Searching for Molecular Solutions – **Cited Notes**

CHAPTER 4

These Files contain details on all references to this ftp site within **Chapter 4** of *Searching for Molecular Solutions*. The page numbers of the book where the reference is made are shown in the Table below, the corresponding page number for this file, and the title of each relevant section.

Contents:

Section 6:   *PCR*

Cited on p. 108 of *Searching for Molecular Solutions*

This section provides a brief description of the polymerase chain reaction (PCR). The general concept of the PCR process is depicted in Fig. 4.Na.

PCR was originated in the 1980s by Kary Mullis at Cetus Corporation, initially using a conventional DNA polymerase from *E. coli* to perform sequential extension reactions in a cycle of target template denaturation, primer annealing, and synthetic polymerase-mediated extensions [1]. But to denature long DNA duplexes without recourse to drastic changes to the solution composition (as with strong bases or other chemical denaturants), it is necessary to use heat. Thermal denaturation unfortunately cannot avoid 'killing' and inactivating DNA polymerases (such as those from *E. coli*) which are normally active at mid-range temperatures (mesophilic). This means that fresh mesophilic polymerase must be laboriously re-added after completion of each denaturation step in a repeated cycle. But enzymes from naturally thermophilic organisms do not suffer from this disadvantage, and the bacterium *Thermus aquaticus* has donated its very useful 'Taq' DNA polymerase for greatly simplifying PCR reactions [2]. Although many other alternatives exist, Taq polymerase and engineered versions of it remain the most extensively used enzymes in nucleic acid amplification.

Kary Mullis was awarded in 1993 Nobel Prize for Chemistry for his work the discovery and development of PCR. This was controversial, in large part because it was considered by many that PCR was built upon previous results in the laboratory of Har Khorana [♥][3].

[♥] Khorana himself received the Nobel Prize in 1968 for his work in synthetic DNA chemistry, which was fundamentally important in the development of the now-routine synthesis of oligonucleotides. Without the ability to make oligodeoxynucleotide primers rapidly and cheaply, the utility of PCR itself would be severely curtailed.

N DNA molecules

Forward primer
5' ——————————— 3'

**Reverse Primer
complementary sequence**

5' ——— 3'
|||||||||||||||||||||||||||||||||||||||||||||||||||||||
3' ——— 5'

**Forward Primer
complementary sequence**

3' ——————— 5'
Reverse primer
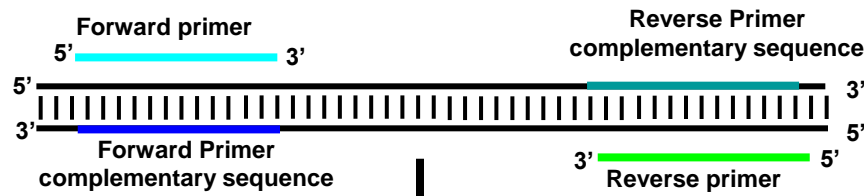
THERMAL CYCLING: *Denature; Anneal Primers; Extend*

Taq DNA Polymerase

5' ———————
||||||||||||
—————————————— 5' --> 3' strand synthesis

Cycle 1: 2N copies

Cycle 2: 4N copies

Cycle *n*: 2^n x N copies
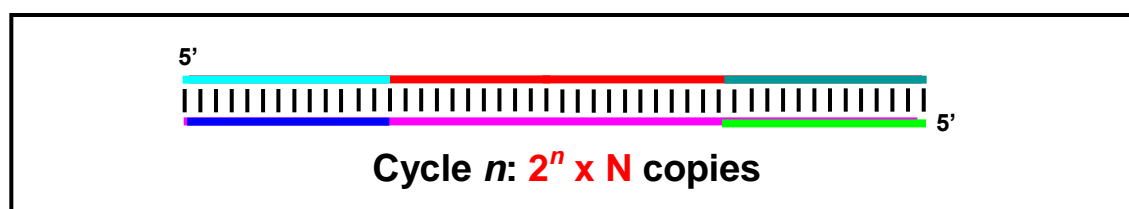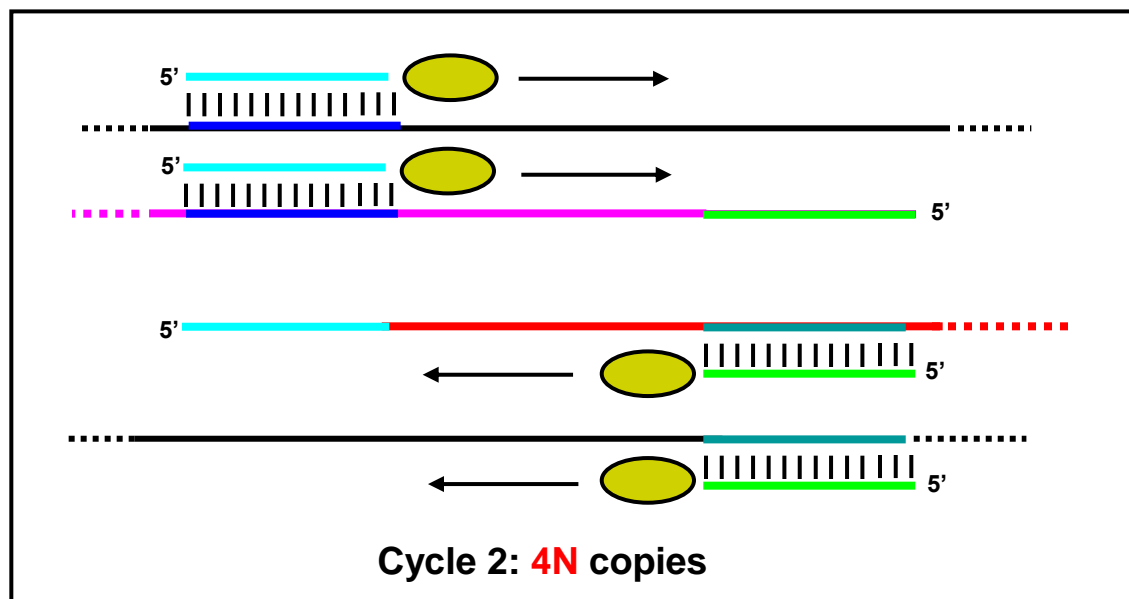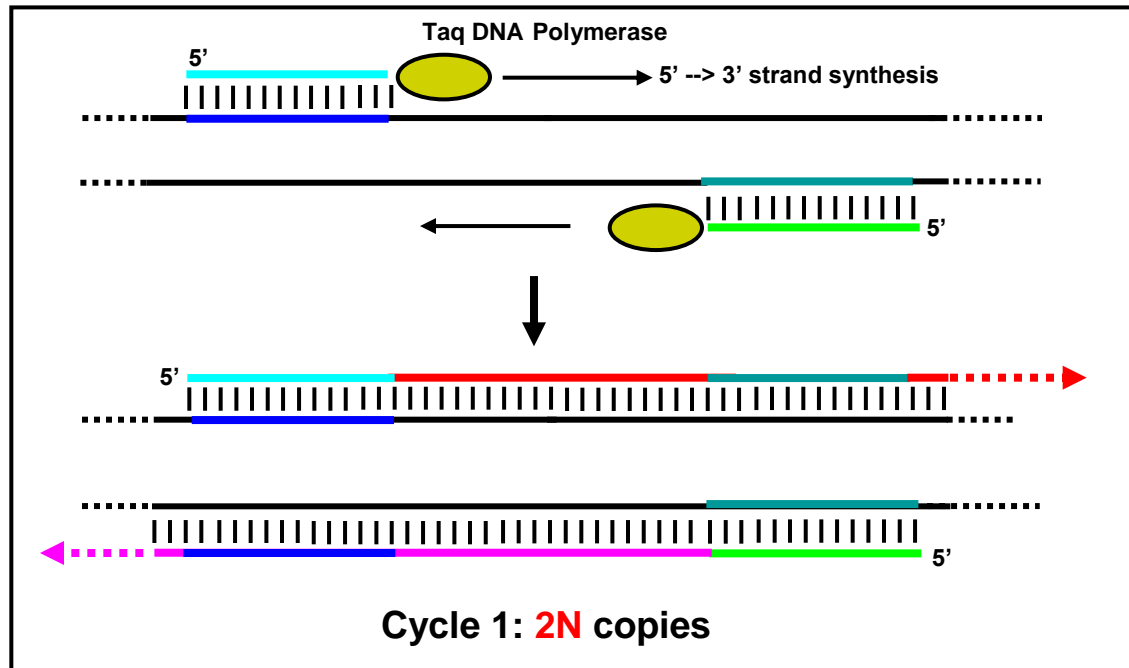
Fig. 4.Na

Schematic depiction of polymerase chain reaction (PCR) process. Primers are designed such that they are complementary to sites within an extended target duplex sequence, and prime polymerization in opposing directions (primer sequences themselves match sequences on opposite strands). In the first cycle, strands are thermally denatured, primers allowed to anneal at their respective complementary sites, and primer extension (DNA synthesis) initiated through the presence of a thermostable DNA polymerase (shown here as Taq polymerase) and deoxynucleotide triphosphates (dNTPs). Newly synthesized strands are shown in red and purple for 'top' and 'bottom' strands respectively, with segments corresponding to complementary sequences for the forward and reverse primers shown in dark blue and dark green, respectively. Repeating the cycle with the newly synthesized strands results in a progressive amplification as indicated (2^n.N DNA molecules for n cycles and N starting molecules). The product is bounded by the priming sites and normally is of a precisely defined length.

Although not relevant to this debate, Mullis's colorful lifestyle and highly unconventional opinions on unrelated subjects [4] tended to amplify the controversy. But there was also a feeling in some quarters that while the achievement itself was significant, it was nonetheless a 'mere technique' and did not warrant the Nobel accolade. Other issues aside, it is quite clear that as a molecular biological method, PCR stands head and shoulders above the rest. In Chapter 4 of *Searching for Molecular Solutions*, some techniques which have not only been widely adopted but inspired a wide range of auxiliary procedures were noted as 'metatechniques'. If any technique is worthy of this title, it is certainly PCR. Its overall impact has been compared with the advent of the internet [5].

At the core of the power of PCR is the requirement for nucleic acid amplification in a huge range of molecular biological applications. Also, although the principle of PCR is in fact quite simple (Fig. 4.Na), this technique has lent itself to an amazing range of applications, and has inspired the devising of other

amplification methods (to note further below). That PCR could be harnessed for such a plethora of uses was not immediately obvious to most observers at the outset. This is suggested by anecdotes reporting that Mullis's superiors at Cetus could not at first see the full value of the technique. More usefully, it can be pointed out that if obvious, the range of PCR applications would have followed closely upon its general availability in the late 1980s, when in fact there has often been a prolonged gap. The spin-offs from a high-level metatechnique can take time to take shape.

Anything approaching a comprehensive review of the wealth and variety of PCR applications is accordingly far beyond the scope of this section (some specific cases are mentioned at various points within *Searching for Molecular Solutions*). Still, to give a hint as to the range of the versatility of PCR, we can note some categories of its applications. Very broadly speaking, PCR is useful either for analytical purposes, or for the directed generation of altered DNA sequences. But in conjunction with either of these generalized applications, PCR can be also used for preparative purposes, simply by the generation of large amounts of desired DNA segments as defined by primer choice (Fig. 4.Na).

What PCR Can Do

Not surprisingly, the ability of PCR to amplify tiny amounts of starting material (in the limit, one molecule of target) is central to its vast array of analytical applications. But even if an initial DNA target is abundant (as from a readily-grown cellular source), by the nature of the process, two primers allow the amplification of a specific region from a huge background of irrelevant sequences, and thus provide a rapid way of obtaining a sequence segment of interest. This inherent feature of PCR in 'finding a needle in a haystack' has made it an invaluable aid in molecular cloning of genes when only a limited amount of sequence information is available [5]. Although the thermostable

polymerases used for the PCR process act on DNA, RNA targets are readily amplifiable by an initial reverse transcription step (thus 'RT-PCR').

All of these advantages underscore the use of PCR in single-cell analyses [6]; forensics [7], identity testing [8] and genotyping [9]. A vast number of PCR diagnostic applications also exist [10-13], which include germline or somatic mutation detection [14,15]. Many genetic analytic procedures involve additional essential processing steps which enable the specific PCR amplification of interest. Without going into detail, examples of these are techniques with DNA ligation-dependent steps: 'inverse PCR' [16]; PCR-amplified genomic footprinting [17], and digestion-circularization PCR for detection of gene deletions [18]

Because product formation during PCR grows at an exponential rate and reaches a saturation point dictated largely by available deoxynucleotide triphosphate (Fig. 4.Na), measurement of the amounts of product at the end of standard cycling is a poor means of quantitating the amounts of starting target. This was circumvented initially by using internal controls and limited cycle numbers for generating quantitative calibration curves, and later through the 'real-time' ♥ generation of fluorescent signals during the amplification process itself by various technological innovations [19,20].

In the Cited Notes for *Searching for Molecular Solutions* for Chapter 2, the Metagenomics section ♣ made brief mention of another application of PCR: the ability to amplify single molecules as 'molecular colonies' or 'polonies' (polymerase colonies) [21,22]. This can be likened to the difference between growth of a mixed bacterial population in liquid suspension compared with isolated

---

♥This indicates that the signals are acquired during the course of the amplification process itself, rather than at the end-point completion of the PCR run. Real-time data acquisition enables accurate measurement of the number of cycles necessary to produce detectable fluorescence, which correlates directly with the initial amount of starting material.

♣ SMS-CitedNotes-Ch2/Section 2, from the same ftp site.

colony growth from single cells on solid media. In the latter case, each colony corresponds to a bacterial clone, and the analogous solid-phase polony technique also provides an *in vitro* molecular clone for subsequent analyses.

While 'metagenomics' is not exactly a household word, a special area of activity within this field has captured even the popular imagination: reconstructing genomes from ancient DNAs of fossil origins. While 'Jurassic Park' scenarios are not going to move from science fiction to reality any time soon, some amazing achievements have indeed been accomplished. Perhaps most notably, a draft genome sequence of the extinct human relative, the Neandthals, is due for release in late 2009 [23]. This feat has been rendered feasible by recent advances in novel DNA sequencing methods, but the 'next generation' technologies now commercially available for such high-throughput sequence determinations all rely on initial PCR-based amplifications [24,25]. Two separate instances of these advanced sequencing approaches use emulsion-based PCR amplification, corresponding to the *in vitro* compartmentalization technology considered in Chapter 4 of *Searching for Molecular Solutions*.

Yet an important caveat must be noted at this point, since recent advances have rendered PCR unnecessary for high-throughput DNA sequencing *per se*. In fact, dense parallel-processing sequencing be performed at the level of single molecules ♥, which actually by-passes the need for an amplification step [26,27]. Moreover, this technology is scalable to the level of rapid human genome sequencing [28]. Avoiding PCR eliminates the potential drawback of amplification-induced errors and artefacts (a subject considered in a little more detail below), and according to the manufacturer of single-molecule technology sequencers ([Helicos Biosciences](#)), this technology is superior in analysis rate and cost-effectiveness to the above-mentioned 'nest-generation' alternatives. While single-molecule sequencing therefore has great appeal, it is hard to predict that it will

---

♥ This technology is based on capture of fluorescent signals from polymerase-incorporated nucleotides by microscopy [26].

entirely eliminate the need for PCR-based amplification to obtain sequencing template. One such application where amplification is likely to remain a requirement in at least some cases is the analysis of fossil DNAs, as considered above.

PCR can be creative in the literal sense of introducing virtually any desired mutations, insertions, or deletions at a pre-determined site within a DNA segment [5]. A basic principle of note here is the fact that homology with target template is important at the 3' end of primers, over a long enough tract such that stable base pairing results at the annealing temperature used for the specific PCR cycle (Fig. 4.Na). As long as this criterion is satisfied, mismatches or even long non-homologous segments at the 5' end of primers are acceptable, and such sequences will be incorporated into the final duplex product. Mutational changes within a continuous segment are arranged by initially amplifying two separate duplexes, each with a 5' homologous segment bearing the desired sequence alterations. A second round of PCR allows strands of the two segments to prime on each other by virtue of their homologous tracts, and the entire fused fragment (with the desired mutation) is then amplified by primers encompassing the entire region. In an analogous manner, fusions between any DNA segments of interest can be readily accomplished in a matter of hours.

PCR Limitations and Problems

It has often been noted that the great power of PCR is also its greatest weakness. Since PCR can amplify even single molecules millions of times up to the threshold of detection, any contaminating DNAs with homologous primer sequences in the reaction can also be subject to amplification. Such artefacts can be a serious problem, and rigorous steps must be taken to prevent contamination from occurring. (A basic dictate is that any PCR test must have 'blank' no-template controls to confirm that nothing is amplified in the absence of deliberately added template). Even if no environmental contaminants are present

during the reaction run, the DNA target source itself can sometimes be problematic. For example, published results of some of the above-mentioned Neanderthal sequencing projects have been questioned, and discrepancies attributed most likely to human contamination of the original Neanderthal DNA source [29].

