MOLECULAR CHARACTERIZATION OF A SUBSET OF KRAB-ZFPs

by

Alain Chamoun

A Thesis Submitted to the Faculty of

The College of Science

in Partial Fulfillment of the Requirements for the Degree of

Master of Science

Florida Atlantic University

Boca Raton, Florida

August 2010

MOLECULAR CHARACTERIZATION OF A SUBSET OF KRAB-ZFPs

by

Alain Chamoun

This thesis was prepared under the direction of the candidate's thesis advisor, Dr. Kasirajan Ayyanathan, Department of Biological Sciences, and has been approved by the members of his supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Master of Science.

SUPERVISORY COMMITTEE:

Kasirajan Ayyanathan, Ph.D.
Thesis Advisor

David Binninger, Ph.D.

James X. Hartmann, Ph.D.

Rodney Murphey, Ph.D.
Chair, Department of Biological Sciences

Gary W. Perry, Ph.D.
Dean, Charles E. Schmidt College of Science

Barry T. Rosson, Ph.D.
Dean, Graduate College

July 16, 2010
Date

ACKNOWLEDGEMENTS

The author wishes to thank Dr. Kasirajan Ayyanathan for his endless help and support during this investigation. Without him, this would not have been possible. The author also wishes to thank Dr. David Binninger and Dr. James X. Hartmann for their continued support and expert insight in the field of molecular biology. The author is also grateful for the members of the Florida Atlantic University's Center for Molecular Biology and Biotechnology. Lastly, the author would like to acknowledge his parents, Tony and Carmen Chamoun.

ABSTRACT

Author:	Alain Chamoun
Title:	Molecular Characterization of a Subset of KRAB-ZFPs
Institution:	Florida Atlantic University
Thesis Advisor:	Dr. Kasirajan Ayyanathan
Degree:	Master of Science
Year:	2010

There are approximately 20,000 genes in the human genome. Around 2% of these genes code for transcriptional repressors known as KRAB-ZFPs. It is already known that Zinc-Finger Proteins contain two main functional domains at either end of the polypeptide. In today's database, you will find a KRAB (Kruppell-associated Box) domain at one end and a tandem array of Zinc-finger repeats at the other end. The carboxyl terminal tandem Zinc-finger repeats function as sequence-specific DNA-binding domains. The amino terminal KRAB domain serves as a repressor domain, which will recruit a co-repressor termed KAP-1 (KRAB Associated Protein-1). Located in between these two domains is a region of uncharacterized DNA referred to as the "Linker Region". This thesis will explore the DNA-binding domains of 6 known KRAB-ZFPs, as well as utilize the linker regions to derive an evolutionary history for this superfamily.

MOLECULAR CHARACTERIZATION OF A SUBSET OF KRAB-ZFPs

LIST OF FIGURES

INTRODUCTION

Eukaryotic cells perform a plethora of functions in order to respond to their environment and communicate with each other. These include cell division, import and export of macromolecules, signal transduction, transcription and translation of the genomic code, and many other specialized functions. Cells must utilize their genetic information to conduct every type of cellular function. The DNA code is 'read' by a large complex of proteins that generate an RNA copy called mRNA (messenger RNA), in a process called transcription. After transcription, the mRNA is exported to the cytoplasm where it is translated by a ribosome, which will ultimately make a protein. Transcription is clearly the pivotal starting point for any cellular function and it is tightly regulated. Regulation of transcription involves many different proteins called transcription factors. Transcription factors can function as either activators or repressors of transcription. A transcriptional activator will modify the DNA structure to create euchromatin, which is diffuse and easily accessible to the complex of proteins involved in transcription. A transcriptional repressor will conversely modify the DNA to create heterochromatin, which is very densely packed and essentially inaccessible to any kind of transcriptional machinery.

Transcription factors are grouped into superfamilies based on their effector domains and their DNA binding domains. Among the transcriptional repressors is a superfamily called KRAB-ZFPs, which contain two main functional domains at either end of the polypeptide. The KRAB-ZNF transcription factors represent the largest superfamily of transcription factors in mammals. Each member of the KRAB-ZFP superfamily contains a KRAB (**Kr**uppel-**A**ssociated **B**ox) domain at one end and a tandem array of zinc-finger repeats at the other end. The carboxyl terminal tandem zinc-finger repeats function as sequence-specific DNA-binding domains. The amino terminal KRAB domain serves as a repressor domain, which will recruit a co-repressor termed KAP-1 (**K**RAB-**A**ssociated **P**rotein-**1**), which functions as a well-characterized repression mechanism (Friedman et al., 1996). Although the mechanism of transcriptional repression has been understood in considerable detail (Ayyanathan et al., 2003), the regulated target genes are known only for a few members.

Located between these two domains is a region of uncharacterized DNA referred to as the "Linker Region". Very little is known about the function of these regions. Comprised is a database of all human KRAB Zinc-Finger Proteins, taking note of specific DNA and protein sequences of each "Linker Region". Extensive research and testing has classified these "Linker Regions" into distinct "Linker Families" based on sequence homology. Further bioinformatics and laboratory research will help to unravel the functionality of

each distinct "Linker Region". These linker regions, which are the main source of variability between members of the KRAB-ZFP superfamily, can also be utilized to derive an evolutionary history for this family of repressors.

There are a few members of the KRAB-ZFP superfamily that have already been documented to have co-expressional splice variants that produce truncated proteins having no KRAB domain. Both ZNF268 and ZNF468 have been reported to produce these truncated variants (Shao et al., 2006; Sun et al., 2005). More importantly, it has been reported that the truncated version of ZNF268, renamed ZNF268s, has cytoplasmic localization where it binds to and regulates IKK (Inhibitor of Kappa-B) (Chun et al., 2008). Such observed cytoplasmic activity, along with the notable linker sequence homology among members indicates that there may be more functionality to these transcriptional repressors than just binding to DNA and recruiting a corepressor. Furthermore, it is likely that these unknown functions are mediated by the linker regions of these transcription factors.

The major goal of this proposal is to decipher target genes for a subset of six KRAB-ZNF transcription factors illustrated in Figure 1. ZNF74 has been described in the development of schizophrenia and the deletion of ZNF74 has been long associated with DiGeorge syndrome (Takase et al., 2001; Ravassard et al., 1999). ZNF328 is known to suppress the MAPK pathway via suppression of SRE and AP-1 (Ou et al., 2005; Cao et al., 2005; Huang et al., 2006). ZNF480 is thought to be important for human heart development and disease

(Yi et al., 2004). Identification of the DNA-binding sites of these members will help to unravel their regulated target genes and to further our understanding of these similar, yet diversely functional proteins.

Currently being investigated are the sequence specific DNA-binding sites of the zinc finger domains of six known KRAB-ZFPs: ZFP36, ZNF74, ZNF136, ZNF141, ZNF328, and ZNF480. The DNA-binding domains are composed of C2H2 Kruppel-type zinc finger tandem repeats and each zinc finger is approximately 25-30 amino acids long. Fusion proteins that were made and folded *in vivo* were successfully isolated by GSH affinity purification for each of the six members (**Figures 11 and 12**). Multiple protocols of *in vitro* refolding were also implemented with little success for any of the six members (**See Appendix)**. However, sufficient *in vivo*-folded protein was obtained, and subsequent binding studies with each of these fusion proteins was carried out with a library of radiolabeled oligonucleotide 49-mers that represents over 68 billion different binding site possibilities. The results revealed a consensus binding-sequence for each of these six members.

MATERIALS AND METHODS

Constructing the Bioinformatics-Based KRAB-ZFP Database

The NCBI online database was utilized to identify all the known human KRAB-ZFPs, and also to retrieve sequence information for each of these members. This was done by entering the keywords "human" and "KRAB-ZFP" into the search query and selecting "Protein" in the pull-down search menu. The search output had to be sifted through to remove any related proteins that were not actual KRAB-ZFPs. Both nucleotide and protein sequences of both the full-length proteins and just the linker regions were collected for each member. Also collected were protein and cDNA accession links, aliases, restriction maps, and expression maps for each member. The linker regions from approximately 350 members were analyzed using a clustalw sequence alignment program. Members were organized into 18 distinct linker families based on sequence homology of their linker regions. Besides being able to test the linker families for their potential functionality, these 18 families were used to derive an evolutionary history for the KRAB-ZFP superfamily. Phylogram trees were constructed for each family using a program called FigTree. The eldest member from each family was chosen to represent its respective family. These chosen members were used to construct another phylogram, which would portray a

basic evolutionary history of how these 18 families came about. This method naturally excluded a percentage of members that did not fall under any particular family. Therefore, one more phylogram tree was constructed to portray relationships between all the linkers individually.

Making GST-ZnF Fusion Proteins

Recombinant clones expressing the zinc finger portion of ZFP36, ZNF74, ZNF136, ZNF141, ZNF328, and ZNF480 were constructed in pGEX plasmids, which provided a GST affinity tag. To grow large scale induced cultures for each member, a 10mL overnight culture was incubated at 30°C and inoculated to 250mL of LB containing ampicillin and kanamycin. The cultures were grown to A_{595} ~0.3-0.4 and then induced with 125µL of 100mM IPTG for 4 hrs. Cells were spun at 7K for 10 min and the cell pellet was stored at -80°C. The cell pellets were resuspended in 8mL of PBS containing 8mg lysozyme, 40µL 10mM $ZnSO_4$, and 60µL 100mM PMSF. The cell suspensions were rotated at 4°C for 1 hr and then subjected to six rounds of sonication. Each round was performed at 4°C for 1 min allowing a 1 min cooling time between rounds. After sonication, the cell lysates were centrifuged at 14K xg for 30 min at 4°C. The supernatant was centrifuged again as above. The clear supernatant was collected for purification of fusion proteins that were folded *in vivo* in a soluble form. The pellet was also stored for DEAE column purification and *in vitro* refolding of inclusion bodies.

Purification of Fusion Proteins folded in vivo by GSH Affinity Chromatography

The soluble supernatant above was passed through a GSH-Sepharose affinity bead column. The columns were packed with 200µL of GSH-Sepharose beads and washed with 10mL of 1x PBS. The samples were loaded onto the columns and the flow through was collected. The columns were then washed with 15mL of 1X PBS and the last 500µL of wash was collected. The columns were then eluted three times with 500µL of glutathione-containing elution buffer (15mM GSH, 50mM Tris-Cl pH 8.5, and 0.1x Triton-X 100). The elution buffer was allowed to sit for 10 min each time and the entire 500µL was collected. The purification procedure was repeated a second time with the flow through fractions for a total of six elutions per sample. Each flow through and elution was run on a 12% SDS-PAGE gel to determine if the fusion protein did in fact correctly fold and bind to the GSH column and get eluted.

Dialysis To Concentrate Usable Protein

The total protein collected was consolidated in separate dialysis tubing for each sample. The tubing was dialyzed against Solution 1 (100mL 10% DPBS; 900mL ddH_20; 100µL 100mM PMSF) with constant stirring overnight. The tubing was then transferred to Solution 2 (same as solution 1) and allowed to dialyze for an additional five hours. The tubing was then transferred to Solution 3 (100mL 10% DPBS; 250mL 100% glycerol; 650mL ddH_2O; 100µL

100mM PMSF). The protein samples were then removed from the dialysis tubing and placed in microcentrifuge tubes as 1mL aliquots.

Bradford Assay To Estimate Protein Concentration

A standard Bradford Assay was implemented to determine the concentration of each protein sample. First, a standard curve was established using BSA at 0, 1, 2, 4, 8, 16, and 32µg increments. Then, both 10µL and 20µL aliquots of each protein sample were run against this standard BSA curve to determine the protein concetration present in both 10µL and 20µL of each protein sample.

Cold Binding To Enrich Oligonucleotide Library

To prepare the randomized 49-mer oligonucleotide library for the radiolabeled binding site selection assay, a non-radioactive binding assay was implemented preliminarily. First, 60µL of GSH sepharose beads were mixed with 90µL of empty sepharose beads. PBS wash buffer (500µL) was added and allowed to rotate for 2 min. The beads were centrifuged at 5K rpm for 2 min and then the supernatant was discarded. These bead-washing steps were repeated two more times. Next, 1mL of PBS wash buffer was added and allowed to rotate for 1 hr at 4^{o}C. The beads were centrifuged at 5k rpm for 2 min and the supernatant was discarded. The washed beads were resuspended in 690µL of PBS wash buffer and aliquoted into 6 individual microcentrifuge tubes (115µL

for each of the 6 protein samples). Aliquots (5μg) of each of the six protein samples were added to each microcentrifuge, respectively. The six microcentrifuge tubes, which contain 115μL of beads and 5μg of protein, were allowed to rotate for 1 hr at $4^{O}C$ to allow the GST-tagged protein to bind to the GSH-coated sepharose beads. After binding, the beads are centrifuged at 5K rpm for 2 min and the supernatant was discarded. The previously described bead-washing steps were again implemented three times, with the third time involving the microcentrifuge tubes being allowed to rotate for 30 min, to ensure that any molecules that may have bound to the beads non-specifically were washed away. The beads were then centrifuged at 5K rpm for 2 min and the supernatant was discarded. Aliquots (750μL) of NEBB+ (Nuclear Extract Binding Buffer +NaCl) were added to the beads, followed by 1μL of the randomized oligonucleotide library. The tubes were allowed to rotate for 30 min at $4^{O}C$ and then an additional 30 min at room temp to allow those oligomers, which have a high affinity for the bead-bound protein, to bind. The beads were then centrifuged at 5K rpm for 2 min and the supernatant is discarded. The beads were washed with 700μL of NEBB+ and the microcentrifuge tubes were allowed to rotate for 2 min. The beads were again centrifuged at 5K rpm and the supernatant was discarded. This NEBB+ wash was repeated another 2 times to ensure that all loosely bound oligomers which have no affinity for our proteins were removed from the beads. Aliquots (20μL) of GSH elution buffer were added to each microcentrifuge tube. The tubes were tapped to mix and

allowed to sit on the bench for 10 min. The beads were then subjected to centrifugation at 14K rpm for 10 min and the supernatant was transferred to fresh microcentrifuge tubes. This supernatant should contain oligomers that were selected by their protein-binding capability. Aliquots (2μL) of this supernatant were used for PCR, and the other 18μL were stored at –20^{O}C. A 50μL PCR reaction (2μL DNA, 5ul DMSO, 25μL 2X Master Mix, 1ul 3' primer, 1μL 5' primer, and 16ul of dd$H_2$0 were cycled 35 times) was set up to amplify the selected oligomers. Aliquots (5μL) of the PCR product were taken for DNA-PAGE analysis (10% polyacrylamide: 8mL 30% polyacrylamide; 2.4mL 5X TBE; 13.6mL ddH_2O; 240μL 10X APS; 33μL TEMED) and the remaining 45μL was subjected to phenol-chloroform cleanup. The PCR products for each of the six protein samples were transferred to a fresh microcentrifuge tubes and 45μL of PCI was added. The tubes were mixed by tapping and centrifuged at 14K rpm for 5 min. The top layer was transferred to a fresh tube and 4μL NaOAc/1μL glycogen was added to each tube before addition of 120μL 100% EtOH. The tubes were vortexed to mix and placed on dry ice for a minimum of 15 min to precipitate the DNA. The tubes were centrifuged at 14K rpm for 15 min and decanted. The pellets were washed with 100μL 70% EtOH, mixed by tapping, and allowed to sit for 5 min. The tubes were then centrifuged at 14K rpm for 10 min and decanted. Once the pellets were dry, they were resuspended in 20μL 1X TE buffer and stored at –20^{O}C. These enriched oligonucleotide libraries were then subjected to the entire selection process again to produce libraries

that should be more enriched for each protein. This "cold-binding" assay was repeated a total of four times in preparation for the radiolabeled "hot-binding" assay.

Hot Binding To Select Specific Binding-Site Sequences

The enriched libraries, which were derived from the cold-binding assays performed for each protein, were then subjected to "hot-binding" EMSA selection to isolate only those oligomers whose sequence has a high affinity for our target proteins. First, a kinasing reaction was assembled to radiolabel our enriched libraries. The reaction contained 2µL enriched DNA, 2µL PNK (Polynucleotide Kinase) buffer, 2µL PNK enzyme, 0.25µL source ^{32}P, and 13.75µL ddH_2O and then it was divided into 6 tubes. The 20µL reactions were incubated at 37^{O}C for 2 hrs. During the incubation, spin columns were packed with G25 beads, and the beads were dried by centrifugation at 2.5K rpm for 1 min. These G25 beads are intended to remove excess ^{32}P from the freshly phosphorylated oligomers. Once the kinasing reactions were finished, 10µL ddH_2O was added to each reaction tube and the 30µL reactions were pipetted onto the dry G25 beads. The spin columns were centrifuged at 2.5K rpm for 10 min and the dirty columns are discarded in radiation waste. Aliquots (2µL) were transferred to clean tubes for scintillation counting and the remaining 28µL of radiolabeled oligomer was stored at –20^{O}C. Next, binding reactions were assembled with 10µL protein, 18µL radiolabeled DNA, and 7µL 5X NEBB. The

products of the binding reaction were electrophoresed on a long DNA-PAGE (5% polyacrylamide: 5.5mL 30% polyacrylamide; 6.6mL 5X TBE; 20.9mL ddH_2O; 330µL 10X APS; 33µL TEMED) and the gel was dried onto a paper membrane. Depending on the amount of radioactivity detected by the scintillation counter, the dry gel was used to expose a piece of film for a varying amount of time. Once the film was developed, the bands that show up on the film were used to find exactly where the protein-DNA complex was "shifted" in the gel. The bands were cut out of the dry gel and eluted in fresh tubes with 500µL of gel band elution buffer overnight at room temp. These eluted complexes should contain only the protein of interest and the specific oligomers that bind it. Once eluted, the tubes were spun at 14K rpm for 10 min and the supernatant was transferred to fresh tubes where they were subjected to two rounds of PCI cleanup, one round of CI cleanup, and EtOH precipitation to clean the eluted DNA. This DNA was PCR amplified in preparation for the next round of "hot-binding". After the last round, the oligomers were PCR amplified and then electrophoresed on a 10% DNA-PAGE. The bands were excised and electro-eluted at 45V for 2 hrs into 400µL of 1X TAE. This "hot-binding" assay was repeated for three rounds to produce DNA libraries for each protein that were highly enriched to a level that they mostly contained only those sequences which physically bind to each respective protein. These enriched libraries were then used for cloning.

Cloning the Selected Sequences

In order to have our selected oligomers sequenced, they had to be cloned into pUC18 vector plasmids. Each oligomer contains conserved *Bam*H1 and *Eco*R1 restriction sites on either side of the randomized region. The oligomers, as well as the pUC18 vectors, were digested with both restriction enzymes. First, 10μL of each enriched library was added with 5μL 10X EcoBuffer, 5μL 10X BSA, 0.5μL EcoR1 enzyme, 0.5μL BamH1 enzyme, and 29μL ddH_2O. Next, 5μL of undigested pUC18 vector was added with 5μL 10X EcoBuffer, 5μL 10X BSA, 1μL EcoR1 enzyme, 1μL BamH1 enzyme, and 33μL ddH_2O. Both digestion reactions were incubated at $37^{O}C$ for 2 hrs. The oligomer digestions were then topped off with 50μL ddH_2O and cleaned using two rounds of PCI cleanup, one round of CI cleanup, and EtOH precipitation before a final resuspension in 10μL ddH_2O. The digested vector was loaded onto a 1% agarose gel and the bands were cut and eluted using the GeneClean kit. Ligation reactions were assembled in fresh tubes and contained 1μL digested pUC18 vector, 4μL of digested oligomer and 5.5μL of ligation cocktail. The tubes were mixed by tapping and centrifuged gently. This was repeated for all six oligomer samples. A control was included that used water instead of DNA. The ligation reactions were incubated for 10 min at room temp. Meanwhile, DH5□ cells were thawed on ice water and 13mL culture tubes were set up in preparation for transformation. Each culture tube was loaded with 40μL competent DH5□ cells and 4.5μL of ligation reaction. The culture tubes were

tapped lightly to mix and then placed on ice for 30 min. These tubes were then warmed in a 37^{O}C water bath for 20 sec and then placed back on ice. LB (900mL) was added to each tube before the tubes were capped and placed in the shaker at 225 rpm for 1 hr at 37^{O}C. Aliquots (100μL and 200μL) of each of the seven culture tubes (six DNA samples and one control) were plated onto agar-containing Petri plates with ampicillin. These plates were incubated at 37^{O}C overnight and then stored at 4^{O}C until the colonies were picked. Thirty colonies were picked from each sample (10 colonies for sample 5 (ZNF328)) and were used to inoculate fresh 13mL culture tubes, which contained 5mL LB and 5μL ampicillin. The culture tubes were lightly vortexed and placed in the shaker at 225 rpm overnight at 37^{O}C. These culture tubes, expected to contain the chosen colonies that have been growing, were taken out, lightly vortexed, and placed back into the shaker for another hour to stimulate maximum growth of the target pUC18 vectors that have been loaded with our selected oligomers. The plasmids were prepared by spinning the 5mL culture at 5K rpm for 3 min. The LB was decanted and 100μL Solution 1 (GTE) was added to the pellets. These were vortexed until the pellets were completely resuspended and 200μL of Solution 2 (SDS/NaOH) was added. The tubes were inverted multiple times to mix and then placed on ice for 5 min. Next, 150μL of Solution 3 (KOAc) was added, and the tubes were again inverted and placed on ice for 5 min. Next, they were centrifuged at 14K rpm for 15 min and the supernatant was poured into fresh tubes. IPOH (300μL) was added to the supernatant and they were

mixed and placed on ice for 15 min. The tubes were again centrifuged at 14K rpm for 15 min and all the IPOH was decanted before washing the pellet with 300µL 70%. The tubes were mixed and centrifuged at 14K rpm for 15 min. The EtOH was decanted and the pellets were dried and resuspended in 20µL ddH_2O. These resuspended pellets contained a bulk amount of our pUC18 vector, which has been loaded with our selected oligomers. Aliquots (5µL) of vector DNA were subjected to double digestion with BamH1 and EcoR1 (0.5µL BamH1, 0.5µL EcoR1, 0.25µL RNase A, 2µL 10X EcoBuffer, 2µL 10X BSA, 9.75µL ddH_2O, and 5µL plasmid per sample) and electrophoresed on a 10% DNA-PAGE to determine the presence of an inserted oligomer. The positive clones were subjected to RNase treatment (0.3µL RNase A, 2.5µL NEB3, and 7.2µL ddH_2O per sample) before being subjected to PCI/PCI/CI/EtOH cleanup and resuspension in 20µL ddH_2O. Aliquots (2µL) of this resuspension were electrophoresed on a 1.2% agarose gel to determine the plasmid concentration. These vectors were then dried onto 96-well plates and sent out to the University of Florida for sequencing.

Deriving Consensus Binding-Site Sequences

Once the oligomers were sequenced, their sequence needed to be extracted from the surrounding pUC18 sequence. The sequences were analyzed manually with the help of an alignment tool, MultAlign. The sequences were organized (based on sequence homology) into a spreadsheet using

Microsoft Excel. The frequency of all four possible nucleotides was measured at each position and these frequency percentages were used to construct consensus sequences. The finalized consensus sequences were used to design and order new oligomers, which would be used in "hot-binding" assays to confirm the binding activity of these sequences. A separate set of oligomers which had their high percentage positions mutated were also ordered to be used in competition binding assays.

Testing Selected Binding-Sites

Testing the binding activity of the selected oligomer sequences and competition by mutated oligomers is still pending.

