MOLECULAR CHARACTERIZATION OF THE RABBIT HKα2 GENE

By

DEBORAH MILON ZIES

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2003

This dissertation is dedicated to the memory of James Byrne McCracken, Jr. (July 19, 1967 to January 13, 1999). Jim's all-too-short-of-a-life has been, and always will be, an inspiration to continue on, even in the toughest of times.

ACKNOWLEDGEMENTS

The work described in this dissertation could not have been completed without the support that I have received on both a professional and a personal level. I would therefore like to acknowledge specific individuals who have been there for me over the past five years.

Professionally, the most significant support came from my mentor, Dr. Brian Cain. He taught me to be an excellent scientist, an outstanding writer, and a supportive colleague and friend. I thank my committee members, Drs. Charles Wingo, Thomas Yang, Peggy Wallace, Sarah Chesrown, and Bruce Stevens, for their suggestions and support throughout this process. Their contributions are evident in both this dissertation and the scientist I have become. I thank the past and present members of the Cain laboratory. I owe a great deal to those people who were members of the lab when I joined, Drs. Paul Sorgen, James Gardner and Tamara Caviston Otto. They all helped make the transition into a new laboratory smooth and fun. In particular, Tammy remained in the the lab long enough to become a true friend. To this day, she continues to offer understanding and encouragement when times are tough. I owe even more to the current members of the Cain laboratory, Michelle Gumz and Tammy Bohannon Grabar. We three have been through so much on every level that there are no words to express my gratitude. I would also like to thank past and present members of our neighbors, the Yang laboratory. These people have made the work environment both stimulating and enjoyable. They are Christine Mione, Sara Jato-Rodriguez, Chien Chen and Sue Lee

Kang. Specifically, I would like to thank Sue for all her help on completing the DNaseI hypersensitivity assay and Christine for her unending patience and guidance while teaching so many things about the computer. Finally, on a professional level, I thank the rest of the Department of Biochemistry and Molecular Biology. The entire group has made my experience as a graduate student one that I will never forget.

On a personal level, the first people I would like to thank have already been mentioned. In addition to all they provided on a professional level, Michelle and Tammy have also provided emotional support and friendship. The rest of my thanks goes to my family. I thank my dad, Frank Milon III, for setting a great example for me by going to law school, and becoming an attorney, later in his life. I thank my mom, Barbara Milon, for being a mom, a cheerleader, a babysitter, a friend, and so much more. The rest of my family, Patricia Milon, Frank Milon IV, Christina Milon, Rebecca Milon, and Joshua Milon, I thank for all of their encouragement, help, and reminders of what is really most important in life. And finally, because she is the most important, I would like to thank my daughter, Lee Ann Zies, just for being her. As a single parent with a five-year-old daughter, it may not have been possible to attend graduate school and make such a big change in my own life. Lee Ann, however, has not only made it possible, she has made it all worthwhile. She is a blessing beyond her years. She has been happy to go the lab, understanding when we have had to be apart, considerate when I have had to study, and loving in every difficult moment. I could not have asked for a better companion. She makes everything worthwhile. I love her and thank her with all my heart.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MOLECULAR CHARACTERIZATION OF THE RABBIT HKα2 GENE

By

Deborah Milon Zies

December 2003

Chair: Brian D. Cain
Major Department: Biochemistry and Molecular Biology

Blood potassium concentration is critical for normal cellular functions. Failure to maintain blood potassium levels within a very narrow range can lead to hypertension, cardiac arrest, respiratory failure and death. The kidney is the major organ responsible for maintaining a constant potassium level despite variations in dietary intake. One group of proteins the kidney employs to carry out this function is the H^+, K^+ - ATPases. These ion transporters use the energy of ATP to absorb potassium ions from the nephron lumen in exchange for hydrogen ions. Evidence suggests that the levels and the activities of this group of transporters are responsive to potassium, sodium and possibly hormones. The rabbit HKα2 gene produces two mRNAs encoding the $HK\alpha_{2a}$ and $HK\alpha_{2c}$ isoforms of the alpha subunit of the renal H^+, K^+ - ATPase. In order to study the regulation of these isoforms, it was first necessary to clone the gene. Three genomic clones (HKα2.1, HKα2.5, and HKα2.8) were identified by screening a lambda genomic library. These three clones span a majority of the HKα2 gene. Subcloning and high-throughput

sequencing of the clone HKα2.5 provided information regarding gene structure and its similarities to the human, mouse, and rat homologs. RNase protection assays mapped two distinct transcription start sites that correspond to $HK\alpha_{2a}$ and $HK\alpha_{2c}$ transcripts. Sequence analysis of the 5' clone, HKα2.1, suggested that this clone contained the gene promoter and 5' regulatory elements. Subsequently, a luciferase reporter gene approach was used to perform a deletion analysis of the region upstream of the transcription start sites. The promoter and two putative negative regulatory elements were identified. Furthermore, sequence alignments and mutational analysis provide evidence for a functional TATA-like element within the promoter region. Finally, RNA and protein analyses suggested that RCCT28A tissue culture cells must be differentiated in order to express the HKα2 gene. Taken together, these data constitute the genomic organization of the rabbit HKα2 gene, the initial characterization of the promoter, the first evidence of cell differentiation as a signal for HKα2 gene expression,and a system for future studies of regulation of the HKα2 gene at the molecular level.

CHAPTER 1
INTRODUCTION

The maintenance of extracellular and intracellular potassium ion (K^+) concentrations is critical for normal cell functioning. The difference between these two concentrations, along with the concentrations of other ions, creates an electrochemical gradient that is known as membrane potential. When the extracellular concentration becomes too low, or too high, the membrane potential is altered and serious side effects occur. In order to maintain a constant blood K^+ level the major organ employed by the body is the kidney. The H^+, K^+ - ATPases comprise one class of proteins found in the collecting duct of the kidney. These proteins use the energy of ATP to bring K^+ from the nephron filtrate into the cell in exchange for hydrogen ions (H^+). This function is thought to play a key role in K^+ conservation. In the rabbit kidney, three isoforms of the alpha (α) subunit of the H^+, K^+ - ATPase have been identified. They are $HK\alpha_1$, the gastric isoform, $HK\alpha_{2a}$, the colonic isoform, and $HK\alpha_{2c}$, a novel splice variant of $HK\alpha_{2a}$ (53). Evidence suggests that when blood K^+ is low, there is little to no change in transcription from the $HK\alpha_1$ gene. On the other hand, transcription from the $HK\alpha_2$ gene appears increase under the same conditions (3). The purpose of our study was to characterize the rabbit $HK\alpha_2$ gene and initiate studies on its regulation. The specific aims were (1) to clone the $HK\alpha_2$ gene from rabbit, (2) to map the transcription start sites for $HK\alpha_{2a}$ and $HK\alpha_{2c}$, (3) to perform promoter deletion analysis of the region 5' of the transcription start sites and (4) to determine the effect of cell differentiation of HKα2 gene expression.

This chapter will review background information on the physiological significance of the H^+, K^+ - ATPases, the evidence for the presence of the H^+, K^+ - ATPases in the kidney, the evidence for regulation of the HKα2 gene products under a variety of cellular conditions, the structure and function of H^+, K^+ - ATPases, and the organization of the known HKα2 cDNAs and genes.

Physiological Significance

Healthy individuals maintain a blood K^+ level between 3.5 and 5.5 milliequivalents per liter (mE/L) (1). If blood K^+ concentration drops below 3.5mE/L (hypokalemia) or if blood K^+ concentration increases above 5.5mE/L (hyperkalemia), serious side effects, and even death, can occur. This section discusses the importance of K^+ concentration and the problems that are associated with low and high blood potassium levels.

Membrane Potential

In a typical cell, the concentration of K^+ inside the cell is much greater (approximately 150 mE/L) than the concentration in the extracellular space and in the blood (approximately 5 mE/L). This concentration difference, along with that of other ions, such as sodium and chloride, creates a slight charge to the plasma membrane known as membrane potential. The charge in an animal cell typically ranges from –50 to –100 millivolts (4). This charge is important for a variety of cell functions including ion transport and cell signaling. In excitable cells, muscle cells and neurons, the membrane potential is absolutely required for stimulation. In a resting state, sodium ion concentration inside the cell is low (15mE/L) while potassium ion concentration is high (150mE/L). When the cell receives a signal, Na^+ channels open and allow Na^+ into the

cell. The influx of Na^+ depolarizes the membrane and the membrane potential becomes temporarily positive (+50mV). The change in membrane potential triggers K^+ channels to open and allow K^+ to exit the cell. The efflux of K^+ causes the membrane to repolarize and the membrane potential once again becomes negative. These local changes in membrane potential are propogated across the entire cell and cause a muscle cell to contract or a nerve cell to fire a signal. K^+ concentration plays a key role in this process, and an imbalance of K^+ concentration leads to some of the side effects associated with hypokalemia and hyperkalemia.

Hypokalemia

Most individuals are capable of maintaining proper K^+ levels. In fact, hypokalemia occurs in less than 1% of the healthy population (1). There are two main groups of people, however, that are very susceptible to developing hypokalemia. There are individuals that suffer primarily from other disease states and acquire hypokalemia as a secondary effect, and there are individuals who acquire hypokalemia as a side effect of taking medications. The most common occurrence of hypokalemia is among patients receiving diuretics, as many as 50% of these patients develop low blood K^+. The second largest group of individuals that develop hypokalemia suffer from hyperaldosteronism associated with heart failure and hepatic insufficiencies. A third significant group of individuals that suffer from hypokalemia are those with renal diseases that effect potassium uptake.

The most predominant effects of hypokalemia are hypertension, muscle weakness, and metabolic alkalosis. Blood volume, and therefore blood pressure, is associated with ion balance in the blood. It appears that low blood K^+ levels lead to increased sodium retention thereby upsetting the normal blood ion balance and leading to

hypertension. Even very mild hypokalemia, a serum potassium level of 3.4 mEq/L, can lead to increased blood pressure and risk for cardiovascular disease (50). When blood K^+ is low, muscle cells can become hyperpolarized, or more negative. The hyperpolarized cell is difficult to depolarize, interfering with the ability of the cell to contract. In cardiac muscle this condition can lead to ventricular arrhythmias. In skeletal muscle, hypokalemia frequently leads to muscle weakness, cramping, and in severe cases, paralysis (1). Hypokalemia also has a profound effect of blood acid/base balance. Alkalosis, high blood pH, results from two main kidney activities. The first is the stimulation of proximal tubule bicarbonate absorption and proton excretion. The second is the increased activity of the H^+, K^+ - ATPases. These complexes exchange K^+ for H^+ eventually leading to acidification of the urine and alkalosis in the blood. A failure to correct metabolic alkalosis can lead to additional cardiac and skeletal muscle weakness as well as liver and brain damage.

Currently, the treatment for hypokalemia consists of oral or intravenous replacement of lost K^+. Unfortunately compliance with potassium supplements is low due to its disagreeable taste and/or the inconvenience of intravenous replacement (50). Additionally, there are cases where potassium supplementation does not result in the desired increase in blood potassium levels (Wingo, personal communication). Finally, overaggressive potassium replacement therapy can lead to hyperkalemia and its severe side effects (see below). An understanding of the regulation of the H^+, K^+ - ATPases may lead to better mechanisms for controlling blood K^+ levels.

Hyperkalemia

Hyperkalemia can result from as a disruption of the normal intracellular and extracellular K^+ concentrations or from a disruption in the balance between K^+ intake and

K^+ excretion (49). The kidney usually has an extraordinary capacity for excreting excess K^+. Therefore, hyperkalemia usually only occurs when there is a combination of increased K^+ uptake or a redistribution of cellular K^+ along with decreased renal efficiency. The most common cause for excess intake of K^+ is its presence in supplements and salt substitutes. The most common causes for a redistribution of K+ include acidosis, anesthetics and hypertonicity. Interestingly, in the case of acidosis, one of the contributing factors to increased blood K^+ is the H^+, K^+ - ATPase found in the collecting duct of the kidney. The H^+, K^+ - ATPase functions in combination with other cellular proteins to remove H^+ from the blood. As part of this process, however, the H^+ is exchanged for K^+, causing an increase blood K^+. Reduced renal efficiency can be caused by a variety of factors including reduction in the number of nephrons due to renal failure, medications, and hormone imbalance (43).

The major adverse effect of hyperkalemia is on the cellular membrane potential. Increased extracellular K^+ makes the resting membrane potential more positive than normal. The result is a decrease in the abililty of the cell to propogate a signal and an increase in the rate at which the cell repolarizes. The most severe effect is in cardiac muscle where there is a delay in conduction of the muscle contraction which can lead to fatal heart arrhythmias. In skeletal and smooth muscle, increased weakness and fatigue are common. Additionally, the weakened muscles in the respiratory tract can lead to severe respiratory depression and respiratory failure.

Currently there are two approaches to treating hyperkalemia. Calcium supplements can be given as a treatment for the myocardial effects of hyperkalemia. This treatment is very rapid, and relieves the major concern associated with hyperkalemia, but

it is not a long term solution to increased blood K^+. Long term treatment of hyperkalemia can be attained by stimulation of cellular uptake of K^+ and stimulation of renal excretion of K^+. An understanding of the regulation of the H^+, K^+ - ATPase may lead to a mechanism for its inhibition thereby decreasing the uptake of K^+ in the nephron and reducing blood K^+.

Mammalian Kidney

The mammalian kidney is an important regulatory organ. By excreting or retaining water and solutes, the kidney can maintain proper blood volume and composition despite changes in diet and activity (25). Figure 1-1A is a diagram of the mammalian kidney. The kidney is divided into three sections: the cortex, the medulla and the pelvis. Running through all three regions are millions of microscopic tubules called nephrons. The nephrons are the functional units of the kidney and are depicted in Figure 1-1B. As blood capillaries enter the nephron they form the glomerulus and come in contact with the nephron at the Bowman's capsule. Blood pressure promotes the free passage of water, urea, ions and solutes from the blood into the lumen of the nephron where they become part of the filtrate. Throughout the nephron there are a variety of proteins that function to pass additional waste products into the filtrate and recover needed materials from the filtrate. The collecting duct of the nephron is the primary location for the reabsorption of K^+ when blood K^+ concentration is low. Figure 1-1C is a diagram of the cells that make up the cortical collecting duct. The three types of cells are principal cells, type A intercalated cells and type B intercalated cells (30). In all three cases, the apical membrane faces the lumen of the nephron and contains a different constellation of proteins than the basolateral membrane, which faces the blood. There is evidence to suggest that all three cell types have the H^+, K^+ - ATPase present on their

apical membrane. Principal cells make up approximately 65% of the cells of the cortical collecting duct. They appear lighter in color under the microscope because they

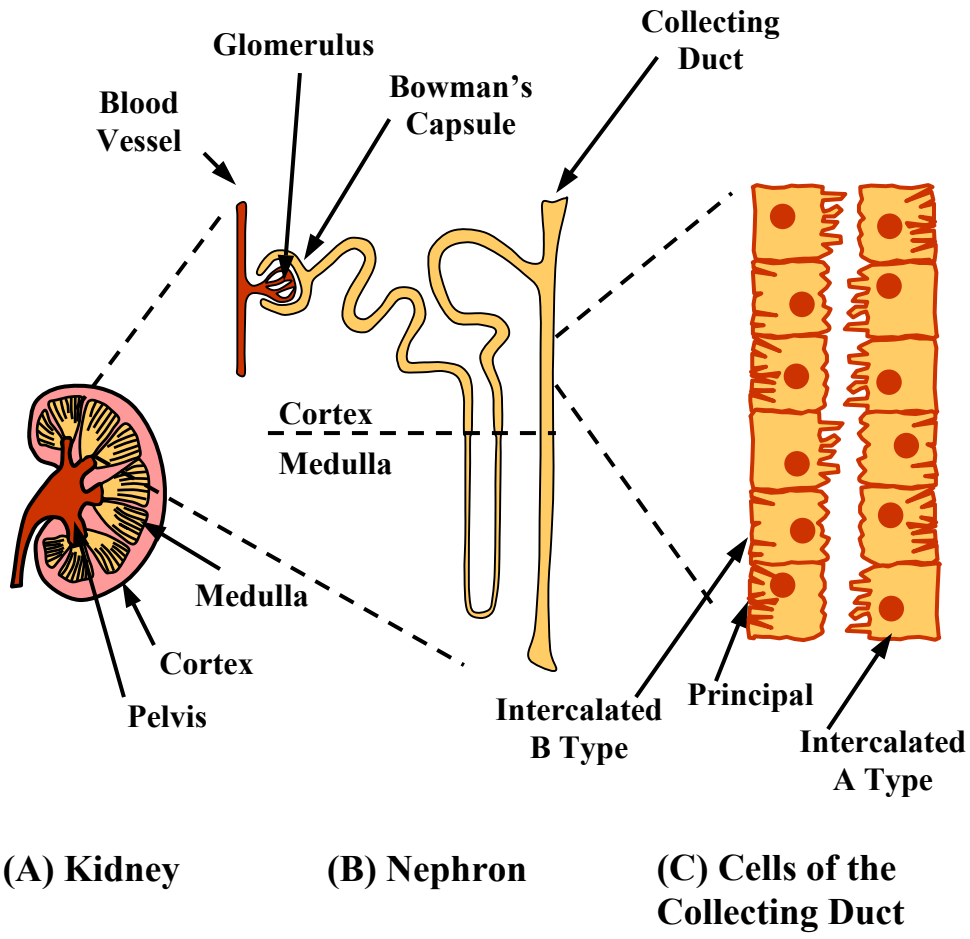


Figure 1-1. Mammalian kidney. A) Cross-section of the mammalian kidney B) The nephron C) Cortical collecting duct cells.

contain fewer mitochondria than either intercalated cell type. They have a highly folded basolateral membrane and a very smooth apical membrane. Type B intercalated cells are the second most abundant cell type, comprising about 25% of the population of cells in the cortical collecting duct. Their basolateral surface is highly folded, similar to the

principal cells. Their apical membrane, however, is more extensive, containing several microplicae. Additionally, tubulovesicular structures can be seen scattered throughout the cell. Type A intercalated cells are the least abundant, making up the last 10% of the cell population. They have extensive microplicae on the apical surface and a numerous tubulovesicular structures just below the apical surface. Their basolateral membrane, in contrast to the principal cells and B type intercalated cells, is smooth. It has been suggested that the three cell types are capable of inter-converting in response to stimuli. In fact, there is a progressive change in cell type along the collecting duct that includes cell types that are intermediate to the three main types.

Evidence for H^+, K^+ - ATPase Expression in the Collecting Duct.

The earliest evidence for the existence of H^+, K^+ - ATPase activity in the collecting duct came from studies involving the perfusion of microdissected rabbit collecting duct tissue (51). By setting up a perfused tubule system, Dr. Wingo was able to measure ion flux under a variety of controlled conditions. In this way, it became clear that the collecting duct contained an apical ATPase that had properties similar to the previously identified gastric H^+, K^+ - ATPase ($HK\alpha1$). Pharmacologically, this ATPase was sensitive to the gastric H^+, K^+ - ATPase inhibitor omeprazole and insensitive to the Na^+, K^+ - ATPase inhibitor ouabain. It differed from the $HK\alpha1$ H^+, K^+ - ATPase in that it was insensitive to the compound Schering 28080. Furthermore, removal of luminal K^+ had a profound effect on the proton secretion by the collecting duct segment of the kidney. This section discusses the molecular evidence that supports these early findings.

Protein

Immunohistochemistry has been used to localize the various isoforms of the H^+, K^+ - ATPase in human (27) and rabbit (48). Kraut et al. (27) used antibodies raised against the human HKα1, the rat HKα2, and the human ATP1AL1 (HKα2) proteins to probe human cortical and medullary collecting duct tissue. Using the HKα_1 antibody, they observed a darker staining in the intercalated cells and a lighter, but consistent staining of the principal cells in both the cortical and medullary collecting ducts. With the rat HKα_2 antibody, this group observed no staining. This result is not surprising because although the rat and human proteins (HKα_2 and ATP1AL1 respectively) are considered homologous proteins (see below), they share only about 87% amino acid identity. It is therefore likely that the antibody against rat HKα2 was unable to recognize the human ATP1AL1 protein. Finally, with the antibody to ATP1AL1, light staining of the intercalated cells and occasional staining of the principal cells in both the cortex and the medulla was observed. In rabbits, Verlander et al. (48) used an antibody raised against the HKα_{2c} isoform of the H^+, K^+ - ATPase and observed intense staining of the apical membrane of both A and B type intercalated cells and a lighter staining of apical membrane of the principal cells in both the corical and medullay collecting ducts. These two reports are consistent and support the localization of the H^+, K^+ - ATPase activity to all the cell types present in the collecting duct.

mRNA

In rat, Ahn et al. (1) used *in situ* hybridization to show that the mRNA for HKα_2 was present in the connecting tubule and intercalated cells throughout the collecting duct. These experiments were not designed to distinguish between the two alternative

transcripts found in rat, $HK\alpha_{2a}$ and $HK\alpha_{2b}$. A contrasting study by Jassier et al. (22), detected HKα1, but not HKα2, in the cortex and the medulla. Although both authors used *in situ* hybridization for the detection of HKα2, Ahn et al. suggest that the digoxigenin method used in their experiments is more sensitive technique than the ^{35}S labeling method used by Jassier et al. In a third report, Marsy et al. (33) consistently found HKα2 mRNA in the cortical collecting duct of rats using quantitative RT-PCR. The use of a different method in this report supports the presence of HKα2 in the collecting duct in rat. Furthermore, two independent groups (6, 12) were able to use 5' rapid amplification of cDNA ends (RACE) to detect $HK\alpha_2$ mRNA in samples from the cortical collecting duct of rabbit.

The cloning of cDNAs for the HKα2 subunits from human (17, 35), rat (9, 26) and rabbit (6, 12) kidney provided the strongest evidence for their expression in the kidney. Additionally, the cDNA's from rat and rabbit were the first indications that there were splice variants of the HKα2 mRNA's. The characteristics of the cDNA's indentified to date are listed in Table 1 and are discussed below.

Regulation of the Renal H^+, K^+ -ATPases

The microperfusion assays performed on rabbit kidney nephrons that led to the discovery of renal expression of the H^+, K^+ - ATPases also provided the first evidence suggesting that the activity of the enzyme was regulated under certain cellular conditions (51). Regulation by ion concentration, acid-base balance, and hormones has since been confirmed predominantly by *in vivo* studies with rat. In considering the earlier work as well as the work discussed below, is important to note that the studies measuring ATPase activity were not designed to distinguish between the activity of pumps containing $HK\alpha_1$

and HKα_2 subunits. Furthermore, the molecular mechanisms responsible for controlling these observed changes in activity have not been studied.

Potassium

In the cortical collecting duct, there is some controversy in the literature regarding the regulation of HKα2 by low K^+. Several investigators (26, 33) have been able to show that K^+ depletion results in an increase in the HKα2 message in the cortical collecting duct. In contrast, Sangan et al. (40) saw a decrease in the amount of HKα2 mRNA with K^+-depletion. The rat model in these experiments, however, was exposed to a low K^+ diet for a longer period of time than previous studies and the effects of chronic hypokalemia may differ from those of acute hypokalemia. Furthermore, Ahn et al. (1) did not observe a change in HKα_{2a} message in the cortical collecting duct. More studies must be carried out in order to clarify the regulation of HKα2 in the cortical collecting duct.

HKα2 transcripts are also present in the outer medullary collecting duct (OMCD) and appear to be subject to upregulation during K^+ restriction. RT-PCR (33), Northern analysis (26), and *in situ* hybridization (1) have all been used to demonstrate increased HKα2 gene activity. H^+, K^+ - ATPase activity measurements reported also indicated increased activity in both rat and rabbit (19, 28, 36). Studies of the type A intercalated cells (23, 1), most abundant in the medullary collecting duct, showed that HKα2 mRNA was present and that low K^+ resulted in an increase in the message.

A greater controversy centers on the presence of H^+, K^+-ATPase activity and the induction of the HKα2 gene in the inner medullary collecting duct (IMCD). HKα2 mRNA was not always detectable (23). Marsy et al. (33) reported the presence of the mRNA by RT-PCR. but she did not observe an upregulation of the message with

potassium restriction. In contrast, Kone and coworkers were able to detect HKα2 message and show an increase in the message during K^+ restriction by both *in situ* hybridization and Northern analysis (1, 26). Moreover, the Northern blot experiments were reproduced independently by Nakamura et al. (36). These investigators were also able to show an increase H^+,K^+-ATPase activity. Taken together, these data suggest that HKα2 ATPase is present in the IMCD and is likely responsive to K^+ restriction.

The observation that HKα2 gene products are likely to play a role in K^+ conservation led to the creation of an HKα2 gene knockout mouse by Meneton et al. (34). The mouse knockout had no observable defects when fed a K^+ replete diet. On the other hand, these animals developed severe hypokalemia when the animals were fed a K^+ free diet. Interestingly, the kidney of the knockout mouse was still able to reduce K^+ loss by 100-fold suggesting that other kidney proteins are capable of compensating for a loss of the HKα2 protein. The knockout mouse maintained on a K^+ free diet, however, developed a more severe case of hypokalemia than the normal mouse on the same diet. Additionally, the bulk of the *in vivo* data indicates that the HKα2 gene is upregulated when K^+ is restricted. It is therefore very likely that the colonic H^+, K^+-ATPase (HKα2) is regulated by low K+ and plays a role in K^+ conservation.

Sodium

A reduction in dietary sodium leads to several alterations including hyperaldosteronism and increased activity of the Na^+, K^+-ATPase. The Na^+, K^+-ATPase is present on the basolateral membranes of principal cells in the cortical collecting duct and functions by bringing K^+ into the cell in exchange for Na^+. As a result of this action, the principal cells must possess mechanisms to remove additional cellular potassium.

The most likely mechanism for the removal of K^+ is the opening of K^+ specific channels and/or increased activity of KCl cotransporters present on the apical membrane. It has been suggested that these conditions stimulate the intercalated cells of the collecting tubule to reabsorb potassium by use of the H^+, K^+-ATPase. Silver et al. (41) identified intercalated cells of the cortical collecting duct by BCECF fluorescence and measured their ability to recover from acid load under sodium depleted conditions. Increases in H^+ and K^+ exchange were observed that could be attributed to either HKα1 or HKα2 containing ATPases. Sangan et al. (40) used a cDNA probe and a polyclonal antibody specific for HKα2 to detect mRNA and protein from rats fed a low sodium diet. Northern and Western analyses of kidney cortex and kidney outer medulla revealed that sodium may have a slight effect on mRNA levels in the kidney cortex but had no apparent influence on the protein level. It is possible that the increase in H^+, K^+-ATPase activity that was observes may be a result of post-translational modification of the pump.

The major hormone released during sodium restriction is aldosterone. It follows, therefore, that if low sodium increases the activity of H^+, K^+-ATPase in the cortical collecting duct, aldosterone could do the same. However, aldosterone levels apparently do affect HKα2 H^+, K^+-ATPase activity. Eiam-Ong et al. (11) used adrenalectomized rats in which aldosterone was replaced at either physiological or pharmacological levels. When H^+, K^+-ATPase activity was measured in microdissected tubules there was no apparent difference between rats that had no aldosterone and those that had either physiological or pharmacological doses of aldosterone. In a similar set of experiments using adrenalectomized rats, Jaisser et al. (23) directly measured HKα2 mRNA. *In situ* hybridization demonstrated that HKα2 mRNA levels were very low in normokalemic rats

and did not increase significantly with the addition of aldosterone or dexamethasone. Interestingly, experiments by Silver et al. (41) showed that when rats on a normal diet were injected with aldosterone in order to simulate levels found during sodium restriction, H^+, K^+-ATPase activity was not increased suggesting that the low sodium induction of H^+, K^+-ATPase activity was not mediated by aldosterone.

Acid-Base

One might expect that blood pH would have a profound effect on H^+, K^+-ATPase activity in the kidney and indeed the evidence for an increase in H^+, K^+ - ATPase activity from alkalosis seems clear. In the rabbit cortical collecting duct, Northern analysis of mRNA derived from rabbits subjected to metabolic alkalosis generated a greater than four-fold increase in HKα2 mRNA (13). On the other hand, metabolic acidosis decreased (13) or had no effect (10) on the level of HKα2 mRNA. Collecting duct tubules taken from animals fed a K^+-depleted diet have 50% less bicarbonate absorption when compared with the tubules from animals fed a K^+-replete diet. At the same time, the K^+- depleted animals have a net increase in K^+ absorption suggesting that an H^+, K^+ - ATPase pump is upregulated under low K^+ conditions (30). Komatsu and Garg also reported that metabolic acidosis increases in H^+, K^+ - ATPase activity and metabolic alkalosis suppresses the same activity (14).

Structure of the H^+, K^+ - ATPase Complex

The H^+, K^+ - ATPases are considered P-type ATPases because they form a phosphorylated intermediate during the reaction cycle. Based on the information known about the structure and function of other P-type ATPases, models for the structure and

function of the HKα2 ATPases have been made. This section describes the P-type ATPases, their reaction cycle, and the models of the HKα2 H^+, K^+ - ATPases.

P-Type ATPases

The H^+, K^+ - ATPases belong to a family of proteins known as the P-type ATPases (32). This family uses the energy of ATP hydrolysis to translocate ions against their electrochemical gradient. They are distinguished from other families of ATPases by forming a phosphorylated intermediate during the reaction cycle. This phosphorylation occurs at a highly conserved aspartate within the amino acid sequence DKTG. All P-type ATPases share a core structure with highly conserved domains known as the ATP binding domain (N), the phosphorylation domain (P) and the transmembrane domain. Outside of these core domains there are several regions that define subtypes of the family. In the P_2-type ATPases, those that translocate non-heavy cations, the non-conserved domains appear to be responsible for activities such as the regulation of ATPase activity, cation specific conformational changes in the protein and proper insertion of the protein into the plasma membrane. One subclass of P_2-type ATPases is the X^+, K^+-ATPases. This family, which includes the H^+, K^+ - ATPases, contains ATPases that are made up of more that one protein subunit (32).

The H^+, K^+ - ATPase Subunits

The H^+, K^+ - ATPase complex is a heterotetramer of two alpha (α) and two beta (β) subunits (8). One of each of the subunits is diagrammed in Figure 1-2. The α subunit is approximately 115 Kilodaltons (Kd) in size and contains 10 transmembrane segments responsible for forming the channel for the translocation of ions. The α subunit also houses the ATP hydrolysis activity located in the large intracellular domain between

transmembrane segments four and five. There are four possible β subunits that may pair with HKα2 gene products. They are the gastric H^+, K^+ - ATPase β (HKβ) subunit and the Na^+,K^+ - ATPase β subunits β1, β2, and β3. It is unclear whether the HKα2 proteins pair with a specific β subunit or if they can pair with any of the four. All of the β subunits, however, share several characteristics. They are approximately 30 Kd in size and contain one transmembrane domain. They contain one large intracellular domain that has a varying number of potential glycolyslation sites. The β subunit has no catalytic activity, but appears to be important for proper positioning and insertion of the α subunit into the plasma membrane (37).

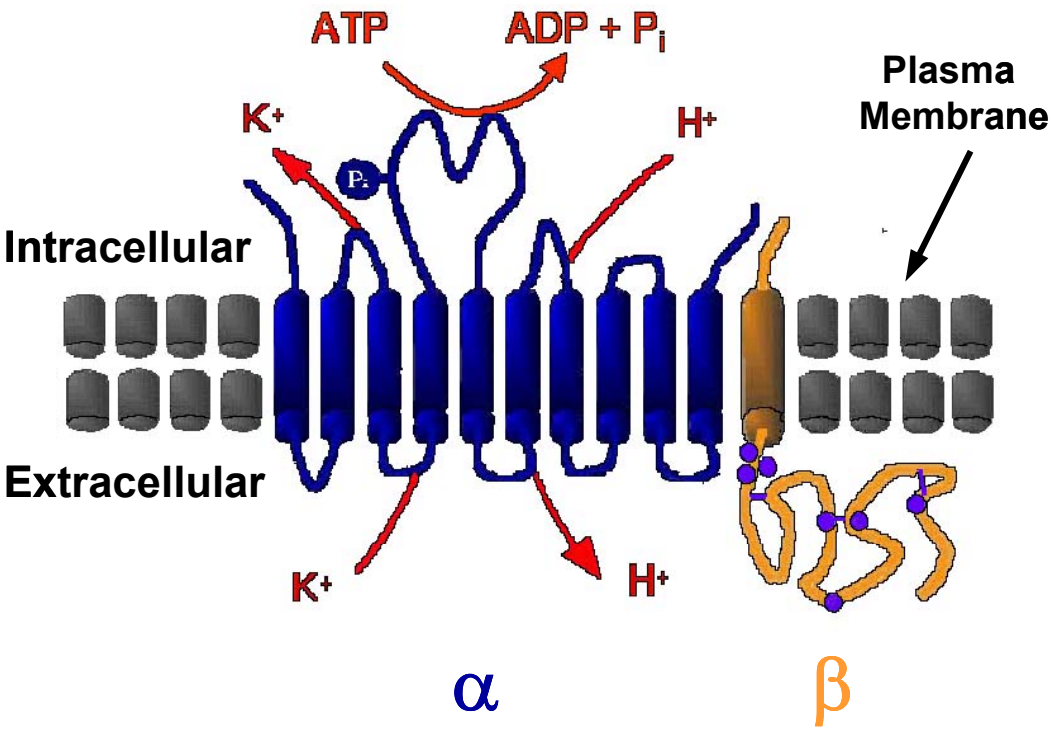


Figure 1-2. Schematic representation of the rabbit H+, K+ - ATPase. Arrows indicate direction of ion transport.

High Resolution Model of Rabbit HKα_{2a}

Recently, the high resolution structure of the E_1 state of the Ca^{++}-ATPase from rabbit sarcoplasmic reticulum was determined (46). Although this enzyme shares only 29% amino acid identity and 47% amino acid similarity with the rabbit HKα_{2a} subunit, the two enzymes are related. It was therefore possible to use the atomic coordinates for the Ca^{++}-ATPase to create a high resolution model of the H^+, K^+ - ATPase α2 subunit (18) (Figure 1-3).

In the transmembrane portion of the protein, the ten transmembrane segments are shown to form a channel for the translocation of ions. The Ca^{++}-ATPase does not have a β subunit. It was therefore not possible to model the β subunit into the H^+, K^+ - ATPase structure. Biochemical and low resolution structural data for the Na^+, K^+-ATPase suggested that the β subunit would be positioned near membrane spanning domains M7 and M10. In the model, there is a space behind M7 that could hold the HK β subunit (18).

In the cytoplasmic portion of the protein, three clear domains are represented: the nucleotide binding domain (N), the phosphorylation domain (P) and the actuator domain (A). The N domain is responsible for binding and hydrolysis of ATP. In Figure 1-3, lysine 517 is highlighted in green. This amino acid forms a portion of the ATP binding pocket. The P domain contains aspartic acid 385, shown in red. This amino acid becomes phosphorylated during the reaction cycle. These two amino acids are far apart in this model because the model represents the E_1 state of the enzyme. When ions bind to the channel, the enzyme goes through a conformational change that brings these amino acids closer together so that the phosphate of ATP can be transferred to the aspartate.

The A domain contains the conserved TEGS loop that appears to play a role in catalysis of the ATP. The A domain interacts with the P domain in the E_1 state and appears to modulate the ability of the P domain to interact with the N domain.

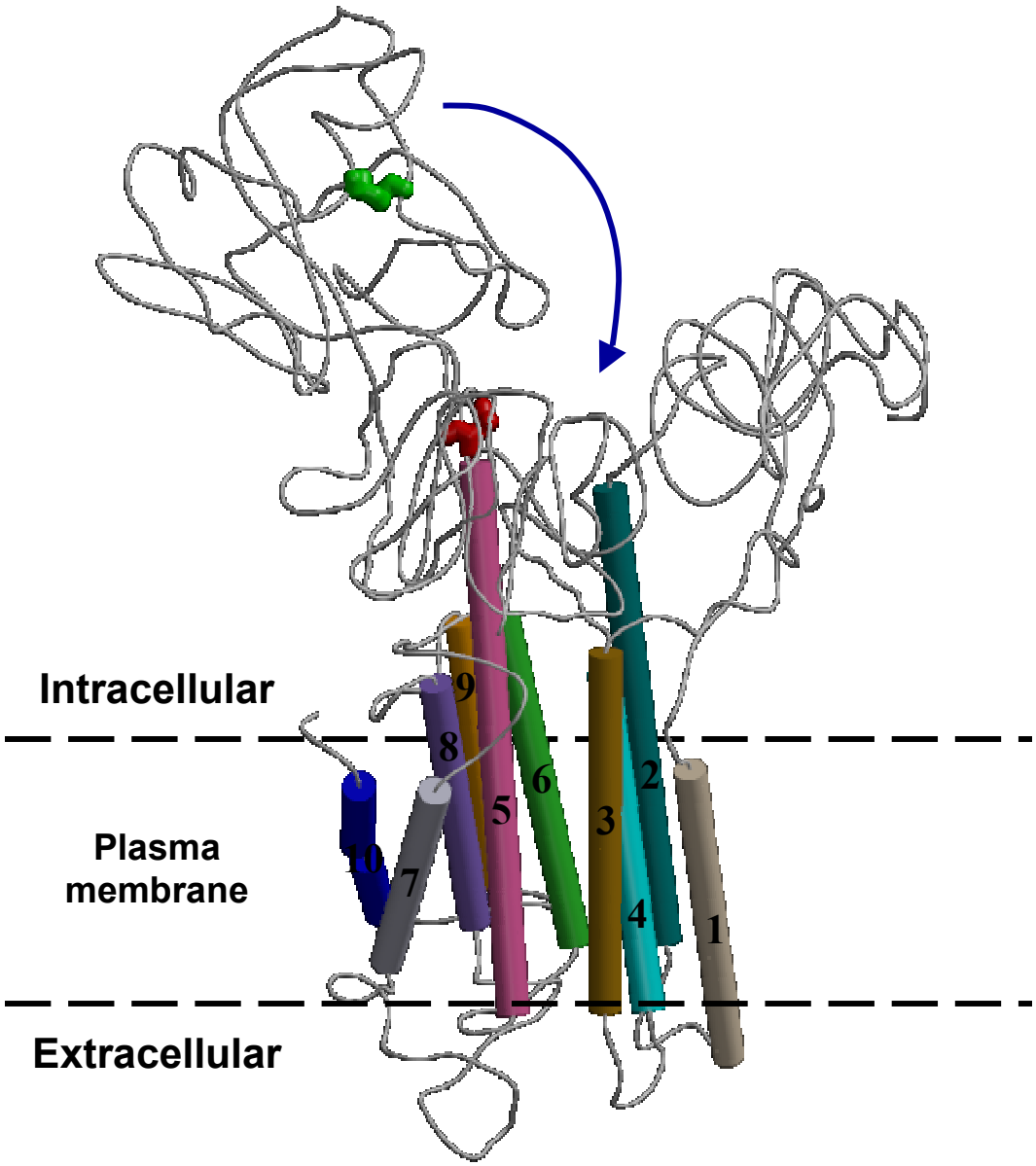


Figure 1-3. High resolution model of the rabbit $HK\alpha_{2a}$ subunit. Courtesy of Michelle Gumz (18). This model is a representation of the E_1 state of the reaction cycle. Each transmembrane domain is numbered. Highlighted in green is lysine 517, part of the ATP binding pocket. Highlighted in red is aspartic acid 385, the amino acid that becomes phosphorylated during the reaction cycle. The arrow indicates the direction of movement of the P domain during the conformational change to the E_2 state.

