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

UMI Number: 1485665

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 1485665
Copyright 2010 by ProQuest LLC.
All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC 789 East Eisenhower Parkway P.O. Box 1346 Ann Arbor, MI 48106-1346

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 Aylyanathan, 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

Date / 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	vii
Introduction	1
Materials and Methods	5
Constructing the Bioinformatics-Based KRAB-ZFP Database	5
Making GST-ZnF Fusion Proteins	6
Purification of Fusion Proteins by GSH Affinity Chromatography	7
Dialysis To Concentrate Usable Protein	7
Cold Binding To Enrich Oligonucleotide Library	8
Hot Binding To Select Specific Binding-Site Sequences	11
Cloning the Selected Sequences	13
Deriving Consensus Binding-Site Sequences	15
Testing Selected Binding-Sites	16
Deriving Potential Target Genes	16
Results and Discussion	18
Initial Purification of GST-Tagged Target Fusion Proteins	19
Final Purification of GST-Tagged Target Fusion Proteins	19
Concentrations of Usable Protein After Dialysis	20
Cold-Binding to Pre-Enrich The Randomized Oligomer Library	20
Hot-Binding (EMSA) to Select Specific Oligomer Sequences	21
Cloning of Selected Oligomers Into pUC18 Vectors	22
Deriving Consensus Binding-Site Sequences	22
Confirming Derived Consensus Binding Activity	23
Deriving Potential Target Genes	23

Appendix	72
Appendix A Protein refolding methods and results	72
Appendix B	77
Appendix C	78
Appendix D	79
References	80

LIST OF FIGURES

Figure 1- Individual family evolution	25
Figure 2- Evolution across all families	41
Figure 3- Individual linker evolution	42
Figure 4- Diagrammatic representation of the KRAB-ZFP Subset	43
Figure 5- Diagrammatic representaion of the GST-ZFP recombinants	44
Figure 6- Expression profiles for each member	45
Figure 7- First GSH-affinity purifications	46
Figure 8- First successful GSH-affinity purifications.	47
Figure 9- Last three rounds of GSH-affnity purifications	48
Figure 10- Cold-binding PCR gels	49
Figure 11- EMSA radiographs	50
Figure 12- Hot-binding PCR gels	51
Figure 13- Plasmid digestion to check for presence of insert	52
Figure 14- Plasmid integrity assay	55
Figure 15- Consensus maps for ZFP36, ZNF74, ZNF136,ZNF141, ZNF480 .	56
Figure 16- Potential target genes for ZFP36.	57
Figure 17- Potential target genes for ZNF74.	59
Figure 18- Potential target genes for ZNF136	61
Figure 19- Potential target genes for ZNF141	64
Figure 20- Potential target genes for ZNF480	67

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 (Kruppel-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), 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 fulllength 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 A595 ~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 ZnSO₄, 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 ddH₂0; 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 ddH₂O; 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°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°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^oC 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°C. A 50µL PCR reaction (2µL DNA, 5ul DMSO, 25µL 2X Master Mix, 1ul 3' primer, 1µL 5' primer, and 16ul of ddH₂0 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 ddH₂O; 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 120uL 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°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 ³²P, and 13.75µL ddH₂O and then it was divided into 6 tubes. The 20µL reactions were incubated at 37°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 ³²P from the freshly phosphorylated oligomers. Once the kinasing reactions were finished, 10µL ddH₂O 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°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 ddH₂O; 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 BamH1 and EcoR1 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 ddH₂O. Next, 5μL of undigested pUC18 vector was added with 5μL 10X EcoBuffer, 5µL 10X BSA, 1µL EcoR1 enzyme, 1µL BamH1 enzyme, and 33µL ddH₂O. Both digestion reactions were incubated at 37^oC for 2 hrs. The oligomer digestions were then topped off with 50µL ddH₂O and cleaned using two rounds of PCI cleanup, one round of CI cleanup, and EtOH precipitation before a final resuspension in 10µL ddH₂O. 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°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°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°C overnight and then stored at 4°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°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 ddH₂O. 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 ddH₂O, 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 ddH₂O per sample) before being subjected to PCI/PCI/CI/EtOH cleanup and resuspension in 20µL ddH₂O. 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 ³²P, 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

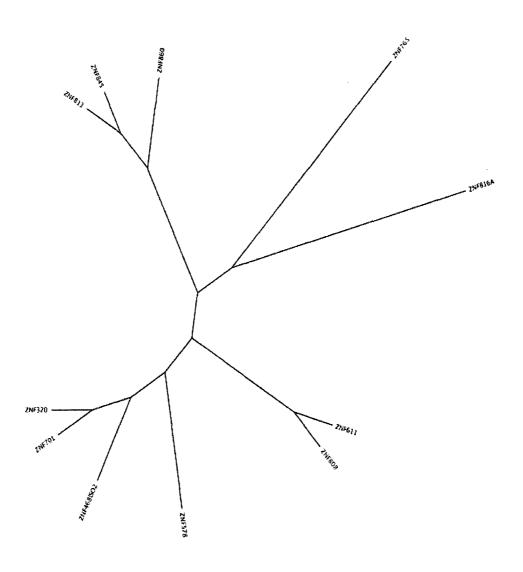


Figure 1. (continued)

Family 2

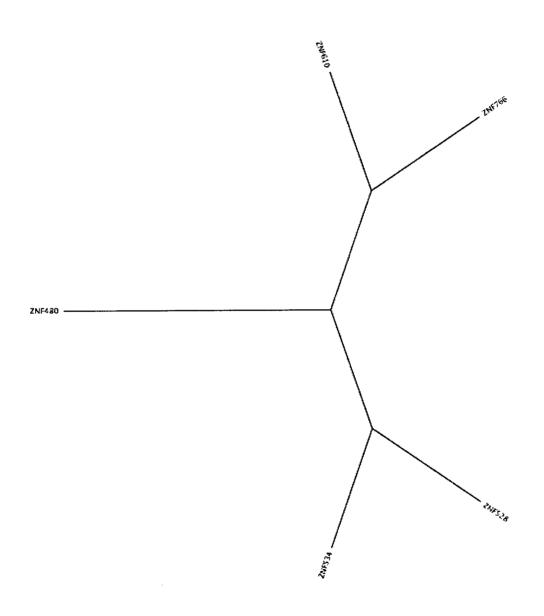


Figure 1. (continued)

Family 3

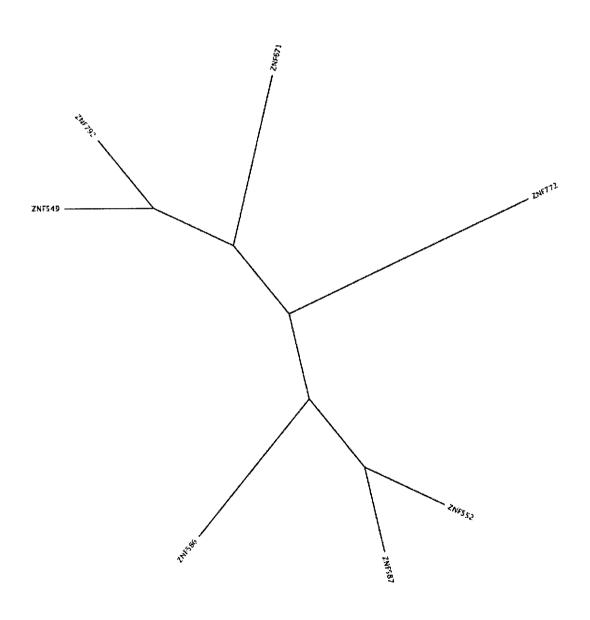


Figure 1. (continued)

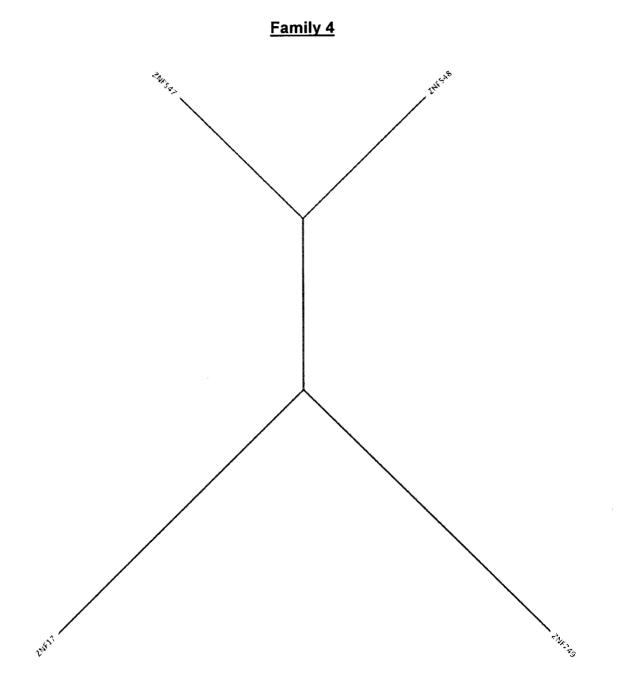


Figure 1. (continued)

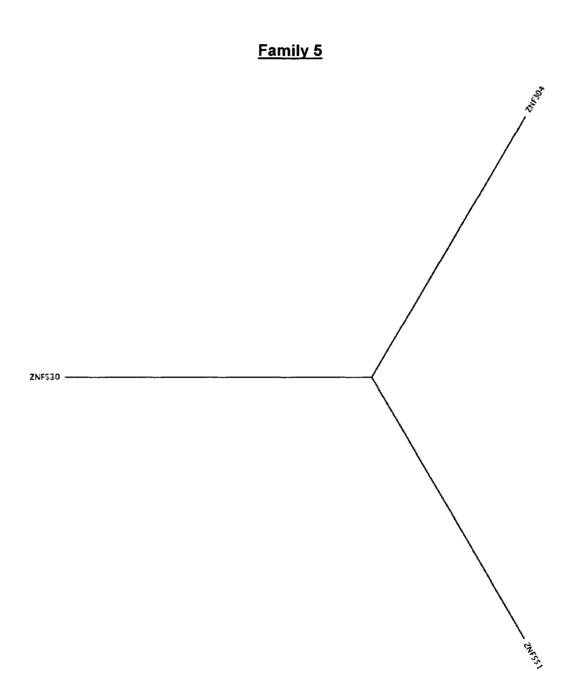


Figure 1. (continued)

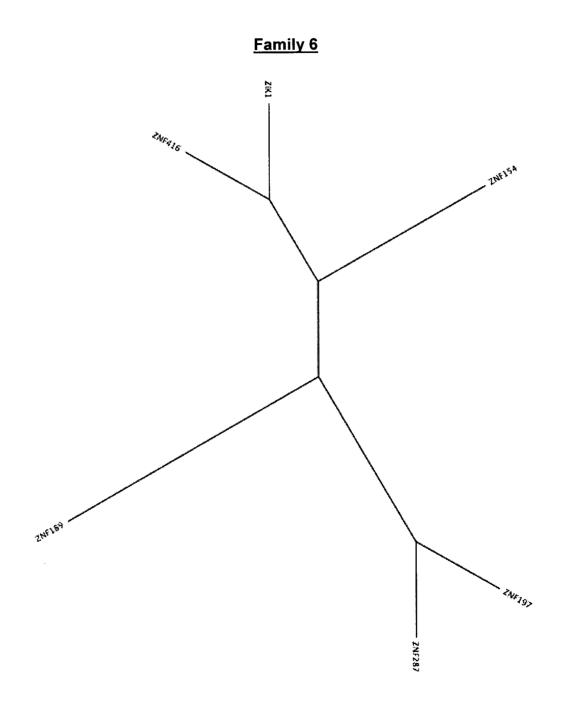


Figure 1. (continued)

