

# A Bacterial One-Hybrid Selection System for Interrogating Zinc Finger-DNA Interactions

Sundar Durai<sup>1,2,3,§</sup>, Allen Bosley<sup>1,§</sup>, Alice Bridgeman Abulencia<sup>1</sup>, Srinivasan Chandrasegaran<sup>\*,2</sup> and Marc Ostermeier<sup>\*,1</sup>

<sup>1</sup>Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA

<sup>2</sup>Department of Environmental Health Sciences, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA

<sup>3</sup>Present Address: Center for Bioinformatics, School of Life Sciences, Pondicherry University, Pondicherry 605014, India

**Abstract:** We have developed two bacterial one-hybrid systems for interrogating and selecting zinc finger-DNA interactions. Our systems utilize two plasmids: a zinc finger-plasmid containing the gene for the zinc finger fused to a fragment of the alpha subunit of RNA polymerase and a reporter plasmid where the zinc finger-binding site is located upstream of a reporter gene—either the gene encoding the green fluorescent protein (GFP) or chloramphenicol acetyltransferase (CAT). Binding of the zinc finger domain to the target binding site results in a 10-fold increase in chloramphenicol resistance with the CAT reporter and an 8- to 22-fold increase in total cell fluorescence with the GFP reporter. The CAT reporter allows for sequence specific zinc fingers to be isolated in a single selection step whereas the GFP reporter enables quantitative evaluation of libraries using flow cytometry and theoretically allows for both negative and positive selection. Both systems have been used to select for zinc fingers that have affinity for the motif 5'-GGGGCAGAA-3' from a library of approximately  $2 \times 10^5$  variants. The systems have been engineered to report on zinc finger-DNA binding with dissociation constants less than about 1  $\mu$ M in order to be most applicable for evaluating binding specificity in an *in vivo* setting.

**Keywords:** Incremental truncation, flow cytometry, one-hybrid system, zinc fingers.

## INTRODUCTION

Zinc fingers (ZF) are ~30 amino acid DNA-binding motifs that fold into a  $\beta\beta\alpha$  structure around a central zinc ion. One finger recognizes a 3-base pair sequence along the major groove of DNA. A larger sequence can be targeted by covalently linking several zinc fingers in tandem. Applications for this type of DNA binding domain range from localizing artificial transcription factors to conferring DNA binding specificity to chimeric nucleases [1-3]. One of the key goals in zinc finger engineering has been to produce proteins that can specifically recognize a pre-determined DNA sequence.

One of the most important and successful strategies for selecting zinc fingers with affinity for the desired target site has been phage display [4-7]. However, this method requires several rounds of selection and these selections do not occur in an *in vivo* setting. A few alternative systems for zinc finger selection have been developed including a yeast one-

hybrid system [8], a mammalian one-hybrid system [9] and a bacterial two-hybrid system [10, 11] derived from a system developed by Hochschild and coworkers [12, 13]. These systems have the advantage that zinc fingers with the desired affinity can be identified in a single round (instead of multiple rounds for phage display) and that the affinity is selected for in an *in vivo* setting. As in all genetic reporter systems, the correlation of protein activity (i.e. binding affinity of a zinc finger) of a particular clone to the transcription of a reporter gene must be done under the assumption that the protein concentration in the cell does not vary greatly from clone to clone.

The bacterial two-hybrid system developed by Pabo and coworkers [10, 11] – derived from a system developed by Hochschild and coworkers [12, 13] – links zinc finger binding to expression of the yeast *HIS3* gene that can complement a defect in growth in *Escherichia coli* cells bearing a deletion of the *hisB* gene. In this system, the zinc finger protein is fused to the yeast Gal11P protein. Gal11P has affinity for the yeast Gal4 protein. Gal4 is fused to a fragment of the  $\alpha$  subunit of *E. coli* RNA polymerase (RpoA[1-248]). The ZF-Gal11P and Gal4-RpoA[1-248] fusions are encoded on separate, compatible plasmids. The desired zinc finger-binding site is positioned on the episome at position -63 in a promoter upstream from the *HIS3* gene. Thus, binding of the ZF-Gal11P fusion to this binding site recruits the Gal4-RpoA[1-248] fusion which in turn recruits the rest of the RNA polymerase. The stimulation of *HIS3*

\*Address correspondence to these authors at the Department of Chemical and Biomolecular Engineering, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA; E-mail: oster@jhu.edu and Department of Environmental Health Sciences, Johns Hopkins University Bloomberg School of Public Health, 615 North Wolfe Street, Baltimore, MD 21205, USA; E-mail: schandra@jhsph.edu

§The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint first authors.

transcription that results allows for growth of a *hisB* strain on selective minimal media plates. The stringency of the system can be increased by the addition of the HIS3 competitive inhibitor 3-aminotriazole to the plates to select for zinc fingers with very high affinity. This system has many advantages over phage display including: (a) single-step isolation of candidates in an *in vivo* context from libraries  $>10^8$  in size and (b) the bypass of complications associated with the export of proteins to the cell membranes. This bacterial two-hybrid system, combined with a domain shuffling strategy, has been successfully used to select zinc fingers with sub 100 pM dissociation constants for their desired target [11].

Although affinity is very important, another important parameter for evaluating zinc fingers is DNA binding specificity. Although "low-stringency" selections have been described with the two-hybrid system, the level of binding affinity that can be detected has not been reported [11]. A genetic system that can report on zinc finger-DNA binding for a wide range of affinities would be very useful for evaluating a zinc finger's specificity (e.g. whether or not a particular zinc finger binds to related sites with dissociation constants in the high nanomolar range). To further emphasize this point, consider a zinc finger that has a high affinity for only one particular sequence but has moderate affinities for a large number of sequences. An interrogation system that can only detect high affinity binding would label this zinc finger "specific". On the other hand, our system can detect moderate binding affinities and would be capable of giving a more complete description of the zinc finger's binding behavior promiscuity. In addition, since the two-hybrid system requires the integration of the reporter operon by homologous recombination into the episome, the evaluation of multiple DNA binding sites or a library of DNA binding sites to address issues of specificity becomes cumbersome.

We have developed a one-hybrid system that can be used quickly and easily to select for or report on zinc finger-DNA binding specificity with about 1  $\mu$ M or less dissociation constants for a given DNA binding site. Our system is less complex and less cumbersome since it avoids the requirements for a specific *E. coli* strain (and growth for several days on minimal media) as well as the integration of the reporter operon into the episome.

## MATERIALS AND METHODS

### Materials

All enzymes were purchased from New England BioLabs (Beverly, MA). The Electromax DH5 $\alpha$ -E was purchased from Invitrogen Life Technologies (Carlsbad, CA). All DNA purification kits were purchased from Qiagen (Valencia, CA).

### Plasmid Construction

Plasmid pDB1-CAT(-62) was constructed in three steps. First, oligonucleotide d(5'-AATTCAGTGGGGCAGAAGCATGCTTAGGCACCCCGGGCTTACACTTTATGCTTCCGGCTCGTATGTTGTG-3') was phosphorylated using T4 polynucleotide kinase and annealed to its corresponding complementary oligonucleotide to make a

double stranded DNA fragment bearing a *EcoRI* and *Sall* sticky ends. Annealing was performed at a concentration of 0.4 mg/ml of each oligonucleotide using a thermalcycler. Fifty  $\mu$ l aliquots of the mixed oligonucleotides were dispensed into PCR tubes (500  $\mu$ l size) and were then placed in a thermal cycler and a program was set up to perform the following profile: (i) heat to 95°C and remain at 95°C for 2 minutes, (ii) ramp cool to 22°C over a period of 30 minutes, (iii) proceed to a storage temperature of 4°C. The tubes were spun in a microfuge to draw all moisture from the lid. The samples were pooled into a larger tube and stored at 4°C. The oligonucleotide duplex was used in a 3 to 1 molar excess over *EcoRI/Sall*-cleaved Hochschild plasmid pFW11-OR2-62 for ligation. The inserted segment bears the DNA binding site for ZF- $\Delta$ QNK (5'-GGGGCAGAA-3') positioned at -62 region of the promoter relative to the transcription start site in the Hochschild plasmid pFW11-OR2-62. The DNA binding site is preceded and followed by two unique restriction sites *SpeI* and *SphI* (5'-*SpeI*-GGGGCAGAA-*SphI*-3'), which were useful for cloning variant binding sites into the plasmid. Next, a *BsaAI/ScaI* fragment containing the already existing chloramphenicol acetyltransferase (*cat*) gene was removed, resulting in a large fragment with blunt ends. The blunt ends were ligated using 1  $\mu$ l of T4 DNA ligase in a 20  $\mu$ l reaction at room temperature for 12 h. The final step was to clone the 660 base pair gene for CAT between the sites *NdeI/HindIII* to locate *cat* downstream from the engineered promoter. The chloramphenicol acetyltransferase gene was PCR-amplified [14] using oligonucleotide d(5'-GAAGCTCATATGGAGA AAAAATCACT-3') (start primer) and d(5'-AAGCTTCA TTTACGCCCCGCCCTGCCA-3') (stop primer) from pFW11-OR2-62. The PCR product was digested with *NdeI/HindIII* and then ligated into *NdeI/HindIII*-cleaved variant of pFW11-OR2-62 vector prepared in the previous two steps to create pDB1-CAT(-62). This construct was confirmed by sequencing.

