Methods in Enzymology

Volume 328

Applications of Chimeric Genes and Hybrid Proteins

Part C
Protein–Protein Interactions and Genomics

EDITED BY

Jeremy Thorner

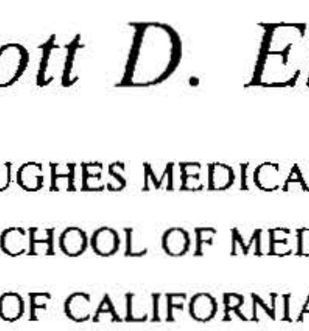
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA

Scott D. Emr

HOWARD HUGHES MEDICAL INSTITUTE
AND SCHOOL OF MEDICINE
UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA

John N. Abelson

CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA



ACADEMIC PRESS
San Diego London Boston New York Sydney Tokyo Toronto

[22] Selectively Infective Phage Technology

By KATJA M. ARNDT, SABINE JUNG, CLAUS KREBBER, and ANDREAS PLÜCKTHUN

reprinted from:
K. M. Arndt, S. Jung, C. Krebber and A. Plückthun, Selectively infective phage technology. Methods Enzymol. 328, 364-388 (2000).

Introduction

In this chapter we describe a selection technology related to phage display, the selectively infective phage technology (SIP). In both SIP and phage display, libraries of proteins or peptides of interest are displayed on the tip of a filamentous phage fused to the gene 3 protein (g3p). However, in contrast to conventional phage display, in which separation of binders with an immobilized ligand is required, in SIP technology the selection step is carried out in solution and directly coupled to the infection event. Therefore, no elution of binders is ever necessary, and incomplete elution of the most tightly binding molecules is not a limitation.

This is achieved by taking advantage of the modular structure of g3p, which occurs probably in five copies on the tip of the filamentous phage and consists of the N-terminal domains N1 [68 amino acids (aa)], N2 (131 aa), and the C-terminal domain CT (150 aa), connected by the glycine-rich linkers G1 (18 aa) and G2 (39 aa) (Fig. 1A and Fig. 2A).[1,2] In a SIP phage the N-terminal domains required for infection (N1 or N1–N2) are replaced by a protein or peptide of interest in all copies of g3p. This leads to a noninfective phage that displays the protein or peptide to be selected as a fusion to the C-terminal domain of g3p. The "adapter," consisting of either N1 or N1–N2, is fused to the second protein or peptide of interest or chemically coupled to a nonpeptidic ligand (Fig. 1B and C). No wild-type g3p must be present on a SIP phage. Therefore, in its simplest form, the CT-fusion protein directly replaces the g3p in the phage genome. However, we discuss below other strategies for carrying out SIP with phagemid/helper phage systems. To simplify the nomenclature, we refer to the protein or peptide library fused to the CT domain of g3p, and thus displayed on the phage, as "A" and denote the target fused to or chemically coupled to the N-terminal part of g3p, the soluble adapter, as "B." Cognate interaction between these two molecules (A and B) restores the gene 3 protein in a noncovalent form and thus allows for phage infection. There are two variants of this selection system, termed *in vitro* SIP and *in vivo* SIP.

In the *in vitro* SIP procedure only the fusion of library A to the C-

[1] I. Stengele, P. Bross, X. Garces, J. Giray, and I. Rasched, *J. Mol. Biol.* **212,** 143 (1990).
[2] J. Armstrong, R. N. Perham, and J. E. Walker, *FEBS Lett.* **135,** 167 (1981).

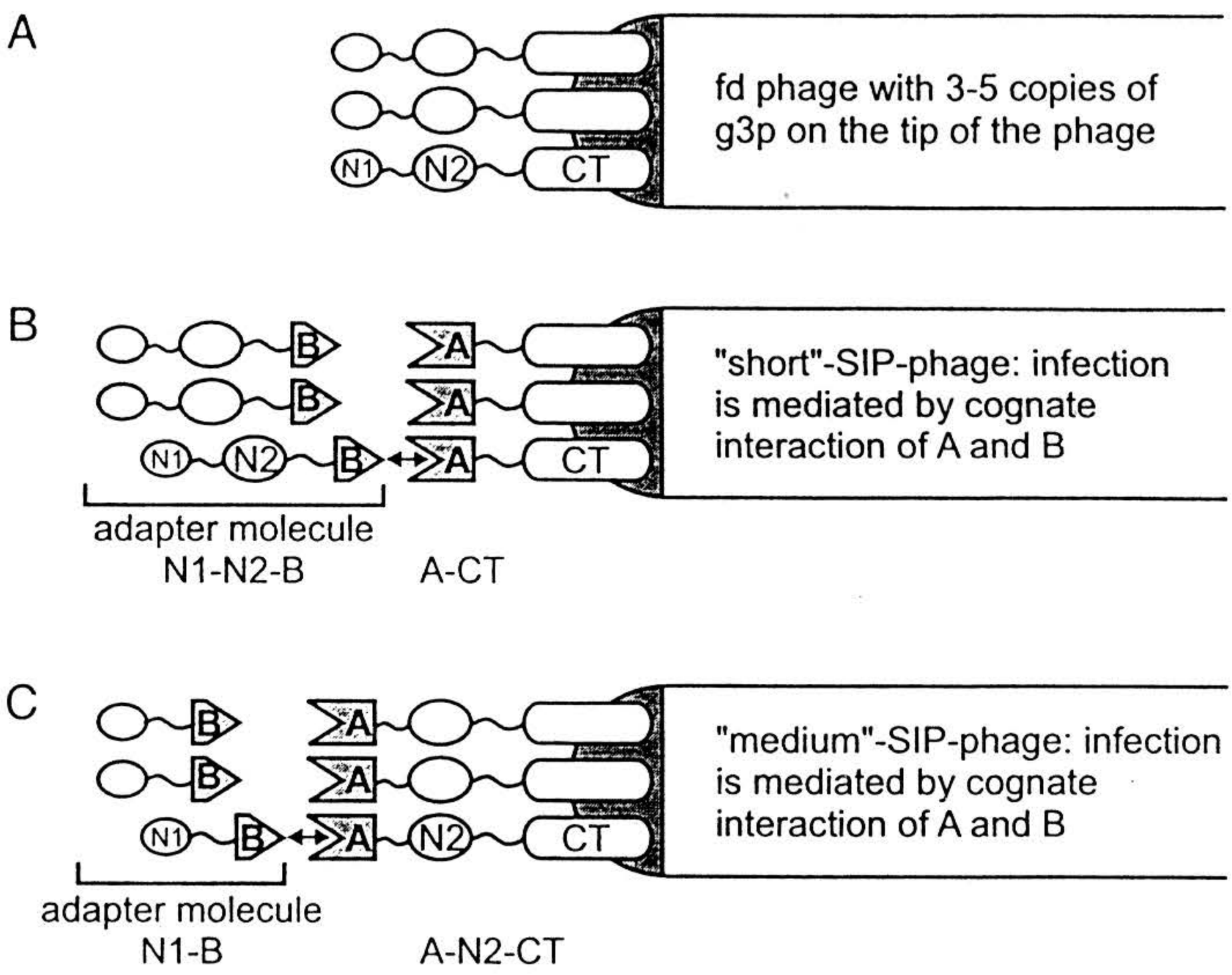


FIG. 1. Overview of the arrangement of the gene 3 protein in various phages. (A) Fd phages carry on their tip three to five copies of the gene 3 protein, which consists of three domains, the two N-terminal domains N1 and N2, required for infection and binding to the F pilus, respectively, and the C-terminal domain CT, which anchors the N-terminal domains in the phage particle. (B) SIP phage as used for *in vivo* and *in vitro* SIP. One protein of interest is displayed as a fusion to CT on the phage ("short" phage), whereas N1–N2 are expressed as a fusion to or chemically coupled to the other protein/molecule of interest. In *in vitro* SIP, the adapter molecule is expressed and purified separately and added later for infection, whereas in *in vivo* SIP both proteins are expressed simultaneously on the phage genome. (C) Alternative SIP phage used in *in vitro* SIP. In this construct, the "medium" phage displays the protein of interest fused to N2–CT, whereas the adapter consists only of N1 and the second protein of interest.

terminal part of g3p is encoded by the phage or phagemid vector (Fig. 2C and D). The adapter molecule with the N-terminal part of the gene 3 protein is expressed and purified separately and can then be chemically coupled to a great variety of targets, including nonpeptidic ones. The separate expression of the adapter molecule allows for better control of the infection experiments by enabling quality control of the adapter molecule, titration of adapter, and measurement of background infectivity of purified phage particles. For *in vivo* SIP, both parts of the gene 3 protein are expressed simultaneously *in vivo,* and thus the genetic information of both parts remains linked to the phage phenotype. This means that, in principle, a library-versus-library approach should be feasible, in which one library is fused to the CT domain (library A) and the other library is fused to the N-terminal (adapter) part of the gene 3 protein (library B) (Fig. 2E).

