Proc. KMB International Symposium
June 24-26, 2003, Muju, Korea

Artificial Transcription Factors for Genomics Studies and Industrial Applications

Jin-Soo Kim, Kyung-Soon Park, Horim Lee, and Young-Soon Jang

ToolGen, Inc., 461-6 Jeonmin-Dong, Yuseong-Gu, Daejeon, 305-390, South Korea

We have developed a method that uses randomized libraries of zinc finger-containing, artificial transcription factors to induce phenotypic variations in yeast and mammalian cells. The modular structure and diverse DNA-binding specificities of zinc finger domains make them ideal building blocks for artificial transcription factors. By linking multiple zinc finger domains together, we constructed more than 100,000 zinc finger proteins with diverse DNA-binding specificities and fused each of them to either a transcriptional activation or repression domain. The resulting transcriptional regulatory proteins were expressed individually in cells, and the transfected cells were screened for various phenotypic changes such as drug resistance, thermo-tolerance, or osmo-tolerance in yeast. Genes associated with the selected phenotypes were identified. Our results show that randomized libraries of artificial transcription factors are powerful tools for functional genomics and phenotypic engineering.

Introduction

In the course of differentiation, development, and response to environmental challenges, cells and organisms display myriad phenotypes by altering the expression of specific genes in a timely manner. A key goal for biologists in this post-genomic era is to identify the genes and gene products responsible for phenotypes of interest. Therefore, we sought to develop a method by which researchers can modify the phenotypes of cells and organisms at will for further study. To accomplish this, we decided to mimic the method that cells and organisms use to achieve new phenotypes: selective regulation of gene expression on a genome-wide scale. Our method involves (i) the introduction of a large collection of novel, artificial transcription factors into a population of cells and (ii) screening of the population for the desired phenotype.

Zinc fingers are well-characterized, highly specific DNA-binding domains found in a wide variety of transcriptional regulatory proteins. Each individual finger recognizes a specific three-base pair DNA sequence, and a single transcription factor can contain multiple zinc fingers. Together, these fingers determine the DNA-binding specificity of the protein and, in part, its target genes. Because of their diversity and modular structure, zinc finger DNA-binding domains are ideal building blocks with which to construct large numbers of transcription factors with diverse DNA-binding specificities.

Here we describe a novel approach for altering diverse phenotypes in yeast and *E. coli* with artificial transcription factor libraries. The ZFP library approach involves regulation of gene expression on a genome-wide scale using randomized libraries of zinc finger proteins (ZFPs).

Results

Construction of Randomized Libraries of Zinc Finger Proteins

A typical ZFP consists of three or more zinc fingers, and, as first approximation, most zinc fingers recognize 3-base pair (bp) subsites. Thus, a 3-finger protein recognizes and binds to a 9-bp DNA sequence, and a 4-finger protein binds to a 12-bp sequence. Scores of zinc fingers with diverse DNA-binding specificities have been identified in many naturally-occurring transcription factors or created artificially via protein engineering techniques such as phage display and site-directed mutagenesis.