Reporter plasmid pDB1-GFP(-62) was constructed by replacing the *NdeI/HindIII*-segment of pDB1-CAT(-62) with a fragment encoding the GFPmut2, in which an internal *NdeI* restriction enzyme site was removed by overlap extension PCR [14], using the following oligodeoxyribonucleotides: d(5'-GATATACATATGAGTAAAGAAGAA-3'), d(5'-CT GTTTCATGTGATCTGGGTATCTCGCA-3'), d(CCCAGATCACATGAAACAGCATGACTTT-3'), and d(5'-TATTTG AAGCTTTTATTGTATAGTTCATC-3'). PCR amplification was achieved through 30 cycles of 94°C for 1 min, 57°C for 1 min, and 72°C for 1 min, utilizing the Taq polymerase.

Fusion plasmid (pA-3- $\Delta$ QNK) containing the zinc finger  $\Delta$ QNK fused to a fragment of  $\alpha$ -subunit of RNA polymerase (RpoA[1-248]) was constructed as follows: the zinc finger  $\Delta$ QNK was PCR amplified using oligodeoxyribonucleotides d(CCGCGCGCGGCCGCAATGGAAAAACCTTACAAA) (start primer) and d(ACCGCCGGATCCTTACTTCTTATTCTGATGCGT) (stop primer) from pET15b- $\Delta$ QNK-F<sub>N</sub> [15]. The PCR product was digested with *NotI/BamHI* and then ligated into *NotI/BamHI*-cleaved pBR $\alpha$ -LN [13], which contains the sequence encoding the  $\alpha$ -subunit of RNA polymerase and a three alanine linker. The pA-3- $\Delta$ QNK fusion construct was confirmed by sequencing.

### Evaluation of the One-Hybrid Systems

To test the antibiotic system, overnight cultures were diluted and approximately 1000 CFU (determined on plates without Cm) were plated on LB plates containing increasing amounts of Cm and incubated for 16 hours at 37 °C.

To evaluate total cell fluorescence, cells were grown in 50 mls of 2x YT (16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl), 0.2% glucose, 100 µg/ml ampicillin and 50 µg/ml Kan in 250 ml shake flasks at 37 °C until the OD<sub>600</sub> was greater than 0.5. The production of the fusion protein was induced by the addition of IPTG to a final concentration of 3 mM and the culture shaken at 37°C for two and a half hours. These cells were then diluted 1:10 in PBS (8 g/L NaCl, 0.201 g/L KCl, 1.15 g/L Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 1.9 g/L KH<sub>2</sub>PO<sub>4</sub>) and their fluorescence was measured utilizing a Becton-Dickinson FACS Caliber flow cytometer. The excitation wavelength was 488 nm and the emission wavelength was 530 nm.

### Optimization of the Fusion and Reporter Plasmids

Incremental truncation was performed as described [16]. The fusion plasmid pA-3-ΔQNK was altered to be suitable for incremental truncation of the (G<sub>4</sub>S)<sub>3</sub> linker between the RpoA[1-248] fragment of the α-subunit of RNA polymerase and the gene for the zinc finger ΔQNK. An *Nsi*I site (that leaves a 4-base 3' overhang, which is protected from *Exo*III digestion) was positioned just upstream of the ΔQNK gene and a *Spe*I site (leaves a 3' recessed end, which is susceptible to *Exo*III digestion) was located just downstream from the linker. The two sites were spaced apart far enough such that plasmid could be digested efficiently with both enzymes.

In order to create an incremental truncation library where the distance between the QNK binding site and the start of transcription was varied, plasmid pDB1-GFP(-62) was digested with *Sph*I and the following annealed oligonucleotides were inserted: d(5'-CCTGCAGATACTTGATCCGTTACACATG-3') and d(5'-TGTAACGGATCCAAGTATCTGCAGGCATG-3'). The restriction enzyme sites *Pst*I and *Bam*HI were incorporated within these oligonucleotides for use in creating the incremental truncation library.

### Generation of Synthetic Binding Sites

The following oligonucleotides were synthesized: d(5'-CTAGTGGGTCAGAAGCATG-3'), d(5'-CTAGTGGGTCGGAAGCATG-3'), d(5'-CTAGTGGGGTGAAGCATG-3'), d(5'-CTAGTGGCGCAGAAGCATG-3') and d(5'-CTAGT TAGATCTTTGCATG-3'). The oligonucleotides were phosphorylated with T4 polynucleotide kinase and annealed to the following oligonucleotides d(5'-CTTCTGACCCA-3'), d(5'-CTTCCGACCCA-3'), d(5'-CTTCCACCCA-3'), d(5'-CTTCTGCGCCA-3') and d(5'-CAAAGATCTAA-3'), respectively, such that an *Spe*I overhang and *Sph*I overhangs were created. The oligonucleotide duplexes were used in a 3 to 1 molar excess over *Spe*I/*Sph*I-cleaved pDB1-CAT(-55) vector for ligation. The recombinant DNA was ethanol-precipitated and was electroporated into DH5α-E cells. The transformed cells were grown on LB agar plates containing 50 µg/ml Kan. Colonies were picked, plasmids were isolated using Qiagen plasmid purification kit (Qiaprep Miniprep) and the clones were sequenced.

### Sorting Based on Total Cell Fluorescence

Cells were grown in the same manner as when analyzing total cell fluorescence levels. The sort gate was set to isolate any library members resulting in a fluorescence higher than that resulting from the control fusion plasmid and the reporter plasmid pDB1-GFP(-62). The gate was also designed so as to prevent selection of cells resulting in a high level of fluorescence due only to their larger size. The sort type was set to exclusion, and dilutions of cells in PBS were prepared and analyzed to determine which yielded a sort rate of approximately 100 events per second. The isolated cells, suspended in PBS, were mixed with 10x YT, glucose, Amp and Kan. The concentrations of the media components were the same as when the cells were grown prior to sorting. After overnight growth at 37°C, cells were diluted 1:1000 in 2x YT and plated on LB plates supplemented with 100 µg/ml Amp and 50 µg/ml Kan. Colonies were selected to prepare inoculums and plasmids were isolated by the Qiaprep Miniprep kit and sequenced.

### Removal of Recombination-Prone Site to Minimize the Occurrence of False Positives

The reporter plasmid pDB1-CAT(-55) was amplified using PCR using oligonucleotides d(5'-GACTTCAGCGGC GCATTTGAAGAGATAAATTGCACTG-3') and d(5'-CGTTGTGCGGCCGCTGCGTGAGATCTTCCTTCAACT ACTCAGCAAAAGTT-3'). This amplification excluded the region responsible for the recombination while incorporating several restriction sites. The PCR product was then digested with the restriction enzyme *Not*I—present on both of the primers—and ligated to cyclize the vector. This vector was then digested *Xba*I and *Xho*I and the small fragment was isolated. *Xba*I and *Xho*I sites flanked the deleted recombination prone site. This fragment was incorporated into *Xba*I/*Xho*I digested pDB1-CAT(-55) that had not been subjected to PCR to create pDB2-CAT(-55). This was done to avoid the occurrence of point mutations, incorporated through PCR, elsewhere on the reporter plasmid.