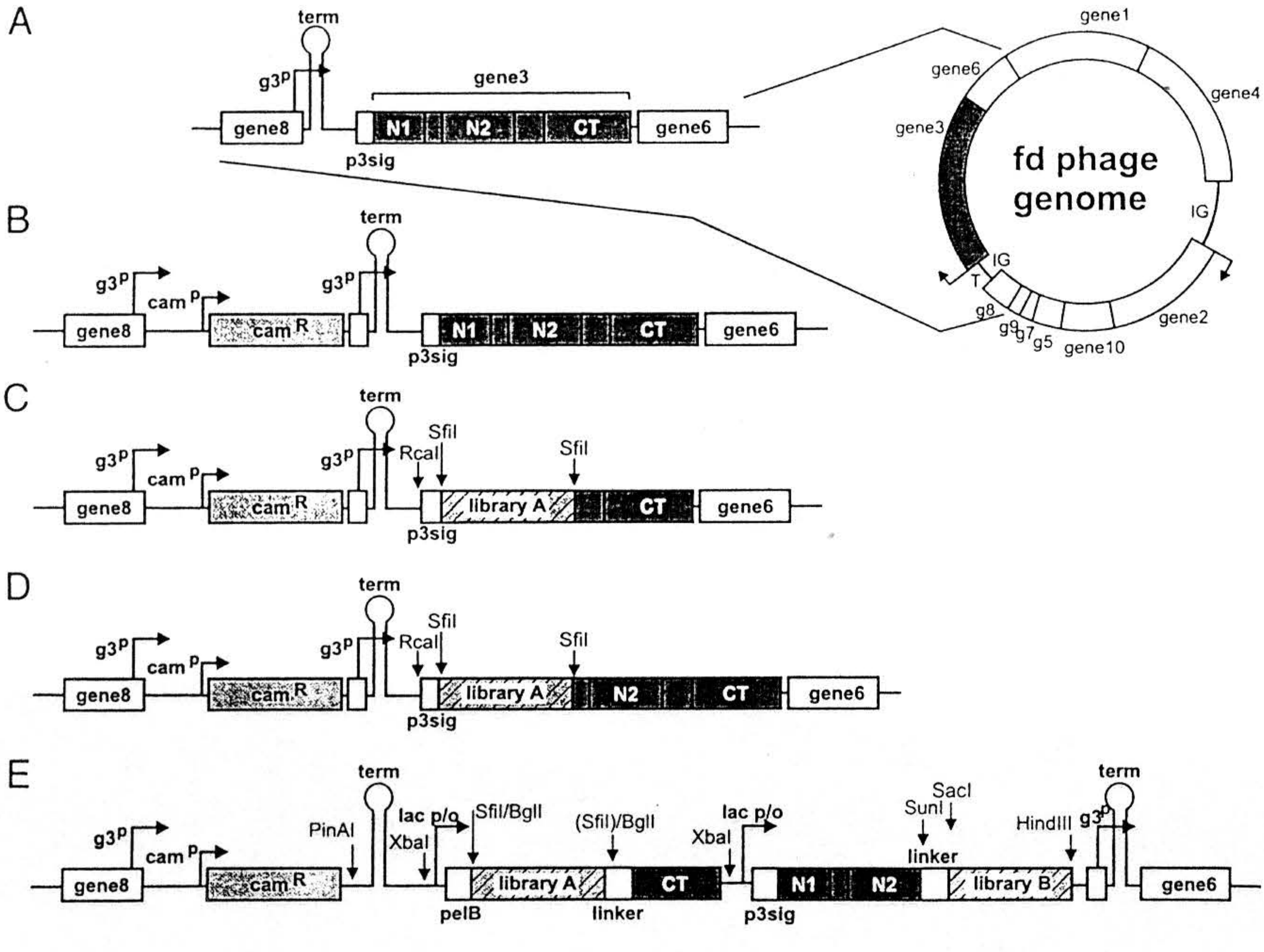


FIG. 2. (A) Organization of the fd phage genome. The genome of the wild-type phage is subdivided by two intergenic regions (IG) into a region of strongly expressed genes (genes 2, 5, 7, 8, 9, and 10) and a region of weakly expressed genes (genes 1, 3, 4, and 6). A terminator (*term*) separates the weakly expressed genes from the strongly expressed ones. The organization of the gene segment of the fd phage showing gene 8, gene 3 promotor ($g3^p$), gene 3 terminator (*term*), gene 3 (dark gray) encoding the three domains N1, N2, and CT of the wild-type gene 3 protein, and gene 6 is shown in more detail. *p3sig* indicates the signal sequence for the gene 3 protein, necessary for transport into the periplasm. (B) In the derivative of the fd phage fCKC, which served as a starting construct for all other constructs, a chloramphenicol resistance gene (Cam^R, light gray) is inserted downstream of gene 8 and upstream of the natural terminator. The end of gene 8 is repeated, as it carries the promotor for g3p.[7] (C) Gene arrangement of the "short" phage (library A-CT) and (D) "medium" phage (library A-N2-CT) as used for *in vitro* SIP. The fusion is expressed under the natural gene 3 promotor ($g3^p$), and library A can be exchanged by two *Sfi*I sites with different cutting sequences. (E) Gene arrangement of the two parts of the gene 3 protein used in *in vivo* SIP. Libraries A and B mark the places where the proteins/peptides of interest are encoded, which can be a single sequence or a library. In this case, both fusions are under the control of the *lac* promotor/operator system (*lac* p/o), and a second terminator is inserted in front of gene 6. Both libraries can be cloned by unique restriction sites as indicated. Restriction enzymes in parentheses indicate sites that are not present in all constructs.

However, this system, which is more powerful but also less controllable, can be used only with proteins and peptides for both partners.

In the following section we provide several general protocols, which apply for *in vitro* as well as *in vivo* SIP. In the second and third part, we focus on *in vitro* and *in vivo* SIP, and compare in more detail the use of both methods by means of a few examples. Two other names, DIRE (direct rescue interaction)[3] and SAP (selection and amplification of phage),[4] have also been used for this technology.

General Procedures

Bacterial Strains

XL1-Blue (Stratagene, La Jolla, CA) (*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [*F′ proAB lacI*^{q}*ZΔM15* Tn*10* (Tet^R)]) is the preferred host for electroporation and for infection experiments because of its *recA1* genotype. For infection, it is possible to switch resistances by alternatingly using the tetracycline- or the kanamycin-resistant strain XL1-Blue MRF′ Kan (Stratagene; Δ(*mcrA*) *183* Δ(*mcrCB-hsdSMR-mrr*) *173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI*^{q} *ZΔM15* Tn5 (Kan^R)]). This ensures that during phage infection only phages but no cells are carried over. An additional precaution is to filter the phage-containing supernatant directly after phage production and prior to polyethylene glycol (PEG) precipitation (see Phage Production, below). BL21 (DE3) (Novagen, Madison, WI) [F$^-$ *ompT hsdS*$_B$ (r_B^- m_B^-) *gal dcm,* harboring DE3, a λ prophage carrying the T7 RNA polymerase gene] is used for the cytoplasmic production of N1–N2 adapter molecules for *in vitro* SIP. BL21 (DE3)pLysS (Novagen) [same as BL21 (DE3), but harboring the plasmid pLysS Cam^R, which encodes T7 lysozyme, a natural inhibitor of T7 RNA polymerase; resistant to chloramphenicol at 40 μg/ml] is used for the cytoplasmic production of N1 adapter molecules for *in vitro* SIP.

Vector Constructs: Phage versus Phagemid System

In SIP, the infection is strictly dependent on the reconstitution of the gene 3 protein. Therefore, when using a phagemid/helper phage system, it is absolutely necessary that the helper phage does not deliver any wild-type g3p to the phage particle. For this purpose, Δg3 helper phages were constructed.[5,6] These phages, while not carrying the gene, do need to carry

[3] K. Gramatikoff, O. Georgiev, and W. Schaffner, *Nucleic Acids Res.* **22,** 5761 (1994).
[4] M. Dueñas and C. A. Borrebaeck, *Bio/Technology* **12,** 999 (1994).
[5] M. Dueñas and C. A. Borrebaeck, *FEMS Microbiol. Lett.* **125,** 317 (1995).
[6] F. K. Nelson, S. M. Friedman, and G. P. Smith, *Virology* **108,** 338 (1981).

functional g3p on their coats, because the helper phage must be able to infect the bacteria harboring the phagemid library for which it needs the g3 protein. There are two ways to produce such a phage. First, the helper phage can be produced in cells harboring an unpackageable g3p expression plasmid, which does not contain an F1 origin.[4,5] Nonetheless, we still observed some packaging of the g3p expression plasmid, which increased the background in subsequent infection experiments, as the population will produce phages carrying wild-type g3p. This problem has more recently been solved by engineering an *Escherichia coli* strain with a g3p expression cassette integrated in the chromosome (S. Rondot and F. Breitling, personal communication, 1999). An alternative way is to transform the phagemid library into cells that already contain Δg3 helper phage.[3] However, at least in our experience, the helper phage Δg3 M13KO7 showed genetic instabilities and lowered the viability and transformation frequency of bacteria.[7]

A third approach, favored by the authors, is to directly use a phage genome, in which the complete genetic information of the phage is encoded. The circular genome of the filamentous phage consists of two main transcriptional units that are separated by a central terminator on one side and the origin of replication on the other (Fig. 2A).[8] The wild-type gene 3 is modified accordingly for *in vitro* and *in vivo* SIP, and the resulting gene 3 cassettes are described further below. In addition, an antibiotic resistance gene was inserted into the fd genome, immediately downstream of gene 8 and upstream of the original gene 3, to enable detection of infection events by screening for resistance (Fig. 2B). Several antibiotic resistances were tested, and resistance against chloramphenicol was chosen, because it gave, along with ampicillin resistance, the highest level of infectivity, and it had the advantage that it does not allow growth of plasmid-free cells due to enzyme leakage as observed for ampicillin. This resulted in the phage vector fCKC,[7] which is the basis for all other constructs.