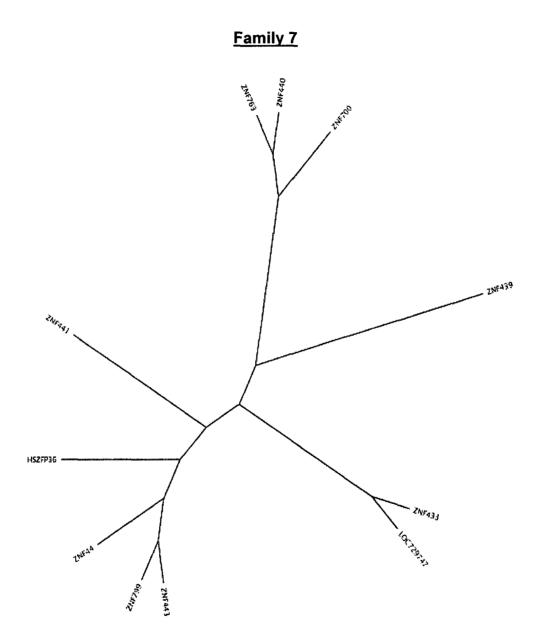


Figure 1. (continued)

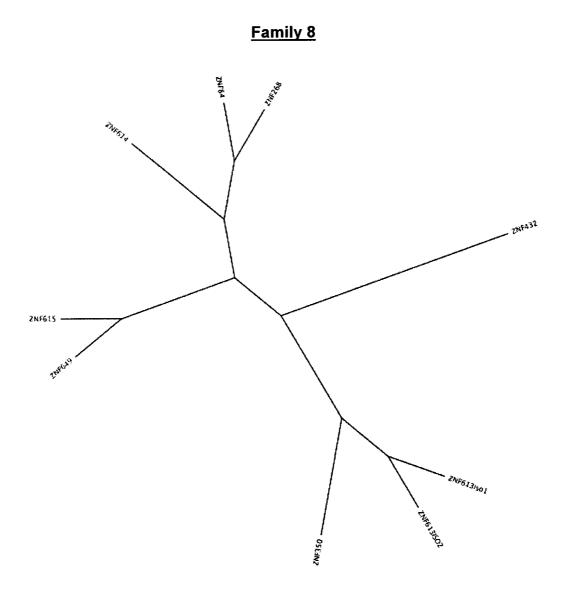


Figure1. (continued)

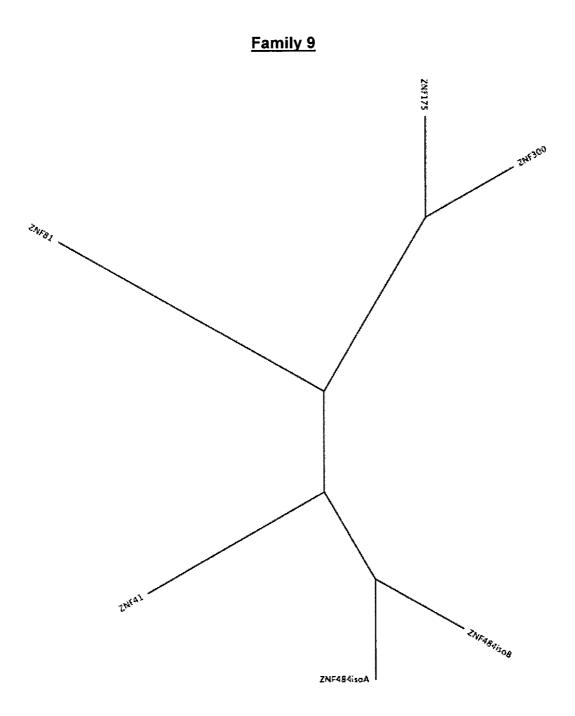


Figure 1. (continued)

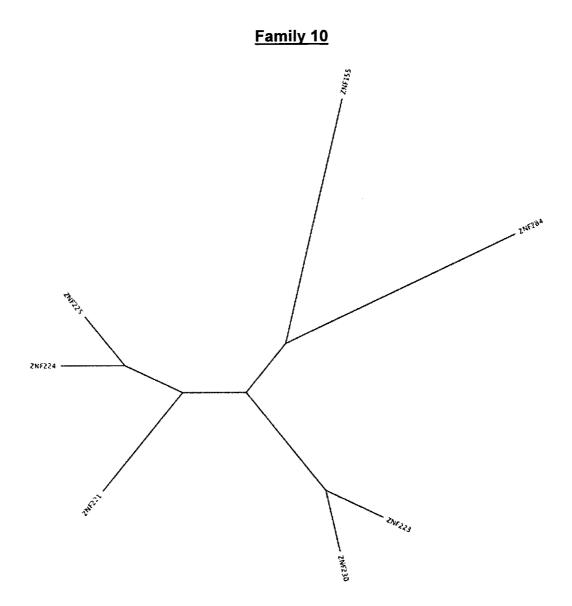


Figure 1. (continued)

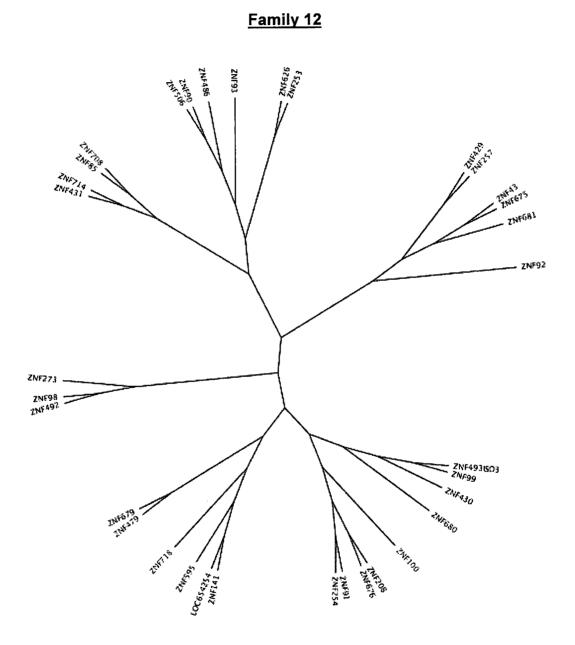


Figure 1. (continued)

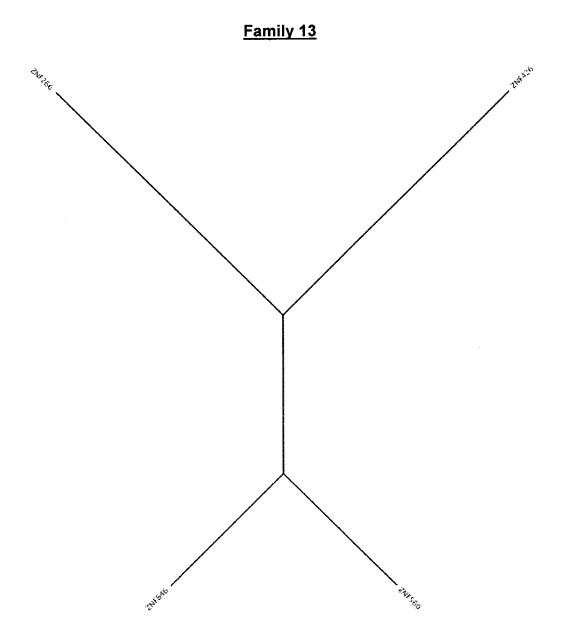


Figure 1. (continued)

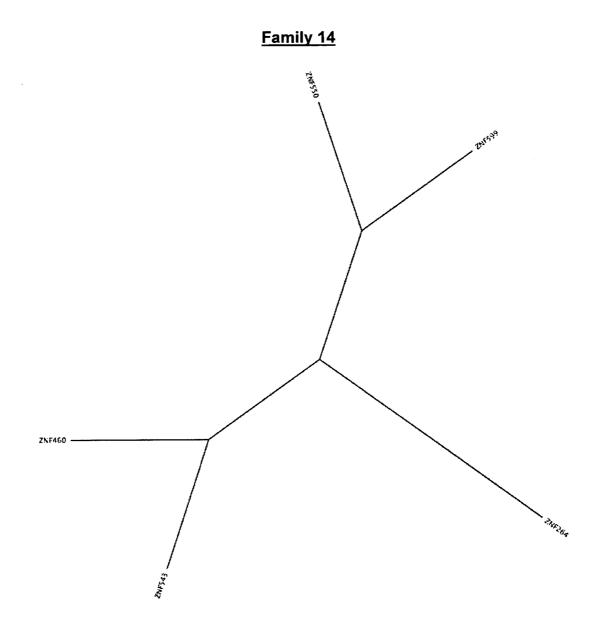


Figure 1. (continued)

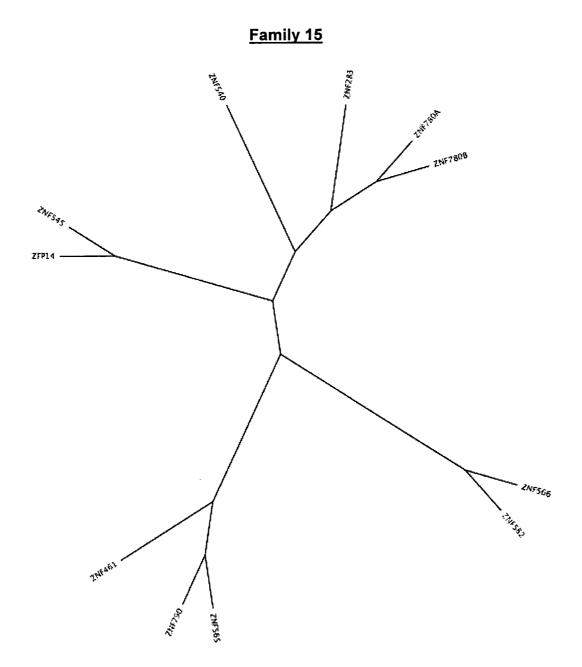


Figure 1. (continued)



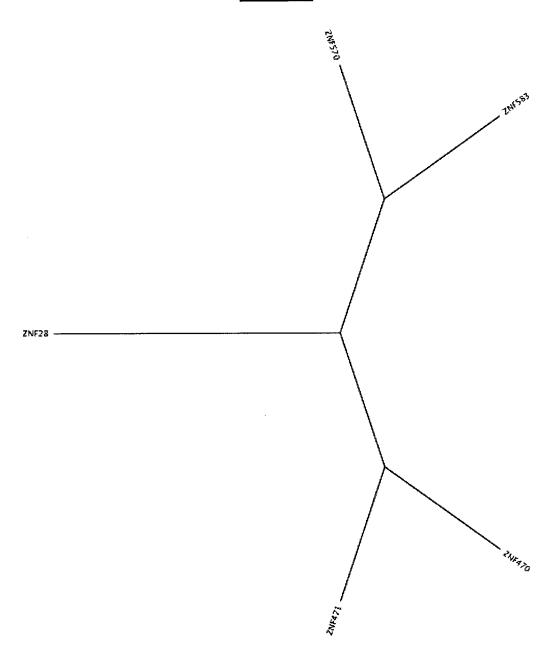


Figure 1. (continued)



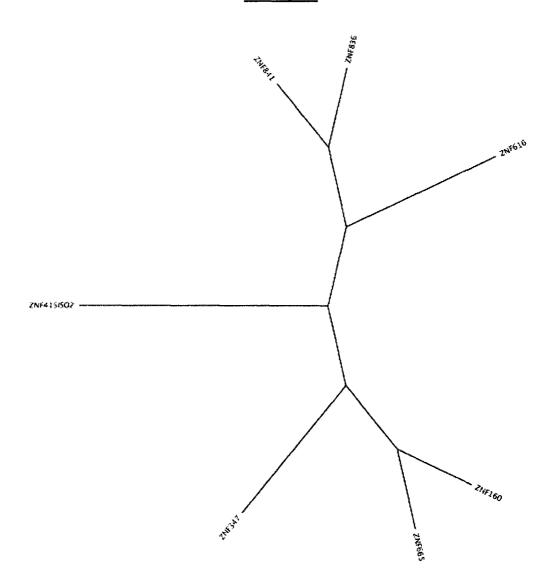


Figure 2. Phylogram tree showing the evolutionary history between all 18 "linker families".