The Reaction Cycle

It has long been established that the P-type ATPases go through significant conformational changes while translocating ions (44). Biochemical data suggested that the enzymes exist in two main conformational states. The E_1 state binds ions and transfers phosphate from ATP to an amino acid residue within the enzyme. The hydrolysis of ATP causes a conformational change in the enzyme to the E_2 state. This confirmation has low affinity for ions and releases them to the opposite side of the membrane. The recent crystal structures that represent the E_1 (46) and E_2 (47) states of the Ca^{++} ATPase taken together with the biochemical data and lower resolution structures of several P-type ATPases suggest that the reaction cycle is much more complicated, consisting of a series of stable intermediate steps (44). The reaction cycle described below and depicted in Figure 1-4 is for the Ca^{++} ATPase, but is likely representative of the P-type ATPases as a whole.

The starting conformation is the closed (E_2H) state (Figure 1-4A). The Ca^{++} binding sites were protonated during the previous reaction cycle and face the lumen of the sarcoplasmic reticulum. Deprotonation of these sites is accompanied by a rotation in the P domain and a reorientation of the Ca^{++} binding sites towards the cytoplasm of the cell. The P domain and the A domain are now able to interact with each other. This is the state is the open conformation (E_1) (Figure 1-4B). Once Ca^{++} is bound to the channel, the N domain binds ATP ($E_1MgATPCa_2$) and the P domain loses its interactions with the A domain (Figure 1-4C). The P domain is then ready to accept a phosphate from an ATP that is hydrolyzed in the N domain ($E_1MgP(Ca_2)ADP$) (Figure 1-4D). The release of ADP from the enzyme causes a major conformation shift that orients the Ca^{++} ions to the lumen (E_2MgPCa_2) (Figure 1-4E).

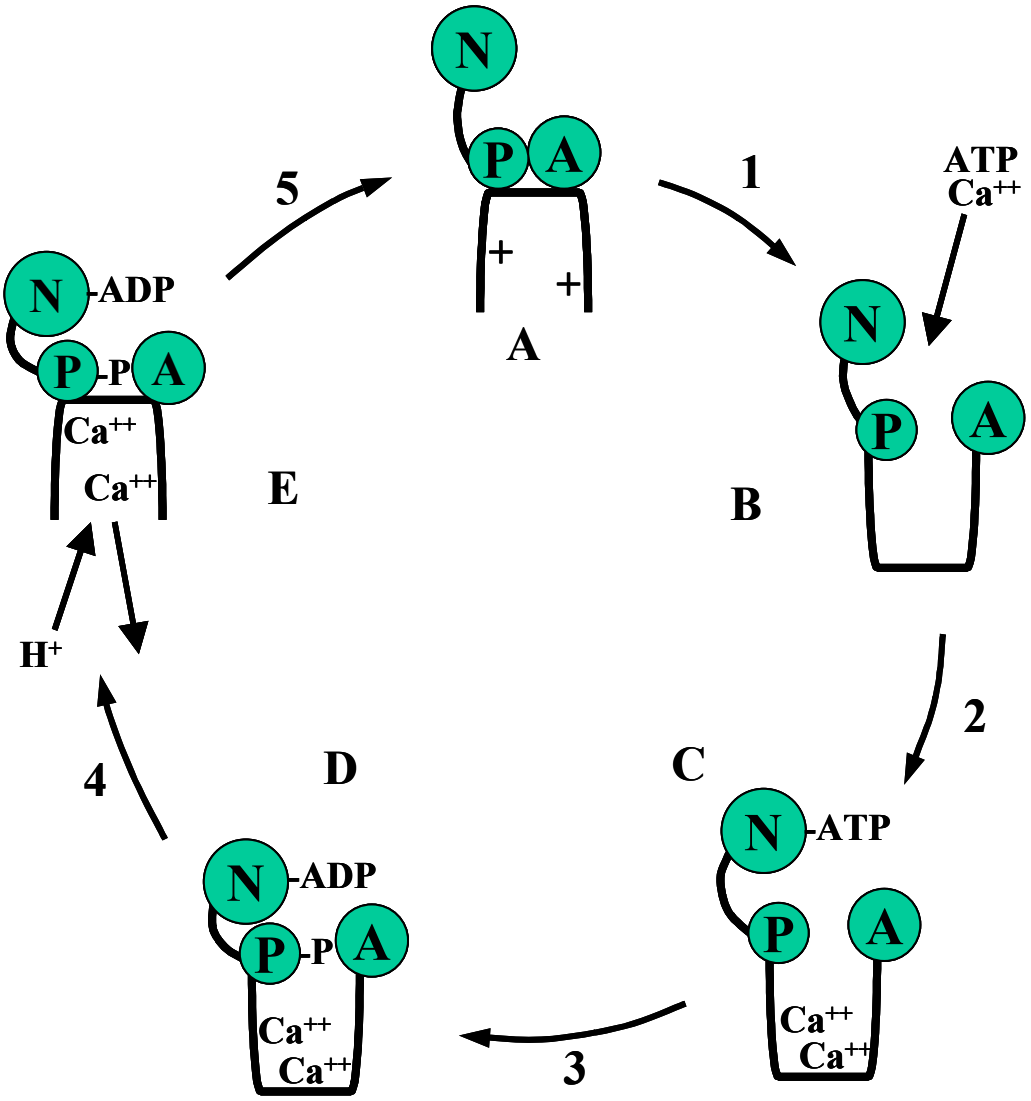


Figure 1-4. Reaction Cycle of the Ca++-ATPase as a representative of P-type ATPases. The reaction intermediates are (A) E2H (B) E1 (C) E1MgATPCa2 (D) E1MgP(Ca2)ADP (E) E2MgP. Numbers represent transitions between states as described in the text.

This conformation has a low affinity for Ca^{++} and it is released (E_2MgP). The sequential release of H_20, Mg^{++}, and inorganic phosphate returns the enzyme in the E_2 state ready for protonation and the beginning of another reaction cycle (Figure 1-4A).

Structure of the HKα_2 Gene

The gene that encodes the rabbit HKα_2 subunit was previously unknown. One of the aims of our study was to clone the gene from a rabbit genomic library. The genes for the human and mouse gene, however, have been cloned. This section will review general characteristics of eukaryotic genes and the genomic organization for the human and mouse HKα2 genes.

Eukaryotic Gene Organization

By definition, a gene is considered to be the entire nucleic acid sequence that is necessary for the synthesis of a functional polypeptide or RNA (31). According to this definition, the nucleotide sequence that codes for the protein or functional RNA is only a portion of the entire gene. The rest of the gene consists of non-coding sequences that play a role in creating the final product. This section describes common features of eukaryotic genes such as intron/exon organization, core promoter elements and transcriptional control elements.

The nucleotide sequence that codes for a protein is not found as a continuous sequence. Instead, the coding regions (exons) are interrupted by non-coding regions (introns). Once the entire region is transcribed, RNA splicing machinery recognizes specific splice junction sequences, removes the introns and joins together the exons creating a complete mRNA sequence. Importantly, this arrangement allows for alternative splicing. It is possible for the RNA splicing machinery to join some exons and remove others creating alternative RNA transcripts from the same gene. The production of alternative transcripts may be regulated by different stimuli or in a tissue specific manner.

The core promoter elements are those that are necessary for basal transcription (7). They are located near the transcription start site. Common core promoter elements are diagramed in Figure 1-5. They include the TATA box, the initiator element (Inr), the downstream promoter element (DPE), the TFIIB recognition element (BRE), and the CpG island. The TATA element, with the consensus sequence TATAAA, is usually located about 25-30 bases upstream of the transcription start site. This element is the binding site for the TATA binding protein (TBP). TBP and its associated factors (TAFs) make up the general transcription factor TFIID. Once bound, TFIID is capable of recruiting and/or interacting with other general transcription factors and RNA polymerase. The entire complex is then capable of determining the specific transcription start site and initiating basal level transcription. In the absence of a TATA box, an initiator element may be present and act to initiate transcription. This element has a very loose consensus sequence of PyPyA+1NT/APyPy and usually overlaps with the transcription start site. It is thought that the consensus sequence is recognized by TBP associated factors and directs TBP to bind upstream in a TATA box independent manner. The rest of transcription initiation occurs similarly to TATA box containing promoters. The downstream promoter element is found about 30 bases downstream of the transcription start site. It has a consensus sequence of RGA/TCGTG. Similarly to the initiator, the DPE is thought to bind a TBP associated factor and direct specific initiation of transcription. It is known to work in conjuction with the initiator sequence at TATA-less promoters, but may also function to stabilize core promoters with weak TATA elements. The TFIIB recognition sequence, G/CG/CG/ACGCC, is located immediately upstream of the TATA box (29). As its name implies, this sequence binds the general

transcription factor TFIIB. This element is present at many, but not all, eukaryotic promoters. Finally, some promoters do not have TATA box or an initiator element. Instead these promoters contain a region of high GC content upstream of the coding region. This GC rich region, known as a CpG island, can form multiple binding sites for SP family members. The stimulatory protein family can bind and direct the formation of preinitiation complexes. This process, however, is often imprecise and allows for multiple transcription start sites. In addition to basal level transcription, many eukaryotic genes are activated and repressed by environmental stimuli. This change in regulation is modulated by the binding of sequence-specific DNA-binding proteins that can interact with the core promoter proteins and either activate or repress transcription.

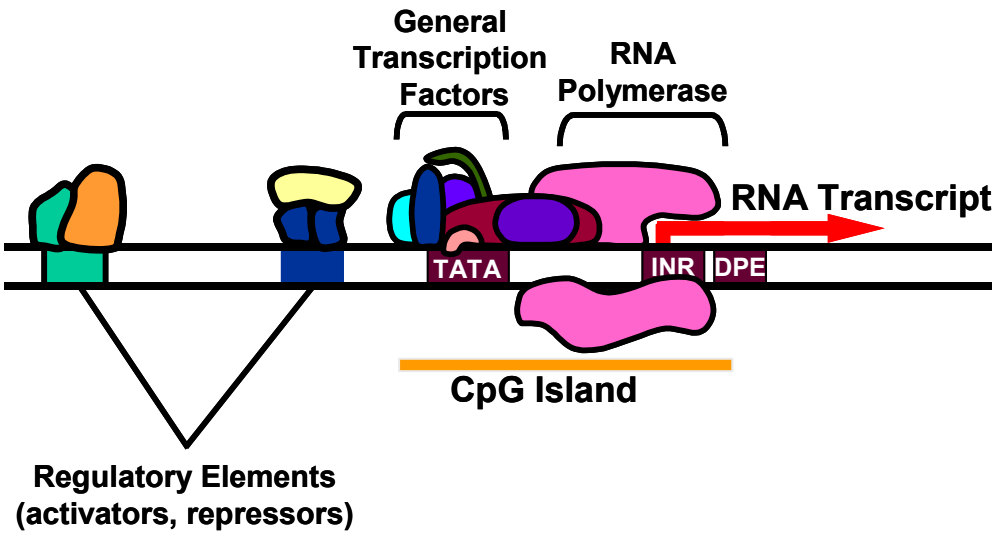


Figure 1-5. Common eukaryotic gene promoter elements. TATA represents the TATA element. INR represents the initiator element. DPE represents the downstream promoter element. Adapted from Lodish et al. (31).

HKα2 cDNAs

Table 1-1 lists all known HKα2 cDNAs along with the sizes of their 5' untranslated regions (UTR), open reading frames and the 3'UTRs. The $HK\alpha_{2a}$ cDNA

sequences are very similar. The first rat $HK\alpha_{2a}$ cDNA was obtained by screening a colonic cDNA library (9). Subsequently, Kone and coworkers (26) reported a cDNA with an identical open reading frame that they cloned from kidney. The 3' UTRs were also identical, but Kone obtained a longer 5'UTR by using primer extension analysis. In addition to the $HK\alpha_{2a}$ sequence, a second HKα2 cDNA termed $HK\alpha_{2b}$ was found (see below). Two groups independently cloned the $HK\alpha_{2a}$ cDNA from rabbit kidney and performed 5' and 3' Rapid Amplification of cDNA Ends (RACE) to determine the length of the $HK\alpha_{2a}$ cDNA (6, 12). RACE has the capacity to produce a full-length cDNA, but the length of the PCR product is dependent upon the efficiency of the reverse transcriptase reaction. Therefore, a full-length cDNA is not necessarily produced. For the rabbit $HK\alpha_{2a}$ cDNAs, Fejes-Toth et al. (12) acquired more 5' UTR sequence while Campbell et al. (6) obtained the longer 3' UTR sequence. Our study used the RNase protection assay to determine the true 5' end of the $HK\alpha_{2a}$ and $HK\alpha_{2c}$ transcripts. The 3'UTR obtained by Campbell et al. extended to the likely poly A signal, and was viewed as complete. It is notable that the complete 3' UTR of the human ATP1AL1 transcript was apparently much shorter than that of other mammalian species. Whether this effects mRNA stability or has some other regulatory significance has not been studied.

cDNA clones representing splice variants of HKα2 gene transcripts have been obtained for rat ($HK\alpha_{2b}$) (26) and rabbit ($HK\alpha_{2c}$) (6). $HK\alpha_{2b}$ and $HK\alpha_{2c}$ cDNAs differ from the $HK\alpha_{2a}$ cDNA only at the extreme 5' end but the deduced protein products differ significantly.

Table 1-1. cDNA's for mammalian HKα2 transcripts.

Organism Subunit	Apparent 5'UTR	Open Reading Frame/ Amino acids	Probable 3'UTR	Genebank Accession Number	Reference
Human					
ATP1AL1	168	3117 / 1039	290	U02076	[Grishin, 1999 #8]
Rat					
$HK\alpha_{2a}$	202	3108 / 1036	650	M90398	[Crowson, 1992 #1]
$HK\alpha_{2a}$	275	3108 / 1036	650	U94912	[Kone, 1998 #66]
$HK\alpha_{2b}$	739	2787 / 929	348	U94913	[Kone, 1998 #66]
Rabbit					
$HK\alpha_{2a}$	190	3101 / 1034	875	AF106063	[Fejes-Toth, 1999 #5]
$HK\alpha_{2a}$	39	3101 / 1034	939	AF023128	[Campbell, 1999 #3]
$HK\alpha_{2c}$	198	3285 / 1095	939	AF023129	[Campbell, 1999 #3]
Guinea Pig					
$HK\alpha_{2a}$	145	3101 / 1034	996	D21854	Genbank Database
Mouse					
$HK\alpha_{2a}$	253	3108/1036	629	AF350499	[Zhang, 2001 #31]].

Figure 1-6 is a schematic representation of the likely mechanism for the formation of the alternative transcripts. In both species, $HK\alpha_{2a}$ transcription initiates at exon 1 and the 5' end of the mRNA is produced by splicing exon 1 to exon 2. The rat $HK\alpha_{2b}$ and the rabbit $HK\alpha_{2c}$ transcripts arise from transcription initiation within what had been designated as intron 1 of the HKα2 gene. The rat $HK\alpha_{2b}$ cDNA had a peculiar organization. The 5' UTR was longer than any of the other HKα2 5' UTRs (739bp) and is in fact much longer than a typical mammalian 5'UTR. In addition, it contained eight short open reading frames prior to the $HK\alpha_{2b}$ translation start site that may offer a mechanism for

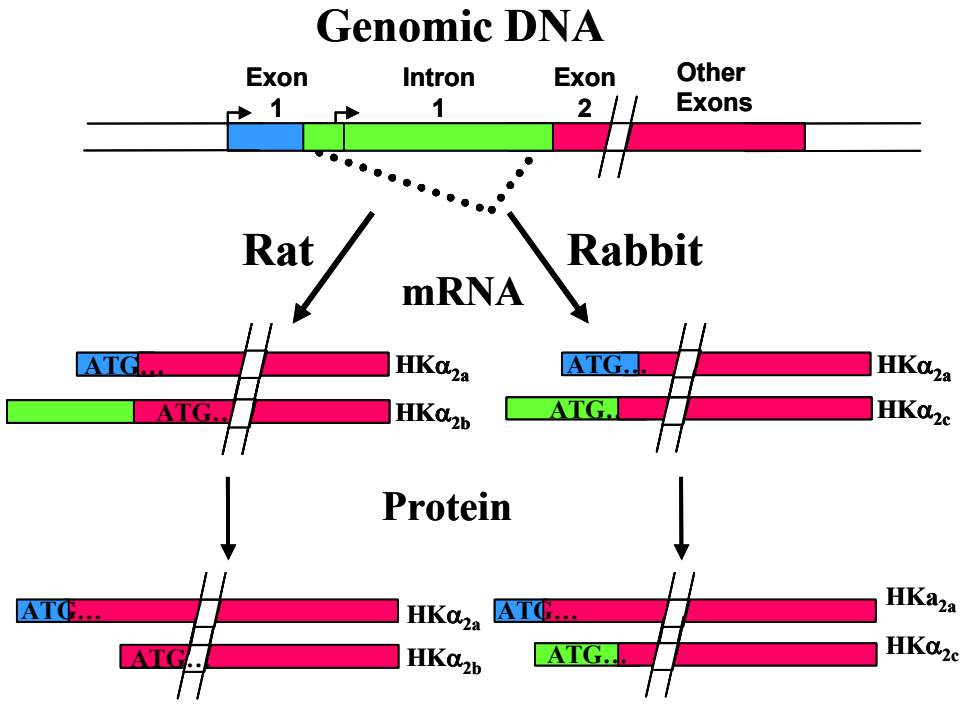


Figure 1-6. Formation of the alternative transcripts from the rabbit and rat HKα2 genes.

translational regulation (26). Northern analysis of total RNA from various rat tissues shows that the HKα_{2b} transcript was present *in vivo*. The HKα_{2b} protein appeared to be truncated by 108 amino acids, but the presence of the protein *in vivo* has not been reported. In rabbit, the HKα_{2c} 5' UTR contains only two short open reading frames prior to the translation start and is of similar length to other HKα2 transcripts. Unlike the truncated HKα_{2b} of rat, the HKα_{2c} protein contains a 61 amino acid extension on the N-terminus of HKα_{2a}. Fejes-Toth et al. (12) failed to detect HKα_{2c} in their 5' RACE experiments. However, Campbell et al. (6) demonstrated by both Northern and Western analyses that HKα_{2c} mRNA and protein were present in rabbit kidney tissue as well as in tissue culture cells derived from the rabbit cortical collecting duct (RCCT28A). The

cloning of homologous cDNAs from other species will be necessary to determine if the formation of splice variants is common among mammalian species or whether it is unique to rat and rabbit.

Human HKα2 Gene

The human HKα2 gene was originally identified as the ATP1AL1 gene (45). It was initially unclear as to whether or not this H^+, K^+ - ATPase should be considered homologous to the rat $HK\alpha_{2a}$ (8). The amino acid identity for these two proteins was much lower (87%) than that of the $HK\alpha_1$ proteins from human, rat and rabbit (97%). The cloning of the rabbit $HK\alpha_{2a}$ cDNA (6) suggests that all three proteins are homologous as it had approximately 87% amino acid identity with both rat $HK\alpha_{2a}$ and human ATP1AL1 sequences (8). The genomic structure of the mouse HKα2 gene (below), the rat gene (NCBI database), and the rabbit HKα2 gene (this study) confirm that these genes are homologous genes.

The complete genomic organization of the human ATP1AL1 gene was reported by Sverdlov et al. in 1996 (45). The gene is approximately 32Kb in length and contains 23 exons and 22 introns. The sizes of the exons and introns are included in Table 2-1. The transcription start site was mapped using S1 nuclease protection and primer extension. The S1 nuclease protection assay produced 4 clustered bands from –185 to –188. The primer extension produced a single band marking the transcription start site at –187 with respect to the ATG start codon. Analysis of the region immediately 5' of the transcription start site identified possible regulatory elements including a TATA box, SP family binding sites, AP-2 binding sites and NF-kB. Additionally, the region from –484 to +369 met the criteria for a CpG island. Analysis of the 3' region of the gene identified

3 possible polyadenylation sites. The authors suggest that these sites may be used in a tissue specific manner.

Mouse HKα2 Gene

Recently, the complete sequence and structure of the murine HKα2 gene was reported (52). Similarly to the human gene, the murine gene spans 23.5Kb and contains 23 exons. The transcription start site was mapped using primer extension. It is located at –253 with respect to the ATG start codon. These authors did not observe an alternative transcription start site as seen in rat (26) and rabbit (6). Computer analysis of 7.2Kb of sequence immediately 5' of the start site identified many possible transcription factor binding sites including a TATA box, CEBP, NF-kB, cAMP and glucocorticoids. There appears to be one poly A signal designating the 3' end of this gene.

Summary

Intracellular and extracellular K^+ concentrations play a critical role in normal cell functioning. The collecting duct segment of the nephron is the major location where K^+ ions are reabsorbed when blood K^+ becomes too low. The HKα2 subtype of the H^+, K^+ - ATPase, located on the apical membrane of collecting duct cells, appears to play a major role in K^+ reabsorption since its activity and its expression appear to be increased when blood K^+ concentration is low. There is additional evidence that the expression of the HKα2 gene products are also regulated by Na^+ levels, acid/base balance, and hormones. All of the studies on HKα2 gene expression, however, have been performed *in vivo*. There is nothing known about the molecular mechanisms responsible for the change in gene expression. The goal of our study was to characterize the rabbit HKα2 gene and initiate studies on the molecular regulation of the gene.

CHAPTER 2
CLONING OF THE HKα2 GENE

The first specific aim of this dissertation was to clone the rabbit HKα2 gene. There were two important reasons for carrying out this goal. First, cloning the gene would provide sequence information essential for the future experiments designed to study the gene's promoter and regulatory elements. The cDNAs for $HK\alpha_{2a}$ and $HK\alpha_{2c}$ have been previously identified (6) and shown to be products of the same gene (5). However, there is currently nothing known about the genomic sequence that is 5' of the cDNA ends. This dissertation, therefore, provides the first data regarding the upstream sequence that contains the gene's promoter and its regulatory elements. The second purpose for cloning the rabbit HKα2 gene was to determine its genomic organization. There has been some controversy over whether or not the HKα2 proteins identified from several species were in fact homologous proteins (8). The α subunit proteins from rabbit ($HK\alpha_{2a}$), rat ($HK\alpha_{2a}$), mouse (HKα2), guinea pig (HKα2) and human (ATP1AL1) share an amino acid identity of 87%. This percentage is much lower than the amino acid identity that the same species share for the HKα1 subunit (97%). One method for resolving this controversy is to compare the intron and exon sizes of the genes that encode the proteins since genes that were derived from common ancestors should maintain a consistent organization. The genomic organization of the human ATP1AL (45) and mouse HKα2 (52) genes have been determined independently from this dissertation. The genomic organization of the rat gene was determined from the rat

genome database at the National Center for Biotechnology Information (NCBI). The cloning and sequencing of the rabbit gene is the subject of this chapter. The comparison of the organization of the four genes supports the conclusion that these genes are homologous and have been derived from a common ancestor.

The rabbit HKα2 gene was cloned from a bacteriophage λ library of the rabbit genome. Three clones were identified that span a majority of the rabbit HKα2 gene. The polymerase chain reaction (PCR) was used on rabbit genomic DNA in order to amplify the remaining portion of the gene. This chapter discusses the cloning, sequencing and analysis of the bacteriophage λ clones and the PCR products that contain the HKα2 gene.

Materials and Methods

Screening the Lamda Genomic Library

A bacteriophage λ library containing 15Kilobase pair (Kbp) inserts of rabbit genomic DNA was purchased from Clontech, Inc. (Catalog #TL1008j). Two approaches were used to screen the library. The first method was a PCR-based screen (21) and the second was a more conventional plaque lift/hybridization screen (Clontech, Inc).

The PCR approach used primers designed to the $HK\alpha_{2a}$ cDNA to identify aliquots of the λ library that contained bacteriophage with inserts that corresponded to the HKα2 gene. The primers BC334 (5'-TATCTGTAGCTGCATGGTGCTCCAC-3') and BC386 (5'-ACCCGCGCGCTCCAGCGCGACAT-3') were used in the PCR reaction. These primers correspond to base pairs 69-93 and base pairs 16-40 the $HK\alpha_{2a}$ cDNA and are known to amplify a 647 base pair fragment from rabbit genomic DNA (5). The amplified fragment is larger than expected because it includes intron sequence that is not present in the cDNA. This reaction was repeated as a positive control for the PCR approach to

screening the λ library. Additionally, 2µl of the PCR reaction was ligated into the TOPO cloning vector (Invitrogen, Inc. Cat# K4574-J10). This construct was designated pDZ6 and is referred to as the 5' probe for all future experiments. To screen the library, 500µl of *E.coli* strain K802 (Clontech, Inc.) was infected with 1 X 10^6 plaque forming units (pfu) of the λ genomic library and incubated at 37°C for 15 minutes. The infected bacteria were brought to a volume of 10mL with LB broth (tryptone, yeast extract, NaCl) supplemented with 10mM $MgSO_4$. 100µl of infected bacteria were placed into each well of a 96 well plate and amplified by growth at 37°C for 5 hours. 25µl from each row and each column were pooled to form 16 samples representing the 96 wells. 10µl of the pooled samples were used in a PCR reaction (250pm each primer, 250µM dNTP mix, 1X PCR buffer, 5U Taq polymerase, 10µl of pooled template, dH_2O to 40µl). The PCR conditions were 94°C for 1 minute 30 seconds; 94°C for 15 seconds + 72°C for 2 minutes times 5 cycles; 94°C for 15 seconds + 70°C for 2 minutes times 5 cycles; 94°C for 15 seconds + 68°C for 2 minutes times 25 cycles; 68°C for 8 minutes. Southern analysis of the PCR products was carried out as described in Maniatis et al. (39). The PCR reactions were run on a 1% agarose gel and visualized with ethidium bromide stain and UV light. The gel was soaked gel in denaturation solution (1.5M NaCl, 0.5M NaOH) for 30 minutes, neutralizing solution (1.5M NaCl, 1.0M Tris-Cl pH 8.0) for 30 minutes, and equilibrated in 10X SSC (1.5M NaCl, .15M sodium citrate) for 30 minutes. The DNA was transferred to a nylon membrane by capillary action overnight. The DNA was UV crosslinked to the nylon membrane, placed in a hybridization tube and incubated with 5mL of buffer (.25M Na_2HPO_4 pH 7.2, 1mM EDTA, 1% BSA, 7% SDS) at 65°C for at least 15 minutes. A radioactive probe was prepared by digesting pDZ6 with *Eco*RI and

running the digestion on a 1% agarose gel. The 647bp band was cut from the gel and extracted using the Gel Extraction Kit from Qiagen, Inc.(Catolog #28706). 25ng of the DNA fragment were used to create a ^{32}P labeled probe. The method used was the Prime-It RmT Random Primer Labeling Kit from Stratagene, Inc. (Catolog #300392). The probe was added directly to the prehybridization solution and incubated with the nylon membrane overnight at 65°C. The next day, the membrane was washed three times with 300mL of wash buffer (20mM Na_2HPO_4 pH 7.2, 1% SDS, 1mM EDTA) at 65°C, wrapped in saran wrap and exposed to autoradiograph film for an appropriate length of time. Positive rows and columns were aligned to determine possible positive wells. The PCR was repeated on samples from individual wells. The contents of positive wells were used to create plaques on agar plates. Plaque lift hybridizations were performed to purify positive plaques (see below).

The second method for screening the library was to perform a series of plaque lift hybridizations. The protocol was modified from the procedure recommended by the rabbit genomic library manufacturer, Clontech, Inc. The initial screens were performed with approximately 20,000 pfu per 150mm agar plate. The λ library was diluted 1/500 with dilution buffer (1M NaCl, 0.1M $MgSO_4$, 1M Tris pH 7.5) and 100µl was used to infect 600µl of *E. coli* strain K802 at 37°C for 15 min. 7mL of warmed soft agar (LB broth + 7g/L agar) was added to infection and poured over a 150mm LB agar plate. After overnight incubation at 37°C, plaque lifts were performed by placing circular nylon membranes on top of each plate for 2 minutes, the filter was removed placed plaque side up on top of Whatman paper #3 soaked in denaturing solution for (1.5M NaCl, .5N NaOH) for 5 minutes, neutralizing solution (1.5M NaCl, 1M Tris pH 8.0) for 5 minutes

and 2X SSC (.3M NaCl, .03M Sodium Citrate) for 5 minutes. The filters were baked dry at 80°C for 1 hour. The filters were probed with either a 5' probe (pDZ6), a mid probe ($HK\alpha_{2a}$ cDNA bases 1264 - 1569) or a 3' probe ($HK\alpha_{2a}$ cDNA bases 3265 - 4073) as described for the Southern blot filters. Positive plaques were pulled from the agar plates with a Pasteur pipette and diluted into 1mL of dilution buffer. Serial dilutions of the plaque were used to infect *E.coli* strain K802, mixed with 3mL of soft agar and poured over 100mm agar plates. The plaque lift procedure was repeated. Isolated positive plaques were picked and the procedure was repeated until all plaques on a plate were positive and each plaque was considered pure.

Pure plaques that were isolated by both screening methods were grown in large scale and DNA was isolated using the Qiagen Lambda Maxi Kit (Cat# 12562). A 5mL overnight culture of *E.coli* strain K802 was pelleted and resuspended in 1.5mL of bacteriophage dilution buffer. Approximately 1 X 10^7 bacteriophage were added to the bacterial cells and incubated at 37°C for 20 minutes. The infected bacteria were added to 250mL of LB supplemented with 10mM $MgSO_4$ and 0.2% maltose. The culture was allowed to grow at 37°C until the bacteria lysed (approximately 4 hours). The bacterial debris was pelleted while the bacteriophage that remained in the supernatant was used for DNA isolation. 400µl of buffer L1 (300mM NaCl, 100mM Tris-Cl pH 7.5, 10mM EDTA, .2mg BSA, .2mg/mL RNaseA) was added to the lysate and incubated at 37°C. This step digests away any bacterial RNA. In order to precipitate and pellet the bacteriophage, 50mL of buffer L2 (30% polyethylene glycol, 3M NaCl) was added and the mixture was incubated on ice for 60 minutes and centrifuged at 10,000 rpm for 10 minutes. The pellet was resuspended in 9mL of buffer L3 (100mM NaCl, 100mM Tris-

Cl pH 7.5, 25mM EDTA) and 9mL of buffer L4 (4% sodium dodecyl sulfate (SDS) and the mixture was heated to 70°C for 10 minutes. This step denatured the bacteriophage proteins and released the bacteriophage DNA. After cooling on ice, 9mL of buffer L5 (3M potassium acetate) was added and mixture was centrifuged for 30 minutes at 15,000 rpm in order to pellet bacteriophage proteins. The supernatant that contained the bacteriophage DNA was poured over a Qiagen column that was equilibrated with buffer QBT (750mM NaCl, 50mM MOPS, 15% isopropanol, .15% triton X-100). The DNA bound to the column. The column was washed with 60mL of buffer QC (1M NaCl, 50mM MOPS, 15% isopropanol) and then the DNA was eluted off the column with 15mL of buffer QF (1.25M NaCl, 50mM Tris-Cl pH 8.5, 15% isopropanol). 10.5mL of isopropanol was added to the eluate and centrifuged at 15,000 rpm for 30 minutes to precipitate and pellet the bacteriophage DNA. The DNA was washed with 70% ethanol and resuspended in TE (10mM Tris pH 7.5, 1mM EDTA). The DNA from each λ clone was digested and used in Southern analysis in order to identify overlapping clones that span the entire HKα2 gene.

λHKα2.1 Sequence

Clone HKα2.1 was the 5' most clone. Southern analysis showed that the *Eco*RI fragment that hybridized to the 5' probe was attached to the λ SP6 arm and sequencing with the SP6 promoter primer (5'-ATTTAGGTGACACTATAG-3') indicated that the orientation of the insert was such that the rest of the clone contained sequence 5' of the HKα2 gene. In order to obtain the sequence of the region immediately 5' of the cDNA, a 6.3Kbp *Xho*I fragment was subcloned into pBluescript (pBS, Stratagene, Inc.) creating plasmid pDZ10. The sequence of the entire fragment was determined by walking along

the sequence in both 5' and 3' directions. Sequencing was carried out by the University of Florida Interdisciplinary Center for Biomedical Research (ICBR) sequencing core facility. Complete sequences from both directions were obtained by compiling sequences from individual primers. The two complete sequences were compared and any base pair mismatches were resolved by additional sequencing through the region. The location of each primer on the complete pDZ10 sequence and the sequence of primers can be found in Figure 2-1 and Table 2-1 respectively.

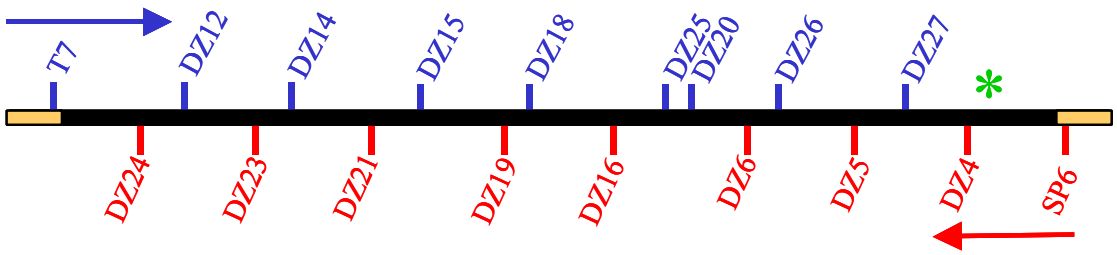


Figure 2-1. Location of Sequencing Primers on the Complete pDZ10 sequence. Black represents the 6.3Kbp *Xho*I fragment subcloned from λHKα2.1. Orange represents the ends of pBS (Stratagene, Inc.). Blue represents primers used for sequencing in the 5' to 3' direction. Red represents primers used for sequencing in the 3' to 5' direction. The star indicates the location of the 5' probe.

λHKα2.5 Sequence

Clone HKα2.5 was the middle clone. It hybridized to both the 5' and the mid probes ($HK\alpha_{2a}$ cDNA base pairs 16-93 and 1264-1569 respectively). Since this clone was being used to determine the intron/exon boundaries within the rabbit HKα2 gene, it was necessary to obtain the sequence of the entire clone. The λ DNA arms were

removed by digestion of the clone with *Xho*I. The digestion was run on a 1% agarose gel and visualized with ethidium bromide. The genomic DNA insert was cut from the gel

Table 2-1. Primers used for pDZ10 sequencing.