Other PCR artefacts can arise from cross-priming of sites within amplified templates with other DNA segments present in the initial target mixture. The starting material (as noted above) for many PCR applications is dependent upon a series of sometimes complex initial steps, which can also be potential sources of spurious 'amplicons' (amplified segments) unless rigorous control measures are taken. As an example, the reverse-transcriptase step in RT-PCR is known to be capable of generating artefactual 'splicing' by template switching, leading to the subsequent amplification of spurious PCR products [30]. A simple means for improving the specificity of PCR amplification is to use a 'nesting' system, where the products of an initial amplification round are subjected to another amplification with two additional primers both bounded within the initial amplified segment. (Theoretically, only the correct amplified segment will possess the additional second priming sites; spurious material arising from PCR with the first primer set will fail to be amplified with the second).

A PCR limitation observed virtually from the outset was the maximal size of amplified segments, which appeared to be only feasible up to ~2000 base pairs. To a large extent, this has been overcome by 'long-distance' PCR protocols, whose basic innovation is the provision of small amounts of a proof-reading thermostable polymerase ♥ [31]. The rationale for this is that Taq DNA polymerase and its kindred do not possess a 3' → 5' exonuclease activity, and cannot remove non-extendable nucleotide derivatives present at low levels in any reaction. Chains which incorporate such 'terminating' nucleotide products are

---

♥ See SMS-CitedNotes-Ch2/Section 1, from the same ftp site, for more details regarding polymerase proof-reading.

thus blocked from any further extension, and such chance incorporations serve to create a limit on the average longest possible amplification product. Inclusion of proof-reading activity can solve this problem, allowing segments in the tens of thousands of nucleotides to be amplified.

A related problem in general is PCR fidelity, since products amplified by Taq polymerase have a high error rate resulting from lack of correction of misincorporated (though still extendable) nucleotides. (Indeed, this has been exploited for deliberate introduction of errors for directed evolution and other related applications, although often the desired error rate has to be boosted above the Taq-induced level). The error background is not usually a problem if a PCR product is directly sequenced, since any single error occurs at a low frequency and the 'consensus' sequence corresponding to the original target template will be revealed. But if PCR products are cloned, then specific errors within single amplified molecules are fixed and therefore a potential confounding factor. A variety of additional polymerases from thermostable prokaryotes have been introduced which do possess proof-reading activities, and these can have dramatically improved fidelities over conventional Taq polymerase [32].

PCR extensions and analogies

Thermal cycling has been an inherent and essential feature of PCR since its inception, requiring the use of machines to accurately and rapidly switch between temperatures as required by specific reaction cycles. Yet DNA amplification *per se* is not necessarily restricted to such conditions, as is immediately apparent from considering the replication of living systems themselves. Indeed, prior to the advent of PCR, molecular biology profited handsomely from exploiting bacterial cells as vehicles for the accurate amplification of extrachromosomal plasmids. These systems clearly work at constant mid-range temperatures, and the example of PCR as an *in vitro* amplification system has stimulated searches for artificial isothermal amplification systems, a number of which have been

successfully developed [33]. One such example was also noted in the Cited Notes for *Searching for Molecular Solutions* for Chapter 2 ❥ in the form of isothermal amplification with a DNA polymerase from the bacteriophage ⌀29.

In Chapter 8 of *Searching for Molecular Solutions*, the idea of nucleic acids as encoding tags in various forms for chemical syntheses was considered. Many of these require the tag amplification through a PCR step, and a related concept is to marry the ability to amplify such a tag with the combining specificity of antibodies. Conjugates of specific antibodies and DNA tags have thus been exploited for 'immuno-PCR' [34]. In most such applications, the desired feature is a PCR-based increase in detection sensitivity of immunodetection assays, and in such instances the read-out is of a digital nature where the nature of the DNA sequence tag is secondary to the reporting of an amplification signal itself. (A positive signal indicates a theoretical antibody binding event). While the DNA tag sequence need not in such circumstances report any further information, multiplexing of such systems is possible with different fluorescent-labeled primers for tag amplification.

A rather more direct 'protein amplification' has also been inspired by PCR. The unusual proteins termed prions were referred to in Chapter 5 of *Searching for Molecular Solutions*, whose defining feature is their ability to impart their aberrant folding state to normally-folded proteins of the same primary amino acid sequences. (The importance of prions in human pathologies such as Creutzfeldt-Jakob disease is now well-recognized). The inherent transmissibility of the aberrant folded state has suggested that cycles of conversion of normal protein by prion versions could allow the detection of trace amounts of potentially dangerous prions. This has been realized in the 'Protein Misfolding Cyclic Amplification' assay, or PMCA [35,36].

❥ SMS-CitedNotes-Ch2/Section 2 (Metagenomics); from the same ftp site.

Section 7; Part A   *Alternative Phage Display*

Cited on p. 135 of *Searching for Molecular Solutions*. This section provides some details on alternative display technologies. The first subsection (Part A) considers some additional phage systems, and then Part B looks at a wider range of systems.

Alternative phage display technologies

Since the maturation of filamentous phage requires extrusion through the bacterial cell membrane rather than lysis, displayed peptides must be compatible with this process, and certain peptide sequences are consequently under-represented in filamentous display libraries ♥. This prompted the development of a number of alternative lytic phage display system, including the much-studied phage lambda ♣. Broadly speaking, the morphology of lambda is shared with many types of lytic phage, where the phage genome is contained within an icosahedral protein capsid 'head' attached to a tubular tail through which the phage nucleic acid is ejected during passage into the host cell. Both head and tail proteins of lambda can be used for display purposes separately [43-45], or together [46] (Fig. 4.Nb). Other lytic phage display systems have also been

---

♥Such peptides can be considered a special case of 'restricted sequences', which can be compared with the more general list in Table 3A2.1 in *Searching for Molecular Solutions* Extras, (SMS-Extras-Ch3/Section A2; from the same ftp site). Yet at least some polypeptides which are normally poorly displayed on filamentous phage can have their display levels greatly improved by altering signal sequences for protein translocation [37-39], or through mutations in a component of the *Sec* secretion pathway [40]. Also, certain mutations in gp8 themselves improve display of proteins [41]. Overexpression of certain periplasmic chaperones may also be beneficial in some circumstances [42]. Therefore, the exclusion of polypeptides from display by filamentous phage is conditional for at least a subset of the total.

♣ Lambda can undergo both lytic and lysogenic life cycles (where the latter involves stable integration of the phage genome into the host *E. coli* chromosome). Many lambda vectors have the lysogenic genes removed, but sometimes the capacity for lysogeny is a useful feature.

developed [47,48], and in the phage T7 system, less sequence bias has been noted in peptide display libraries than for corresponding libraries in filamentous phage [49].
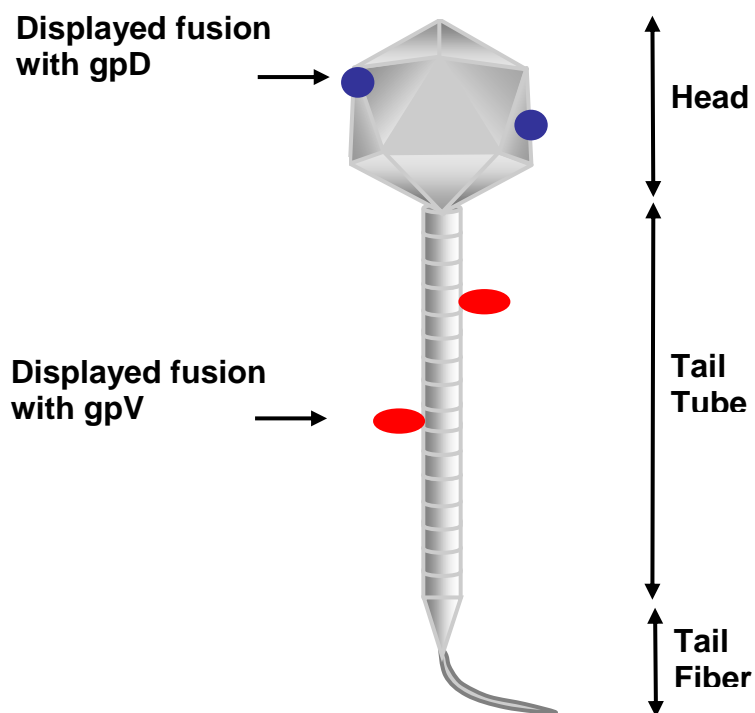
Fig. 4.Nb

Schematic representation of peptides or proteins displayed on the bacteriophage lambda. The capsid is stabilized by the product of gene D (gpD) which can display fusions at either its N- or C-termini [50]; the tail tube is composed of the product of gene V (gpV) which can display fusions at its C-terminus [43,44].

Display has been achieved with the lytic phage T4 by means of a protein which is dispensable for capsid formation [51-53]. This also has the interesting feature that foreign fusions with the dispensible protein ('small outer capsid protein', or SOC) can be incorporated *in vitro* after the capsid has been pre-assembled. In such circumstances, there is no necessary link between the phage genotype and the novel displayed molecule, since a SOC fusion can be prepared elsewhere (encoded by an entirely separate genetic element) and independently inserted into the capsid. The contrast between such T4 display with and without a genotype-phenotype link is depicted in Fig. 4.Nc, which helps to illustrate the point that display *per se* does not necessarily involve the latter linkage. As emphasized in *Searching for Molecular Solutions*, it is essential that the genotype-phenotype correspondence is preserved where display technologies are applied towards directed evolution.
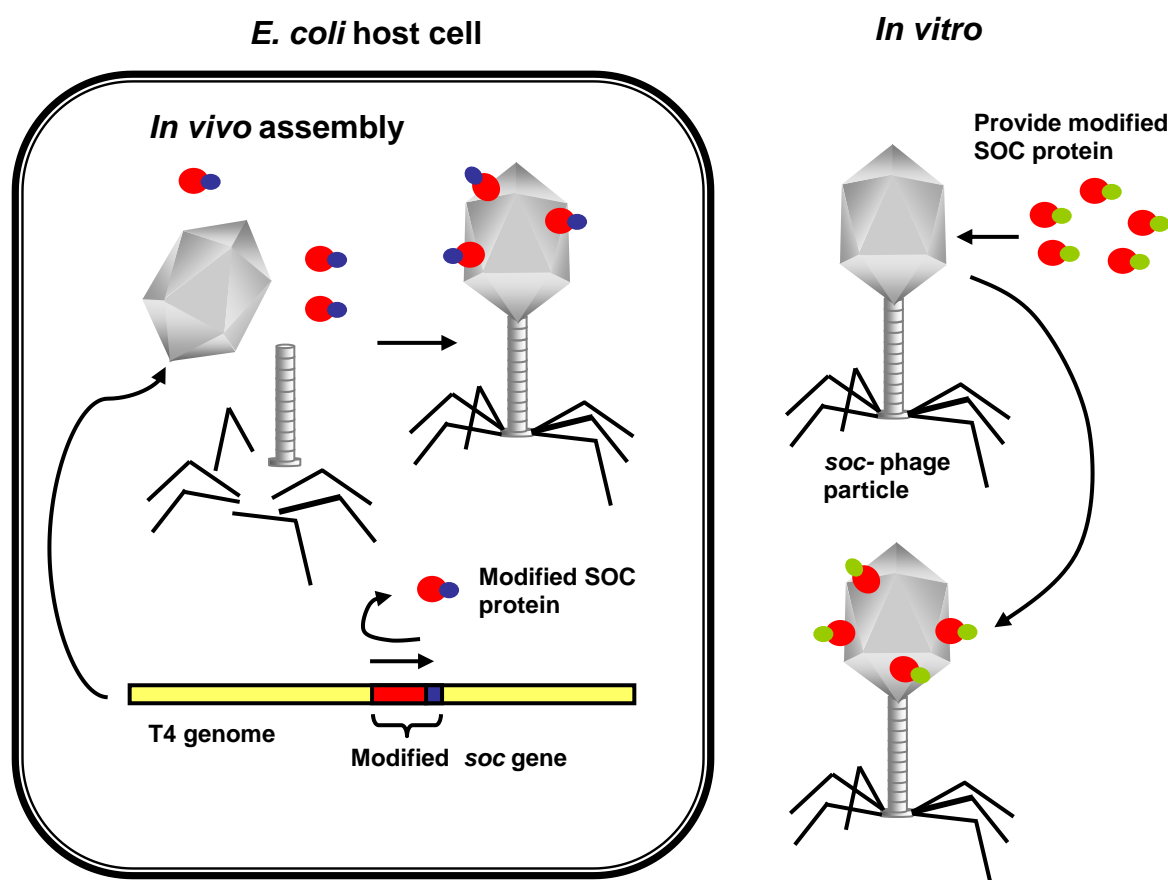
Fig. 4.Nc

Depiction of surface display involving the Small Outer capsid (SOC) protein in bacteriophage T4 with a genotype-phenotype linkage (*in vivo* in host cell) and without such a linkage (*in vitro*, as shown). With the *in vivo* case, a SOC fusion is incorporated during virion assembly and displayed on the capsid surface, and the corresponding phage genome is packaged within the same capsid structure. Since SOC fusions can bind to pre-formed capsids which have initially assembled in the absence of SOC, any structurally-compatible SOC fusion can be used for *in vitro* capsid incorporation irrespective of the phage genotype.

Beyond peptides, sizable proteins have been successfully displayed on phage [43,44] (detailed further in Section 8A below), and a variety of phage systems have been used for the display of cDNA libraries [54-57]. The latter type of display libraries can be conveniently probed to discover the receptors for a ligand of interest [58]. A natural extension of this has been the use of phage display systems as a general tool for identifying and evaluating protein-protein interactions, which is a very useful adjunct to the yeast two-hybrid system (described below in Section 9) [59-63].

Thus far we have a picture of phage display as a robust, versatile, and continually developing technology. In many cases, phage display has been used as a means for empirically obtaining information for which it is unnecessary to repeat the experimentation once acquired. In other words, in fields where phage display has 'done its job', its subsequent usage may decline. (A possible example of this is the deployment of tailor-made zinc finger binding proteins, (also described briefly below in Section 8B). On the other hand, new applications of phage display, such as in materials science [64,65] continue to emerge, making prediction of its long term prospects more difficult.

Section 7; Part B   *Alternative General Display*

Cited on p. 135 of *Searching for Molecular Solutions*

Alternative General Display Technologies

Let us return to the central issue of display, which is represented in Fig. 4.9 ♥ of *Searching for Molecular Solutions*. Phage are certainly not the only avenue by which genotype and phenotype can be linked, nor the most direct. Consider the following arrangement: a library of variant peptides or proteins are expressed as fusions with a functional DNA-binding domain. These are encoded as a library within a bacterial vector (usually a multicopy plasmid) which also contains at least one copy of the target binding site for the DNA-binding domain. As such, a covalent linkage exists between this target binding site and the coding sequence for the fusion protein itself. Upon expression of the polypeptide-DNA protein fusion in bacterial host cells, the plasmid is bound by the fusion protein, and thus a genotype-phenotype link is established.

This can be regarded as yet another example of the modularity of DNA-binding domains ♣. In principle, any DNA-binding domain would suffice for this purpose, as long as its specificity and binding affinity for the target plasmid DNA sequence were acceptable. (Clearly it is essential to have strong binding for this application, to preserve a *bona fide* link between genotype and phenotype). The above general display consideration has indeed been realized, using DNA-binding motifs as diverse as bacterial repressors [66,67] or eukaryotic transcription factors [68]. At the DNA level, a library of variable polypeptide coding sequences fused in-frame to a constant DNA-binding region can be prepared, somewhat analogously to the natural situation of immunoglobulin or T cell receptor variable

---

♥ See also the color version of this Figure, at SMS–ColorFigures.pdf; from the same ftp site.

♣ Considered further in SMS–CitedNotes-Ch3/Section 4; from the same ftp site.

regions and their corresponding constant region genes (Chapter 3 of *Searching for Molecular Solutions*). If each library member is expressed in a separate host cell compartment (as is normally the case), then the DNA binding protein 'grip' on the specific plasmid allows the corresponding coding sequence to be isolated in parallel with selected protein or peptide function expressed within the variable library.

There are some additional special cases of DNA display, but these are best considered as categories of *in vitro* display, as covered in Chapter 3. While the most direct link between genotype and phenotype is made by linking a polypeptide with its encoding nucleic acid, an intervening protein coat still allows this linkage, as we have amply seen with bacteriophages. And furthermore, we have already seen that an entire cell can in principle act as a display 'tag' bearing the genetic information encoding a selected phenotype (Fig. 4.9 of *Searching for Molecular Solutions*), provided of course each cell provides a defined environment for the expression of a specific library member. Surface expression on bacterial cells has indeed been realized for display purposes, using a number of alternative outer membrane proteins [69-71] (which naturally span the boundary of the outer environment into the inner periplasmic space) as protein scaffolds for presentation of foreign inserts.