### Plasmids for Evaluating DNA-Binding Site Locations

We designed the following oligonucleotides to position the QNK-binding site at positions -58, -57, -56, -54, -53 and -52 relative to the transcription start site in the reporter plasmid (Fig. 3A). The following oligos were synthesized:

d(5'-AATTCAGTGGGGCAGAAGCATGCGCACCC CGGGCTTACACTTTATGCTTCCGGCTCGTATGTTG-3') for -58 location; d(5'-AATTCAGGGGCAGAAT AGCATGCCCCCG GCTTACACTTTATGCTTCCGGC TCGTATGTTGTG-3') for -57 location; d(5'-AATTC A CTAGGGGCAGAATGCATGCCCCGGGCTTACAC TTTATGCTTCCGGCTCGTATGTTGTG-3') for -56 location; d(5'-AATTCAC TAGTTGGGGCAGAACATG CCCCCGGGCTTACACTTTATGCTTCCGGCTCGTAT GTTGTG-3') for -54 location; d(5'-AATTCAGT TAGGGGCAGAAATGCCCCGGGCTTACACTTTATG CTTCCGGCTCGTATGTTGTG-3') for -53 location; and d(5'-AATTCAGT TAGGGGCAGAAATGCCCCCG GCTTACACTTTATGCTTCCGGCTCGTATGTTGTG-3') for -52 location. The oligonucleotides were phosphorylated with T4 polynucleotide kinase and annealed

to their corresponding complementary oligonucleotides such that *EcoRI* and *SalI* overhangs were created. The oligo duplexes were used in a 3 to 1 molar excess over *EcoRI/SalI*-cleaved pDB2-CAT(-55) vector for ligation. The recombinant DNA was ethanol-precipitated and was electroporated into DH5 $\alpha$ -E cells. The transformed cells were grown on LB agar plates containing 50  $\mu$ g/ml Kan. Colonies were picked, plasmids were isolated Qiagen plasmid purification kit (Qiaprep Miniprep) and the clones were sequenced.

### Construction of Random Library of the Middle Zinc Finger

A diverse library of zinc fingers in which the amino acids at helix positions -1, 2, 3 and 6 [17] of the middle finger (Finger 2) was created by the reassembly of the 3 different top strand oligodeoxyribonucleotides: d(5'-ATGGA AAAACCTTACAAGTGTCCGGAATGTGGGAAGTCCTTTAGT-3'), d(5'-CAGCGTACGCATACCGGTGA GAAGCCCTACAAATGCCCCGAATGCGGAAAATCATTTTCG-3') and d(5'-CAACGAACCCACACAGGCGAGA AACCATATAAATGTCCTGAGTGTGGTAAGAGCTTTAGC-3'); and the 3 different bottom strand oligodeoxy-ribonucleotides: d(5'-ACCGGTATGCGTACGCTGGTGCT TCTGCAGGTTGCTAGACTGACTAAAGGACTTCCACATTCC-3'), d(5'-TGTGTGGGTTTCGTTGGTGNNNCTGCAGNNNNNACTNNNCGAAAATGATTTTCCGCATTTCG-3') and d(5'-TACTTCTTATTCTGATGCGTACGTTGATGCTTGACAGATGATCACTTCTGCTAAAGCTCTTACCACTCA-3'). These oligonucleotides were added, at a concentration of 2  $\mu$ M each, to 1 mM dNTP's and 10x EcoPol reaction buffer. This mixture was heated to 95°C for 10 minutes, and then gradually cooled to 22°C over a period of 30 min. Klenow was added and the mixture was incubated at room temperature for an additional 20 min. The mixture was then heated to 37°C for 10 min. The products of this annealing step were mixed with the two outside primers d(5'-GGTTCCTACTGCGGCCGCAATGGAAAAACCTTACAAG-3') and d(5'-CGTAGAGGATCCTTACTTCTTATTCTGATG-3'). Amplification of the library was achieved through 45 cycles of PCR utilizing the Taq polymerase. One cycle is defined as 94°C for 1 min, 52.5°C for 1 min, and 72°C for 1 min.

The resulting DNA fragment was gel purified and digested with *BamHI* and *NotI*. This library was ligated into a similarly digested fusion plasmid encoding the  $\alpha$ -subunit and 22 amino acid linker but lacking a zinc finger DNA binding domain. The ligation mixture was ethanol-precipitated and electroporated into Electromax DH5 $\alpha$ -E cells (Invitrogen). The electroporated cells were plated on a large LB agar plate containing 100  $\mu$ g/ml Amp. The plasmid DNA containing the zinc finger library was recovered from the large plate as described [16].

### Selection of Zinc Fingers

For the Cm system, frozen stocks of the library were diluted and  $2 \times 10^8$  CFU (determined on plates without Cm) were plated on 245 mm x 245 mm LB plates supplemented with 0.2% glucose and 500  $\mu$ g/ml Cm. The plates were incubated for 16 hours at 37°C. In the GFP system, the library underwent two rounds of sorting (as described

previously). The gates were the same for both sorts and were constructed to include cells fluorescing greater than those containing the control plasmid pA-3 and the reporter plasmid pDB2-GFP(-55).

### Separation of Selected Zinc Finger Plasmids

The zinc finger plasmids were separated from the reporter plasmids by isolating the plasmid DNA from a positive clone and transforming a small amount of this DNA into fresh DH5 $\alpha$ -E cells and plating on LB agar plate containing 100  $\mu$ g/ml Amp (the antibiotic of the zinc finger plasmid). Individual colonies from this transformation were then streaked first across a plate with 50 mg/ml Kan (the antibiotic of the reporter plasmid) and then (without re-flaming the loop) streaking across a plate supplemented with 100  $\mu$ g/ml Amp to assure that no reporter plasmid was present (i.e. no growth on the first plate but growth on the second plate). DNA was then prepped from a culture grown from this second streak and cotransformed with pure reporter plasmid into DH5 $\alpha$ -E cells. The culture from the second streak was also tested to confirm it lacked Cm resistance or fluorescence depending on the system being used. This ensures that any potential modified reporter plasmid cannot be carried along.

### Construction of Library of Binding Sites

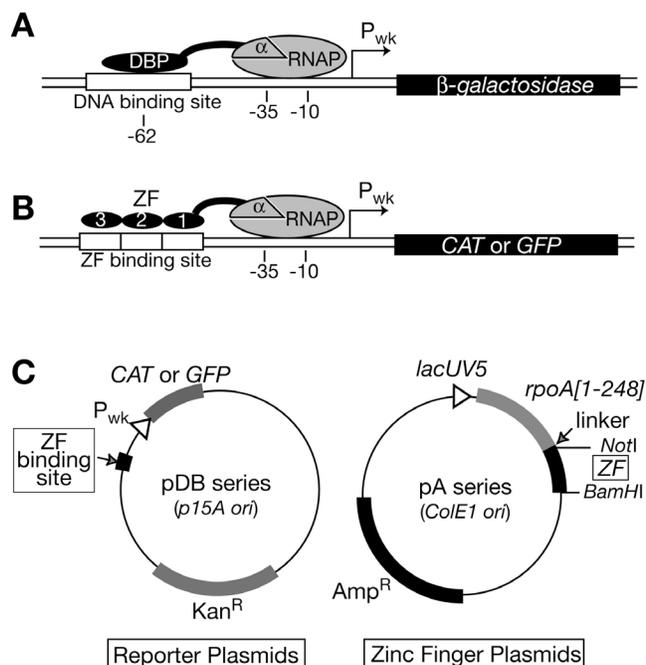
A library of binding sites was constructed in the reporter plasmid by the method of Abel-Santos and coworkers [18] using a primer encoding the 5'-end of the binding site d(5'-TAATCGCGCCTCGAGCAAGACGTTT-3') and a second primer encoding the 3'-end d(5'-TAAAGCCCGGGGCATGCNNNNNNNNNACTAGTGAATTCTTGAAG-3'). The resulting PCR was subjected to a second round of amplification with the third zipper primer d(5'-TAAAGCCCGGGGCATGC-3') to eliminate mismatches from the library sequences. The amplified sequences were digested with *XhoI/SphI* and were used in a 3 to 1 molar excess over *XhoI/SphI*-cleaved pDB2-CAT(-55) or pDB2-GFP(-55) vector for ligation resulting in 1.7 million and 2 million colonies, respectively. The libraries were recovered from the plate as before and the plasmid DNA was isolated using the Qiagen midi prep. Both of these libraries were cotransformed with the fusion plasmid pA-21- $\Delta$ QNK into Electromax DH5 $\alpha$ -E cells resulting 850,000 colonies in the GFP system and 416,000 colonies in the Cm system.