Name	Primer Sequence 5' to 3''	Location
T7	CCCTATAGTGAGTCGTATTA	λ arm
DZ12	CAATCCACGTTGCCCGCATGGG	784-805
DZ14	CCAGTCCGGATACGGAGCAGG	1437-1457
DZ15	CCCCACCAACAGCCCAGACG	2228-2247
DZ18	CCTCCAGGTGAGGACTACTCC	2974-2994
DZ25	CTCTCCCCTCCAACTCTGAAGG	3689-3710
DZ20	GAACGGCCGGCGCTGCGG	3774-3791
DZ26	GTGTCCCATGTGGGAAGCCAGG	4378-4399
DZ27	CTTGGGGGCTCCGGATCCTGG	5064-5083
SP6	ATTTAGGTGACACTATAG	λ arm
DZ4	CGCATGTCGCGCTGGAGG	5587-5570
DZ5	CTGCACTCTCAGAGTGAAGG	4889-4870
DZ6	GGCTATGGGACAGGGATGACG	4165-4145
DZ16	GGCACAGAGAAGTAGTGCCC	3469-3450
DZ19	GAAACCTACTCATGCCAGGCTC	2752-2731
DZ21	GATGAGTTCTCAGGACTCTGAC	1977-1956
DZ23	GCTGCAGCCTAGCACAC	1237-1221
DZ24	GGGGAGTAAACCTCAGGATGGG	568-547

Blue represents primers used for sequencing in the 5' to 3' direction. Red represents primers used for sequencing in the 3' to 5' direction.

and extracted from the agarose using the QIAquick Gel Extraction Kit (Qiagen, Inc. Cat # 28706). This procedure was repeated until 25µg of insert DNA was obtained. The insert fragment was then sheared and shotgun subcloned into the TOPO cloning vector according to Invitrogen, Inc. (Cat.# K7000-01). 25µg of DNA was added to 750µl shearing buffer (TE, 20% glycerol) and placed in a nebulizer attached to a compressed air pump. The DNA was sheared twice at 10psi for 90 seconds. The sheared DNA was precipitated (80µl 3M NaOAc, 4µl glycogen, 700µl 100% isopropanol) on dry ice for 15 minutes, pelleted by centrifugation at 12,000 rpm for 15 minutes, washed with 80% ethanol and resuspended in 200µl of sterile dH_2O. In order to repair the sheared ends for

cloning, 2µg of DNA was added to a blunt-end repair reaction (20µl DNA, 5µl 10X blunting buffer, 1µl BSA, 5µl dNTP mix, 2µl T4 DNA polymerase, 2µl Klenow DNA polymerase) and incubated at room temperature for 30 minutes. The enzymes were deactivated by heating the reaction mix to 75°C for 20 minutes. Dephosphorylation of the repaired ends was carried out by adding 35µl sterile dH_2O, 10µl 10X dephosphorylation buffer, and 5µl calf intestine alkaline phosphatase (CIP) to the blunt end repair reaction and incubating the reaction at 37°C for 60 minutes. The reaction was phenol/chloroform extracted, precipitated and resuspended in 20µl of sterile dH_2O. Shotgun cloning of the λ DNA was carried out with 3 concentrations of DNA (60ng, 20ng, 5ng), 1µl salt solution, and 1µl pCR4-blunt TOPO vector (Invitrogen, Inc.). The ligations were incubated at room temperature for 5 minutes and then transformed into chemically competent *E. coli* strain DH5α. The resulting bacterial colonies were screened to identify 48 plasmids containing inserts of approximately 1500bp. Each colony was grown overnight in 3mL of LB broth and then miniprep DNA was isolated using the QIAprep Spin Miniprep Kit from Qiagen, Inc. (Catolog # 27106). The DNA was digested with *Eco*RI, run on a 1% agarose gel and visualized with ethidium bromide stain. When 48 bacterial colonies that contained plasmids of the appropriate size were identified, 200µl of an overnight culture of each colony was placed in an ELISA plate and taken to ICBR for high throughput sequencing. Approximately 250bp of sequence was obtained from each end of the plasmid inserts. These sequences were assembled into ten contiguous fragments (contigs) by ICBR. The high-throughput sequencing core used the Helix Finch program distributed by Giospiza, Inc. in orger to assemble the sequences. The order of the contigs was determined by alignment of the contigs with the $HK\alpha_{2a}$

cDNA and by determination of plasmids that contained sequence in two contigs. The gaps between the fragments were closed by additional sequence from the plasmids that spanned the gaps as well as sequence from the original λ clone HKα2.5. The primers used for the additional sequencing are listed in Table 2-2.

Table 2-2. Primers used for HKα2.8 sequencing

Name	Sequence 5' to 3'	Location
DZ83	CCCCGCTCTAAAGAAGGCCG	2418
DZ94	GGGCTTTCGGCCGACCTCACTG	2873
TC4	CCTGGAATGGACAGGCT	2983
DZ93	GCCTTCTGCCTCCAGGGC	3181
DZ96	GCCCCCGTTTTGACTCCC	3815
DZ95	GAGCGGGGGTGTCATTCACTCCG	2190
MG45	CGTCCATTCCTGTCCATAGCTATCTTCCAAGTCGTTCAGGTG	2897
MG49	CATCGTATACCCAGATCAGGATGGCATGGGGTACGGCCAC	3188

Blue represents primers used for sequencing in the 5' to 3' direction. Red represents primers used for sequencing in the 3' to 5' direction. Location indicates the position of the primer on the $HK\alpha_{2a}$ cDNA.

λHKα2.8 Sequence

Clone HKα2.8 was the 3' clone. It hybridized to the 3' probe ($HK\alpha_{2a}$ cDNA base pairs 3265-4073). The approximate intron/exon boundaries were determined by alignment of the cDNA from rabbit $HK\alpha_{2a}$ to the exon sizes of human ATP1AL1 gene. Primers were then designed near the expected end of each exon (Table 2-3) and were used for partial sequencing of λ clone HKα2.8. The sequencing was carried out by the ICBR sequencing core.

PCR Amplification of the Missing Fragment

In order to obtain the exon boundaries for the portion of the HKα2 gene that was missing from the three λ clones, PCR primers were designed to the approximate ends of the exons. The sequences of the primers, their orientations, and their location along the

$HK\alpha_{2a}$ cDNA are listed in Table 2-3. Proofstart DNA polymerase (Qiagen, Inc. Cat# 202203) and RCCT28A genomic DNA were used for the PCR reaction. The PCR reaction mix consisted of 1X Proofstart PCR buffer containing 15mM $MgSO_4$, 300µM each dNTP, 1µM each primer, 2.5U ProofStart DNA polymerase, 0.5µg DNA template, and dH_2O up to 50µl. The PCR cycle conditions were 95ºC for 5 minutes times one cycle and 94ºC for 1 minute, 60ºC for 1 minute, 72ºC for 2 minutes times 40 cycles. The PCR products were run on a 1% agarose gel and visualized with ethidium bromide. The most intense band was cut from the gel and extracted using the QIAquick Gel Extraction Kit (Qiagen, Inc. Cat# 28706) as described by the manufacturer. In order to obtain the sequence of the exon boundary, 50ng of DNA was sent to ICBR sequencing core along with the primers used to create the PCR product.

Table 2-3. Primers for genomic PCR and sequencing

Name	Sequence 5' to 3'	Location
DZ98	CCTTGGGTGCGGGGGGACAG	Intron11
DZ99	GGGCGGCCTGGGCGAGC	1852
TC80	GGCTCTCTTATCAATGATTCATCC	1979
DZ81	GCCATTGCCAAGAGTGTAGGG	2093
BC231	GCTTGTCATTGGGATCTTCC	1702
DZ100	CTGAGTCAAATGAGTAGGTCTCTGG	1913
DZ101	CCCTACACTCTTGGCAATGGC	2093
DZ97	CTGGGGAAACTTTGCCCTCC	Intron14

Blue represents primers used for sequencing in the 5' to 3' direction. Red represents primers used for sequencing in the 3' to 5' direction. Location indicates the position of the primer on the $HK\alpha_{2a}$ cDNA.

Results

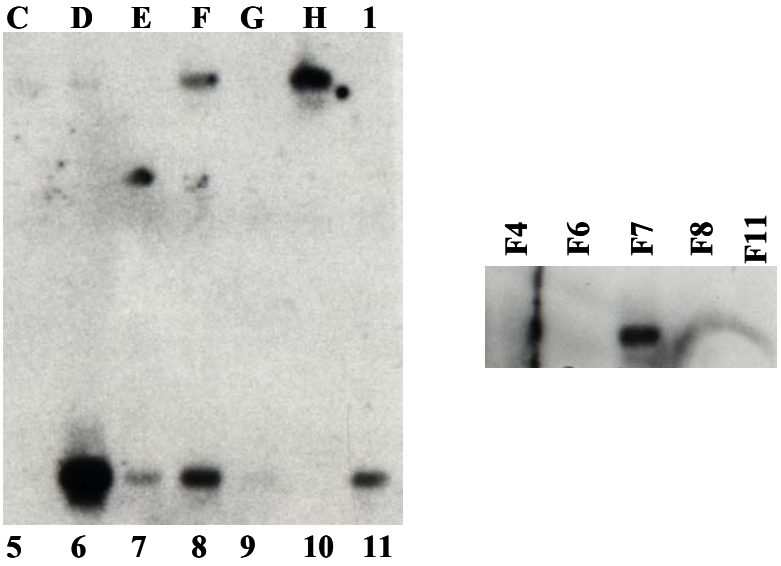
Screening the λ Genomic Library

A majority of the rabbit HKα2 gene was cloned by screening a rabbit genomic library using PCR and plaque lift hybridization methods. These two techniques identified nine λ clones. Two of the clones were identified by the PCR method (HKα2.2, HKα2.4)

and seven of the clones were identified by traditional plaque lift hybridization (HKα2.1, HKα2.3, HKα2.5, HKα2.6, HKα2.7, and HKα2.8).

Figure 2-2 is an example of the PCR screen that identified λ clone HKα2.2. Figure 2-2A shows a Southern analysis of the initial screen. Samples were pooled across rows and down columns and PCR was performed on 25μl of the pooled sample. The PCR products were run on a 1% agarose gel and Southern analysis was performed with the 5' probe. In this example, rows F and H and columns 6, 7, 8 and 11 contain positive clones. Figure 2-2B is the Southern blot that was performed on PCR products from the screen of individual wells in row F. Well F7 was identified as a well containing a positive λ clone. The sample in well F7 was diluted and used in plaque lift experiment to purify the positive clone that was designated HKα2.2. Additional PCR screening with the 5' probe identified λ clone HKα2.4. PCR amplification of positive fragments of DNA quickly caused cross-sample contamination problem that was difficult to overcome. This method was therefore abandoned and the remainder of the λ clones were isolated using standard plaque lift techniques.

Figure 2-3 is example of the plaque lift and purification procedure that used the 5' probe to identify clone HKα2.1. In the first screen (Figure 2-3A) approximately 20,000 plaques were plated on a 150mm dish. After transfer to nylon and hybridization to the 5' probe, only one positive plaque was identified. An agar plug was taken from that region of the plate, diluted, and used to create a new plate with approximately 200 plaques. About one half of the plaques on this plate hybridized to the 5' probe (Figure 2-3B). An agar plug of an isolated plaque was taken from that plate and used to create a new plate with a similar number of plaques. All of the plaques on the new plate hybridized to the



(A) First screen of PCR product pools

(B) Second screen of PCR products from individual wells

Figure 2-2. Southern analysis of PCR screen of the λ genomic library. A) First screen 1kb represents the 1kb ladder, GD represents genomic DNA, A-H represent samples pooled across the rows of an ELISA plate, 1-12 represent samples pooled down the columns of the same ELISA plate. Samples 2 and 10 were cut of before DNA transfer. B) Second screen. Individual wells from positive pools were screened for positive clones.

probe (Figure 2-3C). Any plaque from this plate was considered pure and could be used for a λ DNA maxiprep. This plaque lift method was used to identify seven clones; two hybridized to the 5' probe (HKα2.1, HKα2.3), three hybridized to the mid probe (HKα2.5, HKα2.6, 17-1), and two hybridized to the 3' probe (HKα2.7, HKα2.8).

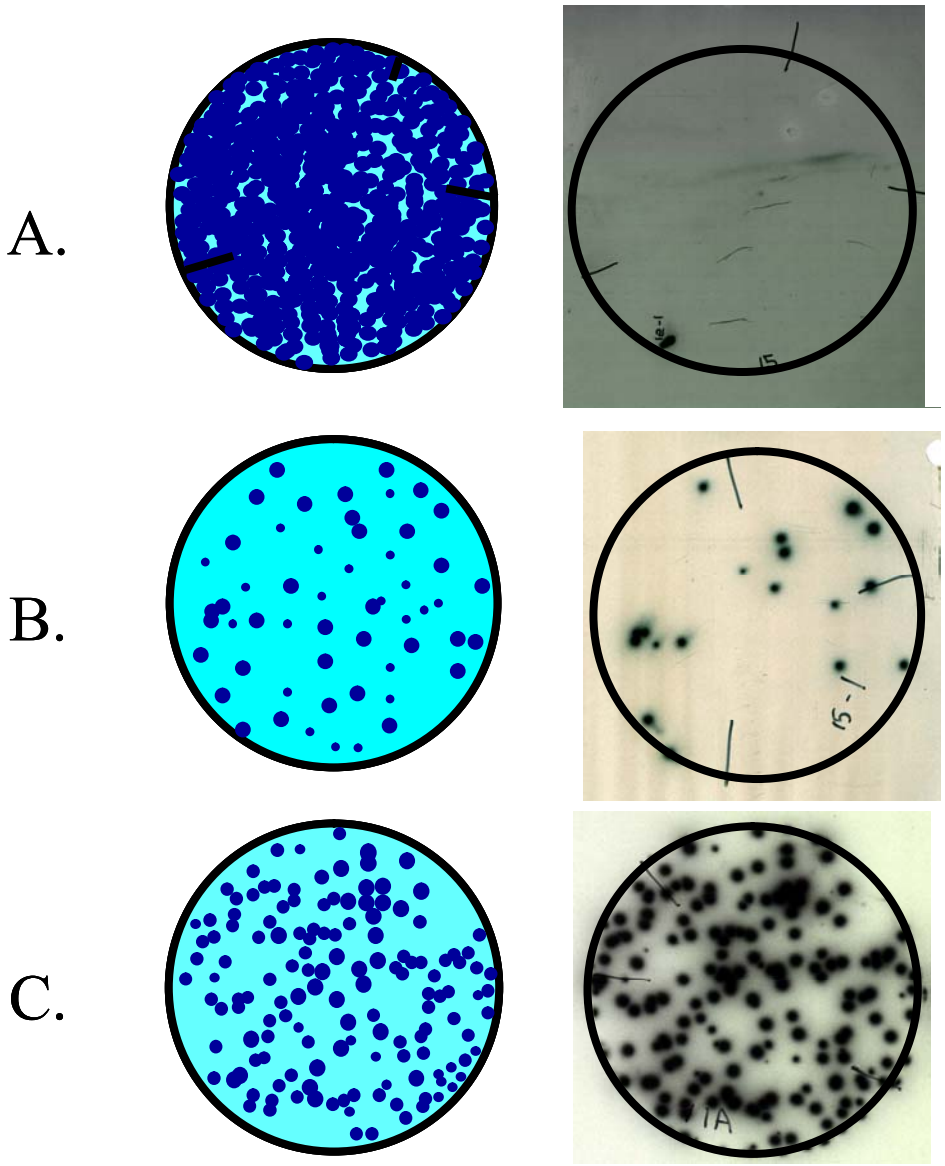


Figure 2-3. Plaque lift screen of the λ genomic library. A) First screen. 20,000 plaques per plate, one hybridizes to the 5' probe. B) Second screen. Plug from first screen is diluted to 200 plaques per plate, ½ the plaques hybridize to the 5' probe. C) Third screen. Individual plaque picked from second screen diluted to 200 plaques per plate, all plaques hybridize to the 5' probe

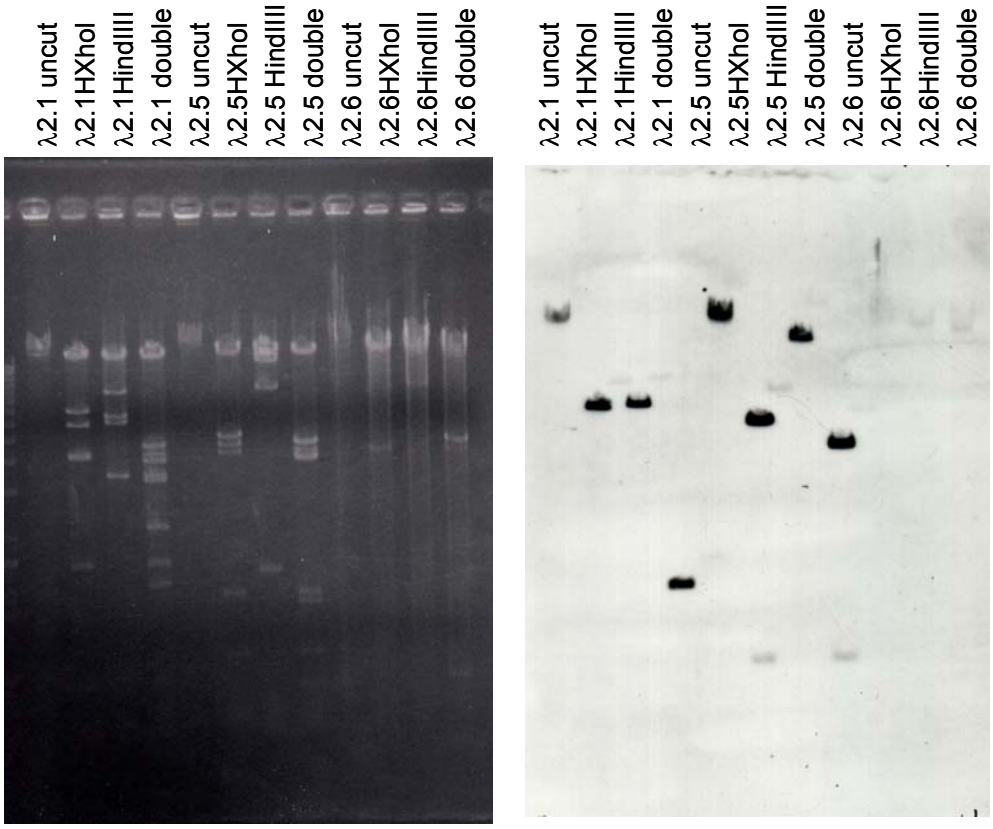
Determination of Overlapping Clones

In order to determine which clones overlapped and spanned the HKα2 gene, Southern analysis was performed on digested DNA from each clone. Figure 2-4 is an example of a Southern analysis showing that HKα2.1 and HKα2.5 overlap in the region of the 5' probe. HKα2.1, HKα2.5 and HKα2.6 DNA were digested with *Xho*I and *Hind*III individually and in combination. The digests were run on a 1% agarose gel and visualized with ethidium bromide (Figure 2-4A). The DNA was transferred to nylon membrane and hybridized to the 5' probe. The membrane was washed and exposed to autoradiograph film (Figure 2-4B). In the lanes representing HKα2.1 and HKα2.5 a single band appears that hybridizes to the 5' probe while in the lanes representing HKα2.6 no band appears. These data clearly show that HKα2.1 that was isolated using the 5' probe and HKα2.5 that was isolated using the mid probe overlap in the region of the 5' probe. HKα2.6, that was isolated using the mid probe, does not extend to the 5' probe. A similar analysis was carried out for the remainder of the λ clones and with all three probes. It was determined that three clones, HKα2.1, HKα2.5 and HKα2.8 spanned a majority of the HKα2 gene, but a gap existed in between HKα2.5 and HKα2.8. Genomic PCR was carried out in order to obtain the missing portion of the gene (see below).

λHKα2.1 Sequence

Clone HKα2.1 hybridized to the 5' probe and contains approximately 14Kbp of sequence upstream of the HKα2 gene. Appendix A contains all of the known sequence from the rabbit HKα2 gene. The 6300bp *Xho*I fragment subcloned from λ HKα2.1 is represented in base pairs 1 – 6298. This sequence was used to determine potential

promoter and regulatory elements. The complete analysis of this sequence is discussed in Chapter three of this dissertation.



(A) Ethidium bromide stained Agarose gel

(B) Autoradiograph of gel probed With the 5' probe

Figure 2-4. Southern analysis to determine overlapping clones. (A) Each λ clone was digested as indicated and run on a 1% agarose gel. The DNA was visualized by staining the gel with ethidium bromide. (B) The DNA was transferred to nylon membrane and probed with the 5' probe (B).

λHKα2.5 Sequence

Clone HKα2.5 hybridized to both the 5' and the mid probes suggesting that it contained many of the 5' exons for the HKα2 gene. ICBR high through-put sequencing of the ends of 48 plasmids that were subcloned from HKα2.5 yielded 144 sequences.

The M13 forward primer (5'-GTAAAACGACGGCCAG-3') sequencing reaction was performed twice and the M13 reverse primer (5'-ACAGGAAACAGCTATGAC-3') sequencing reaction was performed once. The average read-length for the reactions was 307 bases. ICBR used a computer alignment algorithm to assemble the sequences into 10 contiguous fragments labeled 1-10 based upon size. Using the $HK\alpha_{2a}$ cDNA and subclones in which sequence from opposite ends mapped into different contigs, nine of the ten sequences were placed in order. Figure 2-5 depicts the nine sequences and the subclones that spanned the gaps. In order to obtain the remainder of the HKα2.5 sequence, the gaps between the contigs were filled with additional sequence from the indicated plasmids as well as with sequence directly from the HKα2.5 clone. Table 2-3 lists the DNA template and the primers that were used to complete the sequence. Appendix A contains the known sequence of the rabbit HKα2 gene. The sequence determined from HKα2.5 overlaps with the sequence from pDZ10 and is represented in base pairs 4616 to 19766.

λHKα2.8 Sequence

Clone HKα2.8 hybridized to the 3' probe. Partial sequences were determined for this clone. The purpose was to determine the precise intron/exon boundaries of the remainder of the gene as well as the 3' end of the gene. The partial sequence revealed that HKα2.8 contained sequence from exons 15-23. The sizes of the exons and introns identified are listed in Table 2-5. A portion of the sequence of exon 25 is shown in Figure 2-7. The red bases represent the 3' end of the $HK\alpha_{2a}$ cDNA cloned by Fejes-Toth et al. (12). Just upstream of the cDNA end there is a poly A signal sequence (blue) and just downstream there is a T-rich region of DNA. It seems likely that this poly A signal

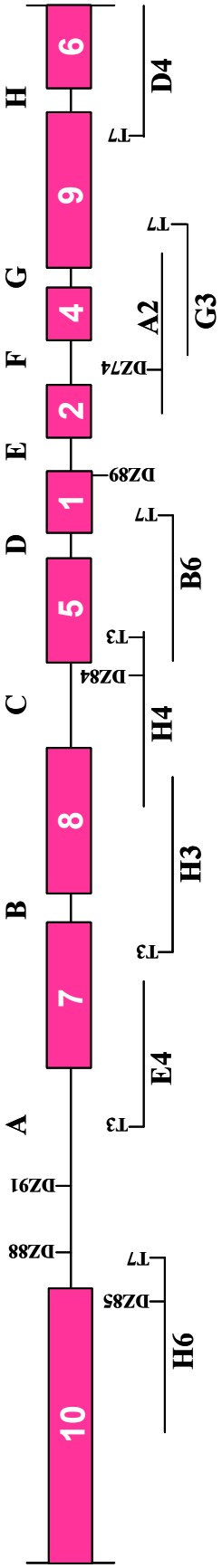


Figure 2-5. Clone λHKα2.5. Pink boxes represent 9 continuous sequences numbered according to size. Letter A-H represent the gaps between the continuous sequences. Lines below represent plasmid used to fill the gaps. Perpendicular lines represent primers used for sequencing.

represents one true end to the HKα2 mRNAs. Campbell et al. (6), however, cloned a cDNA of $HK\alpha_{2a}$ and $HK\alpha_{2c}$ that was slightly longer at the 3' end (green). Just downstream of this cDNA end there are two possible poly A signals and T-rich region of DNA. Therefore it seems likely that one of these two poly A signals, or both represent alternative ends to the $HK\alpha_{2a}$ and $HK\alpha_{2c}$ mRNAs. The partial sequence obtained from HKα2.8 is located in Appendix A.

```
301  AGGTTTTTTT TTTTAATAA AAGATGTTTT TAAGTAAAAT GTTTTATGAA
351  ACAAAATCTA ATTGTGATGT TTTACTTAAT TCAAGTTTTT CCAGAGGCAG
401  GCACGGAAAA TACCAAAAAA ATAAAATAAA ATAAGATTCT GGGTTTTTTT
451  TCTTTTTTGC TCCTTCTGGT CATTTTCTTT ACACACAGAG TGTCTGGAAA
501  TACAGGCTTT TCCTCGTGAG TGCTTCCCGC ACCTGTGCCC CCTCCCCCCC
```

Figure 2-6. Partial sequence from λ clone HKα2.8. Blue represents 3 possible poly A signals. Red represents the last three bases of the Fejes-Toth cDNA for $HK\alpha_{2a}$. Green represents the last three bases of the Campbell cDNA for both $HK\alpha_{2a}$ and $HK\alpha_{2c}$.

Completion of the HKα2 Gene Sequence

The complete sequencing of λ clone HKα2.5 and the partial sequence of λ clone HKα2.8 revealed that three HKα2 exons were not contained in either clone. Therefore, genomic PCR was performed on RCCT28A cell DNA in order to amplify four DNA fragments that contained the missing exons. Primer set DZ98/BC231 amplified a band of approximately 4000 base pairs. This fragment contained a portion of intron 11 and the 5' boundary of exon 12. Primer set DZ99/DZ100 amplified a band of approximately 1500 base pairs. This fragment contained the 3' boundary of exon 12, intron 12 and the 5'

boundary of exon 13. Primer set DZ80/DZ101 amplified a band of approximately 700 base pairs. This fragment contained the 3' boundary of exon 13, intron 13 and the 5' boundary of exon 14. Primer set DZ81/DZ95 amplified a band of approximately 4000 base pairs. This fragment contained the 3' boundary of exon 14 and a portion on intron 15. The sequences obtained from these PCR products are located in Appendix A.

Discussion

The screening of the λ library generated three λ clones that contained 87% of the HKα2 exons and 65% of the HKα2 gene. Most importantly, clone HKα2.1 hybridized to the 5' probe and contains approximately 14kbp of sequence upstream of the gene. Obtaining this clone was a necessary first step in the study of the regulation of the HKα2 gene, which is the subject of the remaining chapters of this dissertation. Additionally, the genomic organization of the rabbit HKα2 gene was determined using the complete sequence of HKα2.5, the partial sequence of HKα2.8 and the genomic PCR fragments that spanned the gap between the two λ clones. HKα2.5 contained 15150bp of gene sequence including the HKα2 exons 1-11, HKα2.8 contained exons 15 – 23, and the PCR fragments contained exons 12, 13, and 14. The entire gene spanned approximately 30Kbp of genomic DNA. It is notable that intron 11 and intron 14 are approximately 4200bp each. The mid probe hybridizes to exon nine and the 3' probe hybridized to the 3' UTR. The size of the intervening DNA (18Kbp) is the likely reason why a clone containing these three exons was not obtained when the bacteriophage λ library was screened.

Table 2-4 compares the sizes of the exons and introns of the rabbit HKα2 with those of the rat HKα2 gene (ICBR database), the mouse HKα2 gene (52), and the human

ATP1AL1 gene (45). The exon sizes for the four genes are absolutely identical except for the three 5' exons one, two and four. It is not suprising that the 5' end of the gene is the most variable since it is in this region where the rabbit and rat genes can undergo alternative splicing to create HKα_{2c} and HKα_{2b} while the mouse and human gene apparently do not. The controversy over whether or not these genes are homologous was partially resolved by a distance analysis of the HKα and NaKα subunit proteins (8). The analysis showed that the three HKα2 proteins were more closely related to each other than to any of the other X^+,K^+ - ATPase α subunits suggesting that they are homologous.

Table 2-4. Exon and intron sizes for the known HKα2 genes.

Exon	Rabbit[a]	Rat[b]	Mouse[c]	Human[d]		Intron	Rabbit[a]	Rat[b]	Mouse[c]	Human[d]
1	208	287	262	195		1	567	677	658	(700)
2	141	153	150	159		2	2702	2133	2286	(2300)
3	60	60	60	60		3	3061	2511	2321	(2900)
4	204	201	201	204		4	1049	746	762	738
5	114	114	114	114		5	1032	742	726	937
6	135	135	135	135		6	113	124	129	123
7	118	118	118	118		7	173	233	227	258
8	269	269	269	269		8	854	939	970	1187
9	199	199	199	200		9	141	144	134	157
10	110	110	110	110		10	2282	1324	1269	(1700)
11	135	135	135	135		11	(4200)	4231	2010	(4200)
12	135	193	193	193		12	1600	1484	1471	(1600)
13	176	176	176	176		13	600	834	639	(900)
14	137	137	137	137		14	4200	1599	1597	(4200)
15	151	151	151	151		15	365	*	1419	557
16	169	*	169	169		16	88	*	90	87
17	155	*	155	155		17	(700)	1425	1311	(1900)
18	124	124	124	124		18	172	184	195	193
19	146	145	146	146		19	445	594	590	(600)
20	134	134	134	134		20	167	174	168	195
21	103	102	102	102		21	(1155)	434	388	431
22	92	92	92	92		22	87	137	161	83
23	658	658	658	905		23	-	-	-	-

Sources: a this dissertation, b NCBI database, c (52), d (45).Notes: () indicates introns with sizes determined by estimating the size of restriction fragments. * indicates sized that could not be determined due to incomplete database sequence.

The exon/intron sizes, compared in Table 2-4, support the exisiting evidence and confirm that these four genes are homologous and were derived from a common ancestor.

CHAPTER 3
MAPPING THE TRANSCRIPTION START SITES FOR THE HKα2 GENE

The second specific aim of this dissertation was to map the transcription start sites for the two alternative mRNAs produced by the HKα2 gene. The determination of the transcription start sites was an important step in characterizing the HKα2 gene for two reasons. The main goal of our study was to initiate an investigation of the regulation of the rabbit HKα2 gene. The core promoter and regulatory elements directing transcription from the gene are likely to be found just 5' of the transcription start sites. Additionally, there was some controversy over the existence of $HK\alpha_{2c}$. Identification of the transcription start site for $HK\alpha_{2c}$ would resolve this controversy.

The first experimental goal was to identify potential promoter and regulatory elements 5' of the transcription start sites. The 5' ends of the cDNAs for $HK\alpha_{2a}$ and $HK\alpha_{2c}$ were previously identified by studies that used 5' Rapid Amplification of cDNA Ends (RACE) (5). This method, however, is not likely to determine the true transcription start site. 5'RACE uses a reverse transcription reaction to extend a primer annealed near the 5' end of the mRNA to the true end of the mRNA. The reverse transcriptase reaction often terminates before reaching the absolute end of the RNA and the cloned cDNA will therefore end 3' of the true transcription start site. In fact, the $HK\alpha_{2a}$ 5' RACE performed by two independent groups produced two different 5' ends. Campbell et al. (6) obtained a 5'UTR of 39 bp, while Fejes-toth et al. (12) obtained a 5'UTR of 190 bp. Furthermore, Campbell et al. was able to obtain a 5' cDNA end corresponding to the

splice variant HKα_{2c} while Fejes-toth et al. did not. The RNase protection assay used in this dissertation more accurately determines the transcription start sites because it does not rely on primer extension. This chapter describes the construction of RNase protection probes using the analysis of the sequence from clone HKα2.1 and the use of the probes in mapping the transcription start sites the for HKα2 gene.

The second experimental goal in mapping the transcription start sites for the HKα2 gene was to confirm the existence of the HKα_{2c} transcript. Our laboratory previously showed that the HKα_{2c} transcript and protein were present in both rabbit kidney and colon (6). A second laboratory, however, was unable to detect the HKα_{2c} transcript while using similar detection techniques (12). This chapter describes the successful mapping of the transcription start sites for both the rabbit HKα_{2a} and HKα_{2c} transcripts.

Materials and Methods

Analysis of λ Clone HKα2.1

Clone HKα2.1 was identified by hybridization to the 5' probe (see Chapter 2). A 6.3Kbp *Xho*I fragment from HKα2.1 was subcloned into pBS (creating pDZ10) and completely sequenced. The sequence was analyzed for elements commonly found at eukaryotic promoters.

RNase Protection Assay

The RNase protection assay was used to map the transcription start sites for HKα_{2a} and HKα_{2c}. This assay was performed in three steps. First, an antisense radioactive RNA probe was created from a fragment of genomic DNA likely to contain the transcription start sites as well as sequence 5' of the start sites. Second, the

radioactive probe was annealed to the specific mRNA thereby protecting it from RNase digestion. And third, the protected fragment was run on an acrylamide gel along with a sequencing ladder of known size. Once the size of the protected fragment was determined, the transcription start site could be mapped on the genomic DNA sequence.

In order to create the RNA probe, a 1.1Kbp *Hinc*II fragment likely to contain both transcription start sites was cloned into the pGEM vector (Promega, Inc) creating plasmid pDZ12. The pGEM vector contained the SP6 polymerase promoter that was used in an *in vitro* transcription reaction to create a radioactive RNA probe. Preliminary experiments showed that one probe could not be used to map both start sites. Therefore, pDZ12 was modified to create one plasmid with the region likely to contain the start site for HKα_{2a} (pDZ44) and a second plasmid with the region likely to contain the start site for HKα_{2c} (pDZ43). Figure 3-1 depicts the construction of these two plasmids, the expected sizes of the full-length probes, and the predicted size of the protected fragment based on the 5' end of the cDNAs. Plasmid pDZ44 was created by digesting pDZ12 with *Sac*II and *Sph*I, filling in the vector ends with Klenow DNA polymerase, and religating the vector. A 450bp fragment containing the HKα_{2c} start site was removed from the vector creating a shorter HKα_{2a} RNA probe (Figure 3-1A). The full-length probe was 235 base pairs and the protected fragment was expected to be approximately 87 base pairs. Plasmid pDZ43 was created by digesting pDZ12 with *Xmn*I and *Hinc*II. The resulting 168bp fragment contained the HKα_{2c} transcription start site and had blunt ends. The fragment was cloned directly into the *Hinc*II site of pGEM vector (Figure 3-1B). This plasmid produced a 182 base pair full-length probe and an expected 105 base pair protected fragment.

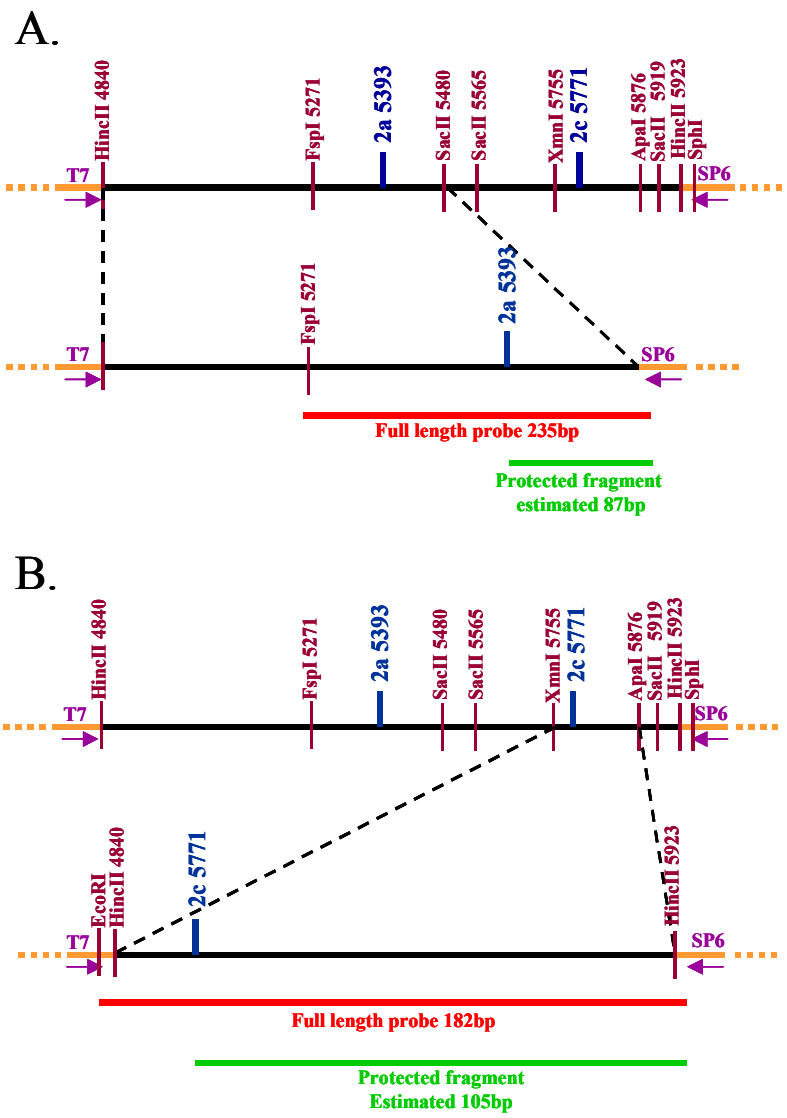


Figure 3-1. Construction of RNase protection probes for HKα_{2a} (A) and HKα_{2c} (B). Orange represents the cloning region of pGEM vector (Promega, Inc.) Black represents 1.2Kbp fragment of rabbit genomic DNA subcloned into the pGEM vector. Brown represents pertinent restriction sites. Blue represents the 5' ends of the HKα_{2a} and HKα_{2c} cDNAs. Purple represents the binding sites for T7 and SP6 polymerase. Red represents the expected sizes of the in vitro transcription products. Green represents the sized of the protected fragments estimated based on the end of the cDNAs.

Each plasmid was used in the MAXIscript *in vitro* transcription kit (Ambion, Inc. Cat. # 1308) in order to create radioactive RNA probes. 1μg of the plasmid DNA was added to the *in vitro* transcription reaction (2μl 10X transcription buffer, 1μl 10mM each ATP, CTP and GTP, 2.5μl 10mCi/ml [^{32}-αP] UTP, 2μl SP6 polymerase, and dH_2O up to 20μl). The reaction was incubated at 37°C for 10 minutes. 1μl of DNaseI was added and the reaction was incubated at 37°C for an additional 15 minutes. After incubation, the entire reaction was loaded on a 5% acrylamide gel and run at 300 volts for 30 minutes. The probe fragment was visualized by wrapping the gel in saran wrap and laying down a piece of Polaroid type 57 high-speed film. When developed, a white band appeared on the film at the position of the probe. The film was aligned with the gel, the band was excised, and gel fragment was pressed through a 1mm syringe containing 250μl of elution buffer (.5M NH_4Acetate, 1mM EDTA, 0.2% SDS). The probe was eluted from the gel fragments by incubation of the mixture at 37°C for one hour. The specific activity of the probe was measured on using a Beckman LS3801 scintillation counter. An aliquot of probe containing a specific activity of 8 x 10^4cpm was used in the ribonuclease protection assay.