The enzyme phosphotriesterase (a member of TIM barrel family, a recurring theme in *Searching for Molecular Solutions*, with an example in Fig. 5.2) is one example of a protein which has been displayed on the surface of *E. coli* for directed evolution. In this case, an alternative surface protein (an ice nucleation protein from another bacterium) was used as the display vehicle [72,73]. Whole cell display has the convenience that the presentation system and replication compartment coincide, and allows fluorescence-mediated sorting and high-throughput screening of displayed peptides or proteins [74]. A fluorescence-based screening strategy using bacterial display of proteases has been notably successful in isolating enzyme variants with altered substrate specificities [75,76].

Bacterial display is also useful for certain specific applications such as live vaccines [77]. On the other hand, displayed structures on bacterial cells have similar constraints as filamentous phage, where they must be compatible with membrane transport in order to progess to the external environment. Phage display systems in general have been used to a far greater extent than whole-cell display, partly through the greater versatility of available phage options and partly owing to their earlier historical development.

But both bacterial and phage surfaces are by definition prokaryotic systems, and while expression in such environments is very versatile, it has certain limitations. For example, glycosylation and certain other post-translational modification patterns seen in eukaryotic cells will not normally be recapitulated in bacteria. Sometimes eukaryotic protein folding itself is inefficient in these circumstances, especially for large proteins with complex subunits ♥. Accordingly, eukaryotic display systems have been developed. By analogy with phage, it would seem a logical step to evaluate eukaryotic viruses as display vehicles, and this has been successfully applied. Display systems for baculoviruses [78,79], retroviruses [80,81], and adeno-associated virus [82] have been developed. Also by analogy with the prokaryotic world, surface display in eukaryotic cells with suitable transmembrane proteins is possible, initially in yeast [83] and subsequently in mammalian cells for antibody display [84,85]. The decision to use a eukaryotic display technology for directed evolutionary purposes in preference to a more standard prokaryotic system will depend in large degree on the nature of the displayed target library. For example, a mammalian cell line would not be the first choice for many display applications due to the relative difficulties (over prokaryotic alternatives) in making and maintaining large diverse libraries, but

---

♥ As a relevant example, we could note the case of land plant Rubisco enzyme of photosynthesis, which has been refractory to functional bacterial expression. (Discussed further in SMS–Extras-Ch5/Section A6; from the same ftp site)

may be competitive where mammalian codon usage, folding, and processing are over-riding factors [84].

Section 8, Part A: *Protein Display on Phage*

Cited on p. 137 of *Searching for Molecular Solutions*

General Protein Display with Phage Systems

This section provides an overview of phage display technologies for proteins. Within this general area we find display of zinc finger proteins and enzymes, which are considered in more detail in Section 8B and Section 11 below. Phage display of proteins is a powerful tool for directed evolution by repeated rounds of *in vitro* selection in favor of altered protein-protein or protein-ligand interaction specificities.

Discrete protein domains [86] or whole proteins of small to moderately large size [87,88], (including hormones [89,90]) have been displayed on phage, and variants selected with improved binding of receptors or desired ligands. The *in vitro* 'tuning' of antibody specificities by display and selection for binding is a special case of this kind of process (as noted in *Searching for Molecular Solutions* Chapter 7), which is analogous to natural *in vivo* affinity maturation (Chapter 3).

Protein binding and protein folding are inter-related. An implicit requirement for the success of a phage display binding experiment is that the protein of interest expressed on the phage surface must be capable of assuming the correct folded state. This principle has been validated [91] and has actually been used as the basis of folding assays [91,92]. Destabilization of protein binding function will tend to occur if long foreign sequences are grafted into loop regions, while corresponding retention of protein activity is associated with the uptake of a stable folded state by the foreign inserted polypeptide [92,93]. Through phage display and selection for functional binding retention, this principle is accordingly a means for identifying 'foldable' sequence grafts [92]. Random fragments of native binding proteins which retain ligand recognition are likewise selectable, and this

constitutes an effective means for identifying the minimal regions of a protein required for a defined functional binding task [94].

A properly folded protein will also be more resistant to proteases than a sequence variant whose folding is even partially disrupted (such resistance is also a correlate of thermal stability, as has been exploited by the 'Proside' strategy [♥]). Proteases can thus be used in conjunction with phage display to probe folding for a number of basic and applied studies. Here the Proside approach can be instituted, or alternatively a related technique where the protein of interest is also linked with a selectable binding tag (for positive selection of relevant gp3 display phage) which is removed if proteolysis occurs [96-98]. Proteolytic selection and filamentous phage display have been exploited for finding folded proteins resulting from a known protein domain fused with combinatorial assortments of random coding sequences, the success of which has implications for protein evolution from non-homologous recombination events [96]. Using proteins initially showing sub-optimal folds, the same approach has been used to select for variants from random mutant libraries with enhanced folding stability [97], a finding applicable to protein design in general. By combining display / proteolytic selection with high-throughput sequencing, it is possible to gain detailed information regarding the sequence correlates of specific protein folds [99].

Despite considerable progress, accurate prediction of protein folding still eludes us in general (considered in more detail in Chapter 9). In Chapter 5 it was noted that the same catalytic activity in numerous cases can be performed by proteins with quite distinct folds (analogous enzymes; Fig. 5.3), an observation which raises many general questions relating to protein sequences, function, and

---

[♥]This procedure relies on differential susceptibility to proteases as a correlate of partial unfolding. Displayed protein segments with improved folding will also exhibit increased protease resistance, and are accordingly selectable on this basis [95]. The Proside approach will be considered further in below.

folding. As an example of one of these, we could ask: what is the maximum possible similarity of primary sequences between proteins with radically different folds? This question should make us think again of the 'Paracelsus Challenge' raised in Chapter 5, which called for design of proteins with ≥50% sequence similarity but divergent folding [100]. Although this challenge has been met [101], the limits of this sequence similarity / structural divergence were not defined in themselves. Here phage display again raises its useful (if not pretty) head. By starting with two separate proteins with distinct folds but equivalent binding functions, it is possible (by means of mutagenesis, phage display, and selection for retention of binding) to iteratively make changes in both of the proteins such that progressive sequence convergence results in the altered 'heteromorphic' pair. Obviously, with the stringent selection applied by the display system, only folded and functional proteins will be 'passed', and thus the limits of sequence convergence (between a specific protein pair, at least) can be probed. In this manner, the structurally distinct proteins streptococcal protein G and staphylococcal protein A (both of which bind a specific region of human immunoglobulin G) were evolved towards a heteromorphic pair with almost 60% sequence identity from an initial identity of only 14% [102]. Selection of protein folding in general by phage display takes us into the realm of *de novo* protein design, also considered further in Chapter 5 of *Searching for Molecular Solutions*.

Section 8, Part B: ***Zinc-finger Protein Display and Zinc-finger Nucleases***

Cited on p. 137 of *Searching for Molecular Solutions*

This section provides more detail on the specialized aspect of protein display concerning zinc finger DNA-binding motifs. In order to accomplish this, some general background concerning this motif is also provided. In addition, the establishment of a 'finger code' for DNA recognition by zinc fingers has been associated with the development of zinc finger nucleases as an important tool in cellular and molecular biology, with numerous therapeutic implications. As an adjunct to the DNA-binding protein Cited Notes Section 4 ♥, a brief account of zinc-finger nucleases is also given here, especially for contrasting with analogous applications of homing endonucleases.

Fingering Phage and Beyond

Of all the proteins which have been successfully displayed on bacteriophages, there is another set which is worth highlighting in a little more detail. These are the category of DNA-binding proteins known as 'zinc finger proteins', and are very significant from a scientific and biotechnological point of view. 'Zinc fingers' are small protein domains stabilized by coordination of zinc ions (Fig. 4.Nd), which were initially characterized as a major class of eukaryotic DNA-binding transcriptional regulators, including both activators [103-105] and repressors [106]. It has become apparent that beyond this role, zinc finger proteins may also bind RNA [107] and participate in protein-protein interactions [103,108]. There are also numerous different structural classes of zinc finger proteins [108,109], but for our purposes here we will be concerned with the 'classic' zinc finger motif [110,111], which has a '$\beta\beta\alpha$' structure based on its β-strand and α-helical content (Fig. 4.Nd). The 'classical' zinc finger is also referred to as a conserved 'Cys$_2$-His$_2$'

♥ SMS-CitedNotes-Ch3/Section 4; from the same ftp site.

motif, based on two histidine residues (in the C-terminal region of the α-helix) and two cysteine residues (near a turn in the antiparallel β-strands) which coordinate the zinc ion [111,112].

From structural studies [113-115], it became evident that each 'finger' of a zinc-finger DNA-binding protein contacts a specific triplet DNA base sequence, based on amino acid sequence variation outside the conserved Cys_2-His_2 motif. This in turn suggested that zinc finger DNA recognition had a modular nature. It was realized by three separate groups at around the same time (in 1994) that if zinc finger proteins could be displayed on phage, a powerful means should emerge for correlating specific finger protein sequences with target DNA sequence. This was borne out in practice in filamentous phage systems [116-118].
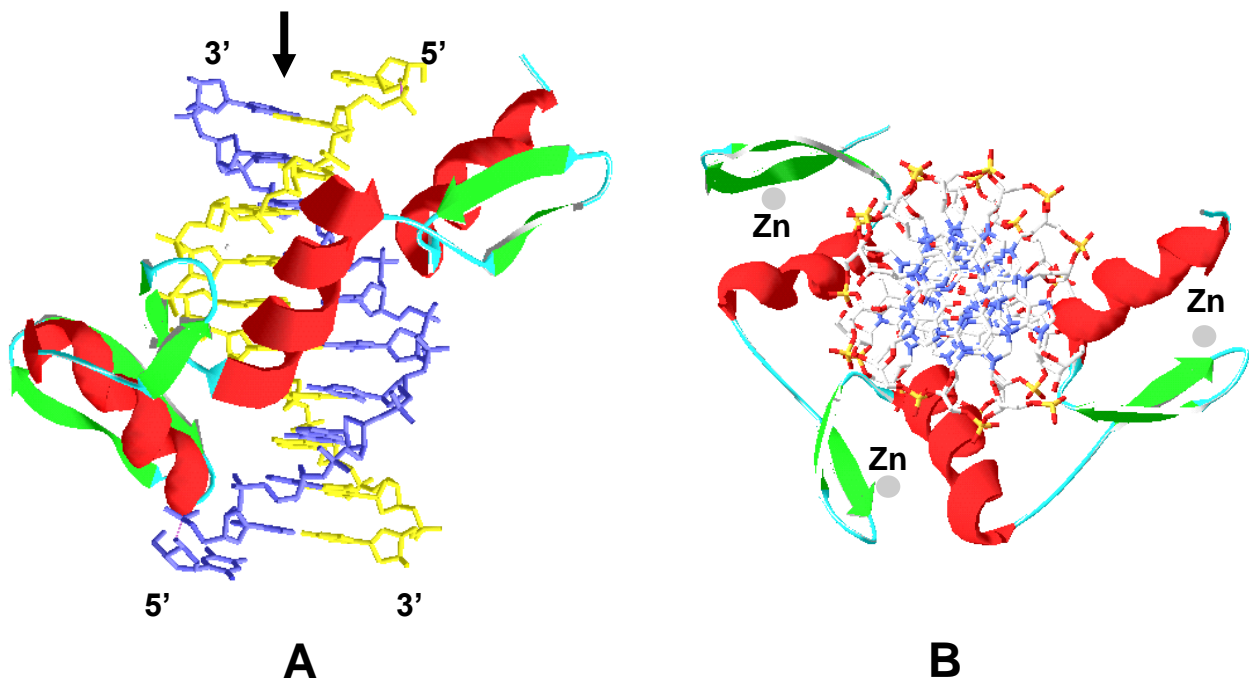

Fig. 4.Nd

Structure of a zinc-finger protein (the mouse intermediate-early protein Egr1 or Zif268) in complex with its DNA target sequence [113]. **A,** Showing DNA duplex (whole strands with

distinguishing colors) and orientation of the three bound zinc fingers where the α-helix of each fits into the major groove of the DNA. Arrows in structural ribbons point in C-terminal direction. **B**, as for A but looking down the end of the duplex as shown by the arrow in A, with the chelated zinc ions as indicated. Here the DNA duplex base pairs and phosphodiester chain are differentially colored for clarity. Source: Protein Data Bank [119] 1ZAA. Images generated with Swiss-pdb viewer [120].

Phage display-mediated selection for finger domains recognizing specific base triplets was generally successful [112,121-123]. The aim here was to attempt to establish a 'recognition code' for DNA by zinc fingers through which any DNA sequence could be targeted. If this was not quite the ultimate 'holy grail', it remained a rather important quest nonetheless. If one then linked multiple defined fingers together, theoretically it would produce a DNA-binding protein with a pre-defined recognition specificity. Multiple zinc fingers can indeed be assembled together to produce a multi-finger protein (or a protein with a 'polydactyl' nature, if you fancy a fancier label) with sequence-specific DNA binding capacity [124]. An important aspect of the modular 'stringing together' of separate zinc finger domains is the nature of the inter-finger linkage, which can be in the form of a physical linker, or alternatively through a non-covalent dimerization domain [112]. Extended linkers also allow the possibility of non-contiguous DNA target sites [125,126].

Given that some additional complexities have suggested that DNA recognition by each finger is not completely independent [112,127,128], generation of specific 'designer' zinc finger proteins (recognizing a longer sequence of 9-18 base pairs) can itself be optimized by phage display [112,129]. Such strategies assume that triplet finger recognition is not sufficiently modular to use a simple recognition code and 'in context' effects need to be taken into account. Yet if zinc finger domains for each of the 64 possible base triplets can indeed be found with independent binding, a much easier design strategy emerges. With a specific

collection of such recognition fingers developed for a subset of these triplet blocks, acceptable modularity has been found [130], and the aim continues to complete the process for all 64 triplets [131]. For the great majority of targets, it is already very feasible nonetheless to design a specific zinc finger DNA binding protein [131].

Artificial engineering of such binding specificity is a breakthrough which then opens up the possibility of making a variety of possible effector proteins. This too is enabled by general protein modularity, where in a great many cases separate functional domains can be joined with retention of their original activities. Of undoubted evolutionary importance, it is clearly profoundly useful for biotechnology. For productive joining with designer DNA-binding proteins, functional modules with specific tasks include domains for nucleases [132] and methylases [133], and domains for transcriptional activation or repression [134]. Specific tools of these types convey tremendous potential power for a variety of biotechnological and therapeutic applications ♥ *in vivo* [134,136,137]. From within this huge and immensely promising field, let's examine further the application of zinc finger proteins to form zinc finger nucleases.

Firstly we should refer to the discussion of restriction and homing endonucleases in the Cited Notes for Chapter 3 ♣, where the utility of Type IIA restriction enzymes (exemplified by FokI) was noted. These enzymes are characterized by functional domain separation of their DNA recognition and cleavage activities, allowing modular artificial recombination of novel recognition domains with the original protein region which mediates double-strand cleavage. Joining a designed zinc finger domain with the FokI cleavage domain thus creates an

---

♥It should also be noted that in many cases, competing alternative technologies exist. For example, control of gene expression by artificial transcription factors based on zinc finger domains is but one approach of a number of alternatives, including RNAi (Chapter 9) and small molecules such as polyamides [135].

♣ SMS-CitedNotes-Ch3/Section 4; also from same ftp site.

endonuclease with a tailor-made recognition specificity. Such 'zinc finger nucleases' can function in mammalian cells, which has enabled technologies for greatly increasing the frequency of both knock-outs and site-directed alterations ('genomic editing'). Repair of double-stranded breaks introduced into genes of choice will introduce mutations at a high frequency which ablate expression, and this result becomes stabilized as a heritable phenotype. This has been used for the generation of transgenic knock-outs [138,139]. But double-stranded breaks introduced by artificial nucleases can also catalyze recombinational repair with co-introduced homologous sequences bearing mutations of choice, resulting in efficient genome editing [140,141]. Therapeutic applications of this technology have already been tested, by converting lymphoid cells from patients genetically susceptible to HIV into resistant cellular derivatives by knock-out of expression of an HIV co-receptor [142].

It is useful at this point to compare zinc finger nuclease technology with that of homing endonucleases adapted for analogous genomic applications (SMS-CitedNotes-Ch3/Section 4). At present zinc finger domains have more versatility towards targeting a very wide range of target choices than the present capabilities of rationally engineered or evolved homing enzymes. On the other hand, there are real concerns as to the absolute recognition specificities of zinc finger nucleases, an area where homing endonucleases appear to have the advantage [143]. In other words, zinc finger domain recognition may result in a level of background non-specific DNA cleavage which might prove unacceptable for some applications. As always, both technologies themselves are rapidly evolving, and performance improvements in both can be expected in years to come.