To produce libraries of zinc finger proteins (ZFPs), we cloned DNA segments that encode zinc fingers with diverse DNA-binding specificities and shuffled them randomly to make composite DNA segments that encode 3-finger or 4-finger proteins (Fig. 1A). These assembled DNA segments were then linked to DNA segments that encode either a transcriptional activation or repression domain to produce libraries of transcription activators or repressors, respectively. Because a ZFP that does not contain an activation or repression domain can function as an efficient transcriptional repressor when it competes with endogenous transcription factors for binding to DNA sequences near the site of transcription initiation, we also included in our transcription factor library assembled zinc finger domains that had not been fused to any additional domains. We chose 25 zinc fingers with which to build 4-finger proteins and 40 zinc fingers with which to build 3-finger proteins. The zinc fingers used to generate these two sets of multi-finger ZFPs recognize 25 3-bp subsites out of 64 (= 4 x 4 x 4) possible subsites. Therefore, we estimate that our ZFP transcription factor (ZFP-TF) libraries contain, on average, 358 3-finger proteins [= $(25/64)^3$ x 2,000 (bp) x 3 (TFs)] or 140 4-finger proteins per a given 1-kbp stretch of DNA. [Note: The probability that at least one ZFP exists in the 3-finger protein library that matches a given 9-bp stretch of DNA is $(25/64)^3$. In a given 1-kbp stretch of DNA, the typical size of a eukaryotic promoter, there are approximately 2,000 overlapping 9-bp sites on both the Watson and Crick strands. Finally, in our libraries, there are three different types of transcription factors (TFs); isolated ZFPs, ZFP-activation domain fusions, and ZFP-repression domain fusions.]. However, it is unlikely that all of these transcription factors will actually bind to a given stretch of DNA in the genome, if one considers the local chromatin structure. Still, these numbers seem large enough that one could expect to find transcription factors that regulate most, if not all, of the genes in virtually any eukaryotic genome.

Phenotypic Changes in Yeast

First, we chose Baker's yeast, *Saccharomyces cerevisiae*, as a model organism in which to test our approach. In all subsequent experiments in yeast, cells were transformed with plasmid libraries encoding ZFP-TFs, grown under selective or stressful conditions, and screened for several desired phenotypes (Fig. 1B). Plasmids that encoded ZFP-TFs were rescued from cells that were selected in the screen, retransformed back into yeast cells to confirm the phenotypic alteration, and sequenced to reveal the identities of the ZFPs. The first phenotype that we screened for was growth arrest, and a number of cells showed growth arrest only in the presence of galactose (Fig. 2A). Because expression of the ZFP-TFs in yeast had been put under the control of a galactose-inducible promoter, we concluded that the ZFP-TFs expressed in the yeast cells were responsible for the growth arrest phenotype.

We also screened for ZFPs that induced industrially important traits in yeast. During fermentation, osmotic pressure is increased and excessive heat is generated. These stressful conditions often limit the productivity of yeast culture systems. We screened for cells that became resistant to heat treatment or osmotic pressure upon expression of the ZFP-TFs in our libraries. For example, more than 99.6% of wild-type cells died upon heat treatment at 55°C for 2 h (Fig. 2B). In contrast, 10% of cells transformed with certain ZFP-TFs survived under these extreme conditions, a 25-fold increase in the thermo-tolerance

phenotype [that is, the percentage of cells expressing ZFP-TFs that survived under stress conditions (10%) divided by the percentage of control wild-type cells that survived under the same conditions (0.4%)]. These phenotypic improvements were observed only when galactose was present in the growth media.

Next, we used lithium chloride (LiCl) to induce osmotic stress. The presence of 100 mM LiCl in the culture media almost completely inhibited the growth of wild-type yeast cells. However, we were able to isolate surviving cells from the pool of ZFP-TF transformants grown under the same conditions (Fig. 2C). Various levels of osmo-tolerance were observed in the selected cells, some displaying up to a 100-fold increase in the phenotype. Again, these phenotypic improvements were observed only in the presence of galactose. These two examples clearly demonstrate that screening ZFP-TF libraries could rapidly yield improvements in the traits of industrially important organisms.

We then applied our approach to a drug resistance phenotype in fungi. Although ketoconazole is a widely used anti-fungal drug, strains that have developed resistant to this drug limit its utility. An understanding of the mechanisms of drug resistance and the identification of genes associated with this phenomenon are of interest to the biomedical community. Thus, we screened for ZFP-TFs that conferred ketoconazole resistance upon yeast cells. Among 10 million transformants screened, 120 colonies grew on agar that contained 35 μM ketoconazole. We determined the DNA sequences that encoded the ZFP-TFs from 23 randomly chosen ketoconazole-resistant colonies. This led to the identification of 11 different ZFP-TFs. The degree of drug resistance was variable among these ZFP transformants (Fig. 3A). Interestingly, co-expression of two selected ZFPs often resulted in an enhancement of the drug resistance phenotype. For example, each of the ZFP-TFs, K4, K5, and K11, alone conferred only partial (10-fold to 100-fold) ketoconazole resistance upon yeast cells. When two of these ZFP-TFs were co-expressed, yeast cells became completely resistant to ketoconazole; this was an ~1,000-fold enhancement in the resistance phenotype (Fig. 3B).

