AND ERKKI RUOSLAHTI

The Burnham Institute, La Jolla Cancer Research Center, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA

Introduction

During the 1970s and early 1980s biochemists were entrenched in the daunting task of understanding the nature of the connection between genetic material (DNA and RNA) and biological function. However, the rapid expansion in the number of known biomolecules has changed the face of the problem for the biochemist. With every new gene that is cloned there is a protein that must be understood, a biochemical pathway to which that protein interfaces, and ultimately an effect of the protein at the cellular level. In 1980 there were only a few thousand known protein sequences, now there are many times that number. In fact, the inevitable success of the human genome project promises to fill the databases with tens of thousands of new sequences. This astronomical expansion in information will continue to change the face of biology. We will soon enter an era when cloning novel genes will be obsolete.

As more sequence information is compiled, the conventional approaches toward understanding function will be too cumbersome to keep pace. Consequently, the current challenge is to devise more efficient methods of studying protein function.

Fortunately, there has been a line of thought which has changed the way in which biology, biochemistry and drug design can be approached. The focus is now centered on creating molecular diversity. Novel molecules are currently being created in the laboratory at a pace that far outdistances that of natural evolution. It may be fair to say that biochemistry is bringing much more evolutionary pressure to bear on proteins *in vitro* than has ever been done by nature itself. These novel strategies are allowing *us* to probe molecular function at an ever quickening pace.

Two advances in technology have been the key to the creation of biomolecular diversity. The first advance to shape this change was the development of the polymerase chain reaction (PCR) (Mullis and Faloona, 1987), an *in vitro* method of amplifying large amounts of double stranded DNA using infinitesimal amounts of a template. The fidelity of the polymerase is a major reason for widespread adoption of PCR. The error rate of the polymerase is so low that DNA can be faithfully replicated with great

Biotechnology and Genetic Engineering Reviews – Vol. 14, April 1997
0264-8725/97/14/51-65 $20.00 + $0.00 P.O. Box 716, Andover, Hampshire SP10 1YG, UK

confidence. However, the real advantage of PCR may be its ability to create random mutations within any given segment of DNA. Therefore, although not originally recognized as such, PCR has become an instrumental tool in the creation of biomolecular diversity.

Another important step in the creation of biomolecular diversity was taken by Smith (Smith, 1985) and then extended by Scott and Smith and others in 1990 (Cwirla *et al.*, 1990; Devlin *et al.*, 1990; Scott and Smith,1990). They taught that vast libraries of random peptide sequences could be displayed on the surface of filamentous bacteriophage and screened for binding activity. These seminal papers provided the means to link an active peptide motif to the genetic material which encodes it, thereby creating heritability, a crucial property in combinatorial approaches.

Subsequent to these two advances, numerous strategies have been put forth toward the creation and screening of novel biomolecules. This article will review the many ways that biomolecular diversity can be created in the laboratory. We will also discuss selected applications of these approaches, phage display in particular, as a means of illustrating the power of the new technology.

Creating molecular diversity

PHAGE DISPLAY

The first widely adopted approach toward creating biomolecular diversity was phage display (for review see Bradbury and Catteneo, 1995; Barbas, 1993; Medynski, 1994; Makowski, 1993). Phage display involves the presentation of a library of peptides, or even proteins, on the surface of a filamentous bacteriophage. Importantly, the phage genome, which is physically connected to the presented peptide, contains the DNA which encodes the peptide. Thus, the key facet of phage display, and most other combinatorial strategies, is the 'heritable' nature of the displayed moiety.

The life cycle of filamentous phage is particularly advantageous for displaying libraries of random peptides. The phage gene three product, also called the gene three protein (or p3) is the phage component that is normally modified for use in phage display of a foreign protein. The p3 protein is presented on the phage surface, where it mediates the binding and entry of the phage particle into bacteria. The p3 protein contains three domains, each separated by what appear to be flexible glycine linkers. The first two domains are essential for pilus binding and entry of the viral genome into the bacteria. The third domain anchors the p3 protein into the phage membrane. It is the first domain, or N1, that is modified for phage-display. The DNA encoding the N1 domain is manipulated so that an additional segment of protein is joined to the N-terminus of p3, yielding a fusion protein in which a peptide or protein is displayed at the phage surface with p3. Between one and five copies of a fusion protein with p3 can be expressed per phage particle, and this can, under some conditions, be considered 'monovalent' display. There are other methods of displaying peptides on the surface of phage in a multi-valent form where the peptide library is fused to the gene VIII protein which is displayed in several hundred copies on the phage surface (Makowski, 1993; Bradbury and Catteneo, 1995).

As demonstrated by three groups in 1990 (Cwirla *et al.*, 1990; Devlin *et al.*, 1990;

Scott and Smith, 1990), when a small peptide is fused to the phage p3 protein, its sequence can be randomized to contain a library of different peptide sequences. Normally a library can be created to contain between 10^7 and 10^9 different peptide sequences. Because the infectivity of the phage is retained throughout this manipulation, the randomized library can be screened in a biochemical assay and then the sequence of the active peptide can be determined by propagation and sequencing of the phage DNA. Scott and Smith originally demonstrated the utility of the 'library approach' toward the identification of the amino acid sequence of an antibody epitope; however, the general strategy of phage display is now widely used in a number of imaginative ways. We provide examples of the use of phage display later in this chapter.

POLYSOME DISPLAY

The utility of a combinatorial library can be a function of the number of different conformers or shapes that are incorporated into the library. Consequently, the 'complexity' of a library, or its degree of diversity, could be a limiting factor in the selection of ligands from the library (see aptamers below). The biology of filamentous phage and the transformation efficiencies of *E. coli* limit the calculated diversity within a library to approximately 10^9 different combinations of a peptide or protein. Because of these limitations, only six to seven amino acid residues can be completely randomized with confidence in a phage library.