To resume the major theme of this section, we should amplify the fact that unlocking of the information embodied in zinc finger protein DNA recognition was greatly accelerated by the concerted application of phage display. It should be noted that this has generally not been via directed evolution by the re-iterative

definition put forward at the beginning of Chapter 4, but rather through single-library randomization and repeated rounds of selection from within this initial set of variants. An important reason for this has been the rational application, from the very beginning of the phage display work, of specific structural information (as depicted in Fig. 4.Nd). This has enabled targeting of regions for mutation to a limited number of sites in the zinc finger α-helices [117,144]), such that a single library of manageable size can realistically encompass all DNA binding diversity. In practice, optimization of zinc finger preferences has required site-directed mutagenesis analyses [121-123] which can themselves be regarded as 'evolutionary' steps of sorts, albeit again significantly guided by rational information. The development of zinc finger DNA binding protein technology is thus a case in point for the effective interplay between empirical and rational strategies, as considered in Chapter 10 of *Searching for Molecular Solutions*.

An arbitrary 18 base pair sequence recognized by a designer zinc finger protein has an excellent chance of being unique in the human genome (occurring on average only once every 4^{18}, or 6.9×10^{11} base pairs). With 49/64 triplets currently usable for zinc finger design [131] (also see Zinc Finger Tools site) and the ability to use discontinuous target sequences [126], few if any endogenous target regions would be entirely refractory for the design of zinc finger proteins which would recognize them (and the remaining triplet gap will probably be closed in the near future). This near universal coverage raises the intriguing question of the diversity of other DNA binding motifs in nature [145] and whether alternative scaffolds are indeed superior over zinc fingers for certain specific sequences [112] (even if by a small but naturally selectable margin). An interesting and instructive exercise for advanced molecular biologists would be to test the fitness changes in a simple bacterium with progressive artificial adaptation and replacement of its endogenous DNA binding proteins with equivalent zinc fingers; ultimately leading to an organism with homogeneity in its DNA-binding

scaffolding, if it was possible ♥. We can also return to our considerations in Chapter 3 of the extent to which antibodies could be generalizable as sequence-specific binding proteins. A judgement could be readily made to the effect that antibodies could not match the modularity of zinc finger domains for extended DNA sequence recognition. On the other hand, although zinc fingers participate in protein-protein interactions [108] it would seem unlikely that they could rival the dazzling protein (and other) recognition abilities of immunoglobulins.

♥ Another variant on the theme of 'molecular lipograms' as raised in Chapter 5 of *Searching for Molecular Solutions*. This would not be simple in some instances where a DNA-binding protein binds to regions within its own conding sequence; see the section describing Lambda O protein in Extras for Chapter 5 (SMS-Extras-Ch5/Section A7; from the same ftp site).

Section 9:   ***Two-hybrid System and Protein-Protein Interaction Assays***

Cited on p. 141 of *Searching for Molecular Solutions*

This section provides some detail on the two-hybrid system and its extensions, and also briefly notes some additional methods of interest for the identification and study of protein-protein binding partners.

The Two-Hybrid System and its Extensions

The two-hybrid system was developed two decades ago in the laboratory of Stanley Fields for identifying novel protein-protein interactions [146,147], and is rendered possible by the modularity of protein domains. The general principles of this system are depicted in Fig. 4.Ne. Although originally derived within yeast host cells, analogous systems have also been adapted to bacterial [148] or mammalian [149,150] counterparts. As in most molecular biological applications, bacterial systems have advantages of larger library sizes and processing speeds, but yeast and especially mammalian systems are more likely to ensure that participating eukaryotic proteins fold correctly and receive appropriate post-translational modifications.

The original and most widely-used application of the two-hybrid system has been searches for binding partners of proteins of interest. The inherent background to this kind of aim is the realization that, as complex self-interactive systems, the operation of biological cells and higher organisms is fundamentally dependent on internal molecular interactions. As such, biological systems cannot be understood unless this network of interactions is defined. And a great deal of this 'interactome' ♥ (although by no means all of it) is based on a web of transient or stable protein-protein recognition events. Taking analysis of the protein

---

♥ The interactome itself is considered further in the file SMS-Extras-Ch9/Section A16; from the same ftp site.

interactome to its logical limits asks for global, high-throughput analyses, and large-scale two-hybrid system-based screening projects have been undertaken, usually in yeast hosts [151-153], but also with mammalian systems [154].
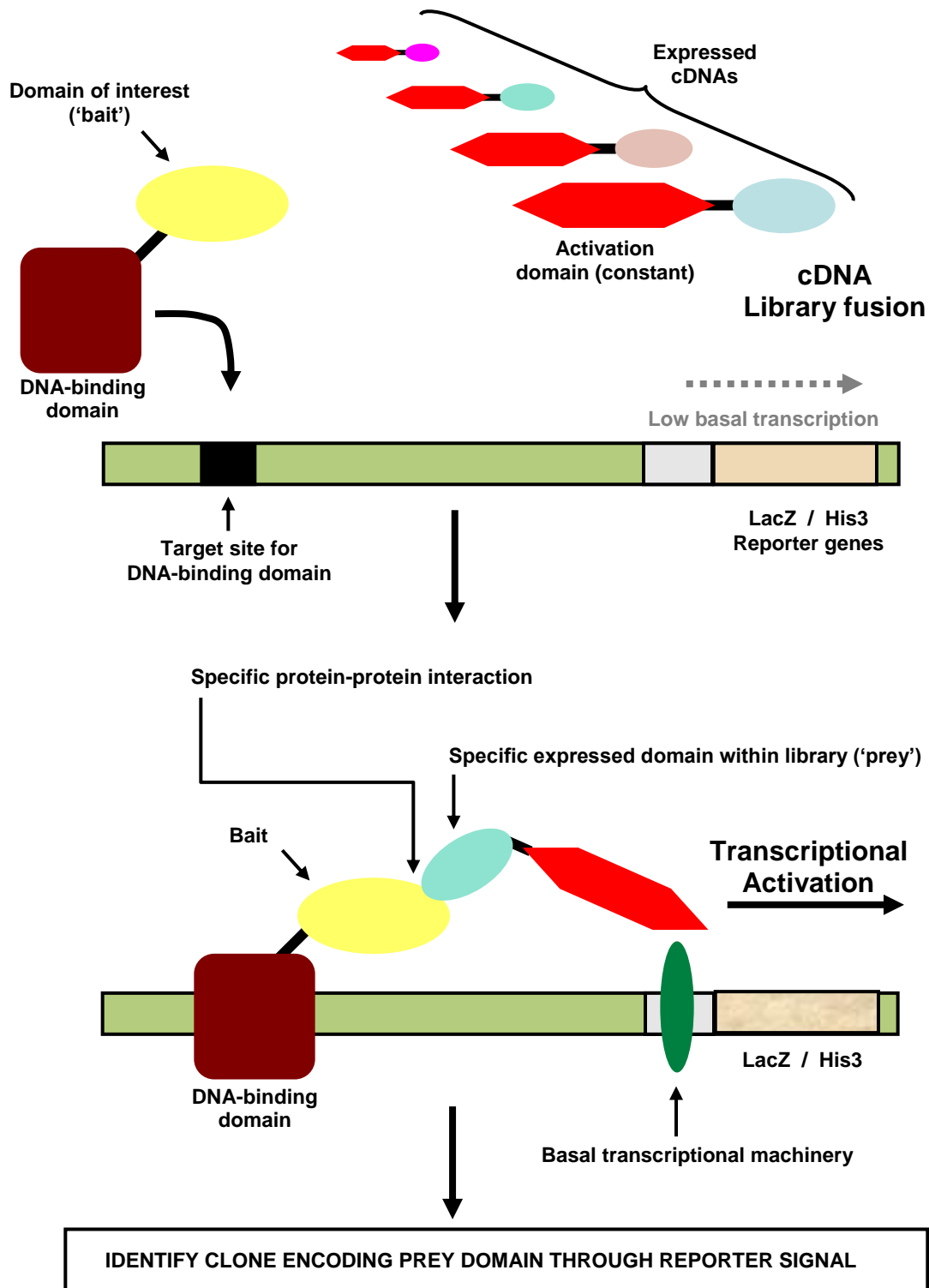
Fig. 4.Ne

Depiction of the two-hybrid system in yeast cells for identifying protein-protein interactions. An initial 'reporter' construct has coding sequence for a marker (*his3*) which allows yeast growth on a selective medium, or a marker (*lacZ*) which allows chromogenic screening based on enzyme activity. These are initially only expressable at very low levels through the basal transcriptional apparatus. Upstream of the reporter gene sequence is a recognition site for a specific DNA-binding protein domain. The latter is fused with a target sequence for which unknown binding partners are sought (hence the common term for such a segment as the 'bait'). Cells harboring the reporter construct are co-transfected with plasmids expressing the DNA-binding protein / bait fusion, and also plasmids expressing a cDNA library expressed as a fusion between variable cDNA segments and a constant strong transcriptional activation domain. (Transfection conditions are designed such that on average, each cell receives a copy of the DNA-binding protein / bait fusion plasmid and one plasmid member of the cDNA library / activation domain fusion). Expression of each component from their respective plasmids ensures that when a specific binding partner for the bait is expressed in the same cell, the activation domain will be focused in close proximity to the basal transcriptional cell protein complex. This then results in transcriptional activation and reporter gene expression. Selection or screening for the reporter then enables identification of the clone expressing the binding partner domain of interest (the 'prey').

As always, fundamental biological information can often rapidly be exploited for applied purposes, but the two-hybrid system has many quite direct applications in biotechnology. A useful example of this can be found in experimental systems designed for selection of intrabodies ♥, or antibodies which retain antigen-binding properties when expressed within cells.

♥ Further details of this topic are provided in the file SMS–Extras-Ch7/Section A11; from the same ftp site.

The basic principle of the two-hybrid system could be summarized with the concepts of specific interactions / modularity / functional bridging, and these ideas can be extended in ways considerably further than the original model. Going in one direction, a cDNA library fused with an activation domain can be used to search for novel domains binding to defined DNA motifs, which has been termed the 'one-hybrid system'. In this case, the specific DNA sequence inserted upstream of the reporter gene is in effect the operational 'bait' [155]. In the other direction, the essential principle lies in the observation that more complex bridging links between a DNA-binding protein and an activation domain can still allow the reporting of the desired information. Alternatively, other functional molecules can potentially replace proteins in forming a functional bridge, provided that the net effect is activation of the reporter gene used in the system. This burgeoning diversity of highly useful assays has often led to the categorization of the original two-hybrid arrangement as a subset of the larger field of 'N-hybrid systems'. Given its interest for some topics within *Searching for Molecular Solutions*, we can spend a little time looking at one of these 'spin-offs', the three-hybrid system, in more detail. A variety of three-hybrid arrangements have been devised, but let's focus on systems which include small-molecule mediators.

Three-Hybrid Systems

The modularity of protein domains allow DNA-binding function to be joined with other functional domains, in particular domains which serve to transactivate transcription. Small bifunctional molecules acting as cross-linkers are well-known, and there is no reason in principle that a known small DNA-binding molecule could not be linked with a small activation molecule, thereby generating a specific low molecular weight transactivator. Tethering a small molecule to a promoter site need not involve direct DNA binding, since a protein can be used as an effective adaptor if it possesses the needed DNA recognition motif and the capacity for binding a low molecular weight ligand. Nuclear hormone receptor-

ligand fusion protein systems (such as the glucocorticoid receptor binding of dexamethasone fused with the yeast Gal4 DNA binding protein domain) have been used in this regard [156] (Fig. 4.Nf). This kind of system allows screening for small organic molecules which can activate transcription.
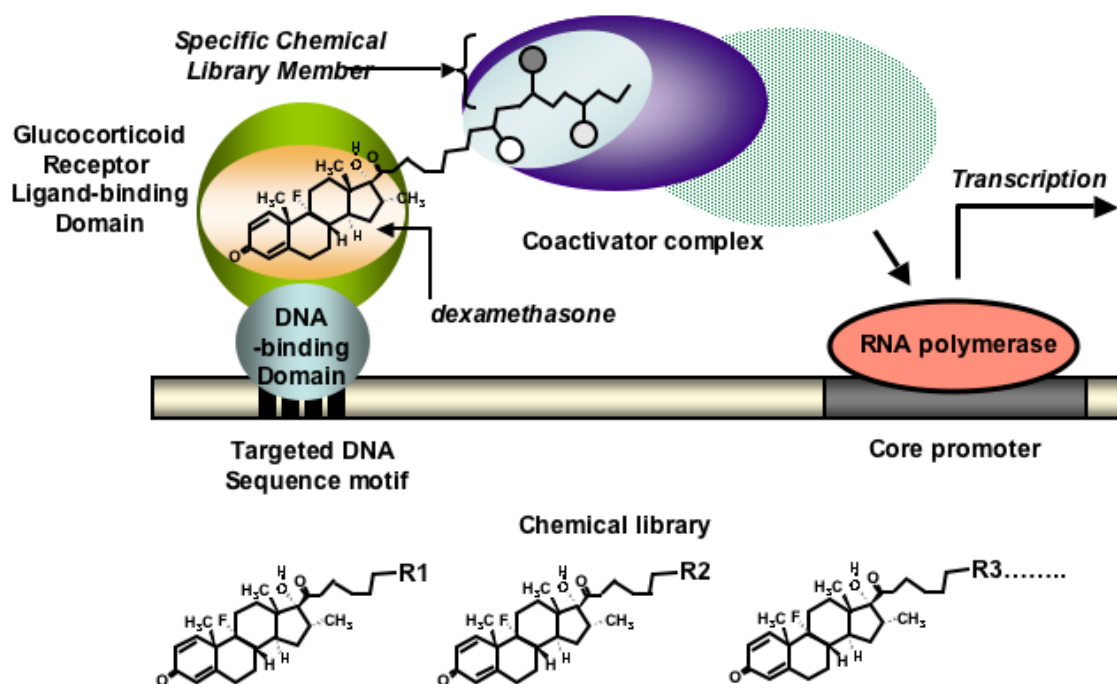
Fig. 4.Nf

Three-hybrid screening for a chemical group capable of binding transcriptional activating complexes. Here a specific chemical group binds to a pocket of a component of an coactivational complex, with resulting activation of RNA polymerase-mediated transcription from a chosen promoter in *cis*. Tethering of a bifunctional small activation molecule can be effected through a moiety which binds to an adjacent transcription factor, in this example dexamethasone associating with the ligand-binding domain of the glucocorticoid receptor. This domain is fused with a DNA-binding domain (such as Gal4 [157,158]) directed towards an introduced DNA site proximal to the promoter for the reporter construct used for the screening assay.

An interesting adaptation of the three-hybrid procedure termed chemical complementation [159,160] enables a reporter read-out for potentially a wide range of enzymes breaking or forming specific covalent bonds (depicted in Fig. 4.Ng).
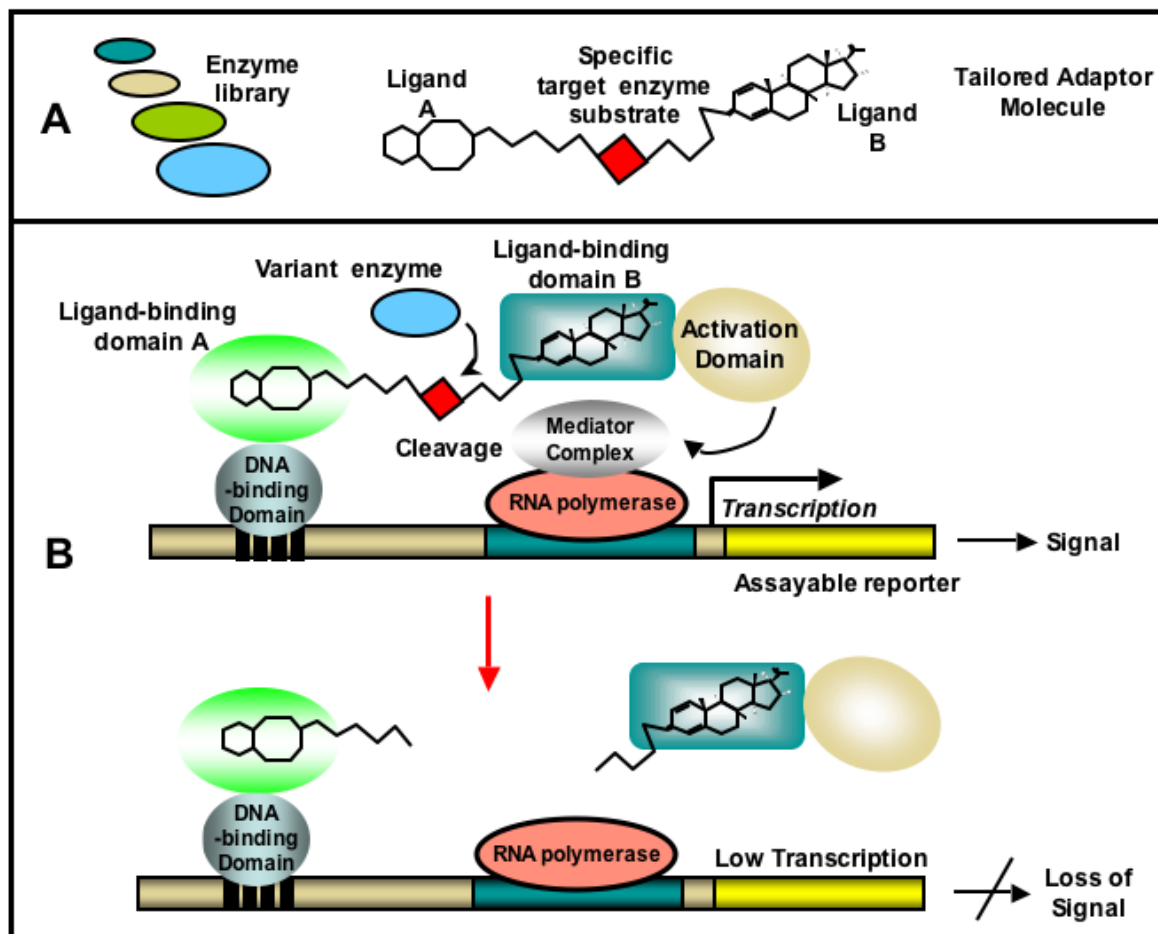
Fig.4.Ng

Depiction of chemical complementation screening for enzyme activity, based on the three-hybrid system (Fig. 4.Nf). A synthetic adaptor molecule is used with specific structures (Ligands A and B) at each end joined by a linker in which a desired enzyme substrate is inserted. Each of the ligand A and B moieties in the adaptor molecule bind to specific protein binding domains, which are fused with a DNA-binding (A) or activation

domain (B). The adaptor molecule allows the formation of a bridge between the DNA-binding and activation functions, and transcriptional activation. Cleavage at the enzyme substrate site by a specific member of an enzyme library destroys the bridge and knocks down expression of the screenable reporter gene. The process can also be performed in the reverse direction to search for enzymes catalyzing specific bond formation which will activate expression from a starting low level.

