

Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method

Xiangdong Meng¹, Stacey Thibodeau-Beganny³, Tao Jiang^{3,4}, J. Keith Joung^{3,4} and Scot A. Wolfe^{1,2,*}

¹Program in Gene Function and Expression, ²Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation St, Worcester, MA 01605 USA, ³Molecular Pathology Unit, Center for Cancer Research and Center for Computational and Integrative Biology, Massachusetts General Hospital, 149 13th Street, 7th floor, Charlestown, MA 02129 USA and ⁴Department of Pathology, Harvard Medical School, Boston, MA 02115 USA

Received February 9, 2007; Revised April 28, 2007; Accepted April 30, 2007

ABSTRACT

The C₂H₂ zinc finger is the most commonly utilized framework for engineering DNA-binding domains with novel specificities. Many different selection strategies have been developed to identify individual fingers that possess a particular DNA-binding specificity from a randomized library. In these experiments, each finger is selected in the context of a constant finger framework that ensures the identification of clones with a desired specificity by properly positioning the randomized finger on the DNA template. Following a successful selection, multiple zinc-finger clones are typically recovered that share similarities in the sequences of their DNA-recognition helices. In principle, each of the clones isolated from a selection is a candidate for assembly into a larger multi-finger protein, but to date a high-throughput method for identifying the most specific candidates for incorporation into a final multi-finger protein has not been available. Here we describe the development of a specificity profiling system that facilitates rapid and inexpensive characterization of engineered zinc-finger modules. Moreover, we demonstrate that specificity data collected using this system can be employed to rationally design zinc fingers with improved DNA-binding specificities.

INTRODUCTION

The Cys2His2 zinc finger is the most abundant class of DNA-binding domain found in human transcription factors (1). The considerable expansion of this domain in higher eukaryotes is likely related to its modularity and to its ability to specify a diverse range of different DNA sequences. Multiple research groups have shown that individual fingers can be re-engineered to recognize novel target sequences and that synthetic zinc-finger proteins (ZFPs) composed of multi-finger arrays fused to effector domains can be used to modify gene expression or genomic structure in human cells (2–6). The ultimate utility of engineered zinc fingers will depend upon developing the capability to rapidly generate synthetic ZFPs that are highly specific for their intended *in vivo* target sequence.

Although various methods for characterizing the DNA-binding specificity of a ZFP have been previously described, each of these methods has certain limitations that make them less than optimal for higher-throughput applications. *In vitro* approaches for determining the DNA-binding specificity of a transcription factor, such as Systematic Evolution of Ligands through EXponential enrichment (SELEX) or protein-binding microarrays (7,8), require the purification of each protein of interest and either multiple rounds of selection or specialized reagents and equipment; these requirements create barriers to the utilization of these systems for high-throughput applications. Other approaches that seek to comprehensively assess all possible variants of a given

*To whom correspondence should be addressed. Tel: 508 856 3953; Fax: 508 856 5460; Email: scot.wolfe@umassmed.edu

Present address:

Tao Jiang, Department of Biochemistry, Tufts University School of Medicine, Jaharis 601C, 150 Harrison Ave., Boston, MA 02111 USA

© 2007 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

binding site, such as ELISA or yeast two-hybrid reporter assays, are useful for the analysis of short binding sites (9–11), but they become cumbersome when the number of positions varied in a target sequence increases to more than 4 bases (i.e.—more than $4^4 = 256$ combinations). Recently, we described a bacterial one-hybrid (B1H) system that provides a simple method for determining the DNA-binding specificity of a transcription factor (12,13). This methodology requires only basic molecular biology expertise to perform and it does not require protein purification. However, like SELEX, this approach requires the sequencing of many binding site clones to generate a composite recognition motif for a transcription factor, thereby decreasing its utility for high-throughput applications. The number of sequences required to generate an informative motif is inversely dependent on the length of the binding site recognized by the factor, but is also dictated by the nuances in specificity that need to be detected, e.g. 20 sequences will provide a rough approximation of the 10 base pair DNA-binding specificity of the Zif268 zinc-finger domain.

To facilitate its use in high-throughput applications, we modified the B1H system so that base pair preferences present in a pool of selected binding sites can be assessed in a single sequencing reaction. Analysis of binding sites in this manner is possible if all of the selected sequences in a population are in register with one another. With this restriction, the summation of the sequencing outputs from each clone within the population will be coherent. The register of the sequences can be fixed by using ‘constant’ zinc fingers of defined specificity to anchor the zinc finger to be characterized over the randomized portion of the experimental binding site (Figure 1). This pre-positioning ensures that the selected portion of each binding site is in an identical position in each DNA molecule. Gogos and colleagues (14) demonstrated the feasibility of this type of approach for naturally occurring ZFPs when using SELEX to characterize the DNA-binding specificities of individual fingers present in different splice variants of the transcription factor CF2.

In this report, we describe the implementation and validation of a fixed binding site library approach in the context of the B1H site-selection system. We use this modified B1H system to profile the specificities of 34 engineered zinc-finger domains obtained using a previously described ‘bi-partite’ zinc finger selection strategy (15). We also demonstrate how specificity information generated from this method can be used to further optimize the binding specificities of engineered multi-finger domains.

MATERIALS AND METHODS

Construction of a fixed DNA-binding site library to characterize the specificities of artificial ZFPs

The ‘Zif23*-binding site library’ (Figure 1A) was constructed using a specially synthesized oligonucleotide that minimizes bias within its randomized regions (Integrated DNA Technologies, Coralville, IA, USA). NotI and Ascl restriction sites are present at each end of the library