Another technique called polysome display has been used to obtain libraries that are substantially more complex, and presumably offer more opportunity for the selection of active molecules. To achieve heritability, polysome display takes advantage of the fact that nascent polypeptides are physically linked to their cognate mRNA on the polysomal complex. Therefore, the active peptide can be identified while it is physically connected to its parent mRNA. Using this approach, investigators at Affymax were able to create a polysome library estimated to contain 10^{12} different random peptide motifs (Mattheakis *et al.*, 1994), a degree of complexity that is at least 10^3 greater than that obtained in most phage libraries. Thus, polysomes can be used to display libraries of vast complexity. This added complexity enables the creation of a peptide library in which ten amino acid residues can be displayed in all possible combinations with great confidence.

The method for conferring heritability is different in polysome display than in phage display. The nascent polypeptide connected to the polysome is stabilized by the addition of chloramphenicol. Then the library of displayed peptides is screened for binding activity. The RNA linked to the active peptide is also recovered during the panning scheme. This RNA can be transcribed to DNA, and then amplified by PCR, yielding sufficient cDNA for sequence analysis.

A major advantage of polysome display is its ability to incorporate great diversity into the library. It has also been suggested that this approach could be adapted for incorporating unnatural amino acids into the displayed polypeptide. Clearly, a disadvantage of the polysome approach is the lack of any cellular machinery to assist in the folding of the nascent polypeptide. Consequently, the utility of the approach is probably limited to smaller peptides which do not require folding or are capable of folding on their own.

PEPTIDES ON PLASMIDS: LIBRARIES OF PEPTIDES FUSED TO THE LAC REPRESSOR

In a third method of constructing peptide libraries, the peptides are displayed as a fusion protein with the lac repressor protein LacI (Cull *et al.*, 1992). LacI binds to a segment of DNA within the lac operon. The dissociation constant of this interaction is 1×10^{-13}M. Consequently, the binding of LacI to plasmid DNA is so tight that the interaction can confer heritability by linking the fused peptide directly to its plasmid DNA. It is important to emphasize that the peptide-plasmid complex is generated intracellularly, ensuring that the peptide is linked only to its 'own' plasmid DNA. The library of peptide-plasmid complexes can be harvested from a broth of *E. coli* and the complexes are then panned for activity using procedures similar to those used for phage display. Three features of the 'peptides on plasmids' strategy are unique. First, the peptides are expressed as N-terminal fusion proteins. Therefore, the peptide presents a free C-terminus, in all other display strategies the peptide displays a free N-terminus. This feature ensures that any stop codons present within the randomized segment of DNA still generate screenable (albeit truncated) peptides. Second, the proteins fused to the LacI protein are expressed in the cytoplasm of *E. coli*, and need not be compatible with the apparatus of *E. coli* used to export proteins to periplasm. Export from the cytoplasm is a stringent requirement in phage-display. Third, the peptides on plasmid system is inherently multi-valent because each repressor acts as a tetramer. Moreover, each plasmid binds to two tetramers. Consequently, this approach is likely to be most valuable in identifying low affinity interactions ($K_d \cong 10^{-6}$ M), which would only be observed using a multi-valent display system.

SYNTHETIC RNA AND DNA LIBRARIES

The drive for more biomolecular diversity is not limited to peptide backbones. The use of oligonucleotide polymers as libraries has also been explored. Szostak and coworkers were the first to test the hypothesis that random libraries of polynucleotides could be used to identify novel ligands (Ellington and Szostak, 1990). Their approach involved the synthesis of single stranded RNA or DNA molecules, which they term 'aptamers'. Their study also addressed an important question in biomolecular diversity: what is the probability that an entity with a random sequence will fold to a stable conformation with the capacity to interact with a ligand? This is a key consideration in the field of molecular diversity because it addresses the issue of how much complexity is required in a random library.

Szostak and co-workers synthesized a randomized segment of DNA consisting of 100 bases. This polymer was flanked by specific sequences for PCR amplification of each DNA molecule. The authors estimated the complexity of the library to be 10^{15} unique sequences. An important step in the construction of the library was the conversion to a single stranded RNA molecule using T7 RNA polymerase. The authors reasoned that a single stranded oligonucleotide would present more potential binding contacts than a double stranded molecule, where many of the binding surfaces are already occupied by strand-strand interactions. As a consequence of this conversion, the complexity of the library was diminished by about 100-fold. The final RNA library contained about 10^{13} unique aptamers and each was represented about 40 times in the library.

To test the ability of the aptamers to interact with a ligand, the authors screened the

library of aptamers for the ability to bind a series of hydrophobic dyes. Dyes were chosen as model ligands because their chemical nature is compatible with polynucleotide stacking interactions. The authors identified aptamers that could bind specifically to several dyes, proving that polynucleotide libraries can be used in a manner similar to peptide libraries.

The quantitative approach taken by the authors allowed them to estimate the probability that a random sequence polymer will bind to any given ligand, yielding an answer that is rather daunting. Their rationale was as follows. The number of potential sequences displayed by a random library of 100mers is 10^{60}. Thus, their pool of 10^{13} different aptamers is only a minute fraction of all possible sequences or conformations. Given these values, and the number of binding aptamers recovered from the library, the authors conclude that only between one in 10^{49} and one in 10^{52} molecules are represented in this kind of a random library and can bind a specific ligand. In essence, this result might imply that with the current approaches, harvesting the best molecular ligand from libraries is really an insurmountable problem. However, we can expect that libraries of peptides would yield considerably better odds, given that the oligonucleotide aptamers are composed of only four different nucleotides and a random peptide could be derived from twenty different amino acid residues. Moreover, similar amino acids tend to have at least partially redundant functions in proteins, making it possible to find 'first generation' ligands even from libraries that may not display complete diversity.