Despite its power and versatility, alternative technologies to N-hybrid systems certainly exist. One of the simplest protein-ligand analysis approaches in principle is often termed the 'pull-down' approach, where a protein of interest taken directly from its *in vivo* source is precipitated with a specific binding agent (usually an antibody). By so doing, other proteins or ligands associated *in situ* with the direct target protein are likewise 'pulled-down', and are then subject to characterization. Often this is via mass spectrometry ♥, a field whose rapid advances in sensitivity have made direct pull-downs often a highly competitive option [161].

Phage display (considered in some detail in *Searching for Molecular Solutions* and in Sections 7A, 7B, 8A and 8B above) can also be used for cloning binding partners, if a surface-displayed cDNA library is probed with a ligand of interest [162]. *In vitro* display technologies can also be put to this kind of use. As noted in Chapter 4, *in vitro* mRNA display employs a novel linkage between mRNA and polypeptide translated from the same mRNA, and affords a powerful approach for directed evolution [163]. It has also been used for searching expressed genomes for binding proteins towards a desired protein or ligand [164-166]. Since the 'identity tag' for translated proteins in mRNA display is relatively small (an mRNA molecule itself), this display technology is well-suited for general searches for

---

♥ A little more will be said about mass spectrometry and its biological applications in the file SMS–CitedNotes-Ch9/Section 33; from the same ftp site.

protein interaction partners [164,167]. Among such applications are the identification of partner proteins for DNA-binding proteins which function as heteromers [167,168].

To conclude this section, though, let's look at some other specific protein-protein interaction assays. The first, based on phage display, is not widely used at present, but worthy of attention given the interest of this technological area in *Searching for Molecular Solutions*.

Selectively Infectious Phage and Protein-Ligand Interactions

We have already seen that the gp3 protein of filamentous phage, which mediates infectivity, has a highly modular domain structure (Fig. 4.10 of *Searching for Molecular Solutions*). In a remarkable confirmation of the modular nature of gp3, the N1 domain function can be separated from the remainder of the protein, provided a non-covalent interaction of sufficient affinity serves to bridge the N1 and N2 domains together. This has been exploited in the development of protein-protein or protein-ligand interaction systems [169,170], usually termed the selectively infectious phage (or SIP) system [171-173]. The principle of this system is depicted in Fig. 4.Nh.
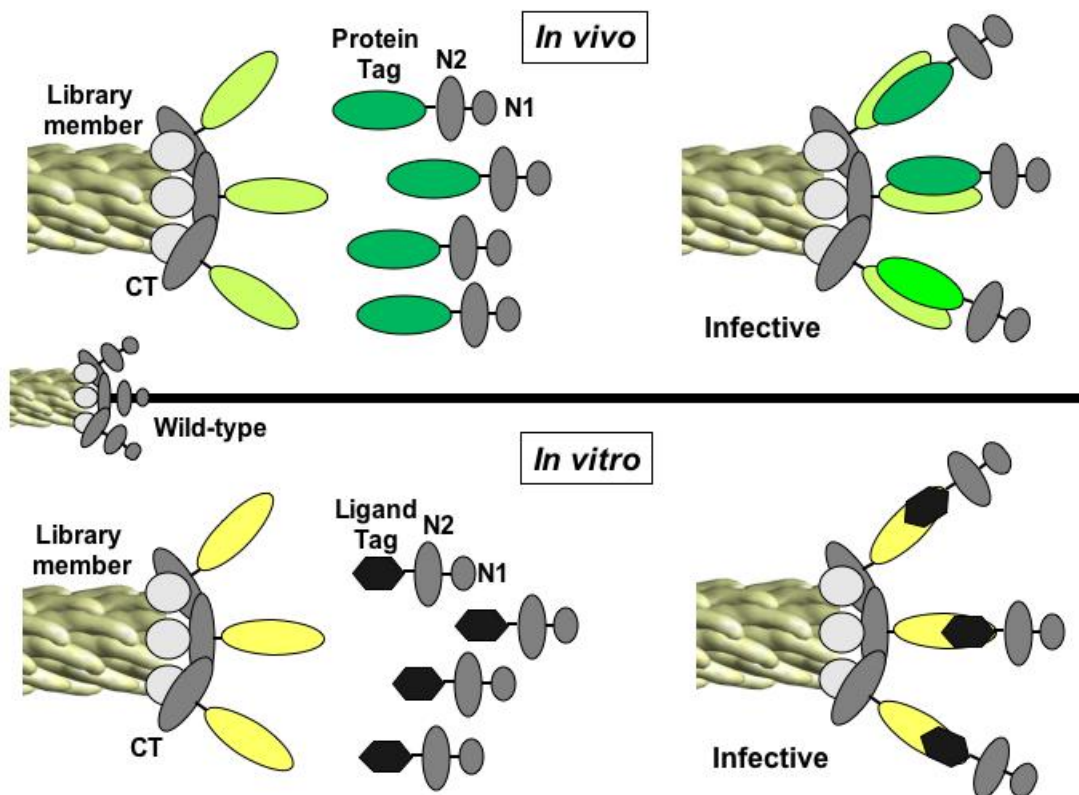
Fig. 4.Nh

Principles of selectively infectious phage (SIP) systems. In the top panel for *in vivo* systems, a library of protein domains (light green) fused to gp3 C-terminal (CT) domains is expressed and assembled *in vivo* into (non-infectious) phage. (See Fig. 4.10 of *Searching for Molecular Solutions* [and the corresponding color version of this figure from the same ftp site; SMS-ColorFigures.pdf] for more details on filamentous phage gp3 display). Within the same cell, interaction of a specific library member with an expressed protein fusion of an interacting 'bait' domain (dark green) linked with N1 and N2 domains allows infectivity to be restored. The specific library member(s) expressing a protein domain interacting with the bait are thus identifiable. *In vitro*: A CT-gp3 fusion library expressed as above can be extruded from host cells but is non-infectious. Infectivity can be activated *in vitro* by binding of an N1-N2-ligand preparation as shown. Note that for the *in vitro* arrangement, the ligand can be in principle any chemical structure, since the N1-N2-fusion itself can be prepared *in vitro*.

Initial enthusiasm for the SIP system was tempered by realization of its limitations. In the case of the *in vivo* version (Fig. 4.Nh) recombination can sometimes be a confounding factor, and as with any filamentous phage display system, biological constraints on certain displayed polypeptides exist. (See Section 7A above). The size itself of the complex between displayed polypeptide and binding ligand has been also defined as an important factor limiting the applicability of this technique [174]. There is also strong competition from the alternative technologies noted in this section. Still, it remains a viable contender assay for specific applications as long as its strengths and limitations are understood [172].

Let us now look at one final assay which appears to have wide appeal for future applications.

The Protein Complementation Assay

Techniques which enable direct visualization of specific binding interactions in living cells are of great future promise [175]. One such method of interest is called the Protein Complementation Assay (PCA) [176], whose name is a little deceptive because the PCA principle is distinct from other protein complementation effects which have been long known. Most of the latter cases involve the reconstitution of a protein functional property (usually an enzymatic activity) by two protein fragments which are already pre-folded into defined structures, as illustrated in the top panel of Fig. 4.Ni.

One of the most widely used 'conventional' protein complementation systems is called the α-complementation of β-galactosidase, where a defective mutant of

the latter enzyme exists as a non-functional dimer, unlike the tetrameric wild-type enzyme. If a relatively small N-terminal fragment of the original enzyme termed the 'α-peptide' is co-expressed with the defective dimeric protein, complementation occurs such that the assembly of active β-galactosidase tetramers ensues [177]. (This effect can also be engineered to occur via surface phage display [178]). The phenomenon has been widely exploited by using coding sequence for the α-peptide as a convenient site for insertional cloning in plasmids, allowing loss of complementation-mediated enzyme activity to serve as a useful screening guide for the presence of cloned inserts. β-galactosidase complementation effects have also been used as tools for monitoring protein-protein interactions in eukaryotes [179].

In contrast, the PCA principle involves the active folding of otherwise disordered reporter protein segments, whose spatial juxtaposition allows a productive folding pathway to commence, with resulting reconstitution of protein activity. The enforced positioning of the protein fragments is engendered by means of linkage with mutually interactive protein domains, as depicted in the bottom panel of Fig. 4.Ni. The PCA has the advantage of negligible background activity, since without initiation of folding the reporter signal is effectively zero, unlike domain-based protein complementation methods, which usually have detectable backgrounds [180].
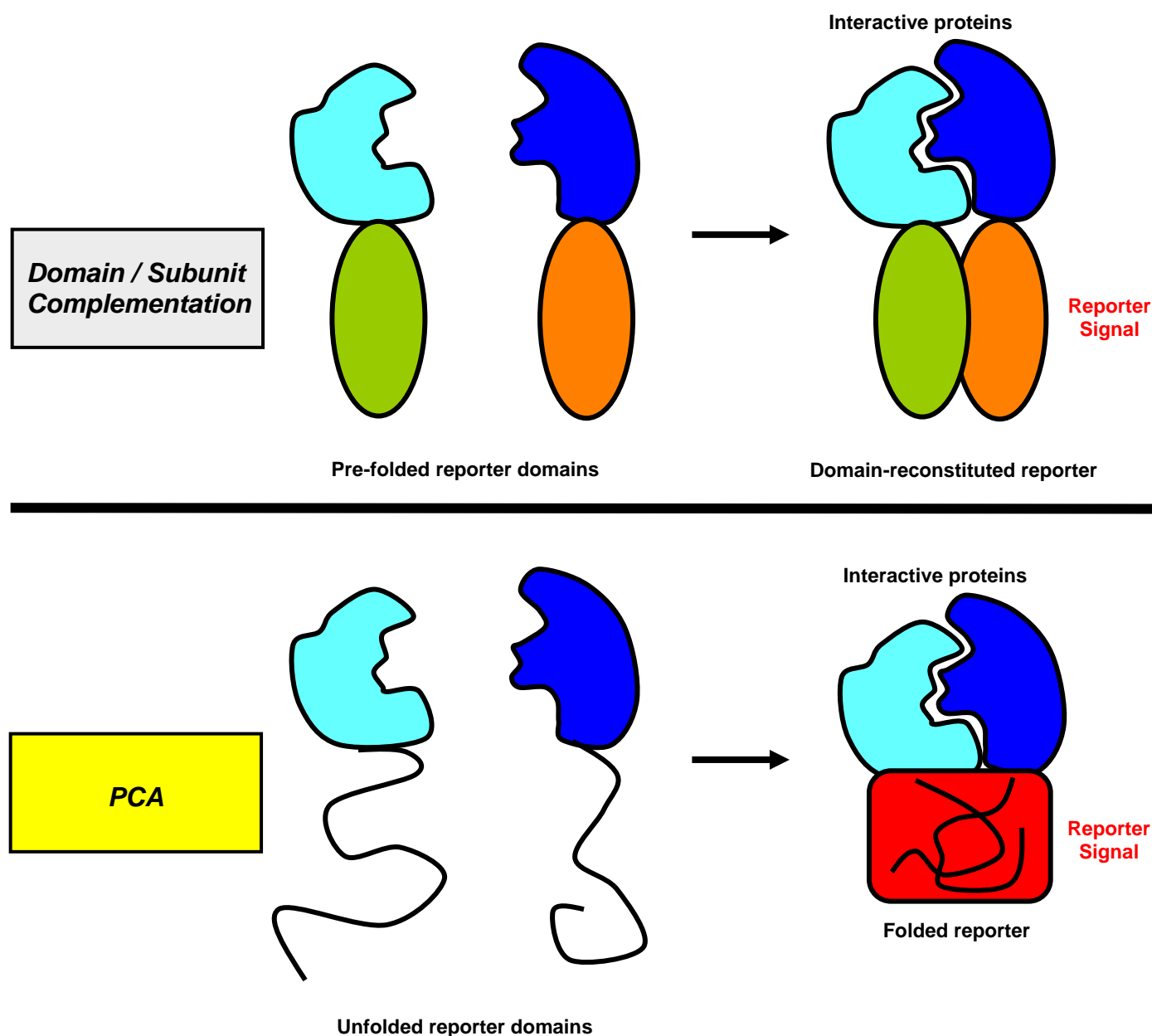
Fig. 4. Ni

Protein complementation assays. In the top panel (Domain / Subunit Complementation), pre-folded protein domains associate and reconstitute a measurable function, such as a specific catalysis. In contrast, (bottom panel) the Protein fragment Complementation Assay (PCA) uses interactive protein domains to drive the folding pathway of two separate (initially unfolded) fragments of a reporter protein. Reconstitution of the reporter than acts as a gauge of the levels of the protein-protein association of interest, with an extremely low background.

In certain cases, PCA-mediated fragment folding is irreversible, but reversible systems have also been derived, and these are invaluable for the above-mentioned *in vivo* monitoring of protein-protein interactions [180,181]. The assembly of the reporter fragments is dependent on the interaction of the two protein domains of interest to which they are linked, and relative intracellular concentrations of the former pair therefore determine whether the reporter activity will be reconstituted. If protein participants within interactive and dynamic intracellular pathways are engineered as components of a PCA system, a powerful and sensitive means for real-time monitoring of the results of artificial systems perturbations is enabled [180]. This principle has been ingeniously exploited for analyzing the 'off-target' effects of drugs [181], or general study of the effects of drugs on cellular protein interaction networks [182]. Both of these applications are potentially of great value to the pharmaceutical industry.

Section 10:   ***CIS Display and Related Areas***

Cited on p. 141 of *Searching for Molecular Solutions*

How could a direct 'display' link be established between a DNA molecule and its encoded protein product, translated via an mRNA molecule? It was noted earlier in these Chapter 4 Cited Notes (Section 7B above) that this can be done intracellularly by fusing an expressed library with a DNA-binding protein domain, where the latter is designed to target a site in the same vector encoding the library itself. In principle, the role of the bacterial cellular 'compartment' can be mimicked by *in vitro* microcompartmentalization (as described in Chapter 4 of *Searching for Molecular Solutions*). In both cases, the genotype-phenotype linkage is mediated through non-covalent interactions (via specific recognition of DNA sequences though the special structural features of the appropriate protein DNA-binding domains). But other *in vitro* systems have been developed for generating linkages between expressed polypeptides and their encoding DNA sequences.

These ingenious systems depend on certain features of both general and specific prokaryotic biology. The most fundamental (and useful) property in this regard is the *coupling* of bacterial transcription and translation, unlike eukaryotic systems. What this means is simply that bacterial translation of an mRNA molecule on ribosomes can begin even before the transcription process is finished ♥. In some bacterial replication systems, this coupling has become a significant factor in DNA replication. For the RI plasmid, the RepA protein recognizes a sequence at the origin of replication (*ori*) via transient interaction with another upstream sequence (CIS; Fig. 4.Nj).