Vector Constructs: Different Gene 3 Fusions

Similar to phage display, a peptide or protein library of interest can be fused to different parts of the gene 3 protein of the phage. In the "short" fusion, the protein of interest is expressed as a fusion to the CT domain, while in the "medium" fusion, the protein or peptide library of interest is fused to the N terminus of the N2–CT fragment. Conversely, the adapter can consist either of only the N1 "bait" fusion or an N1–N2 "bait" fusion (Fig. 1B and C). In one combination of phage and adapter, no N2 domain at all is present, while in another N2 is present both on the adapter and

[7] C. Krebber, S. Spada, D. Desplancq, and A. Plückthun, *FEBS Lett.* **377,** 227 (1995).
[8] E. Beck and B. Zink, *Gene* **16,** 35 (1981).

the phage. All four combinations lead to infectivity as tested for *in vitro* SIP, albeit at different levels and with a different concentration dependence of the adapter.[9] Only the "short" fusion has been tested for *in vivo* SIP so far.

Library Construction

The library will of course depend on the particular question to be addressed with SIP. A number of standard libraries is available nowadays, including peptide libraries, antibody libraries, or cDNA libraries, some of which are even commercially available. In addition, several techniques have been developed for generating genetic diversity using, e.g., PCR (polymerase chain reaction) techniques,[10–13] recombination sites,[14] mutator strains or degenerated codons,[15,16] and trinucleotides[17] for randomization of synthetic genes. These methods are numerous and well established, and therefore are not detailed here. However, we discuss the issue of library complexity and quality in connection with the *in vivo* SIP procedure.

Phage Production

Phage expression in XL1-Blue cells carrying the phage DNA is performed in shake flasks in 2× YT medium[18] supplemented with 1% (w/v) glucose, chloramphenicol (Cam; 30 μg/ml), and tetracycline (Tet; 5 μg/ml). Overnight cultures are grown at 37° (or 25° when there is reason to believe that the *in vivo* folding yield of the ligands might be increased on lowering the temperature). For *in vitro* SIP with the constructs shown, in which the gene 3 fusion protein is expressed under the natural gene 3 promotor, phages are obtained directly from the supernatant of the overnight culture. For *in vivo* SIP, where both parts of gene 3 are under *lac* promoter/operator (*lac* p/o) control, a main culture containing Cam (30

[9] C. Krebber, S. Spada, D. Desplancq, A. Krebber, L. Ge, and A. Plückthun, *J. Mol. Biol.* **268,** 607 (1997).

[10] R. C. Cadwell and G. F. Joyce, *Genome Res.* **3,** S136 (1994).

[11] A. Crameri, S. A. Raillard, E. Bermudez, and W. P. C. Stemmer, *Nature* (*London*) **391,** 288 (1998).

[12] W. P. C. Stemmer, *Nature* (*London*) **370,** 389 (1994).

[13] H. Zhao, L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold, *Nature Biotechnol.* **16,** 258 (1998).

[14] N. Tsurushita, H. Fu, and C. Warren, *Gene* **172,** 59 (1996).

[15] S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik, and M. H. Hecht, *Science* **262,** 1680 (1993).

[16] E. Wolf and P. S. Kim, *Protein Sci.* **8,** 680 (1999).

[17] B. Virnekäs, L. Ge, A. Plückthun, K. C. Schneider, G. Wellnhofer, and S. E. Moroney, *Nucleic Acids Res.* **22,** 5600 (1994).

[18] J. Sambrook, E. F. Fritsch, and T. Maniatis, "Molecular Cloning: A Laboratory Manual," 2nd Ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989.

μg/ml), Tet (5 μg/ml), and glucose (1%, w/v) is inoculated from the overnight culture to an initial OD_{550} of 0.15 (typical dilution of 1:25–1:30) and grown at 37° until an OD_{550} of 0.5 to 0.6 is reached. The cells are then pelleted by centrifugation (3000*g*, 10 min, 4–10°) and resuspended in fresh 2× YT medium without glucose containing the appropriate antibiotics and 0.5 m*M* isopropyl-β-D-thiogalactopyranoside (IPTG) to induce the expression of the two fusion proteins, protein A–CT and N1–N2–protein B, which results in phage packaging. Expression is continued for at least 6 hr at 37°, or in the case of less stable proteins at 24–26° overnight. In both cases, phages are separated from the cells by centrifugation for 10–15 min at 5000*g* and 4°. The supernatant containing the phages can then be optionally filtered (0.22-μm pore size filters) to remove any remaining cells. The yield is usually in the range of 5×10^{11} to 2×10^{12} phages per milliliter of culture supernatant.

Phage Purification

For the infection experiments, phage particles are enriched by two precipitations with polyethylene glycol (PEG). The most convenient way is to prepare a five-times concentrated stock solution with 16% (w/v) PEG 6000 and 3.3 *M* NaCl, passed through a 0.22-μm pore size filter. For the PEG precipitation, one part of this solution is added directly to four parts of the supernatant obtained after phage production and cell precipitation, and incubated on ice or at 4° for at least 1 hr or overnight. Phages are pelleted by centrifugation at least at 5000*g* for 30 min at 4° and redissolved in about 1 ml of TBS buffer [Tris-buffered saline, 50 m*M* Tris-HCl (pH 8.0), 150 m*M* NaCl]. If higher purity or concentrations are required, a second PEG precipitation can be performed similar to the first one. For this purpose, PEG–NaCl solution is added in the same ratio as for the first PEG precipitation, and after 1 hr of incubation at 4° phages are pelleted in a 2-ml tube by centrifuging at maximal speed for 20 min at 4°. Phages are then dissolved in 50–200 μl of TBS buffer. They can be stored at 4° for several months, depending mainly on the stability of the fusion protein.

However, if phages of greater purity are required as, e.g., for a more accurate quantification or electron microscopy, phages are best purified by ultracentrifugation in a CsCl gradient.[19] Phages are first enriched 80-fold by PEG precipitation, and then added to a solution of 1.6 g of CsCl in TBS, filled to a final volume of 4 ml. The CsCl solution is transferred to a 13 × 38 mm polyallomer tube (Quick-Seal; Beckman, Fullerton, CA) and centrifuged at 100,000 rpm (541,000*g*) for 4 hr in a TLN-100 rotor (Beckman) at 4°. After ultracentrifugation the phages become visible as a white

[19] G. P. Smith and J. K. Scott, *Methods Enzymol.* **217,** 228 (1993).

band through light scattering when shining light from the top to the bottom of the tube and by looking through the wall of the tube at a right angle to the light beam. Phages are removed by puncturing the side of the tube with a hypodermic needle and gently removing the banded phages. To remove the remaining CsCl, phages are then pelleted by ultracentrifugation (50,000 rpm/135,000*g*, 1 hr, 4°, TLA-100.3 rotor), redissolved in 3 ml of TBS buffer, pelleted again (same conditions), and redissolved in 50 μl of TBS buffer.

Determination of the Phage Concentration

Phage particles are quantified spectrophotometrically by measuring the absorption between 200 and 350 nm. A broad peak between 260 and 280 nm is obtained due to the presence of both DNA and proteins, and the absorption at 269 nm is used to estimate the phage concentration[20] as follows:

$$\frac{\text{Phage particles}}{\text{ml}} = \left(\frac{OD_{269}}{\text{number of base pairs of ssDNA}}\right)(5.98 \times 10^{16})$$

This method, however, gives only a crude estimate and upper limit of phage concentration for phages that are prepared merely by PEG precipitation, because proteins from the culture supernatant can be coprecipitated by PEG. Especially in cases of low phage production, coprecipitated proteins can influence the absorption significantly, as seen by a red shift of the absorption maximum. This can be avoided by using phages purified by CsCl gradient ultracentrifugation. However, in our experience, "average" phage titers (about 10^{12} phages per milliliter of supernatant), prepared by two subsequent PEG precipitations, have a "normal" absorption spectrum with no red shift (comparable to purified phages), and usually give reliable results in concentration determinations, when compared with phages purified by ultracentrifugation.

Another possibility for phage quantification is by using an enzyme-linked immunosorbent assay (ELISA), by immobilizing the phage and detecting with an antibody against gene 8 protein. However, this method does not yield absolute phage concentrations, but relative numbers, and is therefore best suited for comparative infection experiments. In this case, phages prepared by PEG precipitation are coated overnight at 4° on an ELISA plate in a twofold dilution series. The ELISA plate is washed three times with TBS and blocked with 4–5% (w/v) milk powder in TBS for 1 hr at room temperature. After washing twice with TBS, phages are detected with a 1:5000 dilution of anti-M13–peroxidase conjugate (Pharmacia, Pis-

[20] L. A. Day, *J. Mol. Biol.* **39**, 265 (1969).

cataway, NJ) in TBST [TBS with 0.05% (v/v) Tween 20]. The amount of bound anti-M13–peroxidase is measured with peroxidase (POD) soluble substrate (Boehringer Mannheim, Mannheim, Germany) at 405 nm after various incubation times. To obtain absolute concentrations, a control phage dilution needs to be included with known concentration. Preferably, the concentration of this calibration standard should have been determined by UV absorption after purification by CsCl gradient ultracentrifugation.