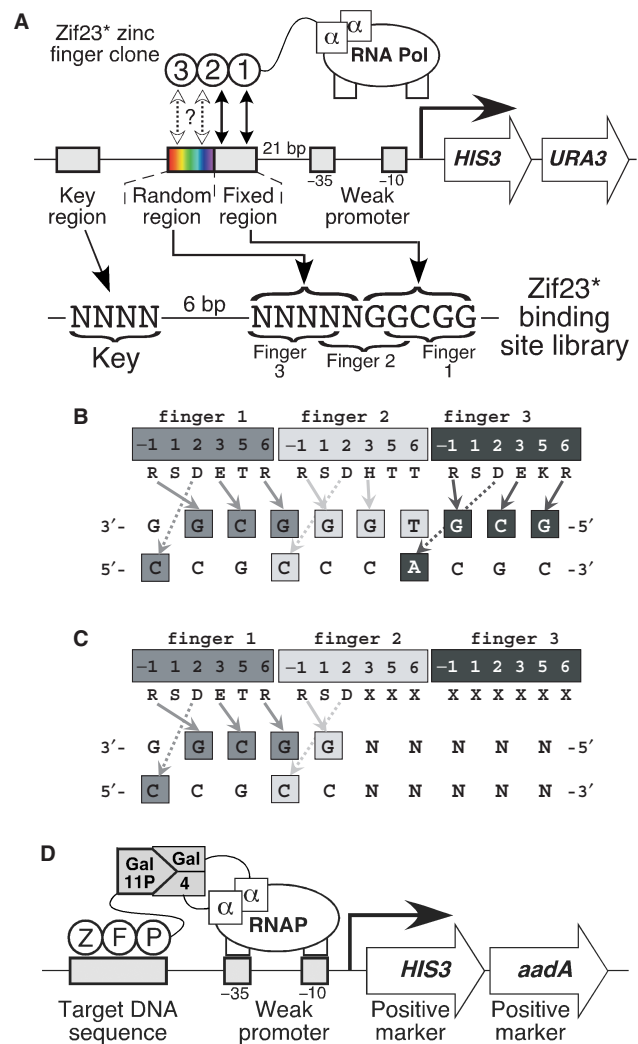


Figure 1. Schematics of the B1H binding site selection system and the ‘bi-partite’ ZFP B2H selection system. (A) The B1H Zif23* binding site library in pH3U3 reporter contains two randomized regions: a 5 bp portion of the zinc-finger binding site and the clone discrimination Key. The ZFP-binding site is positioned 21 bp upstream of the -35 box of the promoter, an optimal binding site position for reporter gene activation in the B1H system. The position of the fingers relative to the randomized region is dictated by the anchor fingers in the ZFP and by the constant portion of the binding site (GGCGG). Each ZFP is expressed as a fusion to the α -subunit of RNAP. The α -ZFPs and members of the Zif23 binding site library are combined such that each cell contains only one member of the binding site library. If an α -ZFP efficiently recognizes the binding site present in the reporter, it will activate reporter gene expression thereby permitting the cells containing these binding sites to survive on selective medium. (B) Schematic of the base-specific contacts observed in the Zif268 structure, where contacts to the primary strand of DNA are indicated as solid arrows and contacts to the complementary strand are indicated as dotted arrows (22,23). (C) Schematic of the randomization strategy for the Zif23* library in which recognition helix residues in Zif268 fingers 2 and 3 are partially randomized (denoted by an X). Because the non-randomized portions of fingers 1 and 2 bind to the sequence 5'-GGCGG-3', the Zif23* library can be used to identify variants capable of binding to sequences of the form 5'-NNNNNGGCGG-3'. (D) In the B2H selection system, RNAP recruitment is mediated by an interaction between a Gal11P fragment fused to the ZFP and a Gal4 fragment fused to the RNAP α -subunit. Like the B1H system, if Gal11P-ZFP binds its target DNA sequence it will recruit RNAP to the weak promoter to activate transcription of a pair of reporter genes.

oligonucleotide (5'-CCGGGCGGCCGCTNNNNAAAA ATNNNNNGGCGGCCGGGTAGGCGCGCCGAATT C-3') for cloning into the pH3U3 reporter vector (12,13). The complementary strand was generated by primer extension using a complementary primer (5'-GAATT CGGCGCGCTACCC-3') in a 50 μ l reaction volume containing 1 \times Taq DNA polymerase buffer (NEB), 0.3 mM dNTP, 10 μ M library oligonucleotides, 30 μ M primer, and 2 units of Taq DNA polymerase (NEB) in a single reaction cycle (95°C for 2 min, 60°C for 3 min, and 72°C for 10 min). Following extension, the duplex DNA product was purified on a 3% TAE low-melting temperature agarose gel, and this DNA was recovered by electroelution (16). The duplex library was ethanol-precipitated to remove excess salts and digested with NotI and AscI sequentially. The digested product was purified by the same procedure described above for the double-stranded library DNA. The pH3U3 reporter vector was similarly digested with NotI and AscI. This digested backbone was purified on a 1% TAE agarose gel and isolated using a QIAquick Gel Extraction Kit (Qiagen). Test ligation reactions of different ratios of reporter plasmid and library insert were performed to optimize the ligation conditions. The optimized plasmid/insert ratio (1:6) was used to ligate the digested library into the pH3U3 reporter vector in 20 μ l containing 1 \times T4 DNA ligase buffer (NEB) and 0.5 μ l T4 DNA ligase (400 units/ μ l NEB) at room temperature for 30 min. This ligation mixture was ethanol-precipitated, washed with 70% ethanol, air-dried, resuspended in 10 μ l of ddH₂O, and then used to electroporate 70 μ l of electrocompetent XL1-Blue cells (Stratagene). Electroporated cells were immediately recovered for 1 h in 10 ml SOC at 37°C with shaking at 120 r.p.m. Following recovery, 20 μ l of the culture was removed and 5 μ l drops of a series of 10-fold serial dilutions were placed on a 2 \times YT plate containing 30 μ g/ml kanamycin to determine the number of independent transformants in the library. Kanamycin was added to the remaining cell culture to a final concentration of 30 μ g/ml and growth of the culture was continued at 37°C with shaking at 200 r.p.m. After 4 h of amplification, the cells were collected by centrifugation at 5000g for 10 min and a plasmid miniprep protocol was used to isolate pooled binding site library DNA from half of the cells (QIAprep Miniprep Kit, Qiagen). A fraction of the recovered library was transformed into the bacterial selection strain (US0 Δ pyrF Δ hisB) and counter-selection was performed on 10 150 mm \times 15 mm round plates containing 2.5 mM 5-FOA YM agar medium to remove self-activating clones from the library (12). The number of transformants subjected to counter-selection ($\sim 6 \times 10^6$) exceeds the theoretical size of the library by more than 20-fold. Surviving colonies were washed off each plate using 10 ml 2 \times YT medium, 15–20 3 mm sterile glass beads and gentle agitation. The recovered cells were pelleted by centrifugation (5000g) at 4°C for 15 min. Half of these cells were resuspended in 3.75 ml P1 buffer from the QIAprep Miniprep kit and then aliquoted into 15 1.5 ml-ependorf tubes. Plasmid DNA was isolated using the standard Qiagen miniprep protocol except that the cell lysates were loaded onto five miniprep columns

in the purification step and the optional PB wash step was performed.