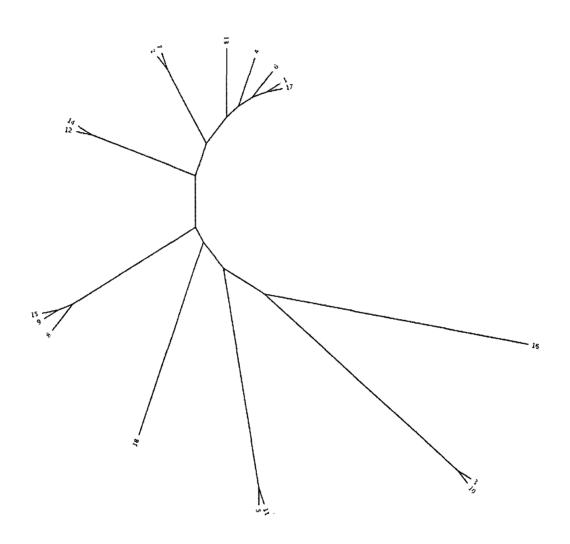


Figure 3. Phylogram tree showing an evolutionary history between each of the 334 individual linker regions of the KRAB-ZFP superfamily.

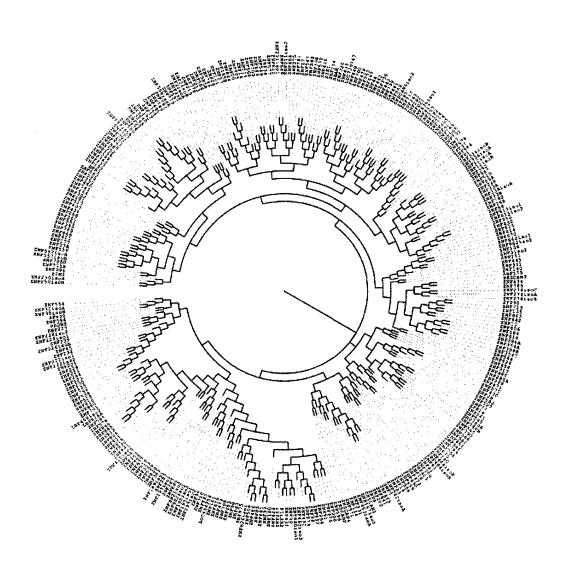


Figure 4. Diagrammatic representation of the KRAB-ZFP Subset. The N-terminal KRAB module serves as a repression domain. The C-terminal Zn-Finger repeat serves as a DNA-binding domain. Each member has a linker region of various length which may serve as a mediator for unknown and unconventional functions.

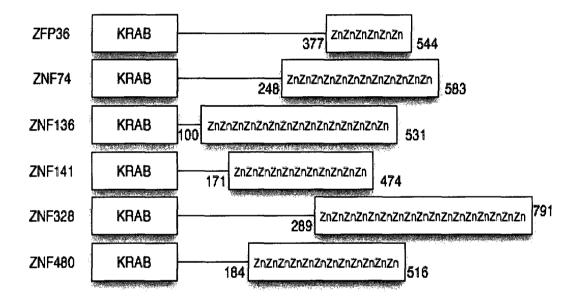


Figure 5. Diagrammatic scheme of the fusion constructs made from the DNA-binding regions of the 6 members of the KRAB-ZFP superfamily. The N-terminal GST tag is used for GSH-affinity purification. The C-terminal Zn-finger repeat is the region of our protein-of-interest which is under investigation.

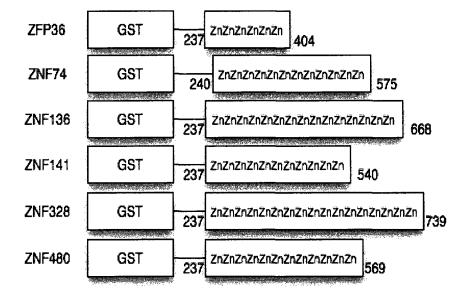


Figure 6. Expression profiles for each of the six expressed members. M=Marker, U-Uninduced Sample, I-Induced Sample, U-Uninduced Control, C-Induced Control. Black arrows - Fusion protein; Red arrows - GST.

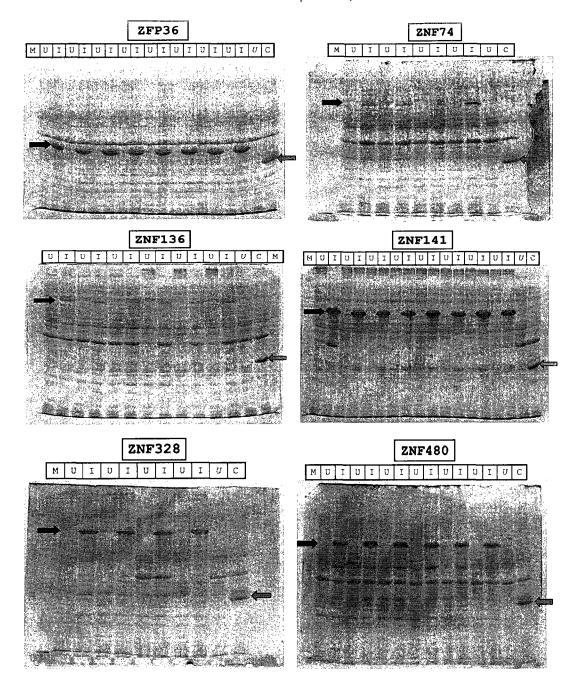


Figure 7. First GSH Affinity Purifications. I-Induced Sample, *I*-Induced Sample, F-Flowthrough, Lanes marked 1, 2 and 3 are Elutions 1, 2 and 3 respectively. Black arrows - Fusion proteins; Red arrows - GST pure protein.

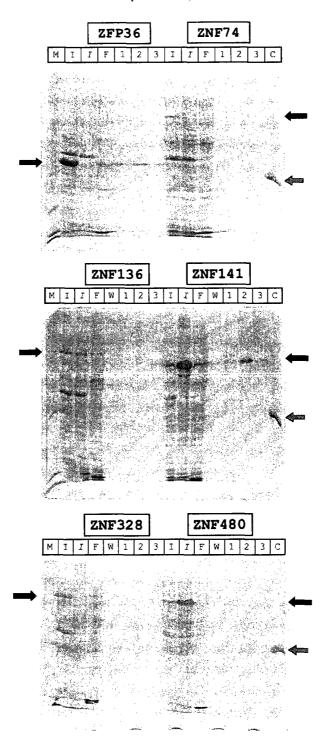


Figure 8. GSH Affinity Purifications. After reworking the protocol, all members were successfully purified and eluted. I=Induced Sample, F=Flowthrough, Lanes marked 1 through 6 are Elutions 1 through 6 respectively. 'C' denotes GST control. Black arrows - Fusion proteins; Red arrows - GST pure protein.

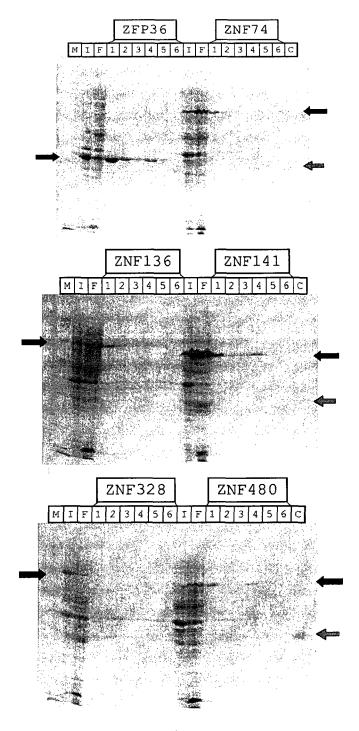


Figure 9. Final three GSH purifications of proteins that were folded *in vivo*. All elutions were used regardless of band intensity. Lanes marked as 1 are elutions 1-3 combined; Lanes marked 2 are elutions 4-6 combined. Black arrows/boxes denote target fusion proteins. Red arrows denote pure GST as a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

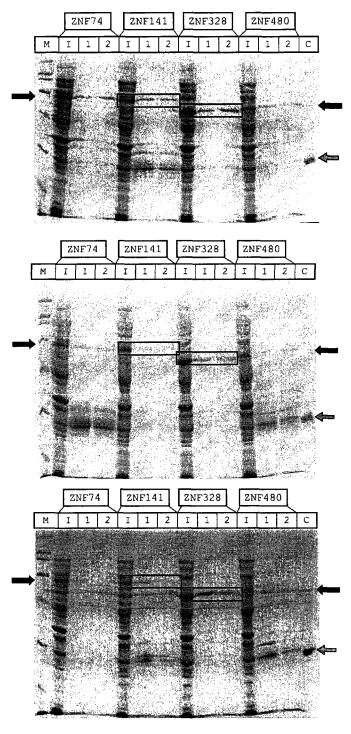


Figure 10. Cold-binding to pre-enrich the randomized oligomer library before hot-binding. The bands in lanes marked 1,2,3,4,5, and 6 are oligomers that were selected by ZFP36, ZNF74, ZNF136, ZNF141, ZNF328, and ZNF480, respectively. Black arrows denote the oligomer bands.

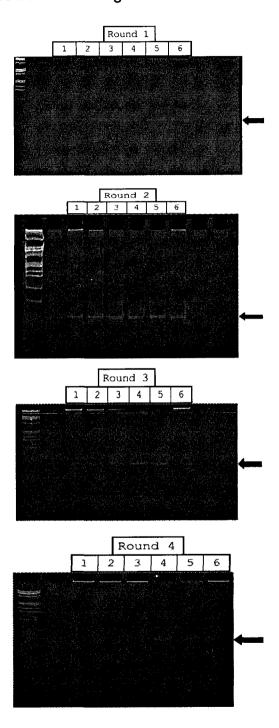


Figure 11. Radiographs from 3 rounds of "Hot-Binding" (EMSA). A radiolabeled randomized library of oligonucleotide 49-mers was subjected to binding with the DNA-binding domains of 6 known KRAB-ZFPs. These DNA-protein complexes are large and will be retarded in their movement through a DNA-PAGE gel. This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

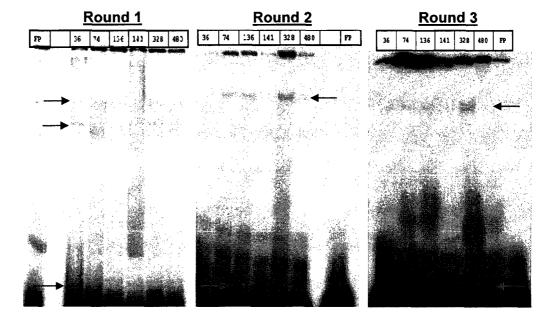


Figure 12. Hot-binding to select oligomers that specifically bind to the DNA-binding domains of our KRAB-ZFP subset. Bands represent selected oligomers after PCR. Round 3 involved loading entire PCR reaction and electro-eluting the oligomers from the cut bands.

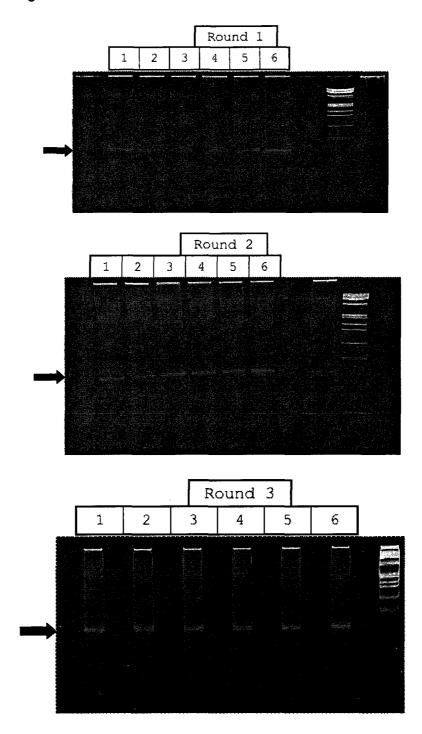


Figure 13. Digestion of selected clones to check for the presence of the oligomer insert. Black arrow denotes the insert that has been cut out of the pUC18 plasmid. The plasmid remains in the well. ZFP36: 1-30; ZNF74: 31-60; ZNF136: 61-90; ZNF141: 91-120; ZNF328: 121-130; ZNF480: 131-160.