In Vitro Selectively Infective Phage Procedures

In the *in vitro* SIP procedure, the adapter molecule is purified separately, and later added to phages displaying the protein or peptide library of interest, which is fused to the C-terminal part of the gene 3 protein. Thus, nonpeptidic "baits" can also be used, which can be chemically coupled to N1 or N1–N2, respectively. In addition, the concentration of the adapter molecule can be varied.

Design of Vectors Suitable for in Vitro Selectively Infective Phage

The phage vector used in these examples is based on fCKC[7] and encodes the protein or library of interest with the gene 3 leader sequence as fusion to either N2–CT ("medium" phage) or CT ("short" phage) under the control of the natural gene 3 promotor. The phage production is carried out in XL1-Blue as described in the general procedures.

Native, periplasmic expression of the N-terminal domains of gene 3 was found to be difficult, as the native protein always led to complete lysis of the culture and relatively low yields requiring a tedious concentration step. On the other hand, the N1 domain alone, lacking the glycine-rich linker, can be obtained by secretion.[21] However, we favor cytoplasmic expression of the gene 3 N-terminal domains under the control of the strong T7 promoter, using the vector pTFT74,[22] as this appears to be more reliable and simpler for downstream processing. The deletion of the signal sequence and the T7 expression system leads to cytoplasmic inclusion bodies, which are easily refolded. Interestingly, the toxicity effects of secreted proteins vary significantly for the various constructs. The highest toxicity is observed for the N1 domain with a C-terminal glycine-rich linker, whereas the expression of the N1–N2 complex caused significantly fewer problems. However, N1–N2 with a C-terminal glycine-rich linker showed again more problems with expression (lower yield and higher toxicity) than N1–N2 without the

[21] P. Holliger and L. Riechmann, *Structure* **5,** 265 (1997).
[22] C. Freund, A. Ross, B. Guth, A. Plückthun, and T. A. Holak, *FEBS Lett.* **320,** 97 (1993).

A

```
   1  AETVESCLAK SHTENSFTNV WKDDKTLDRY ANYEGCLWNA TGVVVCTGDE
  51  TQCYGTWVPI GLAIPENEGG GSEGGGSEGG GSEGGGTKPP EYGDTPIPGY
 101  TYINPLDGTY PPGTEQNPAN PNPSLEESQP LNTFMFQNNR FRNRQGALTV
 151  YTGTVTQGTD PVKTYYQYTP VSSKAMYDAY WNGKFRDCAF HSGFNEDLFV
 201  CEYQGQSSDL PQPPVNAPSG CPHHHHHH*
```

B

```
   1  AETVESCLAK SHTENSFTNV WKDDKTLDRY ANYEGCLWNA TGVVVCTGDE
  51  TQCYGTWVPI GLAIPENEGG GSEGGGSEGG GSGCPHHHHH H*
```

FIG. 3. Amino acid sequence of (A) N1–N2 and (B) N1 as used for *in vitro* SIP. The initiator methionine, which is needed for expression in the T7 system and that becomes cleaved afterward, is not shown. The engineered cysteine is underlined and the hexahistidine (His_6) tag is shown in boldface. (A) N1–N2 domain: N1 (amino acids 1–68) followed by G1 (amino acids 69–86), N2 (amino acids 87–217), and the Cys–His_6 tag (228 residues; 24,854.9 Da, p*I* 4.58; $\varepsilon = 41{,}100\ M^{-1}\ cm^{-1}$). (B) N1 domain: N1 (amino acids 1–68) followed by the first 14 amino acids of G1 and the Cys–His_6 tag for purification and coupling (91 residues; 9,658.37 Da; p*I* 4.87; $\varepsilon = 21{,}210\ M^{-1}\ cm^{-1}$).

long glycine-rich linker. To obtain cytoplasmic expression, the gene 3 signal sequence (amino acids 1–18) is deleted, and a methionine is added as a start codon, which becomes cleaved, as detected by mass spectrometry and N-terminal sequencing. In addition, a C-terminal Ser-Gly-Cys-Pro-His_6 tag is introduced to allow chemical coupling (Cys) and easy purification (hexahistidine, His_6) (Fig. 3A and B). We use the N1–N2 domain with only three additional residues before the engineered cysteine (Fig. 3A) and the N1 domain alone followed by the first 14 amino acids from the linker G1 before the cysteine (Fig. 3B) for *in vitro* SIP experiments. In addition, we could produce in the same way the N1–N2 complex followed by a glycine-rich linker and the N2 domain alone followed by its glycine-rich linker. Similar expression constructs have been used to produce N1–N2 for X-ray crystallography,[23] and N1–TolA for X-ray crystallography.[24] As mentioned above, N1 lacking the glycine-rich linker and Cys-His_6 tag could be purified from culture supernatant for nuclear magnetic resonance (NMR).[21]

[23] J. Lubkowski, F. Hennecke, A. Plückthun, and A. Wlodawer, *Nature Struct. Biol.* **5,** 140 (1998).

[24] J. Lubkowski, F. Hennecke, A. Plückthun, and A. Wlodawer, *Structure Fold. Des.* **7,** 711 (1999).

Expression of the N1–N2 Domain of the Gene 3 Protein

Plasmids carrying the gene for cytoplasmic expression of N1–N2 (Fig. 3A) are transformed into BL21 (DE3), plated onto 2× YT agar plates,[18] containing Amp (200 μg/ml) and 1% (w/v) glucose, and grown overnight at room temperature. A single colony of each host/plasmid is grown in 2× YT [containing Amp at 200 μg/ml, 1% (w/v) glucose] for 8 hr. Subsequently, a two-thirds volume of 50% (v/v) glycerol is added and the cells are frozen as glycerol stocks at −80°. For the expression culture, 50 μl of the glycerol stock is used to inoculate 50 ml of SB medium [2% (w/v) tryptone, 1% (w/v) yeast extract, 0.5% (w/v) NaCl], containing Amp (200 μg/ml), 1% (w/v) glucose, 2% (w/v) glycerol, 50 m*M* K_2HPO_4, and 10 m*M* $MgCl_2$, and the culture is grown overnight at 37°. The overnight culture is used to inoculate the main culture (2 liters of SB medium with the same additives). The main culture is grown at 37° to an OD_{550} of 0.9 to 1.2, and then IPTG is added to a final concentration of 1 m*M*, and the cells are grown for another 3–4 hr.

Expression of the N1 Domain of the Gene 3 Protein

Because of the high toxicity and thus instability of the expression culture for the N1 domain alone (Fig. 3B), it is produced in BL21(DE3) harboring the plasmid pLysS, which encodes T7 lysozyme, a natural inhibitor of T7 RNA polymerase.[25] This plasmid also confers resistance to chloramphenicol (up to 40 μg/ml). After transformation, cells are plated onto 2× YT agar plates, containing Amp at 200 μg/ml, Cam at 30 μg/ml, and 1% (w/v) glucose, and grown overnight at room temperature. For glycerol stocks (see preceding section), 2× YT medium supplemented with Amp (200 μg/ml) and 1% (w/v) glucose is inoculated with a single colony and grown for 8 hr. For production of N1 alone a more elaborate protocol must be used than in the case of N1–N2 production. A dilution of 50 μl of glycerol stock with 250 μl of SB medium is plated on three 100-mm SB–agar plates, containing Amp (200 μg/ml), Cam (30 μg/ml), 1% (w/v) glucose, and 5% (v/v) glycerol. After incubation overnight at 37°, the colonies (near confluence) are scraped from the plates with 20 ml of SB medium, which is then used to inoculate 2 liters of the main culture. The main culture [2 liters of SB medium containing Amp (200 μg/ml), Cam (30 μg/ml), 1% (w/v) glucose, 2% (v/v) glycerol, 50 m*M* K_2HPO_4, 10 m*M* $MgCl_2$] is grown at 37° to an OD_{550} of 0.9 to 1.2, IPTG is added to a final concentration of 1 m*M*, and the cells are grown for another 3–4 hr.

As the expression of the N1 domain with the single cysteine is not

[25] F. W. Studier, *J. Mol. Biol.* **219,** 37 (1991).

totally reliable even in the described experimental setup, two or three main cultures (2 liters each) are started separately and small aliquots of each are analyzed. In contrast, fusion proteins of N1 with other proteins give reliable inclusion bodies. Cells from 50 ml of culture are centrifuged and resuspended in 1 ml of 25 m*M* Tris-HCl, pH 7.5. Four milliliters of 6 *M* guanidine hydrochloride, 25 m*M* Tris-HCl, pH 7.4, is added to lyse the cells. Because of the liberated nucleic acids the lysate becomes viscous and must be sonicated. After centrifugation (SS34, 20,000 rpm/48,000*g*, 10 min, 4°) the supernatant is filtered through a 0.45-μm pore size filter. The filtrate is analyzed by coupled IMAC-AIEX as described in detail below. Small aliquots of the peak fractions can be loaded onto a sodium dodecyl sulfate (SDS)–polyacrylamide gel, as the sample is eluted from the AIEX in a nonionic urea buffer.