## RESULTS

### Modification of Hochschild's One-Hybrid System

We have developed a one-hybrid system for reporting on zinc finger-DNA binding *in vivo*. This system was derived from a bacterial one-hybrid genetic screen developed by Hochschild and coworkers [13, 19] (Fig. 1A). In the Hochschild system, a DNA-binding protein is fused to RpoA(1-248), a fragment of the  $\alpha$ -subunit of RNA polymerase, *via* a three alanine linker. Binding of this fusion protein to its cognate DNA target sequence (centered at -62 relative to the start of transcription) recruits the rest of the RNA polymerase and stimulates transcription of the *lacZ* reporter gene. This promoter-*lacZ* fusion is present on a specially engineered F' episome. Introduction of a new

promoter-*lacZ* fusion requires first constructing the new promoter-*lacZ* construct on a companion plasmid and then transferring it from the companion plasmid to the episome by homologous recombination. The disadvantages of this system from the perspective of interrogating zinc finger-DNA interactions are that the *lacZ* reporter must be used as a screen and that integration of the promoter-*lacZ* fusion into the episome is an extra step that severely limits constructing libraries of DNA binding sites.



**Fig. (1).** One-hybrid system for detection of zinc finger-DNA interactions (A) Schematic of transcription stimulation in a one-hybrid genetic screen developed by Hochschild and co-workers [13]. (B) Genetic selection system for interrogating zinc finger-DNA interactions (C) Plasmids for one-hybrid genetic selection system. The reporter gene—either chloramphenicol acetyltransferase (*CAT*), or green fluorescent protein (*GFP*)—is located downstream from a weak *lac* derivative promoter ( $P_{wk}$ ) on pDB series plasmids. A nine-base target site for binding by the zinc finger is located at a specific distance from the start of transcription. On the pA series of plasmids, the gene for the zinc finger (ZF) is fused to a fragment of the  $\alpha$ -subunit of RNA polymerase (*rpoA*[1-248]) via a sequence coding for an amino acid linker. Binding of the RpoA[1-248]-ZF fusion to the nine base site in the reporter plasmid recruits the other RNA polymerase subunits to stimulate transcription of the reporter gene.

We initially made two modifications to the Hochschild system (Fig. 1B). First, we decided to change the reporter gene to either chloramphenicol acetyltransferase (*cat*) or the gene coding for the green fluorescent protein (*GFP*) in order to be able to select for stimulation of the reporter gene by plating on increasing levels of chloramphenicol or by fluorescence activated cell sorting (FACS), respectively. Antibiotic resistance genes and genes encoding fluorescent proteins have been previously shown to be useful as reporters in two-hybrid systems [10, 20-22]. For the green fluorescent protein we used GFPmut2, a variant optimized for expression in *E. coli* [23]. Second, we chose to have the reporter gene on a low copy plasmid instead of integrating it

into the episome (Fig. 1C). Although this would potentially raise the background level of reporter gene expression because of the increased copy number of the reporter gene, this would speed up the process of constructing new promoter-reporter gene constructs and enable the creation of large DNA-binding site libraries. In addition, a higher copy number of the reporter operon might allow the detection of zinc finger-DNA binding with lower affinities.

In an initial test of the new system we used the zinc finger  $\Delta$ QNK. Desjarlais and Berg [15, 24] designed  $\Delta$ QNK based on a zinc-finger-framework sequence that was derived from a consensus sequence of 131 zinc finger sequence motifs. The consensus framework sequence was combined with specificity rules derived from previous native and mutant versions of Sp1 zinc fingers to create  $\Delta$ QNK. All zinc finger motifs were identical in sequence except for changes in one to four residues in its DNA-recognition regions. Thus,  $\Delta$ QNK, which binds the DNA sequence 5'-GGGGCAGAA-3', was designed and not selected for its sequence specificity.

The gene for  $\Delta$ QNK was fused to *rpoA*(1-248) via a sequence coding for the three alanine linker and the fusion protein was named RpoA-L(3)- $\Delta$ QNK. The plasmid pA-3- $\Delta$ QNK contains the gene for RpoA-L(3)- $\Delta$ QNK under the control of the IPTG-inducible *lacUV5* promoter. For the reporter plasmid, we modified the Hochschild companion plasmid pFW11 by centering the  $\Delta$ QNK binding site (5'-GGGGCAGAA-3') at -62 relative to the start of transcription and flanking the site with restriction enzyme sites. In addition we replaced the *lacZ* reporter gene with either the *cat* gene or the *GFPmut2* gene. These plasmids were named pDB1-CAT(-62) and pDB1-GFP(-62), respectively.

The level of gene expression was determined by cotransforming pA-3- $\Delta$ QNK with either pDB1-CAT(-62) or pDB1-GFP(-62) and plating cells on increasing levels of Cm or measuring total cell fluorescence utilizing a flow cytometer, respectively. The background level of gene expression was determined from identical experiments with cells cotransformed with pA-3, which expresses RpoA(1-248)-A-A-A without any DNA binding protein fused to it, and either pDB1-CAT(-62) or pDB1-GFP(-62). The level of increased gene expression as a result of the presence of the zinc finger was very modest, but detectable. Cells containing pDB1-CAT(-62) or pDB1-GFP(-62) and pA-3 could not grow on plates containing greater than 50  $\mu$ g/ml Cm whereas cells harboring pDB1-CAT(-62) and pA-3- $\Delta$ QNK grew on plates containing 100  $\mu$ g/ml Cm. Cells containing pDB1-GFP(-62) and pA-3 had a geometric mean total cell fluorescence of  $\sim$ 1.5 whereas cells bearing pDB1-GFP(-62) and pA-3- $\Delta$ QNK had a geometric mean total cell fluorescence of  $\sim$ 1.9. Increases in gene expression were not seen when a control reporter plasmid was used in which the  $\Delta$ QNK binding site was mutated to a sequence that  $\Delta$ QNK is known not to bind.

### Improving the Level of Transcriptional Activation of the One-Hybrid Systems

Although we were able to observe a statistically significant increase in gene expression, this increase was clearly insufficient to reliably analyze libraries of zinc finger

proteins or zinc finger binding sites. We reasoned that the modest stimulation of transcription may result from presenting the RpoA(1-248) domain in an suboptimal position for recruiting the rest of the RNA polymerase, perhaps because the RpoA(1-248) was presented in a non-optimal location (e.g. too far away) or that the three alanine linker was too short such that the zinc finger interfered with the required interaction of RpoA(1-248) with the other subunits of the polymerase. We decided to explore improvements to both the linker length between the zinc finger and the RpoA(1-248) and the distance between the 9 base target binding site and the start of transcription.

A library of fusions of RpoA(1-248) and  $\Delta$ QNK was created in which the linker length ranged from 3 to 22 amino acids. The longest linker in the library contained the sequence A-A-(G<sub>4</sub>S)<sub>3</sub>-T-S-A-A-A. Members of the library consisted of truncations in the C-terminal to N-terminal direction starting from the most C-terminal serine residue of the linker. When the linker library was co-transformed with the reporter plasmid pDB1-CAT(-62) and plated on increasing concentrations of Cm, colonies formed at 200  $\mu$ g/ml but not at 300  $\mu$ g/ml. Sequencing of 10 randomly selected linker library members from the 200  $\mu$ g/ml plates indicated that two different linkers conferred this increased level of antibiotic resistance: the full-length linker and the linker A-A-(G<sub>4</sub>S)<sub>3</sub>-T-A-A-A. The fusion containing the latter linker was used in subsequent studies and was named RpoA-L(21)- $\Delta$ QNK. The plasmid bearing this fusion gene was named pA-21- $\Delta$ QNK.

We next constructed a library to vary the target binding site location from approximately -80 to -30 in the pDB1-GFP reporter plasmid. Cotransformants of this library and pA-21- $\Delta$ QNK were subject to one round of FACS to select for an improved target binding site location. Two different locations were selected: -62 and -55. Of these two library members, the -55 target site (plasmid pDB1-GFP(-55)) yielded the highest level of fluorescence. Moving the DNA binding site to -55 did not result in a significant increase in the background level of transcription in the absence of the fusion plasmid or in the presence of control plasmid pA-3. Cotransformants of pDB1-GFP(-55) and pA-21- $\Delta$ QNK had a 7-fold greater geometric mean of total cell fluorescence than cotransformants of pDB1-GFP(-55) and pA-3. Likewise, cotransformants of pDB1-CAT(-55) and pA-21- $\Delta$ QNK had a 6-fold greater MIC for Cm than cotransformants of pDB1-CAT(-55) and pA-3.