♥ In eukaryotic cells mRNA molecules are extensively processed in the nucleus and transported to the cytoplasm for translation. The coupling of the processes in bacteria results in very efficient overall expression and allows very rapid cell growth under optimal conditions.
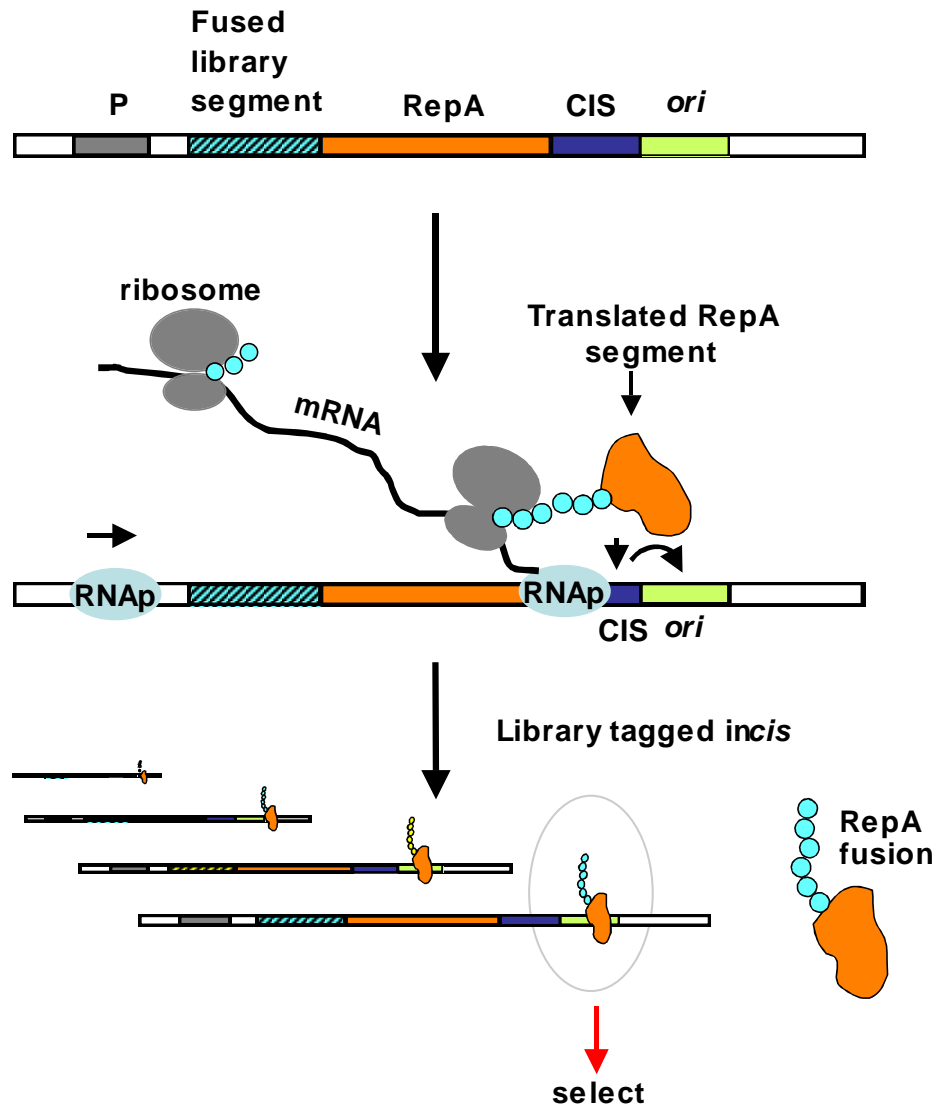
Fig. 4.Nj

Principle of CIS non-covalent DNA display *in vitro* [183]. When a polypeptide library fused with the RepA coding sequence undergoes coupled transcription / translation *in vitro*, the RNA polymerase pauses upon encountering a specific 'CIS' sequence. Following completion of its translation, the RepA protein segment is directed to the CIS sequence and then origin of replication (*ori*) sequence in *cis* (on the same DNA molecule). In this manner an expressed library is linked to a DNA tag carrying its own coding sequence. Abbreviations: P, promoter; RNAp, RNA polymerase.

Non-covalent binding of RepA to the *ori* sequence provides a direct link between sequences fused to RepA itself and their encoded information at the DNA level [183]. The key feature of this system which enables its application with *in vitro* transcription-translation systems is the faithful binding of the RepA protein in *cis* (to the same DNA encoding it and its linked specific library member). It is entirely this property which maintains the all-important genotype-phenotype linkage, and allows a large and diverse library to be expressed *in vitro* with preservation of the encoding informational linkage when functional selections are applied. Without this *cis*-effect, *trans*-binding in a dispersed *in vitro* system would destroy any correspondence between genotype and phenotype ♥.

Another *in vitro* DNA display system actually generates a covalent link between the coding DNA and the expressed library [184]. This operates by means of the phage P2A endonuclease, which initiates phage DNA replication by making a nick at a site within its own coding sequence, and then covalently attaches itself. The arrangement then is analogous to the CIS system, except for the site and covalent mechanism of association of the library fusion protein with the encoding DNA. This display process has been used for antibody libraries [184] and achievable library sizes for both the P2A exonuclease and CIS systems are theoretically comparable to mRNA display.

It should be noted here that the term 'DNA display' has also been applied to another process involving the use of DNA sequences as 'tags' for chemical syntheses [185], which is in considered in Chapter 8 of *Searching for Molecular Solutions*. Although not to be confused with the above biological processes, the chemical-tag display has clear parallels with the gamut of biological display systems we have covered in this chapter to date, and is therefore not a misnomer from this point of view.

♥ In fact, the above-noted *in vitro* compartmentalization is a generalizable solution to this very problem, where preservation of linkage between trans-acting effects and their encoding nucleic acids is desired.

Section 11:   *Enzyme Display on Phage*

Cited on p. 141 of *Searching for Molecular Solutions*

Before phage display can become an attractive technology for directed evolution of an enzyme, obviously the enzyme of interest has to be compatible with the display process *per se,* without sacrificing initial catalytic activity. (We have already seen above (Section 7A above) that several options exist for improving filamentous phage display if a protein should prove recalcitrant, and lytic phage further extend the range of displayable polypeptides). In fact, without needing to resort to heroic means, many enzymes have been successfully displayed through phage systems. A quick sampling of different enzymes so used includes glycosyltransferases [186], DNA polymerase fragments [37,187], lipase [188], bacterial ribonuclease [189], amylases [190], and β-lactamases [191]. The principles of enzyme directed evolution discussed in Chapters 4 and 5 of *Searching for Molecular Solutions* still apply; the display technology powerfully enters the picture at the selection level.

Let's look briefly at how selections for changes in enzyme behavior can be applied in this context. If directed evolution is aimed at altering an enzyme's substrate binding, then a relatively simple selection based on binding of mutant enzymes from within a display library to a transition-state analog compound can be instituted (Fig. 4.Nk-A). This kind of approach has been applied towards the selection of catalytic antibodies [192]. An alternative (but conceptually related) strategy uses 'suicide substrates' which are activated by catalysis and then bind irreversibly to an enzyme's active site [192]. This principle has been used to select high-activity variants from displayed β-lactamase libraries [193,194].
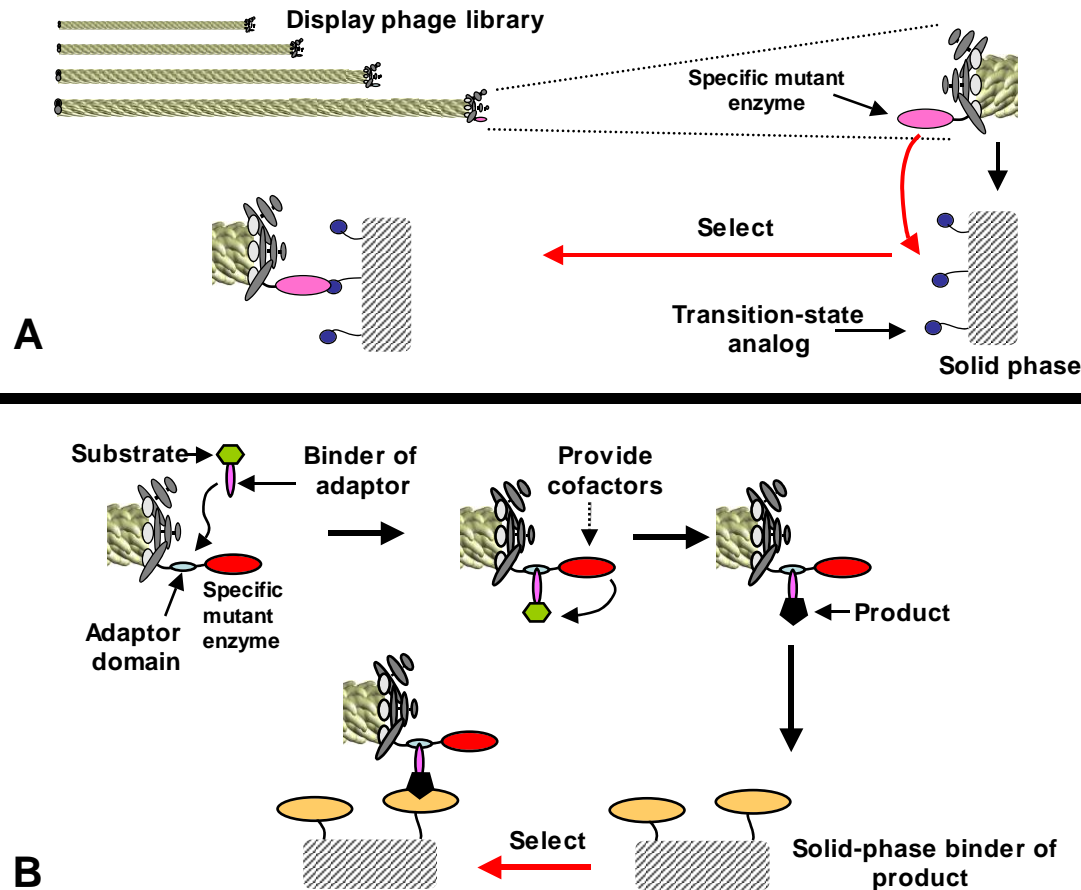
Fig. 4.Nk

Some strategies for directed evolution of enzymes by phage display with gp3 systems of filamentous phage. **A,** Monovalent display of an enzyme library fused to the C-terminal region of gp3 [192,195]. Mutants with affinity for the appropriate transition-state analog compound can be selected on solid-phase supports as indicated. (Bound phage can be eluted by excess free analog or by engineering a specific protease site in a linker region between the gp3 domain and the displayed enzyme). **B,** A more complex system for directly selecting for catalysis. Here an adaptor domain is positioned between the gp3 C-terminus and the enzyme, where the adaptor bears a peptide recognition signal for another protein ('binder of adaptor') which is fused or complexed with the desired novel enzyme substrate. When all cofactors (metal ions, etc.) are provided for catalysis, the closely proximal enzyme can catalyze product formation as shown, which again is potentially selectable on a solid-phase support.

Direct selection for catalysis in a display system is more of a challenge. To exploit the genotype-phenotype linkage offered by display systems, somehow a signal from an active enzyme catalyzing a desired substrate also has to linked to the same phage surface. This might seem daunting, but with a little applied ingenuity some important steps in the right direction have been taken. The basic concept involves the engineering of some form of adaptor system whereby a desired substrate can be coupled with the phage-displayed enzyme. One way to do this is to co-display (with the enzyme library) a peptide ligand sequence for a specific binding protein, such that the display library can be 'activated' by binding of a fusion or conjugate of the specific peptide-binding domain with a substrate of interest (Fig. 4.Nk-B). When this is arranged, and the enzyme supplied with all necessary cofactors, the product is generated in spatial proximity to the display enzyme itself, and the specific phage as a whole can be selected with a solid-phase binder of the product (Fig. 4.Nk-B).

An example of this system has used calmodulin and calmodulin-binding peptide as the adaptor and binder of adaptor respectively [196]. This has been applied towards the evolution of biotin ligase, where the substrate (a specific short peptide sequence recognized by this enzyme) is coupled to the calmodulin-binding peptide, which thus serves to enable *in vitro* 'charging' of displayed phage enzyme with a biotinylatable substrate. Following successful catalysis by the displayed biotin ligase, the biotinylated product is then selectable with streptavidin (a protein with high binding affinity for biotin) [197]. Though elegant in design, such systems have not been notably successful in the selection of altered enzymes from libraries [192,197]. In this context we should recall the availability of the competing alternative *in vitro* technologies of mRNA and ribosome display, which are considered in Chapter 4 of *Searching for Molecular Solutions*.

The flip side of enzyme display is substrate display [198]. Where an enzyme recognizes a specific peptide sequence as its substrate, this sequence (or a library of variants of it) can be displayed, and sequence changes which impinge on enzyme activity can be identified if an appropriate selection can be devised for the products of catalysis. This is relatively easy for selection of proteases, by displaying a library where the C- and N- termini of gp3 are separated by a randomized peptide segment (see Fig. 4.10 of *Searching for Molecular Solutions* for a structural depiction of gp3 domains). Also, by various strategies, the N-terminal domain in the same system can be engineered to bind a specific solid-phase support ♥. When the library is prepared and immobilized in this manner, only phage bearing target sequences for the protease of interest will be released from the support upon digestion with the specific protease itself. Now, if all displayed gp3 proteins on the phage surface were so cleaved, then released phage would lose infectivity, but this can be circumvented by ensuring that only one of the five gp3 copies bears the target modification. (In other words, *monovalent* display). After cleavage, infectious phage can be thus recovered and an another cycle begun. This process allow characterization of the specific peptide sequences within the random library which serve as substrates for the protease of interest.

A more recent improvement in the technique avoids the solid phase immobilization and relies on protease removal of an additional (protease-resistant) displayed peptide which interferes with phage infectivity [199]. Peptide substrate display in general has been applied to the determination of optimal recognition specificities for both natural proteases [200-202] and their engineered counterparts [203]. Information obtained in this manner can be used for finding novel protease targets, with a variety of potential applications [204,205].

♥For example, the N-terminal domain can be modified to display a specific peptide sequence recognized by an antibody, where the antibody is immobilized on a solid gel or resin.

We should also consider selections for other desirable enzymatic properties which use phage display, and stability comes to the forefront in this regard. In [Section 8A](#) above, we made brief note of the Proside technique for stability selection, which is based on the observation that protease resistance and thermal stability are very frequently correlated. As with the above substrate phage, this procedure exploits the domain modularity of filamentous phage gp3, where it is known that insertions can be made between the N1, N2, and CT domains without disrupting phage infectivity. If a gp3 fusion protein with the arrangement: (N1 – N2 – (Protein of interest) – CT) is displayed on a filamentous phage, the presence of the protein insert in effect provides a link between protease resistance (and thereby indirectly a link with thermal stability) and the ability of phage to infect host cells [95,206]. This is so because cleavage of the intervening foreign protein sequence by a protease cuts off the essential N-terminal domains and ablates infectivity ♥. Conversely, resistance of the inserted protein to proteolytic cleavage allows normal infection to proceed, and thereby provides a powerful selection for protease resistance. A non-trivial issue which enables the success of this procedure is the great stability of filamentous phage particles, such that treatment with moderate levels of proteases will not significantly affect phage infectivity but can discriminate between susceptible / resistant inserts into gp3. Application of Proside selection has yielded numerous candidate proteins which show enhancement of thermal stability [207,208], and this approach has been claimed to be superior to computational methods [209]. Nevertheless, we should remember in this context the caveat that exceptions to

---

♥Note that the Proside technique is in essence a negative selection exerted against unwanted protease-sensitive phage, whereas the above substrate-phage technique positively selects for phage with the desired protease sensitivity. So in a sense they are mirror-image approaches, and a key difference again is at the level of display valency for gp3. As noted above, for the substrate phage technique, monovalent display is essential, but for Proside a logical requirement is that *all* displayed gp3 copies should be modified, and equally susceptible or resistant to the protease used as the discriminator. (Were this not the case, cleavage of a specific gp3 molecule bearing an unfolded segment on a Proside phage would not ablate activity, through remaining wild-type gp3 on the same phage particle).

the protease resistance / thermal stability correlation can be found [189,210]. Also, for undefined reasons sometimes enzyme insertions between the CT and N2 domains themselves are incompatible with phage maturation (as with prolyl isomerase) [211].

References

1. Saiki, R. K. et al. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-4 (1985).

2. Saiki, R. K. et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-91 (1988).

3. Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I. & Khorana, H. G. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J Mol Biol* **56**, 341-61 (1971).

4. Mullis, K. B. *Dancing Naked in the Mind Field* (Vintage, 2000).

5. Bartlett, J. M. & Stirling, D. (eds.) *PCR Protocols (Methods in Molecular Biology)* (Humana Press, 2003).

6. Hahn, S. et al. Current applications of single-cell PCR. *Cell Mol Life Sci* **57**, 96-105 (2000).

7. Morling, N. PCR in forensic genetics. *Biochem Soc Trans* **37**, 438-40 (2009).

8. Hohoff, C. & Brinkmann, B. Human identity testing with PCR-based systems. *Mol Biotechnol* **13**, 123-36 (1999).

9. Gibson, N. J. The use of real-time PCR methods in DNA sequence variation analysis. *Clin Chim Acta* **363**, 32-47 (2006).

10. Watzinger, F., Ebner, K. & Lion, T. Detection and monitoring of virus infections by real-time PCR. *Mol Aspects Med* **27**, 254-98 (2006).