Deriving Potential Target Genes

The derived consensus sequences were analyzed manually to isolate a region of 8bp or more that contained the largest number of high percentage positions possible. This same region was also isolated from each of the derivative sequences to create a list of potential binding site sequences. These 8-9bp sequences were used to search through two well-known binding-site databases: The TRANSFAC module from Biobase, and the Eukaryotic Promoter Database (EPD). The TRANSFAC module is a database of previously observed binding sites. If our sequences fall in close proximity to another previously observed binding site, the database provides the associated gene

name and binding site information. It also provides the previously observed binders, if any, and denotes their binding sites in capital letters. The EPD database searches the human genome and provides any genes that contain our potential binding sites in the 60nt region upstream of their promoter. For every example, the searched sequence is highlighted in red letters.

All of the potential target genes were located and verified using the sequence viewer on the NCBI database. Once located, 26nt upstream and 26nt downstream were collected. For the TRANSFAC entries, the capitalized letters represent the previously observed binding sites, and the red letters represent our newly discovered binding sites. For the EPD entries, the capitalized letters represent the designated promoter region, and the red letters represent our newly discovered binding sites. For both cases, ant ATG start-site is denoted in blue, and any of our sequences that are repeated are denoted in green.

RESULTS AND DISCUSSION

One main goal of this proposal is to determine an evolutionary history for the superfamily of human KRAB-ZFP transcriptional repressors. Since the "linker region" is the only variable region in these proteins, all 334 linker regions from this superfamily were used to derive an evolutionary history. Most of these 334 linkers fell into 1 of the 18 families derived by sequence homology but some were left as outliers. **Figure 1** shows cladogram trees derived from each of the 18 "linker families". Family 18 only contained two members of exact homology and therefore could not derive a phylogram tree with clustalw. This is the only family that does not have a tree shown. **Figure 2** shows an evolutionary history among these 18 families. **Figure 3** shows an evolutionary history among all 334 linker regions as individuals.

Another main goal of this proposal is to determine the DNA-binding sequences of a subset of the six KRAB-ZFP superfamily members, which are diagrammatically shown in **Figure 4**. The DNA-binding domains of each member were expressed as fusion proteins tagged with GST by inserting them in-frame into pGEX vectors and diagrammatically represented in **Figure 5**. At least five independent recombinant clones were tested for proper protein

expression and the expression profiles for six members are shown in panels presented in **Figure 6**.

Initial Purification of GST-Tagged Target Fusion Proteins

Before refolding was attempted, purification of the fusion proteins that were folded *in vivo* was the main goal. There were five purification attempts all with results similar to the ones shown in **Figure 7**. ZFP36 was behaving well and binding to the GSH-Sepharose affinity column indicating that it was properly folded *in vivo*. Similarly, human ZNF141 was also a well-behaved protein binding to and eluting from the affinity columns in almost every attempt.

Final Purification of GST-Tagged Target Fusion Proteins

After multiple attempts of purification of *in vivo* folded fusion proteins with success in only two members, the protocol was reassessed and modified. A few details were changed:

1. A 10mL overnight culture was induced in 250mL of media instead of 50mL overnight cultures into 500mL media. This allowed the induction of the cells at ~0.3-0.4 O.D. *after* approximately one hour of growth.
2. The sonication of the cells was carried out with a higher wattage output than was originally used. This allowed complete shattering of the membrane components and nucleic acids and hence better release and purification of the target fusion proteins.

Figure 8 shows SDS-PAGE gels from the first successful round of purification, which indicates the presence of all six fusion proteins in the collected elutions. **Figure 9** shows SDS-PAGE gels from the last three rounds of purification. These last rounds only involved ZNF74, ZNF136, ZNF141, and ZNF480 because there was already sufficient usable protein for ZNF36 and ZNF328.

Concentrations of Usable Protein After Dialysis

The dialysis assay was successful in concentrating all six consolidated protein elutions. The final concentrations were 17.4, 14.4, 24.2, 16.6, 20.9, and 19.3μg/μL for ZNF36, ZNF74, ZNF136, ZNF141, ZNF328, and ZNF480, respectively. The actual volume for each sample ranged from 4-12mL, and this amount of protein was enough to perform four "cold-bindings" and three "hot-bindings". This protein will also be used in the "hot-binding" protocols to test our designed oligomers in binding and competition assays.

Cold-Binding to Pre-Enrich The Randomized Oligomer Library

The "cold-binding" assays did work quite efficiently, and the observation that oligomer bands were observed on each post-assay gel confirms that there was indeed a subset of oligomers that were in some way binding to the GSH-coated beads. Whether the oligomers were binding to the protein or just associating in some other way would be deciphered in later "hot-binding"

assays. The post-assay analytical DNA-PAGE gels showed successful isolation of a subset of oligomers each round (**Figure 10**). This indicates that each of the six *in* vivo-folded protein samples were able to bind to a specific subset of oligomers, thereby successfully isolating them from the rest of the randomized oligomer library. Each round served to further enrich the library further than was achieved in the previous round. A decrease in band intensity can be seen from round one to round four. This was expected, since the actual sample of oligomers becomes smaller and more enriched after each round of "cold-binding". After four rounds of "cold-binding", the original randomized library had become six individual enriched libraries (one for each protein) that were now ready to be further enriched with "hot-binding" assays.

Hot-Binding (EMSA) to Select Specific Oligomer Sequences

The "hot-binding" assays can also be referred to as electrophoretic mobility shift assays (EMSA). The oligomers are incubated with their respective protein to form DNA-protein complexes. This binding reaction is run on a long DNA-PAGE gel. The large DNA-protein complexes will be retarded in their movement through the gel and will form bands that will be "shifted" from the rest of the small oligomers that can easily migrate to the bottom of the gel. Since the oligomers are radiolabeled with ^{32}P, a piece of film can be exposed to view where the bands containing the DNA-protein complexes are. These radiographs are shown in **Figure 11**. Since the bands containing the DNA-protein

complexes were physically cut out of the gel, the risk of the enriched library being contaminated by oligomers that do not belong is very low. After each "hot-binding" assay, a DNA-PAGE gel was used to determine the amount of isolated oligomer. Those gels are shown in **Figure 12** and the same decrease in band intensity can be seen as in the "cold-binding" gels.

Cloning of Selected Oligomers Into pUC18 Vectors

All six enriched libraries were successfully cloned into pUC18 vectors for sequencing. The ligation involved digestion of both the vectors and the oligomers on the same day to ensure efficiency. This resulted in a usable amount of individual colonies on the agar plates. Each colony represents one DH5□ cell that successfully transformed with the pUC18 vector plasmid. Once the clones were processed, a digestion and DNA-PAGE analysis confirmed that over 90% of the selected clones were positive for containing an oligomer insert (**Figure 13**). Next, each of the 160 clones were assayed for plasmid concentration. These gels are shown in **Figure 14**, and suggest that there is a usable amount of plasmid for nearly every sample.

Deriving Consensus Binding-Site Sequences

As expected, analysis of the returned sequences proved to derive multiple consensus sequences for each protein. All except ZNF328 had more than three consensus sequences. Only the consensus sequences that were

derived from the most number of sequences were chosen to design the oligomers that would be used to more directly test for binding. These consensus sequences and their derivatives were also used to derive target genes. Thus, only ~10 of the 30 selected sequences were used for each protein. Those sequences and their derivatives are mapped out in **Figure 15**. The sequences selected for ZNF328 returned no usable binding-site information, therefore, consensus sequences were only derived for the remaining five KRAB-ZFPs. These sequences must be tested for their binding activity and their binding competition capacity.

Confirming Derived Consensus Binding Activity

Two more "hot-binding" assays must be performed to confirm that these derived consensus sequences do selectively and specifically bind to our proteins of interest. One must be done to confirm that these derived sequences will indeed bind, and another in the presence of a mutant to confirm their competition capability.

Deriving Potential Target Genes

The binding site sequences derived for each protein produced a long list of potential target genes. These target genes were organized into tables according to their respective protein binders. **Figure 16** shows the list of potential target genes for ZFP36. The potential target genes for ZNF74,

ZNF136, ZNF141, and ZNF480 are shown in **Figures 17, 18, 19, and 20**, respectively.

LIST OF FIGURES

Figure 1. Phylogram trees constructed from the 18 linker families. These trees were used to determine which member was the most ancient in each family.

Family 1

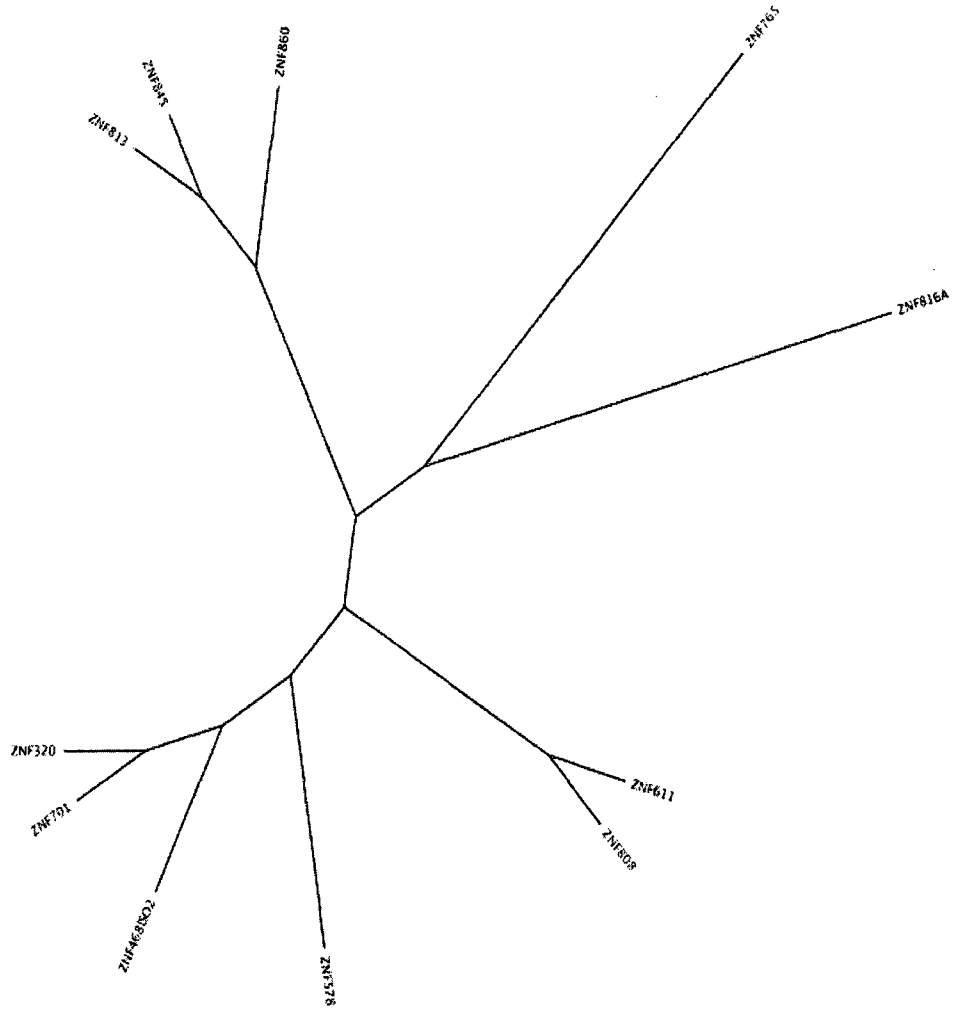


Figure 1. (continued)

Family 2

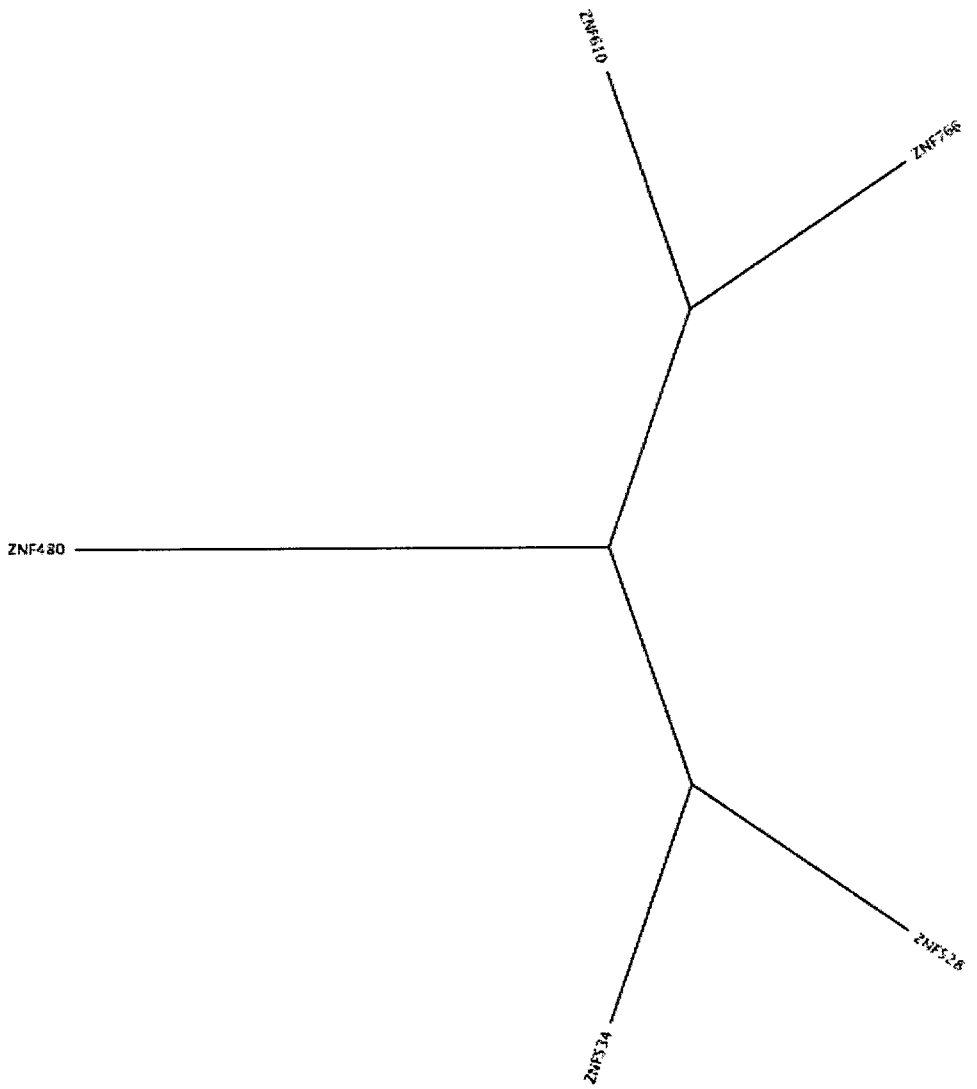


Figure 1. (continued)

Family 3

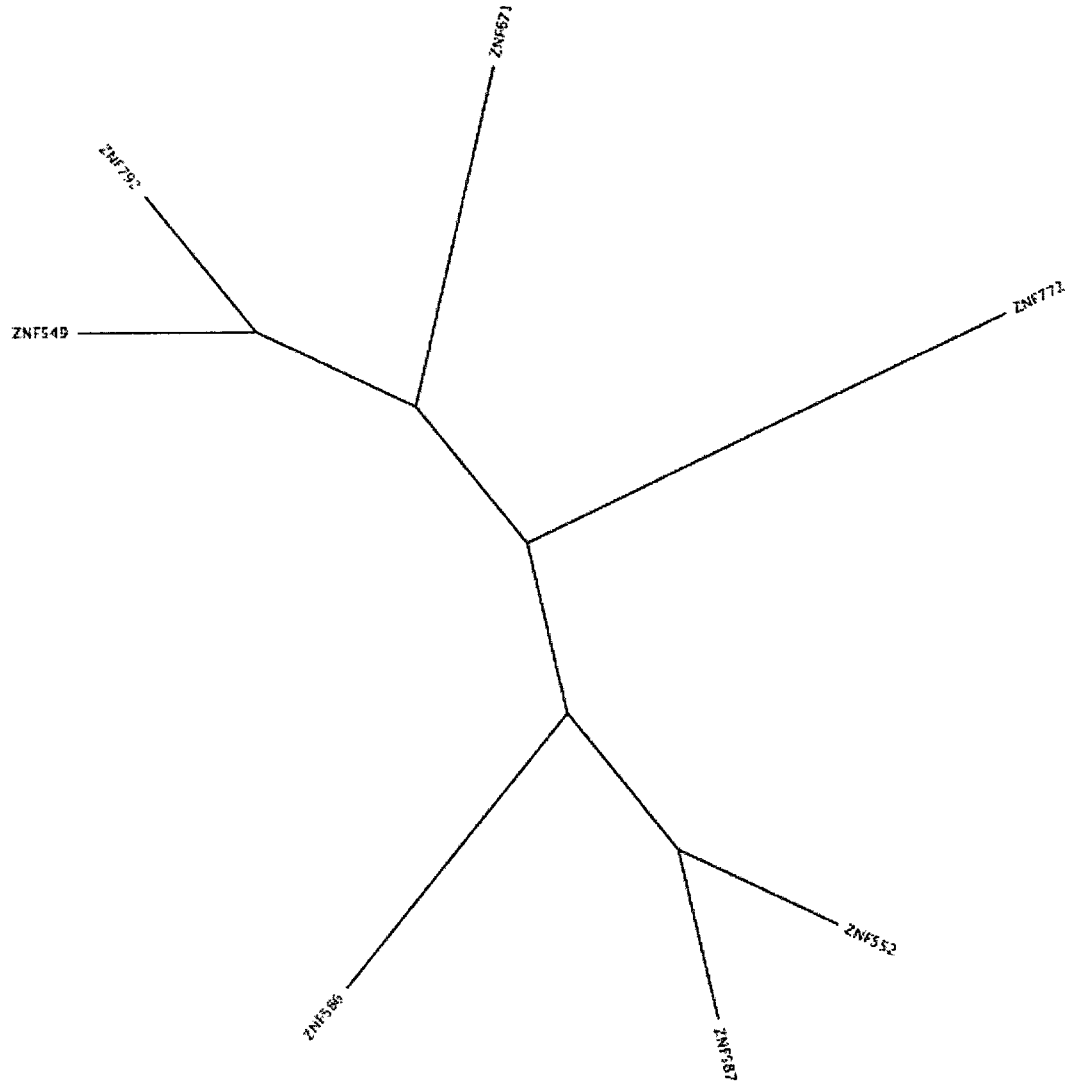


Figure 1. (continued)

Family 4

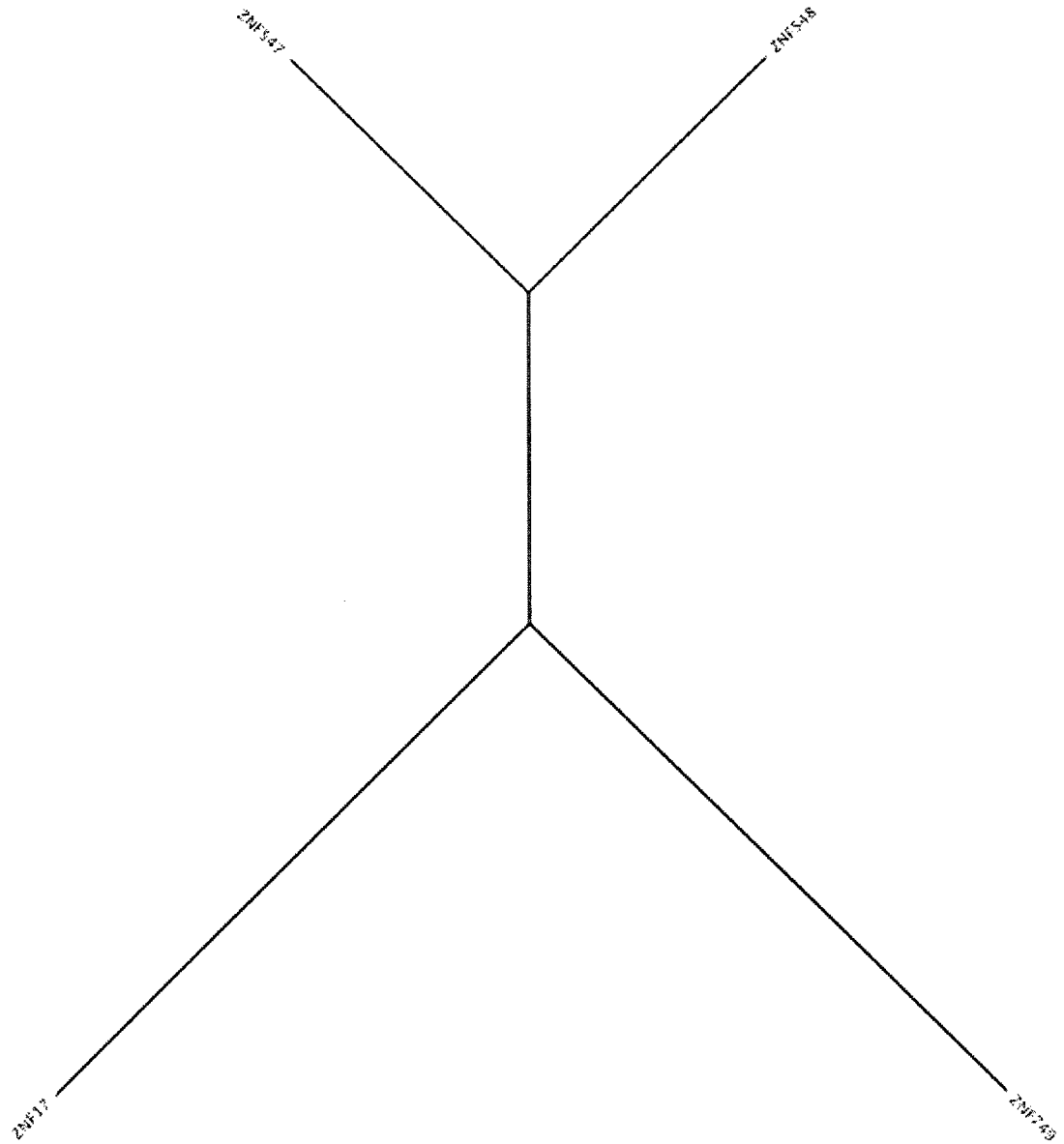


Figure 1. (continued)