Construction of a randomized zinc-finger library of two-finger modules

We used a previously described bacterial two hybrid (B2H) selection system to identify artificial C2H2 zinc fingers from a randomized library which was built using a strategy described by Isalan and colleagues (15). In this approach, randomized libraries are constructed by partially randomizing the recognition helix residues of two zinc fingers from the three-finger Zif268 DNA-binding domain. In these libraries, the unaltered portions of the Zif268 domain anchor the randomized finger recognition helices over a target DNA sequence. We constructed a C2H2 ZF library analogous to the 'Zif23 library' previously described by Isalan and colleagues (15) in which three residues within finger 2 and six residues within finger 3 of Zif268 are partially randomized (Figure 1B and C). Our library differs from the original Zif23* library in that it possesses greater variability in the amino acid residues present at each randomized position (Supplementary Figure 1). We termed this improved library the 'Zif23* finger library'. In contrast to Isalan and colleagues (15), we then used the B2H system previously described by Joung and colleagues (17,18) (instead of phage display) to isolate artificial C2H2 ZFs with novel specificities from this Zif23* finger library. The Zif23* finger library was built using cassette mutagenesis in a plasmid which expresses each library member as a Gal1P-fusion, thereby permitting the DNA-binding activity of each library member to be interrogated using the B2H selection system (Figure 1D). Our new Zif23* finger library was constructed from $\sim 1.3 \times 10^8$ independent transformants and the library contains reasonable distributions of amino acids within the randomized recognition helix positions as determined by inspection of 10 random candidates from our Zif23* library (data not shown). Zinc finger proteins within the Zif23* library recognize a target DNA sequence of the form 5'NNNNNGGCG(g/t)3' in which the constant recognition helix residues from Zif268 in finger 1 and in part of finger 2 bind to the constant bases of this sequence. Seventeen complementary selection strains that each harbor a different binding site of the form 5'XXXXXGGCGT3' (Table 1) were constructed and sequence-verified (data not shown). Briefly, each target DNA sequence was cloned upstream of a weak promoter controlling the expression of a co-cistronic *HIS3/aadA* gene cassette and then transferred to a single copy F' episome by homologous recombination using modified and miniaturized versions of previously described methods (17,19).

Profiling the DNA-binding specificities of engineered zinc-finger modules

The B1H selection system was used to rapidly characterize the DNA-binding specificities of ZFPs identified from selection experiments performed with the Zif23* finger library (13). Briefly, each ZFP to be characterized was

Table 1. Recognition helix sequences of selected C2H2 ZF proteins used for profiling

Zif23* Modules	Finger 1 -1123456	Finger 2 -1123456	Finger 3 -1123456	Target site 5' to 3'
B1	RSDELTR	RSDHLTR	DNRNRIR	GACTG GGCGT
B2	RSDELTR	RSDNLKK	DSTNRIR	GACTG GGCGT
C1	RSDELTR	RSDALKN	ARNDRIR	GCTAC GGCGT
C2	RSDELTR	RSDTLKN	TSGIRKI	GCTAC GGCGT
D1	RSDELTR	RSDSLKV	QSSNRTV	GAATT GGCGT
D2	RSDELTR	RSDVLIT	NNSNRTR	GAATT GGCGT
E1	RSDELTR	RSDVLKI	KSGNRIR	GAGCT GGCGT
E2	RSDELTR	RSDALKE	KSSNRKK	GAGCT GGCGT
F1	RSDELTR	RSDGLTR	RNSNRKR	GAGGG GGCGT
F2	RSDELTR	RSDGLTN	RKDNRRK	GAGGG GGCGT
G1	RSDELTR	RSDGLKR	QNQHRIR	GATGA GGCGT
G2	RSDELTR	RSDNLTR	GSQNRTR	GATGA GGCGT
H1	RSDELTR	RSDHLTT	RRATRKR	GCACG GGCGT
H2	RSDELTR	RSDHLKD	QRSTRKN	GCACG GGCGT
I1	RSDELTR	RSDILKN	DRNSRKR	GCCAT GGCGT
I2	RSDELTR	RSDSLTR	DRKTRIV	GCCAT GGCGT
J1	RSDELTR	RSDNLKD	RRANRIR	GAAAA GGCGT
J2	RSDELTR	RSDNLKV	RKANRTR	GAAAA GGCGT
K1	RSDELTR	RSDNLKT	QKKNRIR	GGAAA GGCGT
K2	RSDELTR	RSDHLII	QRGHRTR	GGAAA GGCGT
L1	RSDELTR	RSDHLTT	QRSNRIT	GCTCG GGCGT
L2	RSDELTR	RSDNLKD	RSTDRTT	GCTCG GGCGT
M1	RSDELTR	RSDTLTD	HRKNRIA	GCCAC GGCGT
M2	RSDELTR	RSDTLTA	DRRARIR	GCCAC GGCGT
N1	RSDELTR	RSDNLIK	RNDVRIR	GTGTC GGCGT
N2	RSDELTR	RSDDLTV	RRDSRIR	GTGTC GGCGT
O1	RSDELTR	RSDNLIK	NSSHRII	GGTTC GGCGT
O2	RSDELTR	RSDGLII	TSSHRII	GGTTC GGCGT
P1	RSDELTR	RSDHLKQ	QNSNRKT	TAACG GGCGT
P2	RSDELTR	RSDHLKT	QNSNRTK	TAACG GGCGT
Q1	RSDELTR	RSDTLKN	RSQHRTA	TGGAC GGCGT
Q2	RSDELTR	RSDSLKN	KRANRTI	TGGAC GGCGT
R1	RSDELTR	RSDSLTR	TRHRTG	TGTGT GGCGT
R2	RSDELTR	RSDSLTR	QRASRKI	TGTGT GGCGT