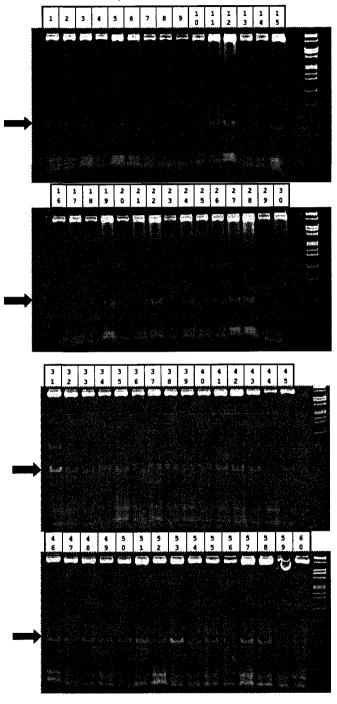


Figure 13. (Continued)

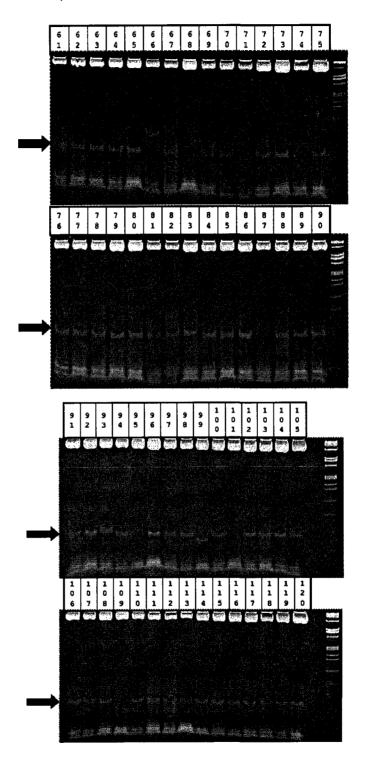


Figure 13. (Continued)

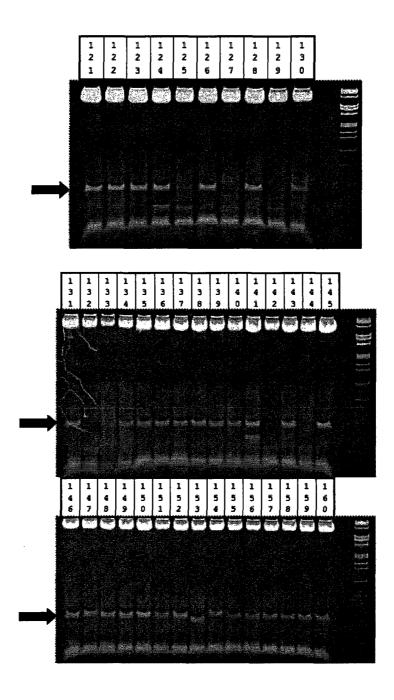


Figure 14. Plasmid concentration prior to sequencing. After the insert check, each clone was run on 1.2% agarose to determine the concentration of plasmid DNA. ZFP36: 1-30; ZNF74: 31-60; ZNF136: 61-90; ZNF141: 91-120; ZNF328: 121-130; ZNF480: 131-160.

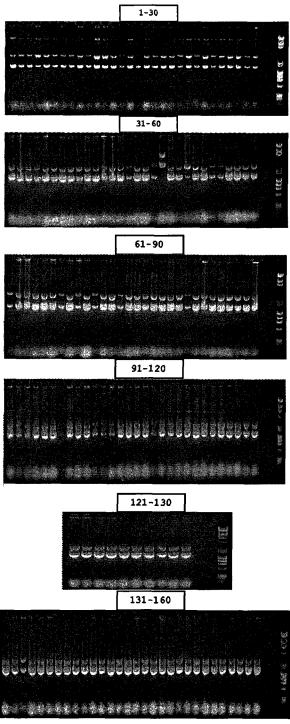


Figure 15. Consensus maps for ZNF36, ZNF74, ZNF136, ZNF141, & ZNF480. Numbers are percentages of occurrence.

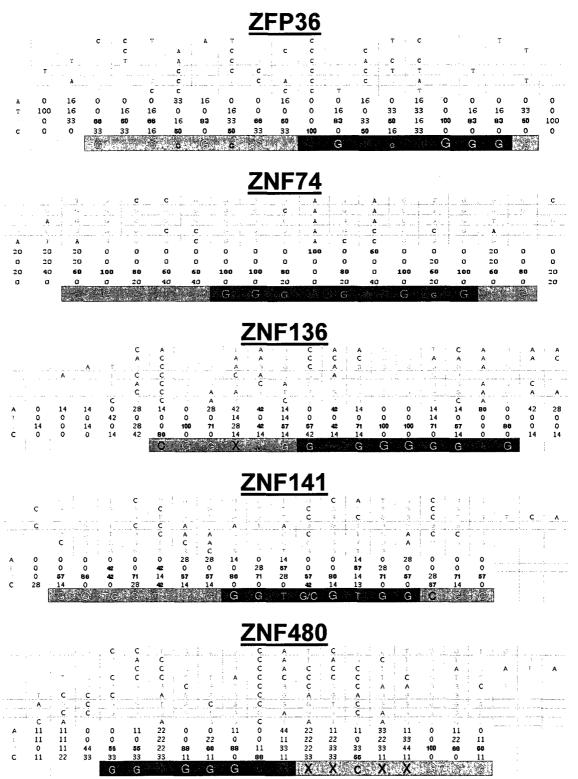


Figure 16. List of potential target genes for human ZFP36. The first box is for the consensus binding site, and each box following represents the derivative sequences. Target genes for all sequences were derived from two sources: Transfac and EPD. The gene name, binding site location, and associated 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.

Consensus 1	Section 15 Continue						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Stran
Transfac	PTK2 protein tyrosine kinase 2	-142011549	cadadussadiddcdasaddacylgogoldddddddddadcasccadscccadscccaciad	-142011495	X-KYr	8	-ve
	N-mys downstream regulated 1	-134209574	actilicáaláid.czádáfedádcijadacijygganggdááccdzáádacserereficác	-134309314	EGR-I	ŧ	-75
	N-mys disministread regulated 1	-134209574	acigodadagicciqqaeqqqqqqqqqqqqqqqqqqqqqqqaaeqqaaaaeqcqc	-134309514	SP-1	é	-74
	chemokine (C-X-C motif) receptor 4	-136876203	congaengygongognengongaeagogogoGAAhggognhGGAAgonhopochaoth	-126876143	c-Ets-1		-75
	high-wability group nucleosame binding domain)	-407211079	adadaddadadcostaddocatpppy//www.com.cogaddadaddadddoccorast.codd	-407211019	ZFP219	21	-V2
	parathyroid hormone 1 receptor	46923627	ātādādatalādasasaiatecēdāsēdācātēda;;;;yggasāsadatācecēdēcādādad a	46323687	MAZ72P-1	. 3	÷ve
<u>epo</u>	high-mobility group nucleosome binding domain 1	-407211079	3333ad3ddaadccaacdcodgggggcgcgggggggggggggggggg	-407211019		. 21	: ~ve
	eniothelin receptor type B	-78493001	ogdataciattiacoccoccocggoracgeggggaagAAAAACACCTGAGAGACAACACC	-78492941		13	-ve
	processe, serine, 23	86511509	Ad3000000000000000000000000000000000000	86511569		11	÷४€
	solute carrier family 35 (CMM-stalic acid transporter), member Al	88152641	ekchassesserecteration of responsered analysisserecteration	88182701		. 6	tva
	protein disulfide isomerase family A, member 6	-10952928	23461500079336653364635365555555555555566664646555665555	-10952868		?	~ ve
	H3 histome, family 38 (H3.38)	-73775899	aagcgigccialaaaacggaggcgangcgjgggciigGAGCGCACACCGGTTTGGTCET	-73775839		17	-ve
	structural maintenance of chromosomes 2	106556908	ADMINISTRACO ADDROBAÇÃOS PO DO RECONO PROCEDE A REGLADO ADDROBAÇÃO A RECONOCIDADA	106856868		9	tva
	processone (prosoze, macropain) subunit, alpha type, 4	78832750	CCCCCAACCCACCCCCCCCCCCCCCCCCCCCCCCCCCC	78832810		15	ŧ49

ZFP36

Derivative :	execution and con-				
	Official Gene Name	Start Pos.	Binding Site	End Pos. Binders	Chr. Strand
Transfac	rone 				
<u> </u>	serine/threonine kinese 16	22011VIBS cggac	- CARTON TOTAL NOTATE CONTENSES FUNDS GUICARPROSE (CONTROCCIONARCE)	AGG 220110245	2 +ve

<u> 27936</u>

Derivative 2	An enthance - successes
	SAME AS CONSENSUS

Figure 16. (continued)

****	rin.	150
11	М	č
C U	u	

Derivative 3	ngir Josephiniae kapini na ari and antifediaet position antifed						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	<u>Chr.</u>	<u>Strand</u>
Transfac	rone						
EPD	POU class 2 homeobox 1	167190077	dadaracadadcaadddaddanraradascaddcaarriiddiraasararraraaand	167199137		1	178

<u> 2</u>7736

Desivative 4	anapagnapanan meridi dipological production						
	Official Gene Name	<u>Start Pos.</u>	Binding Site	End Pos.	Binders	Chr. St	rand
<u>Transfac</u>	tryptophan bydroxylase 2	72332121	accolatologaticatologoa Poglas goolgogotaasabalosoda ogtgood babasa	72332191	none	.2 ` +	ive
<u>epo</u>	serine palmicoyltransferase, long chaim base subtmit l	-945777116	gagcagcgacgcgcaccolligggacGCVCT.GLGACCCCCTCCGCAACCAACCAACCAC	-94977656		19 -	-ve
	picultary tumor-transforming (interacting protein	-46293617	GRECCECURE (C. TECHOCCERAGECTO) (C. HOTGONG VOGRAC POTANOSCICUAD	-46291557		21 -	-V e
	inal (Har4C) homolog, subfamily A, member 2	-47007579	contradorateleorocogocycocogni a ladascogycocognicatori casci lod	-47007519		.6 -	-Ve
	CDES regulatory subunit associated protein 1	-31989357	gootgaagtegoggaadGGAAGTG;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;;	-31989297		20 -	~V€

ZF236

Derivative 5	ง กำกับกับกับกับการแบบการการการการการการการการการการการการการก						
Ì	Official Gene Name	Start Pos.	Binding Site	End Pos.	<u>Binders</u>	Chr.	5trand
Transfac	v-Ait Hardy-Ruckestan 4 feline sarcoma viral choogene homolog	55523982	czásáerdesácárászczádázádásádásátádádádátárárádásggggggggggggggg	55524042	85°.	4	446
	thrombonedulin	-23039692	actemicocoliciadótémicosmició de operación de adiocóliciada de acce	-23029542	SP1	25	-A5
EPD	nuclear receptor binding protein 1	27651443	graccocca.cgttgg.ccgggcdCGnAAAAAAACCDAAAATCCGTTGCGTTGCGTTGC	27651508		1	÷yę
	ATD-kinding cossette, sub-family 8 (MDR/TAr), member 8	150725474	tgtcttttesceatgaccaaggggigaggggitGGCVAGATGGGGGGGGAAGCA	150723534		7	172
	proceasome (prosome, macropain) 246 subunit, non- Alfase, 7	74330667	actorgggagoggaschmanscolocolocolocolacchmanscolacchactic	74330727		16	÷ve
	SWI/SWF related, matrix associated, actin dependent regulator of chromatin, subfamily e, member l	-38904122	dasfáciadsfáciádssgujuchas couras consecuents accusately ny daten	-38604062		17	-ve
	M-phase phosphoprotein 6	-62203890	dddieddddiddiddddddodcaeceaedriaaiddaadeadcrieddaarddddorocdda	-92203930		16	-ve
	EM-domain containing 1	-64646137	Encontrol of the second of the	-64646977		11	-78

2FP36

Derivative 6	edelin bada ningi ingkar dalam bada ningi ingkar						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	<u>Strand</u>
Transfac	tyrosinase (oculorucaneous albiniss IA)	88910975	trocacticaecateggoClA9COCATTOS.COSA1ACCACCCAAfrogaaagaeeagro	88910935	CORR	11	176
						•	
EPD	complement factor A	21913831	MERCHETTO, ATAMARICE CONSCITATE, SEGMICOLONICAMENTAL SEGMICACION	31913891		5	łya

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 sequences. 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.