Family 5

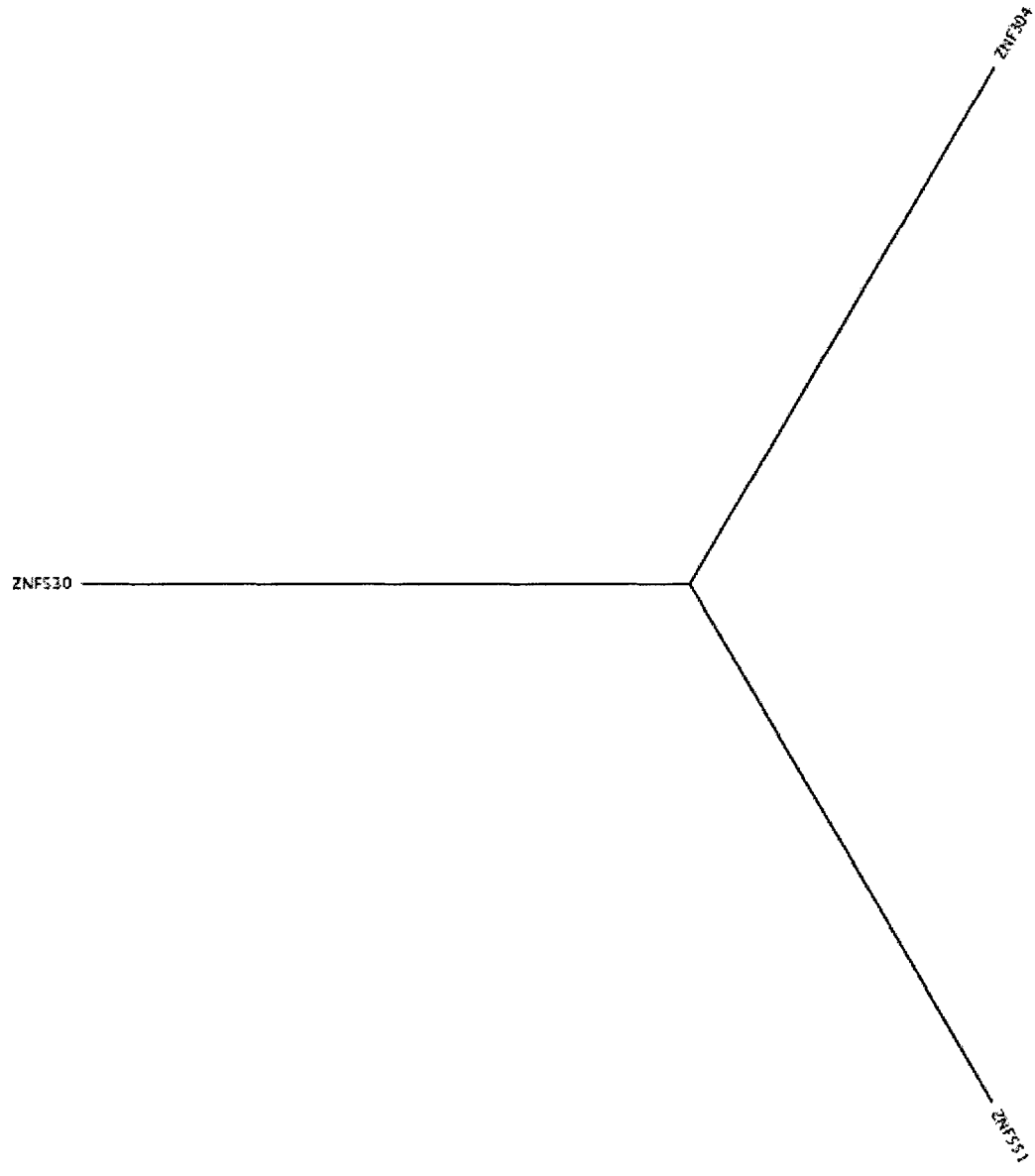


Figure 1. (continued)

Family 6

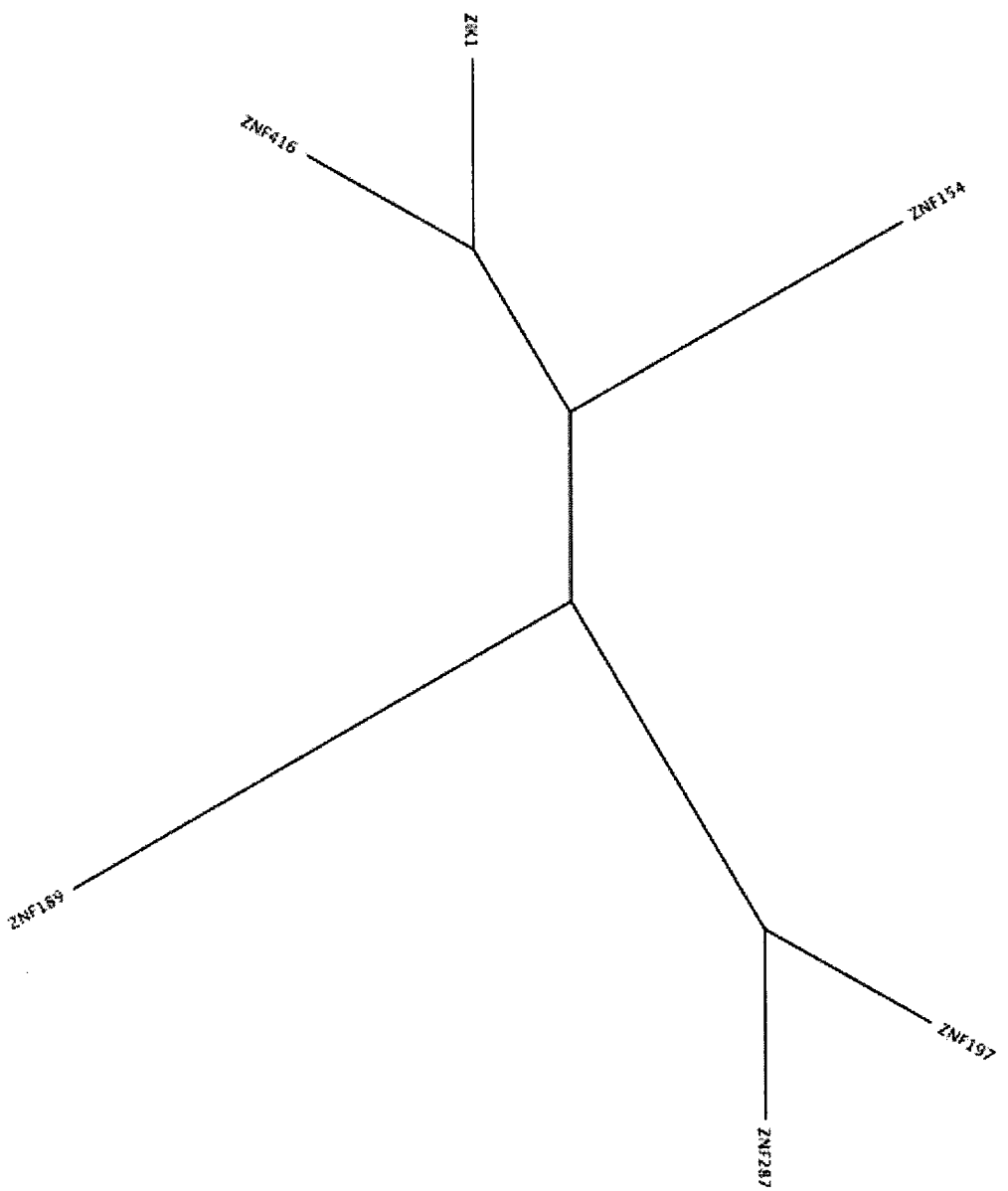


Figure 1. (continued)

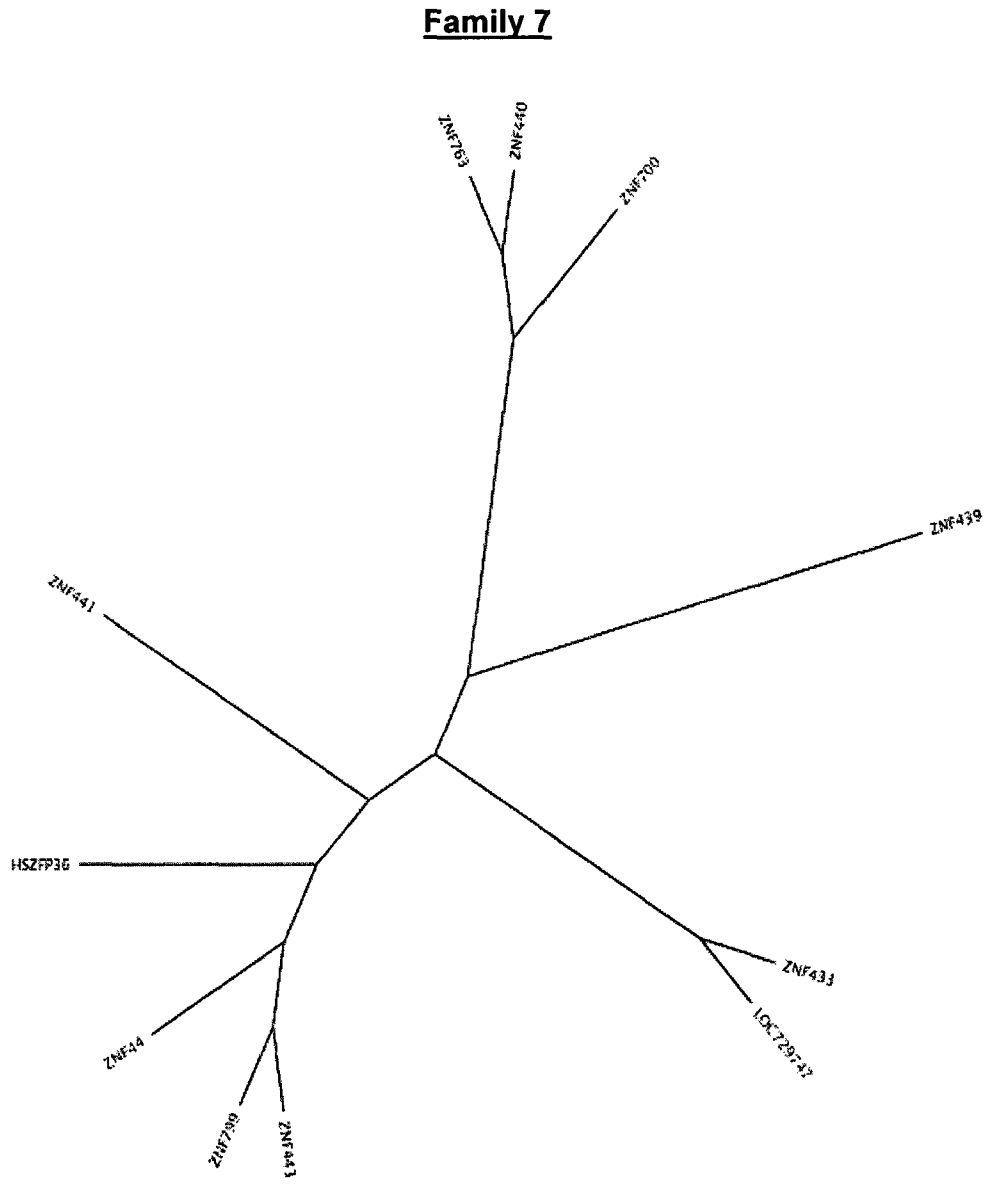


Figure 1. (continued)

Family 8

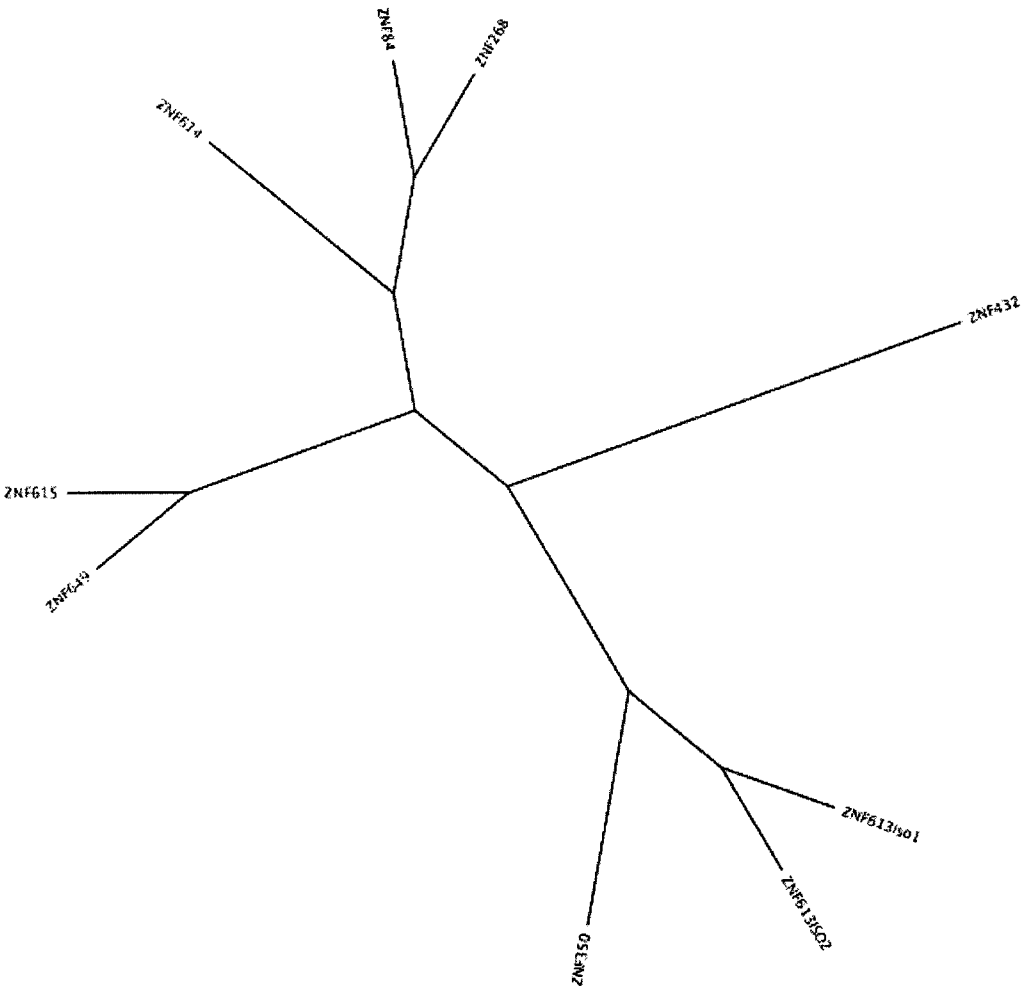


Figure1. (continued)

Family 9

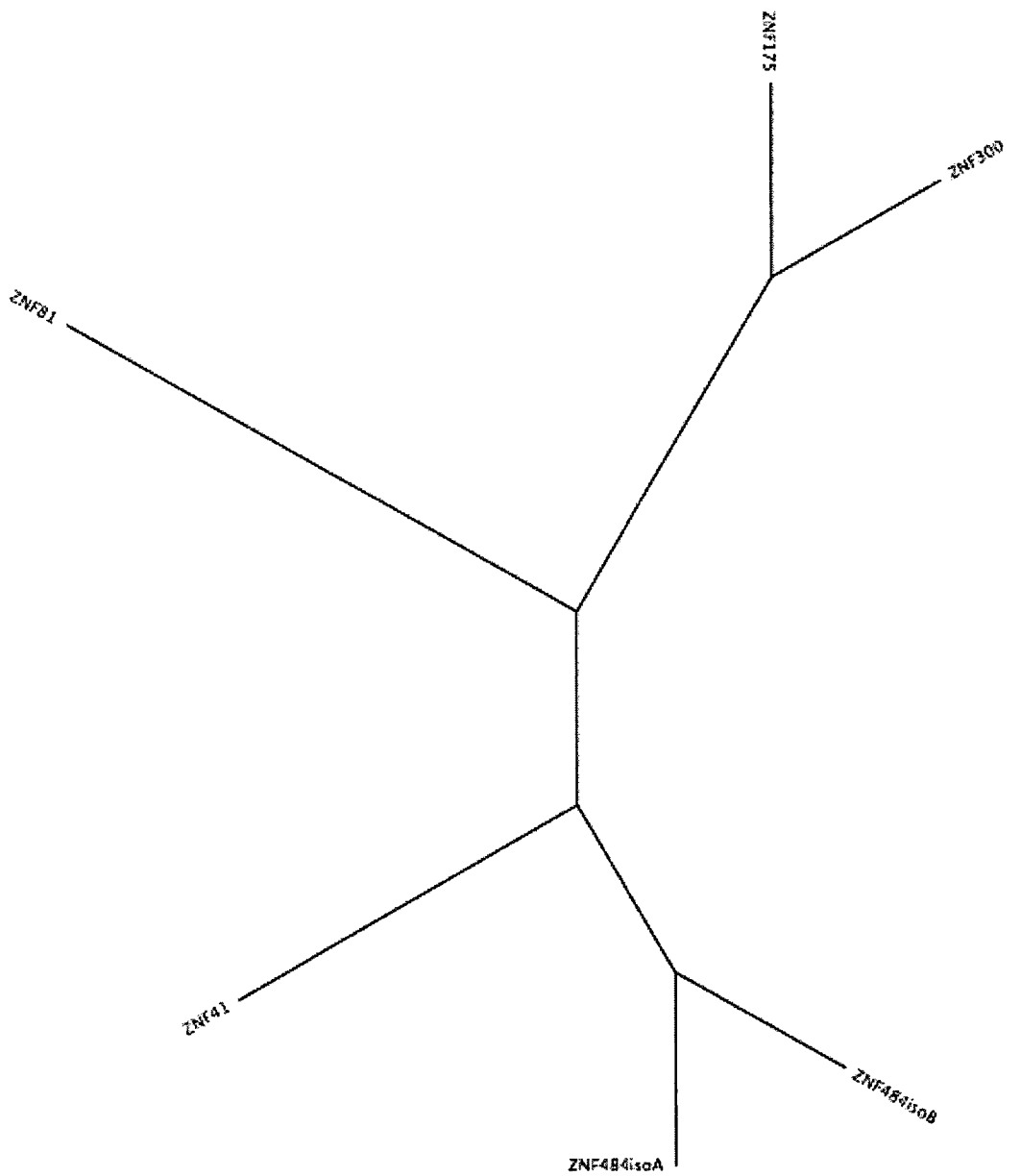


Figure 1. (continued)

Family 10

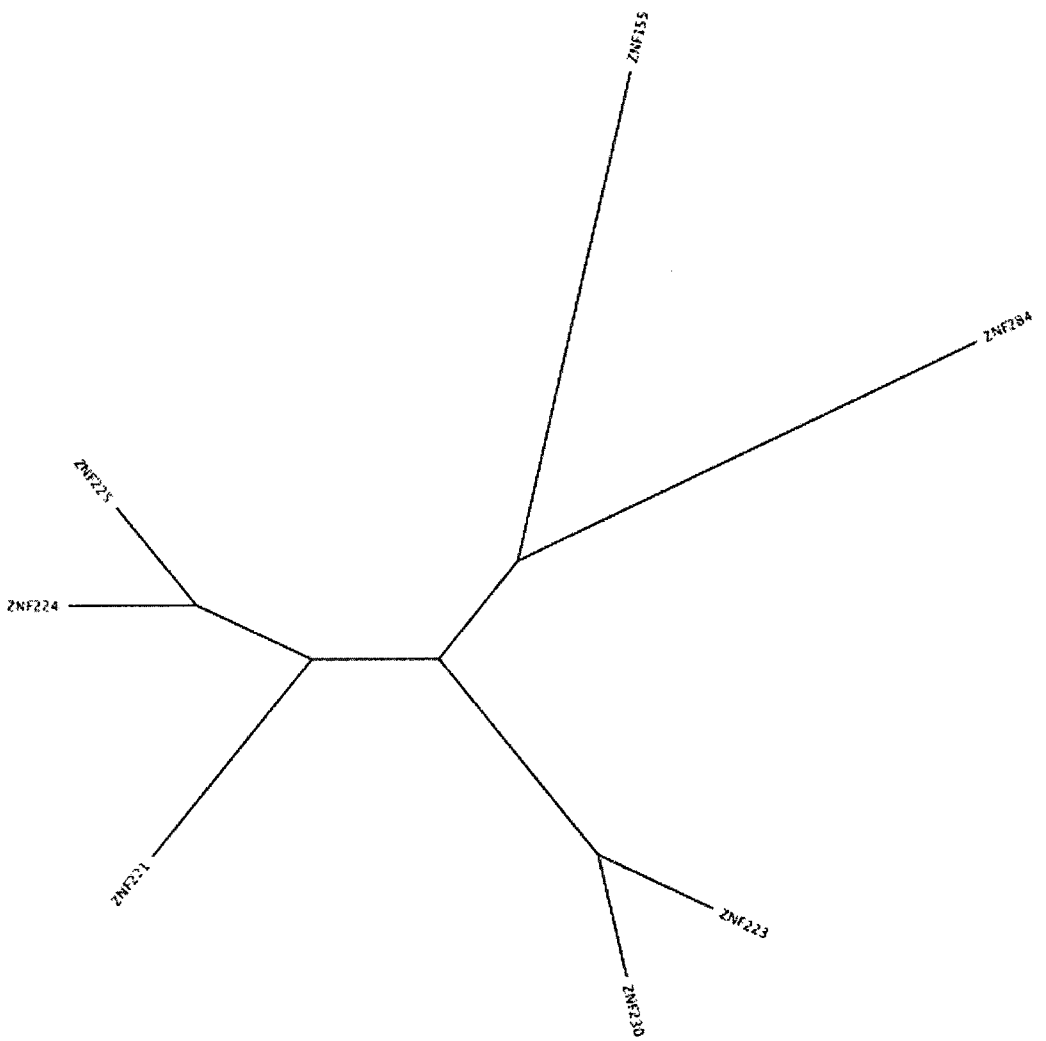


Figure 1. (continued)

Family 12

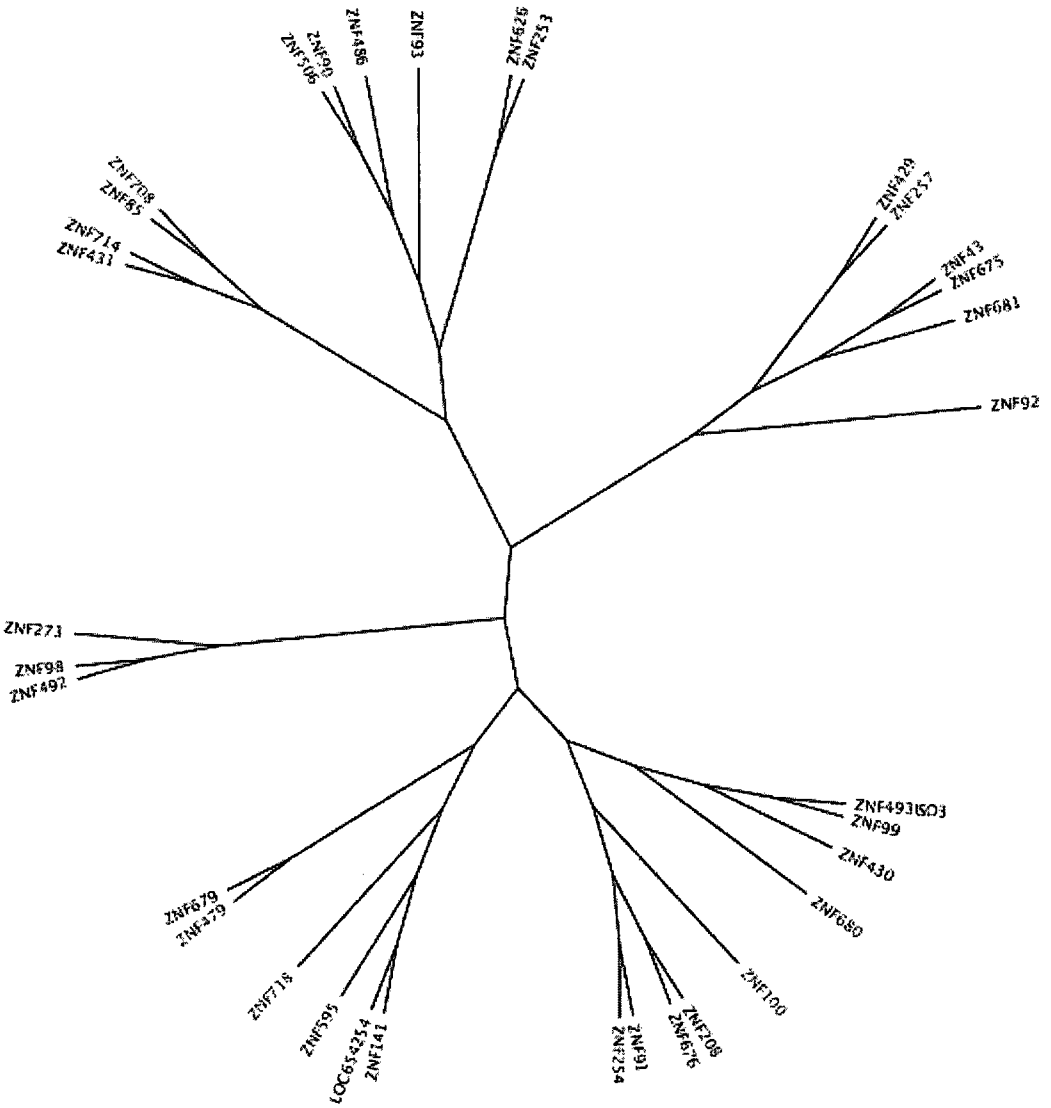


Figure 1. (continued)