The RPAIII kit from Ambion, Inc. (Cat # 1414) was used for the ribonuclease protection assay. The probe was co-precipitated with 10μg of rabbit colon total RNA by bringing the volume of the probe and RNA to 100μl, adding 10μl NH_4OAc and 250μl 100% ethanol, incubating at -20°C for 15 minutes and centrifuging at 15,000 rpm for 15 minutes. The pellet was air dried, resuspended in hybridization solution (Ambion, Inc.), heated to 95°C for 5 minutes and incubated at 42°C overnight. During this time, the probe annealed to its specific mRNA. The next day 1.5μl of RNaseA/RNase T1 cocktail

was diluted 1:100 in RNase Digestion buffer (Ambion, Inc.), added to the hybridization reaction and incubated at 37°C for one hour. During this incubation, all the single stranded nucleic acids were degraded and only the double stranded protected fragment remained intact. After the incubation, the protected fragment was precipitated by adding 225µl of RNase inactivation/precipitation buffer (Ambion, Inc.), incubating the tube at -20°C for 15 minutes and centrifuging the tube at 15,000 rpm for 15 minutes. The pellet was air dried and resuspended in 5µl of gel loading buffer (95% formamide, .025% xylene cyanol and bromophenol blue, 18mM EDTA, .025% SDS).

In order to visualize the protected RNA fragment, the each sample was loaded onto a 6% polyacrylamide sequencing gel along with a sequencing reaction of a known size. The sequencing reaction was carried out using Sequenase 7-deaza-dGTP DNA Sequencing Kit (USB, Cat # 70990) with the control M13 single stranded DNA provided with the kit. The M13 single stranded template (1.0µg) was annealed to the –40 primer (0.5pM) by mixing with 2µl of sequenase reaction buffer (200mM Tris HCl pH 7.5, 2mM DTT, 0.1mM EDTA, 50% glycerol) and dH2O up to 10µl and then heating to 65°C for two minutes. After cooling to room temperature, the labeling reaction (annealed DNA, 0.1M DTT, 2µl labeling mix (1.5mM 7-deaza-dGTP, 1.5µM dCTP, 1.5µM dTTP), 0.5µl [α-^{32}P]dATP, 2µl Sequenase polymerase (1U/µl)) was incubated at room temperature for five minutes. The reaction was terminated by adding 3.5µl of the labeling mixture to each of four pre-warmed termination tubes. All termination tubes contained 80mM of each 7-deaza-dGTP, dCTP, dATP, dTTP, and 50mM NaCl. Additionally, each tube contained 80µM of either ddGTP, ddATP, ddTTP, or ddCTP. The termination reaction was incubated at 37°C for five minutes. The termination reaction was stopped with 4µl

of stop solution (95% dien, 20mM EDTA, 0.05% xylene cyanol). The sequencing reactions were run on a 5% polyacrylamide gel along with the protected fragments from the RNase protection assay. The gel was run at 65 volts for approximately five hours, dried for two hours and exposed to autoradiograph film overnight at -80°C.

Results

Analysis of Clone HKα2.1

In order to determine the region most likely to contain the HKα2 gene promoter, the sequence of the 6.3Kbp *Xho*I fragment was analyzed for characteristics common to eukaryotic promoters. First, it was determined that the 3' end of the sequence contains a CpG island. Figure 3-2 is a graph that shows the number of CpG dinucleotides found in 50 base pair windows of the sequence. Most of the sequence contained very few CpG dinucleotides, but there was a clear peak in the number of CpGs at the 3' end. Next, the computer program TFSearch was used to determine if any possible transcription factor binding sites were present along the sequence. The results showed a wide variety of possible binding sites. Appendix B contains the entire search results. Figure 3-3 is a cartoon depicting the possible transcription factor binding sites that seemed most rational based on previously known data about the regulation of the HKα2 gene (see discussion). These include a TATA-like element, five SP family member binding sites, a downstream promoter element, a cyclic-AMP response element (CRE) and a steroid response element (SRE).

Transcription Start Sites

The transcription start sites for $HK\alpha_{2a}$ and $HK\alpha_{2c}$ were determined using the RNase protection assay. In each case two protected fragments were observed. Figure 3-4

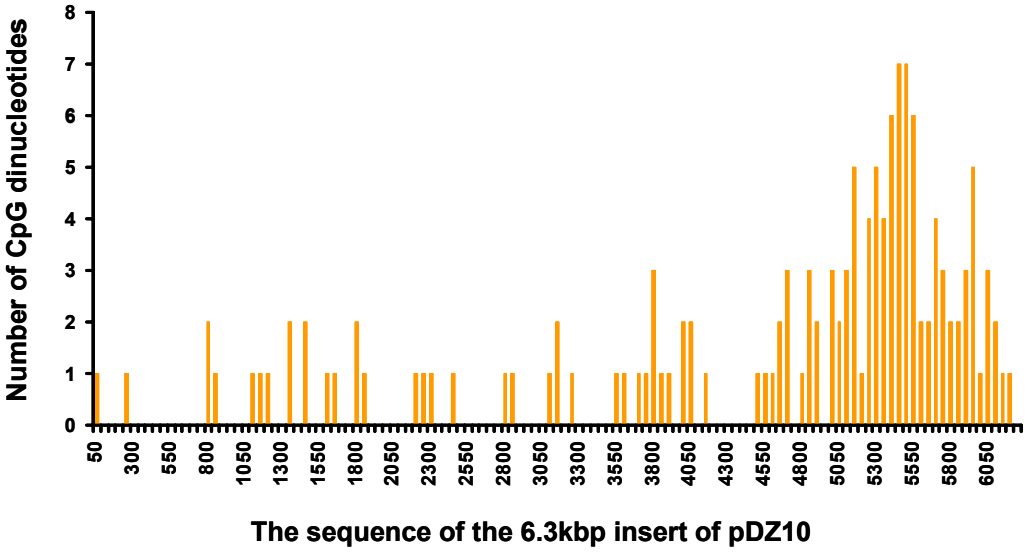


Figure 3-2. CpG dinucleotide analysis of subclone pDZ10.

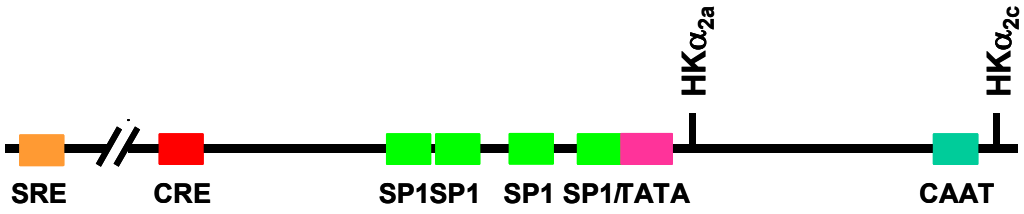


Figure 3-3. Putative transcription factor binding sites determined by TFSearch. The 5' ends of the HKα_{2a} and HKα_{2c} cDNAs are indicated. Blue represents a CAAT box. Pink represents a sequence with weak homology to the TATA box. Green represents possible binding sites for SP family members. Red represents a cyclic AMP response element. Orange represents a sterol response element. Slashes indicate a break in the sequence of approximately 500bp.

is an example of a polyacrylamide sequencing gel in which the protected fragments for $HK\alpha_{2a}$ and $HK\alpha_{2c}$ were run next to a sequence of known size (M13 single stranded DNA). By comparison to the known sequence, it was determined that the two protected fragments for $HK\alpha_{2a}$ were 94 and 95bp. These fragments correspond to the bases of genomic DNA indicated in Figure 3-4. They are 10 and 11 bases upstream of the cDNA end obtained by Fejes-toth et al. (12), making the 5'UTR for $HK\alpha_{2a}$ 200 and 201 base pairs. Similarly the $HK\alpha_{2c}$ protected fragments were 116 and 118bp and correspond to the bases of genomic DNA indicated in Figure 3-4. They are five and seven bases upstream of the cDNA end obtained by Campbell et al. (6), making the 5' UTR for $HK\alpha_{2c}$ 203 and 205 base pairs. For the remainder of this dissertation, the first transcription start site for $HK\alpha_{2a}$ was designated as +1 and all other positions are designated relative to that transcription start site. The $HK\alpha_{2c}$ transcription start sites were therefore designated +382 and +384.

Discussion

Mapping the transcription start site was an important step in the characterization of the HKα2 gene promoter. The RNase protection assay was used to map the transcription starts sites for $HK\alpha_{2a}$ and $HK\alpha_{2c}$. The results of this assay yielded several interesting observations. First and foremost, the protected fragments observed with the $HK\alpha_{2c}$ probe confirmed the existence of the transcript in rabbit colon. Second, each probe yielded two protected fragments. Third, the sequence upstream of the two start sites contained a variety of possible core promoter elements. And finally, further upstream, the sequence revealed several possible *cis* acting regulatory elements.

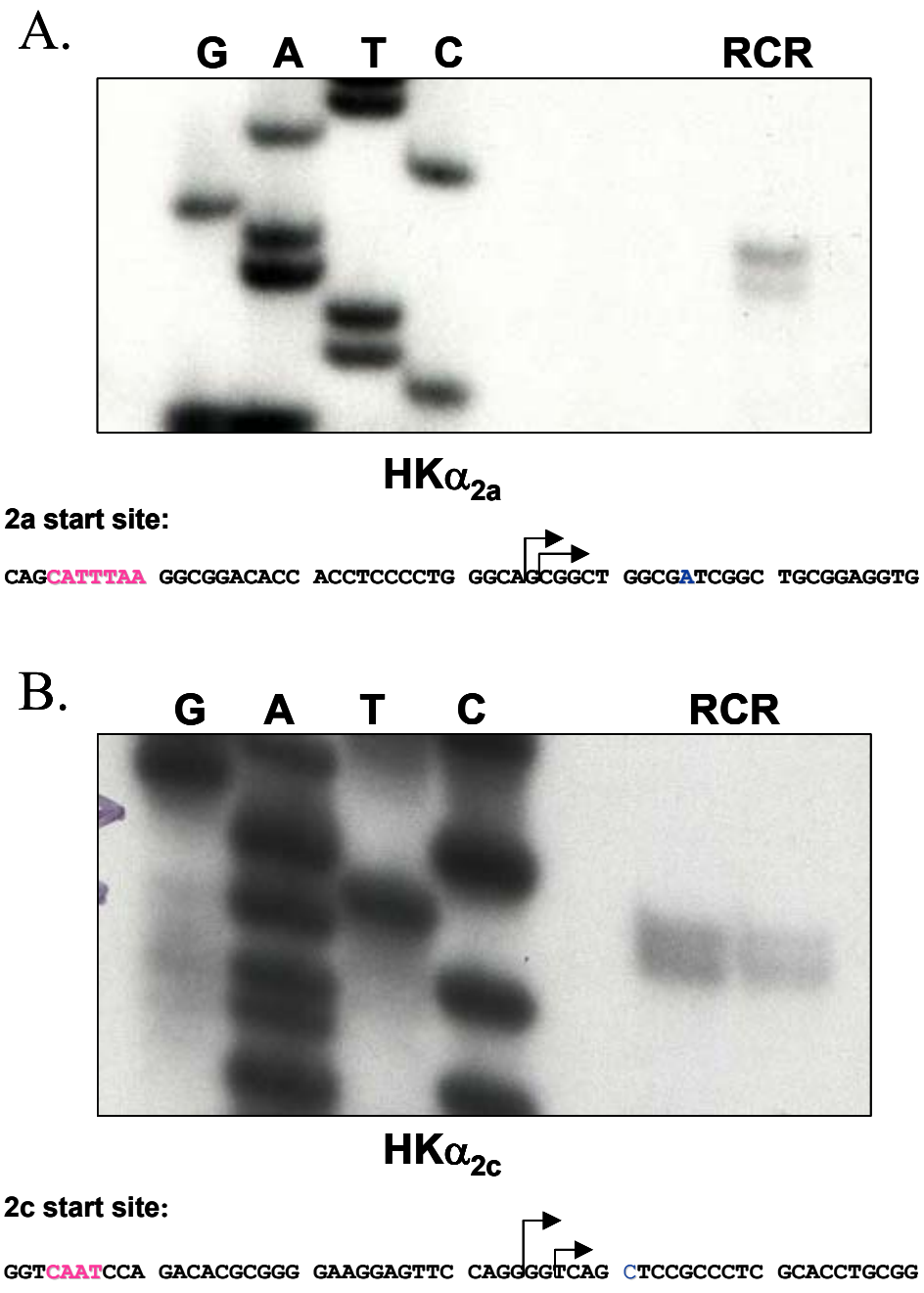


Figure 3-4. RNase protection assay for HKα_{2a} (A) and HKα_{2c} (B). GATC represents those nucleotides for the M13 control sequence. RCR represents the protected fragment from the RNase protection assay performed with rabbit colon RNA. A portion of the genomic sequence 5' of the HKa2 gene is shown below each figure. Arrows indicate the position of the transcription start sites. Pink represents putative core promoter elements and blue represents the 5' end of the respective cDNAs.

The RNase protection assay was used to map two transcription start sites for $HK\alpha_{2c}$. Previously, Fejes-Toth et al. (12) questioned the existence of the alternative transcript identified by Campbell et al. (6). Fejes-Toth states that their 5' RACE experiments generated one amplicon corresponding to the 5' end of $HK\alpha_{2a}$. They go on to state that the convergence of their data from that of Campbell may be due to the fact that the 5' RACE of Campbell was carried out in tissue culture cells instead of rabbit tissue and furthermore that Campbell et al. was unable to detect $HK\alpha_{2c}$ mRNA in the renal cortex. Although these statements are true, Fejes-Toth failed to recognize that $HK\alpha_{2c}$ mRNA was detected in rabbit colon, and $HK\alpha_{2c}$ protein was detected in both rabbit renal cortex and colon. These facts alone substantiate that $HK\alpha_{2c}$ was not an artifact of working with tissue culture cells. The RNase protection assay performed in this dissertation, however, confirmed the existence of the $HK\alpha_{2c}$ transcript at least in rabbit colon.

There are several explanations for the fact that each RNase protection probe yielded two protected fragments in very close proximity for both $HK\alpha_{2a}$ and $HK\alpha_{2c}$. One possibility is that RNA polymerase had difficulty lying down in an exact position and starting transcription at a precise site because the GC content of a region is high. The genomic sequence surrounding the $HK\alpha_{2a}$ and $HK\alpha_{2c}$ start sites are 67% and 68% GC respectively (Figure 3-4). Additionally, the putative TATA box upstream of the $HK\alpha_{2a}$ start site (see below) had very weak homology to the consensus TATA box. It is therefore likely that the putatitive core promoter elements surrounding the TATA box play a role in transcription initiation and my not precisely position RNA polymerase. A second explanation for the two protected fragments comes from the RNase protection

technique itself. It is possible to observe a fragment slighty longer than the true protected fragment because the RNases used in the assay (A and T1) are endonucleases and may leave bases on the end of a protected fragment. It is also possible to get a protected fragment shorter that the true fragment because the end of the RNA-RNA duplex may occasionally separate. It might be possible to distinguish between these possibilities with a primer extension assay. This assay, however, also has inherent problems with distinguishing one correct start site as it relies on reverse transcription similarly to the 5' RACE (see Chapter 3 Introduction).

The putative core promoter elements found upstream of the HKα_{2a} transcription start site were a weak TATA box at -31, four SP family binding sites at –47, -102, 154, and -170, a downstream promoter element (DPE) at +17, a TFIIB responsive element at -42, and a CpG island that extends from –49 to +504. Additionally, directly upstream of the HKα_{2c} transcription start site a single CAAT box at +351 was observed (31 bases upstream). The fact that this is the only promoter-like element immediately upstream of the HKα_{2c} transcription start site suggests that the HKα2 gene has one core promoter that is able to direct transcription from the two alternative starts. The CAAT box may be important in directing the initiation of transcription form HKα_{2c}. The weak TATA element (CATTTAA) may be serving as a binding site for the general transcription factor TFIID. Additionally, the other core promoter elements found surrounding this element may serve to stabilize the preinitiation complex at the weak element. The function of the TATA-like element is further investigated in Chapter 4 of this dissertation.

Further upstream of the transcription start sites, a possible cyclic AMP response element (CRE) at –187 and a possible sterol response element (SRE) at –852 were

identified. There is evidence *in vivo* that cyclic AMP is increased in hypokalemic rats (24). The CRE could provide a mechanism for upregulating the HKα2 gene. Additionally, there is evidence from our laboratory that aldosterone may upregulate the HKα2 gene (5). The SRE may provide a binding site for aldosterone and its hormone receptor.

In summary, the transcription start sites for HKα_{2a} and HKα_{2c} were mapped using the RNase protection assay and rabbit colon total RNA. HKα_{2c} was confirmed as a transcript in rabbit colon. Upstream of the transcription start site several putative transcription factor binding sites were observed. This work is the first analysis of the rabbit HKα2 gene 5' of the transcription start site. The sequence contains many putative transcription factor binding sites. Additionally, determination of the sequence allowed for the design of future experiments regarding the regulation of the HKα2 gene.

CHAPTER 4
REPORTER GENE ANALYSIS OF THE HKα2 GENE PROMOTER

The third specific aim of this dissertation was to analyze the HKα2 gene promoter using a reporter gene system. At the time that this study was proposed, there was *in vivo* evidence that expression of HKα2 gene products was regulated by a variety of cellular conditions including ion concentration, acid-base balance and hormones. There was, however, nothing known about the mechanisms by which the expression was altered. cDNAs for rabbit, rat, guinea pig and human had been identified, but only the human ATP1AL1 gene was known. There have been no studies undertaken to determine promoter elements for the human gene. Recently, the mouse gene was identified and a reporter gene analysis of its 5' flanking region was carried out in mouse inner medullary collecting duct cells (mIMCD3) (52). In their reporter gene experiments, Zhang et al. found that thier longest deletion construct had significant luciferase activity and the deletion of bases -177 –7265 had little to no effect on activity. The authors suggest that the core promoter elements as well as positive regulatory elements are located between bases +235 and –177. Although putative regulatory elements were identified in a database search, there were no attempts made to determine the functionality of any specific core promoter or regulatory elements. Additionally, Zhang et al. tested their promoter constructs in outer medullary collecting ducts cells (mOMCD1) and medullary thick ascending limb cells (ST-1). All of the deletion constructs had significant activity in the second collecting duct cell type (OMCD) but little to no activity in the ST-1 cells. The results suggested that either positive regulatory elements are absent in ST-1 cells or

that negative regulatory elements, including a closed chromatin structure, are present in ST-1 cells.

The experiments described in this chapter used the luciferase reporter gene assay to analyze rabbit HKα2 promoter constructs in a rabbit cortical collecting duct cell line (RCCT28A). It was determined that λ clone HKα2.1 contained the HKα2 gene promoter (see Chapter 2). Portions of the 6.3Kbp *Xho*I fragment from HKα2.1 were cloned in front of the luciferase reporter gene in the pGL3 basic vector (Promega, Inc.). The constructs were transfected into RCCT28 cells and reporter gene activity was measured. Our goals were to provide the first data regarding the regulation of the rabbit HKα2 gene, to identify possible regulatory elements, and to test the functionality of those elements by mutating specific bases within the identified elements. In this way, important regulatory regions would be identified for future studies.

Materials and Methods

The Promega dual luciferase reporter gene assay (Promega, Inc. Catalog # E1960) was chosen for the promoter analysis. Each promoter construct was cloned in front of the firefly luciferase reporter gene in the pGL3 reporter gene plasmid (Promega, Inc. Catalog # E1751). The plasmids were then transfected into RCCT28A tissue culture cells using the Superfect transfection reagent (Qiagen, Inc. Catalog #301305). The cells were simultaneously transfected with the pRL control plasmid which contained the Renilla luciferase reporter gene driven by the thymidine kinase promoter (Promega, Inc. Catalog # E2241). After 24 hours the cells were lysed and both the firefly luciferase activity and the Renilla luciferase activity were measured using a Berthold Sirius Luminometer. These data were normalized using the Renilla luciferase activity and represented as a

percentage of the highest normalized activity observed. The results identified fragments of DNA 5' of the HKα2 transcription start sites that may play a role in the regulation of HKα2 gene transcription.

Reporter Gene Constructs

Four sets of reporter gene plasmids were constructed. The first two sets were promoter deletion plasmids and the second two sets were mutation plasmids. The first set of deletion constructs contained both the $HK\alpha_{2a}$ and the $HK\alpha_{2c}$ transcription start sites cloned into the pGL3 reporter vector (Figure 4-1). These constructs had little to no luciferase activity. Therefore, a second set of deletion plasmids that contained only the $HK\alpha_{2a}$ start site were created (Figure 4-2). These constructs had varying amounts of activity as expected in a promoter deletion experiment. Based on the data obtained from the deletion analysis, two sets of mutation constructs were created. The first set tested the functionality of two potential repressor elements (Figure 4-3) and the second set tested the functionality of a potential core promoter element (Figure 4-4).

The plasmid pDZ10, which contained the 6.3Kbp *Xho*I fragment of clone HKα2.1, was used to create the first set of deletion constructs (Figure 4-1). A 5259bp *Stu*I/*Xho*I fragment was cloned into the pGL3 vector (pDZ15). This sequence extended from –4339 to +930 and contained the transcription start sites (+1 and +382) and the translation start sites (+200 and +585) for both $HK\alpha_{2a}$ and $HK\alpha_{2c}$. In order to create a plasmid that contained upstream DNA, but did not produce the $HK\alpha_{2a}$ protein, the Quikchange Mutagenesis Kit (Stratagene, Inc. Cat# 200-518-5) was used to mutate the ATG start codon at +200. The primers created for the mutation were DZ24 (5'CTCCAGCGCGACACGTGCCAGGTGTGTGAGG3') and DZ25 (5'CCTCA

CACACCTGGCACGTGTCGCGCTGGAG3'). This plasmid was designated pDZ28. Deletion plasmids were made from pDZ15 and pDZ28 by removing a 3463bp *Nhe*I/*Aat*II fragment, filling in the ends using Klenow DNA polymerase, and religating the vector fragment using T4 DNA ligase. These plasmids were designated pDZ29 and pDZ30 respectively. The construction of pDZ29 and pDZ30 inadvertently placed a potential stop codon in the 5' UTR. In order to create a construct that removed the stop codon, two *Pac*I sites were inserted into pDZ28 and pDZ29 by Quikchange with primer sets DZ41/42 (5'CAGAGAAAGCTGTTAATTAACTCCGTGGAGCAC CATGCAGC3', 5'GCTGCATGGTGCTCCACGGAGTTAATTAACAGCTTTCTCTG3') and DZ43/44

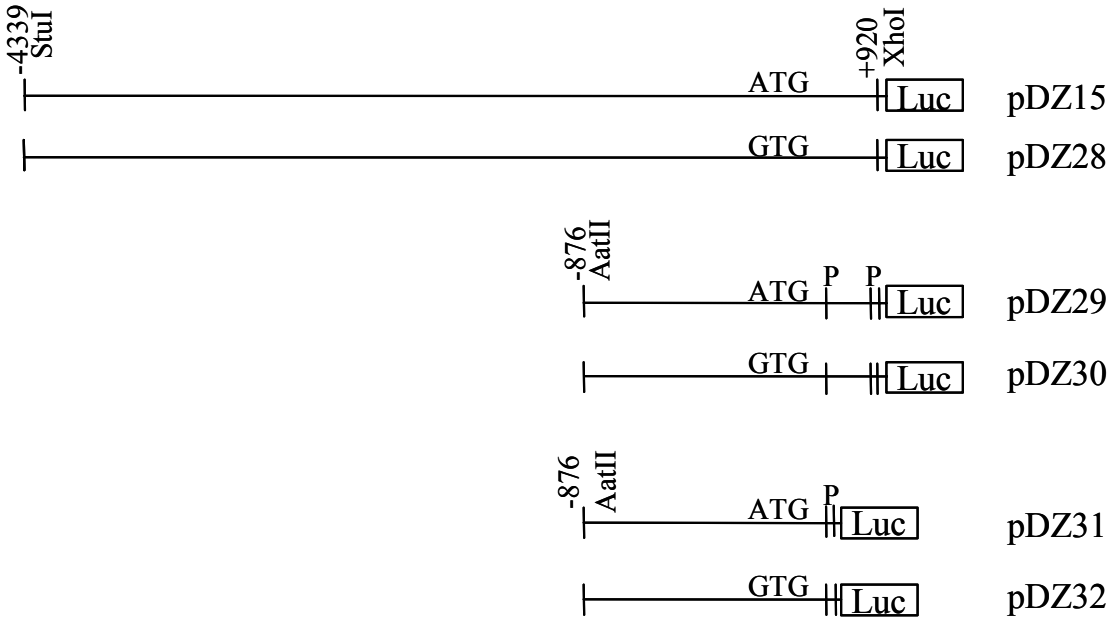


Figure 4-1. Deletion constructs containing the HKα_{2a} and HKα_{2c} transcription start sites. Lines represent the HKα2 gene 5' DNA. Base pair numbers indicate the position of the restriction enzyme recognition site with respect to the HKα_{2a} transcription start site. ATG represent the HKα_{2a} translation start site. GTG represents the mutation created to abolish translation from the HKα_{2a} translation start site. P represents the *Pac*I sites inserted by Quikchange. Luc represents the cDNA for the luciferase reporter gene.

(5’CAGCTTGGCATTCCGGTACTTTAATTAAAGCCACCATGGAAGACGCC3’, 5’GGCGTCTTCCATGGTGGCTTTAATTAAAGTACCGGAATGCCAAGCTG3’). The 85bp fragment was removed by digestion with *Pac*I, and the vectors were religated creating plasmids pDZ30 and pDZ31 (Figure 4-1).

The second set of deletion constructs contained only the transcription start site for HKα_{2a} (Figure 4-2). Plasmid pDZ10 was used as a starting plasmid for these constructs as well. A 5459bp *XhoI*/*Sac*II fragment was cloned into the pGL3 vector (pDZ11). This sequence extended from –5367 to +93. A series of plasmids were then constructed by digestion of pDZ11 with *Nhe*I for the 5’end and a second enzyme for the 3’ end. The overhangs on the vector ends were filled in using Klenow DNA polymerse and the vector was religated using T4 DNA ligase. These plasmids digested with the indicated enzymes were designated pDZ16 (*Stu*I), pDZ20 (*Nde*I), pDZ22 (*Spe*I), pDZ21 (*Msc*I), pDZ19 (AatII), pDZ18 (*EcoR*I), and pDZ23 (*Sma*I). There were no convenient restriction enzymes recognition sequences that could be used to make deletions intermediate to plasmids pDZ21 and pDZ22. Therefore, two plasmids of an intermediate size were created by introducing *Mlu*I sites into pDZ22 by Quikchange (Stratagene, Inc.). Primer set DZ31/32 (5’GGGTAGGGGATGTCACGCGTGGCCAAATGAAGTTG3’, 5’CAACTTCATTTGGCCACGCGTGACATCCCCTACCC3’) introduced an *Mlu*I site at position –2464. Plasmid pDZ26 was then created by digestion with *Mlu*I and *Nhe*I, filling in of the overhangs and religating the remaining vector fragment. Primer set DZ33/34 (5’CTTCTCTGTGCCACGCGTGGCCCAAAAGTTGG3’, 5’CCAACTT TTGG GCCACGCGTGGCACAGAGAAG3’) created an *Mlu*I site at postion –1916.

Digestion and religation of this vector fragment created pDZ27. Furthermore, a plasmid intermediate to pDZ21 and pDZ19 was created by inserting an *Mlu*I site at position –1241. In this case, the *Mlu*I site was introduced as part of a primer set (DZ48 5'**ACGGCT**CCCTGTCCCATAGCCAGAGAATCCC3') used for PCR. The PCR reaction contains10μM of primers DZ48 and DZ5 (5'CTGCACTCTCAGAGTGA AGG3'), 10ng pDZ21, 50μl of Qiagen PCR Master Mix (*Taq* DNA polymerase, Qiagen PCR Buffer with 3mM $MgCl_2$, 400μM each dNTP) and dH_2O up to a volume of 100μl. The PCR conditions were one cycle of 95ºC for 5 minutes, 30 cycles of 95º for 30 seconds, 68ºC for 30 seconds, 72ºC for 30 seconds, and one cycle of 72ºC for 5 minutes. The PCR reaction was run on a 1% agarose gel and visualized with ethidium bromide. The reaction produced a single 700bp band which was gel extracted (Qiagen, Inc.) and cloned into the TOPO cloning vector (Invitrogen, Inc.). In order to create the deletion construct, the TOPO clone was digested with *EcoR*I and *Mlu*I and the overhangs were filled in. The blunt ended fragment was then cloned into the *Sma*I site of pDZ23 in order to create plasmid pDZ36. The shortest construct (pDZ25) was created by performing Quikchange on plasmid pDZ18. Primer set DZ35A/35B (5'CGCGCAGCATTTAACGC GTACAC CACCTCCCC3', 5'GGGGAGGTGGTGTACGCGTTAAATGCTGCGCG3') inserted an *Mlu*I site at –26. Digestion of the plasmid with *Mlu*I and *Nhe*I, filling in the overhangs and religation of the vector fragment completed pDZ25. One final construct was made to ensure that the size of the deletion plasmid was not having and effect on reporter gene activity. A 4.2Kbp *Hind*III fragment of non-specific DNA (from *E. coli* F_1F_0 plasmid pAES9) was ligated into the HindIII site of pDZ18 located at –26. This

fragment made the construct pDZ49 approximately the same size as the largest construct, pDZ11. The sizes of all of these deletion constructs are indicated in Figure 4-2.

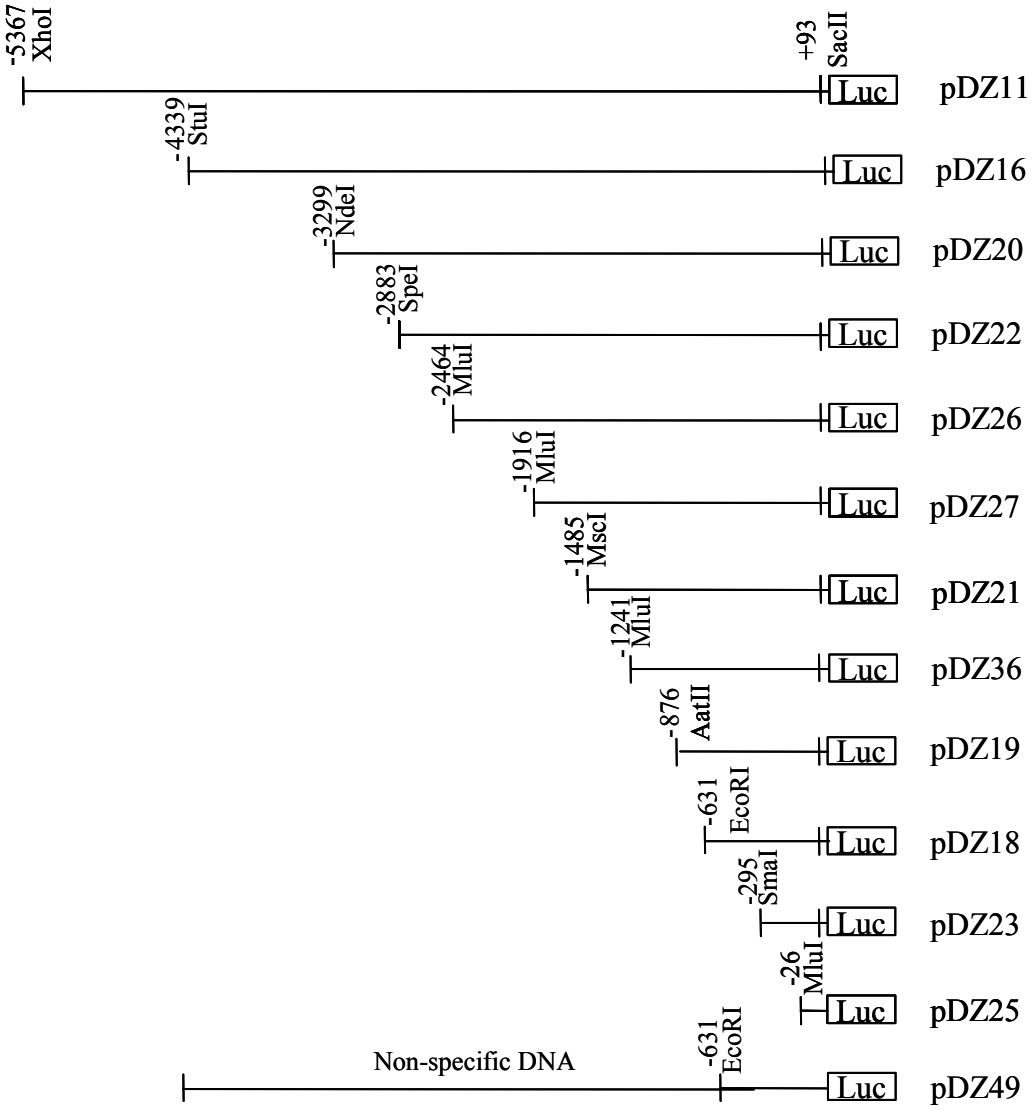


Figure 4-2. $HK\alpha_{2a}$ deletion constructs. Lines represent the HKα2 gene 5' DNA inserted into the pGL3 reporter gene plasmid. Base pair numbers indicate the position of the restriction enzyme recognition site with respect to the $HK\alpha_{2a}$ transcription start site. Luc represents the cDNA for the luciferase reporter gene.

The third set of luciferase reporter gene constructs were created to test the functionality of two putative repressor elements identified by the deletion analysis (see results). Three mutations were made using Quikchange. Primer set DZ55/56 (5'GCAGCACCACGCAGCCCGGGACCATTAATTAAAAACGCTCACTGACCCAG ACCCTCC3', 5'GGAGGGTCTGGGTCAGTGAGCGTTTTTAATTAATGGTCCC GGCTGCGTGGTGCTGC3') mutated the element located at –700. Primer set DZ59/60 (5'GCCCTCCACGCTCACTGACCATTTAAATACATCCCCACCCC TCTCTCC3', 5'GGAGAGAGGGGTGGGGATGTATTTAAATGGTCAGTGAGCG TGGAGGGC3') mutated the element at –680. The two putative elements were close together, so a third primer set, DZ63/64 (5'GGACCATTAATTAAAAACGCTCACTGACCATTTAAA TACATCCCCACCCCTCCTCTCC3', 5'GGAGAGAGGGGTGGGGATGTATTTAAA TGGTCAGTGAGCG TTTTTAATTAATGGTCC3') was used to mutate both elements.

The fourth set of luciferase reporter gene constructs were created to test the functionality of the TATA-like element found at –31 (see results). The second smallest deletion construct that contained the element (pDZ18) and the smallest deletion construct that did not contain the element (pDZ25) were used as positive controls for this set of experiments. Three mutation constructs were made using Quikchange with three primer sets. Primer set DZ53/54 (5'GCGGGGCGCGCAGCGGCCGCGGCGGACACCACC3', 5'GGTGGTGTCCGCCGCGGCCGCTGCGCGCCCCGC3') created a GC box in place of the TATA element (pDZ34). Primer set DZ86/87 (5'GCGGGGCGCGCAGCGAT CGAGGCGGACACCACC3', 5'GGTGGTGTCCGCCTCGATCGCTGCGCG CCCCGC3') created a random sequence in place of the TATA element (pDZ48). And finally, primer set DZ57/58 (5'GCGGGGCGCGCATATAAAGGCGGACACCACC3',

5'GGTGGTGTCCGCC TTTTATATGCGCGCCCCGC3') created a consensus TATA sequence at the same location (pDZ39).

Tissue Culture Cells

RCCT28A cells were chosen for the majority of the reporter gene analyses. These cells were isolated from rabbit cortical collecting duct and transformed with the SV40 virus by Arend et al. (2). They were characterized by their ability to bind several antibodies and by their response to specific hormones. It was determined that RCCT28A cells maintain characteristics most similar to the intercalated cells of the cortical collecting duct. Furthermore, Campbell et al. (6) showed that these cells express the HKα2 mRNA's and proteins. RCCT28A cells were therefore known to contain factors required for expression of the reporter gene driven by the HKα2 gene promoter.

In addition to RCCT28A cells, three other cell types were used for one of the reporter gene assays. The activity of several of the deletion constructs were tested in HIG-82 cells, HEK293 cells and HK2 cells. HIG-82 cells were established by spontaneous transformation of fibroblasts from rabbit soft tissue lining the knee joint (15). They are not likely to express the HKα2 gene products. HEK293 cells are human embryonic kidney cells transformed from sheared human adenovirus type 5 (42). They display general characteristics of renal tubular cells, but it is not possible to relate their characteristics to a specific renal segment. Grishin et al. (16) used these cells for the functional expression of cloned ATP1AL1 cDNAs. Antibodies raised against a portion of the ATP1AL1 protien did not react with untransfected HEK293 cells suggesting that these cells do not express the HKα2 gene product. HK2 cells are human adult proximal tubule epithelial cells immortalized by transduction with human papiloma virus (38). It

has been shown that rabbit proximal tubule cells do not express the HKα2 gene products (12). It is therefore unlikely that HK2 cells express the ATP1AL1 gene product.