If the various phenotypic changes observed in these selected cells resulted from the activity of ZFPs as transcription factors, DNA-binding activities of ZFPs and the presence of appropriate functional domains would be essential. To address this issue, we generated ZFP mutants with one of the two following alterations: (i) a key amino acid residue that is expected to function in DNA base recognition was mutated (K5-M1), or (ii) an activation domain was either removed or replaced with a repression domain (K5-M2). In one mutant, an asparagine residue in the second zinc finger was mutated to alanine. In addition, this mutant protein did not confer the ketoconazole resistant phenotype upon yeast cells (Fig. 3C). In another mutant, the Gal4 transcriptional activation domain present in the K5 ZFP-TF was deleted by inserting a stop codon in front of the DNA sequence that encode the activation domain. This protein also failed to render cells resistant to ketoconazole (Fig. 3C). Interestingly, when the activation domain fused to the K5 ZFP-TF was replaced with the Ume6 repression domain (K5-M3), the transformant showed a reversed phenotype (that is, the cells became sensitive to ketoconazole) (Fig. 3C). This result implies that, at least in some cases, it is possible to induce the desired phenotype by first selecting ZFP-TFs that confer the opposite phenotype and then replacing the effector domain with another domain with the opposite function. This is useful for cases where the opposite phenotype is more amenable to screening than the desired phenotype. From these mutagenesis analyses, we conclude that many of our selected ZFP-TFs do indeed function as transcriptional regulators *in vivo*. However, we note that some naturally-occurring ZFPs exert their effects by binding to RNA or protein. Therefore, for some of our ZFP-TFs, it is possible that they induce the selected phenotypes by these alternative mechanisms.

Identification of Ggenes Associated with Anti-Fungal Drug Resistance

We performed DNA microarray experiments to identify genes that are associated with the ketoconazole resistance phenotype. We reasoned that different ZFP-TFs that confer the identical phenotype may regulate

similar sets of genes whose differential expression is directly or indirectly associated with the phenotype. Thus three ZFP-TFs, K5, K6, and K7, were chosen for genome-scale expression profiling analyses. All of these transcription factors contained the Gal4 activation domain. Out of 6,400 yeast open reading frames (ORFs) tested, ten ORFs were co-activated more than two fold by at least two of the ZFP-TFs tested, and four ORFs were co-activated by all three ZFP-TFs. None of these genes appear to be direct targets of the K5, K6, or K7 ZFP-TF, as we were not able to identify the predicted DNA binding sites in the promoters of these genes. First, we noticed that *PDR5*, a gene known to pump ketoconazole out of the cell, was activated by two ZFPs, K6 and K7, but not by K5. Thus it appears that at least two different biological pathways, one of which involves the activation of *PDR5*, participate in conferring ketoconazole resistance in yeast. In order to identify new genes associated with the drug-resistant phenotype, we over-expressed each of the four co-activated genes and tested for the drug resistance phenotype. One of the genes induced ketoconazole resistance when over-expressed in yeast cells (Fig 3D, 3E). This gene sequence was identified as a hypothetical ORF (GenBank accession number: YLL053C) that is highly homologous to plasma membrane and water channel proteins in another yeast, *Candida albicans*, a pathogenic fungus. It is not yet clear whether the product of the YLL053C gives rise to the observed phenotype by pumping out ketoconazole, as seen with the *PDR5* gene product.