Family 13

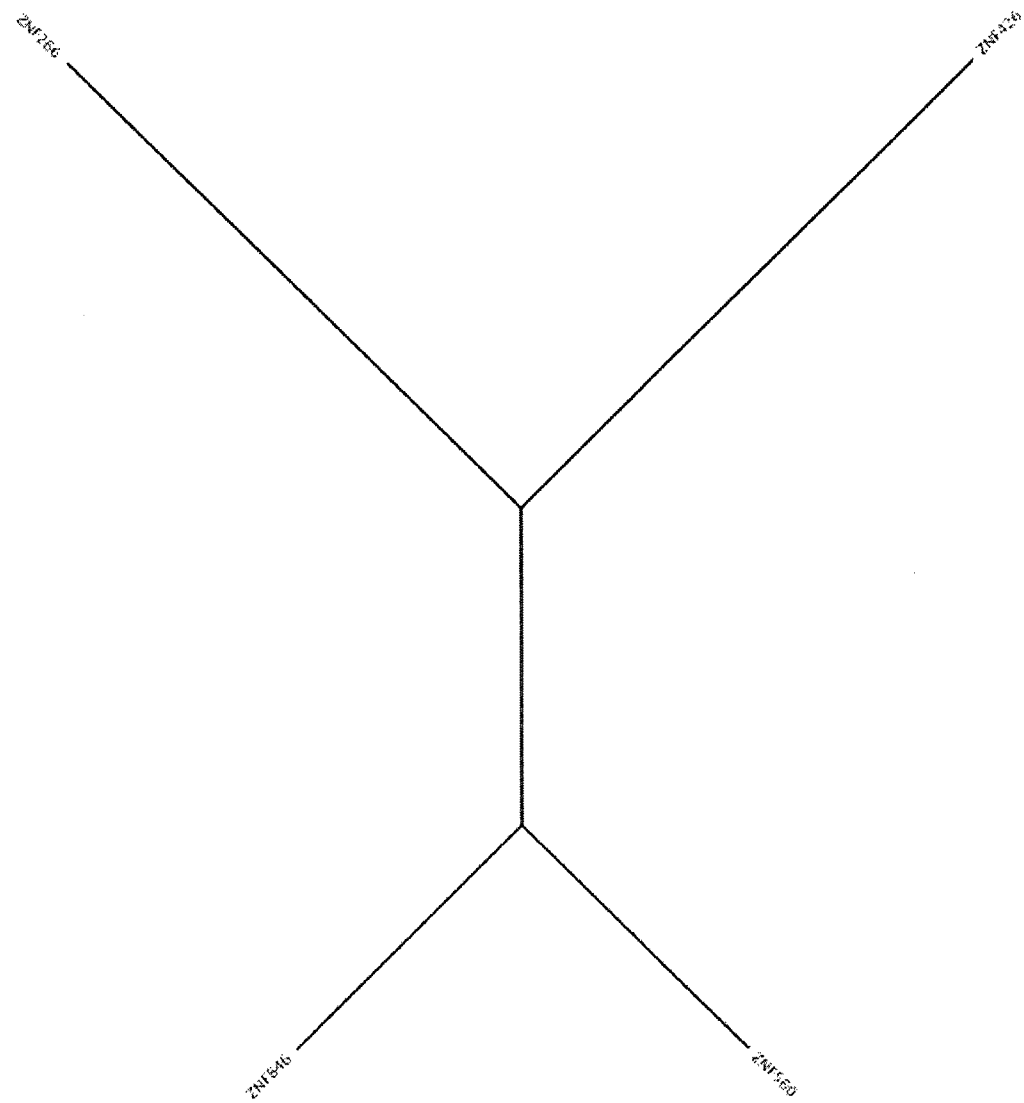


Figure 1. (continued)

Family 14

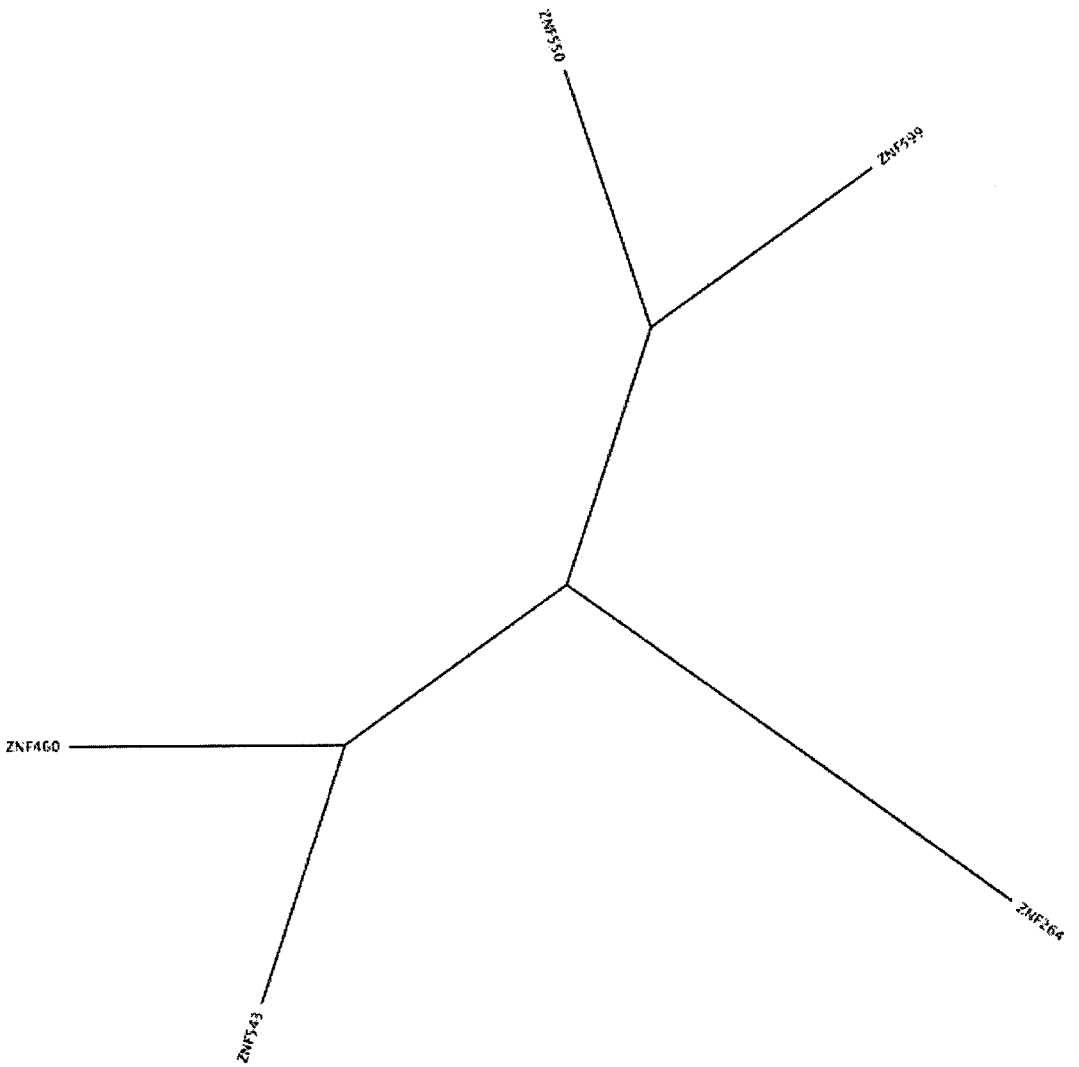


Figure 1. (continued)

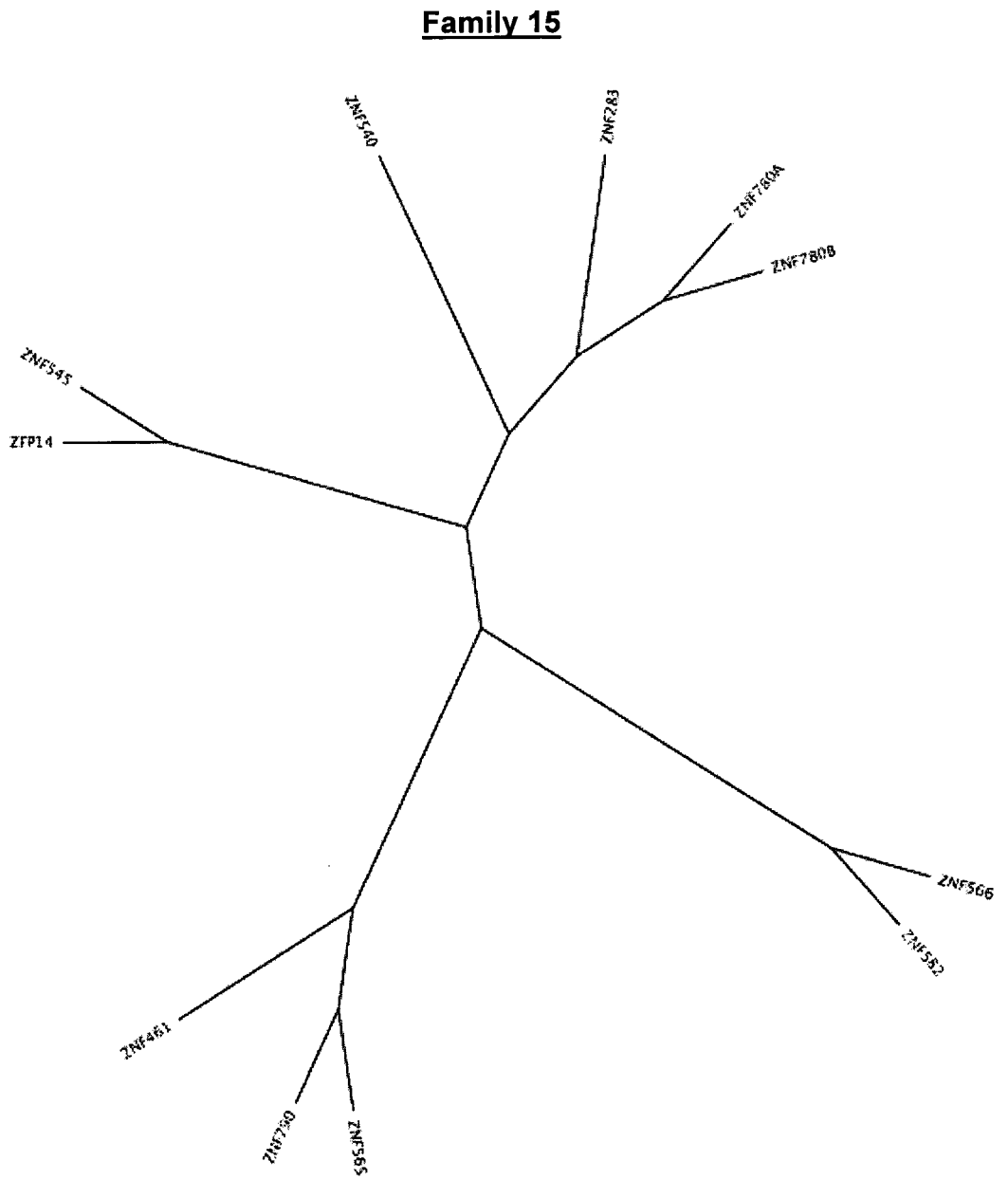


Figure 1. (continued)

Family 16

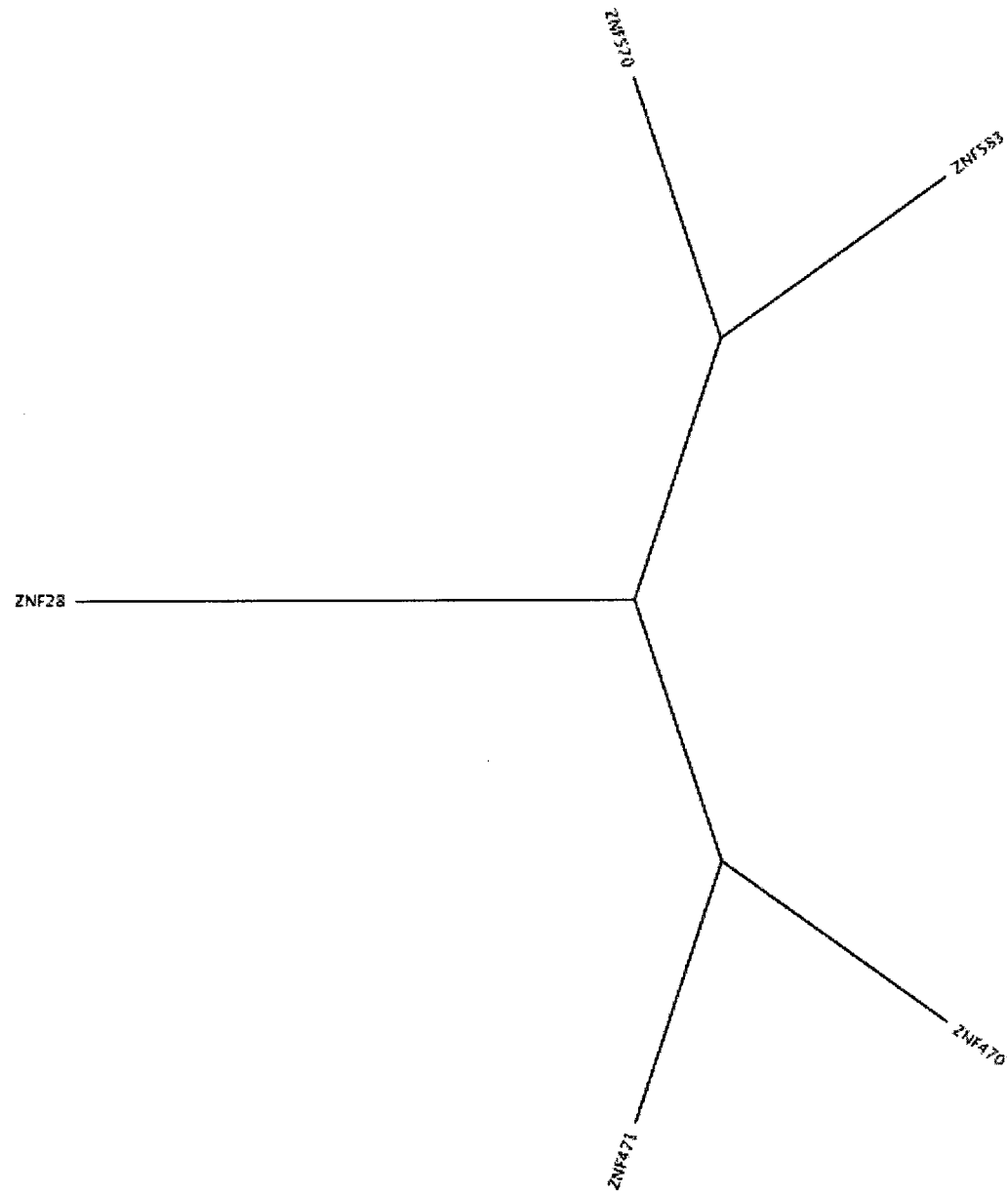


Figure 1. (continued)

Family 17

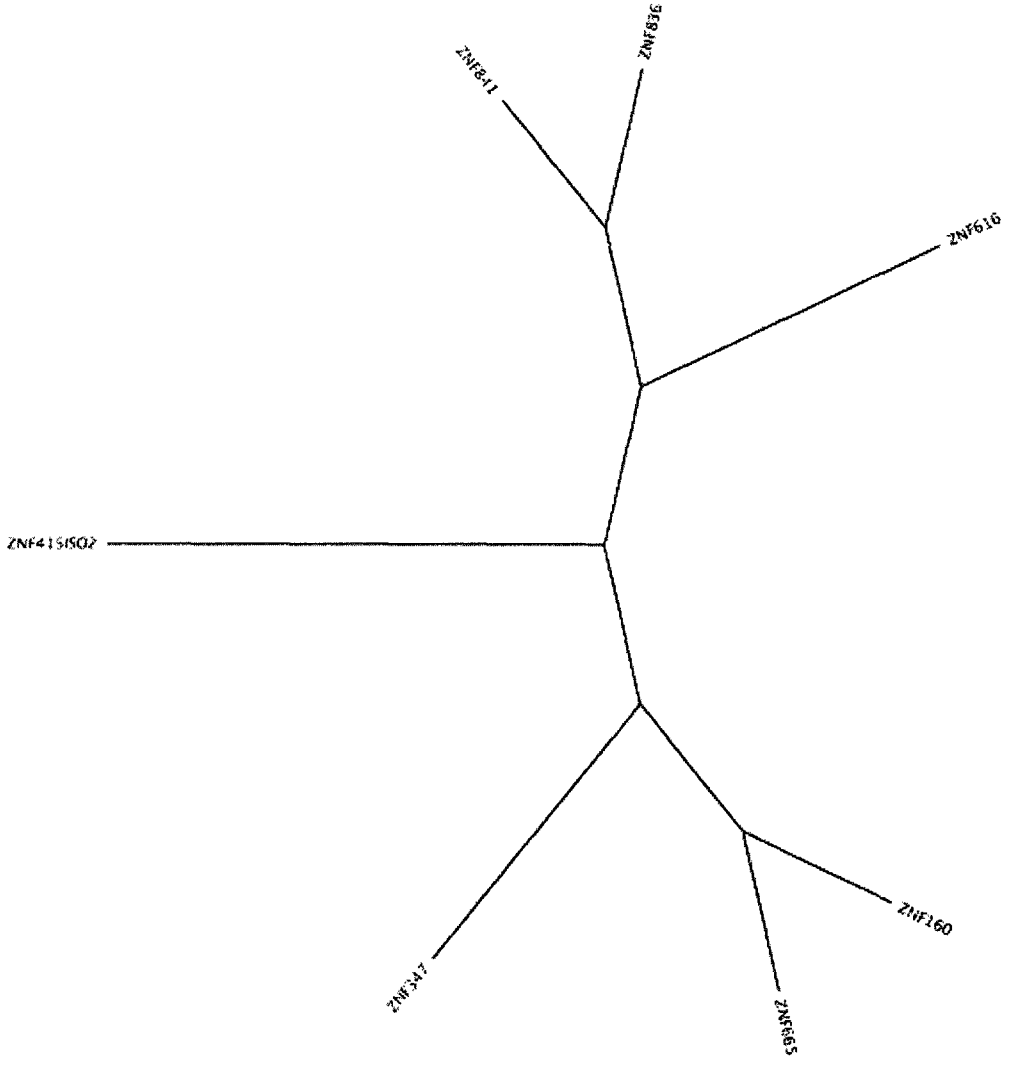


Figure 2. Phylogram tree showing the evolutionary history between all 18 "linker families".

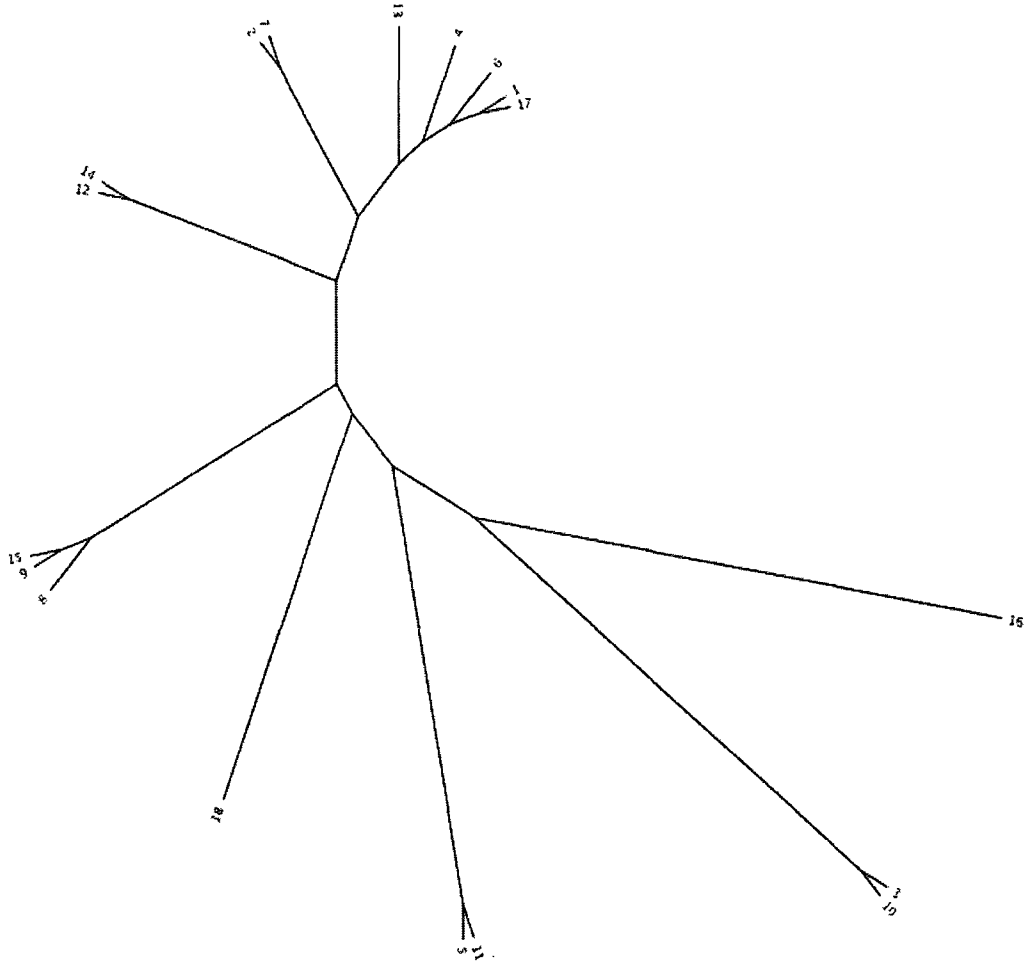


Figure 3. Phylogram tree showing an evolutionary history between each of the 334 individual linker regions of the KRAB-ZFP superfamily.