Recognition helix positions are numbered relative to the start of the zinc finger alpha-helix. The amino acids in bold are those present at recognition helix positions randomized in the Zif23* finger library. The target site (bold) used for each selection is listed to the right of each Zif23* clone where the constant 3' portion of the sequence (GGCGT) is specified by the constant residues in finger 1 and 2.

excised from its B2H expression vector by simultaneous digestion with NotI (NEB) and XbaI (NEB). This DNA fragment, which contains the engineered and anchor fingers, was purified on a 1% agarose gel, isolated using a QIAquick Gel Extraction Kit (Qiagen) and sub-cloned into an α -fusion expression vector (pB1H1) using the unique NotI and AvrII restriction sites (12,13). Electrocompetent *US0 Δ pyrF Δ hisB* cells containing each α -ZFP clone in pB1H1 were prepared and transformed with ~100 ng of the Zif23* binding site library. Upon further investigation we found that the selection procedure can be streamlined by cotransforming the selection strain *US0 Δ pyrF Δ hisB* with the Zif23* binding site library (100 ng) and the α -ZFP expression vector (1 μ g) using electroporation. The transformants were recovered in 1 ml SOC for 1 h at 37°C on a roller wheel at 80 r.p.m. Following recovery, the cells were pelleted by centrifugation at 4000g for 10 min at 25°C and resuspended in 1 ml NM medium (a modified minimal medium (16)) supplemented with 0.1% histidine, 30 μ g/ml kanamycin and 30 μ g/ml chloramphenicol. The cells were grown for an additional hour at 37°C on a roller wheel. Following the NM media recovery step, the cells were transferred into a sterile 1.5 ml eppendorf tube and pelleted at 18 000g for

1 min at 25°C. The cell pellet was washed by resuspension in 1 ml sterile water and pelleted again at 18 000g for 1 min at 25°C. This wash step was repeated two additional times to remove all traces of histidine and then the cells were resuspended in 0.5 ml NM medium lacking histidine. The cells were plated on NM medium plates (150 \times 15 mm) containing 30 μ g/ml kanamycin, 30 μ g/ml chloramphenicol, 10 μ M IPTG, and either 1 or 2 mM 3-AT at a density of $\sim 5 \times 10^6$ transformants per plate (16). These plates were incubated at 37°C for 1–2 days until well-defined colonies appeared. Surviving colonies were washed off the plate in 10 ml of 2 \times YT medium using 15–20 3 mm sterile glass beads and gentle agitation by hand. Harvested cells were pelleted by centrifugation at 5000 g for 10 min at 4°C. Plasmid DNA was isolated from the collected cells, or only a fraction of the cells if the cell pellet size exceeded the amount of cells recovered from a typical 5 ml saturated culture, using a QIAprep Miniprep Kit treating the cell pellet as a single plasmid miniprep. The sequence preference of each ZFP was determined by sequencing the harvested pH3U3 reporter plasmids as a pool using a primer of sequence 5'-GAAATATGTATCCGCTCATGAC-3'. DNA prepared from the surviving colonies contains a mixture of both the pH3U3 reporter plasmids

and the pB1H1 α -ZFP expression plasmid, and therefore approximately four times the amount of DNA typically required for a standard sequencing reaction was used. The pH3U3 reporter vector is a low copy number plasmid and occasionally the sequencing results from the plasmid mixture were unsatisfactory due to a high level of baseline noise. When this occurred the binding site region from the pool of selected pH3U3 plasmids was PCR-amplified by primers that flank this region (forward primer 5'-GAA ATATGTATCCGCTCATGAC-3'; reverse primer 5'-CC AGAGCATGTATCATATGGTCCAGAAACCC-3') in a 50 μ l reaction with 1 \times Taq DNA polymerase buffer (NEB), 0.3 mM dNTP, 0.5 μ M of each primer, 2 u of Taq DNA polymerase (NEB) using 30 amplification cycles (94°C for 20 s, 55°C for 30 s and 72°C for 30 s) followed by a 6 min final extension at 72°C. The PCR product from this amplification was isolated by running a 1% TAE agarose gel and purified using QIAquick Gel Extraction Kit (Qiagen). This purified PCR product was sequenced using the forward primer. In retrospect, sequencing the PCR products generated from the pool of selected binding sites provided a more robust method for the analysis of the pooled binding sites.

Optimization of the DNA-binding specificities of ZFP modules based on their specificity profiles

Two chimeric ZFP modules (B2* and D1*) were generated by combining the recognition helices from two different pairs of proteins (B1/B2 and D1/D2). Single amino acid substitutions were introduced into clone B2 and D1 using a two-step PCR procedure with the mutant codon integrated into the overlapping primers (primer information is available upon request). Each newly assembled ZFP module was digested with NotI and XbaI (NEB) and inserted into pB1H1 between the NotI and AvrII sites for characterization using the B1H-binding site selection system as described above.

RESULTS

Construction of a fixed binding site library for the profiling of DNA-binding specificities

We tested the feasibility of using a fixed binding site library for the characterization of zinc-finger modules by using it to determine the binding specificity of two of the three zinc fingers from the transcription factor Zif268. To accomplish this, we constructed a 'Zif23*-binding site library' that contained two elements: a 10 base pair Zif268 binding site in which five base pairs contacted by fingers 2 and 3 are randomized and five base pairs recognized by finger 1 and a portion of finger 2 are fixed and a clone discrimination 'Key' (Figure 1A), which is an additional four base pair randomized segment upstream of the Zif268 binding site that serves as an internal control to detect any sequence bias in the selected clones that is not due to interaction with the ZFP. The Key region also allows one to determine whether identical binding sites obtained more than once from a given selection are independent isolates should the sequencing of individual clones be desired. The sequences flanking each randomized region

were chosen to avoid the creation of alternate binding sites for fingers 1 and 2 of Zif268 that might complicate interpretation of the selection results.

