Biochemical and Biophysical Research Communications **266,** 243–247 (1999)
Article ID bbrc.1999.1806, available online at http://www.idealibrary.com on IDEAL®

A Genetic Screen to Identify Sequences That Mediate Protein Oligomerization in *Escherichia coli*

Roberto Jappelli[1] and Sydney Brenner
Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, California 94704

Received October 10, 1999

Many proteins assemble as oligomeric complexes and in several cases a distinct domain mediates the interaction between the subunits. The identification of new oligomerization modules is relevant to comprehend both the architecture and the evolution of protein sequences and also for protein engineering applications. Using the bacteriophage λ repressor dimerization assay, we searched *Escherichia coli* genomic libraries for sequences able to mediate protein oligomerization *in vivo*. We identified short peptides that can substitute very effectively the dimerizing domain of the repressor. Most of these peptides belong to open reading frames that are normally not expressed in the bacterial cell.

An increasing number of proteins are known to consist of stable assemblies of either identical or different subunits. A protein quaternary state confers features that are not present at monomeric level such as greater stability, formation of new binding sites, substrate channeling, cooperativity and allosteric regulation. In addition, the presence of multiple subunits reduces synthetic errors, facilitates the evolution of new proteins and reduces the amount of genetic information required by an organism (1, 2).

The three-dimensional structure of both homomeric and heteromeric protein complexes has been the subject of comprehensive analytical surveys aiming to define the principles that govern protein-protein interactions (3–5). An automatic procedure to determine the probable quaternary state for crystal structures contained in the Brookhaven Protein Data Bank is now available as a file server (6). The development of several methods for the prediction of subunit interfaces and protein binding sites have paralleled these analyses. The rational *de novo* design of proteins displaying specific quaternary states, particularly α-helical coiled coils and bundles, represents another current approach to understand the determinants of protein association (7).

In this work we have tried to identify new protein domains able to mediate protein oligomerization *in vivo*. To this end we have turned to a random library approach combined with a genetic screen which utilizes the cI repressor of bacteriophage λ. We have found that some relatively short peptides can substitute very efficiently the dimerizing domain of the repressor. The majority of these peptides correspond to fragments of open reading frames that are normally not expressed by the bacterial cell. Recently, a method similar to the one presented here has been used to search a yeast genomic library for peptides that support protein oligomerization (8).

The dimerization properties of the λ repressor have been used for an extensive analysis of the sequence requirements of GCN4 leucine zipper (9). Since then, with some variations, the repressor assay has been incorporated into several experimental designs to detect protein-protein interaction *in vivo*. The protein consists of an amino-terminal domain (1–92), which binds DNA, and a carboxy-terminal domain (93–236), which mediates dimerization (10). The repressor binds to DNA as a dimer, but the DNA binding domain alone dimerizes inefficiently and requires the carboxy-terminal domain to be functional. Heterologous domains able to dimerize can functionally substitute the carboxy-terminal domain. The wild-type repressor blocks the transcription of the genes responsible for the phage lytic cycle so that a very low level of the molecule confers immunity to λ phage (11). Production of a repressor fusion protein confers a level of phage immunity that depends upon the dimerization constant of the particular domain attached to the repressor.

MATERIALS AND METHODS

E. coli genomic DNA extraction, recombinant DNA procedures and DNA sequencing were performed according to standard methods (12). The first genomic library was constructed by introducing *E. coli Nla*III genomic fragments into the vector pAC117 linearized with *Sph*I. Clones R1, R2, R3 and R6 derived from this library. The next library was

[1] To whom correspondence should be addressed. Fax: (858) 792-8712. E-mail: rjappelli@lynxgen.com.