Cell Rupture and Enrichment of the Inclusion Bodies

Cells are harvested by centrifugation (GS3, 5000 rpm/4000*g*, 10 min, 4°). The cell pellet of a 2-liter culture is resuspended in 25 ml of 50 m*M* Tris-HCl, 1 m*M* $MgCl_2$ (pH 8.0), and 10 mg of RNase A, 10 mg of DNase I, and 25 mg of hen egg white lysozyme are added. The cell suspension is ruptured in a French press (Aminco). After centrifugation (SS34, 20,000 rpm/48,000*g*, 30 min, 4°) the pellet containing the inclusion bodies is resuspended in 25 ml of 20 m*M* Tris-HCl, 23% (w/v) sucrose, 0.5% (v/v) Triton X-100, 1 m*M* EDTA, pH 8.0. The suspension is stirred by a magnetic stir bar at 4° for 30 min, pelleted again, and the wash step is repeated. During the washing procedure the inclusion body pellet should change its color from brownish-yellow to white. To the final washed inclusion body pellet, 20 ml of 5.5 *M* guanidine hydrochloride, 25 m*M* Tris-HCl, pH 7.5, is added and the pellet is solubilized by stirring at room temperature. Insoluble material is removed by centrifugation (SS34, 20,000 rpm/48,000*g*, 30 min), and the supernatant is filtered through a 0.45-μm pore size filter.

Protein Purification

Two routes can be chosen to obtain purified gene 3 N-terminal domains from the inclusion bodies. The first procedure is the more classic way: refolding followed by purification. For the second, the protein is first purified under denaturing conditions. Both procedures are equivalent in yield. However, the second procedure may be advantageous especially when proteases are of concern and when a more defined and pure denatured sample is necessary (e.g., to test refolding parameters for more complicated fusion proteins). Using the second route, we were able to refold the protein by dialysis and did not have to refold by dilution, thus saving an additional

concentration step. Overall, the second method is faster and more straightforward. Chromatography can be performed with any equipment; however, the automated two-column format simplifies the procedure significantly.[26] Immobilized metal-ion affinity chromatography (IMAC) is usually able to enrich the His-tagged adapter protein to a high extent, although some contaminants are still present. Anion- or cation-exchange chromatography is a good second step after IMAC, as the relatively pure IMAC-purified protein can often be baseline separated. In both cases the chromatography, performed with a BioCAD 60 system (PerSeptive Biosystems, Framingham, MA) takes about 30 min to obtain pure protein.

Route 1: Refolding and Subsequent Purification of the Native Protein

Step A: Refolding by Dilution

Twenty milliliters of the solubilized inclusion bodies (in 5.5 *M* guanidine hydrochloride, 25 m*M* Tris-HCl, pH 7.5) is reduced by addition of dithiothreitol (DTT) to 100 m*M* and EDTA to 10 m*M* final concentration. Most conveniently, this is done overnight at 4°, but reduction can also be performed for 4 hr at room temperature. As the DTT interferes with the formation of the disulfide bonds in the gene 3 N-terminal domains, the DTT concentration is reduced to 0.1 m*M* by three dialysis steps for 2 hr each at room temperature in 10 volumes (200 ml) of 5.5 *M* guanidine hydrochloride, 25 m*M* Tris-HCl, 10 m*M* EDTA, pH 6.0. The reduced, denatured protein is then slowly added dropwise to 1 liter of refolding buffer [0.4 *M* L-arginine, 0.2 *M* Tris-HCl, 0.2 *M* guanidine hydrochloride 0.1 *M* $(NH_4)_2SO_4$, 2 m*M* EDTA, pH 8.5] which contains 1 m*M* oxidized and 0.2 m*M* reduced glutathione. After overnight stirring at 10°, the refolding mixture is concentrated in an RA2000 concentrator (Amicon, Danvers, MA) to 120–150 ml, and further concentrated to a volume of 10–15 ml with a Centriprep YM10 (Amicon). Precipitates (mainly made up of contaminating proteins) are removed by centrifugation. Finally, the protein is dialyzed against 25 m*M* HEPES, 900 m*M* NaCl, pH 7.5, and filtered through a 0.45-μm pore size filter.

Step B: Coupled Immobilized Metal-Ion Affinity–Anion Exchange Chromatography under Native Conditions

To purify the refolded N-terminal domains of gene 3 protein, the protein is first bound to an IMAC column, using the C-terminal His_6 tag. After

[26] A. Plückthun, A. Krebber, C. Krebber, U. Horn, U. Knüpfer, R. Wenderoth, L. Nieba, K. Proba, and D. Riesenberg, *in* "Antibody Engineering: A Practical Approach" (J. McCafferty and H. R. Hoogenboom, eds.), p. 203. IRL Press, Oxford, 1996.

the protein is eluted directly onto the anion-exchange column (AIEX) under such conditions that it binds there. On the AIEX, the protein can be purified to a high degree by an NaCl gradient. The whole procedure is automated on a BioCAD 60 (PerSeptive Biosystems): First, the 1.66-ml HQ/M AIEX column (PerSeptive Biosystems) is equilibrated with 16 ml of 50 m*M* Tris-HCl, pH 7.5. The switches are then set such that the column is taken out of the flow and the 1.66-ml MC/M IMAC column (PerSeptive Biosystems) is switched in-line, washed with 20 ml of water, loaded with 2 ml of 100 m*M* $NiCl_2$ through one of the water-washed sample lines, and subsequently washed with 20 ml of water. After Ni^{2+} charging the MC/M column, it is equilibrated with 16 ml of 25 m*M* HEPES, 900 m*M* NaCl, pH 7.5. Five milliliters of sample is loaded onto the MC/M column. It is then washed first with 12 ml of 25 m*M* HEPES, 150 m*M* NaCl, 1 m*M* imidazole, pH 7.5, and then with 24 ml of 25 m*M* HEPES, 12 m*M* imidazole, pH 7.5. After the wash steps, the HQ/M column is set in-line such that the flow is directed from the MC/M to the HQ/M column. To transfer the protein from the MC/M onto the HQ/M column it is directly eluted from the MC/M onto the HQ/M column by 12 ml of 120 m*M* imidazole, pH 7.5, which results in strong binding of the protein to the HQ/M column. The MC/M column is then set off-line and the HQ/M column is washed with 16 ml of 50 m*M* Tris-HCl, pH 7.5. A 20-ml gradient of 0 to 500 m*M* NaCl in 50 m*M* Tris-HCl, pH 7.5, is employed to elute the N-terminal domain protein. As part of the protein dimerizes in the refolding reaction due to the free cysteine in the tag (to be used for chemical coupling), two major gene 3 N-terminal protein domain peaks are observed. The monomer and dimer peak fractions are combined and DTT is added to 2 m*M* and EDTA to 10 m*M* final concentration, respectively. The protein solution is passed through a 0.22-μm pore size filter and stored at 4°. If necessary, the protein solution can be concentrated with a Centriprep YM10 (Amicon).

Route 2: Purification and Subsequent Refolding

Step A: Coupled Immobilized Metal-Ion Affinity–Anion Exchange Chromatography under Denaturing Conditions

The purification under denaturing conditions is basically the same as under native conditions. However, because the purification starts with freshly solubilized inclusion bodies, the reduction with DTT and the subsequent dialysis to remove it are not necessary. Oxidized protein is reduced on the column and only the monomeric species must be separated from contaminants. As the pure protein can be refolded at higher concentrations, the protein is refolded by dialysis and not by dilution, which further simplifies downstream processing.

First, the 8-ml HQ/M AIEX column (PerSeptive Biosystems) is equilibrated with 60 ml of 6 *M* urea, 50 m*M* Tris-HCl, pH 7.5. The column is switched out of the flow and the 4-ml MC/M IMAC column (PerSeptive Biosystems) is set in-line, washed with 40 ml of water, loaded with 5 ml of 100 m*M* $NiCl_2$ through one of the water-washed sample lines, and subsequently washed with 80 ml of water. After charging the MC/M column it is equilibrated with 30 ml of 8 *M* urea, 25 m*M* HEPES, 1.5 *M* NaCl, pH 7.5. Five milliliters of solubilized inclusion body sample is loaded onto the MC/M column. It is then washed first with 20 ml of 8 *M* urea, 25 m*M* HEPES, 1.5 *M* NaCl, 4 m*M* imidazole, pH 7.5, and then with 60 ml of 8 *M* urea, 25 m*M* HEPES, 15 m*M* imidazole, pH 7.5. After the wash step the HQ/M column is set in-line such that the flow would be directed from the MC/M to the HQ/M column. To transfer the protein from the MC/M onto the HQ/M column it is eluted from the MC/M onto the HQ/M column by 40 ml of 8 *M* urea, 100 m*M* imidazole, pH 7.5, which results in strong binding of the protein to the HQ/M column. The MC/M column is then set off-line and the HQ/M column is washed, and at the same time the protein is reduced on the column with 20 ml of 8 *M* urea, 100 m*M* Tris-HCl, 100 m*M* DTT, pH 9.0, at a low flow rate of 2 ml/min (10 min). Subsequently, the column is washed with 60 ml of 6 *M* urea, 25 m*M* Tris-HCl, pH 7.5. The reduced, denatured protein is eluted by a 60-ml gradient from 0 to 250 m*M* NaCl in 6 *M* urea, 50 m*M* Tris-HCl, pH 7.5.