Transfection and Reporter Gene Activity

RCCT28A cells were grown in a 24 well plate format with each well containing 1mlL of media (Dulbecco's Modified Eagle Medium – F12 (DMEM-F12), 10% Fetal Bovine Serum (FBS)) until they reached approximately 70% confluency (usually over one night). For each transfection, 250pmoles of the deletion construct to be tested, non-specific DNA to 1μg and 0.2μg of pRL control plasmid were mixed with 140μl DMEM and 40μl of Superfect reagent (Qiagen, Inc.) and incubated at room temperature for 10 minutes to allow for complex formation between the transfection reagent and the DNA. During this incubation, the RCCT28A cells were washed two times in PBS. The complexed solution was brought up to 400μl with DMEM-F12 plus FBS and 200μl was pipetted on top of each of two duplicate wells. The transfection proceeded at 37°C for 2 hours. The transfection reagents were washed off the cells with PBS and 1mL of fresh media was added to each well. After 24 hours at 37°C, the media was removed and the cells were lysed by adding 100μl of lysis buffer (Promega, Inc.). Each well was scraped with a 20μl pipette tip to facilitate lysis. Ten microliters of the lysed cells were added to 100μl of firefly luciferase substrate and the raw firefly luciferase activity was measured for 10 seconds. One hundred microliters of quenching buffer plus Renilla luciferase substrate was added and the raw Renilla luciferase activity was measured for 10 seconds. Each construct was transfected into RCCT28A cells at least three times and in duplicate each time in order to obtain statistically relevant data. Each round of transfections included the plasmid that initially had the most activity (pDZ18) and the plasmid with the

least activity (pGL3 empty vector). All the constructs in one transfection were normalized to the Renilla control data (see below). The pDZ18 activity was set to 100% and data for the other constructs were calculated as a value relative to 100%.

Normalization of the Luciferase Data

Table 4-1 is an example of the calculations required for normalization of the luciferase data. For each set of transfections, the raw firefly reading and the raw Renilla reading are taken directly from the luminometer (Column one and Column two respectively). The first Renilla reading was divided into each of the subsequent Renilla readings to create a normalizing factor for each sample (Column three). Each firefly reading was multiplied by its normalizing factor to obtain the normalized firefly reading (Column four). The average of the two readings for the plasmid that initially gave the highest activity reading (pDZ18) was divided into the normalized firefly reading for each sample and multiplied by 100 to obtain a percentage of the highest activity (Column five). The background activity observed for the empty vector (pGL3) was subtracted out (Column six) and the activity for the plasmid with the highest activity (pDZ18) was reset to 100% (Column seven). Each plasmid was transfected at least three time and in duplicate. The average percent relative activity for each plasmid was graphed and error bars were added to indicate plus and minus the standard of error.

Results

HKα_{2a} and HKα_{2c} Reporter Gene Activity

Figure 4-3 is a graph of relative luciferase activity for the constructs that contain both HKα_{2a} and HKα_{2c} transcription and translation start sites. All constructs were transfected into RCCT28A cells in duplicate and at least three times. These data were normalized using the Renilla luciferase internal control. Plasmid pDZ18 contains only

Table 4-1. Example of the normalization of raw luciferase data

Vector	Raw firefly	Raw Renilla	Normalizing factor[a]	Normalized firefly[b]	% activity[c]	Subtract pGL3	Reset to 100%
pGL3	616	13166	1.00	616	7.66	-0.45	-.050
	623	11912	1.11	689	8.56	0.45	0.50
pDZ11	4540	34459	0.38	1735	21.56	13.45	14.66
	3887	26942	0.49	1899	23.61	15.50	16.90
pDZ18	12923	20904	0.63	8139	101.16	93.05	101.42
	10513	17404	0.76	7953	98.84	90.73	98.90

a) Normalizing factor equals first renilla reading/each raw renilla reading
b) Normalized firefly equals raw firefly * normalizing factor
c) % activity equal average for(normalized firefly/pDZ18)*100

the HKα_{2a} transcription start site and contained the most firefly activity of all the constructs created. This reading was set to 100% activity and the relative activity for the remaining constructs was calculated. The original construct containing the HKα_{2a} and HKα_{2c} transcription and translation start sites (pDZ15) had no activity. Since the ATG start codons for HKα_{2a} and HKα_{2c} were part of this construct, it seemed possible that the HKα2 amino acids added to the N-terminus of the luciferase protein could affect luciferase activity. Therefore, Quikchange was used to create a plasmid in which the ATG start codon for HKα_{2a} was mutated (pDZ28). This plasmid also had no luciferase activity. At this point, reporter gene activity data from the second set of constructs, those that only contained the HKα_{2a} transcription start site, had shown that shorter plasmids contained more activity (see below). Therefore, plasmids pDZ15 and pDZ28 were digested with *Aat*II and *Nhe*I, the ends were filled in, and the vectors were religated. These two plasmids, pDZ30 and pDZ31 still had no activity. A portion of the sequence from plasmid pDZ30 is shown in Figure 4-4A. Analysis of the two alternative mRNA

transcripts that would be produced by this sequence (Figures 4-4B and C) revealed a potential stop codon that was introduced by the ligation of the HKα2 gene fragment into the pGL3 vector (shown in red). This stop codon would terminate translation from both

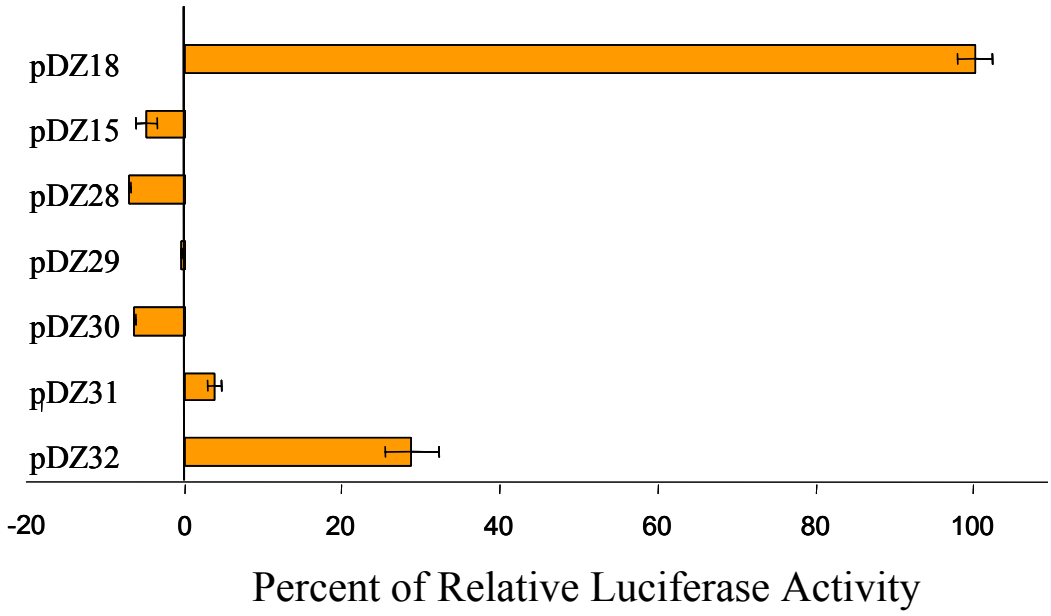


Figure 4-3. Percent activity for reporter gene constructs that contain the HKα_{2a} and HKα_{2c} transcription and translation start sites. Error bars respresent the standard of error.

the HKα_{2a} translation start and the HKα_{2c} translation start before reaching the ATG start codon for the luciferase gene. A new set of constructs were created to fix this problem and to remove some of the HKα2 amino acids that were added to the N-terminus of the luciferase protein. *Pac*I sites were introduced into pDZ29 and pDZ30 using Quikchange (Figure 4-1). Subsequent digestion with *Pac*I and religation of the vector resulted in plasmids pDZ31 and pDZ32. The sequence for pDZ31 and the mRNA transcripts produced by this plasmid are shown in Figure 4-5. Although the potential stop codon was successfully removed, plasmid pDZ31 had no luciferase activity suggesting that the

A.

```
-44  GCGGGGCGCG CAGCATTTAA GGCGGACACC ACCTCCCCTG GGCAGCGGCT GGCGATCGGC
 17  TGCGGAGGTG CGCGCAGGGC CCGCGTGGCT GTGGGTACCT CCTTCGCCAG CACCGTCGCC
 77  ACTACCAACG CCGCCACCGC GGGACCCTAC CCCGCATCGG TCGCCGCCGC CACCGCAGGT
137  CCCACGACCC CTCCTGCCCT CCGCGCCCCC TGCCCGCCGA CCCGCGGCGC CTCCAGCGCG
197  ACATGCGCCA GGTGTGTGAG GAAGTGACGC GGTGCGGACT GGAGAGAAGT GCGGGAAAGG
257  GTGAAGGGCT CCGTCCGGGG GTCTTTACTC TGCAACCCTG TTCCAGCCGC CGAGCACCCG
317  TGTGTCACTC GGGAACTGGC TGGGTAAAGA GGTCAATCCA GACACGCGGG GAAGGAGTTC
377  CAGGGGTCAG CTCCGCCCTC GCACCTGCGG GCTCGGATTC GGAGAAAAGT GCTAGACTGG
437  AGCTACACGT ATGCGTAGCG GTCTGGAAAA TGCCCCAGGC TCGGGTCTGA GGGGCCCAAG
497  TCTATGCACC GCTGGTGTGA CCCCGCAGGG CAACCCCGCG GTTAACTTCT CTCCTGCCCA
557  CCCCTAGAGG TGTCTTCCTG GGAAGACGAT GGCAGGCGGT GCCCACCGAG CCGACCGTGC
617  AACAGGGGAA GAGAGGAAGG AGGGAGGTGG GAGGTGGCGC GCTCCCCACA GCCCTTCCCC
677  TCCTGGCCCG CGAGGGTGTC CGGTCCCACT CAAGGCAGCT GCGCAGAGCC TGTGCAGAAA
737  TACCACCTGG GGCCGGTATT GCACTCTGCT TCTCTTTCAG AGAAAGCTGG AAATTTACTC
797  CGTGGAGCAC CATGCAGCTA CAGATATCAA GAAGAAGGAG GGGCGAGATG GCAAGAAAGA
857  CAATGACTTG GAACTCAAAA GGAATCAGCA GAAAGAGGAG CTTAAGAAAG AACTTGATCC
917  TCGAGATCTG CGATCTAAGT AAGCTTGGCA TTCCGGTACT GTTGGTAAAG CCACCATGGA
977  AGACGCCAAA AACATAAAG
```

B.

```
ATG CGC CAG AGA AAG CTG GAA ATT TAC TCC GTG GAG CAC CAT GCA GCT ACA GAT
ATC AAG AAG AAG GAG GGG CGA GAT GGC AAG AAA GAC AAT GAC TTG GAA CTC AAA
AGG AAT CAG CAG AAA GAG GAG CTT AAG AAA GAA CTT GAT CCT CGA GAT CTG CGA
TCT AAG TAA GCT TGG CAT TCC GGT ACT GTT GGT AAA GCC ACC ATG GAG ACG CCA
AAA ACA TAA AG
```

C.

```
ATG GCA GGC GGT GCC CAC CGA GCC GAC CGT GCA ACA GGG GAA GAG AGG AAG GAG
GGA GGT GGG AGG TGG CGC GCT CCC CAC AGC CCT TCC CCT CCT GGC CCG CGA GGG
TGT CCG GTC CCA CTC AAG GCA GCT GCG CAG AGC CTG TGC AGA AAT ACC ACC TGG
GGC CGG TAT TGC ACT CTG CTT CTC TTT CAG AGA AAG CTG GAA ATT TAC TCC GTG
GAG CAC CAT GCA GCT ACA GAT ATC AAG AAG AAG GAG GGG CGA GAT GGC AAG AAA
GAC AAT GAC TTG GAA CTC AAA AGG AAT CAG CAG AAA GAG GAG CTT AAG AAA GAA
CTT GAT CCT CGA GAT CTG CGA TCT AAG     GCT TGG CAT TCC GGT ACT GTT GGT
AAA GCC ACC ATG GAA GAC GCC AAA AAC ATA AAG
```

Figure 4-4. Sequence analysis of plasmid pDZ15. (A) Sequence of the plasmid. Black indicates HKα2 sequence while green indicates pGL3 sequence. Dark purple represents the transcription start sites and light purple the translation start site for $HK\alpha_{2a}$. Dark pink represents the transcription start site and dark pink the translation start site for $HK\alpha_{2c}$. Blue represents the translation start site for the luciferase gene. Orange indicates the two sets of bases mutated to *PacI* sites for future experiments (see text). (B) mRNA transcribed by initiation from the $HK\alpha_{2a}$ start site. Red indicates stop codon introduced before the luciferase translation start site. (C) mRNA transcribed by initiation from the $HK\alpha_{2c}$ construct.

A.

```
-44  GCGGGGCGCG CAGCATTTAA GGCGGACACC ACCTCCCCTG GGCAGCGGCT GGCGATCGGC
 17  TGCGGAGGTG CGCGCAGGGC CCGCGTGGCT GTGGGTACCT CCTTCGCCAG CACCGTCGCC
 77  ACTACCAACG CCGCCACCGC GGGACCCTAC CCCGCATCGG TCGCCGCCGC CACCGCAGGT
137  CCCACGACCC CTCCTGCCCT CCGCGCCCCC TGCCCGCCGA CCCGCGGCGC CTCCAGCGCG
197  ACATGCGCCA GGTGTGTGAG GAAGTGACGC GGTGCGGACT GGAGAGAAGT GCGGGAAAGG
257  GTGAAGGGCT CCGTCCGGGG GTCTTTACTC TGCAACCCTG TTCCAGCCGC CGAGCACCCG
317  TGTGTCACTC GGGAACTGGC TGGGTAAAGA GGTCAATCCA GACACGCGGG GAAGGAGTTC
377  CAGGGGTCAG CTCCGCCCTC GCACCTGCGG GCTCGGATTC GGAGAAAAGT GCTAGACTGG
437  AGCTACACGT ATGCGTAGCG GTCTGGAAAA TGCCCCAGGC TCGGGTCTGA GGGGCCCAAG
497  TCTATGCACC GCTGGTGTGA CCCCGCAGGG CAACCCCGCG GTTAACTTCT CTCCTGCCCA
557  CCCCTAGAGG TGTCTTCCTG GGAAGACGAT GGCAGGCGGT GCCCACCGAG CCGACCGTGC
617  AACAGGGGAA GAGAGGAAGG AGGGAGGTGG GAGGTGGCGC GCTCCCCACA GCCCTTCCCC
677  TCCTGGCCCG CGAGGGTGTC CGGTCCCACT CAAGGCAGCT GCGCAGAGCC TGTGCAGAAA
737  TACCACCTGG GGCCGGTATT GCACTCTGCT TCTCTTTCAG AGAAAGCTGT TAATTAAAG
797  CCACCATGGA AGACGCCAAA AACATAAAG
```

B.

```
ATG CGC CAG AGA AAG CTG          A GCC ACC ATG GAA GAC GCC AAA  AAC ATA
AAG
```

C.

```
ATG GCA GGC GGT GCC CAC CGA GCC GAC CGT GCA ACA GGG GAA GAG AGG AAG GAG
GGA GGT GGG AGG TGG CGC GCT CCC CAC AGC CCT TCC CCT CCT GGC CCG CGA GGG
TGT CCG GTC CCA CTC AAG GCA GCT GCG CAG AGC CTG TGC AGA AAT ACC ACC TGG
GGC CGG TAT TGC ACT CTG CTT CTC TTT CAG AGA AAG CTG          A GCC ACC
ATG GAA GAC GCC AAA AAC ATA AAG
```

Figure 4-5. Sequence analysis of plasmid pDZ31. (A) Sequence of the plasmid. Black indicates HKα2 sequence while green indicates pGL3 sequence. Dark purple represents the transcription start sites and light purple the translation start site for $HK\alpha_{2a}$. Dark pink represents the transcription start site and dark pink the translation start site for $HK\alpha_{2c}$. Blue represents the translation start site for the luciferase gene. Orange indicates the remaining *PacI* site after digestion and religation. (B) mRNA transcribed by initiation from the $HK\alpha_{2a}$ start site. (C) mRNA transcribed by initiation from the $HK\alpha_{2c}$ construct.

HKα2 amino acids added to the N-terminus of the luciferase protein were inactivating the enzyme. Plasmid pDZ32, which has a mutation at the $HK\alpha_{2a}$ ATG translation start was the first plasmid in this series to have luciferase activity. The amount of activity, however, was about 50% of that seen for the similar sized plasmid that contains only the

$HK\alpha_{2a}$ transcription start site (pDZ19). At this point in the study, it became clear that the second set of deletion constructs, those that contain only the $HK\alpha_{2a}$ transcription start site had significant reporter gene activity and would produce the desired deletion data (see below). Therefore, attempts to restore luciferase activity to these constructs were terminated.

$HK\alpha_{2a}$ Reporter Gene Activity

Figure 4-6 is a graph of the relative reporter gene activity for the all of the plasmids in the second set of deletion constructs. These plasmids contain the $HK\alpha_{2a}$ transcription start site and varying amounts of 5' DNA. These constructs were transfected into RCCT28A cells and after 24 hours, the luciferase activity was measured. Once again the normalized activity from plasmid pDZ18 was set to 100% and the percent activity for the remaining constructs was calculated. The clear result was that progressively shorter promoter fragments contained progressively more luciferase activity. The one exception was that the shortest plasmid (pDZ25) contained activity similar to background (pGL3). It is notable that the luciferase activity increased gradually as the plasmid length decreased. In order to eliminate the possibility that the size of the plasmid was affecting luciferase activity, plasmid pDZ49 was created. This plasmid contained a random fragment of DNA placed in front of the pDZ18 fragment resulting in a plasmid the length of pDZ11. The luciferase activity of this construct was similar to pDZ18, not pDZ11, suggesting that plasmid size does not affect reporter gene activity. A one way ANOVA analysis was performed on the deletion data. The red stars in Figure 4-6 indicate plasmids with a statistically significant difference in the level of

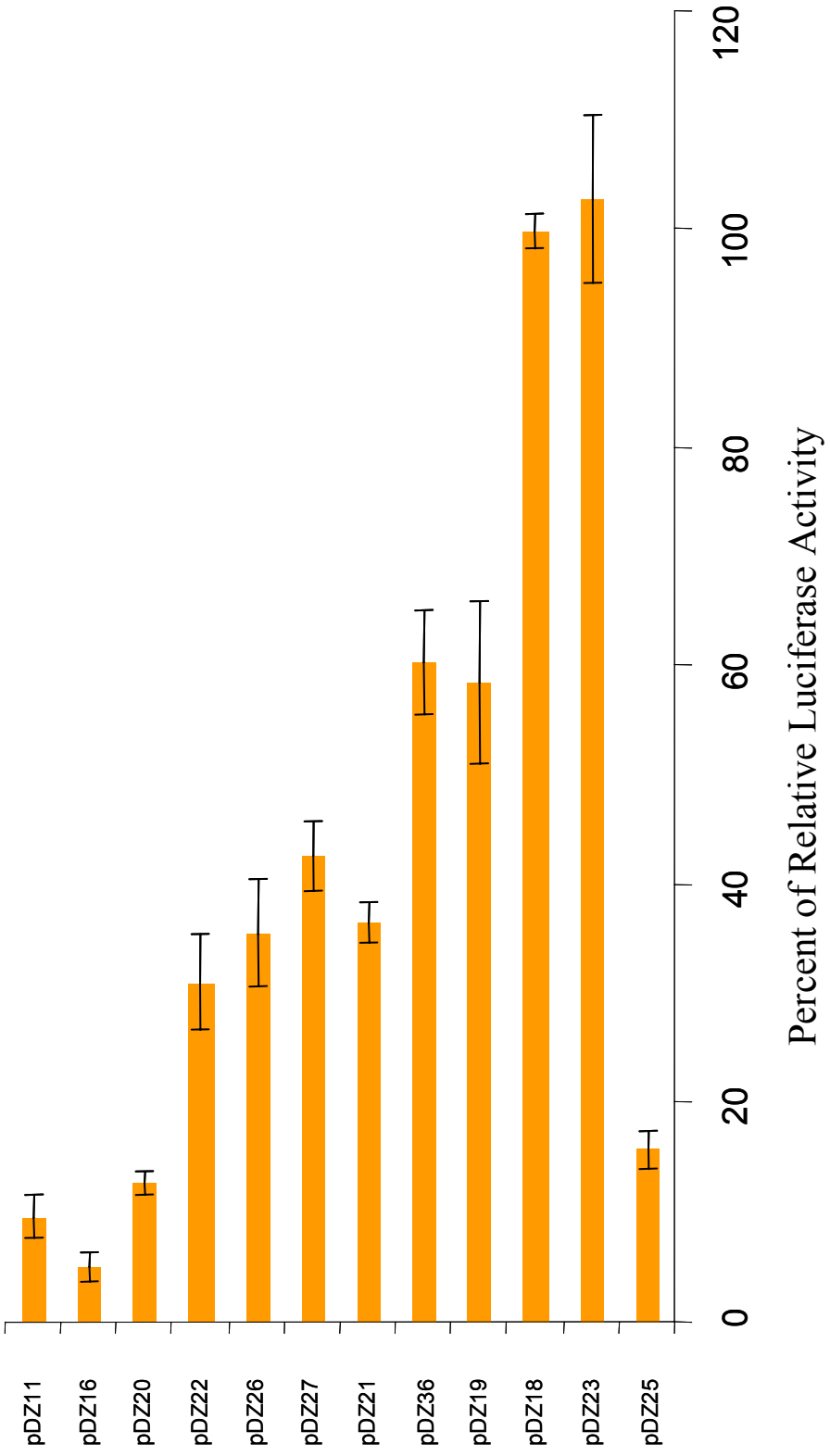


Figure 4-6. Percent activity for constructs that contain the HKα_{2a} transcription start site.

luciferase activity when compared to the preceding plasmid. The differences between pDZ19 and pDZ18, and pDZ23 and pDZ25 were the most significant and therefore their sequences were analyzed for possible regulatory elements.

Putative Repressor Mutations

The difference in luciferase activity between plasmids pDZ19 and pDZ18 was examined. The sequence difference between these two plasmids came from a 245bp deletion that extended from bases –631 to –876. A transcription factor binding site database (TFSEARCH) analysis did not reveal any known repressor binding sites in this region. The human ATPAL1 gene was the only HKα2 gene that also had known sequence in this 5' region. An alignment of the rabbit sequence and the human sequence in this region did show a short sequence that was well conserved between the two species (Figure 4-7). Quikchange mutagenesis was used to randomize the sequence at –680

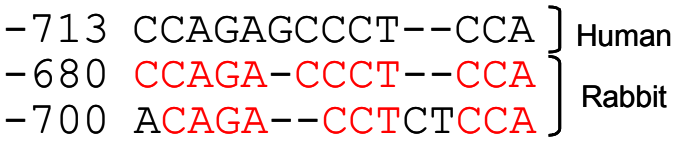


Figure 4-7. Alignment of possible repressor sequences from human ATP1AL1 and rabbit HKα2 genes.

(pDZ38), the sequence at –700 (pDZ40), and both sequences together (pDZ41). These constructs were transfected into RCCT28A cells along with pDZ18 and pDZ19. After 24 hours the luciferase activity was measured. The activity for pDZ18 was set to 100% and the relative activity for the rest of the plasmids was calculated. Figure 4-8 is a graph of these data. Although there was a small increase in activity over that of pDZ19, it did not

appear as though the conserved sequence that was analyzed has a major effect on HKα2 repression.

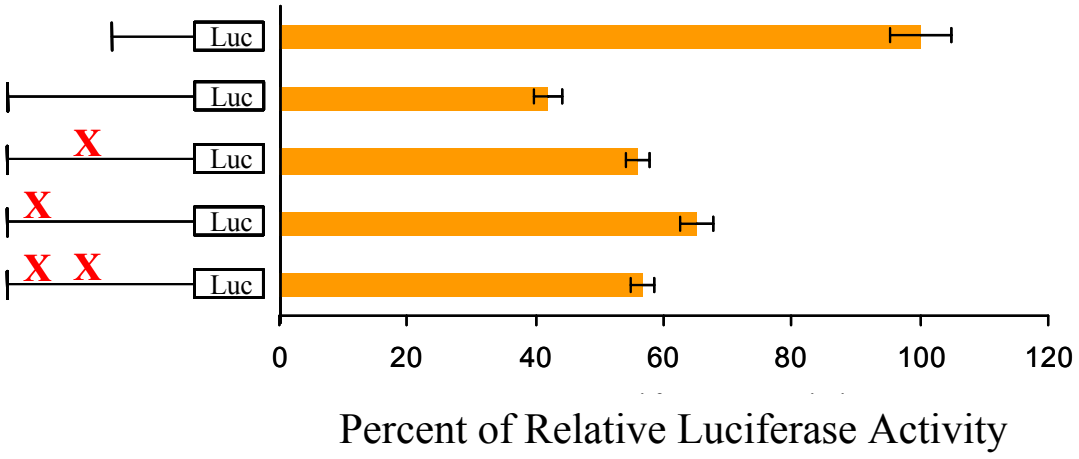


Figure 4-8. Percent luciferase activity in repressor mutation constructs. Red X indicates the position of the putative repressor mutation.

Putative TATA Element Mutations

The dramatic drop in luciferase activity between plasmids pDZ23 and pDZ25 suggested that the core promoter for the HKα2 gene was deleted. An alignment of the 269 bases deleted in pDZ25 with the sequences known for the human ATP1AL1 gene, the mouse HKα2 gene, and the rat HKα2 gene revealed a great deal of homology as indicated by the stars at the bottom of Figure 4-9. In particular, the CATTTAA (red lettering) element near the rabbit transcription start site was completely conserved in human, rat and mouse. In order to test if this element was functioning as a TATA element, pDZ18 was mutated in several ways. These data are represented graphically in Figure 4-10. The mutation constructs that destroyed the element (pDZ34 and pDZ49) show a drop, although not a complete loss, of reporter gene activity. It therefore appears

as though the element is necessary for full activity, but the surrounding bases are capable of initiating an intermediate level of transcription despite the mutation to the TATA box. The mutation that created a consensus TATA box (pDZ39) had activity that was not significantly different from the wild type element, again suggesting that the native element was functioning well as a consensus TATA box.

```
rabbit    ATCCTCGGTTTCCCAGCCCCCTGGGGTGTCTGCAG-GCCGGGCTACTTGCACAGCAGCAG
human     GCCGGCGGGTTTCCTACCCTCCGAGGCGTCCGCTG-GCC--------TGCGCCCTGGCGG
rat       GCACTCGGGTTTCCTATCC-CAGGGGCGTCCCTGGTGCAGTG--AACTGTGTCGCCCCAG
mouse     GTCCTCGGGTTTCCTACCC-CAGGGGCGTCCCCAGTGCAGTG--AGCTGCGTCGCCCCAG
               *** ** **   ** * * ** ***    * **         **        * *

rabbit    GTGCGTAGGCGGGGCGCGCAGCATTTAAGGCGGACACCACCTCCCCTGGGCAGCGGCTGG
human     GGACGTGGGCGGGGCGGGCGGCATTTAAGGCTGGTGCCACCTCCCCGGTGCAGCGGCTGG
rat       GTACGTGGGCGGAGCTCGCAGCATTTAAGGCGCGCTCCACCTCCCAGGGACAGTGGCTGG
mouse     GTACGTGGGCGGAGCGCGCGGCATTTAAGGCGCGCTCCACCTCCCCAGGACAGCGGCTGG
          *  *** ***** **  ** **********     *********  *  *** ******

rabbit    CGATCGGCTGCGGAGGTGCGCGCAGGGCCCGCGTGGCT
human     CGATCGGCCGCGGAGGTGCGTGCAGGGCCCGCGCCGCC
rat       TGATCGGCCGCGGAG-TGCTTG-TGCTTCCGG-TTGGC
mouse     CGGTCGGCCGCGGAGGTGCGTGCTGGGTCCGGGTTGGT
           * ***** ****** ***  *  *   ***    *
```

Figure 4-9. Alignment of the rabbit, human, rat and mouse DNA sequence upstream of the HKα2 transcription start sites. Stars represent completely conserved bases. Red indicates the completely conserved TATA-like element. Dark purple represents the transcription start sites for rabbit $HK\alpha_{2a}$ determined in Chapter 2, human ATP1AL1 (45), rat $HK\alpha_{2a}$ (26) and mouse HKα2 (52).

Effect of Cell Type on Reporter Gene Activity

In order to test the effect of cell type on reporter gene activity, three deletion constructs were transfected into four cell types. Plasmids pDZ18, pDZ19 and pDZ11 were transfected into HEK293, HK2, RCCT28A and HIG-83 cells. Within each cell line, the normalized activity for pDZ18 was set at 100% and the other two constructs are

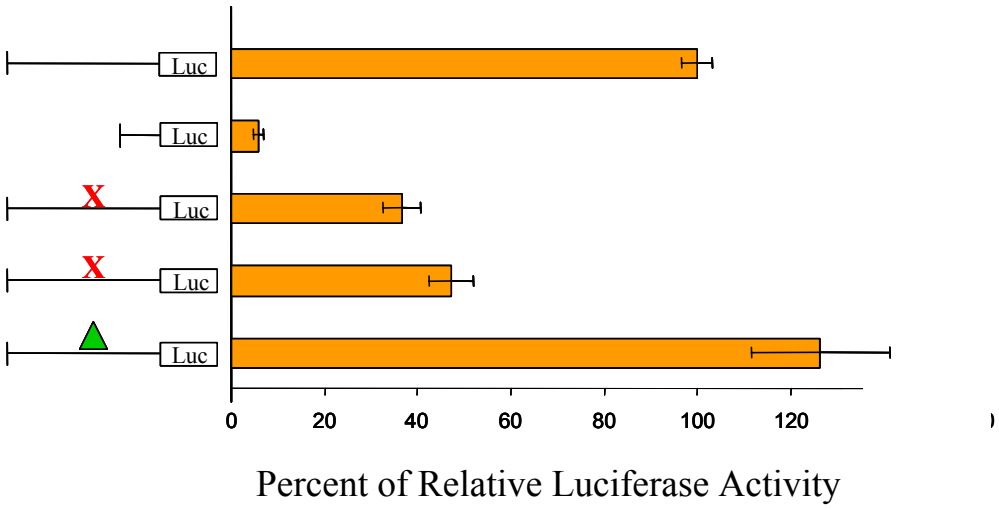


Figure 4-10. Percent activity in reporter gene constructs with mutations in the CATTTAA element. Red X represents mutations that randomize the CATTTAA element. Green triangle represents a mutation that converts the CATTTAA element into a consensus TATA element.

represented relative to 100%. Figure 4-11 is a graph representing the results of the transfections. In the three adult cell lines (RCCT28A, HIG-82, HK2), the expression pattern for the three constructs was similar. The longest construct was repressed and the shorter constructs had increasing amounts of luciferase activity. The embryonic cell line (HEK293), however, had a different pattern of expression. The most dramatic difference was with the longest construct. Plasmid pDZ11 was repressed in all three adult cell lines, but was not repressed in this embryonic cell line. The activity was most similar to pDZ19 suggesting that the embryonic cell line did not contain some or all of the factors required for HKα2 gene repression.

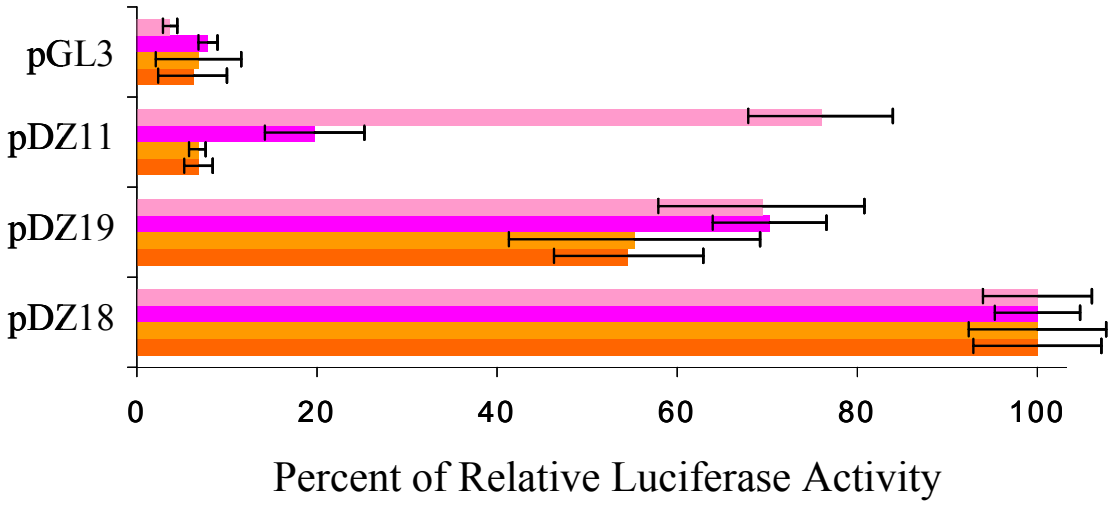


Figure 4-11. Reporter gene activity in various cell types. Plasmid constructs are as indicated. Light pink represents HEK293 cells. Dark pink represents HK2 Cells. Light orange represents HIG-82 cells. Dark orange represents RCCT28A cells.

Discussion

The use of the luciferase reporter gene assay to evaluate the region of DNA 5' of the HKα2 gene provided the first data regarding the regulation of the rabbit HKα2 gene. Four sets of reporter gene constructs and four tissue culture cell types were used in this assay and each provided important information.

The first set of constructs contained the transcription and translation start sites for HKα_{2a} and HKα_{2c} cloned in front of the luciferase reporter gene. These constructs had little to no reporter gene activity. The removal of the stop codon introduced during plasmid construction, and the mutation of the ATG start codon for HKα_{2a} resulted in one construct with significant luciferase activity (pDZ32). This activity, however, was still low when compared to the construct with same deletion (pDZ19) in the series that

contained only the HKα_{2a} start site (60% for pDZ19 vs. 30% for pDZ32). This result was difficult to interpret. The mutation of the HKα_{2a} ATG should have removed the amino acids that were added to the N-terminus of the luciferase protein. The luciferase activity, however, was not restored to the expected level suggesting that there were negative regulatory elements that effect the HKα_{2a} transcription in the downstream region that was eliminated when the second set of constructs were made (+93 to +920). Plasmid pDZ31 removed the stop codon that would terminate translation from the HKα_{2c} transcript. This construct, however, had no luciferase activity. There are several possible explanations for the lack of activity. First, the 72 amino acids added to the N-terminus of the luciferase protein could cause a loss of function (Figure 4-5C). Second, the splicing machinery may not recognize the reporter gene construct and fail to produce the alternative HKα_{2c} transcript. And finally, the HKα_{2c} transcript may be regulated, and without the proper signal, it may completely repressed. As it would be difficult to distinguish between these possibilities, and more importantly, the second set of deletion constructs were producing results, this line of investigation was therefore abandoned.

The second set of reporter gene constructs provided the most information regarding the regulation of the HKα2 gene. The 3' end of these were shorter, ending at +92, and contained only the transcription start site for HKα_{2a}. Interestingly, the longest construct (pDZ11) contained the least reporter gene activity and progressively shorter constructs showed a gradual increase in reporter gene activity. These results suggested that under the conditions of the luciferase assay, the HKα2 gene promoter was repressed and significant reporter gene activity was seen only after deleting the DNA responsible for binding repressor elements. The increase in luciferase activity, however, appeared as

a gradual increase rather than as distinct jumps in activity that would be expected when repressor elements are deleted. It seemed possible, therefore, that the decrease in size of the constructs, rather than the removal of repressor elements, caused the change in luciferase activity. Plasmid pDZ49, however, eliminated this possibility. This construct was made by taking the plasmid with the most activity, pDZ18, and inserting a fragment of *E. coli* DNA into a *Hind*III site upstream of the putative promoter element. The result was a plasmid the size of pDZ11, but devoid of any additional eukaryotic transcription factors. Plasmid pDZ49 has luciferase activity similar to pDZ18, not pDZ11, meaning that the size of the plasmid was not having an effect on luciferase activity. Although the changes in luciferase activity were smaller than one might expect, the one way ANOVA analysis of these data did reveal several deletions that caused a statistically significant change in reporter gene activity. There are clear increases in luciferase activity between constructs pDZ20 and pDZ22, pDZ21 and pDZ36, and pDZ19 and pDZ18. Additionally, there is a dramatic decrease in luciferase activity between plasmids pDZ23 and pDZ25 (Figure 4-6). These changes in activity provided the basis for the construction of the mutation plasmids that are discussed below. The fact that the HKα2 gene was repressed in these assays was a surprise because RCCT28A cells have been shown to express the HKα2 mRNA's and proteins (6). Additionally, Zhang et al. (52) performed a promoter deletion analysis of the mouse HKα2 gene and did not observe repression with their longest constructs. The discrepancy between the results of our study and the study of Zhang et al. may be explained by the differences in the cell types. The RCCT28A cells were derived from the cortical collecting duct while the cells used by Zhang et al. were derived from medullary collecting duct cells. To date, it has been unclear whether or not

the cortical collecting duct normally expresses the HKα2 gene products (see introduction) whereas the medullary collecting duct has been consistently shown to express the HKα2 gene products. The reporter gene assay performed in our study suggested that there may be certain cellular conditions necessary for HKα2 gene expression in the cortical collecting duct. This may, in part, explain the discrepancies present in the literature. One likely explanation for the repression observed in the deletion analysis is the condition under which the RCCT28A cells were grown for the assay. The transfection of the RCCT28A cells necessitated that they were grown to 70% confluency. At this level of confluency the RCCT28A cells are undifferentiated. It was hypothesized that the HKα2 gene was repressed in this state and then when the cells reach 100% confluency and differentiate, they would begin to express the HKα2 gene products. This hypothesis resulted in the experiments that are discussed in Chapter 5 of this dissertation.