Consensus 1	PARAMENT STATE TO SELECT STATE OF SELECT STATE S						
	Official Gene Name	Start Pos.	<u> Minding Site</u>	End Pos.	Rinders	<u>Chr.</u>	Strand
Transfac	placelet derived growth factor C	-157832584	coorteddiddiaddiaddiaddoddiadaladalocyce, yn yn ycchrigol	-157892324	1-803	4	-VR
	placelet derived growth factor C	-157992584	cacciogggggtggggggggggggggggggggcgcCCCCCCCCCC	-157892324	\$P-1	į	-ve
	prostaglandin-endoperoxide synthase 2 (prostaglandin S/H synthase and cyclocxygenase)	-186650036	canagacoagacacgycggcggcggcggagaaaaaGAATECCctgcycccccggacatcag	-196649976	pSC: HelA- pGS	•	-ve
	lymphoid enhancer-binding factor (-109090178	gccaagSASACaagaggaggagggGSASAGggagCaagacgcaagtgggtagc	-109090115	Lefl	4	-V a
	lymphoid enhancer-binding factor l	-109090178	drcsadnyyycoredadidddddadddayyyyddai goriodacdcaediddiadi	-109090118	SP-1	4	-ve
<u>EPO</u>	origin recignition complex, aubinit 4-like (S. cerevisiae)	-148719382	decdavediadd:cecdidiriddd:ldlededddd:ldlededdd:ldledeldrediddidc	-149778322		i	-ve
	advanced glycosylation end product-specific receptor	-32152068	ccccacclagggcggaggccacagcagggaGASSASCAGACACCCCACAACCACCCCCCAAAC	-32152008		Ę	-16
	midkine (neutrice growth-promoting factor 2)	46403159 .	ddidcdddressereicsdddddisiddairiaddiaddedidddeceridsdaecadosea	46403219		11	+1/6

2NF74

EPD	rone		e e e e e e e e e e e e e e e e e e e				
	cheroxine (C-C modif) receptor 5	46411896	ggggg:cgggtigggataggggatacggggagagiAAAaaggggacacagggtiaa	46411946	C/EB/beta	. 3	÷ve
Transfac	polymeric umunoglobulin receptor	-207519885	rdacccidicidicuariduccupulanymistridecentraadadccordinatiddi	-207119929	AP-2		-1/5
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	<u>Chr.</u>	Strand
Derivative 1	THE CONTRACT						

Figure 17. (continued)

2NF74

Derivative 2	acotasti in kiskinda da						
	Official Geog Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strans
Transfac	v-byo byelocytopecosis viral oncogene homolog (avian)	128748496	aacteGCTGTAG.AATtCCAGCGAGASSCAGAGGGAGCCAAGCGGgccggctagggtg	128748556	orce	8	tye
	v-mys syelocytomatceis viral cocogene homolog (avias)	128749496	aactogotątag/AAYTCCAqogagasysaqaqojaqogagogagogqooaqqqog	128748556	U87-1	ş	÷ye
	v-mys myelocytomatosis viral encogene homolog (avian)	128748496	aactogoogfAGtaattocagogagagocagagogagagacotatocaactogoogtag	120748556	PT:3-1	ě	÷v2
	CO44 molecule (Indian blood group)	35160525	acagcotoageagageacggggWWBBARAgaggggccgccogggaggggcgctactacttc	35160588	EGR-:	11	179
	COA4 malecule (Indian bloom group)	35160528	acadocicadesadadesegaggggggggggggadahófidicadociddadddcidiraceic	35160588	sp-1	11	ive
*	epoxide hydrolase Z, cytoplasmic	27348540	cocadocedad recadaceres ocadad rycky grant grant cocad residia da da cocada ce da como como como como como como como com	27348600	\$P-1	ě	÷vē
<u>epd</u>	carcinoembryonic antigen-related cell edhesion molecule 5	42212515	tactoctgcoctageAAGASACTGAGGGCAGAAGAAGGAAGGAAGAAGAAGAAGAAGAAGAAG	42212575		19	175
	MM-domain containing 1	-67505862	geronners and the second of th	-67505802		11	-72
	hypoxamihine phosphoribosyltransferase [133594141	cggggcctgcgggggtgtggcggggggggggggggggg	133594231		ĭ	445
•	\$100 calcium binding protein B	-48025061	geacea;qitteatecatectecet@360A3AC40AATAAGA96CCCCCCTCTGCCCACCA	-48025001		21	-45

2N274

Derivative 3	AGGGGGGGAGAGGGTGG						
	SAME AS CONSENSUS						

<u>20274</u>

Derivative 4	ganarjartan yarigi ka Marang atakan duaha						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
<u>Transfac</u>	sphingosine kinase l	74380526	celedecódástradátramya <u>ccenenysegengegengeg</u> gadáccádodecásád	74350586	\$P-1	Tr.	tye
<u>epo</u>	creatine kinase, brain	-103989214	condicional description of the state of the	-103989154		14	-45
	flap structure-specific endomiclease 1	61560334	FOR THE SECRETARY SECRETARY OF SECRETARY SECRE	61560394		11	÷ve
	NADR dehydrogenase (ubiquinone) flavoprotein 2, 2440a	9102633	cát coddá, cá es rabáca t dábá red désdede de acáca que processo como se cát codo a se como como como como como como como com	9102693		18	÷Vē

<u>2NF74</u>

Derivative 5	Millionary and S					
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr. Strand
Transfac	cyclin-dependent kinase inhibitor 20 (pl0, inhibits tux4)	51434236	adaácóddaálolddaridaddócóckálrolystjányrdáráddiácdadaridedaí	51434296	\$P3/\$P1	1 198
	epidernal grownh factor receptor	15086398	acatacidetdiácetácaagtecidedidiaceinejunyyyyenegegyvátidetádágaacác	35086459	VJR	? łye
<u> </u>	tose					

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 sequences. 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.

Consensus	ensus physicaeths a cathaidhe tarailte						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	
Transfac	v-myb myeloblasicsis viral oncodere homolog (avian)	138502366	sacideaddddelcadarridggggwygggyygidicasaulciairidardarda	135502426	%2F16-C	6	+45
	synapsia 1	-47479564	coccaeactgcqcatcccctatcAGAGAGGGAGAGGGGGAAAcaggatgcggggggggggggggggggggggggggggg	-47479504	none	X	~¥ē
	high-mobility group nucleosoms binding domain)	-40721192	ddaidddddddccddccddcggggyyyyygercdcdddcdddddddddddd	-40721042	217219	21	~ve
	insulin receptor substrace 2	-113439186	Accolácesádacióádourrededá/kamagyeggáseddedálesseárddereddá	-110439126	582/593	13	-45
<u>epd</u>	high-mobility group nucleosome binding domain l	-40721102	ddagdeddaereddaeddaddaeddaeddaeddaeddaedddaeddae	-40721042		 21	-72
	rhodopain	129247411	agatgolgattoagosaggagottaggagggggaggloasttcalaagggloiggggg	129247471	**	3	÷ve

<u>2N£136</u>

Derivative 1	dadahan andadahan					
	Official Gene Name	Start Pos.	<u>Binding Site</u>	End Pos.	Bindera	Chr. Strand
Transfac	anti-Wellerian hormone receptor, type il	53817345	agicaalagagticagcalcticiicoAtBVTCApggaaqggcasagatiigaaattagg	53817405	82-1	12 +ve
	coagulation factor %i	197197022	atytotic:fff/ASACff:GAGATiceAggicagetagaettaatggctaacagctgac	187187082	EM2-telpha	4 +ve
	thrombopoletin	-184096874	totggggacaggggatgacgGGATCASGTCASCCAggaagcccccgaggacagaga	-184096814	Mil	3 -va
	prolime dehydrogenase (oxidase) 1	-18924811	terggggggggggggggggggglCACAEGTCAEGACAACCAAGACCateetqgetaaca	-18924751	Platquasa	22 -ve
<u>epo</u>	core					

ZN£136

Cerivative 2	ACCOMMISSION					
	Official Gene Name	Start Pos.	<u>Binding Site</u>	End Pos.	Binders	Chr. Strand
Transfac	rone					
						•
<u>epo</u>	coce					

Figure 18. (continued)

2M2136

Derivative)	ajandos Nicas cino na se Successión de constituidos			_			
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	<u>Strand</u>
Transfac	fragile X mental retardation, autostral homolog 2	-7518225	rgcker rgcgeyousåådoddeådiddersåeikkaaldraåavåderdidaadodderadodde	-7518165	:-1 <i>x</i>	17	-45
	apolipacrotein £	45400911	erceprergepergerquqqerq3800A3980agaaaaqeceaecroguqaergggga	45498971	AF-2	. 19	iye
	interferon regulatory factor 1	-131027386	erecreared reservancing concessor accordada de deservas and de	-131827326	p50: Rela- p65	į	- <u>ua</u>
EPD	apolipaprotein B	45438983	occtarecetggggggggggggggaraggggayscctataettggaesagtetb20AT	45409043		19	145
	Iransmambrana protein 43	14166510		14166570		3	+98

<u> 207136</u>

Derivative 4	Appropriate production of the control of the contro						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	<u>Bindera</u>	<u>Chr</u>	<u>Strand</u>
<u>Tzansfac</u>	ATPase, Carr transporting, cardiac muscle, alow twitch	110718915	cedacarácadadesigadaesicadasediseis _{toris yaga} dacaticosocodocalasis.	110716975	F123/142	12	₹ ₩
<u> 220</u>	rhodapsin	129247410	ccapaigcigaticagicaggagcilaggaggaggalactilataagggiliggggg	129247470		3	+ye
	titis-cap (telethonis)	37821534	ccármerráccccidamamidádinámákásákászattramadádcciddieddádid	37821594		17	÷45

<u>ZNF136</u>

Derivative 5	ele complete acceptance a mos electronista de la complete de la co						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	cystathionine-beta-synthase	-44496133	adacideddedddaggggyyyydggggyyyddddecedddeiddaldddidagd	-44496073	XZ5-1	21	-V -
	mucin 58, oligomeric mucus/gel-forming	124567.	ceptotygytea@ddiaeec.TopottygyteSchilocoft.gcotgygtt@ddiaeec.	1245731	šr:	11	ŧγş
	cystachionine-beta-synthase	-44496133	#34c.dcd3c3d3d-ggggyydggggggggyddddccd3dccd3dccd3dcdddcdaddad	-44496873	SP1/SP3/SP4	. 11	~Ve
<u>epo</u>	septin ?	25840781	PAROLITECUCIONE CONCUENCACIONA CON CARROLLO DE LO LO ENGICA QUE EN CARROLLO CONCUENTA CONCUENTA CON CONTROLLO CONTRO	35840841		7	÷ve
	heterogenesus suclear ribonucleopsolein R	-23670928	ACCE MADAGES SECRETARIOS ASSESSEDANTO POR ACCESANT	-23670768		1	-ve
	profilin 1	-4851892	rācsastesassercradedādācādā kākā feikaldsadāsadāsadzādācādrācāsādāc	-4851837		17	- 78

Figure 18. (continued)

207136

Derivative &	олове даналения в в в приводения в в приводения		 		
	SAME AS LAST SEQUENCE		 ·	•	

2NF136

Derivative ?	renderalander Veber populativação						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	GLT family zime finger 1	57853966	conagogorougacaquigoqaayiggqqqayacagscqqqqqqackCTTCqqaqcayrqq	57834926	1851/1361/#- Pris:	12	+118
	collagan, type II, alpha l	-48398545	 iddddaedadcoleedadarridddidiaedddcarcrerdcdcoccdcacdar	-48398485	\$2.	12	-95
<u> </u>	CORE						

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 sequences. 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.