Synthetic oligonucleotides encoding the Zif23*-binding site library were introduced into the reporter plasmid pH3U3 upstream of a weak promoter that drives expression of the co-cistronic yeast *HIS3* and *URA3* reporter genes (Figure 1A)(13). The constructed library contains $\sim 2 \times 10^6$ unique clones, ensuring significant oversampling of the 262144 ($=4^9$) potential sequences encoded within the library. Although control experiments demonstrated that the frequency of auto-activating sequences (i.e. binding sites that activate transcription of the weak promoter in the absence of a test ZFP) in our library is only $\sim 0.005\%$, we nonetheless used 5-FOA counterselection to reduce the number of auto-activating sequences as previously described. To do this, the library was transformed into our selection strain (US0 Δ *pyrFahisB*) and plated on YM minimal media containing 2.5 mM 5-FOA, thereby eliminating self-activating sequences due to toxicity associated with expression of the *URA3* reporter gene (12,13,16). DNA from the $\sim 6 \times 10^6$ colonies surviving on these plates was recovered as a pool to generate a purified binding site library for our binding site selection experiments. The purified library was sequenced as a pool and the resulting chromatogram reveals no significant bias within the randomized positions in either the Key or ZFP-binding site regions (Figure 2).

The DNA-binding specificity of fingers 2 and 3 of Zif268 was determined by transforming a selection strain (already harboring a plasmid which expresses an α -Zif268 hybrid protein) with the Zif23*-binding site library ($\sim 2.5 \times 10^6$ transformants) and then selecting for *HIS3* reporter gene expression by plating the transformants on histidine-deficient NM medium containing 2 mM 3-AT (a competitive inhibitor of the *HIS3* enzyme). This selection yielded ~ 3000 colonies after 36 h of growth; these colonies were pooled and the mixture of reporter vectors from this population was isolated and sequenced. The previously defined consensus binding site (5'-GCG(T/G)GGGCGG-3') of Zif268 (20,21) is readily apparent in the sequencing chromatogram from the selected pool of clones (Figure 2; randomized base positions highlighted in bold). This result demonstrates that the B1H-based binding site profiling method can be used to rapidly determine the DNA-binding specificity of a zinc-finger domain using only a single DNA sequencing reaction.

Engineering of Zif268 variants with novel DNA-binding specificities

To further test the utility of the B1H profiling method, we wished to determine the specificities of a series of engineered zinc-finger domains using this approach. To obtain a large number of multi-finger domains with different DNA-binding specificities, we constructed a library of Zif268 variants in which three of the recognition helix residues in finger 2 (positions 3, 5 and 6 numbered relative to the beginning of the

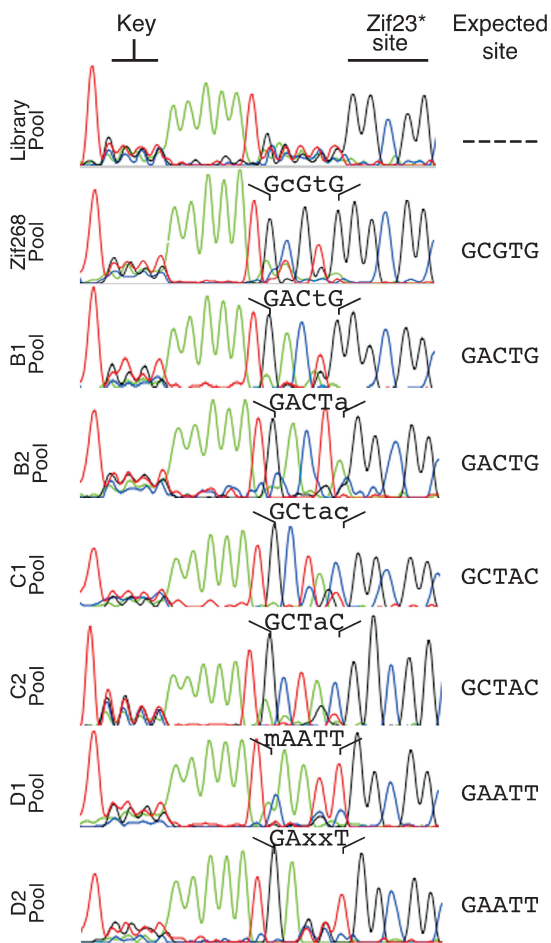


Figure 2. Specificity profiles of six representative Zif23* clones characterized using the BIH selection system. The nucleotides that are enriched at positions within the binding site of each selected pool are listed above the chromatogram. Capital letters indicate positions judged to have a strong sequence bias (>3-fold larger than the peaks for other bases), lower case letters indicate a weaker bias (2–3-fold greater than the other peaks), r, y, m, k, s or w indicate pairs of peaks that are enriched by 2-fold or greater and x indicates no obvious bias at a position (<2-fold difference). The expected target DNA-binding site is listed to the right of each chromatogram. (Pool names correspond to the clones in Table 1). Note that prior to selection the pooled Zif23* DNA-binding site library displays no significant nucleotide bias in the randomized positions (Library Pool) and the Key remains diverse in each of the pools of binding sites following selection.

recognition helix) and six of the recognition helix residues in finger 3 (positions –1, 1, 2, 3, 5 and 6) were randomized (Figure 1B). As previously described by Isalan and colleagues (15), this randomized library can be used to select novel three-finger domains capable of binding to potential target DNA sequences of the form 5'-NNNNNGGCG(t/g)-3' (where N can be any of the four possible nucleotides; Figure 1B). Using a previously described B2H selection system (17,18), we performed selections with this library against 17 different binding sites to identify novel three-finger domains capable of binding to each of these sequences. A small number (~10) of the hundreds of surviving clones from each of the 17 different selections were sequenced to identify the amino acid residues present at each randomized position and to

estimate the diversity that remained in the selected pool of clones. As anticipated, for each target DNA sequence, the sequences of multi-finger domains identified by selection closely resemble one another; in addition, comparison of fingers from different selections reveals significant differences in the amino acid sequences obtained for each binding site, suggesting that the selection protocol identified members of the library with different DNA-binding activities (Supplementary Figure 2).