The third set of reporter gene constructs were designed to test the functionality of putative repressor elements identified in the DNA fragment between plasmids pDZ19 and pDZ18. The increase in luciferase activity between these two plasmids was the largest observed (40%) and seemed a good point to begin a search for novel repressor elements. A transcription factor database search did not identify any known binding sites for transcription factors. An alignment of the rabbit DNA sequence in this region with that of the human ATP1AL1 gene, however, did identify one sequence in human that was well conserved to two sequences in rabbit (Figure 4-7). Unfortunately, random mutation of these sequences both individually and combined did not have a significant effect on luciferase activity (Figure 4-8) suggesting that these sequences are not directly involved

in the repression of the HKα2 gene. In the future, it may be possible to make smaller deletions in this region and identify a short sequence that does play a role in repression. This information could be used to design new experiments, for example, gel shift assays or yeast two hybrid assays, that are aimed at the identification of novel repressor elements.

The fourth set of reporter gene constructs were designed to test the functionality of the CATTTAA element upstream of the $HK\alpha_{2a}$ transcription start site. Two different mutations, both of which changed the element to random sequence, caused a drop in luciferase activity of greater than 50%. The mutation of the CATTTAA to a consensus TATA element (TTTTATAT) had very little, if any, effect. These data suggest that the CATTTAA element plays a necessary role in the initiation of transcription at the $HK\alpha_{2a}$ start site. There are, however, other sequences surrounding the element that also appear to be important for full activity. Additional core promoter elements include an initiator (Inr), a downstream promoter element (DPE), a TFIIB recognition element (BRE) and SP family member binding sites (GC boxes)(see introduction). The sequence surrounding the CATTTAA element contains DNA that has homology to all of these elements except for the BRE, and the alignment of the sequences surrounding the CATTTAA element for rabbit, human, mouse and rat (Figure 4-9) reveals that these putative core promoter elements are at least partially conserved. All four species contain a completely conserved sequence that matches the Inr. The sequence, however, does not overlap with the published the transcription start site for any of the four species. Since there does appear to be a discrepancy between the four species regarding the transcription start site, this sequence should not be discounted as a promoter element. Downstream of the four

transcription start sites there is a completely conserved sequence that matches the consensus DPE in all but one base. Immediately upstream of the CATTTAA element, all four species contain sequence that is extremely GC rich. Although the GC boxes are not completely conserved, it seems likely that SP family member could bind and stabilize transcription initiation for all four species.

In a fifth experiment, several of the HKα_{2a} deletion constructs were used to test the effect of cell type on reporter gene activity. The longest construct, pDZ11, a middle length construct, pDZ19, and the construct with the highest luciferase activity, pDZ18, were transfected into four cell types, RCCT28A, HIG-82, HK2, and HEK293. In the two rabbit cell lines, RCCT28A and HIG-82, the reporter gene activity followed the same pattern as seen in the deletion analysis. Initially, these results were unexpected because RCCT28A was considered an expressing cell line while HIG-82 was considered non-expressing. It later became clear, however, that under the conditions of this assay, the RCCT28A cells were not expressing the HKα2 gene. Therefore, the similarity in the activity levels between RCCT28A and HIG-82 cells was not surprising. The same pattern of expression was also observed for the human adult kidney cell line HK2. This result was expected since HK2 cells were derived from proximal tubule cells that are considered a non-expressing section of the nephron (38). A dramatic difference in expression pattern was seen with the human embryonic cell line HEK293. The longest construct in this case was not repressed, but instead had activity similar to pDZ19. Apparently, the factors required for repression of the HKα2 gene promoter are not present in this cell type. Although it is possible the lack of factors is due to a species difference, the fact that HK2 cells were capable of causing repression, makes it more

likely that the developmental stage of the cell type is the important factor. HEK293 cells, however, do not appear to express the endogenous ATP1AL1 gene product (12). It may be that repressor elements are absent allowing for expression of the reporter gene from a plasmid, but that a repressive chromatin structure inhibits expression for the endogenous gene.

In summary, the reporter gene analysis of the region 5' of the $HK\alpha_{2a}$ transcription start site provided the first information regarding the regulation of the rabbit HKα2 gene. It was determined that under the necessary transfection conditions, the HKα2 gene is repressed. The deletion analysis narrowed the location of two repressive elements to 365bp and 245bp respectively. Furthermore, the observed repression of the HKα2 gene promoter was the driving force for the experiments carried out in Chapter 5 of this dissertation showing that RCCT28A cell differentiation plays an important role in the expression of the HKα2 gene. Additionally, the DNA sequence between bases –26 and –295 as were identified as containing core promoter elements. The CATTTAA element that was completely conserved between the four known HKα2 genes was mutated and found to be a functional TATA box. Additional core promoter elements were identified by alignment, but their functionality was not tested. In addition, the importance of cell differentiation and the developmental stage of the organism were also identified as a possible regulatory factor.

CHAPTER 5
CELL DIFFERENTIATION AND HKα2 GENE EXPRESSION

During the course of our study, it became clear that RCCT28A cells undergo a change in cell morphology once the cells are grown past confluency. Figure 5-1 depicts the RCCT28A cells just as they reach confluency (A) and well after confluency (B). After reaching confluency, the RCCT28A cells appear to differentiate and form ring-like structures that resemble cross-sections of nephron collecting duct tubules. It seemed likely that the change in cell morphology might be accompanied by a change in gene expression. The results of the promoter deletion experiment led us suspect the HKα2 gene may be one example of a gene that is altered in expression when RCCT28A cells differentiate. In the promoter deletion experiments (Chapter 4), the construct with the most 5' DNA (pDZ11) had the least reporter gene activity. This result suggests that under the conditions of the assay, the HKα2 gene was repressed. The transfection of reporter gene constructs into RCCT28A cells necessitated that the cells be grown to about 70% confluency. At this level of confluency, many of the cells were not in contact and no ring structures were apparent. We hypothesized that under these conditions, similarly to the reporter gene driven by the HKα2 gene promoter, the HKα2 endogenous gene would also be repressed. Furthermore, we hypothesized that once the cells come in contact and begin to differentiate the HKα2 gene would become transcriptionally active. In order to test this hypothesis, several experiments were performed using several cellular conditions. RT-PCR was performed in order to evaluate the level of HKa2 transcript in

cells of different confluency. Northern blots were performed in to determine the level of HKα2 transcript in confluent and past confluent cells. Immunocytochemistry was performed to determine the specific RCCT28A cells within a population of differentiating cells that were expressing the HKα2 proteins. Finally, DNaseI hypersensitivity experiments were performed to determine if there was a difference in chromatin conformation at the HKα2 gene promoter under the different cellular conditions. This is the first report of RCCT28A cell differentiation in tissue culture and HKα2 is the first gene in which a change in expression appears to be correlated with RCCT28A cell differentiation.

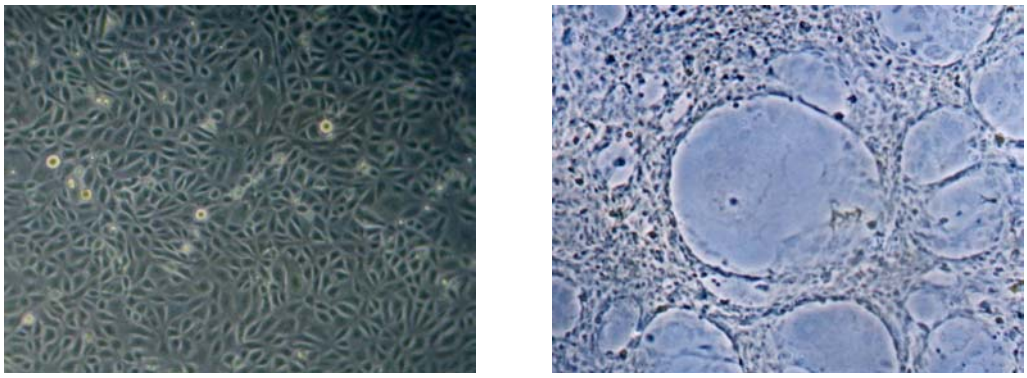


Figure 5-1. Micrographs of tissue culture cells. All cells are photographed at a magnification of 200X. (A) RCCT28A cells undifferentiated. (B) RCCT28A cells differentiated.

Materials and Methods

Detection of HKα2 mRNAs

Total RNA was isolated from RCCT28A tissue culture cells using the Trizol method (Invitrogen, Inc.). The cells were grown in a 60mm dish to 70% confluency, 100% confluency or past confluency until ring-like structures appeared. They were

washed with PBS and 2mL of Trizol was added directly to the 60mm dish. After incubation for 5 minutes at room temperature the cells were pipetted up and down to facilitate lysis and transferred to a 15mL conical tube. 0.4mL of chloroform was added to the tube and then the tube was vigorously shaken for 15 seconds and allowed to incubate at room temperature for 5 minutes. The tube was centrifuged at 4000 rpm for 20 minutes and the aqueous phase was transferred to a new 15mL conical tube. The RNA was precipitated by adding 1.0mL of isopropanol to the tube, incubation at room temperature for 10 minutes and then centrifugation at 4000 rpm for 15 minutes. The pellet was washed with 2.0mL of 75% ethanol, dried briefly, and resuspended in 200µl of RNase free dH_2O. The RNA samples were used for both RT-PCR and Northern blot analysis.

The RT-PCR was carried out with 10µg of the indicated RNAs. The reverse transcriptase reaction (2µl RT buffer, 2µl dNTP, 2µl random decamers, 1.5µl RNase Inhibitor, either minus RT or plus 1µl RT, and volume up to 20µl) was incubated at 37°C for 1 hour. The entire reaction was added to 50µl PCR mastermix (Qiagen, Inc.), which included 26µl dH_2O, 2µl of primer BC230 (5'CCGACACGAGTGAAGACAAT3'), and 2µl of primer BC231 (5'GCTTGTCATTGGGATCTTCC3'). This primer set amplifies a 305 base pair band from the common region of the HKα2 RNAs (HKα_{2a} 1264-1569). The PCR conditions were one cycle of 94°C for 2 minutes, forty cycles of 94°C for 30 seconds and 68°C for 1 minute, and one cycle of 68°C for 5 minutes. 30µl of each PCR reaction was run on a 1% agarose gel and visualized with ethidium bromide.

Northern blot analysis was carried out with 30µg of the indicated RNAs. The RNA was brought to a volume of 50µl with dH_2O and 10µl of dye (0.25% bromophenol

blue, 0.25% xylene cyanol, 30% glycerol, 5% 1μg/mL ethidium bromide) and run on a 1% agarose gel with 1X MOPS (20mM MOPS, 5mM NaOAc, 0.5mM EDTA) as running buffer. After visualizing the RNA under UV light, the gel was denatured in 50mM NaOH for 25 minutes, denatured with 200mM Tris-HCl for 25 minutes, equilibrated to 10X SSC (3M NaCl, 0.3M NaCitrate) for 25 minutes and set up to transfer the RNA to nylon membrane by capillary action. The blots were blocked and hybridized to the HKα2 mid probe as described in Chapter 2.

Detection of HKα2 Protein by Immunocytochemistry

Tissue culture cells were grown to 100% confluency and past confluency to ring-like structures in Nunc Lab-Tek II chamber slides (Fisher CAT# 12-565-7) that contained four wells. The cells were fixed in paraformaldehyde fixative (2.5mL paraformaldehyde, 7.5mL buffer (4.6mL Solution A (0.2M lysine-HCL pH to 7.4 with 0.1M Na_2HPO_4), 3.4mL Solution B (0.1M Na_2HPO_4 pH to 7.4 with 0.1M NaH_2PO4)) for 30 minutes then washed three times with PBS (150mM NaCl, 3mM KCl, 8mM Na_2HPO_4, 2mM KH_2PO_4). In order to remove any endogenous peroxidase activity, the fixed cells were incubated in 3% H_2O_2 for 30 minutes and washed three times five minutes with PBS. The cells were then incubated in blocking serum (50μl donkey serum in 1mL PBS) for 15 minutes and then incubated with LLC26 anti $HK\alpha_{2c}$ antibody (1:100 dilution in PBS) overnight at 4ºC. The next day, the cells were washed two times five minutes in PBS, and incubated for 30 minutes with rabbit α chicken secondary antibody (1:500 dilution in PBS). After washing two times five minutes in PBS, the antibodies were detected using the ABC elite system (Vector Laboratories, Inc.).

DNaseI Hypersensitivity

DNaseI treatment was carried out on T-75 flasks grown until the cells were differentiated. Six flasks were washed with PBS and 1mL of trypsin was added. The cells from the T-75 flasks were combined into one conical tube and pelleted at 1000 rpm for 10 minutes. The pellet was washed with solution A (150mM sucrose, 80mM KCl, 35mM HEPES pH7.4, 5mM K_2HPO_4, 5mM $MgCl_2$, 0.5mM $CaCl_2$) and pelleted at 1000 rpm for 10 minutes. The cells were resuspended in solution B (150mM sucrose, 80mM KCl, 35mM HEPES pH7.4, 5mM K_2HPO_4, 5mM $MgCl_2$, 2mM $CaCl_2$) to a concentration of 10^7 cells/mL, aliquoted into 1mL fractions in 15mL conical tubes and incubated at 37°C. An appropriate concentration of DNaseI (see results) was added to 0.4% NP-40/Solution B mixture and quickly added to the cells. The mixture was incubated at 37°C for 5 minutes. Three milliliters of lysis buffer (50mM Tris-Cl pH 8.5, 150mM NaCl, 25mM EDTA, 0.5% SDS, 300ug/mL Proteinase K) was added to the DNaseI treated cells and incubated overnight at room temperature.

Genomic DNA was isolated from each sample by phenol/chloroform extraction. 3mL of phenol was added to each sample and nutated at room temperature for 1 hour. The tubes were centrifuged at 3000 rpm for 10 minutes and the aqueous layer was transferred to a new 15mL conical tube. Three milliliters of phenol/chloroform/Isoamyl alcohol (IAA)(25:24:1) was added and nutated for 1 hour at room temperature. The tubes were centrifuged at 3000 rpm for 10 minutes and the aqueous layer was transferred to a new tube. Three milliliters of chloroform/IAA (24:1) was added and nutated for 1hour at room temperature. The tubes were centrifuged at 3000 rpm for 10 minutes and the aqueous layer was transferred to a new 15mL conical tube. Each sample was treated with

5µl RNase cocktail (Ambion, Inc.) at 37°C for 1 hour. The samples were then phenol/chloroform extracted as before except that the incubation time was reduced to 30 minutes. Finally, the samples were precipitated with 2X volume of 100% ethanol, washed with 70% ethanol and resuspended in 500µl of dH_2O.

Thirty micrograms of the genomic DNA from each DNaseI treated sample was digested with *Spe*I at 37°C overnight and then run on a 0.7% agarose gel over the next night. The DNA was visualized with ethidium bromide and then the gel was acid washed (0.125M HCl) for 30 minutes, denatured for 30 minutes, neutralized for 30 minutes, soaked in 10X SSC and set up to transfer to nylon membrane by capillary action overnight. The DNA was UV crosslinked to the nylon membrane and hybridized to the *Spe*I 3' probe (see results) at 55°C overnight. The membrane was washed three times at 55°C in wash buffer and exposed to autoradiograph film at –80°C for 5 days.

Results

RT-PCR

RT-PCR was used to investigate the levels of expression of HKα2 mRNAs in 70% confluent and 100% confluent RCCT28A cells. Total RNA was isolated from RCCT-28A cells grown under the two experimental conditions and 10µg of the RNA was used for RT-PCR. The primers BC230 and BC231, used in the PCR reaction, should amplify a 305bp fragment. Figure 5-2 is an ethidium bromide stained gel of the RT-PCR products. A band of the expected size was only seen in the lane that contained the RNA isolated from differentiated cells. This result clearly indicated a relationship between cell differentiation and HKα2 gene expression.

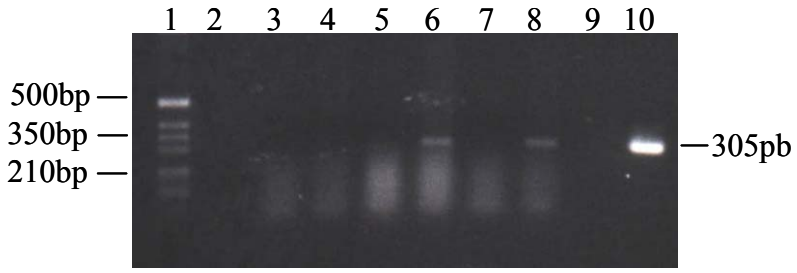


Figure 5-2. RT-PCR products indicating the presence or absence of HKα2 transcripts. lanes are (1) 1Kbp ladder (2) blank (3) 70% confluent RCCT28A cells - RT. (4) 70% confluent RCCT28A cells + RT. (5) 100% confluent RCCT28A cells – RT. (6) 100% confluent RCCT28A cells + RT. (7) 100% confluent RCCT28A cells – RT +Aldosterone. (8) 100% confluent RCCT28A cells + RT + Aldosterone (9) blank (10) plasmid control.

Northern Analysis

Northern analysis was pursued in an attempt to extend the RT-PCR data and test HKα2 expression levels by a second method. Total RNA was isolated from the 100% confluent and the past confluent RCCT28A cells. The RNA was run on a 1% agarose gel and transferred to nylon membrane. The membrane was probed with the HKα2 cDNA mid probe (cDNA base pairs 1264-1569). Figure 5-3 is an ethidium bromide stained agarose gel (A) and an autoradiograph film exposure of the hybridization (B). Although bands appear only in the differentiated cell lane, they are not of the expected 4000bp and 4400bp lengths. Figure 5-3C shows the same blot hybridized to the GAPDH probe. This figure, as well as the ethidium bromide stained gel (Figure 5-3A) show that the RNA preparation as a whole was not degraded and suggests that the HKα2 mRNAs are very rapidly turned over and only degradation products are detectable by Northern analysis. This result is consistent with the RT-PCR and strengthens the argument that RCCT28A cells must be differentiated in order to express the HKα2 gene.

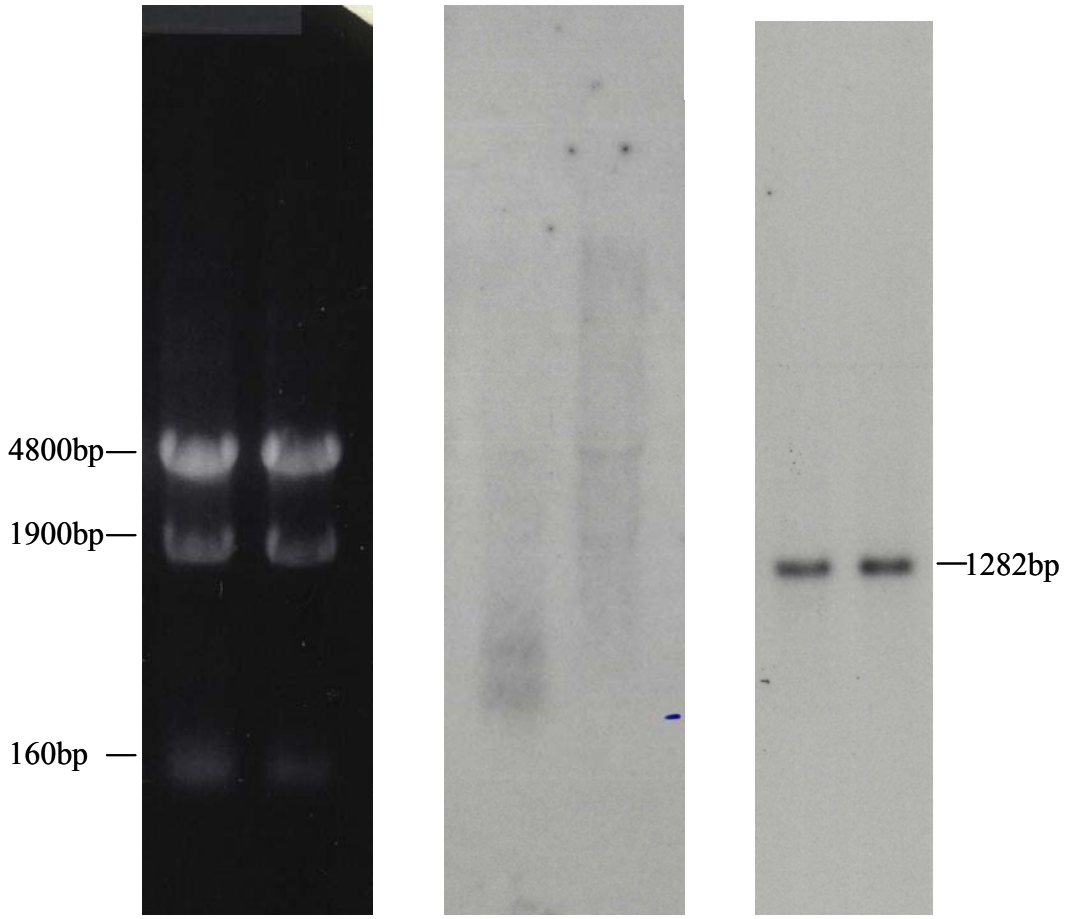


Figure 5-3. Northern blot of RCCT28A total RNA. Lane (1) RCCT28A non-confluent cells (2) RCCT28A confluent cells. (A) Ethidium Bromide stained gel (B) HKα2 mid probe (C) GAPDH probe.

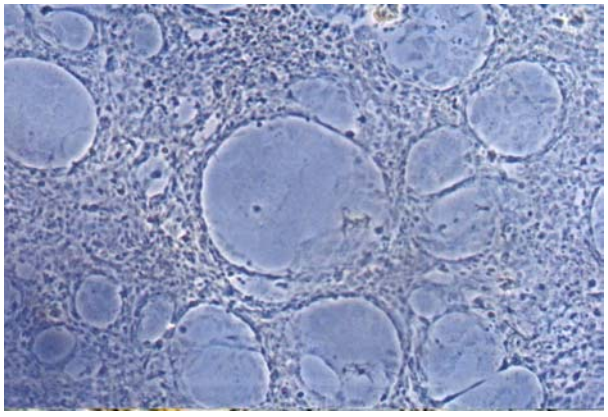
Immunocytochemistry

In order to investigate the HKα2 protein expression in RCCT28A cells in the differentiated state, immunocytochemistry was performed using the $HK\alpha_{2c}$ specific antibody LLC26. This antibody has previously been used to localize $HK\alpha_{2c}$ to the apical

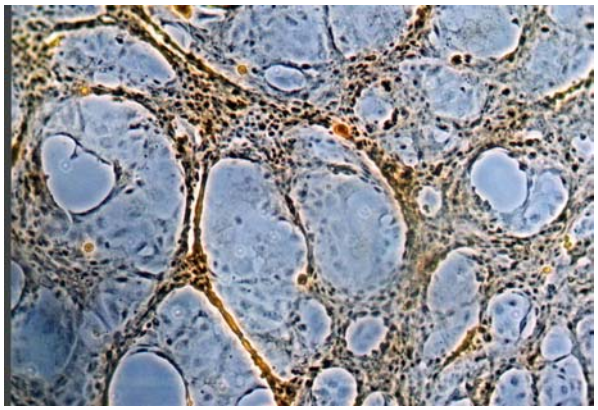
membrane of cortical collecting duct cells in the rabbit kidney (48), and therefore it seemed likely that it would also recognize HKα_{2c} in tissue culture cells. Figure 5-4A is an example of a well containing differentiated cells that were treated with 2° antibody only. There is no apparent background staining. Figure 5-4B is a similar well of RCCT28A cells that were stained with the HKα_{2c} specific 1° antibody as well as the donkey anti-chicken 2° antibody. Within this well, there appear to be two populations of cells. The cells that are differentiated into ring-like structures appear to be stained while and those cells between the structures appear to be undifferentiated and not stained. In Figure 5-4C, a well containing RCCT2A cells grown to 100% confluency, the there also appear to be a two populations of cells. One group of cells that are stained and a larger group of cells that were not stained. Although no ring-like structures were apparent, the cells in the photograph that appeared stained also appeared to be closer together and the staining pattern suggested that they were perhaps beginning to undergo differentiation. The majority of the cells, however, to appeared to be undifferentiated and unstained.

DNaseI Hypersensitivity

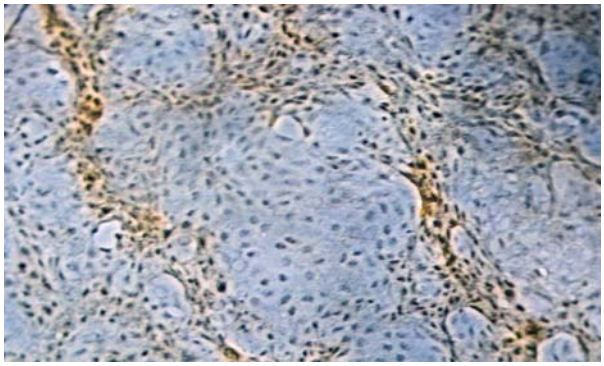
Although the immunocytochemistry experiments suggested that only a portion of the cells in a differentiated culture of RCCT28A cells were in fact expressing the HKα2 gene, DNaseI hypersensitivity was attempted in order to distinguish between an inactive gene (undifferentiated RCCT28A cells) and an active gene (differentiated RCCT28A cells). Tissue culture cells were grown until the cells formed ring-like structures, and then individual aliquots were treated with increasing concentrations of DNaseI (0μg to 120μg). Genomic DNA was then isolated and 5mg of each sample was run on a 0.7% agarose gel to confirm DNaseI digestion (Figure 5-7A). Thirty micrograms of the same



(A) Differentiated cells, secondary antibody only



(B) Differentiated cells, $HK\alpha_{2c}$ specific antibody



(C) Undifferentiated cells, $HK\alpha_{2c}$ specific antibody

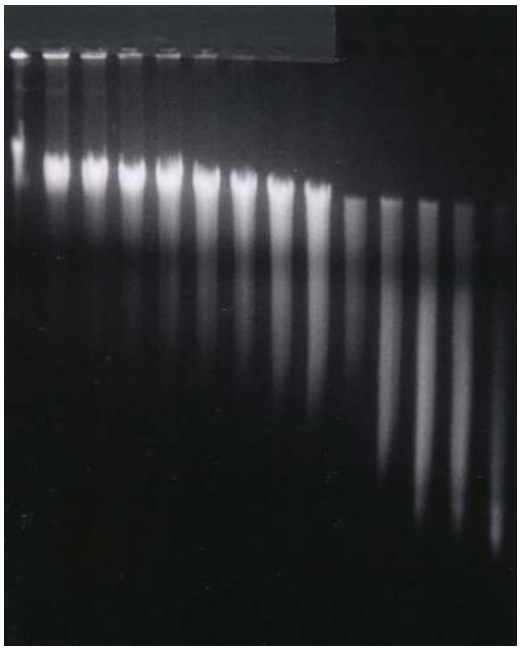
Figure 5-4. Micrographs of the immunostaining of RCCT28A cells. Conditions for each photograph are as indicated. Magnification is 200X.

aliquots of DNAseI treated DNA were digested with *Spe*I and run overnight on a long 0.7% agarose gel (Figure 5-7B). The DNA was transferred to nylon membrane and probed with a fragment of DNA that corresponds to the 3' end of the *Spe*I fragment from pDZ10. This *Spe*I fragment contained the HKα2 gene promoter. Figure 5-5C is an example of an autoradiograph film exposure of a DNaseI hypersensitive blot. The full length *Spe*I fragment (4700bp) can be seen in each lane. As the DNaseI was increased, the intensity of the full length band decreased. As the full length band decreased in intensity, however, smaller bands that correspond to hypersensitive sites were not apparent after several trials.

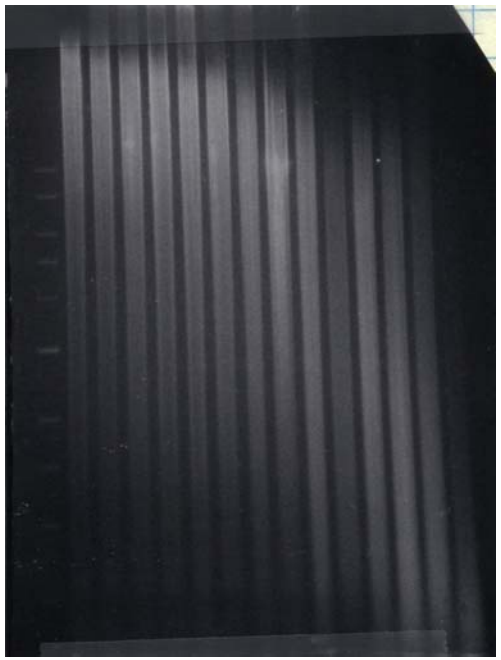
Discussion

The results of the promoter deletion experiment led us suspect that when the RCCT28A cells are undifferentiated the HKα2 gene is repressed. Only when the cells are allowed to differentiate, do they begin to express the HKα2 transcripts at a basal level. In order to test this hypothesis, several experiments were performed on differentiated an undifferentiated RCCT28A cells. The results of each individual experiment are open to several possible interpretations. When taken together, however, the evidence strongly suggests a correlation between RCCT28A cell differentiation and the expression of the HKα2 gene.

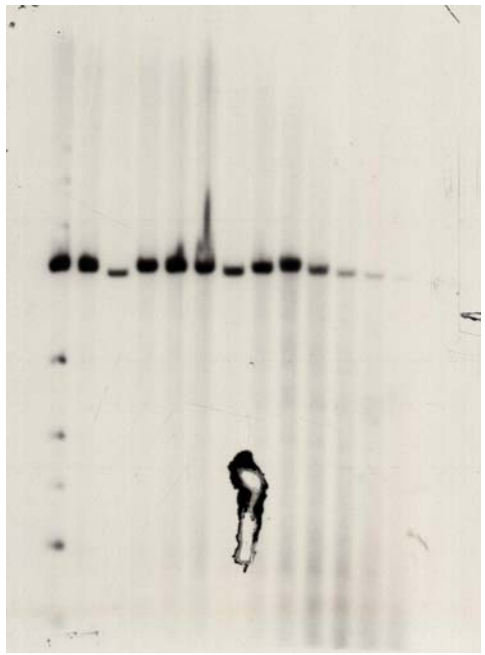
The RT-PCR data regarding the mRNA transcripts for the HKα2 gene products was the most convincing evidence that cell differentiation is a factor that regulates HKα2 gene expression. The promoter deletion analysis performed in Chapter 4 of this study suggested that in RCCT28A cells grown to 70% confluency the HKa2 gene promoter was repressed. The RT-PCR performed on RNA from cells grown to the same level of



(A) Ethidium bromide stained gel of DNaseI treated samples



(B) Ethidium Bromide stained gel of DNaseI treated and *Spe*I digested samples



© Southern Blot probed with *Spe*I 3' probe

Figure 5-7. Genomic Southern of DNaseI treated differentiated RCCT28A cells. Treatments are as indicated.

confluency produced no product. Only in the RT-PCR experiment that contained RNA from 100% confluent cells was a product obtained. This is the first report of a correlation between RCCT28A cell confluency and gene expression.

The fact that so much RNA (10μg) was needed to obtain a RT-PCR product led us to question whether or not allowing the cells to grow past confluency until they started to produce ring-like structures would increase the yield of HKα2 mRNA. Northern blot analysis was performed on undifferentiated and differentiated RCCT28A cells. Although bands that hybridized to the HKα2 mid probe were consistenly observed in the differentiated cell lane, they were not of the appropriate size. No bands were detected in the 100% confluent, but undifferentiated cell lanes. The mRNA that was detectable by RT-PCR was apparently below the dectable level of this assay even though 30μg of total RNA were used to create the Northern blot. It seemed likely that the bands that were observed were degredation products of the full-length HKα2 transcripts. The ethidium bromide stain of the agarose gel, the GAPDH probe, and the repeatability of the result all confirm that the RNA preparation as a whole was not degraded. It appears, therefore, that the HKα2 transcripts in particular were unstable. It may be that once the RCCT28A cells become differentiated, the HKα2 mRNAs are produced and rapidly turned over providing the cells with a very low steady state level of the HKα2 proteins. An appropriate signal, such as ion imbalance or hormone activity, may be necessary to stabilize and/or upregulate the HKα2 transcripts and provide the cells with an increased level of the HKα2 proteins.

Immunocytochemistry was performed on RCCT28A cells in order to specifically determine which cells in a culture of differentiating cells were expressing the HKα2

proteins. Our initial expectation was to find staining only in the cells that had differentiated into ring-like structures. Some of the most intense staining was indeed seen in these cells. There was, however, lighter staining in many of the cells that surrounded the rings. There were also cells within the rings that appeared undifferentiated and were not stained. The RCCT28A cells were derived from rabbit cortical collecting duct tissue, and most closely resemble the intercalated cells that make up the lining of the cortical collecting duct (2). These cells would not normally be found in the tissue surrounding the collecing duct tubule. It may be possible that in the artificial tissue culture environment, the cells that were surrounding the ring structures were also receiving the signal to differentiate and were therefore producing the HKα2 protein. The most convincing evidence for this interpretation was the staining in the RCCT28A cells that were just at 100%confluency (Figure 5-5). In this case, all of the cells that had formed the closest contacts appeared to be expressing HKα_{2c}, not just the single layer of cells that was most likely to form a ring structure. Furthermore, the pattern of staining suggested that the cells expressing the HKα_{2c} protein were beginning to form rings.

DNaseI hypersensitivity assays were carried out on the differentiated RCCT28A cells. The full length *Spe*I fragment (4.7Kbp) can be seen in each lane, and with increasing DNaseI, the intensity of the full-length band decreases. Unfortunately, DNaseI hypersensitive bands were unable to be detected. If the RNA and protein are present in the differentiated cells, it follows that there should be a DNaseI hypersensitive site present at the core promoter. As seen in the immunocytochemistry experiment, however, only a portion of the cells in tissue culture are actually differentiated and expressing the HKα2 gene product. Additionally, the requirement for large amounts of

RNA and protein for the other experiments suggests that the expression in the differentiated cells is very low. The occupancy of transcription factor binding sites at the core promoter is therefore also low. Considering a low percentage of cells expressing the HKα2 gene products at a low levels, the DNaseI hypersensitive site present at the core promoter of expressing cells was probably below a detectable level. The inability to detect a hypesensitive site in a differentiated cell population made it unnecessary to perform the experiment on the undifferentiated cells.

In summary, the experiments in this chapter were designed to test the hypothesis that cell differentiation led to the expression of the HKα2 gene products. The reporter gene experiments carried out in Chapter 4 led to the conclusion that if RCCT28A cells were approximately 70% confluent, the HKα2 gene was repressed. During the course of growing cells for these experiments is it was observed that if RCCT28A cells were allowed to grow past confluency, they began to differentiate and form ring-like structures. We hypothesized that cell differentiation may be the signal required to initiate transcription from the HKα2 gene and the experiments presented in this chapter generally supported the hypothesis. The RT-PCR showed the requirement for above a 70% confluency for expression of the HKα2 transcript. Northern data showed the presence of a transcript only in differentiated RCCT28A cells that hybridized to the HKα2 mid probe. The immunocytochemistry data showed the presence of a protein that is recognized by an $HK\alpha_{2c}$ specific antibody. Taken together, these data strongly suggests a correlation between cell differentiation and HKα2 gene expression.

CHAPTER 6
CONCLUSIONS AND FUTURE DIRECTIONS

The rabbit HKα2 gene produces two splice variants of the alpha subunit of the colonic isoform of the H^+, K^+ - ATPase ($HK\alpha_{2a}$ and $HK\alpha_{2c}$). There is a great deal of *in vivo* evidence that the HKα2 transcript, protein, and activity are increased by a variety of cellular conditions including low blood potassium and sodium, acid/base balance and hormones. The molecular mechanisms by which these levels are increased, however, have not been studied. The purpose of our study was to characterize the rabbit HKα2 gene as a precursor to the initiation of studies of its regulation. To this end I have (1) cloned the HKα2 gene from a rabbit genomic library, (2) mapped the transcription start sites for $HK\alpha_{2a}$ and $HK\alpha_{2c}$, (3) performed a reporter gene analysis of the region 5' of the transcription start sites, and (4) determined the effect of cellular differentiation on HKα2 gene expression.

Cloning the HKα2 Gene

The first specific aim of this dissertation project was to clone the rabbit HKα2 gene. At the time that this project began, the cDNA's for the rabbit and rat HKα2 proteins and the human ATPAL1 protein were known. The only HKa2 genomic organization that was known was for the human ATPAL1 gene. It was unclear whether the rabbit, rat, and human proteins were homologous because the conserved amino acid similarity between the three was lower that that of other homologous P-type ATPases. One of our goals in cloning the rabbit HKα2 gene was to determine its genomic

organization and compare it to the human ATPAL1 gene. Our second goal in cloning the HKα2 gene was to obtain sequence 5' of the previously determined $HK\alpha_{2a}$ and $HK\alpha_{2c}$ cDNA ends because the DNA elements responsible for regulating transcription from the gene would be found in this region. Three bacteriophage λ clones were identified that contain 20-of-23 exons and 65% of the entire HKα2 gene. Four genomic PCR products that contained the missing exons were later obtained. The complete sequence of a 6.3Kbp fragment from clone HKα2.1, the complete sequence of clone HKα2.5 and partial sequences from clone HKα2.8 and the PCR products are listed in Appendix A. The sequence obtained enabled us to determine the genomic organization of the rabbit HKα2 gene. The organization is depicted in Figure 6-1. The HKα2 gene spans 30kbp of genomic DNA. The 23 exons are shown in black while the 22 introns are shown in blue. The location of the three probes used to screen the bacteriophage λ rabbit genomic library, the three λ clones isolated in the screen, and the four PCR products that completed the HKα2 gene are also shown. During the course of this study, the mouse HKα2 genomic organization was published (52), and the rat HKα2 genomic organization became available through the NCBI sequence database. The genomic organization for the rabbit HKα2 gene was determined and found to be nearly identical to the human ATPAL1 gene, the mouse HKα2 gene and the rat HKα2 gene. These data, along with the distance analysis performed by Caviston et al. (8) confirm that these four genes are in fact homologous genes and were derived from a common ancestor (8). Furthermore, the sequence from the 6.3Kbp fragment of clone HKα2.1 provided the necessary data for mapping the transcription start sites for HKα2 gene and initiating studies on its regulation.