Consensus	ARA YARA DI ARRIGA BEBE MET MET ARRIGA						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	<u>Strand</u>
Transfac	v-myc myelomymomenosis viral smoogene homolog (eviam)	128746643	cococquatigicicorricionaments chambanamaticaecranans	128746703	PUMbeta/PUM-	ž	fye
	ν-πγο πγείοσγιοπετοsis viral οποσφεπε homolog (aviem)	128746643	concodasciditinoconnidda@c@clow@cyadadagaaadiisaciiaaaaaf	124746763	MAI/ WA- Factor	8	łya
	adenosine deaminase	-43274987	Eacocatettyacicacatggcagttoggtg;pgaggggAACAAAG;agastgagttica	-43274927	LEFT isol	20	-725
	adequaine deaminase	-42274987	tacosatuttgattoacaTuRCAUTTUgtgyrggaggggaacaaaygagattgagtttca	-43274927	c-Nyio	20	-78
	paterhally expressed 3	-57351408	-Counselvesculler recounsive tell solutions usagger georgects:	-57351348	tone	19	-ve
<u>epo</u>	solute carrier family 31 (copper transporters), masker $\frac{1}{1}$	115983818	GCCGGAAAYCCCCCGCCCCCGCTGGCCGCCGTCAAACTGAAG	115982878		ğ	tye
	DEAL (Asp-Glu-Ala-Asp: box polypertide 25	125774342	CATOUR CAMACION FOR COME CONTRACTOR AND AND ARREST COME CONTRACTOR OF COME CONTRACTOR CO	125774402		11	tye
	ubiquitin specific peptidase 16	1094401	algylcoogggaggtgggggtgggglgglggcgfk%CACTCCCATAACGCCCGF	11196961		21	+7/2
	tripeptidyl peptidase i	-6640719	Cartagtiartaggragaggggtagt@S000000000000000000000000000000000000	-6640659		11	-778
	ras-related C3 bottlinum toxin substrate 1 (rho family, small GTP binding protein Nacl)	6414160	TECHNALAGITANA ANTENNANTENIA A	641.4220 -		7	÷vą
	ribonuclease 82, subunit A	12917440	OTCO-GOAGTATIAGMOT RECAGE TSG (GOT GACRETTCAGRESSES), GGARDICAGRES	12917500		19	175

Derivative 1	selicente ini ann selicente ini ann	·	<u>"</u>	, , , , '	-	
	Official Gene Name	Start Pos.	Binding Site	End Pos	Binders	Chr. Strand
Transfac	5008					
						1
<u>en</u>	COSE					,

207141

Derivative 2	Magagaga - gara ada adagagaga - gara ada					
	Official Gene Name	Start Pos.	Binding Site	End Pos.	<u>Binders</u>	Chr. Strand
<u>Transfac</u>	alconol denyarogenase (C (class)), gamma polymagnide	-10274073	iggaaragicaaggesi (PANACOS), Coo ACCA, T.C. Trangaciigaragi e	-103274010	cone	4 -ve
<u> 110</u>	ECRE					

207141

Derivative 3	toploper manachterite Tidens unschaubt in	-					
	Official Gene Name	Start Pos.	Rinding Site	End Pos.	Binders	Chr.	<u>Strand</u>
<u>fransfac</u>	anti-Mullerian hormose	2249207	rddddedrodddaeridrosecyywerregddeddaddadaraddddrosdea	2249267	or Va Wa •	13	*78
				**		***	
<u>EPD</u>	septin 7	33840772	#CONCREME ACTION	35840332		7	†VB
	OCTA domain containing l	4802460	ANGINANELLANDE SANTONIO ANGINE	45833123		A G	†¥ 8
	MAX-like protein X	40719017	odardorredecerderecranererskitrebskiarrókskikdedeskiktranáródása	40719077		17	1 98
	complement component l, q subcomponent binding protein	-5042484	NTOCONY - Coccapando plane de coccación de cocapa de la fina fila	-5342424	***	17	-98

<u>207141</u>

Derivative 4	desidentes sur code						
	Official Gene Name	Start Pos.	Binding Site	And Pos.	<u>Binders</u>	Chr. St	rand
Transfac	topolsomerase (CNA) 11 alpha 170kOs	-38574295	acaaaccocgyccanii Charla at animee Charataaaaygcaagcaagaatgatt	-38574235	none		-ve
	·		. "				
<u> </u>	tote			*			

<u> 207141</u>

Derivative 5	the selection of the se						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Stran
Transfac	serine/arginine-mich splicing factor 2	-74733758	acijogotuja iggotijagotakasakų pojogi SAAC. Tygagota tujovat gaasakakų	-74733728	c-Myb isol	17	-ve
	ferredaxin 1	110300580	ccocracecer to dosceso à central se de grança granda de de de de contra de contra de de contra de de contra de contra de de contra de co	110300640	\$71	:1	*V e
	ULIS binding protein 1	150284996	cendicidicidicadardacaabicidhahidyyyyyyadarhdaabaabaald	150285056	SP1/3V2	6	tve
***	ULI6 binding protein 1	150284996	ceagacdacatcadeadacdacdagagagagatdacdaeaseada	150285036	AP-Zalpha	6	. ive
	protein phosphatase 2, catalytic subunit, alpha .sozyze	-133561325	renderposedanteridaedindadistatus (Acaredosedindandinapi	-103561765	6460	. 5	-45
<u> </u>	/K536 binding protein IA, 12kDa	-1373784		-1373724		20	-ve
	protein phosphatase 2, catalytic subunit, alpha isozyze	-135561825	production and an experience when production and production of a	-133561765		5	-7 <u>e</u>
	MASSI homolog (BecA homolog, E. coli) (S. cerevisiae)	40987313	şaggoşgoşatacgilecgicgacgcyggcgigaczcigggcgagAGGGT,TGGCGGGAA	40987373		15	ive

2NT141

Derivative 6	Control of the contro						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	<u>Binders</u>	Chr.	Strand
Transfac	fibroblast growth factor 2 (basic)	123747657	rdirrriddirdirdoddddriadiggayygigaerrriddiidararaddddaddrdiai	123747717	EGR-1	4	tve
	fibroblast growth factor 2 (basic)	122747657	rástradádsádsáránádádássegggsádaságsádsádatádadádásádsádsádatáda	123747717	390	ć	+72
	apolipoprotein E	45409201	qqqqtcqqqtcqqqqaaqaqqaqCSGSYSTSASGCAAGCAGCaqqqqacqqaacqq	454092EL	cone	19	+45
	vitania D (1,25- dihydroxyvitamin 13) receptor	-48299207	- gaagtortggcotggtoageecaggtggggtDACGCACotggcotcaggcgtccgcagea	-48209147	ICER-Ligarna	12	-ve
<u>890</u>	emplase 3 (beta, muscle)	4854376	ctcrypyriaan occance tonga obsestiance tonine to tot ence to each c	4854436	٠	17	÷ve
	lysophosphatidylcholine acyltranaferase 3	-7125833	contribution of the state of th	-7125773		12	- va
	heat shock protein 93k0a beta (Grp94), member 1	104324127	reseccicocosáccoásádávadzádáðalavasácádorodvosácsíðssíðsát	104324187		12	+ve

<u>2NF141</u>

Derivative ?	marija ironija i i i i i i i i i i i i i i i i i i			:			
•	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	LONE						
<u>epd</u>	solute carrier family 31 (copper transporters), member	115983816	CONTRACTA DE ESCACACIÓN DE ACADA (SE ESCACACIÓN DE CONTRACTO DE CONTRA	115983878		9	172
	DEA2 (Asp-Glu-Ala-Asp) box polypeptide 25	125774341	can especial contraction of the	125774401		11	tve
	transmambrane emp24-like trafficking protein iC (yeast)	-75643334	control control to the control of th	-75643274		14	-ve
	transcription factor 12	57210825	qqaqqqatcosqasogascoqassosatus oakkoosocotoxogaaaaaaausec	57210985		15	÷ye
	cetraspanic è	-99891744	SCICICIO SE PREMA CONTROCO COSTO POR PROCESTA O CONTROCA A CONTROC	-99891664		X	-y <u>a</u>

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 sequences. 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.

Consensus	SOCEOGRAXIONAGE:	-					
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Transfac	E2P transcription factor 6	-[1606337	redürdrördldaöskirredöradeségggygytAldsésaladöddldoósfraerätä	-11606277	MRF-1	2	-79
	congulation factor III (thromboplestim, tissue factor)	-95007406	asidadoddold lodedddodggggggggggygyyaaldaficdidafadedadcodonllo	-95007346	6GR-1/SP1	1	-7e
	solute carrier family 9 (sodium/hydrogen exchanger), member 3	-324660	āpunācācinādoāaānādoādā,50.550,50.550,750,50.750,00.00.530,00.00.500,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.500,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.530,00.00.500,00.00.0	-524630	SP3/SP1	5	-7€
	solute carrier family 9 (sodium/hydrogen exchanger), member 3	-524660	äteråeåereåeåfädeded366588%5388.säåereeåareeåådåedååaååååde	-524600	EGR-1	5	-ve
	transforming growth factor, alpha	-70781110	täässästetetstassttttttttttttttttttttttt	-70781950	AP-2	2	-ve
	prostaglandin D2 synthase 21kDa (brain)	139871609	adiddasdr@WGMygaddesreddaddesddssaidaeidadaaseaeeefdseee	139871669	1212-1	. 3	+ye
	linker for activation of 7 cells	28996374	adzircorderderdergeggggggggggggggggggggggggggg	28996434	\$81/883	16	+45
	cholesteryl ester transfer protein, plasma	56993468	graarechagracheryqqaqyCVSAP9XWWTxqarcastryapgreaggayrtgagae	36993328	none	15	÷48
<u>epd</u>	Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog	-27961697	CO COCOCIAS CASSOCIOS COS TORBA NECESCIPE SICACOCIAS POLICOCOS AL COCOS	-2796163"		ī	~Ve
	beta-2-micraglobulin	45003691	TOTAATATAAGY DRAKKOG FOOGG TORGOSKA POTTOGAA KOOGACADAA FE	45003741		15	÷ve
	protease, serine, S	-31147121	URSCOLATTORSCOMSCOLATIONAL ASSOCIATION ASS	-3114710El		16	-72
	proteasome (prosome, macropain) 263 subunit, non- Addase, 1	231921529	######################################	231921589		2	† 7 8
	sorting newin 3	-108592233	GONG CONTROL OF THE C	-108582173		ŧ	-46
	insulin-like growth factor binding protein θ	53491414	gcqqctacttaagacaqaqggCCUNCCNCGCCNCCACCTGCSCCNCGACCNCTGGGAA	53491474		12	176
	hypoxembline phosphoribosyloransferase 1	133594137	ääddeädädecidedädaaarèdedääääräääesäadeengaggeongalainettettyy	133594197		X	· •ve

ZNF460

Derivative i	3001/03014703014876						
	Official Gene Name	Start Pos.	<u>Binding Site</u>	End Pos.	<u>Binders</u>	Chr.	Strand
Transfac	damage-specific DNA binding protein 2, 48kDa	4723 6 511	qqaqciccaaqetqqtt:GAACAAGCCC:GGUCATGTTTqqcqqqaaqisqqcttagctc	47236571	p53	11	ive
	solute carrier family 0 (solium/hydrogen exchanger), masher 2	103235959	coleccidececificededdiaedd.comagayggggiiggggiiladdaidaadaica	103236019	\$93/871	2	4A3
<u>epd</u>	peptidylprolyl isomerase P	81107156	adhchdigrcdhdidcadadcciddagocriciccaadcoaragagidcricdirdcacidc	81197216		10	fya
	SPIGE nachear annigen	231286897	TO 161-CHROOD ACT FOR THE TROODUST WORK ACCAMANT ON A DECREASE CHROCEBA	231283957		2	÷Vē
	CKNA-like MARWEL prenemembrane domain containing?	32433326	CCCCCP_CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	32433386		3	tve
	guanine nuclectide binding protein (G protein), beta polypaptide 3	6950069	CONTRACTOR	6950129		12	₹Vē
	8100 calcium binding protein B	-48025065	ggctgeaecaggytteateeateetectctccccGGGAGAGAGAAAAAAGAGGCGGCCGCCC	-48025005		21	-76