The aptamer approach has shown practical application as well. An aptamer with an affinity of 200nM for thrombin was obtained from an aptamer library by affinity selection on thrombin (Bock *et al.*, 1992). Originally an aptamer of 60 nucleotides was isolated, but ultimately the authors were able to narrow the minimal length down to a oligonucleotide consisting of 15 bases. The smaller aptamer displayed a similar affinity for thrombin and equivalent biological potency to the 60mer.

SEXUAL PCR: RECOMBINATION *IN VITRO* THROUGH DNA SHUFFLING

The screening of random peptide and nucleotide libraries is a powerful approach, but it is certainly not the only means of creating new biologically active molecules. Another unique strategy for creating biological diversity, called DNA shuffling, has been pioneered by Stemmer and co-workers (Stemmer, 1994a,b; Crameri *et al.*, 1996). The strategy involves the use of the polymerase chain reaction to recombine and amplify DNA such that the sequences of related proteins are reshuffled into a series of novel hybrid proteins. In a recent review of this strategy (Smith, 1994), G.P. Smith correctly recognized the approach as having sexuality because, like other forms of biological reproduction, it involves genetic recombination. Although under normal circumstances the amplification of DNA by PCR is meant to be of high fidelity, sexual PCR relies on the infidelity of the polymerase chain reaction. Thus, the approach is commonly referred to as 'sexual PCR'. The technique of sexual PCR involves the cleavage of target DNA at random locations by a nuclease like DNAseI. Subsequently, this DNA is amplified by PCR at low stringency, using primers corresponding to the normal 5' and 3' ends of the wild-type gene. The result is the amplification of a mixture, or library, of recombined fragments of the original DNA.

In an elegant demonstration of the utility of sexual PCR, Stemmer and co-workers

were able to promote the 'evolution' of β-lactamase *in vitro* (Stemmer, 1994b). Their goal was to enhance the resistance of bacteria expressing the TEM-1 β-lactamase toward the antibiotic cefotaxime. The gene for TEM-1 was digested with DNAse I and the random mixture was subject to PCR amplification, allowing homologous PCR fragments to substitute for one another in the amplification process. The inherent error rate of PCR also added to the diversification of the library during the amplification. Three rounds of reshuffling were performed and the new versions of TEM-1 were transformed into *E. coli* and the transformants were selected for resistance to cefotaxime. It is important to emphasize the significance of the selection process in this scheme. Biological selection (as opposed to biochemical selection) is used in this iterative manipulation and is a major advantage that is lacking in the screening of peptide libraries. Using this strategy the authors created a novel form of TEM-1 that endows bacteria with 30,000 fold more resistance to cefotaxime than the wild-type enzyme. In principle, this approach could be applied to any protein, or to any two or more related proteins, to obtain recombined and mutated progeny with altered biological function.

PERMUTEINS

Although libraries of random sequences are now widely used to screen for new molecules, there are also strategies for creating diversity from existing molecules. Circular permutation is an example. Circular permutation is based on the hypothesis that a protein can be viewed as a closed circle, much like a circular plasmid in which the N and C-termini are at a break point in the circle. A circular permutation then, involves a change in the break point in a circle. To 'permutate' a protein, the break point in the circle is moved to an alternate location e.g., the protein is engineered to contain a different N and C-terminus, and the circle is 'closed' by creating a peptide bond between the native N and C-termini. Research in this area is still emerging, but at present it is believed that a successful permutation requires the original N and C-termini to be 'proximal' in the three dimensional structure of the native protein. It is also thought that the new break point, or new termini, should be created within an exposed surface loop on the protein. However, neither of these criteria have been rigorously tested.

The function of several permutated proteins has now been examined. Surprisingly, each circularly permutated protein exhibits an activity very similar to the wild-type protein (Zhang *et al.*, 1993; Vignais *et al.*, 1995). The crystal structure of one protein has been obtained for both the native and permutated forms (Hahn *et al.*, 1994). Remarkably, the two structures are nearly identical, even though in the process of permutating the protein, the authors changed the linear spacing between a disulfide bond from 30 residues to over 180 residues. It is interesting to note that most of the structural shifts that occurred in the permutated protein were confined to flexible surface loops. This finding suggests that permutation could be used as means of altering the presentation of surface loops to obtain novel binding activity or to obtain more favorable binding constants.

Although circular permutation has not been exploited extensively, one study has shown that the potency of an immunotoxin consisting of a fusion protein between IL4 and *pseudomonas* exotoxin can be increased by first permutating IL4 and then making

a fusion protein between a permutated version of IL4 and the toxin (Kreitman *et al.*, 1994). The authors interpret their data to mean that the region near the original C-terminus of IL4 is involved in receptor binding, and that fusion of the toxin at this position sterically hinders the binding between IL4 and its receptor. They suggest that the permutated fusion protein is more toxic because the contacts between IL4 and its receptor are not disrupted or masked by the fusion. In principle, the permutein approach has great promise, but at present this technology is lacking the 'library' approach that has made all of the other engineering strategies so successful. If libraries of proteins can be created in which each member of the library represents a different permutation, this technique would be far more valuable.

Harvesting Molecular Diversity from Phage Libraries

PHAGE LIBRARIES AS A SOURCE OF BIOLOGICALLY ACTIVE PEPTIDES

Phage libraries have been used to identify peptides for a number of purposes including identification of epitopes for antibodies (Cwirla *et al.*, 1990; Scott and Smith, 1990), isolating peptides that mimic protein ligands for various receptors (Hammer *et al.*, 1992), peptides that mimic carbohydrate ligands for lectins (Oldenburg *et al.*, 1992; Scott *et al.*, 1992), and peptides that bind to specific DNA sequences (Wang *et al.*, 1995).