### Specificity of Transcriptional Activation

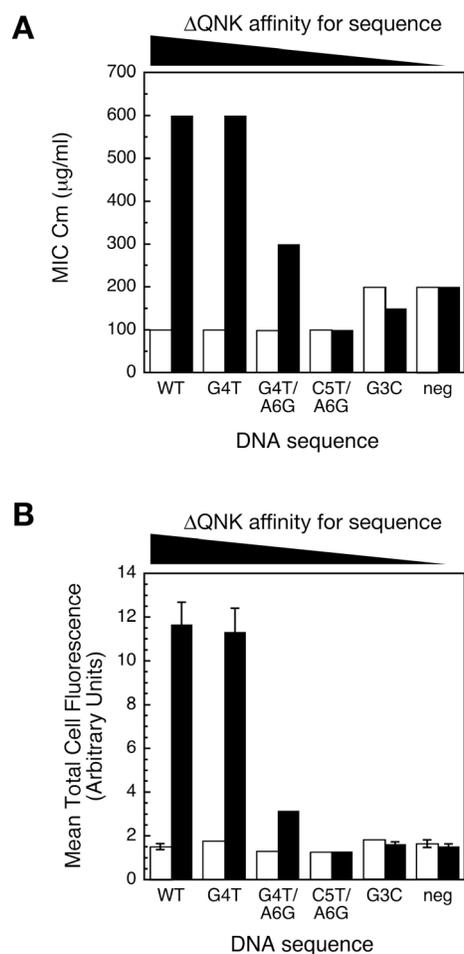
We next sought to confirm that our system could discriminate between target and non-target binding sites and could report on zinc finger-DNA interactions with dissociation constants as high as 1  $\mu$ M. From gel shift assays, the affinity of  $\Delta$ QNK for a variety of related DNA sequences is known [15]. These sequences and affinities are described in Table 1. Transcription levels when using these modified target site sequences were consistent with zinc finger binding expectations (Fig. 2). The single base mutation that has a small negative effect on  $\Delta$ QNK binding *in vitro* resulted in no detectable decrease in the signal (Cm<sup>R</sup> or fluorescence). More importantly, single base mutations or extensive mutations of the binding site that have a large

negative effect on zinc finger binding *in vitro* abolish the stimulation of transcription.

**Table 1. Dissociation Constant of  $\Delta$ QNK and DNA Sequences**

Sequence	Abbreviation	$K_d$ (nM) <sup>a</sup>
5' -GGGGCAGAA-3'	WT	24
5' -GGGTCAGAA-3'	G4T	142
5' -GGGTCGGAA-3'	G4T/A6G	990
5' -GGGGTGGAA-3'	C5T/A6G	7500
5' -GGCGCAGAA-3'	G3C	>>25,000
5' -TAGATCTTT-3'	neg	nd

<sup>a</sup>from Smith et al. [15].



**Fig. (2).** Validation of the bacterial one-hybrid system. Transcription stimulation, as measured by (A) chloramphenicol resistance or (B) total cell fluorescence, correlates with the affinity  $\Delta$ QNK has for the DNA sequence centered at -55. The DNA sequences and the affinity of  $\Delta$ QNK for these sequences are shown in Table 1. White bars, cells bearing pA-3 and either (A) pDB1-CAT(-55) or (B) pDB1-GFP(-55); Black bars, cells bearing pA-21- $\Delta$ QNK and either (A) pDB1-CAT(-55) or (B) pDB1-GFP(-55). Error bars in (B) represent the standard deviation of three independent cultures.

The increases in reporter gene expression observed in both systems were qualitatively similar. The data suggested

that the affinity necessary for “full” induction of expression of the reporter gene corresponded to a zinc finger with an *in vitro*  $K_d$  between 140 nM and 1  $\mu$ M. A zinc finger with an *in vitro*  $K_d$  of 1  $\mu$ M still results in some stimulation but one with an *in vitro*  $K_d$  of 7.5  $\mu$ M fails to induce reporter gene expression above background levels. However, it is important to recognize that the conditions for determining *in vitro* binding data and conditions in the cell are quite different. In addition, deficiencies in affinity can be compensated for by increases in zinc finger expression. Thus, as in all genetic reporter systems for DNA binding, comparisons between reporter output and a DNA-binding protein's affinity for a DNA site must be made with caution. The affinity of  $\Delta$ QNK for these sites *in vivo* is unknown; however, the correlation between *in vitro* affinity and our one-hybrid reporter output is encouraging. As discussed later, the increased background level of expression for cells containing DNA sequences G3C and neg (observed with the Cm system, Fig. 2A) may be the result of the natural RpoA binding directly to these sequences (see last section of Results).

### Minimization of False Positives

At the same time that the linker length and binding site location were being engineered, the frequency of false positives was investigated. DH5 $\alpha$ -E cells containing reporter plasmid pDB1-CAT(-62) and the control plasmid pA-3 have an MIC of 100  $\mu$ g/ml Cm. Increasing numbers of these cells were plated on plates containing 250  $\mu$ g/ml Cm to determine the frequency of false positives. This revealed a high false positive frequency of  $\sim 3 \times 10^{-4}$ . Sequencing of the plasmid DNA of ten false positives revealed a rearranged plasmid in which the *cat* reporter gene was relocated downstream from a specific region on the plasmid. Analogous rearranged plasmids were found when false positives were isolated using the GFP reporter system. These rearranged plasmids conferred constitutive high levels of transcription of the reporters and an MIC  $\geq 500$   $\mu$ g/ml for Cm.

To minimize the number of false positives that result in constitutive high expression of the reporter gene, the recombination-prone region ( $\sim 1100$  base pair region coding for a fragment of the *lacI* gene), was removed from the reporter plasmid pDB1-CAT(-55) to create pDB2-CAT(-55). When 189 million DH5 $\alpha$ -E CFU (determined in the absence of Cm) bearing pDB2-CAT(-55) and pA-3 were plated on a large 248 x 248 mm LB plates containing 500  $\mu$ g/ml Cm, no colonies were visible after 16 hours of incubation at 37 °C (corresponding to a false positive frequency of  $\leq 5.3 \times 10^{-9}$ ).

A positive spiking experiment was performed in order to confirm that true positives could also grow under these conditions. Approximately 1000 CFU (based on no antibiotic) from a culture of cells bearing pDB2-CAT(-55) and pA-21- $\Delta$ QNK (i.e. a true positive) were mixed with approximately 200 million CFU from a culture of cells bearing pDB2-CAT(-55) and pA-3. This mixture was plated on a large plate containing 500  $\mu$ g/ml Cm. The number of colonies that grew exceeded 1000, indicating that the colony forming ability of true positives would not be impaired under library-selection conditions. The plasmid DNA of ten colonies was isolated and confirmed to be pDB2-CAT(-55) and pA-21- $\Delta$ QNK. Thus, the false positive frequency of our

one-hybrid system with the *cat* reporter is very low:  $\leq 5.3 \times 10^{-9}$ .