Discussion

We conclude that it is possible to generate phenotypic variation in microbes and by randomly regulating gene expression using randomized libraries of artificial transcription factors. Our approach, termed GeneGrip Phenomining, provides a novel tool for functional genomics and cell engineering that has several advantages over current techniques (such as random mutagenesis, comprehensive over-expression, gene knockout screening, ribozyme library screening).

First, we predict that this technology will be universally applicable in many, if not all, organisms of commercial and scientific interest.

Second, ZFP-TF libraries encode both transcriptional activators and repressors, increasing the possibility that highly diverse phenotypic changes will be observed.

Third, ZFP-TFs can cause more subtle changes in gene expression. In cases when over-expression or knockout of a gene is lethal to the organism, subtle changes in gene expression might allow one to achieve the desired phenotypes.

Fourth, our method requires significantly fewer rounds of screening than do most other approaches. Conventional mutagenesis typically requires screening of 10^7 to 10^9 clones. The diversity of a ribozyme library must be at least on the order of 10^7. We have isolated various "mutants" with the desired phenotypes by screening only 10^5 to 10^6 ZFP-TFs in yeast.

Fifth, the genes that are associated with the selected phenotypes can be identified. In contrast, it is not straightforward to identify the genes associated with the desired phenotypes when mutants are screened after random mutagenesis.

And finally, phenotypic transfer from one strain to another is straightforward with our approach.

Zinc fingers constitute the most abundant DNA-binding motif encoded in the human genome. Various genetic events, such as DNA duplication, translocation, deletion, or addition, that took place during evolution, are believed to have generated numerous ZFPs. These events might have contributed to phenotypic diversity and, therefore, aided the evolutionary process. Therefore, mimicking this process on a genomic scale, as demonstrated in this report, could provide a new mode of "mutagenesis", which might be useful for biomedical research and biotechnology applications.