2017480

Derivative 2	CALCESCA LAGORICA COM					
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr. Strand
Transfac	core					
EPD	none					

ZNF480

Derivative 3	Secretaria Cara						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
<u>Transfac</u>	v-Ha-ras Harvey rat sercoma viral oncogene homolog	-534789	şşijarigeaggaigdeariACGC18UUCCEUXACSCCGiggegeyeicitgicgiggesaq	-334729	p53 iso1	11	-ve
	ornithine decarboxylase 1	-10588199	ecggggegggesgagggaccag60000000000AUGtGagggggagcescgataggatage	-10588139	AP2alpha/AP2 /Max isc2:c-	2	-v∈
	anti-Mullerian hormone	2249139	aşqcancecqceqqağlqqqqqqcngcqqccqcCAAGGYCGoqqcaqaqqa	2249249	SFI	19	tye
<u>epd</u>	mannan-binding lectin serine peptidase 1 174/01 artivating component of Ra-reactive factor)	-107039521	MAGRIPAGO A CARGO CARGO CONTRACA CONTRACA CONTRACA A CONTRACA CONT	-197009461		3	-VE
	stanmicoaldin Z	-172755450	Tantigagolaccatoro. Goda motosgolacioscoporgotaaccacctostca	-172755390		2	-ve
	electron-transfer-flavoptotein, alpha polypeptide	-76603805	TO A CHESC TO A DESCRIPTION OF THE SECOND STATE OF THE SECOND STAT	-76603745		15	~V4
	surfeit (~136242964	CASCOMACONGAGONGCCOGGCCOCCAGONACTICCTGTGGGGGCCCCAGCGGGTGCG	-136242904	_	9	-ve

ZNE480

parena a araacepena Parena Araace Perena					ļ
Official Gene Name	Start Pos	Sinding Site	<u>En</u>	d Pos. Binders Chr.	. Strand
sone					İ
					i
EQRE					
	Official Gene Name	Official Gene Name Start Pos.	Official Gene Name Start Pos. Binding Site none	Official Cene Name Start Pos. Sinding Site En	Official Gene Name Start Pos. Binding Site End Pos. Binders Chr.

2NF480

2NF480							
Cerivative 5	3:350EGT.VOTCAA320G						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Strand
Trapafac	dystachich.nerbeta-synthase	~44496277	coedicierésidireticaguage discussivité accodédrorte latriasors es a	-44496217	%Z: -:	21	-va
	peroxisome proliferator-activated receptor garma	12329218	crocodacéndedecorgarácide@C@gygynggggdata@docafadacalacecoraces@	12329276	EGR-1	3	176
	cyclin-dependent kiname inhibitor 2A (melanoma, pl6, inhibits OBMA)	-91994483	ngyagong stCACoT atCGTA: tAaa yagogg: yoagonyangongagonaygonon yya	-21994343	79X2/T9X3	9	-7e
	lmn0 (Rap40) homolog, subfamily 3, mamber)	-14628355	AccodderacadadideadqidaaaAabadaqadagagaaqadagaaqadada	-14629295	NE- y	19	-va
	presenilin l	73603247	ctdedcracdedccdcddddaaaeddddcaacagggAydidracdidcaideddaaaddda	73603307	E	14	+7e
	priosephosphate isomerase 1	6974539	as cash plear our odd deld odd show the constraint of the constr	6378579	SPl	12	+45
	cyclin-dependent kinese inhibitor 1A (p21, Cip1)	36646370	$\delta cc + \delta \delta d cc cc \delta 3 \delta \delta a g \chi g \chi cc cc \delta \delta \delta c d \delta c \delta \delta \delta \tau d \delta \delta \delta a c \delta a \delta \delta \delta \delta \tau c c c \delta c \delta \tau d c \delta \delta \delta \delta \tau c c c \delta \delta \delta \delta \tau d c \delta \delta \delta \delta \delta \tau c c c \delta \delta \delta \delta \delta \tau d c \delta	36646430	GKUF	6	≠va
	coll-like receptor adaptor molecule 1	-4831796	ccccacaccccaczcqqgacqcggAAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	-4931736	601	19	-ve
	cystathionine-beta-synthase	-44496277	comproquaryagamayacassayshahagassassayayacooqaytootqoqacoqmacommag	-44496217	203/301	21	-78
	hese shock TOKO's protein IA	31784621	cascondacjaggolgogycolands/WWWWggojonggoonlooupalgyogga	31784701	192		1 V-#
	cyclin-dependent kinase inhibitor 1A (p21, Cigl)	36646370	anoligaaccccaacyeacanayu con kacamayodd adaada adaacaa daacacaa	36646433	2457.648		+46
	cyclin-dependent kinase inhibitor 20 (pid, inhibita CCR4)	51434302	USCCACCAGGGCCCAGGCCCCCCCCCCCCCCCCCCCCCCC	51434362	E2F-I:DP-1	ı	+ve
	cyslin-dependent kinase inhibitor 20 (p18, inhibits	51434302	ogcoagsa0000090a0000000000000000000000000000	51434362	SP3/SP;	1	17 0
	CDK4: pygopus homolog 2 (Drosephila)	-194934260	Cast tottest tentakeateskingestätestäteistainestätaladestainines	-154934200	Elf-1/Elk-1	1	-76
	cAMP reaponsive element binding protein I	108394623	igageggeageigeggeactoggaaci%SCGGageiggaiggaigseiggeigeggaice	208394683	SP1	2	}ye
	insulin receptor substrace l	-227664147	atgggggcogoagaggatgatttttttttttttttttttt	-227654083	tone	2	-ve
	superoxide disputase 2, mitochondrial	-160114385	de:3000-cacded03.48000-cacd-cacquara	-160114325	SP3/SP1	-	-ve
	survival of motor neuron 1, melazeric	70220339	ogiscofffeddwy.ggggaracad margu.yg.gg.gcorrapigcgnarciigagaagggarg	70225999	ELS/PEA1	4	+45
	survival of motor neuron 1, telemente	70220839	carecrarreeds.LCCCVRysEccocedSmaschdamaccarrescentsveeds.aca	70220899	1L-6	5	178
EPD	plasminogen scrivator, proximare	75670812	OGCAGGSAGEOGSGCUGGNGAGUGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	75678872		10	+ve
	transforming growth factor, bena-induced, 68kDa	135364634	DESTENDANT TAARTISSECUSSESSENERASSESTITEACTTEES, SEASSES	, 135364694		. 5	·ve
	matrix metallopercidase [1] (coronelysin []	24114984	gonggggggggaaagonataaggggggggggggggggggg	24115044		22	178
	cysteinyl-thNA synthecase	-3678691	gggcccggcgcgggggcGACTCCCGGGGCGCTCCCTCCACAAT.C.ACGAAGTGTCT	-3078631		11	~7a
	DiStorge syndrome oritical region dene 2	-19109988	ctacgosaggaggaggaggacass (nocoocaccassAAccooAccAasAAccacca	-19109928		22	1.6
	ornithine decarpoxylese antizyme 1	2269473	geniggegeagearsiacaaggeggeggeagagggageaga;T:TGCMAACGCG	2263533		19	186
	phospholipase AZ, group KV	08279287	annggeegggggangeggggyktgegggggggggggtsAAGDUGGGGGACCCCCC	65279267		36	149
	acidic (leucine-rich) nuclear phosphopoctein 30 family, member 9	139745558	manadosantacedancedancedanadosantacedanadosantacedosantacedos com a company de la comp	108745618		9	. tya
	enclase 1, (alpha)	-853878S	gestinaciasonagaagicosocopologaacastatictstokstaccosgas	-8939725		1	-va
	Gardner-Resheed Teline sarcoma mixel (v-fgr) oncogene homolog	-27361711	CTTREATTOCTTCCTCCCTGCCCAGEAGGAGERERERERERERERERERERERERERERERER	-27961650		. 1	-ve
	calponin 3, acidis	-95332709	SEGCORIU, S INVERNACURARITUS ARTERIORIS ACCIRCORGOS STOUTTO CARC	-95392649		1	-ve
	perilipin 2	-19127562	GEOACGEGO FECCIA SECE EGOGGAG POLICAR DAGGEGOGGET. ATAGGE PEGGILOPEC	-19127682		9	-78
	sugeroxide dismutese 2, mitochondrial	-160114318	OC POCCOGNUCT VICTEAACHDOGONNOWN JOACHANIGONACHTONOCTOTEST O	-160114258		ŧ	*74
	PABTA, member RAS oncogene family	128444963	cgtctcgtgacaggtACTTCCSCTCGGGS;;cccGCSGTGGCSGAAGTGGGAGCSGCCCT	128445923		3	+V4
	cylidine sonaphosphate Neutelylneurasinic adid aynthetase	22139136	cqcqqqccagccqccaggqqqqCATCCRNLHGCSHYCDAGCTGAGCGGTGAGCGACCTAGC	22199196		12	•ve
	angulin-like growth factor binding protein ?	-57976618	TST-SST-TOTOGOGGA/AGGCU-SGGGG XSGGG-TGGGTTT-AAGGGSGCGGGGG-SGCCUGACA	-57376358		4	-46
	serino/threonine winase 25	-242447755	aBeddwadurdccrridwadeddrddlaincadurwedddaddddaddcdedddadw	-242447695		2	ve
	celrectculin	: 31.49382	ggcqqqcqgqcataaaqtqcaqqcqqqcqCCCCCCCCCCCCCCCCCCCC	13043442		. 19	tve
	Selba-like i homolog (Eromoghila)	101193180	ccacyaqqagyqaqacqqtacgaaaagqgeyqcqqqqqqqqqqqqqqqqqqayqtccccq	301193246		14	• • • • •
	mcesin	64887476	qoqqaaqooqqqoacataaaqqaqoqqqoqqcqqqqAQAQ;;;;;;;:::::::::::::::::::::::::	64897536		X.	+76
	insulin-like growth factor binding protein 6	53491335	orgacoggrygorarttaagaragagyyseks/19660konA60A60,000609666A6960	53491399		12	974
	amyloid beta (A4) precursor-like protein 2	125939741	Cast Peoget Poststaraacoecosessa kassossaacpoks Tivasai septon 988	120939801		11	tva
	bean shore TORDS promein F (mornalis)	-137511199	FOUADTATOUTE VOGTUAGGENKOEGENING PROUBLYCHOAGEGGAAGAGEGGGEUT OTGEG	-137911139		5	-ve
•	cysteine-rich protein 2	108941113	acagerqqqcaggeqqqctt00ccqcck, corcoccqccbcqA66A6AAC066cq6A66	105941173		14	*ve
	RNA binding protein SI, sering-rich domain	-2318579	adiodrinandress adeportation to a section decrease decreases	-2316019		16	-79
	guanthe spoisotide binding protein (G protein), years	93551319	ARBITACIACTORAL DUTCHROUSE PRIBECUTERRICA DISTUTERA, TRIT. FORGAC	93551379		7	*78
	ll phosphifrustokinase, platele:	3109672					
			áczódálczdláráládásácádásíkólsiktókrzccásársadácisácásásásaá	3109732		10	+ve
	mbiquitin-like 46	-153715314	cadedd Leddi i sewiedd Jedech kild hef ar Sewiedd Laddi radael d	-153754958		x	-ve