0006-291X/99 $30.00

TABLE 1

Sequence and Activity of the λ Repressor Fusion Proteins

Clone	Amino acid sequence[a]	Notes	λ[b]	HLPD[c]	β-gal[d]
R2	[LEA] CSPSGSICESSDLQVLTRFLEGE-"**IBPB**"(M1-S142)	Entire protein	R	10^{-1}	0.5
R3	[LEA] "**AAC70143**"(C45-G184)	Hyp ORF, C-terminal fragment	R	10^{-1}	0.7
R6	[LEA] CLRRIYSS-"**YIGA**"(M1-M58)AYESGK	Hyp ORF, N-terminal fragment	R	10^{-1}	0.6
R1	[LEA] CVCCRKRISGSSVRQARRIIHSIRAK	Out of frame in YJFH sequence	R	10^{-1}	0.9
R25	[LEG] STCRITVKLCVSTAQCRKAGKKNGA	Out of frame in MUTL sequence	R	10^{-1}	1.2
R37	[LEG] SRVTTLKKNWRKKV	E. coli non coding region	R	10^{-1}	5.5
R45	[LEG] SYCRKRSRGRCCYPNGTEVPCQSRRAHRWDVPRR	Out of frame in GIDA sequence	R	10^{-1}	1.5
R51	[LEG] SISFCGGITTVVLNLRGR	Non coding strand in CYSI sequence	R	10^{-1}	1.0
R54	[LEG] SCRRRRQKRKG	Out of frame in YAGF sequence	R	10^{-1}	5.1
pAC117	[LEAC]	Negative control	S	10^{-7}	945.8
pAC130	[LEGSAC]	Negative control	S	10^{-7}	1164.4
pAC115	β-lactamase	Negative control	S	10^{-7}	490.6
pAC116	Glutathione S-methyl transferase	Positive control	R	10^{-1}	4.0
pAC120	cAMP dependent protein kinase regulatory subunit	Positive control	R	10^{-1}	~0
pMTC3	Wild-type repressor	Positive control	R	10^{-1}	~0

[a] Peptide sequences found in frame with the repressor DNA binding domain.
[b] R: λ resistant. S: λ sensitive.
[c] HLPD: Highest Lysing Phage Dilutions that produces a visible lysis of the bacterial lawn.
[d] β-gal: β-galactosidase activity.

generated inserting *Sau*3A1 fragments into the plasmid pAC130 digested with *Bam*HI. Clones R25, R37, R45, R51, and R54 derived from this second library. Before the ligation reaction the two vectors were dephosphorylated. Plasmid libraries were transformed in electrocompetent bacterial strain Q537, (F^- *mcr*A *mcr*B r_K^- m_K^+ *i lac am*U281 *arg*Eam *gal*$^-$ *rif nal sup*^{o}) and plated on LB plates containing chloramphenicol (50 μg/ml) and 5-bromo-4-chloro-3-indolyl β-D-galactoside (40 μg/ml). After overnight incubation at 37°C and a few days of further incubation at 4°C the plates were inspected for the presence of white colonies. Plasmid DNA was isolated from white colony clones.

Freshly transformed bacteria were tested for both repressor and β-galactosidase activity. Repressor activity was quantitated with λ2001, which carries the deletion KH54 in the *cI* gene that prevents lysogenization. Ten-fold serial dilutions of a 10^{10} pfu/ml phage stock containing from 5 to 5(10^7) pfu/ml were spotted on bacterial lawns poured on λ plates containing chloramphenicol (50 μg/ml). Highest lysing phage dilution (HLPD) is the highest phage dilution that produces visible bacterial lysis. β-galactosidase activity was assayed according to the method of Miller (13).

The most important features of the expression vector are explained in Fig. 1. The fusion protein produced by each vector is reported in Tables 1 and 2. All the vectors used in this work are derivative of pACYC184 (14). Plasmid pMTC3 encodes the wild-type repressor and has been constructed before (15). Plasmid pAC115 and pAC120 encode a repressor fusion protein similar to the vectors pMTC9 and pMTC111. Both pMTC9 and pMTC111 have been described before (15). They include the sequence for the first 155 amino acid of the repressor. In contrast, the new vectors pAC115 and pAC120 encode only the 102 amino-terminal residues. In addition the construct pAC120 differs from the previous vector pMTC111 as the sequence for the cAMP dependent kinase regulatory subunit corresponds to the full-length gene. pAC115 was obtained from the previously described pAC55 (16), replacing the fragment *Sac*I-*Xho*I with the fragment *Sac*I-*Xho*I of pMTC9 (15). pAC120 derives from pAC115 replacing the *Kpn*I-*Eag*I fragment with the *Kpn*I-*Eag*I fragment of pMTC16 (15).