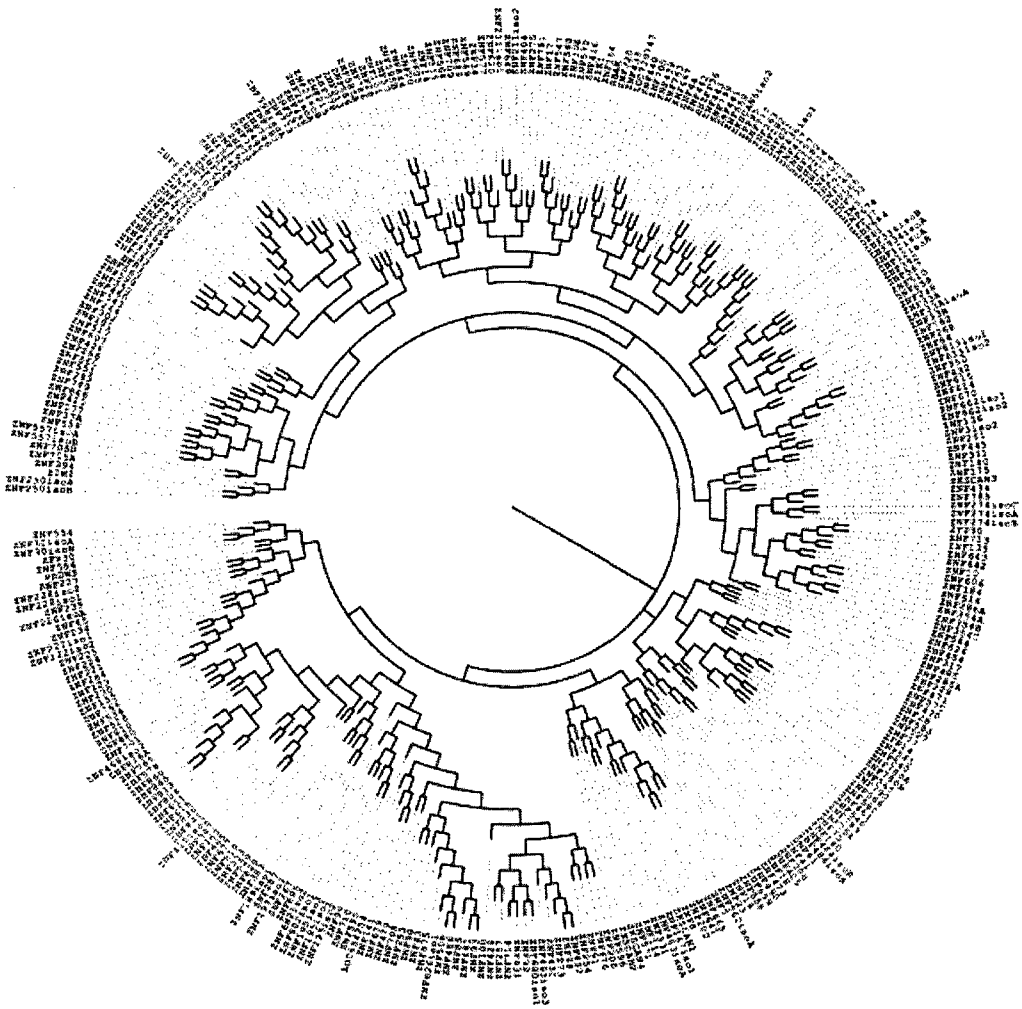


Figure 4. Diagrammatic representation of the KRAB-ZFP Subset. The N-terminal KRAB module serves as a repression domain. The C-terminal Zn-Finger repeat serves as a DNA-binding domain. Each member has a linker region of various length which may serve as a mediator for unknown and unconventional functions.

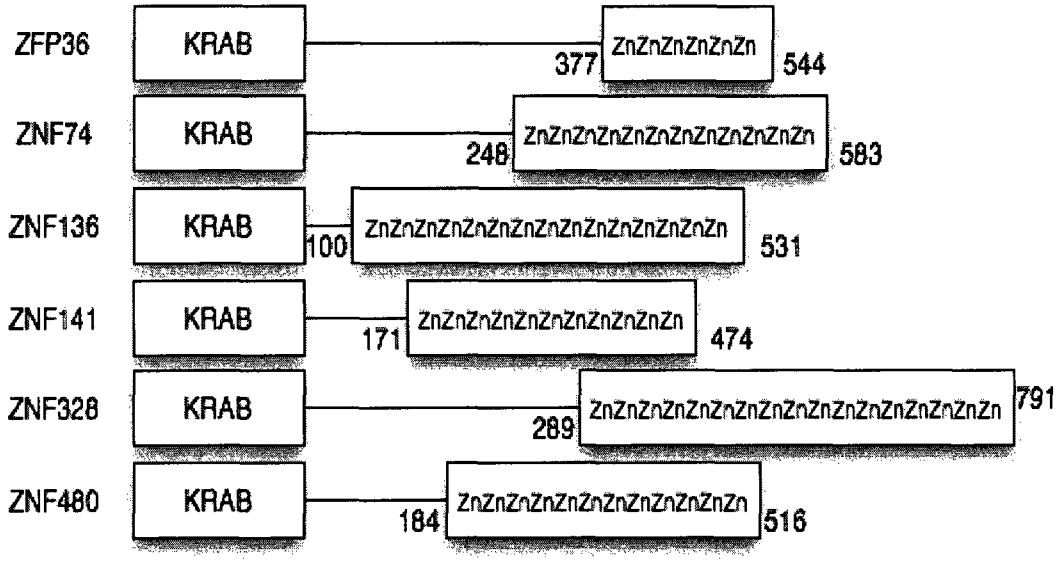


Figure 5. Diagrammatic scheme of the fusion constructs made from the DNA-binding regions of the 6 members of the KRAB-ZFP superfamily. The N-terminal GST tag is used for GSH-affinity purification. The C-terminal Zn-finger repeat is the region of our protein-of-interest which is under investigation.

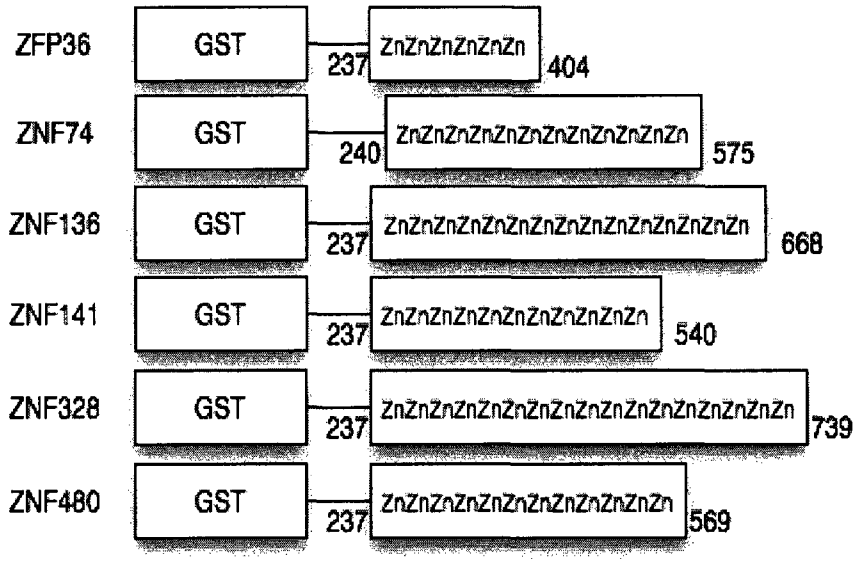


Figure 6. Expression profiles for each of the six expressed members. M=Marker, U-Uninduced Sample, I-Induced Sample, *U*-Uninduced Control, C-Induced Control. Black arrows - Fusion protein; Red arrows - GST.

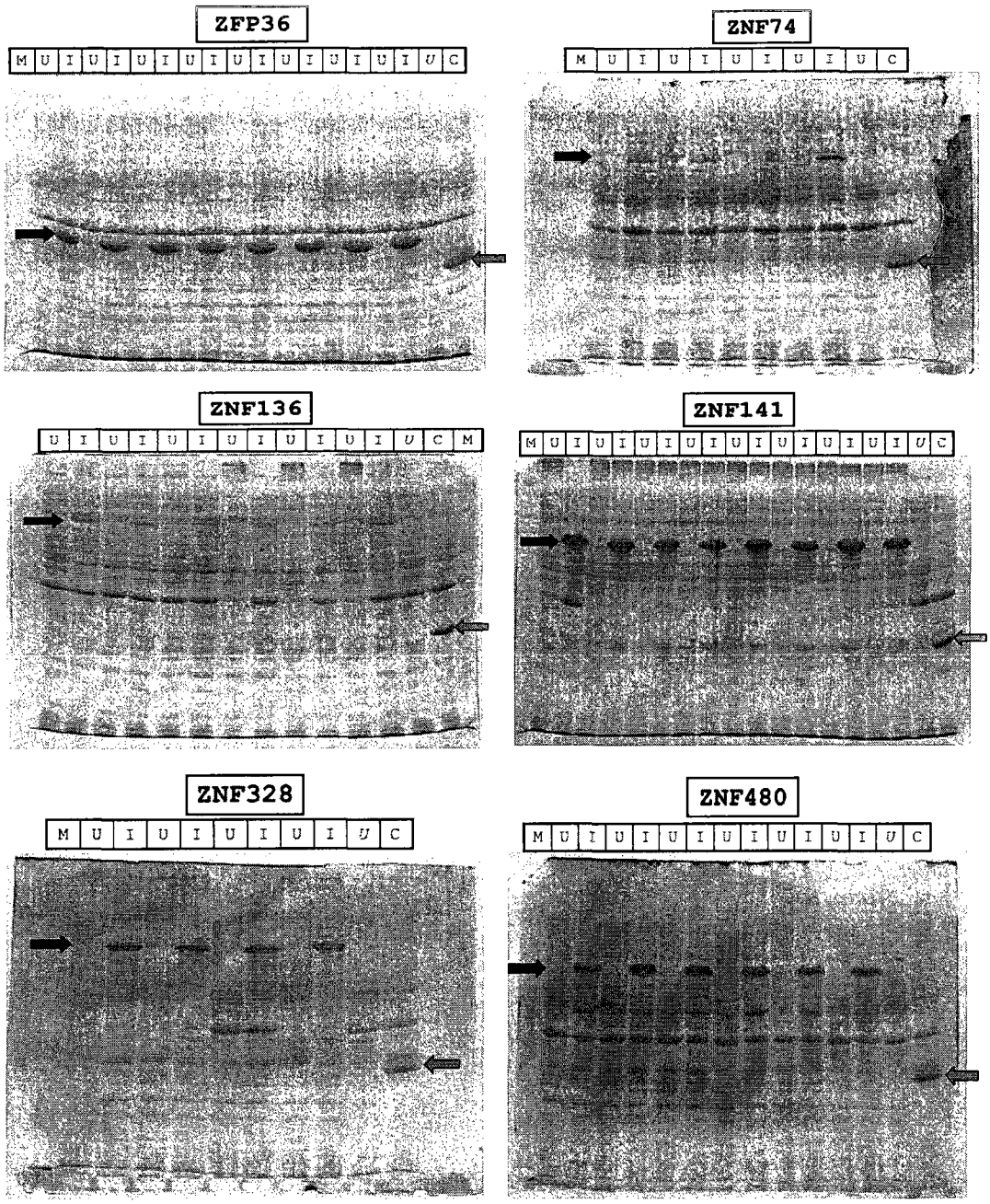


Figure 7. First GSH Affinity Purifications. I-Induced Sample, *I*-Induced Sample, F-Flowthrough, Lanes marked 1, 2 and 3 are Elutions 1, 2 and 3 respectively. Black arrows - Fusion proteins; Red arrows - GST pure protein.

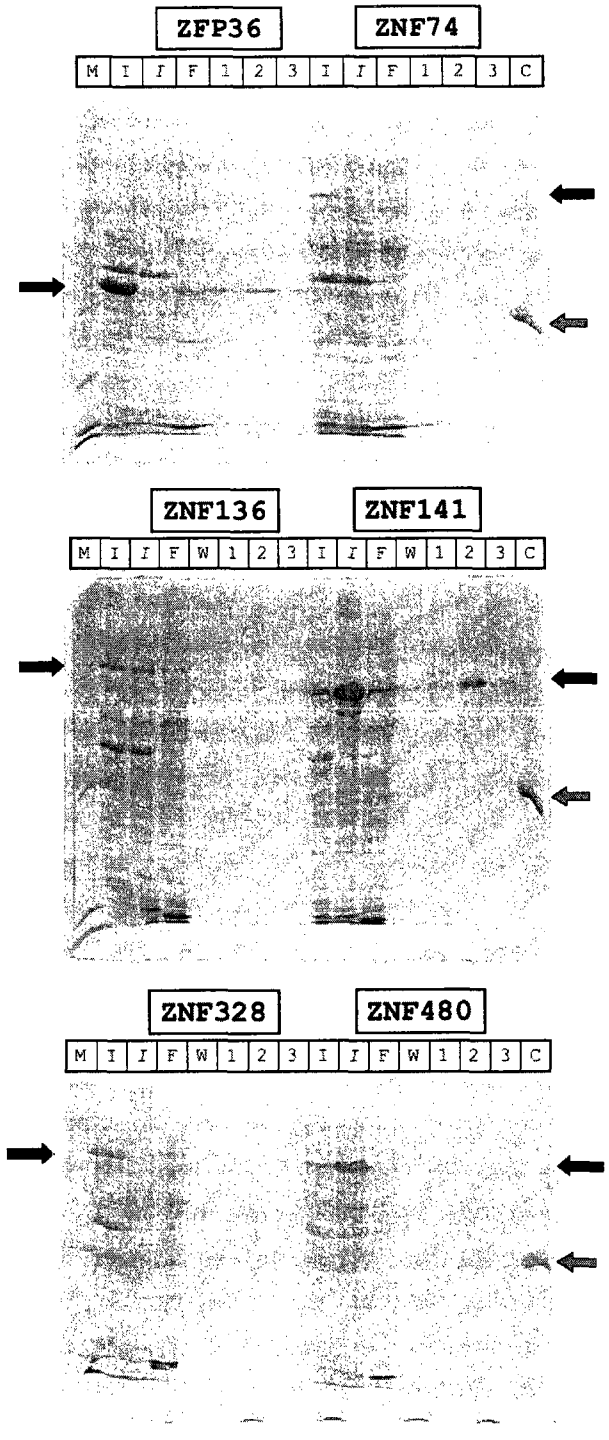


Figure 8. GSH Affinity Purifications. After reworking the protocol, all members were successfully purified and eluted. I=Induced Sample, F=Flowthrough, Lanes marked 1 through 6 are Elutions 1 through 6 respectively. 'C' denotes GST control. Black arrows - Fusion proteins; Red arrows - GST pure protein.

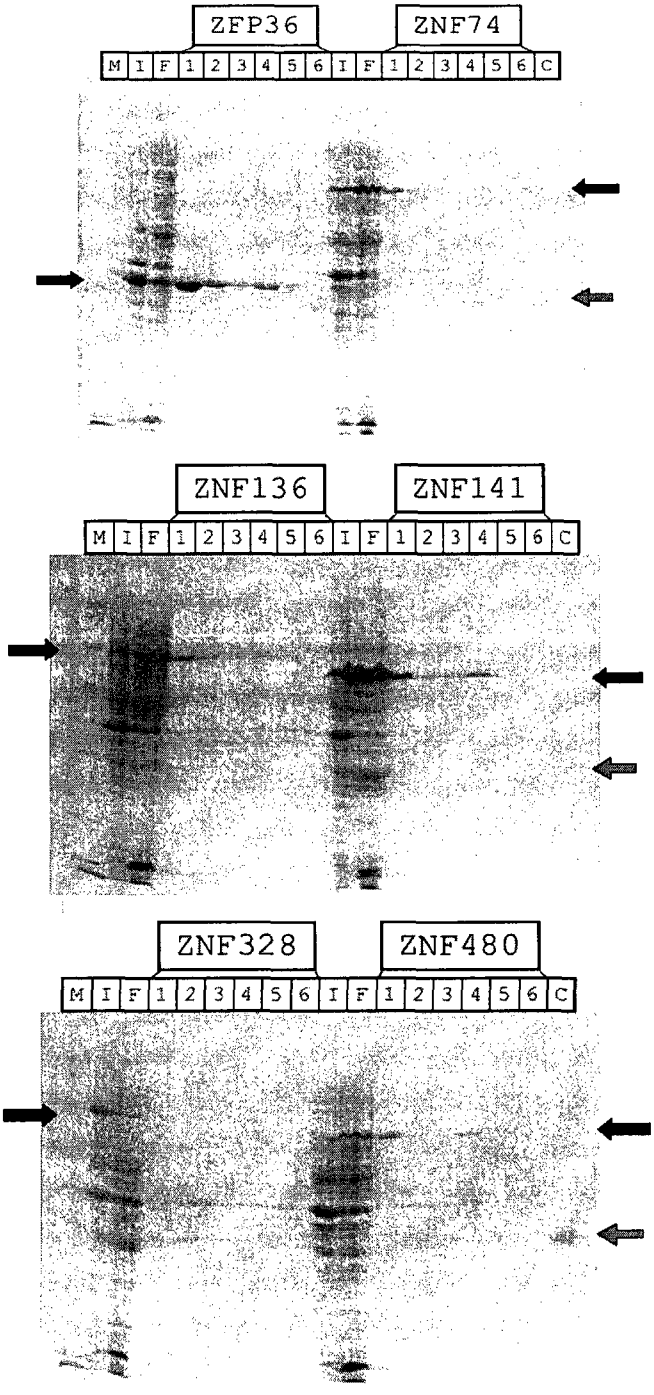


Figure 9. Final three GSH purifications of proteins that were folded *in vivo*. All elutions were used regardless of band intensity. Lanes marked as 1 are elutions 1-3 combined; Lanes marked 2 are elutions 4-6 combined. Black arrows/boxes denote target fusion proteins. Red arrows denote pure GST as a control.

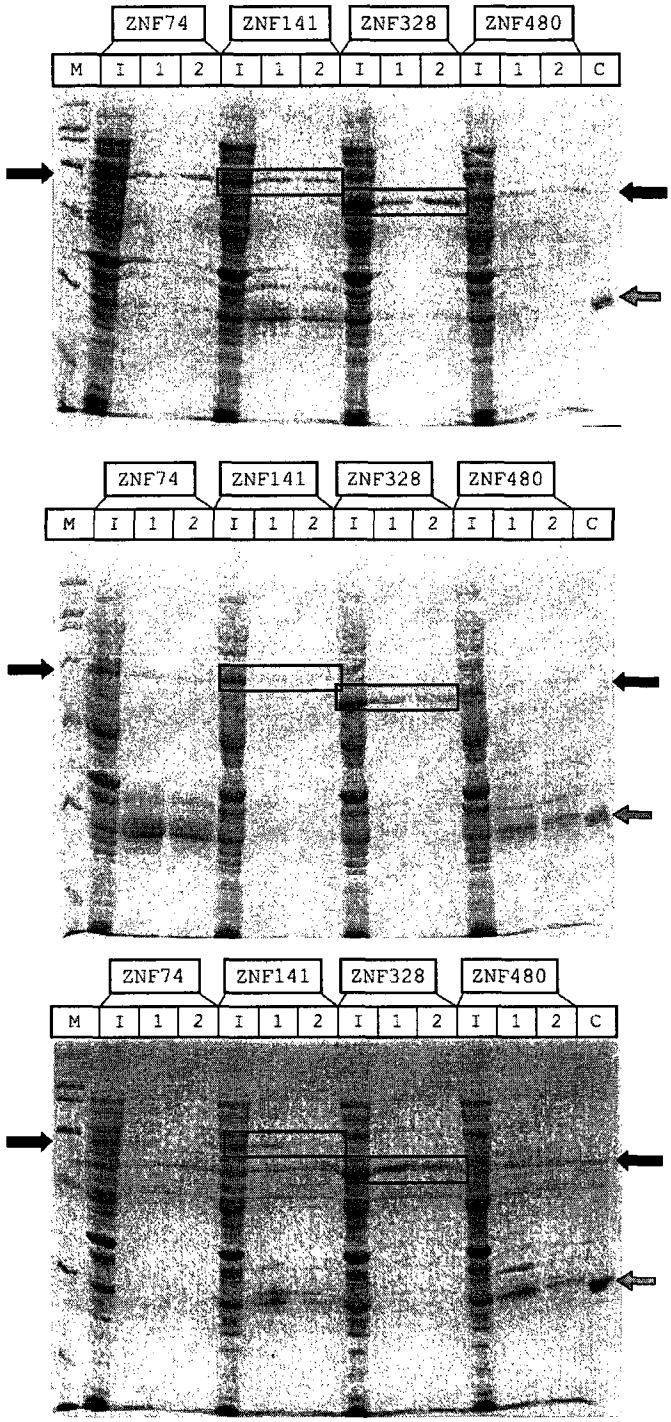


Figure 10. Cold-binding to pre-enrich the randomized oligomer library before hot-binding. The bands in lanes marked 1,2,3,4,5, and 6 are oligomers that were selected by ZFP36, ZNF74, ZNF136, ZNF141, ZNF328, and ZNF480, respectively. Black arrows denote the oligomer bands.

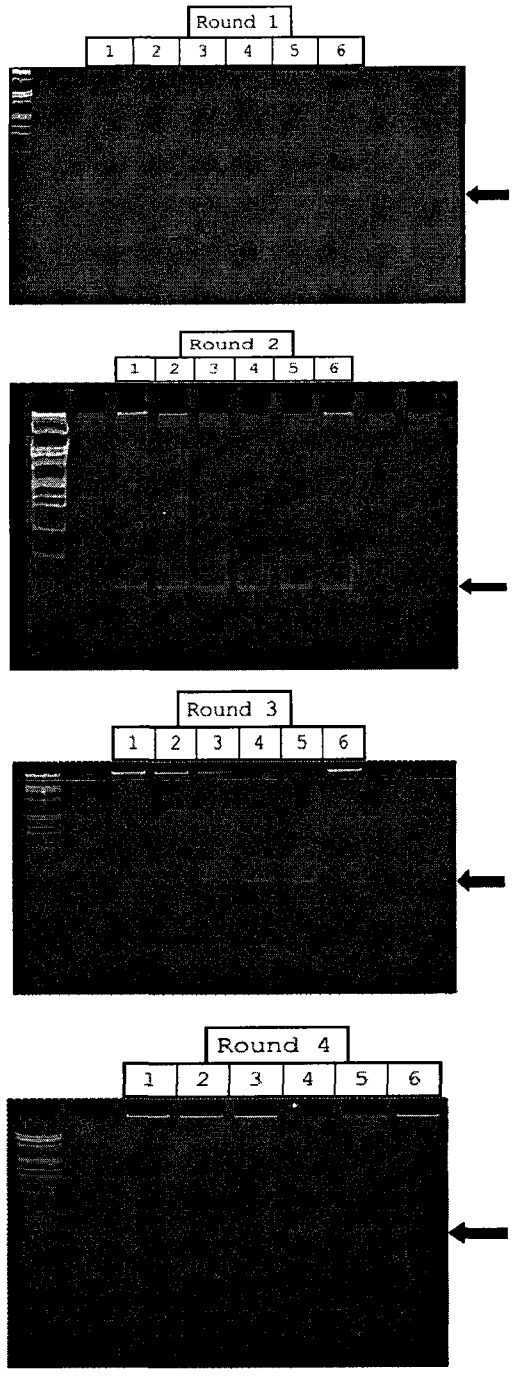


Figure 11. Radiographs from 3 rounds of "Hot-Binding" (EMSA). A radiolabeled randomized library of oligonucleotide 49-mers was subjected to binding with the DNA-binding domains of 6 known KRAB-ZFPs. These DNA-protein complexes are large and will be retarded in their movement through a DNA-PAGE gel. This allows for the selection of specific oligomers that bind to these DNA-binding domains. Black Arrows – DNA-protein complex Red Arrows – Free DNA Oligomer.

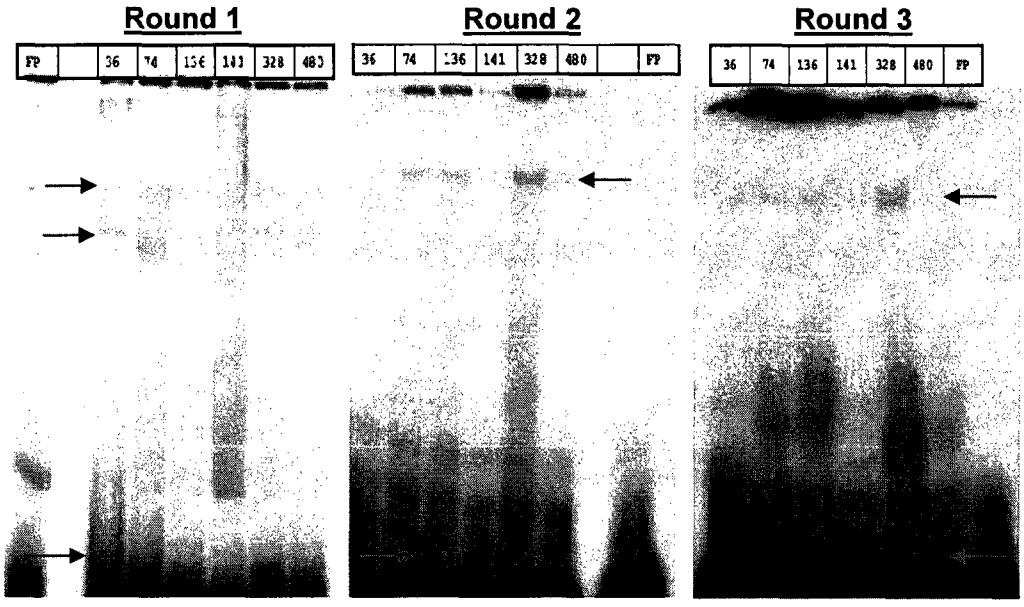


Figure 12. Hot-binding to select oligomers that specifically bind to the DNA-binding domains of our KRAB-ZFP subset. Bands represent selected oligomers after PCR. Round 3 involved loading entire PCR reaction and electro-eluting the oligomers from the cut bands.