Our laboratories have used the phage display methodology to isolate peptide ligands and antibody ligand mimics for integrin adhesion receptors (*Figure 1*). Integrins are αβ heterodimers which are responsible for most cell-matrix interactions. Many integrins, including the ones that we chose to target, will bind to the Arginine-Glycine-Aspartic Acid (RGD) tripeptide motif when it is presented in the proper conformation (Ruoslahti, 1996). Because integrins are involved in adhesive events that contribute to vascular occlusion, inflammation, osteoporosis and tumor invasion, it is generally accepted that antagonists of integrins will be clinically useful. Thus, the use of phage display to identify integrin ligands has two purposes. First, the isolated peptides serve as good lead molecules for further drug design. Second, the information obtained from phage libraries has provided insight into the molecular basis of integrin ligand binding specificity. This information must also be considered in the effort to obtain highly specific integrin antagonists.

In an initial study from one of our laboratories, a hexapeptide phage library was screened for optimal peptide ligands of the α5β1 integrin (Koivunen *et al.*, 1993). All of the more than fifty peptide phage isolated from the library by panning on the α5β1 integrin contained integrin recognition sequences. These were mostly RGD or related motifs (*Figure 1*). There were virtually no phage isolated that were found to be nonspecific binders.

The screening of a hexapeptide library on the α5β1 integrin also illustrates the potential for obtaining conformationally constrained peptides from a library. One of the peptide motifs that was isolated by panning on α5β1 contained two cysteine residues flanking an RGD sequence. In fact, of all of the peptides selected from the hexapeptide library, the one containing the two cysteine residues was the most avid binder. This observation prompted the design and testing of a library in which two cysteine residues were programmed into the displayed peptide and the six residues between the cysteines were randomized. These libraries should express predomi-

nantly disulfide linked cyclic peptides, and practice has shown this prediction to be correct (Koivunen *et al.*, 1993, 1994). As expected, many of the peptides isolated from this library bound tightly to the integrins.

A comparison of the peptides obtained from the linear and cysteine constrained libraries revealed an interesting difference in peptide sequence among two RGD-binding integrins. Most RGD-binders, like $\alpha 5\beta 1$ and $\alpha v\beta 3$, bound better to the cyclic peptides rather than linear peptides. Prior study with a limited number of synthesized peptides also indicated this preference (Ruoslahti and Pierschbacher 1987). These data indicate that the conformational flexibility of the integrin ligand is a key determinant in ligand binding specificity. It is likely that the disulfide constrained peptides assume fewer conformations than their linear counterparts, which are likely to assume many transient conformations that are not compatible with integrin binding.

Interestingly, some of the peptides obtained from phage libraries that were selected on integrin contained four cysteine residues and are likely to contain two disulfide bonds. The sequence of one such peptide, which displays high affinity for the $\alpha v\beta 3$ and $\alpha v\beta 5$ integrins, is shown in ACDERGDCFCG. This peptide, and others containing two disulfides, are likely to exhibit conformations that are highly restricted, and this may explain their high affinity for the integrins.

The use of phage peptide libraries to probe the integrin ligand binding site has provided two other important observations. First, the approach allowed us to identify peptides that potentially represent secondary contact points between the integrin and its ligand. For example, by panning on the $\alpha 5\beta 1$ integrin, we also isolated a few peptides that appeared to mimic sequences situated close to the RGD motif in fibronectin (Koivunen *et al.*, 1993). One of these sites in fibronectin has also been identified in antibody inhibition and mutagenesis studies as an integrin contact point (Obara and Katsutoshi, 1995). Thus, phage screening was capable of identifying a primary and secondary contact points in the ligand for integrin.

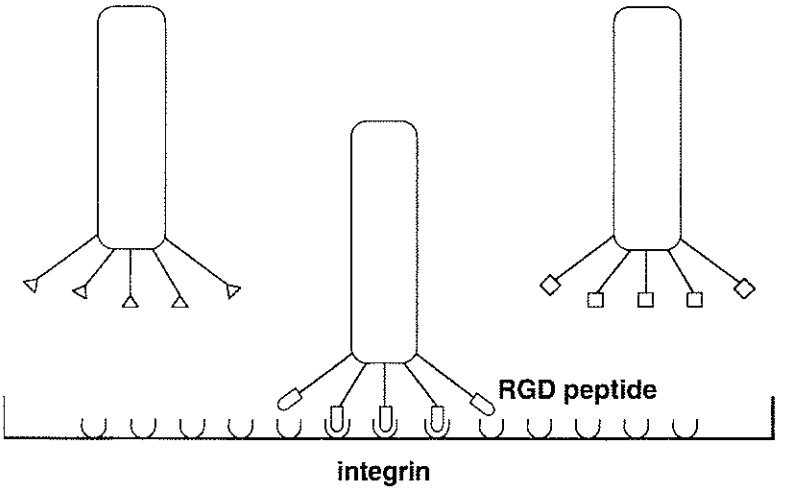

Figure 1. Principle of phage selection by panning.

Phage display was used to determine the domain within an integrin adhesion receptor that bound to the RGD motif (Pasqualini *et al.*, 1995). This problem is a crucial issue in integrin biochemistry because the precise location of this domain has not been mapped. We hypothesized that a selection of a phage library on an RGD ligand may identify sequences that would mimic the receptor binding site. Peptide libraries were panned on the integrin ligand fibronectin, yielding a series of peptides, many of which contained a consensus sequence of CWDDG/LWLC. This consensus sequence has homology to a domain within the β subunit of the integrin (sequence of KDDLW). This region in the integrin had been previously implicated in the receptors ligand binding function (D'Souza *et al.*, 1994). This study used a reductionist approach to obtain a minimal motif with a function homologous to a large and cumbersome heterodimeric transmembrane protein, suggesting that the use of phage-screening to identify binding sites within large and unmanageable receptors may be broadly applicable.