Step B: Refolding by Dialysis

The purified, reduced and denatured protein (10–20 ml) is immediately dialyzed against 500 ml refolding buffer [0.4 *M* L-arginine, 0.2 *M* Tris-HCl, 0.2 *M* guanidine hydrochloride, 0.1 *M* $(NH_4)_2SO_4$, 2 m*M* EDTA, pH 8.5], which contains 1 m*M* oxidized and 0.2 m*M* reduced glutathione, overnight at 10°. The buffer is exchanged by two dialysis steps, each against 500 ml of 50 m*M* Tris-HCl, 50 m*M,* NaCl, pH 7.5. DTT is added to 2 m*M* and EDTA to 10 m*M* final concentration, respectively. The protein solution is passed through a 0.22-μm pore size filter and stored at 4°, and if necessary concentrated on a Centriprep YM10 (Amicon). However, in case of direct peptide or protein fusions to the C terminus of N1 or N1–N2, the Cys residue in the C-terminal tag is replaced by the interaction partner B, and the refolding and purification procedure might have to be varied accordingly.

Chemical Coupling of Nonpeptidic Ligands

Coupling is carried out via the free sulfhydryl group, introduced by the Cys residue in the C-terminal tag (Fig. 3). Coupling chemistry and

purification may vary depending on the bait. We will give an example for coupling fluorescein,[9] which is already coupled to a Lys residue to give the compound FluCad (5-[(5-aminopentyl)thioureidyl]fluorescein). Coupling is achieved by using the heterobifunctional cross-linker *N*-succinimidyl-6-maleimidocaproate (Fluka, Ronkonkoma, NY). The coupling of FluCad to the cross-linker is carried out in dimethylformamide (DMF) for 1 hr at 30° in the dark, at a ratio of FluCad to cross-linker of 3:2. Completion of the reaction is controlled by thin-layer chromatography in 80% (v/v) ethyl acetate–10% (v/v) methanol–10% (v/v) acetic acid. At the same time the DTT in the N1 or N1–N2 stock is removed by gel filtration on a short Sephadex G-25 column (Pharmacia) in 25 m*M* sodium phosphate, pH 6.8. This leads to N1 or N1–N2 molecules that contain a free sulfhydryl group in the cysteine of the tag. The amount of protein is quantified by spectrophotometry, using the calculated extinction coefficient.[27] Three molar equivalents of FluCad are then reacted with one molar equivalent of the free sulfhydryl group of N1 or N1–N2 in 25 m*M* sodium phosphate, pH 6.8, for 1 hr at 25° and then at 4° overnight. The pH of the reaction buffer is critical, as at higher pH the maleimide would start to react with primary amines in the N1 or N1–N2 domain. The resulting adapter molecules are gel filtrated on a Sephadex G-25 column (Pharmacia) in 50 m*M* Tris-HCl, pH 7.5. Because fluorescein tends to noncovalently interact with the adapter protein (N1 or N1–N2), another purification step is needed. Therefore, the fluorescein-coupled adapter molecules are separated from the uncoupled N1 or N1–N2 by anion-exchange chromatography on a perfusion chromatography HQ column NaCl gradient (0 to 600 m*M* NaCl in 50 m*M* Tris-HCl, pH 7.5), using a BioCAD 60 system (PerSeptive Biosystems). Mass spectrometry should be used to verify the success of the coupling reaction and purification. In general, the coupling method and the heterobifunctional linker can be used for the coupling of any primary amine compound that does not contain a free sulfhydryl group. Using an analogous strategy every compound activated by a maleimide should be linkable to the free cysteine containing N1 or N1–N2. The disulfide bonds of N1 or N1–N2 were not affected, neither by the DTT nor the maleimide under the conditions employed.

Phage Infection Experiment

In vitro SIP experiments are performed by incubating 1 μl of SIP-phage supernatant, concentrated 500 times by two PEG precipitations (as described under General Procedures), with 350 n*M* adapter N1–N2 bait

[27] S. C. Gill and P. H. von Hippel, *Anal. Biochem.* **182,** 319 (1989).

at 4°. For infection, 100 μl of an exponentially growing XL1-Blue culture is added to the mixture and incubated for 1 hr at 37° with shaking, and plated subsequently on 2× YT–agar plates containing Cam (25–30 μg/ml) and Tet (15 μg/ml). Addition of 50 m*M* $MgCl_2$ to the cells prior to and during infection increases infectivity four- to sixfold.[9] The number of phages can be varied according to library size and expected infection rate. The concentration of adapter molecule, especially for N1–N2 bait, can influence the infectivity, and thus, when maximal infectivity is needed, a titration of adapter molecules might be advisable (see the next section).

Dependence of Infectivity on Concentration of Adapter Molecule

The concentration of adapter molecules influences the phage infectivity. Because of the law of mass action, higher concentrations of adapter will shift the equilibrium to the bound and thus infective state. However, differences between N1 bait and N1–N2 bait are apparent, presumably because of the ability of N2 to bind to the pili and thus block infection at high concentration. Both variants have previously been tested with fluorescein as bait, fused to N1 or N1–N2, respectively, and a fluorescein-binding single-chain variable fragment (scFv), fused to N2–CT ("medium" phage) or CT ("short" phage), respectively (Fig. 4).[9] The infectivity of 10^{10} medium

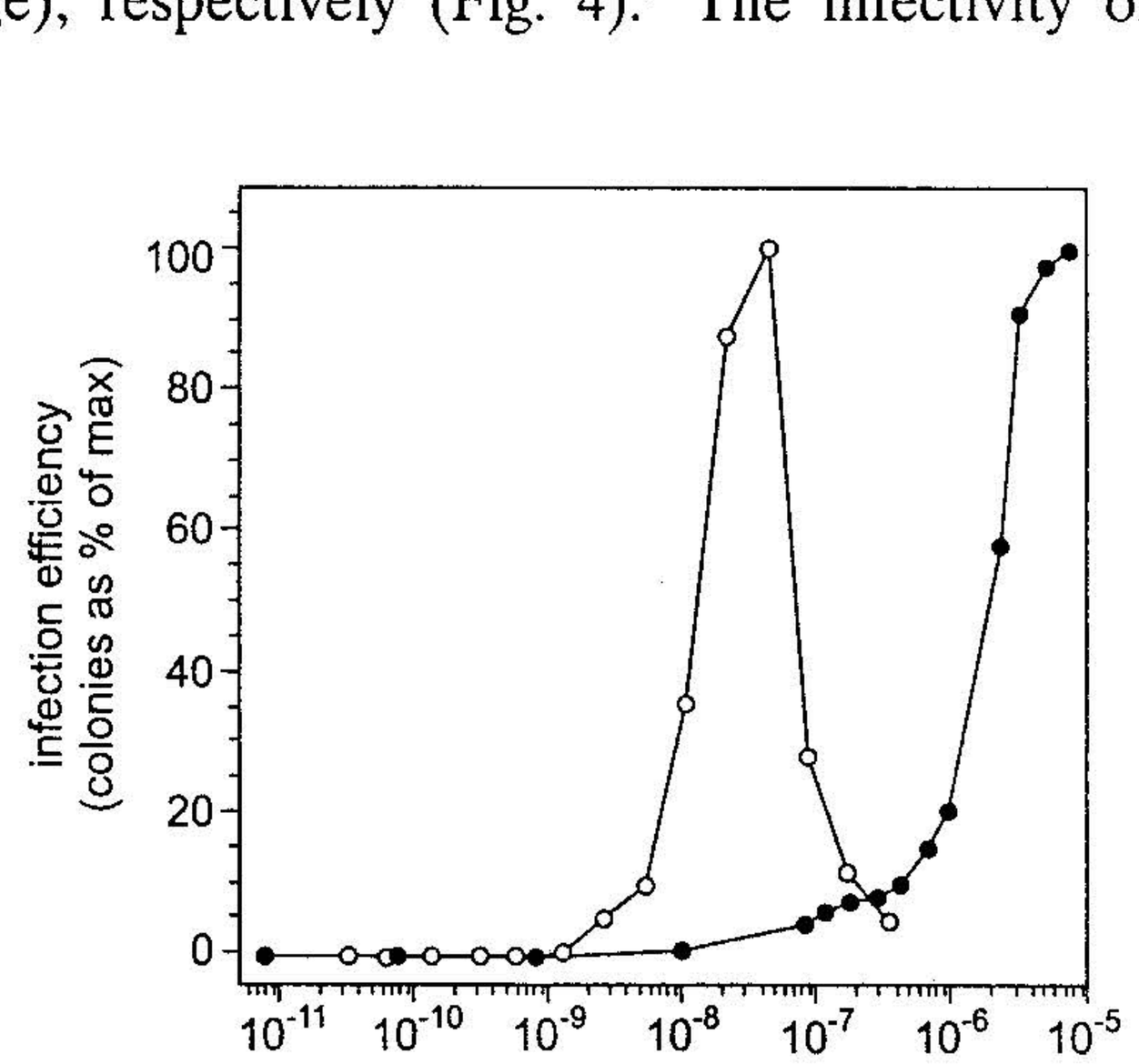


FIG. 4. Dependence of infectivity on concentration of adapter molecules and g3p arrangement in *in vitro* SIP. Open circles represent the situation depicted in Fig. 1C with the "short" phage and N1–N2 bait as adapter molecule. In this case, the same infectivity was observed when using "medium" phages instead (data not shown). Filled circles represent the infectivities obtained with medium phage and N1 bait as adapter molecules (Fig. 1D). [Adapted from Krebber *et al.*[9]]

phages increases as a function of N1–fluorescein up to the highest concentration tested (10^{-5} *M*), at which point it levels off, yielding 1.2×10^6 colonies. In contrast, using the adapter N1–N2–fluorescein with short or medium phages, a maximum of infectivity is reached at 3×10^{-8} *M*, yielding 5×10^4 colonies from 10^{10} input phages. In summary, using N1 bait as adapter and medium phages, the maximal infectivity is about 20-fold higher than with the N1–N2 bait, but a higher adapter concentration is needed. Consequently, the N1–N2 bait is recommended for selection of high-affinity systems, as the adapter concentrations can be kept low, whereas the N1 bait is more suitable for lower affinities, as the adapter concentration can be increased without inhibitory effects.