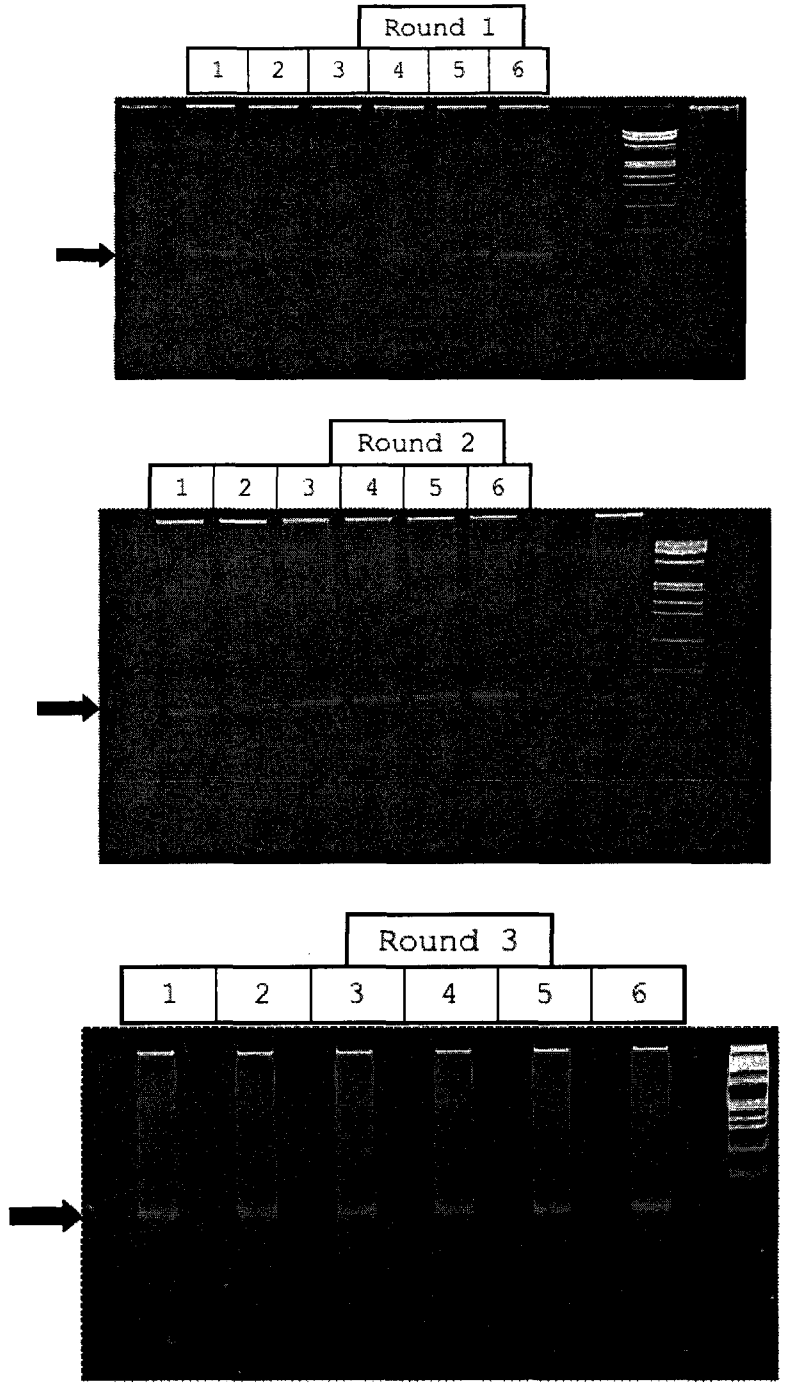


Figure 13. Digestion of selected clones to check for the presence of the oligomer insert. Black arrow denotes the insert that has been cut out of the pUC18 plasmid. The plasmid remains in the well. ZFP36: 1-30; ZNF74: 31-60; ZNF136: 61-90; ZNF141: 91-120; ZNF328: 121-130; ZNF480: 131-160.

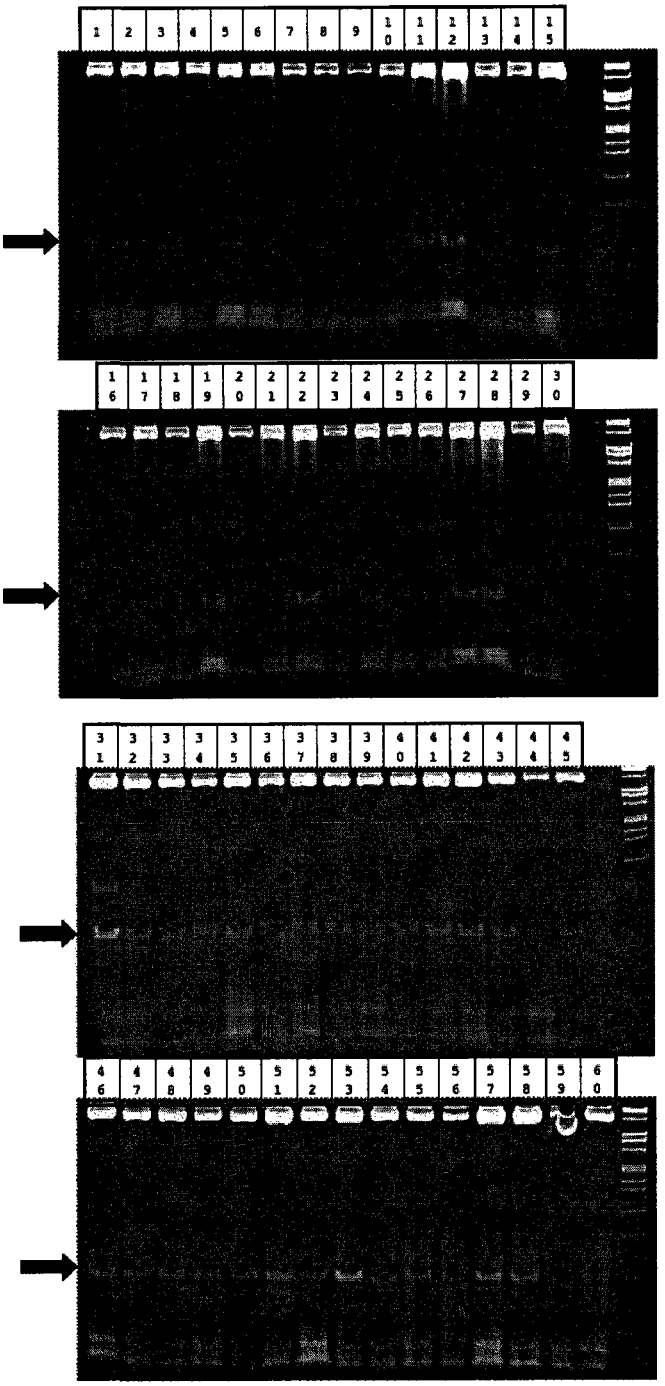


Figure 13. (Continued)

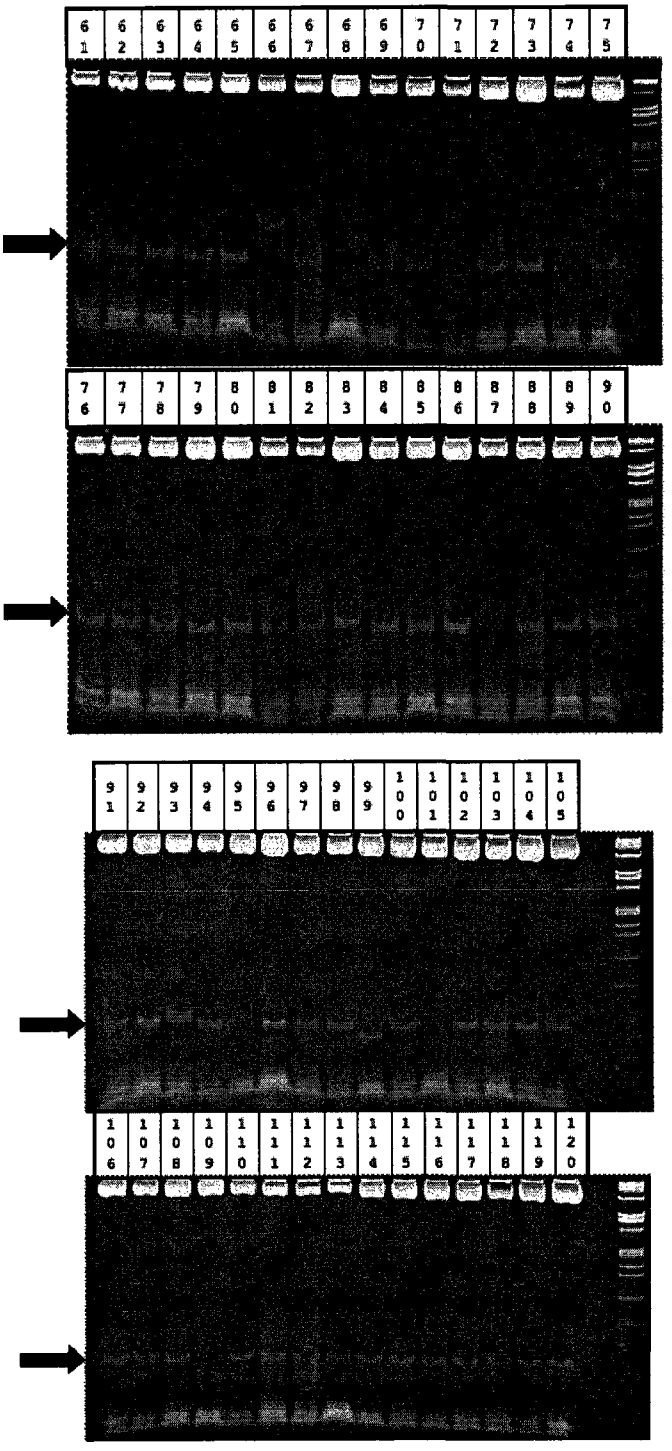


Figure 13. (Continued)

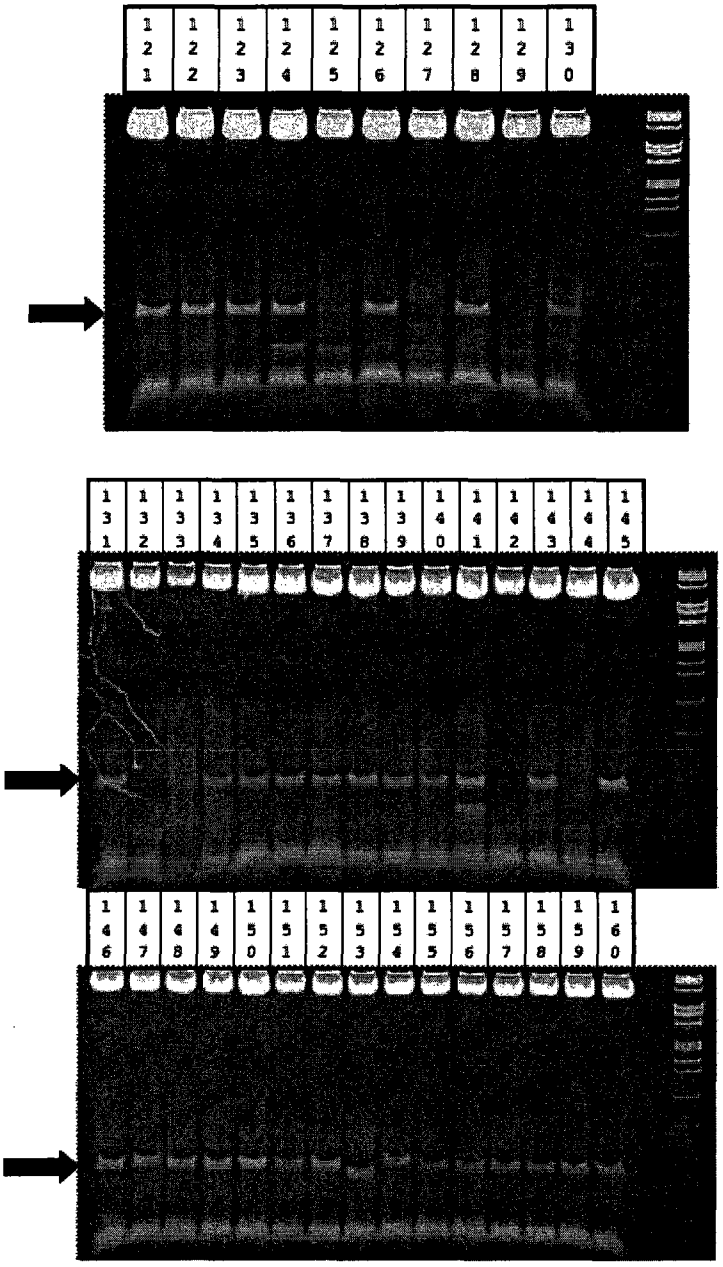


Figure 14. Plasmid concentration prior to sequencing. After the insert check, each clone was run on 1.2% agarose to determine the concentration of plasmid DNA. ZFP36: 1-30; ZNF74: 31-60; ZNF136: 61-90; ZNF141: 91-120; ZNF328: 121-130; ZNF480: 131-160.

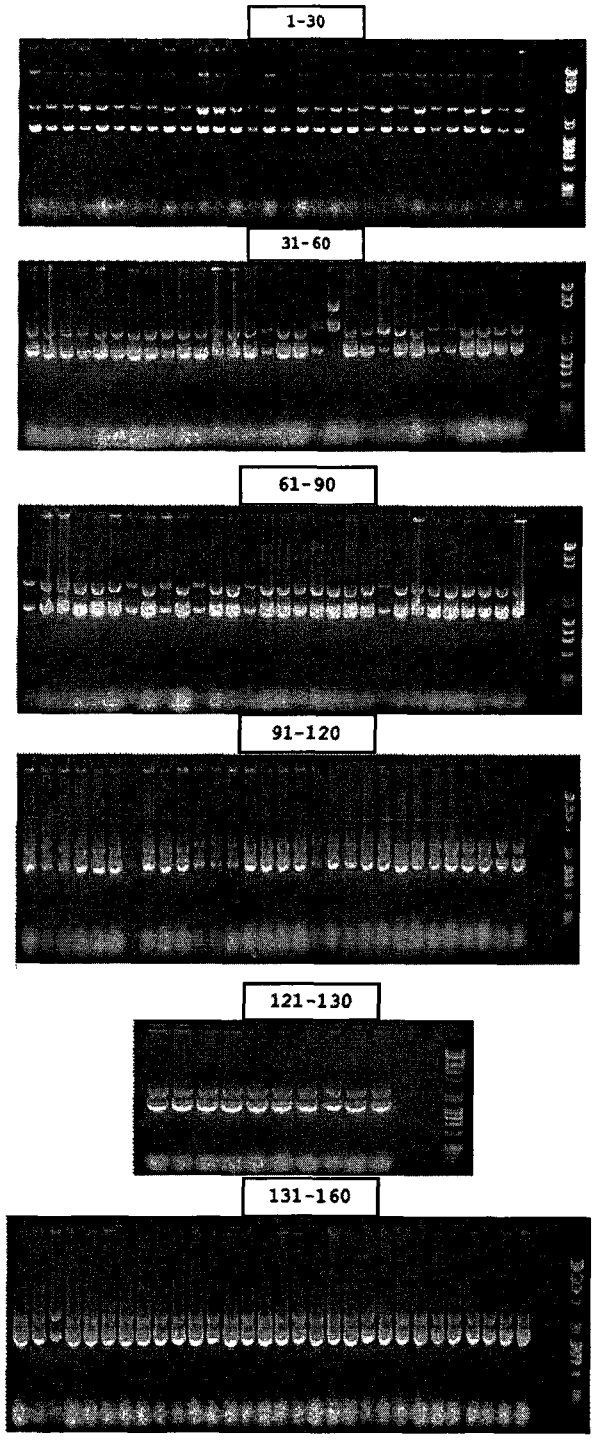


Figure 15. Consensus maps for ZNF36, ZNF74, ZNF136, ZNF141, & ZNF480. Numbers are percentages of occurrence.

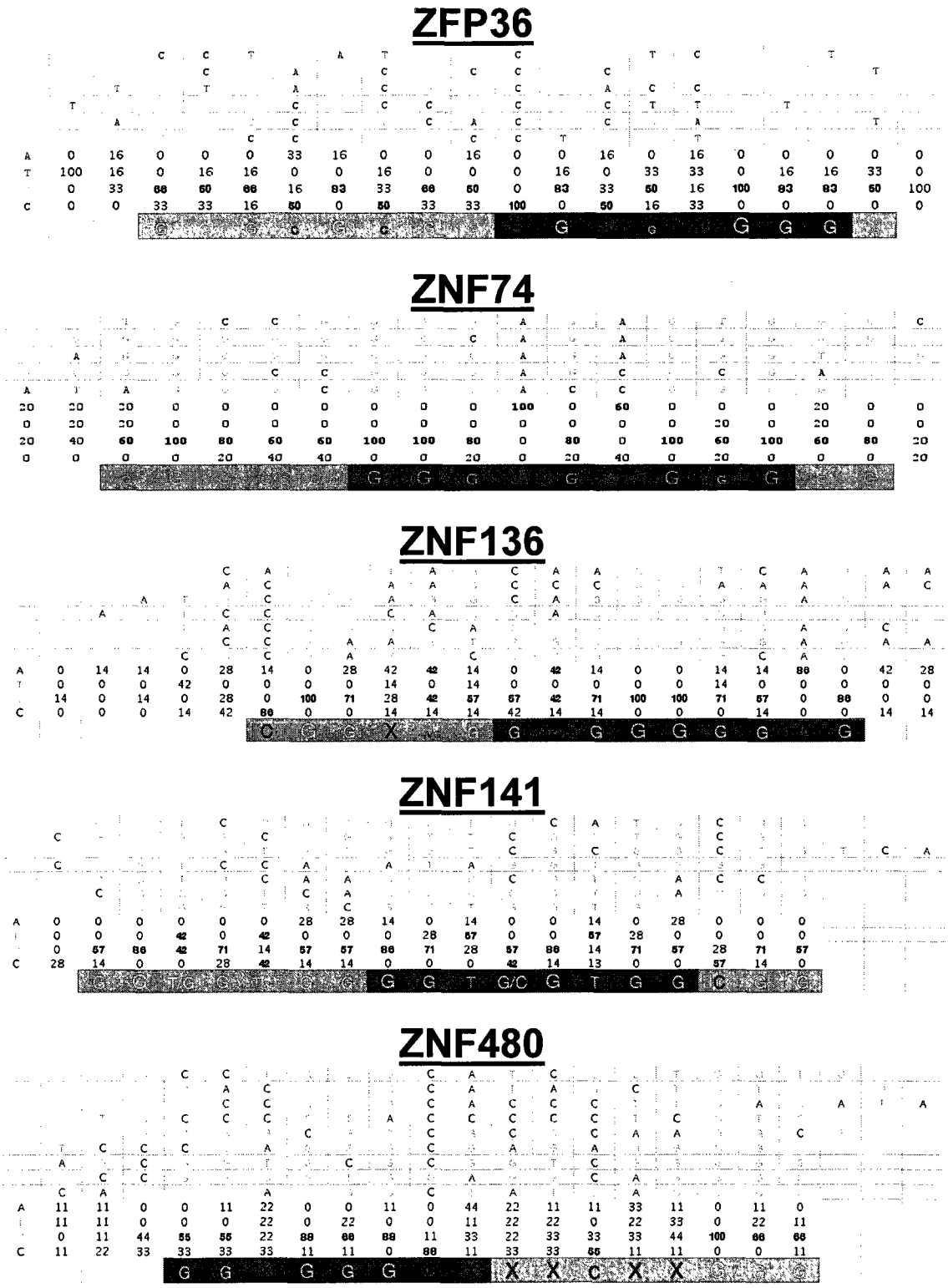


Figure 16. List of potential target genes for human ZFP36. The first box is for the consensus binding site, and each box following represents the derivative seqeunces. Target genes for all sequences were derived from two sources: Transfac and EPD. The gene name, binding site location, and associated binders are listed for Transfac derived entries. For EPD-derived entries, the gene names, genomic and mRNA accession links, and 60nt upstream of the promoter are shown. Letters in red denote our binding site. Capitalized letters denote previously observed binding sites. Letters in blue or green denote some observed importance.

ZFP36

Consensus 1 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	PTK2 protein tyrosine kinase 2	-142011545	[illegible]	-142011485	N-Myc	8	-ve
	N-myc downstream regulated 1	-134309574	[illegible]	-134309514	EGR-1	8	-ve
	N-myc downstream regulated 1	-134309574	[illegible]	-134309514	SP-1	8	-ve
	chemokine (C-X-C motif) receptor 4	-136876203	[illegible]	-136876143	c-Ets-1	2	-ve
	high-mobility group nucleosome binding domain 1	-407211079	[illegible]	-407211019	ZFP219	21	-ve
	parathyroid hormone 1 receptor	46923627	[illegible]	46923687	MAZ/SP-1	3	+ve
EPD	high-mobility group nucleosome binding domain 1	-407211079	[illegible]	-407211019		21	-ve
	endothelin receptor type B	-78493001	[illegible]	-78492941		13	-ve
	protease, serine, 23	86511509	[illegible]	86511569		11	+ve
	solute carrier family 35 (CMP-sialic acid transporter), member A1	88182641	[illegible]	88182701		6	+ve
	protein disulfide isomerase family A, member 6	-10352928	[illegible]	-10352868		2	-ve
	H3 histone, family 3B (H3.3B)	-73775899	[illegible]	-73775839		17	-ve
	structural maintenance of chromosomes 2	106556808	[illegible]	106556868		9	+ve
	proteasome (prosome, macropain) subunit, alpha type, 4	78832750	[illegible]	78832810		15	+ve

ZFP36

Derivative 1 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	serine/threonine kinase 16	220110185	[illegible]	220110245		2	+ve

ZFP36

Derivative 2 [illegible]

SAME AS CONSENSUS

Figure 16. (continued)

ZFP36

Derivative 3 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	POU class 2 homeobox 1	167190077	gggagccagagcgagggagggtttattgacggccgatttggttaaaatattcaaaatg	167190137		1	+ve

ZFP36

Derivative 4 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	tryptophan hydroxylase 2	72332131	acctttatttgattattacacaTtgtacgtttgtgtcaaaatatcacatgtgccttataaa	72332191	none	12	+ve
EPD	serine palmitoyltransferase, long chain base subunit 1	-94577716	gagcagcgacgcgcactttgggac[illegible]	-94577656		9	-ve
	pituitary tumor-transforming 1 interacting protein	-46293617	[illegible]	-46293557		21	-ve
	DnaJ (Hsp40) homolog, subfamily A, member 2	-47307579	[illegible]	-47307519		16	-ve
	CDK5 regulatory subunit associated protein 1	-31989357	gcctgaagtagcggcggaa[illegible]	-31989297		20	-ve

ZFP36

Derivative 5 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	55523982	ccgggaagaagcgagacccaggtggcgtgacggaggggaggcgaagga[illegible]	55524042	SP1	4	+ve
	thrombomodulin	-23030602	actcatcccggcgaggtgatcctatgcgtga[illegible]gcgcaagggcggccagagaaccc	-23030542	SP1	20	-ve
EPD	nuclear receptor binding protein 1	27651443	gcaccgcccatcgttggtccgggc[illegible]	27651503		2	+ve
	ATP-binding cassette, sub-family B (MDR/TAP), member 9	150725474	tgtcttttcaccaatcaccgccaaggcgtgaggg[illegible]	150725534		7	+ve
	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7	74330667	actctgggagcgg[illegible]	74330727		16	+ve
	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member 1	-38804122	gaaagtgcgagtgcggaa[illegible]	-38804062		17	-ve
	M-phase phosphoprotein 6	-82203890	[illegible]	-82203830		16	-ve
	EH-domain containing 1	-64646137	[illegible]	-64646077		11	-ve

ZFP36

Derivative 6 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	tyrosinase (oculocutaneous albinism IA)	88910875	tttcactttaacataggc[illegible]tcgaaagaaaagtc	88910935	none	11	+ve
EPD	complement factor B	31913531	[illegible]	31913591		6	+ve

Figure 17. List of potential target genes for human ZNF74. The first box is for the consensus binding site, and each box following represents the derivative seqeunces. Target genes for all sequences were derived from two sources: Transfac and EPD. The gene name, binding site location, and associated binders are listed for Transfac derived entries. For EPD-derived entries, the gene names, genomic and mRNA accession links, and 60nt upstream of the promoter are shown. Letters in red denote our binding site. Capitalized letters denote previously observed binding sites. Letters in blue or green denote some observed importance.