Profiling the DNA-binding specificities of engineered Zif268 variants

We used the Zif23*-binding site library and the BIH profiling method to characterize two multi-finger proteins isolated from each of the 17 selections described above (34 proteins in total; Table 1). The Zif23*-binding site library is compatible with these engineered zinc-finger domains because the unmodified portions of fingers 1 and 2 from the Zif268-framework serve as anchors to position the re-engineered portions of fingers 2 and 3 over the randomized base positions within the binding site library. To perform these binding site selections, a DNA fragment encoding each engineered three-finger ZFP was introduced into the pBIH1 expression vector to permit its expression as an RNA polymerase α -subunit fusion(13). To determine the DNA-binding specificity of each engineered domain, $\sim 5 \times 10^6$ US0 *ApyrF**AhisB* cells containing both the Zif23* binding site library and one of the 34 different α -ZFP expression vectors were plated on NM minimal medium containing 1 or 2 mM 3-AT. Typically 10^3 – 10^4 colonies appeared after ~ 36 h of growth at 37°C; the plasmid DNAs isolated from pools of these clones were sequenced. The vast majority of the pools displayed nucleotide bias within the randomized portion of the Zif268 binding site region that is partially or completely consistent with the expected specificity of each re-engineered Zif268 variant (Figure 2 and Supplementary Figure 3). Of the 34 Zif268 variants examined using this approach, >80% displayed good specificity for their target sequence (defined as a match between the expected target sequence and the tallest peak in sequence chromatogram at three or more of the five positions; Table 2).

Although some of the engineered variants show partially degenerate specificities, this is not unexpected since many naturally occurring C2H2 ZFs possess partially degenerate specificities (e.g. the known degeneracy in the binding site signature of Zif268 seen in Figure 2). In addition, the B2H selections used to isolate the engineered zinc-finger domains were not performed at the highest possible stringency and only a small number of variants (~10) were sequenced for each selection. Thus, the selections have a very low probability of identifying zinc-finger domains with fully optimized DNA-binding affinities and specificities.

Optimizing the DNA-binding specificities of engineered zinc-finger domains

Because the DNA-binding specificities of some of the 34 engineered zinc-finger domains were not completely optimal, we hypothesized that we might be able to use

Table 2. Specificity of the 34 characterized Zif23* clones

Matches between specificity profile and target sequence ^a	Number of Zif23* clones	Percentage of Zif23* clones
5 bases	7	20.6
4 bases	12	35.3
3 bases	9	26.5
2 bases	4	11.8
1 base	2	5.8

^aUsing the data shown in Figure 2, the base peak with the highest intensity in the chromatogram at each position in the binding site was used to define the specificity of the Zif23* clone at that position.

the specificity data generated using the B1H profiling system to rationally introduce mutations that improve the specificities of these modules. For example, for target binding sites B and D (Figure 2), the two different Zif268 variants we characterized for each site specified different subsets of bases within the intended five base pair target site. If we assume that the engineered three-finger domains utilize a canonical, Zif268-like recognition pattern (22,23), we can infer the amino acid residue in each finger that is responsible for defining the specificity of each base position in the target DNA binding site (Figure 1B). Based on this assessment, we designed new fingers that would be expected to possess improved specificities for target sites B or D. For example, the B1 and B2 three-finger domains were isolated from a selection with the intended target site 5'-GACTG-3'. Our binding site selection results indicate that the B1 module displays good specificity for four bases in its target site but only a weak preference for the T at the fourth position (5'-GACTG-3', Figure 3). The B2 module also displays good specificity but for a different subset of four of the five bases in its target site—it weakly specifies A instead of G in the fifth position (5'-GACTA-3'). If the key recognition residues based on the structure of Zif268 are acting independently at these positions then substituting the histidine at position 3 of finger 2 from the B1 module for the asparagine at the same position in the B2 module should result in a hybrid protein (B2*) with the desired specificity. Information from deterministic and probabilistic recognition codes also predict histidine at position 3 should recognize a guanine in the middle position of a triplet subsite (24–26). Moreover, histidine is the most commonly occurring amino acid at this position in the pool of fingers sequenced to recognize this site (Supplementary Figure 1). As shown in Figure 3, the B1H profiling system shows that the rationally designed B2* module fully specifies all five bases in the 'B' target site 5'-GACTG-3' as expected. A similar result was observed for the D1 and D2 modules isolated from a selection with the target sequence 5'-GAATT-3'. The D1 module specified all of the positions within its binding site except the 5' G (5'-mAATT-3'). The D2 module effectively specified the 5' G in its binding site but poorly specified two of the central positions (5'-GAXxT-3'). Arginine occupies position 6 of the recognition helix in finger 3 of the D2 module, which should specify G, whereas valine

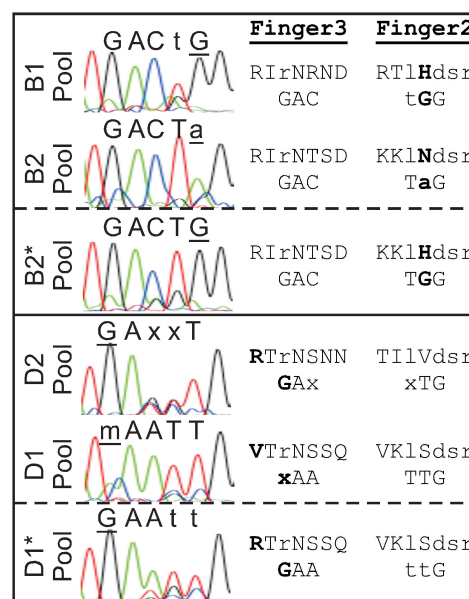


Figure 3. Creating chimeric Zif23* clones with improved specificities. Close-up of the specificity profiles for the characterized clones from the B and D selections (left). Amino acid sequence of the recognition helix for each clone over the DNA sequence specified by the clone based on its specificity profile (right). Upper case amino acids represent positions selected from the randomized Zif23* finger library. Amino acid position of interest and the DNA base it should participate in specifying based on a canonical Zif268 recognition scheme are shown in bold type (22,23). Hybrid proteins were constructed by substituting the amino acid from a clone that properly specifies the desired base into a clone that does not properly specify it. The experimentally determined binding sites of hybrid proteins (B2* and D1*) are shown.

occurs at position 6 of finger 3 in the D1 module, which is not expected to display a strong preference for any particular DNA base based on deterministic and probabilistic recognition codes (24–26). Substitution of arginine for valine at position 6 in the D1 module results in a module, D1*, that displays the desired DNA-binding specificity for the entire binding site (5'GAAtt3'). These experiments demonstrate that the DNA-binding specificity data generated by the B1H profiling system can be used to improve the DNA-binding specificities of engineered ZFPs.