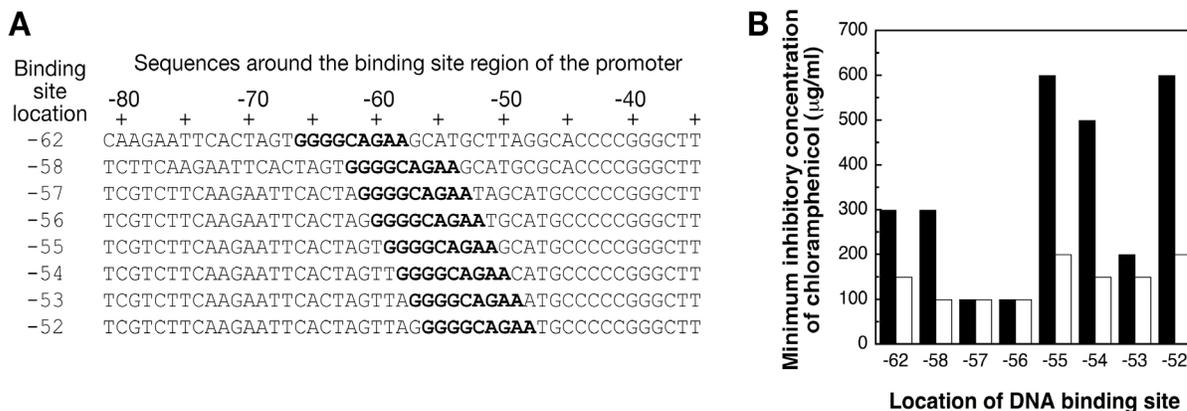
### Transcriptional Activation Observed When the DNA Target Site is Moved to Sites Other than that Centered at -55

Cells containing pDB2-CAT(-55) and pA-21- $\Delta$ QNK have an MIC of 600  $\mu$ g/ml Cm because the RpoA-L(21)- $\Delta$ QNK fusion protein binds to the QNK-binding site centered at -55. However, what if a zinc finger being tested as to whether it binds to a target site centered at -55 actually binds a sequence centered at -54 or -58? Would this result in a MIC of 600  $\mu$ g/ml as well? If so, this would potentially complicate identifying zinc finger-binding site pairs. To address this issue, we constructed a series of related reporter plasmids in which the QNK-binding site was located between -58 and -52 and determined the MIC for Cm of cells bearing these reporter plasmids and pA-21- $\Delta$ QNK (Fig. 3A). The level of Cm resistance varied in a complex manner with the location of the binding site (Fig. 3B), presumably due to the geometric requirements for RpoA-dependent recruitment of the other subunits of the RNA polymerase. Most sites exhibited moderate or no stimulation of transcription. However, centering the binding site at -52 and -54 resulted in transcriptional activation comparable to that observed with the binding site centered at -55. This illustrates that binding of the RpoA-zinc finger at sites other than the intended target site centered at -55 can result in false positives in terms of identifying zinc finger-DNA binding site pairs. This promiscuity does not result from the addition of the long flexible linker between rpoA(1-248) and the zinc finger, as RpoA-L(3)- $\Delta$ QNK (which contains the original A-A-A linker) exhibits a similar pattern of promiscuous transcriptional stimulation (Fig. 3B). None of the reporter plasmids in Fig. 3A, either when transformed alone or when cotransformed with control plasmid pA-3, confer resistance above background levels (MIC for Cm of 100  $\mu$ g/ml).

The reasons that binding sites centered at -52 and -54 were not identified in the initial optimization of the binding site location are not clear. However, an exhaustive sequencing of the colonies that grew was not performed and the -55 site was selected at a much higher frequency than -62, suggesting a bias in the library for certain constructs. In addition, the exact promoter sequences tested for -52 and -54 in Fig. 3A could not have been constructed in the library due to differences in flanking sequences.

### Selection of Zinc Fingers Binding the Site 5'-GGGGCAGAA-3'

To analyze our system's ability to identify zinc fingers binding a target site from a library of zinc finger variants, we constructed a library of variants of  $\Delta$ QNK in which positions -1, 2, 3 and 6 of the  $\alpha$ -helix of the middle finger were randomized. These positions are known to be important in the interaction between the zinc finger and DNA. The degeneracy of the library at the DNA level was  $4^{12} = 1.68 \times 10^7$  and at the amino acid level was  $21^4 = 194,481$ . A library of  $2 \times 10^6$  transformants was constructed and several clones were sequenced to ensure that no significant bias existed. The library was then cotransformed with either pDB2-CAT(-55) or pDB2-GFP(-55), both with the target site 5'-



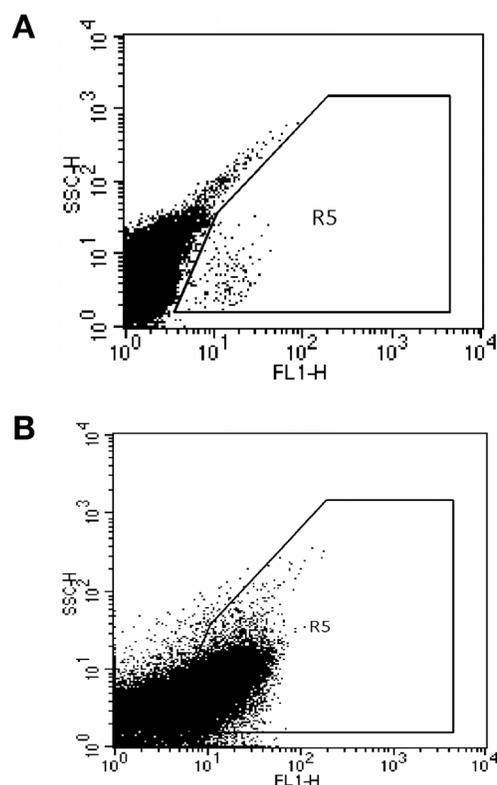
**Fig. (3).** Dependence of transcription stimulation on the location of the zinc finger-binding site. (A) Sequences in the upstream region of the promoter in which the binding site for the zinc finger  $\Delta$ QNK is positioned at different locations from -62 to -52 in the reporter plasmid pDB2-CAT. The  $\Delta$ QNK target binding site sequence is shown in bold. (B) The level of transcription activation – measured as the minimum inhibitory concentration of chloramphenicol – is shown as a function of binding site location in pDB2-CAT for cells bearing pDB2-CAT and either pA-21- $\Delta$ QNK (black bars) or pA-3- $\Delta$ QNK (white bars).

GGGGCAGAA-3' centered at -55. Both of these cotransformations produced about  $2 \times 10^6$  transformants

In the GFP system, members of the library of zinc fingers with total cell fluorescence above that found in cells transformed with pA-3 were selected using FACS (roughly above a fluorescence of 3). Initially, 0.035% of the library resulted in fluorescent levels greater than background. After two rounds of sorting, we enriched our library 1628-fold (Fig. 4). The plasmid DNA from 19 selected library members was sequenced to determine the zinc finger variants that were selected (Table 2). Seven different zinc fingers and ten different DNA sequences were found (there was one false positive which did not show any transcriptional stimulation and contained a stop codon at one of the randomized positions). To eliminate the possibility that these high levels of fluorescence were due to a modification of the reporter plasmid, the pA-21- $\Delta$ QNK variants were separated from the pDB2-GFP(-55) plasmid and then co-transformed with fresh pDB2-GFP(-55) (with the target binding site) to verify that increased fluorescence was not derived from modifications to the reporter plasmid or to chromosomal DNA.

The zinc finger library also underwent selection using the Cm system. Selection was on LB plates supplemented with 500  $\mu$ g/ml Cm incubated 16 hours at 37 °C. Sequencing of 18 randomly selected colonies resulted in the identification of eight different zinc fingers and eight different DNA sequences (Table 2). Controls analogous to those performed with the GFP system confirmed that all zinc finger sequences identified were true positives.

The number of different zinc fingers in the library with sufficient affinity for the target site (about  $\geq 10^6 \text{ M}^{-1}$ ) was estimated to be about 26 using the frequency of colonies on 500  $\mu$ g/ml Cm (Cm system) and about 25 using the frequency of events with high fluorescence (GFP system). This offers a reasonable explanation for incomplete overlap between the sequences identified by the two systems and the fact that the original  $\Delta$ QNK sequence was not selected from the library, since only eighteen sequences of each were



**Fig. (4).** A library of zinc fingers containing > 2.1 million clones was analyzed by flow cytometry. The library consisted of cells transformed with pDB2-GFP(-55) and pA-21- $\Delta$ QNK in which positions -1, 2, 3, and 6 of the middle zinc finger were randomized. (A) Side scatter versus fluorescence of the library before sorting (B) Side scatter versus fluorescence of cells after two rounds of sorting.

sequenced. Several zinc fingers isolated by one system were tested with the other system to evaluate the performance of both systems. Interestingly, several of the selected zinc finger sequences conferred much higher levels of total cell fluorescence than observed with the original  $\Delta$ QNK finger in the GFP system but very little difference in the corresponding MIC values the Cm system. The origins of this difference may reside in the fact that the two systems

Table 2. Selected Zinc Fingers and their Ability to Stimulate Transcription in the One-Hybrid Systems