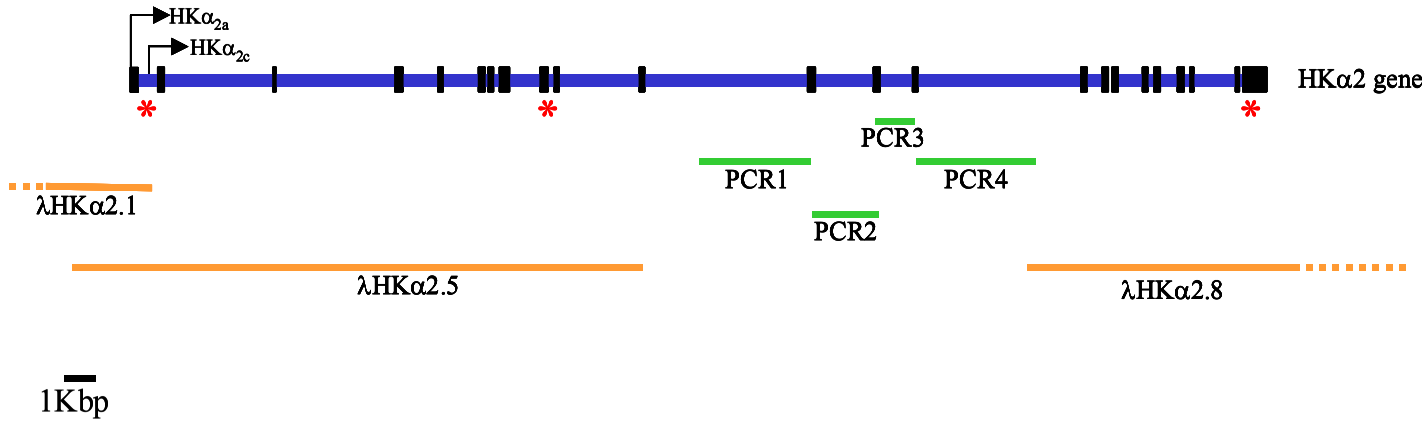


Figure 6-1. The HKα2 gene. The arrows represent the transcription start sites for $HK\alpha_{2a}$ and $HK\alpha_{2c}$ as indicated. The red stars mark the locations of the three cDNA probes used to screen the bacteriophage λ library. Black represents HKa2 exons 1-23. Blue represents HKa2 introns 1-22. Orange represents λ clones indicated. Green represents PCR products generated to complete the HKα2 gene sequence.

Transcription Start Sites for HKα_{2a} and HKα_{2c}

The second specific aim of this dissertation was to map the transcription start sites for the two alternative transcripts produced by the HKα2 gene (HKα_{2a} and HKα_{2c}). The determination of the transcription start site was an important step in characterizing the HKα2 gene for two reasons. First, the region of DNA just upstream of the transcription start sites was likely to contain the core promoter elements and regulatory elements. And second, the existence of the alternative transcript (HKα_{2c}) was called into question by Fejes-Toth et al. (12). Mapping the transcription start site for HKα_{2c} would prove that it was an authentic rabbit transcript. A transcription factor binding site search of the 6.3Kbp of DNA immediately upstream of the cDNA end for HKα_{2a} and HKα_{2c} suggested that there were putative promoter and regulatory elements present. Upstream of HKα_{2c} there was one potential CAAT box. Upstream of HKα_{2a} there was a potential TATA box as well an initiator element, a downstream promoter element and multiple SP family member binding sites. Additionally, a CpG island spanned both cDNA ends. The RNase protection assay performed in our study revealed that the true transcription start sites were just upstream of the previously identified cDNA ends for HKα_{2a} (12) and HKα_{2c} (6). The HKα_{2a} transcription start site was ten to eleven bases upstream and the HKα_{2c} transcription start site was five to seven bases upstream. It seemed likely, therefore, that some of the putative elements identified in the search would be functional. This was the basis for some of the reporter gene assays performed in Chapter 4 of this dissertation. Furthermore, the use of an independent assay to determine the transcription start site for HKα_{2c} confirms that it is an authentic transcript found in rabbit tissue and not an artifact of tissue culture and/or 5' RACE.

Reporter Gene Analysis of the Region 5' of the HKα2 Gene

The third specific aim of this dissertation was to perform a reporter gene analysis of the promoter activity present in DNA 5' of the $HK\alpha_{2a}$ transcription start site. An analysis of the reporter gene activity would provide the first data regarding the molecular regulation of the rabbit HKα2 gene. The luciferase reporter gene assay was chosen to carry out this specific aim. Four sets of reporter gene constructs were made and transfected into the rabbit cortical collecting duct cell line RCCT28A. Two of the four sets of constructs provided significant data on the regulation of the HKα2 gene promoter. The deletion constructs that contained the $HK\alpha_{2a}$ transcription start site and decreasing amounts of 5' DNA revealed that the HKα2 gene was repressed under the conditions of the assay. The two largest increases in reporter gene activity were seen when the DNA between –1485 and –1241 and between –876 and –631 were deleted. A search of the transcription factor database, and the mutation of specific bases in those regions, did not identify any known repressor binding sites. The largest decrease in reporter gene activity was seen when the bases between –295 and –26 were deleted. An alignment of this region with the sequence for the human, mouse and rat HKα2 genes showed conservation of several sequences that could act as core promoter elements. The most likely sequence to serve as a TATA box was a completely conserved CATTTAA element. Mutation of this sequence to a random sequence decreased the reporter gene activity by about 50%. Other putative core promoter elements that are conserved and may contribute to the remaining reporter gene activity include SP1 binding sites, an initiator sequence and a downstream promoter element. The functionality of these elements has not been tested. Several of the luciferase constructs were transfected into four different cell types, and it

became apparent that the developmental stage of the cell type played a role in the ability of the cell type to repress the reporter gene activity. The three adult cell lines, from both rabbit and human, were able to repress reporter gene activity. The human embryonic cell line, however, was unable to repress activity from the same reporter gene constructs. These results suggested that the HKa2 gene may be active in early development. During the course of growing the RCCT28A cells for the reporter gene assays, it was noted that the cells undergo differentiation into ring structures when they are grown past confluency. For the transfection of the RCCT28A cells with the reporter gene constructs, however, the cells were only grown to 70% confluency. The fact that the HKα2 gene was repressed under these conditions led to the hypothesis that the HKα2 gene is only expressed when the RCCT28A cells undergo differentiation. This hypothesis was tested by the next set of experiments.

Cell Differentiation and HKα2 Gene Expression

The fourth specific aim of this dissertation was to investigate the effect of cell differentiation on HKα2 gene expression. In order to carry out this aim, several experiments were performed on RCCT28A cells at various levels of confluency. RT-PCR experiments were performed to determine the level of endogenous HKa2 transcript under the conditions used for the luciferase assays. Northern blots were performed on total RNA preparations in order to determine the level of HKα2 gene transcripts present in the differentiated and undifferentiated cells, and immunocytochemistry was carried out in order to determine which cells in a culture of differentiating cells were expressing the HKα2 protein. The RT-PCR experiment resulted in a band of the expected size only in the RNA samples grown to 100% confluency. The cells grown to the same level as used

in the luciferase assay (70%) were not expressing the endogenous transcript. In the Northern blot experiments, the transcripts from the 100% confluents cells were below the level of detection for this method. In the differentiated cells, however, bands of a smaller size than the full-length HKα2 transcripts were detected. The RNA samples were tested and found to contain full-length transcripts for GAPDH. Therefore, it appeared as though the HKα2 mRNA was a particularly unstable transcript. The instability of the transcript in the cortical collecting duct cell line may explain why some investigators have been able to show HKα2 expression in the cortical collecting duct tissue (6, 26, 33) and others have not (40). The immunocytochemistry performed with the $HK\alpha_{2c}$ specific antibody further substantiated that the HKα2 gene was expressed in the RCCT28A cells that were differentiated into ring structures.

Future Directions

In our study, the rabbit HKα2 gene was cloned, and initial studies on the regulation of the HKα2 gene has generated many possible avenues for future studies. There are several experiments that can be done to support the findings of this dissertation and to expand upon the scientific data regarding gene expression as a whole.

The identification of the transcription start sites for $HK\alpha_{2a}$ and $HK\alpha_{2c}$, the reporter gene assay and the observation that RCCT28A cells undergo cellular differentiation are all discoveries that can be used as a basis for future experiments. Some of these represent immediate experimental opportunities to study regulation of the HKα2 gene. For example, the identification of the transcription start sites allowed for a transcription factor database search of the DNA surrounding the start sites. The functionality of the TATA-like element was tested our study. The functionality of the

other conserved elements (INR, DPE, SP1) could be tested in the future by using the same reporter gene assay. Additionally, chromatin immunoprecipitation (ChIP) can be used to determine the specific proteins that bind to functional core promoter elements. In our study, the reporter gene assay identified two regions of DNA likely to bind novel repressor elements. Smaller deletion construct through these two regions may identify the exact sequences to which a novel repressor protein binds. This experiment could be followed up with gel shift assays and/or yeast two hybrid assays in order to isolate and identify the specific proteins that bind to the sequences. Furthermore, the reporter gene assay can be modified in order to study the effect of cell differentiation on the HKα2 gene promoter. Stable transfections of the reporter gene constructs could be made. The RCCT28A cells containing stable reporter gene constructs could be grown to 70% confluency and past confluency to cell differentiation and then the reporter gene activity can be compared between the two conditions. In this way, the effect of cell differentiation on the HKα2 gene promoter can be studied in a system that may allow for the isolation of factors required for the change in gene expression.

The work presented in this dissertation also generated some larger scientific question for future study. This is the first report of RCCT28A cell differentiation and it appears as though HKα2 gene expression is dependent upon differentiation. It would be very interesting to determine the molecular mechanisms by which gene expression is altered by cell differentiation. Unfortunately, HKα2 gene expression was found to be extremely low and/or the transcripts were unstable even in the differentiating cell population. It may be more advantageous to identify a gene with a higher basal level expression that is upregulated by cell differentiation. Perhaps the molecular mechanisms

for regulating that gene would also apply to the HKα2 gene. Furthermore, now that it has been established that cell differentiation is required for basal level expression of the HKα2 gene, it is possible to attempt to identify a signal for the upregulation of the gene. RCCT28A cells that have differentiated, and also RCCT28A cells with stable transfections of reporter gene constructs, can be treated with a variety of substances that could potentially upregulate transcription from the HKα2 gene promoter. The transcription factor database search carried out as part of this dissertation identified and SREBP and a CRE upstream of the HKα2 core promoter. Steroid hormones and cyclic AMP are therefore two good choices for beginning such a screen. Another interesting mechanism for HKα2 gene expression is related to tissue specificity. Although it is somewhat unclear as to which segments of the collecting duct express the HKα2 gene products, it is clear that HKα2 gene expression is limited to very few tissues beyond the kidney and colon. There have been reports of HKα2 gene expression in uterus, brain and spleen (52) but all at extremely low, and not always detectable, levels. The reporter gene constructs designed in our study could be used in transfection with tissue culture cells derived from a variety of tissues. It could be determined if the same regulatory elements involved in repression in the collecting duct cell line are functional in other tissues, or if there are other elements, such as chromatin structure, responsible for the repression.

Summary

This dissertation project successfully cloned the rabbit HKα2 gene and initiated studies on its regulation. The rabbit HKα2 genomic organization is depicted in Figure 6-1. The data obtained regarding the regulation of the HKα2 gene lead to the model for HKα2 gene expression depicted in Figure 6-2. As seen in the reporter gene assay, the

HKα2 gene repressed when the RCCT28A cells are not confluent and therefore not differentiated (Figure 6-2A). There appeared to be at least two regions of DNA upstream of the HKα2 gene promoter that were necessary for binding repressor proteins and inhibiting the formation of an initiation complex at the HKα2 gene promoter. The mRNA and protein analyses indicated that when the RCCT28A cells were grown past confluency and began to differentiate into tubule-like structures, the HKα2 mRNA and protein became detectable, although they were apparently at a low level and unstable (Figure 6-2B). Furthermore, we hypothesize that given the appropriate signal, perhaps low K^+, activators may either bind to the promoter and increase expression from the gene and/or stabilize the message and protein resulting in increased levels of protein (Figure 6-2C).

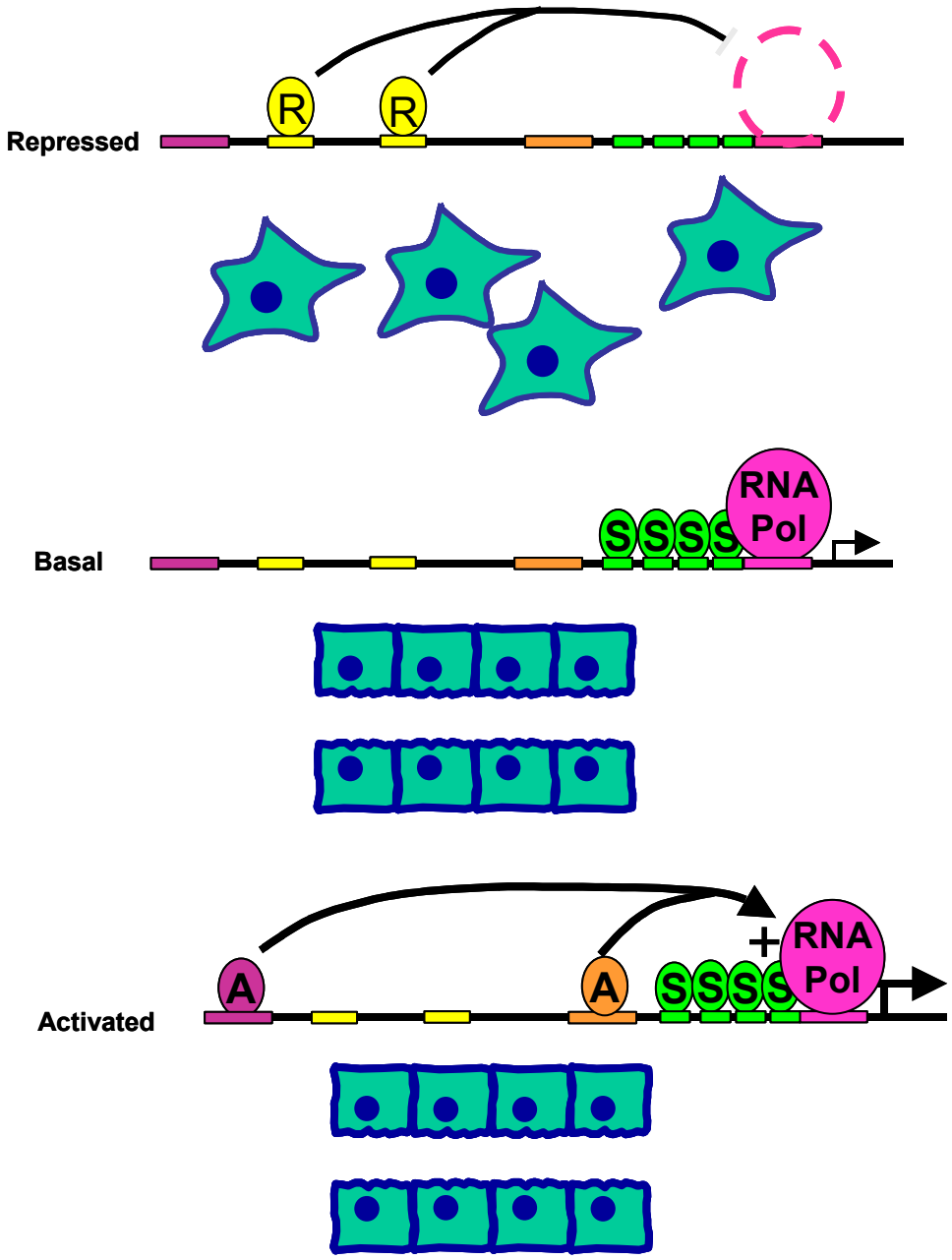


Figure 6-2. Model for HKα2 gene expression.

APPENDIX A
RABBIT HKα2 GENE SEQUENCE

Appendix A contains all of the sequence compiled to determine the intron/exon boundaries for the 23 exons of the HKα2 gene. In all cases, lower case letters indicate intron sequence and upper case letters indicated exon sequence.

Sequence Data from λ Clone HKα2.1 and HKα2.5 Exons 1-11

```
-5384       tgggtaccgg gccccccctc gagttgtaat cttgtgtgat gctcttgaaa tttcctaagg
-5324       gaatagattg ggggttgctt ttatcacaaa aaaagattgt gatgtaatga tgtgagagga
-5264       tggataagtc aatttgctta actactttgt tatgaataag tatacatatt tattttattt
-5204       tattttattt tattttagga accagcactg tggctcagca ggttaagcca ccatatgcga
-5144       agatagcatc ccataggagt gctggtttgg gtcccagtac atgggcccat gccatctaca
-5084       aaggagacct ggaattccag gcttctggtt tcagcctgac ccagtcctag ctattatggc
-5024       catttggaga gtggaccaga ggatagacta tttctctcct ctctctctct ctctctctgt
-4964       aactctgact ttcaaataaa taaattataa aaattataaa tatattagat aatgtataat
-4904       agaatagaaa aataaataat aaaaattatt tcctaaaggc agagagagat agacaaagac
-4844       agatttccca tcctgaggtt tactccccaa atggccacaa cagctgaacc aggagcctga
-4784       aattctattc aggtctccca catggagtcc tttagaataa ataaacctat tttatttatt
-4724       tgaaaggcag agttacagag tgagagtgac agagatggag agagattgcc catcctcagg
-4664       tccactcacc aaaagcctgt aacaaccagg gctgggccag gccaaagtca ggagctggaa
-4604       tctcaatcca cgttgcccgc atgggtgtca gggatcgatg tactttagtc aacacatatt
-4544       acctcccagg gtgcacatta gtgggaaact gatgtagaga gtggatctgg gacccaaacc
-4484       caggcaaact gataggggac acagccatcc caagcagtgg cttaaccact gggccaaaca
-4424       cccaccctta tctgagcata ttaacacagc atgttgtaca cctcctatac agtcaacttt
-4364       ttaaaggtct gtcttagccc cttgcaggcc ttcacctgcc tctgaaaaca ccacagctca
-4304       cggtcctgga ggtctctaac agccccagga aagatcaaca cagcagatta ccagtgattt
-4244       ttggaaagcg tgtttctctc ctgcaggatg tttacatgcc agatatccta taaacgaaga
-4184       aatgaggaaa ccatataaaa gtgtgctagg ctgcagcttt gttttgcttt tggttctcat
-4124       tctgactcca actccaggca gtccccattg gcagagaaat gccggcctag tcatagcctg
-4064       gcacgctggg cagaagcagg tgtgcaggaa gggctccagg tggggatgtc cagtgtgtgg
-4044       agaggaaccc agggagaggg agagggagag ggcaggtggg cagtagaggc accagccagt
-3944       ccggatacgg agcaggaagc agagcttcct cttgtgctca tcacctggat cctcctccct
-3884       gtatgtttgg atcatcagtc acaccccaag ccagctgggg gacagctcca agagggctcc
-3824       cccaactaaa gccccacctc aacggccctt tttgcctctt gtccaccttc cttggagggt
-3764       gcgaagggat agaaaagtat ctgaataata aatcctgaat ttggagatgg ggctgttctc
-3704       cctgtcttct tctgtatttt gttgctgtaa cagaaagcta gaaactgggt gatttataaa
-3644       gaagatggtt taaatcggct cgtggtccta gagactggga gatttgaggt taaggggcac
-3584       ttcccgtgca ggtttcttgc ctgtggggac tttctgcaga gtcccaaggc agtgcagggt
-3524       aagggaagga tacatccaac agagtcaaga ggacttttat attagacccg ctctctagat
-3464       aattcactaa accattaatc caataatgat gaaagtcaga gtcctgagaa ctcatcaact
-3404       attaaagtcc ctctccatgc ctctacaatg gggattacat ttcaatgtga gagctgcaga
-3344       catttggcct aggagctaag atgccagttc cattaccctg tcccatatgg gattgcctgg
-3284       attcaattct catctccagt tactgattcc tgccagtgca ggctctggga ggcagcaggg
-3224       acggctcaag aagttgagtc tctgccaccc aataaggggg acctggatgg agttcccagc
-3164       tcccagcccc accaacagcc cagacgtggc aggaattcta ggtatgagct agtgaatggg
-3104       agctctcgtt ctctatcatc tatctatcta tctatctatc tatctatcta tctacctacc
-3044       tacctaccta tctctaataa acaattttaa catgagattt ggtggggaca ttcaaaccat
-2984       agcggtcccc ccaagcaatc tctttccctt aatttcttcc agcacttaca gcctgataat
-2924       gacaactgca cactgtatta tttcttaccc tgccttattc actagtgctg cttggttaac
```

```
-2864   ttggcctctc caactccagc ccaacttcct ggaggacaga gaatagcata gtttttggta
-2804   ccatttctcc ccaccacccc caccaccccg acccattaac ctcaagttga gctccatgag
-2744   ttgatgaaaa tgaactagga gcagtcacac acatgatgac aggctggccc agatatgtaa
-2684   aatacttgcc caaaagaacc acagtaaaag agcctggcat gagtaggttt cttttgacag
-2624   acagtaaaat tcagcctcaa gcggaagctg ctgctagagt gtgtgggctg gggaggtgtc
-2564   gccagttgca gcctggcttc ctctgtggcc agtcttaaga gaggcattca aaataggtgg
-2504   agtgcaaggt ttgttccctg gaacagggta ggggatgtcc cctgtggcca aatgaagttg
-2444   aatcagtgtc tgaaaggaaa cccagattct tccctccagg tgaggactac tccaggtgct
-2384   cccatcacag acagccagca gccactcagg cacaacagga catccaggcg actagaggca
-2324   gcagggggct ggccccacgt tccctcttca gtacaggtca acgctccggg acacctgaag
-2264   acaccaggga ctgccaaagg ccatggcagc agcagcagca gcagctaaca ctgatccctg
-2204   gcaaaagctg gaaactgtac aaggccagag aacgtgacaa tagaaactct tttgtccagt
-2144   ttgctggtaa gtctacttct tcacacttga caccagcaga gataagcagc ccacttccca
-2084   agtgtagttg taattcacca atttttaaaa ttgcaccatc acttctccat gagtcccctc
-2024   cctccacacc tgccaacaca ctacatccac ccccacctcc ctatattcca tcactaaagc
-1964   ttgaatgggg cttcactctg ggcactctgg gcactacttc tctgtgccac acctggccca
-1904   aaagttggaa caaggctaga aaggtggaaa gtgagcagca gccacctgta caacgactct
-1844   caccagaggt ttccagtaaa cagtgaactg tcagtttcat aaaattgtgt atttcgttac
-1784   ctctttctgc aaatctttct ttggagaaaa gatacaaagc agagctcctt ccgatgaccc
-1724   tgctgcttca gtttagacta gaatctactc tcccctccaa ctctgaagga cctgtgatgt
-1664   gtgatctctg cacaagctgt caatgccatc ttcttgtccc ttaagagtta atgaacggcc
-1604   ggcgctgcgg ctcactaggc taatcctccg ccttgcagcg ccggcacact gggttctagt
-1544   cctggtcagg gcactggatt ctgtcccggt tgcccctctt ccaggccagc tctctgctgt
-1484   ggccagggag tgcagtggag gatggcccaa gcacttggga gaccaggata agtacctggc
-1424   tcctgccatc ctatcagcac ggtgcgctgg ccgcagcacg ccagccatgg cggccattgg
-1364   agggtgaacc aacggcaaag gaagaccttt ctctctgtct ctctctctca ctgtcctctc
-1304   tgcctgtcaa aaaaaaaaaa aaaaaaaaaa aaaaaaagag ttaatgaaga agtaactcac
-1244   tggcgtcatc cctgtcccat agccagagaa tcccagactt aaattctgct ctttggattt
-1184   agggtgtttc taatggattt tctttttgta acttcaagta tttatctatt taatgtattt
-1124   gaaagacaaa gacacagaga tgtcttcccc cttactggtt cactctccaa atgtccccaa
-1064   agcaggggct gggccaagag ggagccaggg atccaagggc tgggagctca acctaggtgt
-1004   cccatgtggg aagccaggac ccaaatactt gagcagtccc tgctgcctcc cagggcatgt
-944    attagcagga agtcagaatt gggaggaggg gcggtgcttg aatccaggta ctctgataag
-884    acaggcagac gtctccagcg ctgcaccaaa tggtcacccc agatcctggc cttgaggagc
-824    caagggaagg ggcgaaaact tgtgtgtgca gtactgtggg tcggggtctc catcgatccg
-764    gagtctgggg gtgtcatcct aaattctgct aacaactggg cgcagcacca cgcagcccgg
-704    gaccccagag ccctccacgc tcactgaccc agaccctcca tccccaaccc ctctctcctc
-644    caaaatggct tcagaattcc caaattctga tgtagatgct gcacagggta taccagcgtc
-584    tccttagagt ctctcgcggc tgggcccaca gtggcggtta acccagatcc gctccccaag
-524    cgacttgacc ttcactctga gagtgcagct gctactggaa ctgcaatttc ctctcctctg
-464    cttacatatc tgtataaacc cctttatggg ttagcaaatg aaattttata aagacaagtg
-404    tgtaggggtt cccacgaaag cttgaacagg gagtgggagc acccggagcg cggagcctca
-344    gcagccccgg ggcgcttggg gacttggggg ctccggatcc tggggccccg gggtgggggt
-284    gctgagcaca gagggctact gcggagctga aggcgttgtt ccaagcgcca aggatttggg
-224    acccggcccg gagacgcccc acgccgctgt gttcggctcc tggaaggaat tggggtcccc
-164    agccccggac tctccctgcc tcttgccata gccagcccgg tcccggactg cgcatcctcg
-104    gtttcccagc cccctggggt gtctgcaggc cgggctactt gcacagcagc aggtgcgtag
-44     gcggggcgcg cagcatttaa ggcggacacc acctcccctg ggcaGCGGCT GGCGATCGGC
17     TGCGGAGGTG CGCGCAGGGC CCGCGTGGCT GTGGGTACCT CCTTCGCCAG CACCGTCGCC
77     ACTACCAACG CCGCCACCGC GGGACCCTAC CCCGCATCGG TCGCCGCCGC CACCGCAGGT
137    CCCACGACCC CTCCTGCCCT CCGCGCCCCC TGCCCGCCGA CCCGCGGCGC CTCCAGCGCG
197    ACATGCGCCA Ggtgtgtgag gaagtgacgc ggtgcggact ggagagaagt gcgggaaagg
257    gtgaagggct ccgtccgggg gtctttactc tgcaaccctg ttccagccgc cgagcacccg
317    tgtgtcactc gggaactggc tgggtaaaga ggtcaatcca gacacgcggg gaaggagttc
377    caggggtcag ctccgccctc gcacctgcgg gctcggattc ggagaaaagt gctagactgg
437    agctacacgt atgcgtagcg gtctggaaaa tgccccaggc tcgggtctga ggggcccaag
497    tctatgcacc gctggtgtga ccccgcaggg caaccccgcg gttaacttct ctcctgccca
557    cccctagagg tgtcttcctg ggaagacgat ggcaggcggt gcccaccgag ccgaccgtgc
617    aacaggggaa gagaggaagg agggaggtgg gaggtggcgc gctccccaca gcccttcccc
677    tcctggcccg cgagggtgtc cggtcccact caaggcagct gcgcagagcc tgtgcagaaa
737    taccacctgg ggccggtatt gcactctgct tctctttcag AGAAAGCTGG AAATTTACTC
797    CGTGGAGCAC CATGCAGCTA CAGATATCAA GAAGAAGGAG GGGCGAGATG GCAAGAAAGA
857    CAATGACTTG GAACTCAAAA GGAATCAGCA GAAAGAGGAG CTTAAGAAAG AACTTGATCT
```

```
917          Ggtcagtagc cttggaaagg gctcccgcgg acttggtcat ttccctctgg gagttggacc
977          cttcaccggc ctcttacaaa agaaaaatca ggctcaaggc tttggaatcc ccctaacagt
1037         cgtttccttg tggattacta tctagatgtg ataatatgta tccgggtctc aggtcctctg
1097         caggatgcct gagctctggt tgaaggcagt cctggggaga atcgggaggc acaaatccct
1157         aactcccaaa ccacaggcat ggacaagagg tctccctgtc tgcttcaacg agatagaaat
1217         aaccagctgc ccgtttttag ttggaaaaac tatttgggct ggcacctgcc aggaagaaaa
1277         cccaagctac tggaccttgt taggtttggg gggtgtgtgc tgcctgggtg agaaggactc
1337         aggcagtgag aatgttgcac ttggaaccac ttaggttcct cacagcccat tctccatagt
1397         ctgagctgta aacgatctag cctcgtggaa gctcacagaa aatgggaaat aaaagaatct
1457         gtagggatct gtgaggagaa aagcgaggta caagtcttta gaaattaaga tgctttctag
1517         gagtgcaaca ttgttattat caaggatgaa ctctggagtt atttcaactg gggttattta
1577         ataagtatta ataaattcat tagattactt actgaataaa gagctgctgg ggatctgaaa
1637         ggtgacagat tgcatctcag gaatgtcagg gagggagaga gattttgctg gattgccctc
1697         cagcccctca cagacgttac ggctctatga ctggtaggca gccgctgggc tggcagaggc
1757         cagcatgctc aggctagaca actgattgtc aaatattcag gaacttggcg agccgactga
1817         gattctgttc acagctggaa atttggctct ggtcagaaat cagtggttgc cttgaaccag
1877         agccatttgt ttatcttgcc tggagttccc atgtgtgaat gacactcact agtttcctat
1937         tttgaaaaaa attgggatta aaaatggtgc ctgccaccag aggttattaa gaagattaag
1997         tgagataaag caggaaaaaa gcactagtca atagtcatgg ttatttccta tgtatctctg
2057         gcatttaaga ttaaaaacac atctgttctt ccttgaaatt gctcttgtac cattctctcg
2117         tatacctatt ggctagatga aaatcccccc tttctgagca ctttatgagg atgtatgata
2177         caatatagaa gaatagcttc cattgtatta ggcaaaaagt tccccttcct ccttgctggg
2237         aaaaagagct cctttctttc cttccttgtt ctcaagctcc tcacaagaat tctgcagttt
2297         gcaagagtgc acttcaatac tgctgcctct ccacctttcc tgcaaatcag agccttgata
2357         actaagcacc cctggcttct gaataaagga ctccgctgca gggtgcttcc ctgtgctaag
2417         ccagctagct ggcagaaggc acccagacac ccagcactcc tcacaccact taactcgcag
2477         gctcccctcc aagtccaggc tggaaagaca agaaaggtag agtgctttgc ctcacagtgc
2537         taccaggtct gcggtccatc cacacaaatc tttactactg tgtctgccca gtgctcacac
2597         agtctgccct gtaaccttcc atgcccctag caacaacaca tcccaatgac actggcacaa
2657         tgcctgatga gcagagcaga atctgactga gaagagagcc agaaacaaaa taacaatacg
2717         tcatctttta ttggaaggct taatttccta atgcaacatg aactattact ttttagcatg
2777         catttaaaaa gagagatggt tctagtacca gtcacagcat aaatcttcat ttcagtgaat
2837         cagggtatga agaatcctga gtgctggggt tttcttgggg gtaggagggt gaggaagata
2897         tgccactgga ttttctgcta taggaagagg tttcttttca tccgttagcc ccgtggttct
2957         tgaaaggaga agagtgagtg gattttgtcc catccctttc ccccacaaaa acgcattggc
3017         ccaaacgtta ggaagtctgc attagtctaa ttagtctaat gattctgtgg caaaatctgc
3077         aagtgtgagc catactggta tggtttgttt gtcgggtttt tagtccctac cagtgaactt
3137         gttgcatttc atagggcaaa taagggaagg tgtgcttctt tctgccagag accacctagg
3197         agcaatgagg agttgtccag agaaagtggg tctggtttgt tttgctcctc agcctcaacc
3257         aatcctagaa ttcaggtctg aatgaagtat ccagcttgct ggtacttgga agatgtgcta
3317         gagatgagat tgacatgata tgtgacttcg ttgataaaaa ttacagatgc accaggtagc
3377         cagcagcagg tgtaacctga gctctttggg ctggccatgg tgcaggacag ggctctcctg
3437         tagcagtgtc ttccctggct gcacagctcc tggcaccaag ggtccctgtg ttcctttcct
3497         gcaccctcca gcccctccta gccacagtgc tactccagct gtcagggggt ggtcagagct
3557         tgggctttga gtcatttagc attatgtcac tcattctgct cctcttctga ttgtgcattt
3617         CAGGATGACC ACAAACTCAG CAATAAGGAG CTGGAAACGA AATATGGCAC AGACATCATT
3677         CGGgtaagtt ataagggaag tgggcatcag gcaaggaagc ctgcggacta caaaggggaa
3737         agctatatgc tctgcatgtg tggaaaagta agagggcacc tttggggcag gtggcctgcc
3797         acacgttgca gtcatgggca cacatgtttg gcttcccatg ctgcaccaag gctatggctg
3857         tgtccctaca cagagggcaa tgcttggttg tagagagttc ttccatctgt gactgcagga
3917         aggctgtggt gcctaggaga tatgcgattg gatgctttta atctctgcct cctgggagtg
3977         taatgaagga gaagctgggg gatccagtat tagaaggaaa agacccaggg gtgcggtgtt
4037         tggtgcaaca gttgagacac cacttggtac actctcctcc catattggag tgcctagatt
4097         tcatatccca ctctggttcc ccactccagc ttcatgacag cacagtccct ctgaggcaac
4157         aggaaatggc tccttaggtt cctaccacct gcgtgagagg cctggattga gttaccagct
4217         cccagtttct tcctggccca gtcttggcca tcatagtagc catttgggga gtgaaccaga
4277         agacaggagt ttctctctct ctctctccct ctccaccccc caccccgccc ccatcacatt
4337         ttttaaaatt acaaagaaga agacacagac ccagagcaac atctgaaacc ctgctggtca
4397         gccagacccc ctgagaaggg atgatagccc caggcacacc tctgtgattt ttctgcctct
4457         tacatattct ctccccactg gagcactgct tccctctggg aattcaagtt caaattacat
4517         ttccatagtt tttgtttatt gacatcagga ggtctggata ctgagtgtac atctgtgtaa
4577         atcaatacaa atccctacta tgtggttcac attggactgg cactggggca ggtgacctcc
4637         ctgaccttaa ggaggcagct gggtagggaa gtgaacatcc agcctatccc atgccaatgg
```

```
4697        ccaatgaatg gttgagacca tctgttggga gaatgatgaa ttctgaccag agggaccagg
4757        gagcgcagag tttctaaaga gggaagcaac ttgatacaat tcactttcac caggcaattc
4817        tttaagtgga aggtggaagc ttctagaaaa atgaaataaa ccaggactgt gaaatgggga
4877        tcccaaagag gggagcccag taagatggga aggaagggga caggtttaag agggagttca
4937        cagaaataac cagtagctct tgggggatac tttttgagtg aatccctaac tcaagttaga
4997        gaagtaggag gcagagattt ggccaagttg gggcgggggg cagctgggtg aaggtgagag
5057        attagaggag gtggtaggga ggagatgatg gtgagtttta ttttaaatta atcagccctc
5117        aatatccata aggtccctcc ttgatagtaa aatccacagg tgctttagtc atttatataa
5177        aatggcatac tacttgtata taacttgtat gcatatcctc ccttatactt taaatcatct
5237        ctagattacc tataataact aatacaatgt gaatgctatg taaataggtg ttgtatggta
5297        ttatttacag tacaaagaaa aaaagtctgt acatgttcac taaatatgca accaccacag
5357        gatccaaggt tcattgaccc catgggcctg gaacccatgg atacaaaggg cagattatat
5417        ggttaggaat gactctgcaa tatctgtgca gaaatgacca acagttattg agaaggatga
5477        gccagggttc tagaaaagga ttgtggttag agggtgatgg catcaacttg gtgaccaaga
5537        atccagctac tcaataatta cgggagttca ctagtcactt aaaaataaaa aagcattaga
5597        ccctaaaaaa gacacagaaa acacagaaat gatcaagaca gagtccttgc cattatgtag
5657        ctggcaacat aattagggag aaaatgtgtg tacaaaacta gtcccataaa aggtagtatt
5715        tggtgggga agccctatgc aaatgcagga ttttggagtg cgggtatcct ctatgcacct
5777        gtgaatgtca ctctagatct tgggtgggat tggtaactgt taaaacaaac aagaactggg
5837        aacttagaat gcatatgtag aaaagtacca gagtaaagat ggaaaaatca gcccacatgg
5897        tctacagagt tccagaatca gttgtatttt cttcagggtc actgcattgt ccttaatcct
5957        gatgacaagc tcctttattt caaatactcc ttatccaagt tacttccttc tgtggatgaa
6017        tctctagaat gctaaaatgc ctaaacatgc agggggcttt gaaccggtca cctggctgag
6077        ttctggtgcc acttttttaac attttttaag attttattta tttatttgag agctacagtt
6137        acagacagag agaaggcaag gagacagaga gaaaggtctt ccttctgctg gtttactccc
6197        cgaatgggtg caatggctgg agatgagcca atctgaagcc aggagccagg agcttattcc
6257        agggcttcca cgtggatgca ggggcacaag cacctaggcg atcttctaac tgctttccca
6317        ggccaccaca gagagctgga taggaagagg acatgaactg gcacccatat gggatgcctg
6377        caccacaggc aaaggcttag cctactacac cacagtgcca gccccctggt gccacttctt
6437        accagccagg tgaccttagg aaggtatgtc acctccgtgt gcctcacatt ccatgtctac
6497        agatgagaat aataatagta tcctcttcag agccattctg aagactaagt gagtaaactc
6557        atggggtata cttgctggct atcagtatta aatattcctt ccagaaaaat gggtattgcc
6617        catatttctt gggactttct ccagtgtgtt aatgcctgtg tgccacctgc agaaacagga
6677        agcacagcct caaagggagc aggatgtctc actccagttt gcttttgtgc tttttgtgtt
6737        tcagGGTCTC TCCAGCACCA GAGCTGCTGA GCTCCTGGCA CAGAACGGAC CCAACGCCCT
6797        CACCCCTCCC AAACAGACCC CAGAGATCAT CAAGTTCCTC AAGCAGATGG TGGGCGGCTT
6857        TTCCATCCTT CTGTGGGTAG GAGCTGTCCT GTGTTGGATC GCATTTGGGA TTCAGTATGT
6917        CAGCAATCCA TCTGCCTCCC TGGACAGCgt aaggctcctg gtgtcaactt cctggttttg
6977        ctctgcaagg ggtggggtcg gttatgggag gggagagcac agctaaaata aaggccatca
7037        gcgacatctg tgagagcagc catgggctgg aggcaggatt ctaccatgga tcaaccactg
7097        ccagggtgtg ggaaaatcag gacttttccc agatgtagaa caaacgctct gtgacttttct
7157        gatgcctcag tgggtagagt aagcatggat tttagaagag ggagagaaag ccaaaacagt
7217        ttttaagttc ctcaaggagt cgacatccat gtgaagtgtc tagggccttc cctccttaaa
7277        gctgacttttt cccttgaacc tggtaagcat ctgttcctgc tgggagccag cgaccaaccc
7337        cacaacacag agctggccgg catggtgctg ggaaaggagc cagatgcaag tgcctcaagt
7397        tgtgtggctc ctcttacaca gagcacccag atgggcatat ctgtggagac agggagctca
7457        ccagtgggga cagggatggg gcaggggcag gaggtgactg cctaggggcc tgatggccat
7517        gctgcagagc tgaatgacag ctgcacagct ctgcacattc aggagaaacc ggtgagttgt
7577        gcacttacaa tgagtacagg ctgggtatca ggatgcaaat gccagcagct catccctgtg
7637        gagagctgta gcccaggacc tggctcatac ttgacccagg ttgatgagcc gtcatctgca
7697        atgcttgggg ccagaatatt tgcacacatc cataacgaga tgtcttgggg atggaaccca
7757        agtctctaca tgaaatttat ccttcacaca cagcttacac acatagcctg taggtgattc
7817        ctgagccata tttttaatgc acttgtttta ttgtggttgt catgtgaggt caagtgtgag
7877        atttttcact tgtggtatca tgtcagtact caataagtgt caactctgga acctttcaga
7937        ttagggatgc acaacccaag catgtgtctc tgactgatgc ttttcttcct taaccagGTG
7997        TACCTGGGCA CTGTACTTGC CGTGGTTGTC ATTTTAACAG GAATCTTTGC CTATTACCAA
8057        GAGGCAAAAA GCACCAACAT CATGGCCAGC TTCTGCAAGA TGATCCCCCA Ggtgagcatc
8117        agttgccttt cctctggtct cagcacaact ttcaggtgag gaagcatcag acaggccctg
8177        ggggcagggt cgctgcgtag ccctgcccca tctgtgctgt ctgagaaatc ttgagcgggt
8237        tatccctgct ttctagaccc cagctcctca cagggtttag aggtgatcga actcagtaag
8297        atcgggcaaa atgcttttca aaacgtgagt tgtcacacaa catcttccat taccaaaaag
8357        tttgcagaaa ttgaattaaa agataagttt attttggtgc aaaaaagaaa ttggaatcca
8417        tgcatagttt ttttttttcat aatccacatt ttccatgaac ttttgatggc ccctcgtact
```

```
8477        gttccttggc atttctgggc cagctgggtg gtgacatggt ggctccaagg agggaaccta
8537        gaattccaag ggctgtttgt gcactggcag ccatggagga ccccttgtgc tcctctccat
8597        gggactgctt ttcctcctat ccctgagcta tagtagaagt tgcccccaaa tgctctgggt
8657        aaggggaccc agcattcccc tcttccccca agagccctgc cagctgcgca cagcagggag
8717        accagtgttc ccaccccaga caacagccag ctattagaag gggaaggctc tggctagtga
8777        agatgcctcc taagaagtgg gccatcaatc acctcatcaa agttccccca gccttatcag
8837        ctgcataagc tcctgggctc ctcccaggca cccagccatg taattgtccc cctctctggg
8897        agatccagag ctctcacttc tccacgtcca cctgacagca gcattttctg gttctgtgag
8957        cacaggaaaa gattagttct tggggtcttc agccccgccc gcggcggaag tcttccattt
9017        catcctcgct cttgggacgg ccatgggtac catacctctg ttctgggatt cggtgaggtg
9077        tggatggaga aggaggccag catcaccgcg ggtctgaggc tctgtcatgc tttgcttcca
9137        cagCAAGCTG TTGTCATCCG TGACTCGGAG AAAAAGGTTA TCCCTGCAGA GCAGCTGGTG
9197        GTGGGGGACA TCGTGGAGAT TAAAGGAGGT GACCAGATTC CTGCCGACAT CAGGCTGCTG
9257        TCTGCCCAGG GGTGTAAGgt aagggccggg ggaaccccga ggatcctgct tggagcgagc
9317        aaacattctg ctttttccct ggccgtggta ggcgaggcag ttccctctga gctgctgtca
9377        aacttttttcc tagGTGGATA ACTCATCTCT TACTGGAGAG TCTGAGCCCC AGTCCCGCTC
9437        AAGTGAGTTC ACCCACGAAA ACCCCCTGGA AACAAAGAAC ATCACTTTCT ACTCCACGAC
9497        CTGCCTGGAA Ggtaagccca gcagctgtca gggtcgcaag ccccaccttg tcactgctga
9557        cccctaaggc cagagggttt ttcgctgcct tccctcgtta caaagctttc ccccacccca
9617        agaagcatac agtggaggga gccaccaagg agggccttga ataaaacgtc tacttctccc
9677        atagGCACGG CAACTGGCAT GGTCATCAAC ACGGGTGACC GGACCATCAT TGGCCGCATT
9737        GCCTCCTTGG CTTCAGGCGT CGGGAATGAG AAGACGCCCA TTGCCATTGA GATCGAACAT
9797        TTTGTGCACA TTGTGGCAGG AGTGGCCGTC TCCGTCGGCA TCCTGTTCTT CATCATCGCA
9857        GTGTGCATGA AGTACCACGT CCTGGACGCC ATCATCTTCC TCATTGCCAT CATTGTGGCC
9917        AACGTGCCTG AAGGCCTCCT GGCCACTGTC ACTgtgagtc tacactgtta ggcagcctgc
9977        agcccaactg tggctgctgg ggaggctcca catgcagacc acaagtctcc ctggcccctg
10037       tcttcagaac agggacaacc aaggaacatt ctggaaagac tgcttcctag agggacaaat
10097       ggaattgctt atttctaatg tttgaatgca acctccttga aaacatcaag atttcatgcg
10157       atatctctgt aattcacaga ggttacttag ctcccttaag atctagagag gacagacctc
10217       caaccaccac ttacagattt tctagaagcc tgacgcctcc ccaaagttaa aaaaaaaaaa
10277       aaaaaaaaag aaagagaaaa agatgtccag gtctcttctg actccaggag atagaatttg
10337       gagaagctgg agccaacaga acccttggag agctgagggc ttaaaagatc tttaaaatat
10397       ggggctcttg acatgaaata caaatgcttc tggcctcccc agagcagcca gtcagggtct
10457       ctctcacaca ttcatgctcc tggtgacagg atcagggtgg cccaggcacc ctaattggat
10517       gcaagcacca gccacagcaa tgtggctgtc tccactacct ataagccctc gggattatat
10577       tcttcaagaa ataaaggaac aagaatagtg tctcatttaa ttttttttaa aaaggattgt
10637       gacctgtaag acttttagtg gcttacacat tggcagttct gccttaatgc taaaacaaac
10697       ccagcccctc ctgaaattcc attgtagctc atcttagctg tttatacacc tgaactccag
10757       atcaaagtta ccttgcccaa agcttctttc cctctctctt ctgccagGTG GCCCTGTCGC
10817       TCACAGCCAA ACGGATGGCC AAGAAGAACT GCCTGGTGAA GAACTTGGAG GCAGTGGAGA
10877       CCCTCGGCTC CACCTCCATC ATCTGCTCTG ACAAGACTGG GACTCTGACG CAGAACAGGA
10937       TGACCGTGGC CCATCTGTGG TTTGACAATC AGATCTTCGT GGCCGACACG AGTGAAGACA
10997       ATTTAAGtaa gacttttagg gtggggcatt gagcctgggc tgaagctgtg gactcaagca
11057       gtggctaggc tgagcactgt gaccagcaca gtgcaggcat ggtgggccca gcacccatag
11117       cagccattct tctttctcct tgtttcagAC CAAGGCTTTG ACCAAAGCTC TGGAACCTGG
11177       ACCTCCTTGT CCAAGATAAT AGCATTGTGT AACCGAGCTG AGTTCAAGCC AGGAGAGGAG
11237       AGTGTCCCCA TCATGAAGgt aaaatttcca catgcttgtc ttaaatcact gctggtttta
11297       ggtcactttc tctgcccacc ttgggtgcgg ggggacagag aatctctctg catccaagca
11357       gagttggtac gagcaagagt aaatctgact tacaagacac atagcaagct ttctccactt
11417       tagttctgag aaatcacaaa gttgcttaaa agatcttggg cccttctgaa aattcacagt
11477       atgggagatg aatgtgcctt tttatttctt tcctaacttc tttaccagct aagtggagat
11537       agtcaaatgt tttgccattc actgaagttt aaaatataat tagatggaaa taggctggcg
11597       ccacagctca ctaggctaat cctccgcctt gcggcgccgg cacactgggt tctagtcccg
11657       gtcggcgcgc cagattctgt cccagttgcc ccttttccag gccagctctc tgctgtggcc
11717       agggagtgca gtggaggatg gcccaagtgc ttgggccctg cactccatgg gagaccagga
11777       gaagcacctg gctcctgcca tcggatcagc gcggtgcgcc ggccacggaa aggaagacct
11837       ttctctctgt ctctctctca ctgtccactc tgcctgtcaa aaaaaaaaaa tatatatata
11897       tatatatata attagatgga aataaaggac ttggtcttag ttcatcttcc tcttgactat
11957       aagatggtag taggtggaac actaaggggc cctgaaattt ggaaatatac aatattccaa
12017       aattccaaaa gtgcagtgac tcttttctct ctccctatct ctgtccctat ctctgtctct
12077       atctctgtct ctacctctac ctctacctct acctctacct ctacctctac ctcaacaagt
12137       caccaaatct gggttctcct tgcttttcca ttcaccatct gtttctttca ctgactcact
12197       ttctttctta tatttttaat ttgcttttcc ttattagaaa gctcatcaat gctccctttg
```

```
12257        gccaattact tctccatccg tacctttga atctctatgt cctatcctga tttctcatct
12317        aaatcccaac ctggattccc agcaagctgc atgtgtccac tcagaagttc cactgagttc
12377        acctatcaac aaattacact tgtccagttc cacctgtatc taccttattt tgtcctgagt
12437        tttctattat actgtatcat aaccattcca catgtcatcc aggcaaatgc ttttagaagt
12497        cttctttgac tgtctccata aatctagtga atcaccaagt cctatcattc ttgcttcaaa
12557        atacacataa ttttatactt ctatgatcat taataccttg catctttttc aaagtttttt
12617        ttatttattt acttgagaag tagagttaca gacagtgaga ggaagagaga gataaaggcc
12677        ttccacctgt tggttcactc cccaaataga cataatggcc aaagctgagc tgatctgaag
12373        ccagcatcca ggagcttctt ccagatatcc caagtgggtg cagaggccca agcacttggg
12797        ccatctttta ctgccttccc aggccataac agagggctgg atcagaagta cagcagctgg
12857        gtctcaaact ggtgcacata cggaatgcca gcactgcagg tggcggcttt atctgctatg
12917        ccacagcact ggccctaata ccttgcatct taatagctgc accaacctcc ccagattcca
12977        gaatatgcct tccacagtct ttcttcaatt ttccttaaaa attgctttta tcttcaccta
13037        actgtgatca aaatttctcc tgcttcccta ttttctgctt cttcagtcta gacctcagcc
13097        tgtctttcag ggacctgtct aaccatactc ggtgtccacc tgtactccca gcactcacag
13157        tgcaatcttc cactccatgc tcggtgtgta gaatcctggc ttcttctgct cctgtgtctc
13217        tcggggaccc atggctatca agttctagtt catagctgct tttttccctg aagctctcca
13277        gtcttcacag agctgccacc catttgaact tccctagtac cggtaagtgg caccactgtt
13337        ttgcactctt gccatggtca tgactgggtc actgatggac aagtgagcac cttgtccttc
13397        tgaccaggta gcaagctgag ctaggagatg agtggcttgc ttctaaaata atgttggtcc
13457        attaaaaaca atcacaaaag aaggtgtggt atttgcaggt gatatactgc aaatatttgt
13517        ttctctcttt tgcttcccag AGAGTCGTGG TTGGAGATGC TTCAGAAACT GCTCTTCTGA
13577        AATTCTCAGA AGTCATTTTG GGTGACGTGA TGGAAATTAG AAAAAGAAAC CACAAAGTAG
13637        TCGAAATCCC TTTTAACTCA ACCAACAAAT TTCAGgtgag catttcctca tagtcgacaa
13697        tctctgttat cactagaaca acatttttac ttgtacacat tctttaatag cccattgttt
13575        gagtaataca tgacttcaaa agtggtcttc agggaacaga gccttggcac ctgttcaacc
13817        caggcaaggt ccattgcttt ccctgatgta acagagggaa aaaagaaaga atctatgcgg
13877        gcctggctgc tgctgtgaat gaccatctct attgcgtcca cactcactca tgctcacaca
13937        cactcataca tgcctccacc accaccacca ccaccaccat caggtaactt ttgcacaggt
13997        gaggagcagc agccctgttt cgtagggtaa atgagaggct ggggaagagg agagacctaa
14057        tttttatggc tcgctcagta cagttgcccc tctgaatctg tgcattcaat attagtgaga
14117        aaaaaatgtt gttgaatttg tgtagacttt tgttcttgtc catattccct aaacaataca
14177        gtacaacagc catcaagcat aacattgaca ttgtgtttgg tagtgtatgt aatctggaga
14237        ttttttaaat ctacaggatt acatgggctc tgcacaaaca ctatcccatt ttaggtgagg
14297        cacttgagca tccccagttc ttagtatctc aagagtcaag ggcgaattcg ttaaacctgc
14357        agatactcct tggatgacta ttttcc
```