ZNF74

Consensus 1

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	platelet derived growth factor C	-1578892584		-1578892524	EGR-1	4	-ve
	platelet derived growth factor C	-1578892584		-1578892524	SP-1	4	-ve
	prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	-1866580036		-1866649976	p50: MelA-p65	1	-ve
	lymphoid enhancer-binding factor 1	-1089090178		-1089090115	Lef1	4	-ve
	lymphoid enhancer-binding factor 1	-1089090178		-1089090116	SP-1	4	-ve
EPD	origin recognition complex, subunit 4-like (S. cerevisiae)	-149778382		-149778322		2	-ve
	advanced glycosylation end product-specific receptor	-32152068		-32152008		6	-ve
	midkine (neurite growth-promoting factor 2)	46439159		46439219		11	+ve

ZNF74

Derivative 1

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	polymeric immunoglobulin receptor	-207116860		-207116920	AP-2	1	-ve
	chemokine (C-C motif) receptor 5	46411306		46411346	C/EBPbeta	3	+ve
EPD	none						

Figure 17. (continued)

ZNF74

Derivative 2 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	v-myc myelocytomatosis viral oncogene homolog (avian)	128748496	[illegible]	128748556	CTCF	8	+ve
	v-myc myelocytomatosis viral oncogene homolog (avian)	128748496	[illegible]	128748556	USF-1	8	+ve
	v-myc myelocytomatosis viral oncogene homolog (avian)	128748496	[illegible]	128748556	PTFG-1	8	+ve
	CD44 molecule (Indian blood group)	35160528	[illegible]	35160588	EGR-1	11	+ve
	CD44 molecule (Indian blood group)	35160528	[illegible]	35160588	SP-1	11	+ve
	epoxide hydrolase 2, cytoplasmic	27348540	[illegible]	27348600	SP-1	8	+ve
EPD	carcinoembryonic antigen-related cell adhesion molecule 5	42212515	[illegible]	42212575		19	+ve
	LIM-domain containing 1	-67505862	[illegible]	-67505802		11	-ve
	hypoxanthine phosphoribosyltransferase 1	133594141	[illegible]	133594201		X	+ve
	S100 calcium binding protein B	-48025061	[illegible]	-48025001		21	-ve

ZNF74

Derivative 3 [illegible]

SAME AS CONSENSUS

ZNF74

Derivative 4 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	sphingosine kinase 1	74350526	[illegible]	74350586	SP-1	17	+ve
EPD	creatine kinase, brain	-103989214	[illegible]	-103989154		14	-ve
	flap structure-specific endonuclease 1	61560334	[illegible]	61560394		11	+ve
	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	9102633	[illegible]	9102693		18	+ve

ZNF74

Derivative 5 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	51434236	[illegible]	51434296	SP3/SP1	1	+ve
	epidermal growth factor receptor	55086393	[illegible]	55086453	VDR	7	+ve
EPD	none						

Figure 18. List of potential target genes for human ZNF136. The first box is for the consensus binding site, and each box following represents the derivative seqeunces. Target genes for all sequences were derived from two sources: Transfac and EPD. The gene name, binding site location, and associated binders are listed for Transfac derived entries. For EPD-derived entries, the gene names, genomic and mRNA accession links, and 60nt upstream of the promoter are shown. Letters in red denote our binding site. Capitalized letters denote previously observed binding sites. Letters in blue or green denote some observed importance.

ZNF136

Consensus

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	v-myb myeloblastosis viral oncogene homolog (avian)	135502366		135502426	M2F1B-C	6	+ve
	synapsin I	-47479564		-47479504	none	X	-ve
	high-mobility group nucleosome binding domain 1	-40721102		-40721042	ZF92.9	21	-ve
	insulin receptor substrate 2	-110439186		-110439126	SP2/SP3	13	-ve
EPD	high-mobility group nucleosome binding domain 1	-40721102		-40721042		21	-ve
	rhodopsin	129247411		129247471		3	+ve

ZNF136

Derivative 1

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	anti-Mullerian hormone receptor, type II	53817346		53817406	SP-1	12	+ve
	coagulation factor XI	187187022		187187082	ENF-4alpha	4	+ve
	thrombopoietin	-184096874		-184096814	XBLB1	3	-ve
	proline dehydrogenase (oxidase) 1	-18924811		-18924751	PPARgamma	22	-ve
EPD	none						

ZNF136

Derivative 2

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	none						

Figure 18. (continued)

ZNF136

Derivative 3							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	fragile X mental retardation, autosomal homolog 2	-7518223	[redacted]	-7518165	NRF-1	17	-ve
	apolipoprotein E	45430411	[redacted]	45430471	AP-2	19	+ve
	interferon regulatory factor 1	-131827386	[redacted]	-131827326	p50: RelA-p65	5	-ve
EPD	apolipoprotein E	45430383	[redacted]	45430043		19	+ve
	transmembrane protein 43	14166313	[redacted]	14166370		3	+ve

ZNF136

Derivative 4							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	ATPase, Ca++ transporting, cardiac muscle, slow twitch 2	110718915	[redacted]	110718975	SPI/SP3	12	+ve
EPD	rhodopsin	129247413	[redacted]	129247470		3	+ve
	citin-cap (calerhonin)	37821534	[redacted]	37821594		17	+ve

ZNF136

Derivative 5							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	cystathionine-beta-synthase	-44496533	[redacted]	-44496673	NRF-1	21	-ve
	mucin 5B, oligomeric mucus/gel-forming	1245671	[redacted]	1245731	SP1	11	+ve
	cystathionine-beta-synthase	-44496533	[redacted]	-44496673	SP1/SP3/SP4	11	-ve
EPD	septin 7	35840781	[redacted]	35840841		7	+ve
	heterogeneous nuclear ribonucleoprotein A	-23670828	[redacted]	-23670768		1	-ve
	profilin 1	-4851892	[redacted]	-4851832		17	-ve

Figure 18. (continued)

ZNF136

Derivative 6	[illegible]	
	SAME AS LAST SEQUENCE	

ZNF136

Derivative 7	[illegible]						
	Official Gene Name	**Start Pos.**	**Binding Site**	**End Pos.**	**Binders**	**Chr.**	**Strand**
Transfac	GLI family zinc finger 1	57853966	cccagcgcccagacagaggtgagaaggggggccaggcgggggacCACCTGggagcagtgg	57854026	USF1/USF2/M-Twist	12	+ve
	collagen, type II, alpha 1	-48398545	ggggcagtgccgcagccccgggttggggcgcaggggccatccctgcgccccgccgat	-48398455	SP1	12	-ve
EPD	none						

Figure 19. List of potential target genes for human ZNF141. The first box is for the consensus binding site, and each box following represents the derivative seqeunces. Target genes for all sequences were derived from two sources: Transfac and EPD. The gene name, binding site location, and associated binders are listed for Transfac derived entries. For EPD-derived entries, the gene names, genomic and mRNA accession links, and 60nt upstream of the promoter are shown. Letters in red denote our binding site. Capitalized letters denote previously observed binding sites. Letters in blue or green denote some observed importance.

ZNF141

Consensus						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	v-myc myelocytomatosis viral oncogene homolog (avian)	128746643		128746723	PUKbeta/PUR-1	8	+ve
	v-myc myelocytomatosis viral oncogene homolog (avian)	128746643		128746703	PUR1/ PUR-factor	8	+ve
	adenosine deaminase	-43274957		-43274927	LEF1 isol	20	-ve
	adenosine deaminase	-43274957		-43274927	c-Myb	20	-ve
	paternally expressed 3	-57351408		-57351348	none	19	-ve
EPD	solute carrier family 31 (copper transporters), member 1	115963818		115963878		9	+ve
	DEAD (Asp-Glu-Ala-Asp) box polypeptide 25	125774342		125774402		11	+ve
	ubiquitin specific peptidase 16	31396301		31396361		21	+ve
	tripeptidyl peptidase I	-6640719		-6640659		11	-ve
	ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)	6414160		6414220		7	+ve
	ribonuclease RZ, subunit A	12917441		12917500		19	+ve

ZNF141

Derivative 1						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	none						

Figure 19. (continued)

ZNF141

Derivative 2 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	alcohol dehydrogenase 1C (class I), gamma polypeptide	-103274070	[illegible]	-103274010	none	4	-ve
EPD	none						

ZNF141

Derivative 3 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	anti-Mullerian hormone	2249207	[illegible]	2249267	SF1	19	+ve
EPD	septin 7	35840772	[illegible]	35840832		7	+ve
	OCIA domain containing 1	48833063	[illegible]	48833123		4	+ve
	MAX-like protein X	40719017	[illegible]	40719077		17	+ve
	complement component 1, q subcomponent binding protein	-5342484	[illegible]	-5342424		17	-ve

ZNF141

Derivative 4 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	topoisomerase (DNA) II alpha 170kDa	-38574295	[illegible]	-38574235	none	17	-ve
EPD	none						

Figure 19. *(continued)*

ZNF141

Derivative 5 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	serine/arginine-rich splicing factor 2	-74733788	[illegible]	-74733728	c-Myb iso1	17	-ve
	ferredoxin 1	110300580	[illegible]	110300640	SP1	11	+ve
	UL16 binding protein 1	150284996	[illegible]	150285056	SP1/SP3	6	+ve
	UL16 binding protein 1	150284996	[illegible]	150285056	AP-2alpha	6	+ve
	protein phosphatase 2, catalytic subunit, alpha isozyme	-133561825	[illegible]	-133561765	CREB	5	-ve
EPD	FK506 binding protein 1A, 12kDa	-1373784	[illegible]	-1373724		20	-ve
	protein phosphatase 2, catalytic subunit, alpha isozyme	-133561825	[illegible]	-133561765		5	-ve
	RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)	40987313	[illegible]	40987373		15	+ve

ZNF141

Derivative 6 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	fibroblast growth factor 2 (basic)	123747657	[illegible]	123747717	EGR-1	4	+ve
	fibroblast growth factor 2 (basic)	123747657	[illegible]	123747717	SP1	4	+ve
	apolipoprotein E	45409201	[illegible]	45409261	none	19	+ve
	vitamin D (1,25- dihydroxyvitamin D3) receptor	-48299207	[illegible]	-48299147	ICER-IIgamma	12	-ve
EPD	enolase 3 (beta, muscle)	4854376	[illegible]	4854436		17	+ve
	lysophosphatidylcholine acyltransferase 3	-7125833	[illegible]	-7125773		12	-ve
	heat shock protein 90kDa beta (Grp94), member 1	104324127	[illegible]	104324187		12	+ve

ZNF141

Derivative 7 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	solute carrier family 31 (copper transporters), member 1	115983816	[illegible]	115983876		9	+ve
	DEAD (Asp-Glu-Ala-Asp) box polypeptide 25	125774341	[illegible]	125774401		11	+ve
	transmembrane emp24-like trafficking protein 10 (yeast)	-75643336	[illegible]	-75643276		14	-ve
	transcription factor 12	57210525	[illegible]	57210585		15	+ve
	tetraspanin 6	-99891744	[illegible]	-99891654		X	-ve

Figure 20. List of potential target genes for human ZNF480. The first box is for the consensus binding site, and each box following represents the derivative seqeunces. Target genes for all sequences were derived from two sources: Transfac and EPD. The gene name, binding site location, and associated binders are listed for Transfac derived entries. For EPD-derived entries, the gene names, genomic and mRNA accession links, and 60nt upstream of the promoter are shown. Letters in red denote our binding site. Capitalized letters denote previously observed binding sites. Letters in blue or green denote some observed importance.

ZNF480

Consensus							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	LZF transcription factor 4	-11606337		-11606277	NRF-1	2	-ve
	coagulation factor III (thromboplastin, tissue factor)	-95007406		-95007346	EGR-1/SP1	1	-ve
	solute carrier family 9 (sodium/hydrogen exchanger), member 3	-524660		-524630	SP3/SP1	5	-ve
	solute carrier family 9 (sodium/hydrogen exchanger), member 3	-524660		-524630	EGR-1	5	-ve
	transforming growth factor, alpha	-70781110		-70781050	AP-2	2	-ve
	prostaglandin I2 synthase 21kDa (brain)	139871609		139871669	HNF-1	9	+ve
	linker for activation of T cells	28996374		28996434	SP1/SP3	16	+ve
	cholesteryl ester transfer protein, plasma	56993468		56993528	none	16	+ve
EPD	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	-27961637		-27961637		1	-ve
	beta-2-microglobulin	45003681		45003741		15	+ve
	protease, serine, 6	-31147121		-31147061		16	-ve
	proteasome (prosome, macropain) 26S subunit, non-ATPase, 1	231921529		231921589		2	+ve
	sorting nexin 3	-108582233		-108582173		6	-ve
	insulin-like growth factor binding protein 6	53491414		53491474		12	+ve
	hypoxanthine phosphoribosyltransferase 1	133594137		133594197		X	+ve

ZNF480

Derivative 1							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	damage-specific DNA binding protein 2, 48kDa	47236511		47236571	p53	11	+ve
	solute carrier family 9 (sodium/hydrogen exchanger), member 2	103235959		103236019	SP3/SP1	2	+ve
EPD	peptidylprolyl isomerase F	81197156		81197216		10	+ve
	SP100 nuclear antigen	231280897		231280957		2	+ve
	CX11-like MARVEL transmembrane domain containing 2	32433326		32433386		3	+ve
	guanine nucleotide binding protein (G protein), beta polypeptide 2	6959069		6959129		12	+ve
	S100 calcium binding protein B	-48025965		-48025905		21	-ve

Figure 20. (continued)

ZNF480

Derivative 2	[illegible]					
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	none						

ZNF480

Derivative 3	[illegible]						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	-534789	[illegible]	-534729	p53 1so1	11	-ve
	ornithine decarboxylase 1	-10598199	[illegible]	-10598139	AP2alpha/AP2 /Max 1sc2/c-	2	-ve
	anti-Mullerian hormone	2249189	[illegible]	2249249	SP1	19	+ve
EPD	mannan-binding lectin serine peptidase 1 (C4/C2 activating component of Ra-reactive factor)	-187009321	[illegible]	-187009441		3	-ve
	stanniocalcin 2	-172755453	[illegible]	-172755393		5	-ve
	electron-transfer-flavoprotein, alpha polypeptide	-76603805	[illegible]	-76603745		15	-ve
	surfeit 4	-136242964	[illegible]	-136242904		9	-ve

ZNF480

Derivative 4	[illegible]					
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	none						
EPD	none						

Figure 20. (continued)

ZNF480

Derivative 5 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	cystathionine-beta-synthase	-44496277	[illegible]	-44496217	MZF-1	21	-ve
	peroxisome proliferator-activated receptor gamma	12329218	[illegible]	12329278	EGR-1	3	+ve
	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	-21994403	[illegible]	-21994343	PAX2/PAX3	9	-ve
	DnaJ (Hsp40) homolog, subfamily B, member 1	-14629355	[illegible]	-14629295	NF-Y	19	-ve
	presenilin 1	73603247	[illegible]	73603307	EKM	14	+ve
	triosephosphate isomerase 1	6976539	[illegible]	6976579	SP1	12	+ve
	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	36646370	[illegible]	36646430	GKLF	6	+ve
	toll-like receptor adaptor molecule 1	-4831796	[illegible]	-4831736	SP1	19	-ve
	cystathionine-beta-synthase	-44496277	[illegible]	-44496217	SP3/SP1	21	-ve
	heat shock 70kDa protein 1A	31784621	[illegible]	31784701	SP1	6	+ve
	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	36646370	[illegible]	36646430	SP3/SP1	6	+ve
	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	51434302	[illegible]	51434362	E2F-1:DP-1	1	+ve
	cyclin-dependent kinase inhibitor 2C (p18, inhibits CDK4)	51434302	[illegible]	51434362	SP3/SP1	1	+ve
	pygopus homolog 2 (Drosophila)	-154934260	[illegible]	-154934200	Elf-1/Elk-1	1	-ve
	cAMP responsive element binding protein 1	208394623	[illegible]	208394683	SP1	2	+ve
	insulin receptor substrate 1	-227664143	[illegible]	-227664083	none	2	-ve
	superoxide dismutase 2, mitochondrial	-160114385	[illegible]	-160114325	SP3/SP1	6	-ve
	survival of motor neuron 1, telomeric	70220839	[illegible]	70220899	Ets/PEA3	5	+ve
	survival of motor neuron 1, telomeric	70220839	[illegible]	70220899	IL-6	5	+ve
EPD	plasminogen activator, urokinase	75670812	[illegible]	75670872		10	+ve
	transforming growth factor, beta-induced, 68kDa	135364634	[illegible]	135364694		5	+ve
	matrix metallopeptidase 11 (stromelysin 3)	24114984	[illegible]	24115044		22	+ve
	cysteinyl-tRNA synthetase	-3078691	[illegible]	-3078631		11	-ve
	DiGeorge syndrome critical region gene 2	-19109988	[illegible]	-19109928		22	-ve
	ornithine decarboxylase antizyme 1	2269473	[illegible]	2269533		19	+ve
	phospholipase A2, group XV	68279207	[illegible]	68279267		16	+ve
	acidic (leucine-rich) nuclear phosphoprotein 32 family, member B	100745558	[illegible]	100745618		9	+ve
	enolase 1, (alpha)	-8938785	[illegible]	-8938725		1	-ve
	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	-27961711	[illegible]	-27961651		1	-ve
	calponin 3, acidic	-95392709	[illegible]	-95392649		1	-ve
	perilipin 2	-19127562	[illegible]	-19127502		9	-ve
	superoxide dismutase 2, mitochondrial	-160114318	[illegible]	-160114258		6	-ve
	RAB7A, member RAS oncogene family	128444963	[illegible]	128445023		3	+ve
	cytidine monophosphate N-acetylneuraminic acid synthetase	22199136	[illegible]	22199196		12	+ve
	insulin-like growth factor binding protein 7	-57976618	[illegible]	-57976558		4	-ve
	serine/threonine kinase 25	-242447755	[illegible]	-242447695		2	-ve
	calreticulin	13049382	[illegible]	13049442		19	+ve
	delta-like 1 homolog (Drosophila)	101193180	[illegible]	101193240		14	+ve
	moesin	64887476	[illegible]	64887536		X	+ve
	insulin-like growth factor binding protein 6	53491335	[illegible]	53491395		12	+ve
	amyloid beta (A4) precursor-like protein 2	129939741	[illegible]	129939801		11	+ve
	heat shock 70kDa protein 9 (mortalin)	-137911199	[illegible]	-137911139		5	-ve
	cysteine-rich protein 2	105941113	[illegible]	105941173		14	+ve
	RNA binding protein S1, serine-rich domain	-2318079	[illegible]	-2318019		16	-ve
	guanine nucleotide binding protein (G protein), gamma 11	93551319	[illegible]	93551379		7	+ve
	phosphofructokinase, platelet	3109672	[illegible]	3109732		10	+ve
	ubiquitin-like 4A	-153715018	[illegible]	-153714958		X	-ve

Figure 20. (continued)

ZNF480

Derivative 6 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	mitogen-activated protein kinase kinase kinase 5	-137113706	[illegible]	-137113646	E2F1/2/3/4	6	-ve
	podoplanin	13910197	[illegible]	13910257	SP3/SP1	1	+ve
	thrombomodulin	-23030480	[illegible]	-23030420	SP1	20	-ve
EPD	nuclear receptor binding protein 1	27651450	[illegible]	27651510		2	+ve
	RAD51 homolog C (S. cerevisiae)	56769929	[illegible]	56769989		17	+ve
	caspase 6, apoptosis-related cysteine peptidase	-110624654	[illegible]	-110624594		4	-ve
	methionine adenosyltransferase II, beta	162932582	[illegible]	162932642		5	+ve

ZNF480

Derivative 7 [illegible]

	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)	-21994502	[illegible]	-21994442	SP1	9	-ve
EPD	carbonyl reductase 1	37442275	[illegible]	37442335		21	+ve
	guanine nucleotide binding protein (G protein), alpha inhibiting activity polypeptide 2	50273578	[illegible]	50273638		3	+ve
	proteasome (prosome, macropain) subunit, alpha type, 7	-60718456	[illegible]	-60718396		20	-ve
	X-box binding protein 1	-29196564	[illegible]	-29196504		22	-ve
	StAR-related lipid transfer (START) domain containing 7	-96874612	[illegible]	-96874552		2	-ve
	ets variant 5	-185826873	[illegible]	-185826813		3	-ve
	ubiquitin-like 4A	-153714988	[illegible]	-153714928		X	-ve

70

Figure 20. (continued)

ZNF480

Derivative 8							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	55523621		55523885	c-bID-2	4	+ve
	glycophorin C (Gerbich blood group)	127413630		127413650	none	2	+ve
	erythropoietin	100321394		100321454	none	7	+ve
	gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)	-25282679		-25282619	POU1F2	8	-ve
	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 1	-128657605		-128657545	SP1	X	-ve
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	120049929		120049989	alpha-enolase	1	+ve
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	119957868		119957928	alpha-enolase	1	+ve
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	119957868		119957928	YY1	1	+ve
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1	120049929		120049989	SP3/SP1	1	+ve
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	119957868		119957928	SP3/SP1	1	+ve
	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 5	95047677		95047737	SP1	14	+ve
	baculoviral IAP repeat-containing 5	76210296		76210236	BTEB2	17	+ve
	suppressor of cytokine signaling 1	-11353743		-11353782	STAT6	16	-ve
	suppressor of cytokine signaling 1	-11353743		-11353783	c-ETS-1/2	16	-ve
EPD	calcium binding atopy-related autoantigen 1	-74385836		-74385826		10	-ve
	interferon, alpha-inducible protein 27	94577013		94577073		14	+ve
	minichromosome maintenance complex component 3	-52149562		-52149532		6	-ve
	AAA1, activator of heat shock 90kDa protein ATPase homolog 1 (yeast)	77924364		77924424		14	+ve
	neural precursor cell expressed, developmentally down-regulated 8	-24701629		-24701569		14	-ve
	Fas (TNFRSF6)-associated via death domain	70049423		70049483		11	+ve

ZNF480

Derivative 9							
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	surfactant protein B	-85855453		-85855393	TTF-1	2	-ve
	surfactant protein B	-85855453		-85855393	TTF-1	2	-ve
	LIM homeobox 3	-139096361		-139096301	NF-1/ CTF	9	-ve
	prepronociceptin	28174704		28174764	SP1	8	+ve
	protease, serine, 56	-46759549		-46759489	SP1	3	-ve
EPD	secreted phosphoprotein 1	88896920		88896980		4	+ve
	interferon induced transmembrane protein 2 (1-8D)	308117		308177		11	+ve

APPENDIX

Appendix A- Protein refolding methods and results

METHODS AND MATERIALS

Solubilization of Inclusion Bodies and Ion-Exchange Chromatography

1L of each individual culture was used for this method. The cell pellets were resuspended in ddH_20 (1g/mL) containing sodium deoxycholate (NaDOC). The cell suspensions were sonicated for three cycles and then centrifuged for 15 min at 14K xg in $4^{O}C$. The supernatant was discarded and the pellets were resuspended in ddH_2O at 1g/mL. Three 100mL aliquots were taken from each sample and spun 14K for 15 min. The pellets were resuspended in 100µL 0.1M Tris-Cl (pH 8.5) containing different concentrations of urea (2M, 5M, and 8M). The aliquots were centrifuged at 14K xg for 15 min and both the supernatant and pellet were analyzed on a SDS-PAGE gel to determine the level of inclusion body solubilization. The analysis indicated that a mild 2M urea solution with a pH of 12.5 must be used to completely solubilize all the inclusion bodies (see results). Hence, the rest of the original cell suspensions were solubilized using 2M urea. Since the starting pellet weight was different for each sample, the final resuspended volumes were matched with inclusion body solubilization buffer 1 and rotated at $4^{O}C$ for 120 min. Protease inhibitor PMSF, and DTT

were added to a final concentration of 1mM and 10mM, respectively. The samples were centrifuged at 12K xg for 15 min and the supernatant was subjected to DEAE ion-exchange column chromatography. The columns were packed with 2mL of DEAE-sepharose beads and equilibrated with 5mL 1X PBS. Elution buffers were made with 8M urea, 0.1M Tris-Cl (pH 8.5), and containing different concentrations of NaCl (0mM, 50mM, 100mM, 200mM, 300mM, 400mM, and 500mM). The samples were loaded onto the columns, setting aside 100µL of each sample for gel analysis. The columns were washed with 4mL of 0mM NaCl elution buffer. Bound proteins were eluted with 1.7mL of each elution buffer twice, in a stepwise manner with increasing NaCl concentration (starting with 50mM and ending with 500mM). The elutions were analyzed on a SDS-PAGE gel along with the load, wash, and flow through fractions.