Sequence <sup>a</sup> - 1236	Cm system				GFP system			
	Occurrences <sup>b</sup>	MIC <sup>c</sup> (µg/ml)			Occurrences <sup>b</sup>	Total Cell Fluorescence <sup>c</sup>		
		WT <sup>d</sup>	G4T/A6G <sup>d</sup>	C5T/A6G <sup>d</sup>		WT <sup>d</sup>	G4T/A6G <sup>d</sup>	C5T/A6G <sup>d</sup>
<b>QSDK</b>	-	600	300	100	-	11.3 ± 1.0	3.2	1.3
QITK	2(1)	600	-	-	6(2)	9.7	-	-
QSNK	2(1)	650	100	250	-	24	1.3	2.0
QTTK	2(1)	650	-	-	1(1)	15	-	-
QSTK	8(1)	650	100	250	6(2)	24	1.3	2.1
QEDR	-	-	-	-	1(1)	10	-	-
QETK	-	-	-	-	2(2)	4.1	-	-
QTFK	-	-	-	-	1(1)	3.1	-	-
QETR	1(1)	600	100	250	-	35	1.4	9.5
ITDR	1(1)	650	-	-	-	22	-	-
NDDR	1(1)	700	-	-	-	27	-	-
RTTK	1(1)	600	-	-	-	20	-	-
DNYR	-	500	250	250	1(1)	7.5	1.8	1.7

<sup>a</sup>Amino acids at positions -1, 2, 3, and 6 of the  $\alpha$ -helix of finger 2. Those amino acids in bold are that found in  $\Delta$ QNK (QSDK).

<sup>b</sup>Number of occurrences of the indicated amino acid sequence out of 18 randomly selected positives. The number in parenthesis is the number of different DNA sequences that coded for that amino acid sequence.

<sup>c</sup>Minimum inhibitory concentration of chloramphenicol.

<sup>d</sup>Sequence of binding site (see Table 1) in reporter plasmid pDB2-CAT(-55) or pDB2-GFP(-55).

<sup>e</sup>Geometric mean of total cell fluorescence (arbitrary units). Standard deviation is of three independent cultures.

use very different conditions for selection. The Cm system uses plating of uninduced, stationary-phase cells whereas the GFP system uses an induced, liquid culture of exponentially growing cells.

### Specificity of Selected Zinc Fingers

A subset of the selected zinc fingers were tested in both the Cm and the GFP systems for their ability to stimulate transcription from promoters containing some of the modified DNA binding sites listed in Table 1. None of the tested zinc fingers could stimulate transcription from promoters containing the negative control site (neg) or the G3C site. The G3C site contains a single base change (in the triplet that interacts with a finger other than the one that was mutated in the library) that abolishes stimulation with  $\Delta$ QNK. This confirms that these zinc fingers were not selected because they were non-discriminating DNA-binding proteins. Furthermore, the pattern of transcriptional stimulation of the fingers QSNK, QSTK and QETR differed from  $\Delta$ QNK (Table 2). The data suggests that  $\Delta$ QNK has enough affinity for G4T/A6G but not enough for C5T/A6G to stimulate expression from promoters containing those sequences; however, the fingers with the sequence QSNK, QSTK and QETR at the four varied positions have the opposite ability to stimulate expression, suggesting that while these fingers retain their high affinity for the desired primary binding site, they exhibit different DNA specificity profiles for their secondary sites.

### Using the One-Hybrid System to Select Binding Sites for which a Zinc Finger has Affinity

We constructed a library of all 262,144 possible nine base sequence target sites centered at the -62 or -55 position

relative to transcription start site in the reporter plasmids pDB2-GFP and pDB2-CAT. The heterogeneity of the nine-base target site library sequence was confirmed by sequencing a few individual members (Table 3) and by sequencing the library as a whole. Cotransformations of these libraries and plasmid pA-21- $\Delta$ QNK underwent selection as described for the library of zinc fingers. We tested our Cm system by plating on increasing concentrations of Cm. When the binding site was centered at -55 or -62 the number of colonies that grew corresponded to 2.9% or 5.5% of the library, respectively. Sequencing of 14 randomly selected individual binding site library members revealed the selected sequences were unlike the recognition site of  $\Delta$ QNK and enriched at certain positions for A/T (Table 3). In addition, when the reporter plasmids were separated from the plasmid expressing RpoA-L21- $\Delta$ QNK and retransformed into cells (without pA-21- $\Delta$ QNK), high levels of Cm resistance were still observed. Similarly, when only the reporter plasmid containing the library of binding sites was present in the cell (no zinc finger present) we found that a similar fraction of the library grew under selective conditions. Similar experiments with the GFP reporter system also revealed a significant population of cells fluorescing higher than the background; however, the fraction of the library was considerably lower (only about 0.4%). Sequencing of seven individual library members with high GFP fluorescence also revealed the selected binding site sequences to be unlike the target site of  $\Delta$ QNK and enriched for A/T at certain positions. This constitutive expression for some of these library members may result from the native RpoA-mediated recruitment of the RNA polymerase subunits, as RpoA is known to bind to A/T rich sequences [25]. This fact limits the ability of the one-hybrid systems to analyze completely random libraries of binding sites. In

addition, it appears that the Cm system, and to a lesser extent the GFP system, cannot readily be used to evolve zinc fingers to bind a small fraction of the 262,144 possible 9-base sequences because of the high background transcription these sequences confer.

**Table 3. Sequences of Randomly Selected Naïve Library Members and Library Members With “Self-Stimulating” Sites Isolated from the Binding Site Library Using the Cm One-Hybrid System**

Naïve or selected?	Binding site centered at	Sequence
naive	-55	5'-GCAATCAC-3'
naive	-55	5'-CATTCCTCC-3'
naive	-55	5'-CCCCGGCAA-3'
naive	-55	5'-CGCCATTGG-3'
naive	-55	5'-CACCACGCT-3'
selected	-55	5'-ACACTATAC-3'
selected	-55	5'-AGAATGTCC-3'
selected	-55	5'-ATAATACAG-3'
selected	-55	5'-TGACTCGTT-3'
selected	-55	5'-TGACCACAC-3'
selected	-55	5'-TGACTCGCC-3'
selected	-55	5'-GAACGACCC-3'
selected	-55	5'-CCACGGGTT-3'
selected	-62	5'-ACATTACA-3'
selected	-62	5'-ACACTCGGT-3'
selected	-62	5'-ACAATGAAA-3'
selected	-62	5'-ACAATGAG-3'
selected	-62	5'-ATCATATTG-3'
selected	-62	5'-ACAATTACC-3'

## DISCUSSION

We have developed a bacterial one-hybrid system for interrogating and selecting zinc finger-DNA interactions *in vivo*. This system is derived from a previously described bacterial one-hybrid system [13, 19]. Our system contains the RpoA[1-248] fragment of  $\alpha$  subunit of RNA polymerase fused *via* a 21 amino acid linker to a zinc finger. This gene fusion is under the control of the *lacUV5* promoter on a pBR322-derived plasmid. A second compatible plasmid contains the target DNA binding site located at position -55 within a promoter upstream from the reporter gene. The use of chloramphenicol acetyltransferase as the reporter gene allows for single-step isolation of zinc fingers that bind the target site on plates containing 500  $\mu$ g/ml of Cm. The use of the green fluorescent protein as the reporter gene allows for flow cytometric analysis of the library as well as isolation of desired zinc fingers using FACS. Based on the success of our system in selecting zinc fingers for an intended target site and its low false positive frequency of  $\leq 5.3 \times 10^{-9}$ , this bacterial one-hybrid system should find use in the

engineering of zinc fingers with affinity for desired DNA sequences.

Our system shares a number of advantages with a previously developed bacterial two-hybrid system [10] including single step isolation of zinc fingers, the potential to analyze large libraries and selection for zinc fingers in an *in vivo* setting. Our system can be used quickly and easily to select for or report on zinc finger-DNA binding specificity with about 1  $\mu$ M or less dissociation constants for a given DNA binding site. However, our system in its present form cannot preferentially select zinc fingers with high affinity (pM  $K_d$ ) over those with moderate affinity (nM  $K_d$ ), as has been demonstrated with the two-hybrid system [10, 11]. Still, our system offers a number of advantages: (a) our system is less complex, requiring one instead of two proteins to establish the link between zinc finger binding and transcriptional activation, (b) our system does not require a specific auxotrophic *E. coli* strain (c) our selection does not require growth for several days on minimal media, (d) the reporter promoter does not need to be integrated into the episome, and (e) the GFP reporter allows for the facile quantitative analysis of any library or individual clone using flow cytometry and has the potential to be used in a negative selection.