Examples from the Literature using in Vitro Selectively Infective Phage Selection

In vitro SIP was used to select for novel, nonrepetitive linkers for antibody scFv fragments.[28] In this case, medium phages and adapters with N1 bait were used, and phage production was performed at 30 and 37°. After one round of SIP selection, 22 of 22 clones gave positive signals in ELISA, whereas before selection, only 1 of 23 clones was positive. Nine clones were characterized further, and shown to be comparable to the parental sequence, which had $(Gly_4Ser)_3$ as the linker, in terms of binding, folding, expression, and solubility.

A selection from an antibody Fab library was carried out and compared with selection results from conventional phage display.[4] In this case, a phagemid system was used, and the adapter, consisting of the first 98 amino acids of the mature gene 3 protein fused to the antigen, was secreted from a separate *E. coli* culture, and the supernatant was used directly for infection. The N2 domain was missing completely in this experimental setup. Infection was carried out by adding 25% (v/v) adapter-containing supernatant to the supernatant from the phage-producing cells and incubating the mixture overnight at 4° before adding XL1-Blue recipient cells. From this library, half the phages gave positive ELISA signals after three rounds of SIP selection, while no antibodies against the antigen could be obtained from phage display after three rounds of phage panning. Thus, SIP is faster than conventional phage display, using the identical library.

In two further model systems, the properties of selection were investigated in more detail by using a small set of molecules with defined thermodynamic and kinetic parameters. The phage infectivity of various point mutants with slightly different K_D values and activities was compared and a

[28] F. Hennecke, C. Krebber, and A. Plückthun, *Protein Eng.* **11,** 405 (1998).

clear correlation was observed.[29] In these experiments, short phages and N1–N2 bait adapter molecules were used. In a competitive SIP experiment with all mutants, only the tightest binders remained after three rounds of selection. In another model selection with six different Fab fragments, the ability to select for affinity or even kinetic constants was investigated.[30] In individual experiments using the phagemid system (mentioned above), phages expressing higher affinity clones were enriched preferentially with low concentrations of antigen. However, it was also possible to select clones with lower affinity by increasing antigen concentration. In another experiment the incubation time was varied. Using short incubation times (30 min), clones with high association rate constants were preferred, whereas after long incubation (16 hr), clones with the lowest dissociation rate constant were preferred.

Two further examples use techniques closely related to *in vitro* SIP. In the first example, linkage of covalent catalysis to infectivity was demonstrated by fusing a catalytic antibody fragment to the CT domain and the substrate was coupled to N1–N2. Only covalent catalysis was able to produce infective phages.[31] In the second example, selection for protease resistance was carried out by fusing the proteins of interest between N1–N2 and CT and performing several rounds of *in vitro* proteolysis, infection, and propagation.[32,33]

In Vivo Selectively Infective Phage Procedures

In the *in vivo* SIP procedure, all steps are carried out with crude *E. coli* supernatant, without the need for purification of any compound or any *in vitro* panning steps. In particular, the target for the library does not have to be expressed and purified. Both interacting partners (protein A–CT and N1–N2–protein B) are encoded on the same phage vector, as described below. Alternatively, a phagemid system[3,4] (see General Procedures) or a combination of phage and phagemid vectors that are copackaged[34] can be used. Although two-vector systems provide for more convenient cloning,

[29] G. Pedrazzi, F. Schwesinger, A. Honegger, C. Krebber, and A. Plückthun, *FEBS Lett.* **415,** 289 (1997).
[30] M. Dueñas, A. C. Malmborg, R. Casalvilla, M. Ohlin, and C. A. Borrebaeck, *Mol. Immunol.* **33,** 279 (1996).
[31] C. Gao, C. H. Lin, C. H. L. Lo, S. Mao, P. Wirsching, R. A. Lerner, and K. D. Janda, *Proc. Natl. Acad. Sci. U.S.A.* **94,** 11777 (1997).
[32] V. Sieber, A. Plückthun, and F. X. Schmid, *Nature Biotechnol.* **16,** 955 (1998).
[33] P. Kristensen and G. Winter, *Fold. Des.* **3,** 321 (1998).
[34] F. Rudert, C. Woltering, C. Frisch, C. Rottenberger, and L. L. Ilag, *FEBS Lett.* **440,** 135 (1998).

it is important to ensure the quality of both vectors and the complete transformation or transfection of one library to the other, which is usually verifiable only by sequencing cotransformants.

Cloning of Two Libraries in One Vector

If a library-versus-library experiment is to be carried out, two libraries must be cloned (library A–CT and N1–N2–library B). If the phage system is used, both libraries must be cloned in one vector with the highest efficiency possible. Two strategies are available. First, to avoid multiple transformations, both libraries can be combined with the interconnecting vector fragment in an assembly PCR, and then cloned in one step into the vector. Thereby, only one transformation is required. Alternatively, both libraries can be cloned one after the other, which then requires two ligations and two transformations, however. We tested both possibilities by sequencing individual clones obtained after cloning. Using the assembly PCR approach, fewer clones analyzed were generally correct, and the remainder were mainly frameshift mutants with deletions of 1 base pair up to a stretch of several base pairs. Also in the region between both libraries, deletions were occasionally observed. However, using the second approach with two subsequent ligations and transformations, the result was significantly better. It appears that the assembly PCR, even though performed with a proofreading polymerase, introduces more errors, and because the overall transformation rates are comparable for both approaches, the sequential cloning appears to be the method of choice for library cloning in one-vector systems.

Design of Vectors Suitable for in Vivo Selectively Infective Phage to Avoid Genetic Instabilities

In the case of *in vivo* SIP, where both parts of the gene 3 protein are expressed in the same cell, whether on one or two vectors, great care must be taken to prevent recombination. In initial constructs with the arrangement N1–N2–protein B and protein A–CT, homologous recombination between duplicated glycine-rich regions was observed, leading to restoration of a wild type-like gene 3 protein at high frequency. Therefore, in more recent vectors[7] the order of the two fusion proteins has been switched to protein A–CT being upstream of N1–N2–protein B, and the glycine-rich linkers between protein A–CT and N2–protein B were shortened such that there were no long identical sequence stretches present in both fusion proteins. By this measure, recombination could be prevented successfully in most cases.

However, the occurrence of recombination of course still depends on the nature of the two interacting proteins or peptides A and B. In the case

that both parts have similar sequence stretches, either in constant parts or in a randomized region, the risk of recombination increases significantly. In an example of a library-versus-library experiment using two semirandomized libraries, recombinations were found only in clones in which an identical stretch of 8 consecutive base pairs (bp) happened to exist in libraries A and B. In these cases, the sequence stretch between those 8 bp, which served as recombination sites, was doubled, which led to a wild type-like arrangement of gene 3. Such clones can therefore easily be recognized by wild type-like infectivity.

However, such recombined clones can be eliminated in a simple and fast way. Phage DNA of the entire library after the appropriate SIP round, where recombined clones are found, is prepared, and the gene 3 cassette is excised by restriction digest (in the vector described here, *Pin*AI/*Hin*dIII; Fig. 2E). The correctly sized band of the gene 3 cassette (1963 bp in our example) can then be easily separated by gel electrophoresis from the larger band created by the recombination (3379 bp). The purified, correctly sized band is cloned back into fresh vector and then used for further SIP selection. With this approach, another round of SIP selection yields only clones with correctly sized gene 3 cassette, as judged by analytical restriction digest, and significant recombination usually does not occur for another two rounds. It is advisable to perform an analytical restriction digest as a control for recombination from the entire library or single clones obtained after selection.

Selective Infection Experiments

For infection, Tet^R or Kan^R XL1-Blue cells are grown in 2× YT medium, supplemented with 1% (w/v) glucose, 50 m*M* $MgCl_2$, and Tet (5 μg/ml) or Kan (50 μg/ml), respectively, to an OD_{550} of 0.5 to 0.8. It has been shown previously that the addition of 50 m*M* $MgCl_2$ increases the infectivity four- to sixfold.[9] The number of phages used for infection should be on the order of 10^8 to 10^{11} phages, and is varied depending on the size of the libraries and the expected infection rate. The appropriate amount of phages is added to 0.5 ml of cells and shaken for 1 hr at 37°. Bacteria are plated on 2× YT–agar containing 1% (w/v) glucose, Cam (30 μg/ml), and Tet (5 μg/ml) or Kan (50 μg/ml), depending on the host strain. For library experiments 245 × 245 mm plates are usually most appropriate. The infectivity is determined by plating a series of 10-fold dilutions on small plates. Plates are incubated overnight at 37 or 25° for more unstable proteins. With this step, one round of SIP selection is finished. To proceed with the library for further selection rounds, the cells are pooled from the plates into 5–10 ml of rich medium, shaken for 5 min at 37° to allow for good

mixing and cell separation, and used for inoculation of a new overnight culture. At the same time, glycerol stocks are prepared as a backup (stored at −80°), by adding glycerol to a final concentration of 30% and making aliquots of 500–1000 μl. Individual clones can be analyzed by restriction digest or sequencing of minipreparation DNA.