In Vitro Refolding of Inclusion Bodies

The solubilized inclusion body samples were adjusted to a protein concentration of 250µg/mL in a total volume of 5mL. The samples were introduced to 50mL of refolding buffer using a drop-wise rapid dilution method (Novinec et al., 2008). The refolding buffer consisted of 50mM Tris-Cl pH 8.0, 1M urea, 150mM NaCl, 1mM GSSG, 3mM GSH, and 50uM $ZnSO_4$. The rapid dilution was performed under constant stirring at $4^{O}C$. Once the samples were

completely added, the entire refolded content was allowed to stir at 4°C for 2 days.

Purification of Fusion Proteins Refolded In Vitro

The refolded solution was loaded onto a GSH-Sepharose column. The columns were packed with 200µL of GSH-Sepharose beads and washed with 10mL of 1x PBS. The samples were loaded onto the columns and the flow through was collected. The columns were then washed with 15mL of 1x PBS and the last 500µL of wash was collected. The columns were then eluted three times with 500µL of glutathione-containing elution buffer. The elution buffer was allowed to sit for 10 min each time and the entire 500µL was collected. The purification procedure was repeated a second time with the flow through fractions for a total of six elutions per sample. Each flow through and elution was run on a 12% SDS-PAGE gel to determine if the fusion protein did in fact correctly fold and bind to the GSH-Sepharose column.

RESULTS AND DISCUSSION

Solubilization of Target Fusion Protein Inclusion Bodies

After many unsuccessful attempts of purification, *in vitro* refolding was assessed. The theory was that these proteins are expressed in misfolded form, aggregated, and sequestered as inclusion bodies. Urea was used as a solubilizing agent. Each inclusion body sample was assayed with buffer containing different concentrations of urea (2M, 5M, and 8M). After incubation

with the buffer, each sample was centrifuged, and the supernatant and pellet were run on a PAGE gel to determine if the target protein did indeed solubilize into the supernatant or if the target protein was still insoluble in the form of inclusion bodies. Only ZFP36 was found solubilized in the supernatant, whereas all other samples were still insoluble in the pellet (**Figure 8**). Since the success rate was not very high, a milder solubilization protocol was attempted, which utilized 2M urea in a buffer with a pH of 12.5. The samples were then subjected to DEAE chromatography and then went directly for *in vitro* refolding.

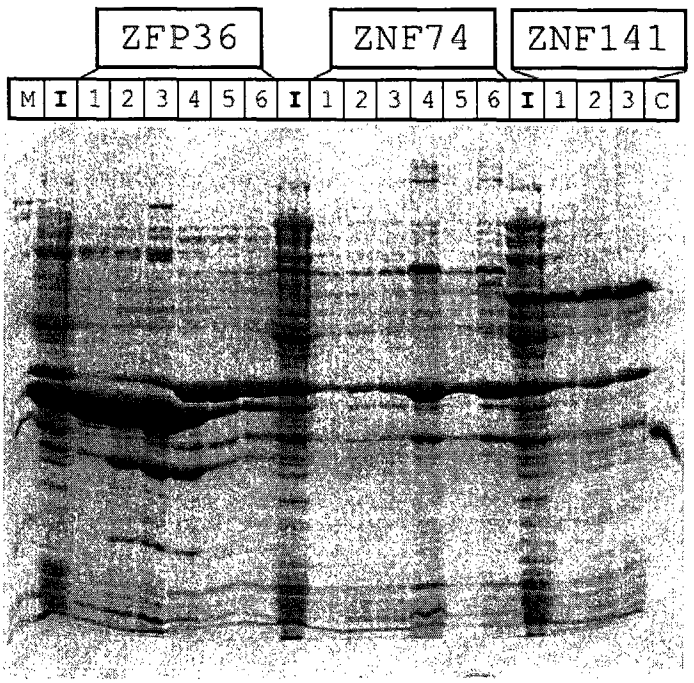
Refolding/Purification of Target Proteins Using DEAE Chromatography

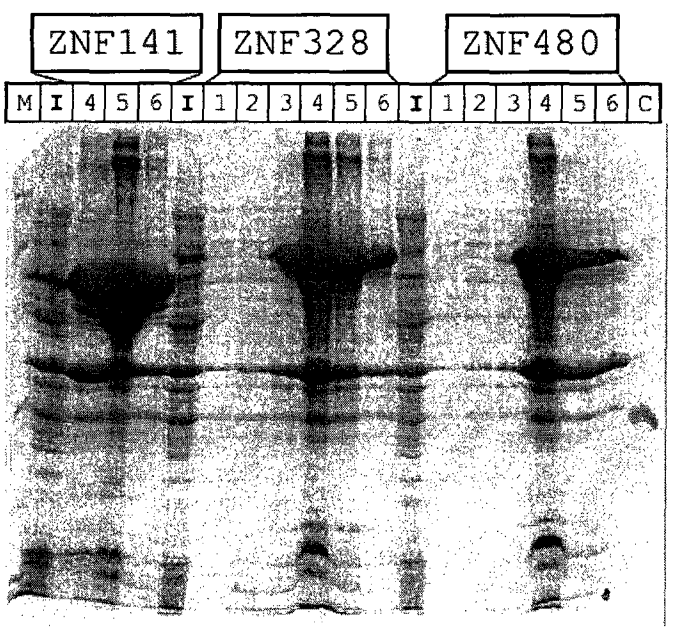
Purification and refolding was also attempted using DEAE ion-exchange chromatography. It has been reported that ion-exchange chromatography can refold proteins without the presence of any other factors (Chen et al., 2009). **Figure 9** shows the SDS-PAGE gel analysis, which indicates that all members except ZFP36 were found only in the "flow-through" and "wash" lanes. ZFP36 was the only sample that bound to the column and was eluted with 50mM NaCl. While this procedure did not help in the refolding of the target fusion proteins, it did successfully separate the target fusion proteins from many of E.coli-derived proteins, which can be seen in the elution lanes.

GSH Affinity Purification of In Vitro-Refolded Target Proteins

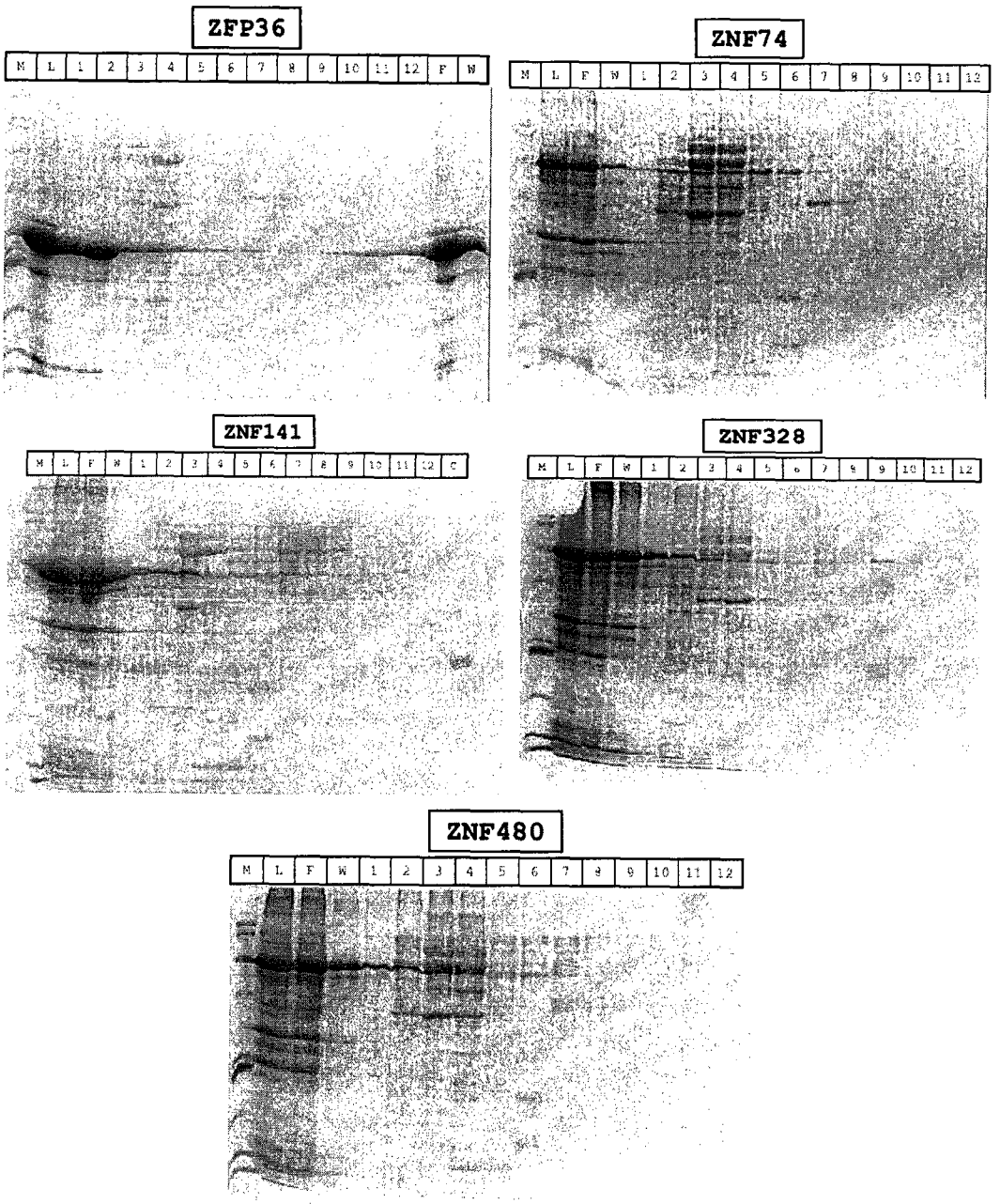
After incubation with 2M urea buffer (pH 12.5) and DEAE chromatography, the samples were immediately introduced to the refolding buffer. SDS-PAGE analysis of the GSH-Sepharose affinity purification of the samples after refolding showed that none of the samples were successful in refolding *in vitro* (**Figure 10**). The two bands that are present in all the elutions are thought to be an E.coli-derived glutathione transferase. The target fusion protein for ZFP36 co-migrates with the E.coli-derived protein during PAGE, and hence, it is difficult to determine if there is refolded ZFP36 present. However, it would be likely, given its small size and behavior in previous purification attempts.

Appendix B- Inclusion Body Solubilization Assay. Each inclusion body sample was solubilized in 2, 5, and 8M Urea. The supernatant (S) and pellet (P) fractions from each sample were analyzed to determine the level of solubilization. 1-2M(S), 2-5M(S), 3-8M(S), 4-2M(P), 5-5M(P), 6-8M(P).

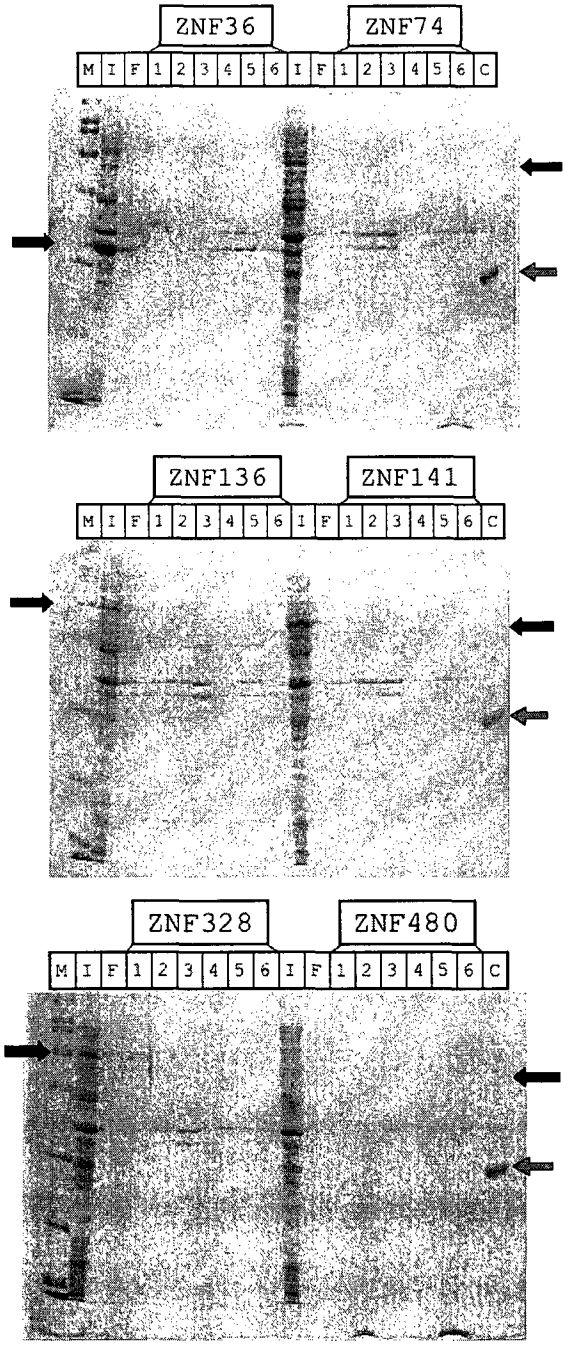




Appendix C- Refolding and Purification by DEAE Chromatography. Each sample was loaded onto an ion-exchange column and sequentially eluted with 50, 100, 200, 300, 400, and 500mM NaCl. M-Marker, L-Loaded Sample, F-Flowthrough, W-Wash, Elution samples in lanes 1&2-50mM; 3&4-100mM; 5&6-200mM; 7&8-300mM; 9&10-400mM; 11&12-500mM.



Appendix D- GSH Purification After Refolding. Each member was refolded from inclusion bodies and analyzed on SDS-PAGE gels. None were refolded correctly. The visible bands in the elution lanes are thought to be E.coli-derived GST and/or fusion proteins that were cleaved during refolding/eluting. I-Induced Sample, F-Flowthrough, Lanes marked 1 through 6 are Elutions 1 through 6 respectively. 'C' denotes GST control. Black arrows - Fusion proteins; Red arrows - GST pure protein



REFERENCES

Ayyanathan K, Lechner MS, Bell P, Maul GG, Schultz DC, Yamada Y, Tanaka K, Torigoe K, Rauscher III FJ (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. Genes and Development 17:1855-69

Cao L, Wang Z, Zhu C, Zhao Y, Yuan W, Li J, Wang Y, Ying Z, Li Y, Yu W, Wu X, Liu M. (2005) ZNF383, a novel KRAB-containing zinc finger protein, suppresses MAPK signaling pathway. Biochem Biophys Res Commun. 333:1050-9.

Chen Y, Leong SS. (2009) Adsorptive refolding of a highly disulfide-bonded inclusion body protein using anion-exchange chromatography. J Chromatogr A. 1216:4877-86.

Chun JN, Song IS, Kang DH, Song HJ, Kim HI, Seo J, Lee KJ, Kim J, Kang SW. (2008) A splice variant of the C(2)H(2)-type zinc finger protein, ZNF268s, regulates NF-kappaB activation by TNF-alpha. Mol Cells. 26:175-80.

Derry JM, Jess U, Francke U. (1995) Cloning and characterization of a novel *zinc finger gene in* Xp11.2. *Genomics*. 30:361-5.

Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ 3rd (1996). KAP-1, a novel corepressor for the highly conserved KRAB repression domain. Genes and Development 15: 2067-78.

González-Lamuño D, Loukili N, García-Fuentes M, Thomson TM. (2002) Expression and regulation of the transcriptional repressor ZNF43 in Ewing sarcoma cells. Pediatr Pathol Mol Med. 21:531-40.

Huang C, Jia Y, Yang S, Chen B, Sun H, Shen F, Wang Y. (2007) Characterization of ZNF23, a KRAB-containing protein that is downregulated in human cancers and inhibits cell cycle progression. Exp Cell Res. 313:254-63.

Huang C, Yang S, Ge R, Sun H, Shen F, Wang Y. (2008) ZNF23 induces apoptosis in human ovarian cancer cells. Cancer Lett. 266:135-43.

Huang X, Yuan W, Huang W, Bai Y, Deng Y, Zhu C, Liang P, Li Y, Du X, Liu M, Wang Y, Wu X. (2006) ZNF569, a novel KRAB-containing zinc finger protein, suppresses MAPK signaling pathway. Biochem Biophys Res Commun. 346:621-8.

Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, Gordon L, Branscomb E, Stubbs L (2006) A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. Genome Res. Genome Res 16:669-77.

Langmann T, Schumacher C, Morham SG, Honer C, Heimerl S, Moehle C, Schmitz G. (2003) ZNF202 is inversely regulated with its target genes ABCA1 and apoE during macrophage differentiation and foam cell formation. J Lipid Res. 44:968-77.

Li J, Wang Y, Fan X, Mo X, Wang Z, Li Y, Yin Z, Deng Y, Luo N, Zhu C, Liu M, Ma Q, Ocorr K, Yuan W, Wu X. (2007) ZNF307, a novel zinc finger gene suppresses p53 and p21 pathway. Biochem Biophys Res Commun. 363:895-900.

Li Y, Yang D, Bai Y, Mo X, Huang W, Yuan W, Yin Z, Deng Y, Murashko O, Wang Y, Fan X, Zhu C, Ocorr K, Bodmer R, Wu X. (2008) ZNF418, a novel human KRAB/C2H2 zinc finger protein, suppresses MAPK signaling pathway. Mol Cell Biochem 310:141-51.

Luo K, Yuan J, Shan Y, Li J, Xu M, Cui Y, Tang W, Wan B, Zhang N, Wu Y, Yu L. (2006) Activation of transcriptional activities of AP1 and SRE by a novel zinc finger protein ZNF445. Gene. 367:89-100.

Novinec M, Kovacic L, Skrlj N, Turk V, Lenarcic B. (2008) Recombinant human SMOCs produced by in vitro refolding: calcium-binding properties and interactions with serum proteins. Protein Expr Purif. 62:75-82.

Ou Y, Wang S, Cai Z, Wang Y, Wang C, Li Y, Li F, Yuan W, Liu B, Wu X, Liu M. (2005) ZNF328, a novel human zinc-finger protein, suppresses transcriptional activities of SRE and AP-1. Biochem Biophys Res Commun 333:1034-44.

Qi X, Li Y, Xiao J, Yuan W, Yan Y, Wang Y, Liang S, Zhu C, Chen Y, Liu M, Wu X. (2006) Activation of transcriptional activities of AP-1 and SRE by a new zinc-finger protein ZNF641. Biochem Biophys Res Commun. 339:1155-64.

Ravassard P, Côté F, Grondin B, Bazinet M, Mallet J, Aubry M. (1999) ZNF74, a gene deleted in DiGeorge syndrome, is expressed in human neural crest-derived tissues and foregut endoderm epithelia. Genomics. 62:82-5.

Shao H, Zhu C, Zhao Z, Guo M, Qiu H, Liu H, Wang D, Xue L, Gao L, Sun C, Li W. (2006) KRAB-containing zinc finger gene ZNF268 encodes multiple alternatively spliced isoforms that contain transcription regulatory domains. Int J Mol Med. 18:457-63.

Stene MC, Frikke-Schmidt R, Nordestgaard BG, Grande P, Schnohr P, Tybjaerg-Hansen A. (2008) Functional promoter variant in zinc finger protein 202 predicts severe atherosclerosis and ischemic heart disease. J Am Coll Cardiol. 52:369-77.

Sun L, Gu S, Li N, Zheng D, Sun Y, Li D, Ji C, Ying K, Xie Y, Mao Y. (2005) A novel zinc finger gene ZNF468 with two co-expressional splice variants, ZNF468.1 and ZNF468.2. Biochem Genet 43:271-86.

Takase K, Ohtsuki T, Migita O, Toru M, Inada T, Yamakawa-Kobayashi K, Arinami T. (2001) Association of ZNF74 gene genotypes with age-at-onset of schizophrenia. Schizophr Res. 52:161-5.

Wagner S, Hess MA, Ormonde-Hanson P, Malandro J, Hu H, Chen M, Kehrer R, Frodsham M, Schumacher C, Beluch M, Honer C, Skolnick M, Ballinger D, Bowen BR. (2000) A broad role for the zinc finger protein ZNF202 in human lipid metabolism. J Biol Chem. 275:15685-90.

Wang D, Shi L. (2009) High-level expression, purification, and in vitro refolding of soluble tumor necrosis factor-related apoptosis-inducing ligand (TRAIL). Appl Biochem Biotechnol. 157:1-9.

Wen Q, Ma L, Luo W, Zhou MQ, Wang XN. (2008) Expression, purification, and refolding of recombinant fusion protein hIL-2/mGM-CSF. Biomed Environ Sci. 21:509-13.

Yi Z, Li Y, Ma W, Li D, Zhu C, Luo J, Wang Y, Huang X, Yuan W, Liu M, Wu X. (2004) A novel KRAB zinc-finger protein, ZNF480, expresses in human heart and activates transcriptional activities of AP-1 and SRE. Biochem Biophys Res Commun. 320:409-15.