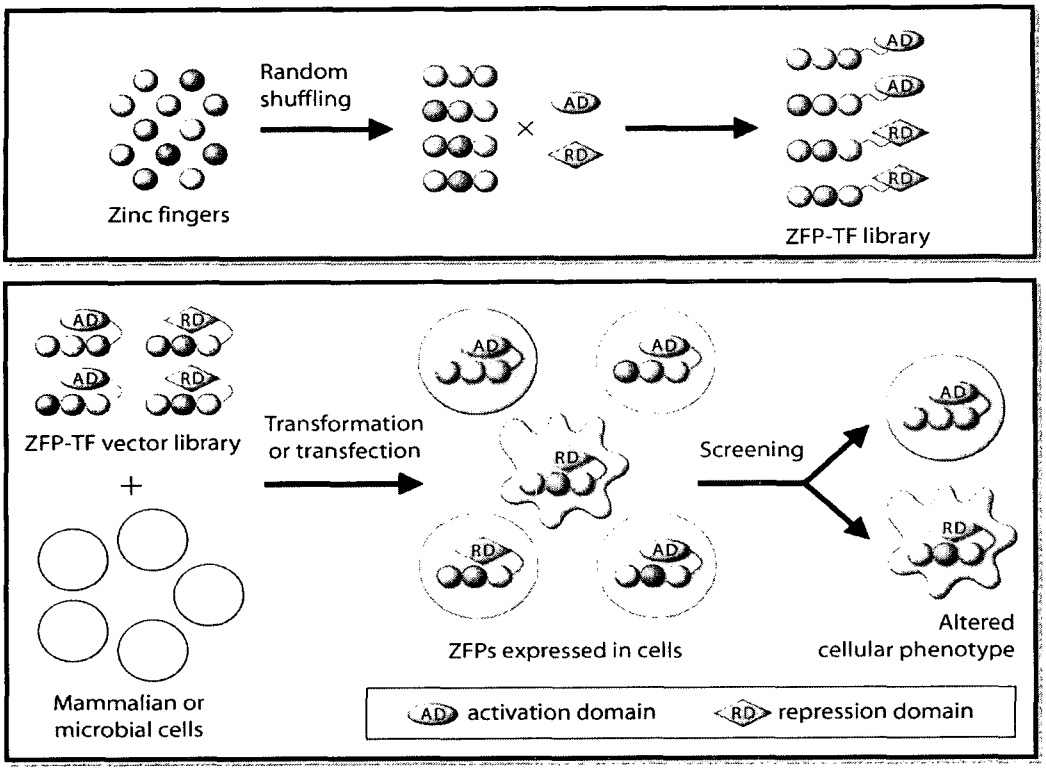

Fig. 1. Schematic representation of the ZFP library approach. (A) Construction of randomized libraries of ZFP-TFs. (B) Steps of the ZFP library approach ZFP, zinc finger protein; AD, activation domain; RD, repression domain.

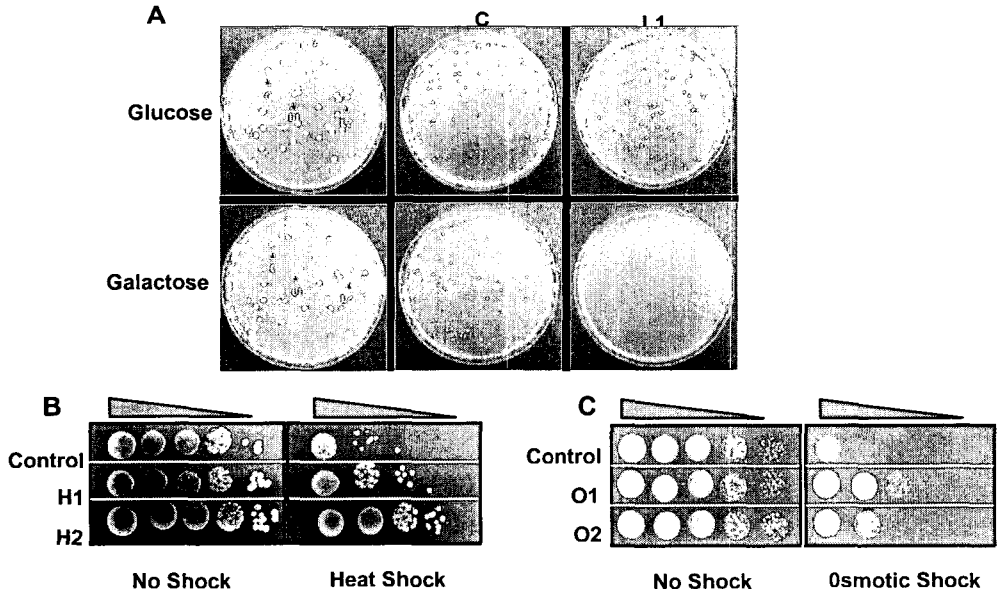

Fig. 2. Various phenotypic changes in yeast induced by artificial ZFP-TFs. (A) Growth-defective cells are observed upon expression of ZFP-TFs. (B) Thermo-tolerant phenotype induced by ZFP-TFs. (C) Resistance to osmotic shock induced by ZFP-TFs. In (B) and (C), the triangles drawn above of each panel indicate 10-fold serial dilutions (1:1 to 1:10,000, left to right) of spotted yeast cells.