pAC116 derives from pAC55 (16), replacing the *Sac*I-*Sal*I fragment with the PCR product amplified with oligonucleotides R84 and R86 and restricted with *Sac*I-*Sal*I. The vector pGEX (Amersham Pharmacia Biotech) was used as template for the amplification of the glutathione S-transferase sequence. pAC117 derives from pAC116 eliminating the *Xho*I-*Xho*I fragment. pAC130 derives from pAC117

TABLE 2

Sequence and Activity of the Deletion Mutants Derived from the Clones R6 and R51

Clone	Amino acid sequence	λ	HLPD	β-gal
R6	[LEA] CLRRIYSS-"**YIGA**"(M1-M58)AYESGK	R	10^{-1}	0.6
pAC133	[LEA] CLRRIYSS-"**YIGA**"(M1-M58)	R	10^{-1}	~0
pAC134	[LEA] CLRRIYSS-"**YIGA**"(M1-H45)	S	10^{-7}	1191.3
pAC132	[LEA] CLRRIYSS-"**YIGA**"(M1-R42)	S	10^{-7}	1230.2
pAC135	[LEA] "**YIGA**"(Q3-M58)	S	10^{-7}	22.2
R51	[LEG] SISFCGGITTVVLNLRGR	R	10^{-1}	1.0
pAC137	[LEG] SISFCGGIITVVLNLR	R	10^{-1}	2.6
pAC140	[LEG] SISFCGGIITVVLNL	S	10^{-7}	638.0
pAC138	[LEG] SISFCGGITTVVLN	S	10^{-7}	1935
pAC141	[LEG] SITTVVLNLR	S	10^{-7}	788.0

Note. For explanations see legend to Table 1.

in two steps. First, a single *BamH*I site was eliminated from pAC117. Next, the fragment *Sac*I-*Sph*I was replaced by a cassette provided by oligonucleotides R97 and R98 which also contains a new *BamH*I site.

The vectors pAC132, pAC133, pAC134 and pAC135 were all constructed from the vector pAC117 replacing the *Sac*I-*Sph*I fragment with PCR fragments amplified from the clone R6 and restricted with *Sac*I-*Sph*I using the primers R103-R104, R103-R105, R103-R106, and R107-R105, respectively.

The vectors pAC137, pAC138, pAC140 and pAC141 all derive from the plasmid pAC130 replacing the fragment *Xho*I-*BamH*I with cassettes provided by the oligonucleotides R113-R114, R115-R116, R119-R120, and R121-R122, respectively.

The structure of all the deletion mutants was confirmed by DNA sequencing.

Oligonucleotide sequences: R84 (5′-GCAGAGCTCGAGTCCCCTATACTAGGTTATT-3′). R86 (5′-CCCGTCGACCTAGCATGCCTCGAGTTTTGGAGGATGGTCGCCAC-3′). R97 (5′-CGAGGGATCCGCATG-3′). R98 (5′-CGGATCCCTCGAGCT-3′). R103 (5′-CGCGAGCTCTGTCTACGACGGATTTATTC-3′). R104 (5′-CGCGCATGCTTAACGTATCGCTTCTACTGCGC-3′). R105 (5′-CGCGCATGCTTACATGTGCCACTCGACCAACG-3′). R106 (5′-CGCGCATGCTTAATGCGGCACACGTATCGCTT-3′). R107 (5′-CATGAGCTCCAACCAGGGGAAGAACTGCA-3′). R113 (5′-TCGAGGGATCGATATCGTTCTGTGGCGGGATCACTACCGTGGTTTTGAATTTACGCTAAG-3′). R114 (5′-GATCCTTAGCGTAAATTCAAAACCACGGTAGTGATCCCGCCACAGAACGATATCGATCCC-3′). R115 (5′-TCGAGGGATCGATATCGTTCTGTGGCGGGATCACTACCGTGGTTTTGAATTAAG-3′). R116 (5′-GATCCTTAATTCAAAACCACGGTAGTGATCCCGCCACAGAACGATATCGATCCC-3′). R119 (5′-TCGAGGGATCGATATCGTTCTGTGGCGGGATCACTACCGTGGTTTTGAATTTATAAG-3′). R120 (5′-GATCCTTATAAATTCAAAACCACGGTAGTGATCCCGCCACAGAACGATATCGATCCC-3′). R121 (5′-TCGAGGGATCGATCACTACCGTGGTTTTGAATTTACGCTAAGATATCG-3′). R122 (5′-GATCCGATATCTTAGCGTAAATTCAAAACCACGGTAGTGATCGATCCC-3′).