11. Gunson, R. N., Bennett, S., Maclean, A. & Carman, W. F. Using multiplex real time PCR in order to streamline a routine diagnostic service. *J Clin Virol* **43**, 372-5 (2008).

12. Day, P. J. Miniaturized PCR systems for cancer diagnosis. *Biochem Soc Trans* **37**, 424-6 (2009).

13. Sontakke, S., Cadenas, M. B., Maggi, R. G., Diniz, P. P. & Breitschwerdt, E. B. Use of broad range16S rDNA PCR in clinical microbiology. *J Microbiol Methods* **76**, 217-25 (2009).

14. Li, J. & Makrigiorgos, G. M. COLD-PCR: a new platform for highly improved mutation detection in cancer and genetic testing. *Biochem Soc Trans* **37**, 427-32 (2009).

15. Milbury, C. A., Li, J. & Makrigiorgos, G. M. PCR-based methods for the enrichment of minority alleles and mutations. *Clin Chem* **55**, 632-40 (2009).

16. Garces, J. A. & Gavin, R. H. Using an inverse PCR strategy to clone large, contiguous genomic DNA fragments. *Methods Mol Biol* **161**, 3-8 (2001).

17. Pfeifer, G. P. & Tommasi, S. In vivo footprinting using UV light and ligation-mediated PCR. *Methods Mol Biol* **130**, 13-27 (2000).

18.   Chu, C. C., Paul, W. E. & Max, E. E. Quantitation of immunoglobulin mu-gamma 1 heavy chain switch region recombination by a digestion-circularization polymerase chain reaction method. *Proc Natl Acad Sci U S A* **89**, 6978-82 (1992).

19.   Joyce, C. Quantitative RT-PCR. A review of current methodologies. *Methods Mol Biol* **193**, 83-92 (2002).

20.   Bustin, S. A., Benes, V., Nolan, T. & Pfaffl, M. W. Quantitative real-time RT-PCR--a perspective. *J Mol Endocrinol* **34**, 597-601 (2005).

21.   Mitra, R. D. & Church, G. M. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res* **27**, e34 (1999).

22.   Chetverin, A. B. & Chetverina, H. V. Molecular colony technique: a new tool for biomedical research and clinical practice. *Prog Nucleic Acid Res Mol Biol* **82**, 219-55 (2008).

23.   Dalton, R. Neanderthal genome to be unveiled. *Nature* **457**, 645 (2009).

24.   Rusk, N. & Kiermer, V. Primer: Sequencing--the next generation. *Nat Methods* **5**, 15 (2008).

25.   Strausberg, R. L., Levy, S. & Rogers, Y. H. Emerging DNA sequencing technologies for human genomic medicine. *Drug Discov Today* **13**, 569-77 (2008).

26.   Braslavsky, I., Hebert, B., Kartalov, E. & Quake, S. R. Sequence information can be obtained from single DNA molecules. *Proc Natl Acad Sci U S A* **100**, 3960-4 (2003).

27.   Harris, T. D. et al. Single-molecule DNA sequencing of a viral genome. *Science* **320**, 106-9 (2008).

28.   Pushkarev, D., Neff, N. F. & Quake, S. R. Single-molecule sequencing of an individual human genome. *Nat Biotechnol* **27**, 847-52 (2009).

29.   Wall, J. D. & Kim, S. K. Inconsistencies in Neanderthal genomic DNA sequences. *PLoS Genet* **3**, 1862-6 (2007).

30.   Roy, S. W. & Irimia, M. When good transcripts go bad: artifactual RT-PCR 'splicing' and genome analysis. *Bioessays* **30**, 601-5 (2008).

31.   Barnes, W. M. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc Natl Acad Sci U S A* **91**, 2216-20 (1994).

32.   Pavlov, A. R., Pavlova, N. V., Kozyavkin, S. A. & Slesarev, A. I. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol* **22**, 253-60 (2004).

33.   Gill, P. & Ghaemi, A. Nucleic acid isothermal amplification technologies: a review. *Nucleosides Nucleotides Nucleic Acids* **27**, 224-43 (2008).

34.   Adler, M., Wacker, R. & Niemeyer, C. M. Sensitivity by combination: immuno-PCR and related technologies. *Analyst* **133**, 702-18 (2008).

35.   Saborio, G. P., Permanne, B. & Soto, C. Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* **411**, 810-3 (2001).

36.   Soto, C., Saborio, G. P. & Anderes, L. Cyclic amplification of protein misfolding: application to prion-related disorders and beyond. *Trends Neurosci* **25**, 390-4 (2002).

37.   Jestin, J. L., Volioti, G. & Winter, G. Improving the display of proteins on filamentous phage. *Res Microbiol* **152**, 187-91 (2001).

38.   Thammawong, P., Kasinrerk, W., Turner, R. J. & Tayapiwatana, C. Twin-arginine signal peptide attributes effective display of CD147 to filamentous phage. *Appl Microbiol Biotechnol* **69**, 697-703 (2006).

39.   Steiner, D., Forrer, P., Stumpp, M. T. & Pluckthun, A. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nat Biotechnol* **24**, 823-31 (2006).

40.   Peters, E. A., Schatz, P. J., Johnson, S. S. & Dower, W. J. Membrane insertion defects caused by positive charges in the early mature region of protein pIII of filamentous phage fd can be corrected by prlA suppressors. *J Bacteriol* **176**, 4296-305 (1994).

41.   Sidhu, S. S., Weiss, G. A. & Wells, J. A. High copy display of large proteins on phage for functional selections. *J Mol Biol* **296**, 487-95 (2000).

42.   Loset, G. A., Lunde, E., Bogen, B., Brekke, O. H. & Sandlie, I. Functional phage display of two murine {alpha}/{beta} T-cell receptors is strongly dependent on fusion format, mode and periplasmic folding assistance. *Protein Eng Des Sel* **20**, 461-72 (2007).

43.   Maruyama, I. N., Maruyama, H. I. & Brenner, S. Lambda foo: a lambda phage vector for the expression of foreign proteins. *Proc Natl Acad Sci U S A* **91**, 8273-7 (1994).

44.   Dunn, I. S. Assembly of functional bacteriophage lambda virions incorporating C-terminal peptide or protein fusions with the major tail protein. *J Mol Biol* **248**, 497-506 (1995).

45.   Sternberg, N. & Hoess, R. H. Display of peptides and proteins on the surface of bacteriophage lambda. *Proc Natl Acad Sci U S A* **92**, 1609-13 (1995).

46.   Zanghi, C. N., Sapinoro, R., Bradel-Tretheway, B. & Dewhurst, S. A tractable method for simultaneous modifications to the head and tail of bacteriophage lambda and its application to enhancing phage-mediated gene delivery. *Nucleic Acids Res* **35**, e59 (2007).

47.   Kozlovska, T. M. et al. RNA phage Q beta coat protein as a carrier for foreign epitopes. *Intervirology* **39**, 9-15 (1996).

48.   Danner, S. & Belasco, J. G. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc Natl Acad Sci U S A* **98**, 12954-9 (2001).

49.   Krumpe, L. R. et al. T7 lytic phage-displayed peptide libraries exhibit less sequence bias than M13 filamentous phage-displayed peptide libraries. *Proteomics* **6**, 4210-22 (2006).

50. Yang, F. et al. Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nat Struct Biol* **7**, 230-7 (2000).

51. Ren, Z. J. et al. Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4. *Protein Sci* **5**, 1833-43 (1996).

52. Ren, Z. & Black, L. W. Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. *Gene* **215**, 439-44 (1998).

53. Li, Q., Shivachandra, S. B., Zhang, Z. & Rao, V. B. Assembly of the small outer capsid protein, Soc, on bacteriophage T4: a novel system for high density display of multiple large anthrax toxins and foreign proteins on phage capsid. *J Mol Biol* **370**, 1006-19 (2007).

54. Hufton, S. E. et al. Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. *J Immunol Methods* **231**, 39-51 (1999).

55. Santi, E. et al. Bacteriophage lambda display of complex cDNA libraries: a new approach to functional genomics. *J Mol Biol* **296**, 497-508 (2000).

56. Rhyner, C., Kodzius, R. & Crameri, R. Direct selection of cDNAs from filamentous phage surface display libraries: potential and limitations. *Curr Pharm Biotechnol* **3**, 13-21 (2002).

57. Faix, P. H. et al. Phage display of cDNA libraries: enrichment of cDNA expression using open reading frame selection. *Biotechniques* **36**, 1018-22, 1024, 1026-9 (2004).

58. Sche, P. P., McKenzie, K. M., White, J. D. & Austin, D. J. Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem Biol* **6**, 707-16 (1999).

59. Allen, J. B., Walberg, M. W., Edwards, M. C. & Elledge, S. J. Finding prospective partners in the library: the two-hybrid system and phage display find a match. *Trends Biochem Sci* **20**, 511-6 (1995).

60. Enright, A. J., Iliopoulos, I., Kyrpides, N. C. & Ouzounis, C. A. Protein interaction maps for complete genomes based on gene fusion events. *Nature* **402**, 86-90 (1999).

61. Rodi, D. J., Makowski, L. & Kay, B. K. One from column A and two from column B: the benefits of phage display in molecular-recognition studies. *Curr Opin Chem Biol* **6**, 92-6 (2002).

62. Tong, A. H. et al. A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* **295**, 321-4 (2002).

63. Sidhu, S. S., Fairbrother, W. J. & Deshayes, K. Exploring protein-protein interactions with phage display. *Chembiochem* **4**, 14-25 (2003).

64. Kriplani, U. & Kay, B. K. Selecting peptides for use in nanoscale materials using phage-displayed combinatorial peptide libraries. *Curr Opin Biotechnol* **16**, 470-5 (2005).

65. Merzlyak, A. & Lee, S. W. Phage as templates for hybrid materials and mediators for nanomaterial synthesis. *Curr Opin Chem Biol* **10**, 246-52 (2006).

66. Cull, M. G., Miller, J. F. & Schatz, P. J. Screening for receptor ligands using large libraries of peptides linked to the C terminus of the lac repressor. *Proc Natl Acad Sci U S A* **89**, 1865-9 (1992).

67. Gates, C. M., Stemmer, W. P., Kaptein, R. & Schatz, P. J. Affinity selective isolation of ligands from peptide libraries through display on a lac repressor "headpiece dimer". *J Mol Biol* **255**, 373-86 (1996).

68. Speight, R. E., Hart, D. J., Sutherland, J. D. & Blackburn, J. M. A new plasmid display technology for the in vitro selection of functional phenotype-genotype linked proteins. *Chem Biol* **8**, 951-65 (2001).

69. Etz, H., Minh, D. B., Schellack, C., Nagy, E. & Meinke, A. Bacterial phage receptors, versatile tools for display of polypeptides on the cell surface. *J Bacteriol* **183**, 6924-35 (2001).

70. Bessette, P. H., Rice, J. J. & Daugherty, P. S. Rapid isolation of high-affinity protein binding peptides using bacterial display. *Protein Eng Des Sel* **17**, 731-9 (2004).

71. Rice, J. J., Schohn, A., Bessette, P. H., Boulware, K. T. & Daugherty, P. S. Bacterial display using circularly permuted outer membrane protein OmpX yields high affinity peptide ligands. *Protein Sci* **15**, 825-36 (2006).

72. Shimazu, M., Mulchandani, A. & Chen, W. Cell surface display of organophosphorus hydrolase using ice nucleation protein. *Biotechnol Prog* **17**, 76-80 (2001).

73. Cho, C. M., Mulchandani, A. & Chen, W. Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents. *Appl Environ Microbiol* **68**, 2026-30 (2002).

74. Becker, S., Schmoldt, H. U., Adams, T. M., Wilhelm, S. & Kolmar, H. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr Opin Biotechnol* **15**, 323-9 (2004).

75. Olsen, M. J. et al. Function-based isolation of novel enzymes from a large library. *Nat Biotechnol* **18**, 1071-4 (2000).

76. Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G. & Iverson, B. L. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc Natl Acad Sci U S A* **102**, 6855-60 (2005).

77. Georgiou, G. et al. Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol* **15**, 29-34 (1997).

78. Oker-Blom, C., Airenne, K. J. & Grabherr, R. Baculovirus display strategies: Emerging tools for eukaryotic libraries and gene delivery. *Brief Funct Genomic Proteomic* **2**, 244-53 (2003).

79. Makela, A. R. & Oker-Blom, C. Baculovirus display: a multifunctional technology for gene delivery and eukaryotic library development. *Adv Virus Res* **68**, 91-112 (2006).

80. Buchholz, C. J. et al. In vivo selection of protease cleavage sites from retrovirus display libraries. *Nat Biotechnol* **16**, 951-4 (1998).

81. Urban, J. H. et al. Selection of functional human antibodies from retroviral display libraries. *Nucleic Acids Res* **33**, e35 (2005).

82. Muller, O. J. et al. Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat Biotechnol* **21**, 1040-6 (2003).

83. Boder, E. T. & Wittrup, K. D. Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol* **328**, 430-44 (2000).

84. Ho, M., Nagata, S. & Pastan, I. Isolation of anti-CD22 Fv with high affinity by Fv display on human cells. *Proc Natl Acad Sci U S A* **103**, 9637-42 (2006).

85. Akamatsu, Y., Pakabunto, K., Xu, Z., Zhang, Y. & Tsurushita, N. Whole IgG surface display on mammalian cells: Application to isolation of neutralizing chicken monoclonal anti-IL-12 antibodies. *J Immunol Methods* **327**, 40-52 (2007).

86. Dalby, P. A., Hoess, R. H. & DeGrado, W. F. Evolution of binding affinity in a WW domain probed by phage display. *Protein Sci* **9**, 2366-76 (2000).

87. Heyd, B. et al. In vitro evolution of the binding specificity of neocarzinostatin, an enediyne-binding chromoprotein. *Biochemistry* **42**, 5674-83 (2003).

88. Drevelle, A. et al. Structures of in vitro evolved binding sites on neocarzinostatin scaffold reveal unanticipated evolutionary pathways. *J Mol Biol* **358**, 455-71 (2006).

89. Bass, S., Greene, R. & Wells, J. A. Hormone phage: an enrichment method for variant proteins with altered binding properties. *Proteins* **8**, 309-14 (1990).

90. Bernat, B., Sun, M., Dwyer, M., Feldkamp, M. & Kossiakoff, A. A. Dissecting the binding energy epitope of a high-affinity variant of human growth hormone: cooperative and additive effects from combining mutations from independently selected phage display mutagenesis libraries. *Biochemistry* **43**, 6076-84 (2004).

91. Gu, H. et al. A phage display system for studying the sequence determinants of protein folding. *Protein Sci* **4**, 1108-17 (1995).

92. Minard, P., Scalley-Kim, M., Watters, A. & Baker, D. A "loop entropy reduction" phage-display selection for folded amino acid sequences. *Protein Sci* **10**, 129-34 (2001).

93. Watters, A. L. & Baker, D. Searching for folded proteins in vitro and in silico. *Eur J Biochem* **271**, 1615-22 (2004).

94. Rudgers, G. W. & Palzkill, T. Protein minimization by random fragmentation and selection. *Protein Eng* **14**, 487-92 (2001).

95. Sieber, V., Pluckthun, A. & Schmid, F. X. Selecting proteins with improved stability by a phage-based method. *Nat Biotechnol* **16**, 955-60 (1998).

96. Riechmann, L. & Winter, G. Novel folded protein domains generated by combinatorial shuffling of polypeptide segments. *Proc Natl Acad Sci U S A* **97**, 10068-73 (2000).

97. Chu, R. et al. Redesign of a four-helix bundle protein by phage display coupled with proteolysis and structural characterization by NMR and X-ray crystallography. *J Mol Biol* **323**, 253-62 (2002).

98. Bai, Y. & Feng, H. Selection of stably folded proteins by phage-display with proteolysis. *Eur J Biochem* **271**, 1609-14 (2004).

99. Distefano, M. D., Zhong, A. & Cochran, A. G. Quantifying beta-sheet stability by phage display. *J Mol Biol* **322**, 179-88 (2002).

100. Rose, G. D. Protein folding and the Paracelsus challenge. *Nat Struct Biol* **4**, 512-4 (1997).

101. Dalal, S., Balasubramanian, S. & Regan, L. Protein alchemy: changing beta-sheet into alpha-helix. *Nat Struct Biol* **4**, 548-52 (1997).

102. Alexander, P. A., Rozak, D. A., Orban, J. & Bryan, P. N. Directed evolution of highly homologous proteins with different folds by phage display: implications for the protein folding code. *Biochemistry* **44**, 14045-54 (2005).

103. Laity, J. H., Lee, B. M. & Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr Opin Struct Biol* **11**, 39-46 (2001).

104. Ladomery, M. & Dellaire, G. Multifunctional zinc finger proteins in development and disease. *Ann Hum Genet* **66**, 331-42 (2002).

105. Thiel, G. & Cibelli, G. Regulation of life and death by the zinc finger transcription factor Egr-1. *J Cell Physiol* **193**, 287-92 (2002).