BUILDING INTEGRIN LIGANDS FROM ANTIBODIES DISPLAYED ON PHAGE

One of the inherent limitations of the immune system is the inability of host to recognize self. In collaboration with Barbas and coworkers, one of our labs recently overcame this limitation by engineering a human antibody to bind to human integrins (Barbas *et al.*, 1993; Smith *et al.*, 1994). We reasoned that an RGD motif could be inserted into the antigen binding site of an antibody, endowing the antibody with the ability to bind integrin. By incorporating phage display into this strategy, we were able to optimize for binding affinity and specificity.

The RGD motif was inserted into the third complementarity determining region of the heavy chain (HCDR3) of a human antibody. The six residues flanking the RGD were randomized and then constrained by including two cysteine residues. When a phage library displaying these antibody constructs was panned on the αvβ3 integrin, several antibodies with sub-nanomolar affinity for the integrin were obtained. These antibodies also bound to the platelet integrin αIIbβ3 with nearly equal affinity. Importantly though, none of the antibodies bound to the αvβ5 integrin, which can also bind to the RGD sequence. Thus, the application of only a single selection step yielded human antibodies with very high affinity and a great deal of specificity for particular integrins.

As an extension of the original study, we tried to modify the best antibody so that it would distinguish the two β3 integrins, αvβ3 and αIIbβ3. The phage antibody with the highest affinity for these integrins is called Fab-9, and it presents the sequence CSFGRGDIRNC within its CDR3. To alter the integrin binding activity of this antibody, we constructed a library of antibodies containing a CDR with the identical flanking sequences of Fab-9, but with a modified core sequence that was not biased toward RGD. The alternate library displayed the motif CSFGXXXXRNC, in which x is any residue. The new phage library was incubated in solution with purified αvβ3 to eliminate phage that recognized this integrin. Then the library was panned on immobilized αIIbβ3. Several antibodies were obtained from this selection that exhibited a preference for the platelet integrin. Two of the most selective antibodies contained the sequences CSFGKGDNRNC and CSFGRTDQRNC within HCDR3. These bound the αIIbβ3 integrin with 30 to 100-fold greater affinity than to αvβ3.

Thus, by extending the selection scheme by a simple manipulation of the CDR within the phage library, we were able to obtain antibodies that could distinguish the two β3 integrins.

We also examined whether synthetic peptides with sequences corresponding to the selected HCDR3 sequences could block ligand binding to integrins. In this study two different attempts were made to recapitulate the conformation of the CDR. In one series of peptides the disulfide bond between the two cysteine residues was maintained. Yet in another series, the peptides were synthesized in linear form, without the disulfide constraint. Interestingly both sets of peptides behaved identically. Although the peptides displayed significantly lower affinity for integrin than the whole antibody, they did maintain integrin binding specificity. This result is surprising given that peptides selected from libraries based on integrin binding affinity show a clear preference for disulfide constrainment (see above). Thus, there are clear differences in the biochemical outcome of selection of antibodies and small peptides from phage libraries.

Extending the Phage Display Technology

There are two important extensions that have arisen from our studies aimed at identifying integrin ligands from phage display libraries. The first is the development of a protein engineering method called loop grafting, an approach that extends the combinatorial nature of protein manipulation. The second extension is the application of phage display to identify peptide sequences that can direct the *in vivo* homing of particles to specific vascular beds. In conjunction, these two methods may ultimately facilitate the targeted delivery of protein and cellular therapeutics.

ENGINEERING A NOVEL TISSUE PLASMINOGEN ACTIVATOR BY PROTEIN LOOP GRAFTING

Although phage display has shown great utility in displaying peptide and some protein libraries, not all proteins can be expressed on the surface of phage in an active conformation. The following example illustrates a unique strategy for circumventing this limitation of phage display. The goal of the study was to engineer a version of tissue plasminogen activator (tPA) that is targeted to blood clots. tPA is a serine protease that catalytically dissolves fibrin in blood clots. Thus, tPA is the primary enzyme with the capability to dissolve thrombi, and recombinant versions of tPA are used as therapy for myocardial infarction (Collen, 1996). The ability to target tPA to platelets, which are a major constituent of thrombi, may localize thrombolytic activity at the site of the clot and enhance the efficacy of tPA.

In collaboration with Tachias and Madison, one of our groups reasoned that tPA could be delivered to clots by endowing the enzyme with the ability to bind the platelet integrin αIIbβ3 (Smith *et al.*, 1995). Since we had previously engineered a protein surface loop with a high affinity for this integrin within the CDR of an antibody (Barbas *et al.*, 1993), we attempted to graft this optimized protein surface loop (or CDR) into tPA. As a recipient site for the loop graft, we chose the epidermal growth factor (EGF) module in tPA (*Figure 2*). This domain is not essential for the catalytic