CONCLUSION

We have shown in this report that combining the B1H binding site selection system together with a fixed register binding site library permits a rapid and inexpensive qualitative assessment of the specificity of an engineered ZFP module. Once an appropriate binding site library is constructed and the nucleotide diversity within the randomized region is verified by sequencing to ensure that there is no significant bias that could adversely affect the interpretation of results, the specificity of an individual ZFP module can be determined in a single selection step: a single sequencing reaction performed on the pool of selected binding sites provides the relative distribution of bases present at each position. Since most artificial ZFPs

are selected in a context where one or more fingers are randomized while others are held constant (due to inherent limitations on library complexity) (15,18,27,28), the B1H profiling system is complementary to a typical zinc-finger selection system. In principle, the binding specificities of the most promising modules from a zinc-finger selection can be rapidly determined using this profiling method and then combined with complementary selected modules to generate ZFPs with entirely novel specificities. Following assembly of the engineered modules, the specificity of the full protein can be interrogated using the standard B1H binding site selection system (13,16).

This type of high-throughput analysis should prove particularly powerful for the characterization of large databases of fingers or modules that are selected from a common finger library framework to specify different sequences. Although the experiments described herein have focused on the characterization of one-and-a-half fingers from a three-finger protein, the profiling system should also be amenable to the analysis of individual fingers (re-engineered by either selection and/or design) provided that they are placed in an appropriate framework with constant 'anchoring' fingers (24–26,29). Moreover, because the B1H system also works with other non-zinc-finger DNA-binding domains, it should be possible to employ this method to evaluate other classes of engineered DNA-binding domains, such as homeodomains (30) or homing endonucleases (31), where information about the specificity of only a portion of the domain is desired. It is important to note that the information obtained by sampling a pool of selected sequences using this method should not be considered quantitative, as each surviving clone may contribute a different amount of DNA to the pool due to differences in colony size. Furthermore, the relative peak heights in a chromatogram are susceptible to potential bias due to variability in the sequencing chemistry, such as the sequence-dependent efficiency of termination (32). However, recent improvements in sequencing chemistry and reaction conditions have resulted in much more consistent peak heights throughout a sequencing chromatogram (33,34). Concerns about potential chromatogram bias due to the sequencing chemistry can be alleviated by sequencing the complementary strand of the selected region to verify the observed base ratios (35). It is also worth noting that other selectable markers—GFP and CAT—have been developed and validated for the B1H system and that they may provide useful alternatives for performing these selections (36). For instance, cell sorting using a GFP reporter could be used to isolate all clones above some desired total cell fluorescence threshold to create a pooled population of cells for analysis. Finally, this method will not provide any information about the co-variation of bases within the binding sites for a pool of selected clones, but this information can be obtained by sequencing individual clones from the selection (7,37).

The ability to use the specificity data produced from this technique to engineer zinc-finger chimeras with improved DNA-binding specificities highlights an additional

application of this method. A similar type of specificity optimization of individual finger modules has been employed by the Barbas laboratory using an Enzyme-linked immunosorbent assay (ELISA)-based assay to assess the specificity of fingers for a set of individual target sequences (27,38). However, this ELISA-based assay typically interrogates the specificity on a sequence-by-sequence basis. In contrast, single-pool analysis via a B1H selection instead of interrogating 4^N different sequences has significant advantages with regards to throughput, although this sacrifices a direct comparison between individual sites. We would also note that the simple single residue substitutions that were successfully employed herein to generate chimeras with improved specificity may only work in certain instances because recognition helix residues in ZFPs do not always act independently to specify bases within their binding site. In some cases, there may be context-dependent effects at the level of both the DNA and protein sequence that influence base preferences within the binding site (21). Ultimately, the ability to rapidly and inexpensively characterize the specificity of a large number of zinc-finger clones should provide an important avenue for generating more comprehensive data sets on the DNA-binding specificities of not only zinc fingers but also other families of DNA-binding domains. Such data sets will provide an important basis for creating a more accurate description of the complex determinants for each family that are responsible for sequence-specific DNA-recognition. This technology should also have practical immediate uses in efforts to construct more specific artificial ZFPs for targeted gene regulation (39,40) or modification through the use of tethered nucleases (4,41–43).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Gary Stormo and Michael Brodsky for their helpful suggestions regarding these experiments. Work by S.A.W. and X.M. was supported by the NIH (NIGMS) grant 1R01GM068110 (S.A.W.). Work by S.T.-B. and J.K.J. was supported by start-up funds from the MGH Department of Pathology. J.K.J. is supported by the NIH (R01 GM069906 and R01 GM072621). Funding to pay the Open Access publication charges for this article was provided by the NIH (NIGMS) grant 1R01GM068110 (S.A.W.).

Conflict of interest statement. None declared.

REFERENCES

1. Tupler, R., Perini, G. and Green, M.R. (2001) Expressing the human genome. *Nature*, **409**, 832–833.
2. Tan, S., Guschin, D., Davalos, A., Lee, Y.L., Snowden, A.W., Jouvenot, Y., Zhang, H.S., Howes, K., McNamara, A.R. *et al.* (2003) Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc. Natl Acad. Sci. USA*, **100**, 11997–12002.