106. Urrutia, R. KRAB-containing zinc-finger repressor proteins. *Genome Biol* **4**, 231 (2003).

107. Lu, D., Searles, M. A. & Klug, A. Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* **426**, 96-100 (2003).

108. Gamsjaeger, R., Liew, C. K., Loughlin, F. E., Crossley, M. & Mackay, J. P. Sticky fingers: zinc-fingers as protein-recognition motifs. *Trends Biochem Sci* **32**, 63-70 (2007).

109. Matthews, J. M. & Sunde, M. Zinc fingers--folds for many occasions. *IUBMB Life* **54**, 351-5 (2002).

110. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *Embo J* **4**, 1609-14 (1985).

111. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**, 635-7 (1989).

112. Pabo, C. O., Peisach, E. & Grant, R. A. Design and selection of novel Cys2His2 zinc finger proteins. *Annu Rev Biochem* **70**, 313-40 (2001).

113. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-17 (1991).

114. Wuttke, D. S., Foster, M. P., Case, D. A., Gottesfeld, J. M. & Wright, P. E. Solution structure of the first three zinc fingers of TFIIIA bound to the cognate DNA sequence: determinants of affinity and sequence specificity. *J Mol Biol* **273**, 183-206 (1997).

115. Lu, D. & Klug, A. Invariance of the zinc finger module: a comparison of the free structure with those in nucleic-acid complexes. *Proteins* **67**, 508-12 (2007).

116. Choo, Y. & Klug, A. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci U S A* **91**, 11163-7 (1994).

117. Rebar, E. J. & Pabo, C. O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-3 (1994).

118. Jamieson, A. C., Kim, S. H. & Wells, J. A. In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33**, 5689-95 (1994).

119. Berman, H., Henrick, K. & Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* **10**, 980 (2003).

120. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714-23 (1997).

121. Segal, D. J., Dreier, B., Beerli, R. R. & Barbas, C. F., 3rd. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**, 2758-63 (1999).

122. Dreier, B., Beerli, R. R., Segal, D. J., Flippin, J. D. & Barbas, C. F., 3rd. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**, 29466-78 (2001).

123. Dreier, B. et al. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **280**, 35588-97 (2005).

124. Liu, Q., Segal, D. J., Ghiara, J. B. & Barbas, C. F., 3rd. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci U S A* **94**, 5525-30 (1997).

125. Nomura, W. & Sugiura, Y. Effects of length and position of an extended linker on sequence-selective DNA recognition of zinc finger peptides. *Biochemistry* **42**, 14805-13 (2003).

126. Yan, W., Imanishi, M., Futaki, S. & Sugiura, Y. Alpha-helical linker of an artificial 6-zinc finger peptide contributes to selective DNA binding to a discontinuous recognition sequence. *Biochemistry* **46**, 8517-24 (2007).

127. Isalan, M., Choo, Y. & Klug, A. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci U S A* **94**, 5617-21 (1997).

128. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**, 183-212 (2000).

129. Wolfe, S. A., Ramm, E. I. & Pabo, C. O. Combining structure-based design with phage display to create new Cys(2)His(2) zinc finger dimers. *Structure* **8**, 739-50 (2000).

130. Segal, D. J. et al. Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry* **42**, 2137-48 (2003).

131. Mandell, J. G. & Barbas, C. F., 3rd. Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-23 (2006).

132. Durai, S. et al. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res* **33**, 5978-90 (2005).

133. Smith, A. E. & Ford, K. G. Specific targeting of cytosine methylation to DNA sequences in vivo. *Nucleic Acids Res* **35**, 740-54 (2007).

134. Jamieson, A. C., Miller, J. C. & Pabo, C. O. Drug discovery with engineered zinc-finger proteins. *Nat Rev Drug Discov* **2**, 361-8 (2003).

135. Dervan, P. B., Doss, R. M. & Marques, M. A. Programmable DNA binding oligomers for control of transcription. *Curr Med Chem Anticancer Agents* **5**, 373-87 (2005).

136. Lee, D. K., Seol, W. & Kim, J. S. Custom DNA-binding proteins and artificial transcription factors. *Curr Top Med Chem* **3**, 645-57 (2003).

137. Klug, A. Towards therapeutic applications of engineered zinc finger proteins. *FEBS Lett* **579**, 892-4 (2005).

138. Doyon, Y. et al. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26**, 702-8 (2008).

139. Meng, X., Noyes, M. B., Zhu, L. J., Lawson, N. D. & Wolfe, S. A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26**, 695-701 (2008).

140. Miller, J. C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-85 (2007).

141. Moehle, E. A. et al. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc Natl Acad Sci U S A* **104**, 3055-60 (2007).

142. Perez, E. E. et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26**, 808-16 (2008).

143. Stoddard, B. L. Homing endonuclease structure and function. *Q Rev Biophys* **38**, 49-95 (2005).

144. Choo, Y. & Klug, A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A* **91**, 11168-72 (1994).

145. Garvie, C. W. & Wolberger, C. Recognition of specific DNA sequences. *Mol Cell* **8**, 937-46 (2001).

146. Fields, S. & Song, O. A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-6 (1989).

147. Chien, C. T., Bartel, P. L., Sternglanz, R. & Fields, S. The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci U S A* **88**, 9578-82 (1991).

148. Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci U S A* **97**, 7382-7 (2000).

149. Dang, C. V. et al. Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization. *Mol Cell Biol* **11**, 954-62 (1991).

150. He, R. & Li, X. Mammalian two-hybrid assay for detecting protein-protein interactions in vivo. *Methods Mol Biol* **439**, 327-37 (2008).

151. Parrish, J. R., Gulyas, K. D. & Finley, R. L., Jr. Yeast two-hybrid contributions to interactome mapping. *Curr Opin Biotechnol* **17**, 387-93 (2006).

152. Stelzl, U. & Wanker, E. E. The value of high quality protein-protein interaction networks for systems biology. *Curr Opin Chem Biol* **10**, 551-8 (2006).

153. Ruffner, H., Bauer, A. & Bouwmeester, T. Human protein-protein interaction networks and the value for drug discovery. *Drug Discov Today* **12**, 709-16 (2007).

154. Fiebitz, A. et al. High-throughput mammalian two-hybrid screening for protein-protein interactions using transfected cell arrays. *BMC Genomics* **9**, 68 (2008).

155. Ouwerkerk, P. B. & Meijer, A. H. Yeast one-hybrid screening for DNA-protein interactions. *Curr Protoc Mol Biol* **Chapter 12**, Unit 12 12 (2001).

156. Jung, D., Choi, Y. & Uesugi, M. Small organic molecules that modulate gene transcription. *Drug Discov Today* **11**, 452-7 (2006).

157. Liu, B., Alluri, P. G., Yu, P. & Kodadek, T. A potent transactivation domain mimic with activity in living cells. *J Am Chem Soc* **127**, 8254-5 (2005).

158. Rowe, S. P., Casey, R. J., Brennan, B. B., Buhrlage, S. J. & Mapp, A. K. Transcriptional up-regulation in cells mediated by a small molecule. *J Am Chem Soc* **129**, 10654-5 (2007).

159.   Baker, K. et al. Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc Natl Acad Sci U S A* **99**, 16537-42 (2002).

160.   Lin, H., Tao, H. & Cornish, V. W. Directed evolution of a glycosynthase via chemical complementation. *J Am Chem Soc* **126**, 15051-9 (2004).

161.   Brymora, A., Valova, V. A. & Robinson, P. J. Protein-protein interactions identified by pull-down experiments and mass spectrometry. *Curr Protoc Cell Biol* **Chapter 17**, Unit 17 5 (2004).

162.   Jestin, J. L. Functional cloning by phage display. *Biochimie* **90**, 1273-8 (2008).

163.   Wilson, D. S., Keefe, A. D. & Szostak, J. W. The use of mRNA display to select high-affinity protein-binding peptides. *Proc Natl Acad Sci U S A* **98**, 3750-5 (2001).

164.   Hammond, P. W., Alpin, J., Rise, C. E., Wright, M. & Kreider, B. L. In vitro selection and characterization of Bcl-X(L)-binding proteins from a mix of tissue-specific mRNA display libraries. *J Biol Chem* **276**, 20898-906 (2001).

165.   McPherson, M., Yang, Y., Hammond, P. W. & Kreider, B. L. Drug receptor identification from multiple tissues using cellular-derived mRNA display libraries. *Chem Biol* **9**, 691-8 (2002).

166.   Shen, X., Valencia, C. A., Szostak, J. W., Dong, B. & Liu, R. Scanning the human proteome for calmodulin-binding proteins. *Proc Natl Acad Sci U S A* **102**, 5969-74 (2005).

167.   Horisawa, K. et al. In vitro selection of Jun-associated proteins using mRNA display. *Nucleic Acids Res* **32**, e169 (2004).

168.   Tateyama, S. et al. Affinity selection of DNA-binding protein complexes using mRNA display. *Nucleic Acids Res* **34**, e27 (2006).

169.   Duenas, M. & Borrebaeck, C. A. Clonal selection and amplification of phage displayed antibodies by linking antigen recognition and phage replication. *Biotechnology (N Y)* **12**, 999-1002 (1994).

170.   Gramatikoff, K., Georgiev, O. & Schaffner, W. Direct interaction rescue, a novel filamentous phage technique to study protein-protein interactions. *Nucleic Acids Res* **22**, 5761-2 (1994).

171.   Krebber, C. et al. Selectively-infective phage (SIP): a mechanistic dissection of a novel in vivo selection for protein-ligand interactions. *J Mol Biol* **268**, 607-18 (1997).

172.   Jung, S., Arndt, K. M., Muller, K. M. & Pluckthun, A. Selectively infective phage (SIP) technology: scope and limitations. *J Immunol Methods* **231**, 93-104 (1999).

173.   Arndt, K. M., Jung, S., Krebber, C. & Pluckthun, A. Selectively infective phage technology. *Methods Enzymol* **328**, 364-88 (2000).

174.   Cebe, R. & Geiser, M. Size of the ligand complex between the N-terminal domain of the gene III coat protein and the non-infectious phage strongly influences the usefulness of in vitro selective infective phage technology. *Biochem J* **352 Pt 3**, 841-9 (2000).

175. Massoud, T. F., Paulmurugan, R., De, A., Ray, P. & Gambhir, S. S. Reporter gene imaging of protein-protein interactions in living subjects. *Curr Opin Biotechnol* **18**, 31-7 (2007).

176. Pelletier, J. N., Campbell-Valois, F. X. & Michnick, S. W. Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci U S A* **95**, 12141-6 (1998).

177. Zabin, I. beta-Galactosidase alpha-complementation. A model of protein-protein interaction. *Mol Cell Biochem* **49**, 87-96 (1982).

178. Dunn, I. S. In vitro alpha-complementation of beta-galactosidase on a bacteriophage surface. *Eur J Biochem* **242**, 720-6 (1996).

179. Rossi, F., Charlton, C. A. & Blau, H. M. Monitoring protein-protein interactions in intact eukaryotic cells by beta-galactosidase complementation. *Proc Natl Acad Sci U S A* **94**, 8405-10 (1997).

180. Michnick, S. W. Protein fragment complementation strategies for biochemical network mapping. *Curr Opin Biotechnol* **14**, 610-7 (2003).

181. MacDonald, M. L. et al. Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nat Chem Biol* **2**, 329-37 (2006).

182. Michnick, S. W., Ear, P. H., Manderson, E. N., Remy, I. & Stefan, E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat Rev Drug Discov* **6**, 569-82 (2007).

183. Odegrip, R. et al. CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. *Proc Natl Acad Sci U S A* **101**, 2806-10 (2004).

184. Reiersen, H. et al. Covalent antibody display--an in vitro antibody-DNA library selection system. *Nucleic Acids Res* **33**, e10 (2005).

185. Halpin, D. R. & Harbury, P. B. DNA display II. Genetic manipulation of combinatorial chemistry libraries for small-molecule evolution. *PLoS Biol* **2**, E174 (2004).

186. Love, K. R., Swoboda, J. G., Noren, C. J. & Walker, S. Enabling glycosyltransferase evolution: a facile substrate-attachment strategy for phage-display enzyme evolution. *Chembiochem* **7**, 753-6 (2006).

187. Strobel, H., Ladant, D. & Jestin, J. L. Efficient display of two enzymes on filamentous phage using an improved signal sequence. *Mol Biotechnol* **24**, 1-10 (2003).

188. Droge, M. J. et al. Binding of phage displayed Bacillus subtilis lipase A to a phosphonate suicide inhibitor. *J Biotechnol* **101**, 19-28 (2003).

189. Pedersen, J. S., Otzen, D. E. & Kristensen, P. Directed evolution of barnase stability using proteolytic selection. *J Mol Biol* **323**, 115-23 (2002).

190. Verhaert, R. M., Beekwilder, J., Olsthoorn, R., van Duin, J. & Quax, W. J. Phage display selects for amylases with improved low pH starch-binding. *J Biotechnol* **96**, 103-18 (2002).

191. Ponsard, I., Galleni, M., Soumillion, P. & Fastrez, J. Selection of metalloenzymes by catalytic activity using phage display and catalytic elution. *Chembiochem* **2**, 253-9 (2001).

192. Fernandez-Gacio, A., Uguen, M. & Fastrez, J. Phage display as a tool for the directed evolution of enzymes. *Trends Biotechnol* **21**, 408-14 (2003).

193. Soumillion, P. et al. Selection of beta-lactamase on filamentous bacteriophage by catalytic activity. *J Mol Biol* **237**, 415-22 (1994).

194. Vanwetswinkel, S., Avalle, B. & Fastrez, J. Selection of beta-lactamases and penicillin binding mutants from a library of phage displayed TEM-1 beta-lactamase randomly mutated in the active site omega-loop. *J Mol Biol* **295**, 527-40 (2000).

195. Forrer, P., Jung, S. & Pluckthun, A. Beyond binding: using phage display to select for structure, folding and enzymatic activity in proteins. *Curr Opin Struct Biol* **9**, 514-20 (1999).

196. Demartis, S. et al. A strategy for the isolation of catalytic activities from repertoires of enzymes displayed on phage. *J Mol Biol* **286**, 617-33 (1999).

197. Heinis, C. et al. Selection of catalytically active biotin ligase and trypsin mutants by phage display. *Protein Eng* **14**, 1043-52 (2001).

198. Matthews, D. J. & Wells, J. A. Substrate phage: selection of protease substrates by monovalent phage display. *Science* **260**, 1113-7 (1993).

199. Chaparro-Riggers, J. F., Breves, R., Maurer, K. H. & Bornscheuer, U. Modulation of infectivity in phage display as a tool to determine the substrate specificity of proteases. *Chembiochem* **7**, 965-70 (2006).

200. Richardson, P. L. The determination and use of optimized protease substrates in drug discovery and development. *Curr Pharm Des* **8**, 2559-81 (2002).

201. Hwang, B. Y. et al. Substrate specificity of the Escherichia coli outer membrane protease OmpP. *J Bacteriol* **189**, 522-30 (2007).

202. Kerr, F. K. et al. Elucidation of the substrate specificity of the MASP-2 protease of the lectin complement pathway and identification of the enzyme as a major physiological target of the serpin, C1-inhibitor. *Mol Immunol* **45**, 670-7 (2008).

203. Lien, S., Francis, G. L., Graham, L. D. & Wallace, J. C. Isolating substrates for an engineered alpha-lytic protease by phage display. *J Protein Chem* **22**, 155-66 (2003).

204. O'Boyle, D. R., 2nd, Pokornowski, K. A., McCann, P. J., 3rd & Weinheimer, S. P. Identification of a novel peptide substrate of HSV-1 protease using substrate phage display. *Virology* **236**, 338-47 (1997).

205.   Coombs, G. S., Bergstrom, R. C., Madison, E. L. & Corey, D. R. Directing sequence-specific proteolysis to new targets. The influence of loop size and target sequence on selective proteolysis by tissue-type plasminogen activator and urokinase-type plasminogen activator. *J Biol Chem* **273**, 4323-8 (1998).

206.   Kristensen, P. & Winter, G. Proteolytic selection for protein folding using filamentous bacteriophages. *Fold Des* **3**, 321-8 (1998).

207.   Martin, A., Sieber, V. & Schmid, F. X. In-vitro selection of highly stabilized protein variants with optimized surface. *J Mol Biol* **309**, 717-26 (2001).

208.   Wunderlich, M. & Schmid, F. X. In vitro evolution of a hyperstable Gbeta1 variant. *J Mol Biol* **363**, 545-57 (2006).

209.   Wunderlich, M., Martin, A., Staab, C. A. & Schmid, F. X. Evolutionary protein stabilization in comparison with computational design. *J Mol Biol* **351**, 1160-8 (2005).

210.   Andrews, S. R. et al. The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca2+-independent stable xylanases. *J Biol Chem* **279**, 54369-79 (2004).

211.   Heinis, C., Alessi, P. & Neri, D. Engineering a thermostable human prolyl endopeptidase for antibody-directed enzyme prodrug therapy. *Biochemistry* **43**, 6293-303 (2004).