RESULTS

We have used an expression system designed to keep a constant low level of fusion protein inside the bacterial cell (17). In this vector, a tightly repressed copy of the 434 phage P_L promoter drives the expression of the λ repressor. In addition, the vector encodes the sequence for an amber suppressor tRNA which is cloned downstream from the major leftward λ promoter (λP_L) and acts as a reporter system (15). Transcription of the tRNA depends on the extent of repression at λ O_L P_L. The expression of the suppressor tRNA is detected after transformation of the plasmid into a bacterial strain which carries an amber mutation in its chromosomal *lacZ* gene. As a result, bacteria expressing a functional repressor fusion protein are identified by a *lac*$^-$ phenotype (Fig. 1).

Two random peptide libraries were constructed by introducing *E. coli* (strain Q537) DNA fragments into a cloning site positioned at the 3′ end of the sequence encoding the repressor DNA binding domain (Ser1-Glu102). We decided to use the genome of the bacterium *E. coli* because its DNA has been entirely sequenced (18).

Plasmid libraries were transformed in the detector strain Q537. Plasmid DNA was prepared from white colonies (40 from the first library and 76 from the second library) and retransformed in strain Q537 to make sure that the bacterial phage resistance phenotype was indeed determined by the presence of the plasmid. The freshly transformed putative resistant clones were tested for the level of β-galactosidase activity. Repressor activity was scored by a spot test applying phage dilutions on bacterial lawns derived from the individual clones. Those that conferred only an intermediate level of phage resistance were not further characterized. Highly resistant clones were found at a frequency of 1 in 2400. After DNA sequencing a total of 9 highly resistant clones were isolated from the two libraries.

The amino acid sequences that were found to be in frame with the repressor DNA binding domain in each of these clones are listed in Table 1. Residues in square brackets derive from the vector cloning site. The functional properties displayed by these clones were similar to those conferred by the wild-type repressor and by