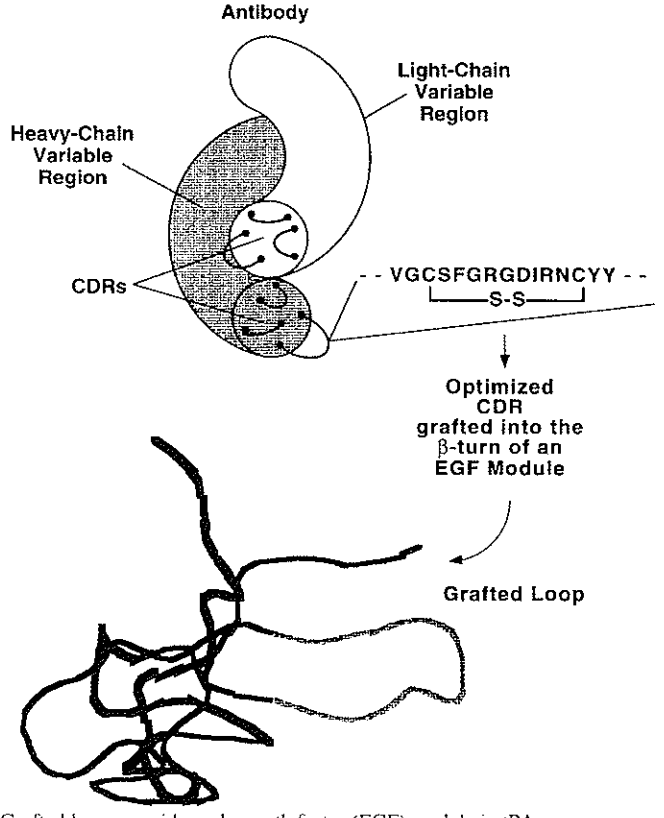

Figure 2. Grafted loop on epidermal growth factor (EGF) module in tPA.

activity of tPA, although it does contribute to the enhancement of activity by fibrin. The EGF module in tPA contains an exposed β-turn of 10 residues flanked by two pairs of cysteines. Thus the size of the loop is similar to the size of the RGD-containing CDR from Fab-9. However, it should be emphasized that the disulfide bond arrangement in the EGF module is quite different than in the CDR. The EGF loop (β-turn) is constrained by two overlapping disulfides in a 1–3, 2–4 arrangement. The CDR of Fab-9 is constrained by only a single disulfide bond.

Despite this difference in disulfide bond arrangements, the loop graft was a success. The insertion of the optimized integrin binding loop of Fab-9 into tPA endowed tPA with the ability to bind the platelet integrin αIIbβ3 with an affinity of near one nanomolar. Thus, nearly all of the binding affinity of the surface loop was recapitulated in the graft recipient (tPA). The strategy of grafting a surface loop that had been optimized by phage display between two proteins is substantially different in approach, and in outcome, from the simple insertion of an RGD motif into a protein. The RGD sequence has been inserted into non-adhesive proteins, endowing them with integrin binding capability. However, the affinity of those proteins for integrin appears to be nearly 100-fold lower than we obtained by grafting the optimized surface loop of Fab-9 into tPA (Hashino *et al.*, 1992; Yamada *et al.*, 1995; Maeda *et al.*, 1989).

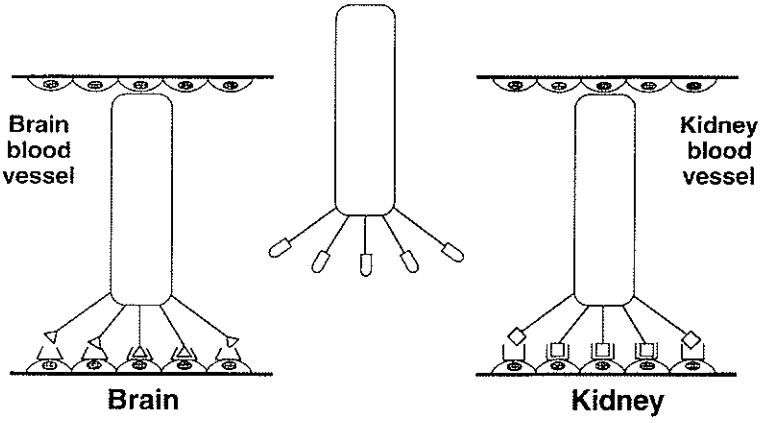

Figure 3. *In vivo* targeting: tissue specific differences (brain and kidney)

IN VIVO SELECTION OF HOMING PEPTIDES FROM PHAGE LIBRARIES

The ability to target drugs, cells and genes to the appropriate locations in the body has been a long-standing goal of researchers and physicians. If it were possible to deliver drugs selectively to their intended targets, many effective, but toxic, drugs could find a use. For example, cancer chemotherapy, which to a large extent relies on highly toxic compounds that are only partially selective for tumor cells, could benefit enormously from technology enabling selective delivery of the drug; higher concentrations of the drug would be reached in the tumor, and fewer side effects would result (Jain, 1996). Similarly, the potential harm from delivering genes into normal cells, which is a serious limitation in gene therapy, could be attenuated by targeting the therapy. Targeting could also make it possible to use drugs that cannot be delivered into the diseased tissue at therapeutically effective concentrations.

Unfortunately, there are only a few situations in which targeted drug delivery is possible. Antibodies prepared against various tumor antigens, such as the carcinoembryonic antigen, and tumor-infiltrating lymphocytes have been used for tumor targeting (Juweid *et al.*, 1996; Salgaller *et al.*, 1994). However, specificity and incomplete tissue penetration tend to limit their effectiveness. Other ways of drug targeting, such as direct injection of a drug into the arterial blood supply of a target organ, are invasive and therefore, less desirable.

Pasqualini and Ruoslahti (1996) have developed a method in which phage libraries of peptides are screened *in vivo* for sequences that have affinities for selective tissues. The method, at least when phage libraries are used, primarily targets tissue-specific differences in endothelial cells (*Figure 3*).

Peptide sequences that direct phage homing to precise organs were obtained by intravenous injection into mice followed by recovery of the phage. To select peptides that home to the brain, phage were injected intravenously, recovered from the brain, amplified *in vitro*, and re-injected to obtain a further enrichment. After three rounds of selection, phage preparations were obtained that accumulated in the brain up to 13-fold more than in the kidney. A similar selection of phage that home to kidney gave phage that strongly favored binding to the vasculature of the kidney over brain. The peptide

sequences isolated from each organ were distinct, implying a specific receptor mediated homing process.