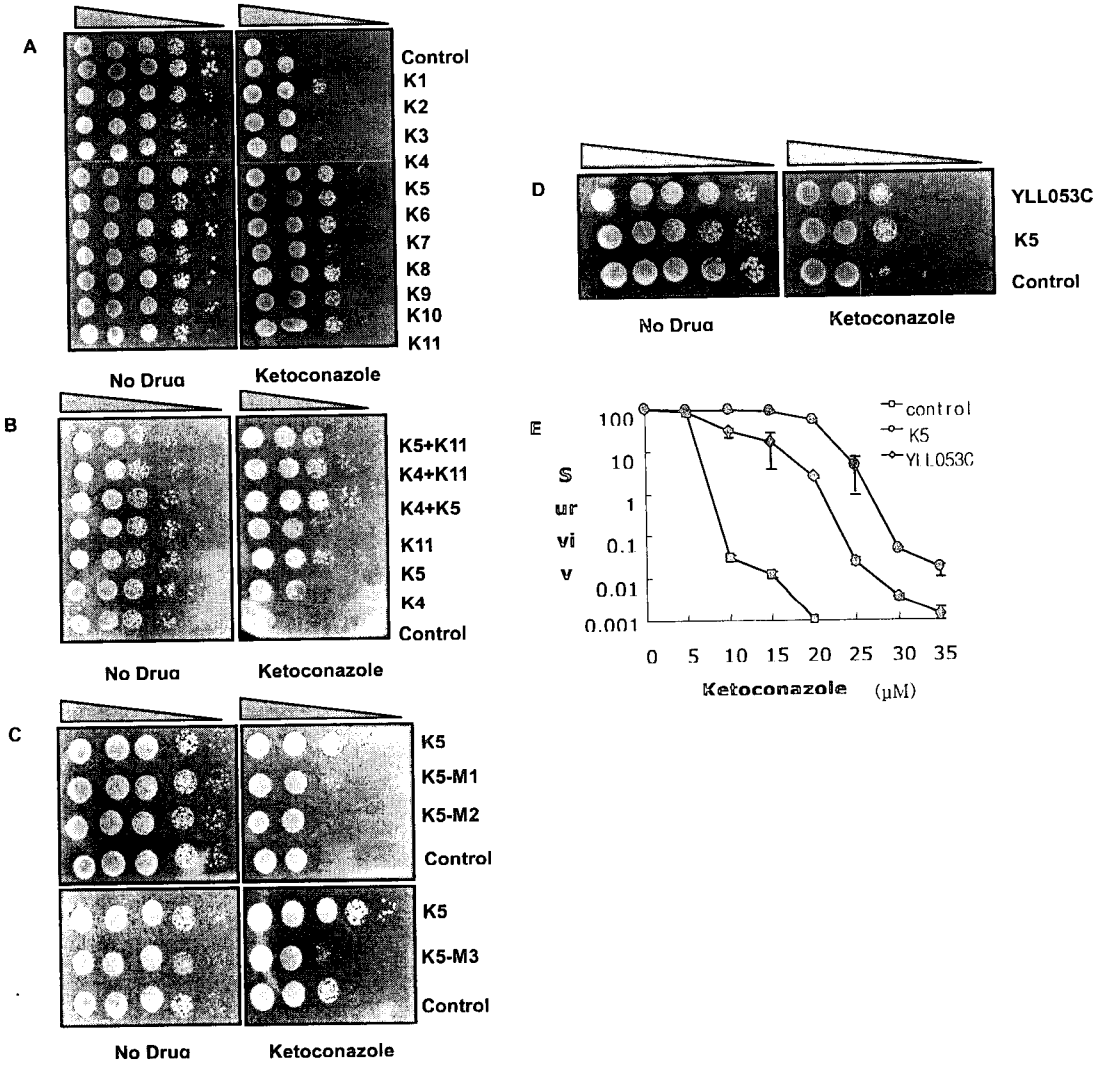

Fig. 3. Characterization of ketoconazole-resistant transformants. (A) Eleven ketoconazole-resistant ZFP-TF transformants were selected by growth on agar plates containing ketoconazole (35 μM). (B) Co-expression of ZFP-TFs in yeast cells leads to phenotypic enhancement. (C) Selective mutagenesis of the K5 ZFP-TF (VSSR-DGNV-VSSR-VDYK-Gal4). (D) Induction of ketoconazole resistance by over-expression of yeast ORF *YLL053C*. (E) Cell viability test on media containing variable concentrations of ketoconazole (0, 5, 10, 15, 20 25, 30 or 35 μM).