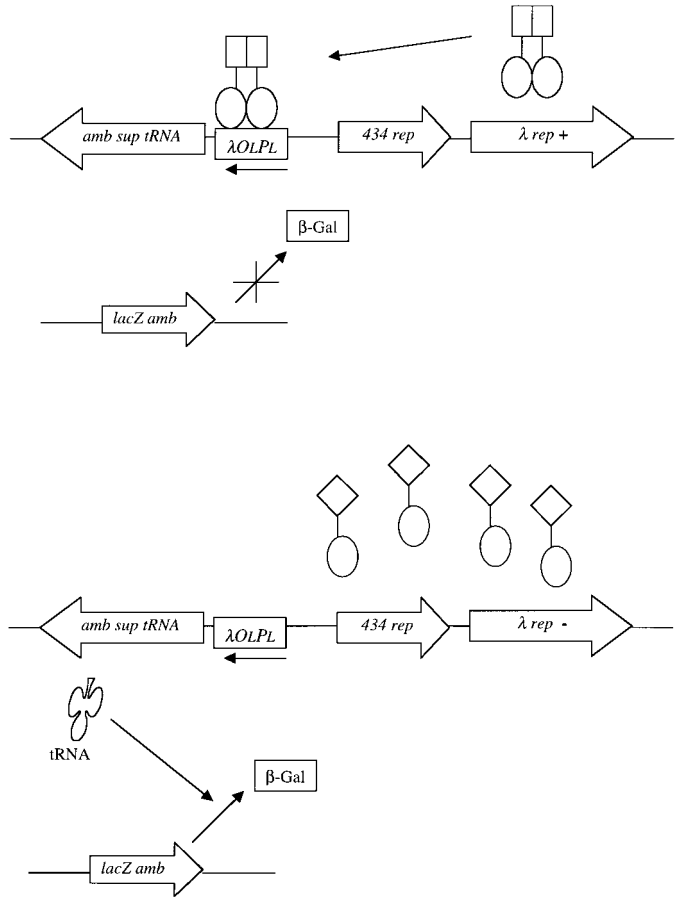

FIG. 1. Rationale of the genetic screen used to isolate oligomerizing peptides. Top part: a functional repressor fusion protein (λ rep$^+$) binds the λ operator, prevents the transcription of the amber suppressor tRNA and confers *lac*$^-$ phenotype. Bottom part: a nonfunctional repressor fusion protein (λ rep$^-$) does not bind the operator, allowing the expression of the suppressor tRNA and the synthesis of β-galactosidase.

two other dimeric fusion proteins, which were constructed as a positive control. This was done fusing the cI DNA binding domain to the cAMP dependent protein kinase regulatory subunit and the glutathione S-methyltransferase. In contrast, bacteria expressing either a truncated form of the repressor lacking the dimerizing domain or the cI DNA binding domain fused to the monomeric enzyme β-lactamase were λ sensitive and served as negative controls (see lower part of Table 1).

Three of the isolated clones encode a known or hypothetical *E. coli* protein sequence. Clone R2 contains the entire sequence corresponding to the heat shock protein IBPB (19). This protein has been characterized and there is evidence that it forms oligomers (20). Clone R3 contains the sequence corresponding to the C-terminal part of the protein AAC70143. The insert of R6 encodes the N-terminal region of the YIGA protein. AAC70143 and YIGA sequences are annotated as hypothetical ORFs in the protein database and their function is unknown. The data suggest that the two ORF segments in question can be expressed inside the bacterium and do fold into a stable structure.

To confirm the validity of our approach we derived a number of deletion mutants from clone R6 with a view of identifying the minimal sequence able to oligomerize. As a consequence of a co-ligation event between two library DNA fragments the protein fusion expressed by clone R6 contained a carboxy-terminal extension (AYESGK) that is not part of the YIGA ORF. Deletion of this peptide sequence indicates that it is not essential for the oligomerization. Instead, two larger carboxy-terminal deletions in the YIGA ORF fragment completely abolished its ability to substitute for the repressor carboxy-terminal domain (Table 2). The repressor fusion encoded by clone R6 includes a short sequence (CLRRIYSS) between the repressor moiety and the ATG initiation codon of YIGA ORF. Removal of this segment together with the two amino-terminal residues of the YIGA ORF caused a marked increase in phage sensitivity together with a moderate increment of β-galactosidase activity (Table 2). This result indicates that all or part of these amino acid residues are either directly involved in the folding-assembly of the YIGA protein segment or act as a linker sequence between the YIGA domain and the repressor domain.

The rest of the clones isolated from the two genomic libraries present DNA sequences that were also identified in the database as belonging to the *E. coli* genome. However, these DNA sequences apparently do not direct the synthesis of a natural product. In fact the inserts of clones R1, R25, R45, R51, and R54 overlap but are not in frame with known or hypothetical protein sequences. In particular, the sequence of clone R51 derives from the "non-coding strand" of a known protein. Finally, in the case of clone R37, the sequence is presumably part of a non-coding region of the *E. coli* genome. The lengths of these peptide sequences range between 11 and 34 amino acids. A BLAST search of these protein sequences did not show a significant similarity to the sequences present in the whole protein database.

To verify to what extent the λ resistance phenotype conferred by this second group of fusion proteins was specifically dependent upon their amino acid sequences we decided to carry out a deletion analysis on one of the peptides (Table 2). Elimination of the two carboxy-terminal residues, GR, from the 18 amino acid peptide encoded by clone R51, did not cause a functional change. In contrast, deletion of the last three amino acid residues, RGR, determined a radical increase in both phage sensitivity and β-galactosidase activity. Thus, we have identified two peptides, pAC137 and pAC140, which differ by the presence of a single arginine at the carboxy-terminus, but display drastic phenotypic differences. It should also be mentioned that DNA sequencing of the two clones, besides the presence of the carboxy-terminal deletions, identified also the unanticipated substitution of a T to I in position 9 of the peptide. The presence of a single arginine at the carboxy-terminus is not sufficient for the activity of the peptide, as the deletion of six amino acid residues from its amino-terminal region led to a high level of phage sensitivity (Table 2).