OBVENZ

						SANGARA CON CARACAS	9 Sylisyliso
basriè	Cpr.	<u>Binders</u>	End Pos.	esis paibais	Start Pos.	SEEN GOBO LELLIZO	
₩	9	¥/5/3/1373	BABETTTEI-	ეგებინაშიმინაშიინაშეეგეგეგების მმენაშინაშინაშინაშიმინაშიმინაში	904611461-	č sesniž sesniž sesniž nistoroj batsvitos-nagonic	<u>Dalanerí</u>
445	í	146.642	19201661	eposésébbsétabbábábáságggggggggggabábássssságábbábábábábataáb	Telcuell	upun (dapad	
9A-	50	148	07706662-	BecopperExtrapolicationerations, many et al. the contribution of t	38768662-	ná (uba ec ámo sido	
446	ζ		21661513	OUT THE TOP TO THE TANKE TO THE TENNESS OF THE TRANSPORT	53 6 27 4 86	i nierarg gribnid torgeren teeloon	CG3
ĐA+	ij		68669495	ATTORACOTENTO PORTINA MENTANTA PROPERTIES AND PROPERTIES OF CONSIDERATION AND PROPERTIES AND PRO	2636959	(SELiziveter .4) 2 polemon 120AN	
-A 5	; ;		-110654294	33334c36c3c3c3c46c3addqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqqq	159779011-	esebingeq enfereyn benefer-kiepongege (è eseges	
974	Ş		162302642	ენი და გენი განები გენი გენი გენები გენი გენები გენი გენები გენი გენი გენი გენი გენი გენი გენი გენ	762932652	sced (Il esavelensvilvaches enincidrem	

0813H2

						WIND OF THU STEERS OF THE STEE	Serivative ?
Strand	CDI.	erabniñ	End Pos.	etis paibnis	Start Pos.	Official Gene Wang	
ēA~	ŧ	748	ZV99661Z-	A6666 de	ZOSPOETZ-	,81q ,amcasiam) AS solididmi aesaik imahasqab-milingo (PAMO alididmi	<u> 263enezî</u>
4A5	12		35442332	CLONG SENSON CLOCOMER MICCOCCO TRODUCCIÓN ES LO RECESSO (RECES LE	33745232	l eastouber lynadiso	<u>043</u>
₹Aŧ	3		66927308	ibezőssztasádőberődőridőtőfrásátőbezőssőssássátasádsszesesses	51361392	enqie "(niesosią 9) miesosyą gnibaid ebstoca zakaesą S ebstogaąyską ytrostoca gnicididai	
ēA-	62		96651209-	ASAMEN ET TIEGETEN ET PROCESSE GEGES GEGES PROCESSE GAS GEGES PROCESSE GAS	92481709-	Figure digis (cinedus (misquader (empedate) empsegar	
8A-	22		7059ET6Z-	orgerocordescondinocorrescontributarios de la pagada de pagada de la p	99596162-	i niezorą taibnia kod-k	
£Y-	7		22271896-	A2000012A201010000000000000000000000000	21991896-	gministnes mismob (CAATE) relament bigil bessier-SATE	
5V~	£		£18928581-	CANDINGED AND ENGINEER OF THE LEGISLAND OF LOWER OF THE MANAGER	E18928581-	i insitav ale	
ēA-	X .		82671LESI-	39:38c3c3c3c36:3836:36:36:36:36:36:36:36:38:38:38:38:38:38:38:38:38:38:38:38:38:	996714551-	Ah skil-minimpido	

ZNP480

lerivative E	Commence of the control of the contr						
	Official Gene Name	Start Pos.	Binding Site	End Pos.	Binders	Chr.	Stran
Transfac	v-kit Bardy-Zuckernan 4 feline sarocza viral onzogene homolog	51723521	geocoageotggcoquocaceccomagggat&AAAAQquggagagagaaaqggggcoogg	55523885	c-bf5-2	4	445
	glycophoric ? (Gerbish blosi group)	127413590	cgggtqcgtgttcaaaasaaaaaaGWNNACGCSAGccagccaagtctqcagasa	127413650	sone	2	+45
	erythropoietin	100321394	pospospycocapycoppgyAAACGACKW.GCAGGGCCCgggccccacgcgcqccgcc	120321454	rere	7	+ve
	<pre>gonedotropin-releasing formine 1 (lateinizing- releasing hormone)</pre>	-25282679	aalaacacgticacggrigcaciiciinggciggaaChTCTATAAAacttagalaaigata	-25282419	203372	ě	-43
	SW1/SSW related, matrix associated, actin dependent segulator of chrobatin, subfamily a, member 1	-128657605	ddadiciodalaicdirgyygggggggggggglyddaddaddddadadddadaa	-125657545	ŠP.	X	-45
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase i	120049929	ggacacaGAACGTTTGCAAAAAAAATGGGGTgaggaaaatgaggcatctgtqtgagtat	120049989	alpha- enslase	1	+ve
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isoperase 2	119957868	ggggteatggastttttgtAAAAASUN9GJggsggassataaggeatstgetgagtgta	119957920	alpha- enclase	. 1	156
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 2	119957858	qqqqtcacqqaattettqtaaXAAT%Cqqtqqaqqaaataaqqcatctqctqaqtqta	119957928	YY)	1	+78
	hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-issmerase :	120049929	ggasacagasigittysaaaaaaaat%%%%ggaaasigaggsatsigiqigagsat	120049959	SP3/SP1	1	: +v
	hydroxy-delta-5-steroid dehydrogenace, 3 beta- and steroid delta-isomerase 2	119957868	qqqquanqqaatttiqtassasaftXXXXXXqqaasataqqqatttqqqqqta	119957928	198/698	. 1	+44
	serpin pepticase inhibitur, diade A (alpha-l antiproteinase, antitrypsin), member 5	95047677	gggalqiccashkkkilkkkkcaqqqqayyygiggaqacagyyicggalqiggilaqigaca	95047727	SP1	14	tVe
	backlowiral IAP respect-containing 5	76210296	codecteracrosomémagacodedalego;@dacadecturdaledadadedeceda	76210236	81232	17	+34
	suppressor of cytakine signaling :	-10319749	odássedátű.CCCMAGEMydsádásoáálósoda Adodsecdodssádosadádássesec	-11350752	State	16	-46
	suppressiv of cytokine signaling 1	-11350743	ecdesealauusgekknyyerddasebblandladseselandestaddassese	-11350793	c-£TS-1/2	. 16	-V4
<u>epd</u>	calcium binding atopy-related autoentigen 1	-74385886	GGAGAG CANGTGAGAGTGGGGGGAGGAGGAGGAGGAT MUTCCGGGGGTT TUACCCGTA	-74385826		10	-V <u>e</u>
	interferon, alpha-inducible protein 27	94577613	csátátosadádosátdorroadroádádráásámrástraasadácosáddaroscac	94577073		14	+ve
	minichromosome maintenance complex component 3	-52149562	OGGEAGAGOTOGGAAGUTT TOPOGGUESIN, GGAGTOATEDEGGEAGOTT (ACGCGAGC	~52149532		٤	-48
	AMAI, activator of heat shock 90k0a groteir Alfaze homolog I (yeas:)	77924364	qaqqqqcccaaactiaaceaccccaccccccccccaccacccacccccccccc	77924424		14	+94
	neural precursor cell expressed, developmentally down- regulated 8	-24701629	totejaaceceagtoggggggggggggggggagdcagacagtgacceggaACTACAAG	-24701569		. 14	-Y
	Fas (IN) ASY6) -associated via death domain	70049423	GAATCAGGCACCGGACCGCAGGTTCGGRESCEGAATCCCTGGGCCGCCGGGAAGCGGREG	70049483		11	: •ve

2NF480

Derivative 3	ALCONTOCONTAINS NACC ADSCINE SELECTION OF SECON						-
	Official Gene Mame	Start Pos.	Binding Site	End Pos.	Binders	<u>Chr.</u>	<u>Strand</u>
Transfac	surfactant protain 8	-85895453	cijojaalagijggarcaagcacif@Aktggstciicagagcaalacaaaacigagj	-85895393	775-1	2	-VB
-	surfaciant protein 8	-95895453	crigiaesagildadacesadosocilda/Clèstot.CrVPadossalacesaracidadi	-85835393	Tr-1	. 2	-ve
	LIM homesbox 3	-139096361	cārtābāsayrestālādābartālijimielējā, gyprašarppyydcadādlādādscale	-139096301	NY-1/ CTE	9	-ve
-	prepronociceptin	26174704	totgrcatggaggtgggcagggaaggswACScrtgrgcgtgggggagaggacararatgc	28174764	SPI	3	tve
	protesse, serine, 50	-46759548	and world decoded and rate added some restanded added to a	-467594488	SP:	3	-ve
-							
EPD	secreted phosphoprotein l	35696920	NEOTOTOLICAREANT HANTOTOLICACIONETTOTOLICACIACIARIA RANCARRA	89896980		4	+75
:	interferon induced transmerizane protein 2 (1-80)	308107	annen er en	308177		11	tya.

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 ddH20 (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^oC. The supernatant was discarded and the pellets were resuspended in ddH2O 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-CI (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°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 ZnSO4. The rapid dilution was performed under constant stirring at 4°C. Once the samples were

completely added, the entire refolded content was allowed to stir at 4^oC 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

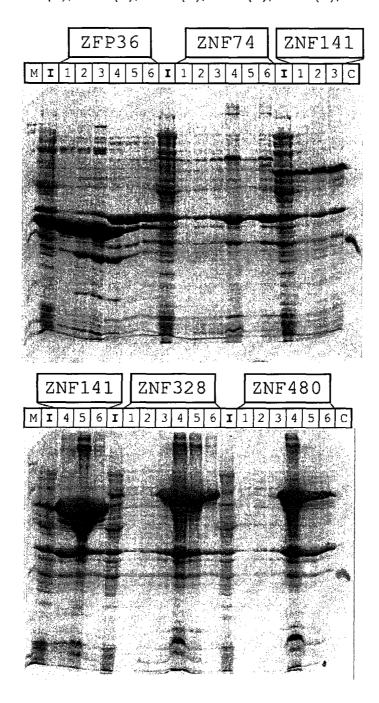
Purification and refolding was also attempted using DEAE ion-exchange chromatography. It has been reported that ion-exchange chromatography can refold proteins without the presence of any other factors (Chen et al., 2009).

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.

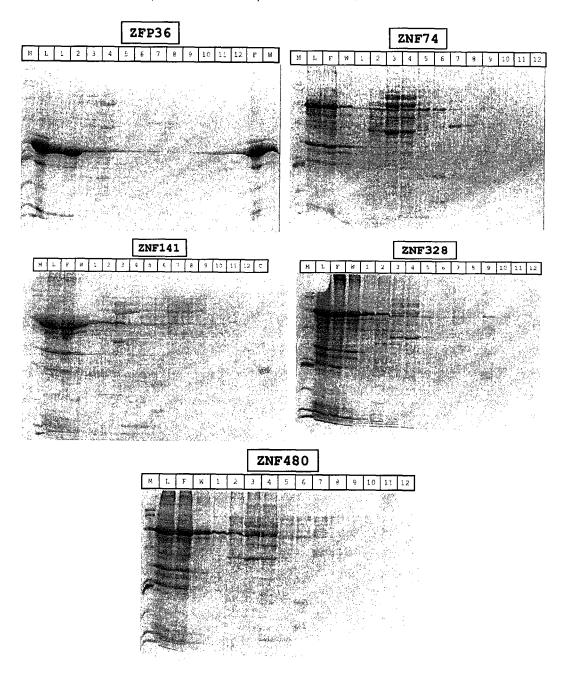
GSH Affinity Purification of In Vitro-Refolded Target Proteins

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.

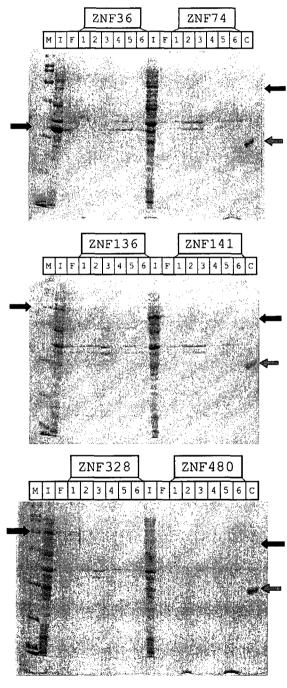
Appendix B- Inclusion Body Solubilization Assay. Each inclusion body sample was solubilized in 2, 5, and 8M Urea. The supernatant (S) and pellet (P) fractions from each sample were analyzed to determine the level of solubilization. 1-2M(S), 2-5M(S), 3-8M(S), 4-2M(P), 5-5M(P), 6-8M(P).



<u>Appendix C-</u> 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.



<u>Appendix D-</u> 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 hlL-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.