Other studies lend further support to the homing capability of the peptide sequences. First, the individual phage clones exhibited the same organ specificity after intravenous injection as the pooled, selected phage. Immunohistochemical staining showed that the phage had been retained within the capillaries of the appropriate organ. Second, a soluble cyclic peptide based on one of the brain-binding phage sequences inhibited the localization of that phage in brain. This synthetic peptide had no effect on the kidney-homing phage. Moreover, coupling the peptide with the brain-homing sequence onto the surface of glutaraldehyde-fixed red blood cells produced coated cells that accumulated in the brain to a greater extent than in the kidney. The preferential brain localization of the coated cells was blocked by co-injection of the soluble peptide, whereas retention in the kidney was not affected.

The Ruoslahti laboratory has also demonstrated that some prior knowledge of the target molecules within a given vascular bed can be incorporated into the *in vivo* homing scheme. The $\alpha v\beta 3$ integrin has recently been described as a marker of angiogenic vasculature in tumors (Brooks *et al.*, 1994). Since, phage carrying an RGD peptide that binds with high selectivity to the $\alpha v\beta 3$ (and $\alpha v\beta 5$) integrin (Koivunen *et al.*, 1993) had previously been isolated, the ability of these phage to home to tumors was tested and found to home into tumors (Pasqualini and Ruoslahti, unpublished results).

The initial success of targeting by (Pasqualini *et al.*, 1996) and the selection of phage on cultured cells (Barry *et al.*, 1996), suggest that combinatorial libraries can serve as a source compounds with the ability to home to specific tissue targets. Peptide targeting motifs identified in this manner could potentially be used further to engineer proteins, drugs and gene vectors with novel homing properties to deliver them to specific tissues.

References

BARBAS, C.F. (1993). Recent advances in phage display. *Current Opinions in Biotechnology* **4**, 526–530.

BARBAS, C.F., LANGUINO, L. AND SMITH, J.W. (1993). High-affinity self-reactive human antibodies by design and selection: targeting the integrin ligand binding site. *Proceedings of the National Academy of Sciences USA* **90**, 10003–10007.

BARRY, M.A., DOWER, W.J. AND JOHNSTON, S.A. (1996). Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. *Nat. Med.* **2**, 299–305.

BOCK, L.C., GRIFFIN, L.C., LATHAM, J.A., VERMAAS, E.H. AND TOOLE, J.J. (1992). Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature* **355**, 564–566.

BRADBURY, A. AND CATTANEO, A. (1995). The use of phage display in neurobiology. *Trends in Neuroscience* **18**, 243–249.

BROOKS, P.C., CLARK, R.F. AND CHERESH, D.A. (1994). Requirement for vascular integrin avb3 for angiogenesis. *Science* **264**, 569–571.

COLLEN, D. (1996). Fibrin-selective thrombolytic therapy for acute myocardial infarction. *Circulation.* **93**, 857–865.

CRAMERI, A., WHITEHORN, E.A., TATE, E. AND STEMMER, W.P.C. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **14**, 315–319.

CULL, M.G., MILLER, J.F. AND SCHATZ, P.J. (1992). Screening for receptor ligands using large libraries of peptides linked to the C-terminus of the lac repressor. *Proceedings of the National Academy of Sciences USA* **89**, 1865–1869.

CWIRLA, S.E., PETERS, E.A., BARRETT, R.W. AND DOWER, W.J. (1990). Peptides on phage: a vast library of peptides for identifying ligands. *Proceedings of the National Academy of Sciences USA* **87**, 6378–6382.

D'SOUZA, S.E., HAAS, T.A., PIOTROWICZ, R.S., BYERS-WARD, V., MCGRATH, D.E., SOULE, H.R., CIERNIESWKI, C., PLOW, E.F. AND SMITH, J.W. (1994). Ligand and cation binding are dual functions of a discrete segment of the integrin b3 subunit: cation displacement from this site is implicated in ligand binding. *Cell* **79**, 659–667.

DEVLIN, J.J., PANGANIBAN, L.C. AND DEVLIN, P.E. (1990). Random peptide libraries: a source of specific protein binding molecules. *Science* **249**, 404–406.

ELLINGTON, A.D. AND SZOSTAK, J.W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822.

HAHN, M., PIOTUKH, K., BORRISS, R. AND HEINEMANN, U. (1994). Native-like *in vivo* folding of a circularly permuted jellyroll protein shown by crystal structure analysis. *Proceedings of the National Academy of Science USA* **91**, 10417–10421.

HAMMER, J., TAKACS, B. AND SINIGAGLIA, F. (1992). Identification of a motif for HLA-DR1 binding peptides using M13 display libraries. *Journal of Experimental Medicine* **176**, 1007–1013.

HASHINO, K., SHIMOJO, T., KIMIZUKA, F., KATO, I., MAEDA, T., SEKIGUCHI, K. AND TITANI, K. (1992). Engineering of artificial cell adhesion proteins by grafting the Arg-Gly-Asp cell adhesive signal to a calpastatin segment. *Journal of Biochemistry(Tokyo).* **112**, 547–551.

JAIN, R. K. (1996). Vascular and interstitial barriers to delivery of therapeutic agents in tumors. *Cancer Metastasis Reviews* **9**, 253–266.

JUWEID, M., SHARKEY, R.M., BEHR, T.M., SWAYNE, L.C., DUNN, R., YING, Z., SIEGEL, J.A., HANSEN, H.J. AND GOLDENBERG, D.M. (1996). Clinical evaluation of tumor targeting with the anticarcinoembryonic antigen murine monoclonal antibody fragment, MN-14 F(ab)2. *Cancer* **78**, 157–168.