DISCUSSION

The data presented here confirm the value of the λ repressor as a tool to detect and map the interactions between protein domains. As a result of our genetic screen we have identified two classes of sequences able to promote protein oligomerization *in vivo*. Some sequences encode proteins which are known or which presumably exist inside the bacterial cell. Since a large number of natural proteins form oligomeric complexes, we suggest that systematic PCR cloning of ORFs combined with the λ repressor dimerization test could be used to prove the expression and the ability of the hypothetical ORFs present in the database to fold-assemble in *E. coli*. Large-scale information about the tendency of proteins to homo-oligomerize may also be obtained with other screening methods to analyze protein-protein interactions, like the yeast two-hybrid system (21).

The second group of clones includes DNA sequences that direct the synthesis of relatively short and "unnatural" peptides. We found it remarkable that such sequences could substitute for the repressor dimerizing domain, although it is possible that not all the peptides isolated by this method assemble in a simple homoligomeric form. The function of some of these protein fusions could derive from the interaction of the peptides with another oligomeric protein present inside the cell, which could link and indirectly bring together

the λ repressor headpieces. Similarly, DNA or RNA molecules could also act as a bridge. This last possibility is reinforced by the fact that several sequences are rich in positively charged residues. Regardless of the specific mechanism of interaction, the identification of novel sequences promoting protein oligomerization may be important to understand the evolution of natural protein structures. In addition, it may be interesting for protein engineering applications. Zhang and collaborators have screened a yeast genomic library for peptides able to support protein oligomerization *in vivo* (8). As in our case, the majority of the functional sequences isolated by these authors do not belong to the complement of the natural proteins.

REFERENCES

1. Phizicky, E. M., and Fields, S. (1995) *Microbiol. Rev.* **59,** 94–123.
2. Goodsel, D. S., and Olson, A. J. (1993) *Trends Biochem. Sci.* **18,** 65–68.
3. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 13–20.
4. Jones, S., and Thornton, J. M. (1997) *J. Mol. Biol.* **272,** 121–132.
5. Larsen, T. A., Olson, A. J., and Goodsel, D. S. (1998) *Structure* **6,** 421–427.
6. Henrick, K., and Thornton, J. M. (1998) *Trends Biochem. Sci.* **23,** 358–361.
7. Kohn, W. D., and Hodges, R. S. (1998). *Trends Biotechnol.* **16,** 379–389.
8. Zhang, Z., Murphy, A., Hu, J. C., and Kodadek, T. (1999) *Curr. Biol.* **9,** 417–420.
9. Hu, J. C., O'Shea, E. K., Kim, P. S., and Sauer, R. T. (1990) *Science* **250,** 1400–1403.
10. Pabo, C. O., Sauer,R. T., Sturtevant, J. M., and Ptashne, M. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 1608–1613.
11. Ptashne, M. (1992) A Genetic Switch, Cell Press and Blackwell Scientific Publications, Cambridge MA.
12. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994) Current Protocols in Molecular Biology. Wiley, New York.
13. Miller, J. H. (1992) *in* A Short Course in Bacterial Genetics, pp. 72–74. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14. Chang, A. C. Y., and Cohen, S. N. (1978) *J. Bacteriol.* **134,** 1141–1156.
15. Cairns, M. T., Green, A. J., White, P. M., Johnston, P. G., and Brenner, S. (1997) *Gene* **185,** 5–9.
16. Jappelli, R., and Brenner, S. (1998) *FEBS Lett.* **423,** 371–375.
17. Koh, C. G. (1992) New Approaches for Mapping Bacterial Genomes Ph.D. thesis, University of Cambridge, England.
18. Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew G. F., Gregor, J., Davis N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) *Science* **277,** 1453–1474.
19. Allen, S. P., Polazzi, J. O., Gierse, J. K., and Easton, A. M. (1992) *J. Bacteriol.* **174,** 6938–47.
20. Shearstone, J. R., and Baneyx, F. (1999) *J. Biol. Chem.* **274,** 9937–9945.
21. Fields, S., and Song, O. (1989) *Nature* **340,** 245–246.