Occurence of Cysteine Mutations and Effects of Dithiothreitol

Libraries created by using synthetic oligonucleotides, even those of high quality, almost always have a low percentage of frameshift mutants due to imperfect oligonucleotide synthesis. Similarly, cysteines are introduced by random mutagenesis techniques such as DNA shuffling or error-prone PCR[12,13,35] and will inevitably be present in cDNA libraries. Most of the frameshift mutants will not be selected, because they lack the ability to fold and bind their partner, but some might still show up in selection: Provided a cysteine is introduced both at the end of the adapter and the beginning of the CT domain, a covalently linked "wild type-like" gene 3 protein is formed. In the following section, we highlight this problem and point out general solutions.

In an example of a library-versus-library selection (library A–CT and N1–N2–library B) for noncovalently interacting pairs, strong selection for a single cysteine in each library peptide was observed, even though the designed sequences contained no cysteines at any place. The cysteines were found to be caused by either point mutations (in library A) or by frameshifts (in library B). The cysteine pairs caused a strong increase in infectivity of four orders of magnitude in one SIP round. The same phenomenon was observed in a one-library approach (library A–CT and N1–N2–peptide B), in which also a point mutation in library A and a frameshift in peptide B led to single cysteines (for details see Ref. 36). These covalent interactions are mostly unspecific but nonetheless provide a strong selective advantage. Consequently, the high infectivity of phages in the absence of DTT was significantly reduced after DTT treatment. Control experiments showed that incubation with DTT has only a minor effect on phage infectivity itself, consistent with previous experiments, where it was reported that only phage production but not phage infection and phage DNA replication is prevented by 5 m*M* DTT.[37] Furthermore, all four disulfide bridges in the native g3p

[35] D. W. Leung, E. Chen, and D. V. Goeddel, *Technique* **1,** 11 (1989).
[36] S. Jung, K. M. Arndt, K. M. Müller, and A. Plückthun, *J. Immunol. Methods* **231,** 93 (1999).
[37] M. Vaccaro, B. Boehler-Kohler, W. Müller, and I. Rasched, *Biochim. Biophys. Acta* **923,** 29 (1987).

are inaccessible to the alkylating agent vinylpyridine after treatment with DTT (50°, 50 min, 100-fold molar excess of DTT over cysteines).[38]

This stability of phages against reducing agents can be extremely useful when working with complex libraries, where the occurrence of single cysteines cannot be excluded. In our experience, the purified wild-type phage tolerates a 3- to 5-hr incubation at 37° with 5 m*M* DTT, pH 8.0, prior to infection very well, with only a slight loss of infectivity. We usually dilute the phages after DTT incubation and prior to infection to obtain a final DTT concentration of about 10 μM during infection, but the infection worked equally well at a 100-fold higher DTT concentration (1 m*M*). However, the exact amount of DTT, the incubation time, and temperature depend not only on the stability of the phage but also on those of the proteins of interest. Nonetheless, DTT incubation can significantly reduce but not fully eliminate the problem as some disulfide bridges might survive the DTT treatment or reform during the infection process.

Detection of Translational Frameshifts by the Appearance of Polyphages

In a further SIP experiment using the same library-versus-library approach previously mentioned, two clones were selected with a −1 frameshift in the library peptide before the CT domain. Extensive studies revealed that neither the frameshift peptide, which would be produced instead of the CT–library A fusion, nor the peptide from library B fused to N1–N2, could replace the function of CT. Most likely the observed −1 frameshift is accompanied by a second, translational +1 frameshift caused by two rare arginine codons (AGG) that were present in both clones due to the −1 frameshift.[39] This was further confirmed by the observation that these clones formed polyphages, most likely due to insufficient capping of the nascent phages caused by the limited amounts of CT domain, created by a translational +1 frameshift (for details refer to Ref. 36).

It can therefore be concluded that the CT domain, and probably the transmembrane helix, is not only needed for infection, but can definitely not be replaced by "sticky peptides," which would nonspecifically attach N1–N2 to the phage. Therefore, the SIP procedure must clearly bring together N1–N2 and CT on the phage.

Examples from the Literature using in Vivo Selectively Infective Phage Selection

In an example of *in vivo* SIP selection, a conserved region in the immunoglobulin variable domain was investigated by using a small synthetic

[38] A. Kremser and I. Rasched, *Biochemistry* **33,** 13954 (1994).
[39] R. A. Spanjaard and J. van Duin, *Proc. Natl. Acad. Sci. U.S.A.* **85,** 7967 (1988).

library.[40] The library was fused to the CT domain, and the peptidic antigen was fused to N1–N2. After only three rounds of SIP, a strong enrichment was found of clones similar to the most abundant sequences in the Kabat database. In another example, the coiled-coil domain of c-Jun with an engineered free cysteine was fused to the CT domain of gene 3 protein and used as bait to select against a human cDNA library fused to the soluble N-terminal domains.[3,41] Several clones were isolated that were predicted to encode potential coiled-coil structures. However, in our hands, constructs with free cysteines, even though they show higher infectivity, showed less specificity in their interaction. In addition, in the one-library approach we prefer to fuse the library to the CT domain rather than to N1–N2. To keep phenotype and genotype connected, it is advantageous to have the library tightly connected to the phage, in case the ligand dissociates and exchanges to a different phage during the infection in the library pool.

A further model experiment showed the feasibility of a two-vector packaging of adapter and CT fusion, which would be useful for library-versus-library selection.[34] In this case, protein A–CT was displayed on phage and encoded on the phage genome and the N1–N2–bait was encoded on a phagemid. After cotransformation both vectors were copackaged in polyphages, which were screened on special filter plates.

Discussion

SIP is a powerful strategy to select for protein–ligand interactions as well as for other desired features such as protein folding and stability, provided the binding function of the protein is limited to the native state. The selection is carried out completely in solution without any need for solid-phase panning steps. Furthermore, the enrichment cycle is fast and thus less time consuming. Most importantly, the selection power is extremely high and enrichment factors of 10^5 to 10^6 have been observed,[42] while in phage display enrichment factors of 10 to 10^3 are normal.[43] Therefore, in most applications SIP requires only one selection round to separate binders from nonbinders,[28,29] whereas phage display usually needs three or four rounds. In three or four SIP rounds, it is moreover possible even to discriminate between more subtle affinity, folding, and stability differences.[29,40]

[40] S. Spada, A. Honegger, and A. Plückthun, *J. Mol. Biol.* **283,** 395 (1998).
[41] K. Gramatikoff, W. Schaffner, and O. Georgiev, *Biol. Chem.* **376,** 321 (1995).
[42] M. Dueñas, L. T. Chin, A. C. Malmborg, R. Casalvilla, M. Ohlin, and C. A. Borrebaeck, *Immunology* **89,** 1 (1996).
[43] G. Winter, A. D. Griffith, R. E. Hawkins, and H. R. Hoogenboom, *Annu. Rev. Immunol.* **12,** 433 (1994).

However, be aware that this extremely strong selection ability can sometimes lead to unwanted solutions, by covalently linking the N terminal to the C-terminal part of g3p. This is only a problem in the *in vivo* SIP procedure, where both partners can potentially acquire mutations, because in *in vitro* SIP the infection-mediating particle is produced separately. With highly complex libraries, products of random mutagenesis, and cDNA libraries, a selection for disulfide bridges and enrichment of recombined clones is possible. Genetic recombination is a rare event, easy to check and to remedy by cutting out and recloning restriction fragments. The selection for disulfide bonds from spurious cysteines can be most effectively controlled by working with high-quality libraries in the beginning. Reducing agents also help, but do not eliminate the problem. Currently, the work with cDNA is still a challenge.

For a library-versus-library selection, only *in vivo* SIP is applicable, but needs to be strictly controlled. It also remains to be determined what the maximum size of the fusion protein can be, whether there are geometric restrictions imposed by the infection process, and what the minimum affinity for the interaction is. In an *in vivo* SIP library-versus-library approach, a weak interaction might lead to the exchange of adapter molecules between different clones and thus destroy the genotype–phenotype correlation. Another limitation for complex libraries is the moderate infection efficiency. The highest observed was 1 infection per 10^4 SIP phages,[9] but much lower infectivities have also been reported.

Taking everything together, SIP is at this point most useful in the fast screening of less complex libraries and in molecular improvement. We generally recommend the use of the *in vitro* SIP procedure over *in vivo* SIP whenever possible, because it is more robust and easier to control. Furthermore, by choosing the right adapter and titrating its concentration, it allows for selection of medium as well as high-affinity interactions. Furthermore, the *in vivo* SIP complements very well the other two techniques available to directly select two libraries simultaneous against each other, the yeast two-hybrid system[44] and the protein fragment complementation assay,[45] which select in the nucleus of yeast or the cytosol of *E. coli*, respectively. In *in vivo* SIP proteins are folded in the oxidative environment of the periplasm, which is essential for selection from libraries containing proteins that require the formation of disulfide bridges.

[44] S. Fields and O. Song, *Nature* (*London*) **340,** 245 (1989).

[45] J. N. Pelletier, K. M. Arndt, A. Plückthun, and S. W. Michnick, *Nature Biotechnol.* **17,** 683 (1999).