A detailed analysis of the performance of our system revealed two minor drawbacks. First, our analysis of binding-site libraries revealed that a small subset of 9-base sequences (~5% in the Cm system, but only 0.4% in the GFP system) result in zinc finger-independent high expression of the reporter gene, presumably by creating a binding site for the chromosome-encoded RpoA. The GFP system appears to be significantly less prone to “self-stimulating” sites as shown in the binding site library experiments as well as the data in Fig. 2. Still, for either system, any target site should first be tested to confirm it is not among these “self-stimulating” sequences. Since the sequences of these “self-stimulating” sites at -55 and -62 are likely to be different, a self-stimulating site at -55 may not be self-stimulating at -62, though we have not yet tested this hypothesis. A related point is that although our system allows the creation of large binding site libraries, the utility of this, in light of the high frequency of self-stimulating sites, is somewhat limited in the systems' current form. The second caveat of our system is the fact that binding to sites centered at -54 and -52 produces reporter gene expression on par with binding centered at -55. This sets up the possible false identification of zinc finger-DNA binding site pairs. An obvious control for this is to create two different reporter plasmids in which the identity of bases -50, -49 and -48 differ. It is reasonable to expect that the drawbacks described above may be shared by the bacterial two-hybrid system; however, neither issue has been experimentally addressed in that system [10, 11].

Our system was not designed to select for high affinity binding (i.e. selection of zinc fingers with sub-nM dissociation constants over those with nM dissociation constants). The lack of discrimination against moderate affinity binding may arise from the higher copy number of our reporter operon, differences between the two-hybrid and one-hybrid systems or in the nature of the reporter proteins. It may be possible to increase the stringency of our systems by integrating the reporter operon into the episome.

However, our system is likely to be most useful for a quick but quantitative evaluation of the specificity of zinc fingers *in vivo*, an application we are currently developing.

We demonstrated our system by selecting zinc fingers that can bind to the site 5'-GGGGCAGAA-3' from a library of zinc finger variants derived from zinc finger  $\Delta$ QNK in which the positions -1, 2, 3 and 6 in the  $\alpha$ -helix of the middle zinc finger were randomized (Table 2). In  $\Delta$ QNK, the amino acids at these positions are Q, S, D, and K, respectively. The sequences of the fingers that were selected by our system were very consistent with the sequence of  $\Delta$ QNK and sequences of Zif268 derived zinc fingers selected by other methods for the triplet GCA [4, 26, 27]. Our selected zinc fingers showed absolute conservation for a basic residue (K or R) at position 6, a preponderance of D or T at position 3 and a strong preference for Q at position -1. Position 2 was often S or T but exhibited more variation than other positions. This is consistent with previous reports that found that a 'S' residue at position 2 is not, in general, responsible for determining binding site preferences [28]. The difference in total cell fluorescence between different zinc fingers in the GFP system, with some sequences resulting in over 3-fold higher fluorescent levels than the original  $\Delta$ QNK sequence, is interesting but the reasons underlying these differences are not known. The increased fluorescence may reflect increases in zinc finger affinity or differences in RpoA-ZF fusion production.

#### ACKNOWLEDGEMENTS

We thank Ann Hochschild and Simon L. Dove (Harvard Medical School) for providing the precursor strains and plasmids of the one-hybrid system and Brendan P. Cormack (Johns Hopkins University) for the GFPmut2 plasmid. We also thank Thomas J. Reichel and Rebecca L. Fagan for their contributions to this work. This study was made possible in part through the support of National Institutes of Health grant GM 53923 (S.C.) and National Science Foundation grant BES 0239088 (M.O.) as well as support from The W.W. Smith Charitable Trust (M.O.) and the Maryland Cigarette Restitution Fund Research Grant to the Johns Hopkins Medical Institutions (FY 03) (M.O.).

#### REFERENCES

- [1] Tan, S.; Guschin, D.; Davalos, A.; Lee, Y.L.; Snowden, A.W.; Jouvenot, Y.; Zhang, H.S.; Howes, K.; McNamara, A.R.; Lai, A.; Ullman, C.; Reynolds, L.; Moore, M.; Isalan, M.; Berg, L.P.;

- Campos, B.; Qi, H.; Spratt, S.K.; Case, C.C.; Pabo, C.O.; Campisi, J.; Gregory, P.D. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 11997-12002.
- [2] Segal, D.J.; Barbas, C.F., 3rd. *Curr. Opin. Biotechnol.*, **2001**, *12*, 632-637.
- [3] Kandavelou, K.; Mani, M.; Durai, S.; Chandrasegaran, S. Engineering and applications of chimeric nucleases. In *Nucleic Acids and Molecular Biology*, A. M. Pingoud ed's Berlin, Germany: Springer Verlag; 413-434, **2004**.
- [4] Choo, Y.; Klug, A. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 11163-11167.
- [5] Beerli, R.R.; Dreier, B.; Barbas, C.F., 3rd. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 1495-1500.
- [6] Dreier, B.; Segal, D.J.; Barbas, C.F. 3rd. *J. Mol. Biol.*, **2000**, *303*, 489-502.
- [7] Greisman, H.A.; Pabo, C.O. *Science*, **1997**, *275*, 657-661.
- [8] Bartsevich, V.V.; Juliano, R.L. *Mol. Pharmacol.*, **2000**, *58*, 1-10.
- [9] Blancafort, P.; Magnenat, L.; Barbas, C.F. *Nat. Biotechnol.*, **2003**, *21*, 269-274.
- [10] Joung, J.K.; Ramm, E.I.; Pabo, C.O. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 7382-7387.
- [11] Hurt, J.A.; Thibodeau, S.A.; Hirsh, A.S.; Pabo, C.O.; Joung, J.K. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 12271-12276.
- [12] Dove, S.L.; Joung, J.K.; Hochschild, A. *Nature*, **1997**, *386*, 627-630.
- [13] Hu, J.C.; Kornacker, M.G.; Hochschild, A. *Methods*, **2000**, *20*, 80-94.
- [14] Dieffenbach, C.; Dveksler, G. *PCR Primer: A Laboratory Manual*. New York: Cold Spring Harbor Press; **1995**.
- [15] Smith, J.; Berg, J.M.; Chandrasegaran, S. *Nucleic Acids Res.*, **1999**, *27*, 674-681.
- [16] Paschon, D.; Ostermeier, M. *Methods Enzymol.*, **2004**, *388*, 103-116.
- [17] Pavletich, N.P.; Pabo, C.O. *Science*, **1991**, *252*, 809-817.
- [18] Abel-Santos, E.; Scott, C.P.; Benkovic, S.J. *Methods Mol. Biol.*, **2003**, *205*, 281-294.
- [19] Dove, S.L.; Hochschild, A. *Cold Spring Harb. Symp. Quant. Biol.*, **1998**, *63*, 173-180.
- [20] Shaywitz, A.J.; Dove, S.L.; Kornhauser, J.M.; Hochschild, A.; Greenberg, M.E. *Mol. Cell Biol.*, **2000**, *20*, 9409-9422.
- [21] Starling, A.L.; Ortega, J.M.; Gollob, K.J.; Vicente, E.J.; Andrade-Nobrega, G.M.; Rodriguez, M.B. *Genet. Mol. Res.*, **2003**, *2*, 124-135.
- [22] Mayer, G.; Launhardt, H.; Munder, T. *Biotechniques*, **1999**, *27*, 86-88, 92-84.
- [23] Cormack, B.P.; Valdivia, R.H.; Falkow, S. *Gene*, **1996**, *173*, 33-38.
- [24] Desjarlais, J.R.; Berg, J.M. *Proc. Natl. Acad. Sci. USA*, **1993**, *90*, 2256-2260.
- [25] van Ulsen, P.; Hillebrand, M.; Kainz, M.; Collard, R.; Zulianello, L.; van de Putte, P.; Gourse, R.L.; Goosen, N. *J. Bacteriol.*, **1997**, *179*, 530-537.
- [26] Segal, D.J.; Dreier, B.; Beerli, R.R.; Barbas, C.F., 3rd. *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 2758-2763.
- [27] Liu, Q.; Xia, Z.; Zhong, X.; Case, C.C. *J. Biol. Chem.*, **2002**, *277*, 3850-3856.
- [28] Kim, C.A.; Berg, J.M. *J. Mol. Biol.*, **1995**, *252*, 1-5.