3. Blancafort, P., Chen, E.I., Gonzalez, B., Bergquist, S., Zijlstra, A., Guthy, D., Brachat, A., Brakenhoff, R.H., Quigley, J.P. *et al.* (2005) Genetic reprogramming of tumor cells by zinc finger transcription factors. *Proc. Natl Acad. Sci. USA*, **102**, 11716–11721.
4. Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **646**, 646–651.
5. Eberhardy, S.R., Goncalves, J., Coelho, S., Segal, D.J., Berkhout, B. and Barbas, C.F.III. (2006) Inhibition of human immunodeficiency virus type 1 replication with artificial transcription factors targeting the highly conserved primer-binding site. *J. Virol.*, **80**, 2873–2883.
6. Jamieson, A.C., Miller, J.C. and Pabo, C.O. (2003) Drug discovery with engineered zinc-finger proteins. *Nat. Rev. Drug Discov.*, **2**, 361–368.
7. Roulet, E., Busso, S., Camargo, A.A., Simpson, A.J., Mermod, N. and Bucher, P. (2002) High-throughput SELEX-SAGE method for quantitative modeling of transcription-factor binding sites. *Nat. Biotechnol.*, **20**, 831–835.
8. Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W.III and Bulyk, M.L. (2006) Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. *Nat. Biotechnol.*, **24**, 1429–1435.
9. Bae, K.-H. and Kim, J.-S. (2006) One-step selection of artificial transcription factors using an in vivo screening system. *Mol. Cells*, **21**, 376–381.
10. Isalan, M. and Choo, Y. (2001) Rapid, high-throughput engineering of sequence-specific zinc finger DNA-binding proteins. *Methods Enzymol.*, **340**, 593–609.
11. Bae, K.H., Do Kwon, Y., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y. *et al.* (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.*, **21**, 275–280.
12. Meng, X., Smith, R.M., Giesecke, A.G., Joung, J.K. and Wolfe, S.A. (2006) Counter-selectable marker for bacterial-based interaction trap systems. *BioTechniques*, **40**, 179–184.
13. Meng, X., Brodsky, M.H. and Wolfe, S.A. (2005) A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat. Biotechnol.*, **23**, 988–994.
14. Gogos, J.A., Hsu, T., Bolton, J. and Kafatos, F.C. (1992) Sequence discrimination by alternatively spliced isoforms of a DNA binding zinc finger domain. *Science*, **257**, 1951–1955.
15. Isalan, M., Klug, A. and Choo, Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nat. Biotechnol.*, **19**, 656–660.
16. Meng, X. and Wolfe, S.A. (2006) Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nat. Protocols*, **1**, 30–45.
17. Joung, J.K., Ramm, E.I. and Pabo, C.O. (2000) A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **97**, 7382–7387.
18. Hurt, J.A., Thibodeau, S.A., Hirsh, A.S., Pabo, C.O. and Joung, J.K. (2003) Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc. Natl Acad. Sci. USA*, **100**, 12271–12276.
19. Whipple, F.W. (1998) Genetic analysis of prokaryotic and eukaryotic DNA-binding proteins in *Escherichia coli*. *Nucleic Acids Res.*, **26**, 3700–3706.
20. Swirnoff, A.H. and Milbrandt, J. (1995) DNA-binding specificity of NGFI-A and related zinc finger transcription factors. *Mol. Cell. Biol.*, **15**, 2275–2287.
21. Wolfe, S.A., Greisman, H.A., Ramm, E.I. and Pabo, C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.*, **285**, 1917–1934.
22. Pavletich, N.P. and Pabo, C.O. (1991) Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, **252**, 809–817.
23. Elrod-Erickson, M., Rould, M.A., Nekludova, L. and Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
24. Wolfe, S.A., Nekludova, L. and Pabo, C.O. (2000) DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 183–212.
25. Benos, P.V., Lapedes, A.S. and Stormo, G.D. (2002) Probabilistic code for DNA recognition by proteins of the EGR family. *J. Mol. Biol.*, **323**, 701–727.
26. Choo, Y. and Klug, A. (1997) Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.*, **7**, 117–125.
27. Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F.III. (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
28. Greisman, H.A. and Pabo, C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
29. Sera, T. and Uranga, C. (2002) Rational design of artificial zinc-finger proteins using a nondegenerate recognition code table. *Biochemistry*, **41**, 7074–7081.
30. Simon, M.D., Sato, K., Weiss, G.A. and Shokat, K.M. (2004) A phage display selection of engrailed homeodomain mutants and the importance of residue Q50. *Nucleic Acids Res.*, **32**, 3623–3631.
31. Ashworth, J., Havranek, J.J., Duarte, C.M., Sussman, D., Monnat, R.J.Jr, Stoddard, B.L. and Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–659.
32. Ewing, B., Hillier, L., Wendl, M.C. and Green, P. (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.*, **8**, 175–185.
33. Zakeri, H., Amparo, G., Chen, S.M., Spurgeon, S. and Kwok, P.Y. (1998) Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing. *BioTechniques*, **25**, 406–410, 412–404.
34. Korch, C. and Drabkin, H. (1999) Improved DNA sequencing accuracy and detection of heterozygous alleles using manganese citrate and different fluorescent dye terminators. *Genome Res.*, **9**, 588–595.
35. Nurpeisov, V., Hurwitz, S.J. and Sharma, P.L. (2003) Fluorescent dye terminator sequencing methods for quantitative determination of replication fitness of human immunodeficiency virus type 1 containing the codon 74 and 184 mutations in reverse transcriptase. *J. Clin. Microbiol.*, **41**, 3306–3311.
36. Durai, S., Bosley, A., Abulencia, A.B., Chandrasegaran, S. and Ostermeier, M. (2006) A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Comb. Chem. High Through Screen*, **9**, 301–311.
37. Fields, D.S., He, Y., Al-Uzri, A.Y. and Stormo, G.D. (1997) Quantitative specificity of the Mnt repressor. *J. Mol. Biol.*, **271**, 178–194.
38. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas, C.F.III. (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466–29478.
39. Beerli, R.R., Segal, D.J., Dreier, B. and Barbas, C.F.III. (1998) Toward controlling gene expression at will: specific regulation of the erbB-2/HER-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *Proc. Natl Acad. Sci. USA*, **95**, 14628–14633.
40. Liu, P.Q., Rebar, E.J., Zhang, L., Liu, Q., Jamieson, A.C., Liang, Y., Qi, H., Li, P.X., Chen, B. *et al.* (2001) Regulation of an endogenous locus using a panel of designed zinc finger proteins targeted to accessible chromatin regions. Activation of vascular endothelial growth factor A. *J. Biol. Chem.*, **276**, 11323–11334.
41. Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell. Biol.*, **21**, 289–297.
42. Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
43. Beumer, K., Bhattacharyya, G., Bibikova, M., Trautman, J.K. and Carroll, D. (2006) Efficient gene targeting in *Drosophila* with zinc-finger nucleases. *Genetics*, **172**, 2391–2403.