KOIVUNEN, E., GAY, D.A. AND RUOSLAHTI, E. (1993). Selection of peptides binding to the a5b1 integrin from phage display library. *Journal of Biological Chemistry* **268**, 20205–20210.

KOIVUNEN, E., WANG, B. AND RUOSLAHTI, E. (1994). Isolation of a highly specific ligand for the a5b1 integrin from a phage display library. *Journal of Cellular Biology* **124**, 373–380.

KREITMAN, R.J., PURI, R.K. AND PASTAN, I. (1994). A circularly permuted recombinant interleukin 4 toxin with increased activity. *Proceedings of the National Academy of Science USA* **91**, 6889–6893.

MAEDA, T., OYAMA, R., ICHIHARA-TANAKA, K., KIMIZUKA, F., KATO, I., TITANI, K. AND SEKIGUCHI, K. (1989). A novel cell adhesive protein engineered by insertion of the Arg-Gly-Asp-Ser tetrapeptide. *Journal of Biological Chemistry* **264**, 15165–15168.

MAKOWSKI, L. (1993). Structural constraints on the display of foreign peptides on filamentous bacteriophages. *Gene* **128**, 5–11.

MATTHEAKIS, L.C., BHATT, R.R. AND DOWER, W.J. (1994). An *in vitro* polysome display system for identifying ligands from very large peptide libraries. *Proceedings of the National Academy of Sciences USA* **91**, 9022–9026.

MEDYNSKI, D. (1994). Phage display: all dressed up and ready to role. *Biotechnology (N. Y).* **12**, 1134–1136.

MULLIS, K. B. AND FALOONA, F. A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods in Enzymology* **155**, 335–350.

OBARA, M. AND KATSUTOSHI, Y. (1995). Possible involvement of the interaction of the a5 subunit of a5b1 integrin with the synergistic region of the central cell-binding domain of fibronectin in cells to fibronectin binding. *Experimental Cell Research* **216**, 273–276.

OLDENBURG, K.R., LOGANATHAN, D., GOLDSTEIN, I.J., SCHULTZ, P.G. AND GALLOP, M.A. (1992). Peptide ligands for a sugar-binding protein isolated from a random peptide library. *Proceedings of the National Academy of Science USA* **89**, 5393–5397.

PASQUALINI, R., KOIVUNEN, E. AND RUOSLAHTI, E. (1995). A peptide isolated from phage

display libraries is a structural and functional mimic of an RGD-binding site on integrins. *Journal of Cell Biology* **130**, 1189–1196.

PASQUALINI, R. AND RUOSLAHTI, E. (1996). Organ targeting *in vivo* using phage display peptide libraries. *Nature* **380**, 364–366.

RUOSLAHTI, E. AND PIERSCHBACHER, M.D. (1987). New perspectives in cell adhesion: RGD and integrins. *Science* **238**, 491–497.

RUOSLAHTI, E. (1996). RGD and other recognition sequences for integrins. *Annual Reviews of Cell and Developmental Biology* **12**, 697–715.

SALGALLER, M.L., WEBER, J.S., KOENIG, S., YANNELLI, J.R. AND ROSENBERG, S.A. (1994). Generation of specific anti-melanoma reactivity by stimulation of human tumor-infiltrating lymphocytes with MAGE-1 synthetic peptide. *Cancer Immunology, Immunotherapy* **39**, 105–116.

SCOTT, J.K. AND SMITH, G.P. (1990). Searching for peptide ligands with an epitope library. *Science* **249**, 386–390.

SCOTT, J.K., LOGANATHAN, D., EASLEY, R.B., GONG, X. AND GOLDSTEIN, I.J. (1992). A family of concanavalin A-binding peptides from a hexapeptide epitope library. *Proceedings of the National Academy of Science USA* **89**, 5398–5402.

SMITH, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228**, 1315–1317.

SMITH, G.P. (1994) Applied evolution. The progeny of sexual PCR. *Nature* **370**, 324–325.

SMITH, J.W., HU, D., SATTERTHWAIT, A., PINZ-SWEENY, S. AND BARBAS, C.F. III. (1994). Building synthetic antibodies as adhesive ligands for integrins. *Journal of Biological Chemistry* **269**, 32788–32795.

SMITH, J.W., TACHIAS, K. AND MADISON, E.L. (1995). Protein Loop Grafting to Construct a Variant of Tissue Type Plasminogen Activator that Binds Platelet Integrin aIIbb3. *Journal of Biological Chemistry* **270**, 30486–30490.

STEMMER, W.P. (1994a). DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proceedings of the National Academy of Science USA* **91**, 10747–10751.

STEMMER, W.P. (1994b). Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature* **370**, 389–391.

VIGNAIS, M.L., CORBIER, C., MULLIERT, G., BRANLANT, C. AND BRANLANT, G. (1995). Circular permutation within the coenzyme binding domain of the tetrameric glyceraldehyde-3-phosphate dehydrogenase from Bacillus stearothermophilus. *Protein Science* **4**, 994–1000.

WANG, B., DICKINSON, L.A., KOIVUNEN, E., RUOSLAHTI, E. AND KOHWI-SHIGEMATSU, T. (1995). A novel matrix attachment region DNA binding motif identified using a random phage peptide library. *Journal of Biological Chemistry* **270**, 23239–23242.

YAMADA, T., SONG, H., INAKA, K., SHIMADA, Y., KIKUCHI, M. AND MATSUSHIMA, M. (1995). Structure of a conformationally constrained Arg-Gly-Asp sequence inserted into human lysozyme. *Journal of Biological Chemistry* **270**, 5687–5690.

ZHANG, T., BERTELSEN, E., BENVEGNU, D. AND ALBER, T. (1993). Circular permutation of T4 lysozyme. *Biochemistry* **32**, 12311–12318.