Sequence Data from Products Exons 12, 13 and 14

```
60           aagggtcgga actaagaatc aagcgtagcc aggtggtccg tgctggaaca taccaggtta
120          gatgcatctt ccaggcatga accagactcc gactcccgct gcagcgccat ctgggacagg
180          aaagggtctt cctgtcatgt gggcagtgaa agaancaggc cccatgagac ctgctgaatt
240          ctgactcagg gtgcccctgt nggctggggt gatgcccaga tgctcctagt tggccacaaa
300          aagaaaaggg ctagaacatt cctctgaatg ncgttcccct tcctccttta cacctcaagg
360          acagtcatga aaaagtactg tgtctccaag tcctgtccta ttgctaattg gaagggctgt
420          ggcgagcagg ctacatctaa cgtgccctat tcatgctctc tcAGCTCTCC ATACACCAGA
480          CGGAAGATCC CAATGACAAG CGCTTCCTGC TGGTGATGAA GGGGGCCCCC GAGAGGATCC
540          TAGAGAAGTG CAGCACCATC ATGATCAACG GCAAGGAGCA GCCACTGGAC AAGAGCATGG
600          CCCAGGCCTT CCACACGGCC TACATGGAGC TGGGCGGCCT GGGCGAGCGC GTGCTGGgtg
660          agtgtggggg cacagcccct gtctctcccc cagaggtgcc aaaccgaggc tcaggagctg
720          gtggagtagc tgcagtatcc acgcatcaca gaaagaagaa tatgatcaac agtatagcaa
780          cagataagag gaggcttcgt gaggggtctt caaaagggca tggaaatgca tattaggaaa
840          gaacaataca ggcatttaaa atttttttgc aacagaatga cttctttttt tgaaagattt
900          atttatttga aaggcaaaaa tatacagaga gaggaaaaag agagagtttc actccgctgt
960          ttcactcccc aaatggctgc aatagccaat gctgggccag gctgaagcca ggagccagga
1020         gcttcttctg ggtctcccat gtgggtgcag gggttcaagc ccttggccat cttcctctgc
1080         tttcccagac acattagcag ggagctcgat tggaagcaga gcagccagaa ctcaaagtga
1140         cacctatatg ggatgctggc atgcagacag aggcttaacc ttctgcacca cagcgccagc
1200         ccagaataac ttatttttta attccatttt caagggtctt ttgaagtcca ctaacataat
1260         ggcattgagg agtgtggggt cacttccaga caagcataat cagggaagac ttcctggagg
1320         gagttccgtt t
```

Gap in intron 12

```
60          gtcttttcca gGTTTCTGCC ATTTCTACCT GCCAGCAGAT GAGTTTCCAG AGACCTACTC
120         ATTTGACTCA GAATCCATGA ACTTCCCCAC CTCCAACTTA TGTTTTGTGG

60          GGCTCTTATC AATGATTGAT CCTCCTCGAT CCACTGTCCC AGATGCAGTC ACCAAATGCC
120         GGAGTGCAGG AATCAAGtgg ccactgagca gtgtgcccag agctcacggg gacag

60          ccttcagctt gcgcgtttta tggcgccggc tcagcacgct cagccccaaa cctctgctct
120         tggcccGTTA TCATGGTTAC AGGTGATCAT CCCATCACAG CCAAAGCCAT TGCCAAGAGT
180         GTAGGGATCA TTTCAGCCAA CAGTGAGACA GTGGAAGACA TTGCAA


60          AACGCTGCAA CATCGCCGTG GAGCAGGTTA ACAAACGgta agcacagacc caatactgcg
120         tacactcagg gtcttactca caggaaccat gctatttgga gcaattgaag caaattgaag
180         caaacctttt ttnaaaaatg tttatttatt tatttattta tttaaaaatc agagttacac
240         agagagagaa ggagaggcag agagagagaa agaagtcttc catccactgg tncactcccc
300         aattggccac aatggtggga gctatgc
```

Gap in intron 14 and to HKa2.8

Sequence Data from λ Clone HKα2.8 Exons 15-23

```
1           aatggctcta gtttaaactt tctgtttact gtcttgaaat ggagggcaaa gtttccccag
60          tagtcataac ccctgagata gccccagagt attcgcccat ctgggatatt aggttccaaa
120         tggagtcaaa aagaattgag gccacatcca gcctagaacc atttcaacca aagatctgcc
180         cttgggcctc actaggtttc ttctggccct tttagGGATG CCAAGGCCGC CGTGGTGACC
240         GGCATGGAGC TGAAGGACAT GAGCCCAGAA CAGCTGGATG AGCTCTTAGC CAACTACCCG
300         GAAATCGTGT TTGCACGGAC GTCCCCCCAG CAAAAGCTGA TCATCGTGGA GGGCTGTCAG
360         AGGCAGgtgg gtgacagcag cacctggaga atcaactacc ctcagccaca tgtccccagc
420         tctgcatttg atcatgggag tctggggttt cgacagcaac acatttgcca ctgtggcagt
480         gccactgtag aaagatgata aaatagtaag tttcaaaaca aagcttgcag agacatagga
540         tatccaacac ctccatttgc acgcacctta ttcggagctc cttgccttgc atctgagatt
600         tcctgatggt ctcttctccc aacgggacct catcctggcc ccagcaatcc ttgcagcccc
660         aaggacagca ggaggctgtg tgtggctttc gcccgctcac tcactcccct cttcccacgg
720         ggcccccgca gGACGCAGTT GTGGCCGTGA CGGGGGACGG AGTGAATGAC TCCCCCGCTC
780         TAAAGAAGGC CGACATTGGC GGTGCCATGG GGATAACGGG TTCTGACGCG GCCAAGAACG
840         CAGCCGACAT GATCCTGCTG GATGACAACT TCTCCTCTAT CGTCACAGGG GTGGAGGAAG
900         gtgagcgggg tccccaaggt ctgccggagg gccagggtcc acgggaggac tggggacaag
960         ccctctaagg agaaccatct ctgcctagGC CGCTTGATAT TTGACAACCT AAAGAAGACC
1020        ATCGCTTACA CCCTGACCAA GAACATTGCC GAGCTCTGCC CCTTTTTGAT TTACATCATT
1080        CTCGGGCTGC CCCTGCCCAT TGGCACCATC ACCCTCCTGT TCATCGACTT GGGCACAGAC
1140        ATAGtaagtg acactaaagg gaacagggtc cttctgcacc gcaggggagt gcacaggttg
1200        cggttgttac cttggagggc caggagatgg gctgaggccc acgctgctca gacttctctg
1260        caatggagtt caaagttttg cctttctgac aagctcccag gacatagcga ttctgctggg
1320        cccacagacc accacttcaa gggactagag acatttgtca agtgtttgga aattcaaact
1380        tcactgttta cggggatttt taaggcagga gcccctcaaa aaatatcttt tgccaatggt
1440        cagttttagc agccaggcag ccctttttctt tggcgcagct gaagggccct gtgctactgg
1500        gggcttggtg gctaatgtga taaccccaca gcgccccctg tggagggaga agaatgggcg
1560        ctcttaaggg cgggccctag ggttctgctc t
```

Gap in intron 17

```
1           agcatcccac atgggtgcca gttcatgacc cagctgttcc acctccgatc cagctctctg
60          ctatggcctg ggaaggcagt ggaagatggc ccaagtgctt gggcccctgc aactgtgtgg
120         gagacctgga ataagctcct ggctccaggt ttcatattag cccagctcca gccattgcgg
180         ccatttgggg agtgaaccag cggatggaag acctctctct ctgcctctgc ctctgccccct
240         ctgtaacact gcctttcaga taaataacta aatttaaaaa aaaaaaaaag acttcatggg
300         ctatctccct ctccagATCC CCTCCATTGC CTTGGCGTAT GAGAAAGCAG AAAGTGACAT
360         TATGAACAGG AAGCCTCGGC ACAAGAAAAA GGACAGACTG GTGAACCAGC AGCTTGCTGT
420         ATACTCGTAC CTGCACATTG gtacgcctgg gctctcctac tgtcactgga gggcctgtcc
480         caggcacagt cggagagtgt tttcctttgg ggctgaggtt gggatggtag gaggggctga
540         gaggagaaag gatgaggcca tcctctggag gcatcaggga gtctaagccc ttagctctct
```

```
600          tgttctttca gGCCTCATGC AAGCCCTGGG AGCTTTCCTG GTGTACTTCA CTGTGTACGC
660          ACAGCAGGGC TTTCGGCCGA CCTCACTGTT TCACCTGCGG ATAGCGTGGG ACAGCGACCA
720          CCTGAACGAC TTGGAAGACA ACTATGGACA GGAATGGgta agcgctgggg cgctctgtgc
780          cagctttgcc tgcttcttct tccctcctct cgctgccctg ccaggactgg ggcagagcct
840          gcattcctcc acatggcctt gacaagaccg actgtggaag cttgctagtt gatagggctg
900          gggacccctg gaggacactg gtttggcctt tgccaggacc cagccttttg tcttgggcca
960          gaggtcaaga gctcaccttt taatgttagc tttgccattg gcgtggaaaa gtaatcatgt
1020         caacacgccc acagttcagc cagagacgtc agcatctttc tctcctcacc tggctgaaga
1080         ccaaagtttc tctctccatc tgggtcaggc agagatgaga ggaaagcttc ccagagatcc
1140         ccccggactc tacaactttc ctccaagctc tcactgtctc ctccttctct gcccatcgac
1200         agACGAGTTA TCAGAGGCAA TACCTGGAAT GGACAGGCTA CACGGCTTTC TTTGTTGGCA
1260         TCATGGTCCA GCAAATAGCA GATCTGATCA TCAGGAAGAC CCGCAAGAAC TCCATCTTCA
1320         AGCAGGGGCT CTTCAGgtat tgctgtatcg gcctcatggg tcagcctcag gcctgcgggc
1380         ttgctggagc ttcctccggc catgcctggc acacacaagg ccttgctcag ctggtggccg
1440         gttctgcgga gggcatgatc acagagccgt gggaactcag aatctgctta ctttgtcttg
1500         cagAAATAAA GTCATCTGGG TGGGGATCGC CTCCCAGATC ATCGTCGCCC TGCTCCTCTC
1560         TTACGGGCTC GGCAGTATCA CAGCCCTAAA TTTCACCATG CTCAAGtgag ttcgccttcc
1740         acagcagcaa ggaaactgac agccctgccc cgagctgtca cagcccaaat aataactagg
1800         atatcccaag gtccctctcc cacccacccc ccctttttt tttaaagatt tatttattta
1860         tttgaaagag ttacacagag aggcagagag agaaagagag tgagagagag aaagaggggt
1920         cttccatctg ctggttcatt ccccaattgg ccgcaatggc cagaggtggg ccaatcttca
1980         gaaaccagag ccaggagctt cttccaggtc tcccacacag ttgcaggggc ccaaggactt
2040         gggccatctt ctactgcttt cccaggccat agcagagagc tggattggaa gtgacacagc
2100         cgggactcca atcggcgtcc atatgggatg ctggcactac aggc
```

Gap in intron 22

```
1            cctgcttcct gctaatgcag accctgaaag cagcagatca cagttcaagt atttcggtcc
60           ctgctaccca agtgggagac tcagattagc atctgagctc cagacttcag cttggcccag
120          ccctggctat tgcagacatt tagggtgtag accagaagat gggagggccc cctcttactt
180          ctctgtctgt ctctctctct ctctttcaaa taaataaaaa catagaagta ctttcttttt
240          ggacaggcag agttagccaa tgagagagag agacagagag aaaggtcttc cttccttcca
300          ttggttcacc ccccagaatg gccgctacgg ccggcacact gcgccaatcc gaagccagga
360          gtcaggtgct tcctcctggt ctcccatgcg ggtgcagggc ccaagcactt gggccatcct
420          ccactgcctt cctgggccac agcagagagc tggactggaa gaggagcagc cgggacagaa
480          tccgacaccc caaccaggac tagaacccgg ggtgccggtg ccgcaggcag aggattagcc
540          tagtgagcca cggtgccggc cagaagtact tattttaagt acaactaaaa taaggaactt
600          tctggtttga gtgtggtgtg acccagtgga tttggtgtgc ttatcagtaa agcagggcca
660          agggccaagg gcaggaggct ccctgcttgc tggcattcgc acgttccagc gctcatcatc
720          ctgccttctg cctccagGGC TCAGTACTGG TTTGTGGCCG TACCCCATGC CATCCTGATC
780          TGGGTATACG ATGAAATGCG GAAACTCTTC ATCAGGCTCT ACCCCGGAAg tgagtagtgc
840          agcggtatct tggaggttct gtgtgcttcc tctaccagac tctccatttc taatttacct
900          tcttctgact ctctagGCTG GTGGGATAAG AACATGTATT ACTGAGACCA GGTCTGTCTC
960          TGAGTCTCCC AGCGGCACCT GCCTGGTGGT CTTCGGCAAG ACCTCTGTGT AGTGTGGATG
1020         TTGCCAAGCT CCACTCGGGA GGAGACTCTC ATCTAGAACA CAGTGGTGAA GCTTCTTACT
1080         GATCTGTTGT ACTTCAAAGC TGAGATTCAG CTGGTTGTAT ATGATTTTCA TCTCTATCTC
1140         CATCTCCTTA CCTTAAAAGA TGTGGATGTC AAGGTCATGG TGTAGGGAAG GATGTGTTTA
1200         TCTGTATATG AAGCTCACTG ATGTCACACA GACTTGTGTA ACCCAGGTGG CTGCTGGAGT
1260         CTGCCATAAG TTGAGCTAGA ATTGCTCAGA TCTCCTTCCA CACCCTGTCA AAGGCCCGGT
1320         GAGCTCCATA GGATTTCTGT GAATCCCCCT GAAACATAAC TTTTGGGGTT TGCTTTGCTC
1380         AGCTGAGGGT GTGAGTTGGA AGTGTGGCAG CAGGAGCACC TCAGAACAGC AAAGACAGCC
1440         CCCGTTTTGA CTCCCAGACA CTTTGTTGCT GTGATGGGTT CCTGGCCATG CGGCCCCAGT
1500         CCGCCTCTCA CAGCACTCCA CCACCTGTTC CTGCAAAGCT GACCTCCAAG TCCATTCCAC
1560         AAACCTTAAC TCAAACATTC GTGGACCCAA AGGGGCTGTC ACTGACTGGG ACTCGGCCTC
1620         TCCGGAAAGC CACTGTGGTT TAGATAGCAC TATTTATTTC TTGTAGATAG GCTGCCAAGC
1680         ACTCTCCAGC AGCCATTTTA TGTCAATCAC ATTTTTGTAA CTTAGATATA TTTGTGTGGG
1740         ACACGAAACA CATACATCCA TGTTGACAGG TTTTTTTTTT TAAATAAAAG ATGTTTTTAA
1800         GTAAAATGTT TTATGAAACA AAATCTAATT GTGATGTTTT ACTTAATTCA AGTTTTTCCA
1860         GAGGCAGGCA CGGAAAATAC CAAAAAAATA AAATAAAATA Agattctggg ttttttttct
1920         tttttgctcc ttctggtcat tttctttaca cacagagtgt ctggaaatac aggcttttcc
1980         tcgtgagtgc ttcccgcacc tgtgcccccc ccccccctca cctgactggg actcctgtgg
2040         gccagatcaa cctgctggca gcagcaccac aggcaaagtc attttcacag actttccatc
2100         aaccacacac ccacactaac tgatttcaca gactttaaaa gctgtacttg gttaaacttt
```

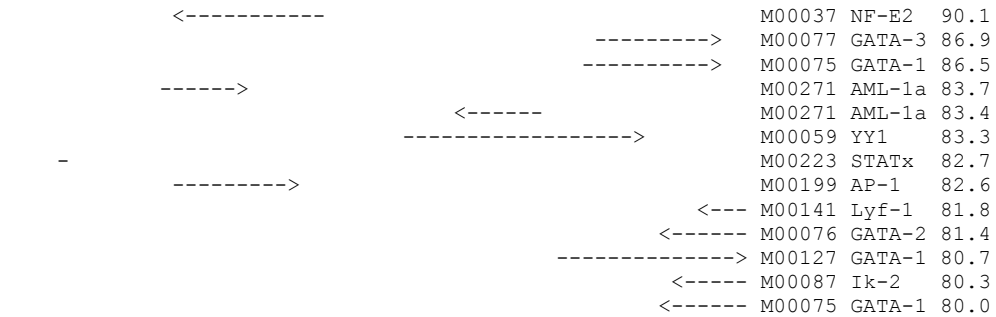
```
2160          gcccattcag tggcctctcc acgccagcca cttgcagctg ctgctgcagg cttcagcttg
2220          gcatctctgg gacaggaaag ggagacgtag agtccat
```

APPENDIX B
TFSEARCH RESULTS

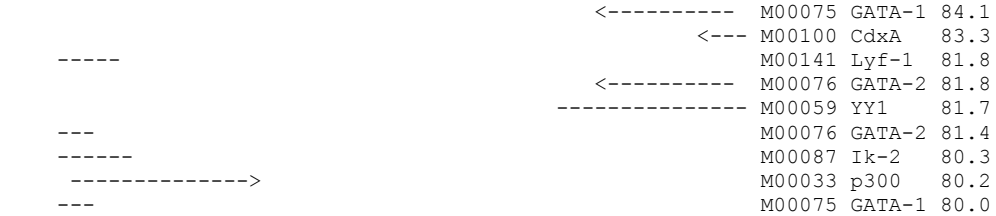
```
** TFSEARCH  ver.1.3 **       (c)1995 Yutaka Akiyama (Kyoto Univ.)

  <Warning> Scoring scheme is so straightforward in this version.
            score = 100.0 * ('weighted sum' - min) / (max - min)
            The score does not properly reflect statistical significance!

 Database:  TRANSFAC MATRIX TABLE, Rel.3.3 06-01-1998
 Query:     untitled (6300 bases)
 Taxonomy:  Vertebrate
 Threshold: 80.0 point


 TFMATRIX entries with High-scoring:

   1 TGGGTACCGG GCCCCCCCTC GAGTTGTAAT CTTGTGTGAT GCTCTTGAAA entry        score
                                <---------                  M00011 Evi-1  89.9
                                                    <----   M00147 HSF2   86.5
                             ------------->                 M00137 Oct-1  84.7
                                                    -----   M00147 HSF2   84.6
                                          ---------->       M00075 GATA-1 84.5
                                                    -----   M00052 NF-kap 84.1
                                                      <--   M00074 c-Ets- 83.8
                                                  <------   M00240 Nkx-2. 83.7
          <----------                                       M00008 Sp1    83.6
                                                    <----   M00146 HSF1   82.6
                                                      <--   M00109 C/EBPb 82.4
                <----------                                 M00008 Sp1    82.2
                                          ----->            M00271 AML-1a 82.1
                                                    -----   M00053 c-Rel  81.8
                 <-------                                   M00083 MZF1   81.7
                                          ---------->       M00076 GATA-2 80.6
                                                     <---   M00261 Olf-1  80.6
                                                    -----   M00146 HSF1   80.4
                               <-----------                 M00079 Evi-1  80.3
                <-------                                    M00083 MZF1   80.0

  51 TTTCCTAAGG GAATAGATTG GGGGTTGCTT TTATCACAAA AAAAGATTGT entry        score
                   --------->                               M00077 GATA-3 89.4
                                                --------   M00077 GATA-3 88.4
                                                     ---   M00045 E4BP4  87.3
                                <-----------                M00203 GATA-X 86.6
     -----                                                  M00147 HSF2   86.5
                 ---------->                                M00106 CDP CR 85.8
                               <-------------               M00128 GATA-1 85.7
                                                --------   M00162 Oct-1  85.7
                                                   <-      M00228 VBP    85.6
                                            <------------   M00130 HFH-2  84.7
     ---->                                                  M00147 HSF2   84.6
                                            ------->        M00148 SRY    84.5
     ---->                                                  M00052 NF-kap 84.1
     ----------                                             M00074 c-Ets- 83.8
                                                      -     M00039 CREB   83.7
                                                    ----    M00109 C/EBPb 83.6
                       <-----------                         M00072 CP2    83.3
                                  ------->                  M00100 CdxA   83.3
                               <--------------              M00127 GATA-1 83.1
                            <------                         M00148 SRY    82.7
     -----                                                  M00146 HSF1   82.6
                                               ---------   M00075 GATA-1 82.4
```

```
           --------------->                          M00126 GATA-1 82.4
------------                                         M00109 C/EBPb 82.4
            ---------->                              M00104 CDP CR 82.1
                                  <-----             M00271 AML-1a 82.1
                                              ---    M00271 AML-1a 82.1
                                    <-------         M00100 CdxA   82.1
    ------------>                                    M00087 Ik-2   82.0
---->                                                M00053 c-Rel  81.8
   --------->                                        M00227 v-Myb  81.8
             ---------->                             M00075 GATA-1 81.6
                                             --------- M00076 GATA-2 81.0
                                                   <-  M00260 HLF    80.9
                              <-------------         M00159 C/EBP  80.8
                               ------->              M00101 CdxA   80.7
                                                <---  M00116 C/EBPa 80.7
                               <---------            M00077 GATA-3 80.6
--------------------                                 M00261 Olf-1  80.6
---->                                                M00146 HSF1   80.4

101 GATGTAATGA TGTGAGAGGA TGGATAAGTC AATTTGCTTA ACTACTTTGT entry        score
    ------------->                                   M00137 Oct-1  93.8
                        ----------->                 M00203 GATA-X 93.6
    >                                                M00077 GATA-3 88.4
                                  --------------->   M00116 C/EBPa 87.3
    -------->                                        M00045 E4BP4  87.3
                                  --------------->   M00117 C/EBPb 87.2
     ------->                                        M00101 CdxA   87.1
                                    <------          M00148 SRY    86.4
                                               <---- M00148 SRY    86.4
                 ---------->                         M00075 GATA-1 86.1
                 ---------->                         M00076 GATA-2 85.8
    ----->                                           M00162 Oct-1  85.7
    --------                                         M00228 VBP    85.6
                                  ------------->     M00159 C/EBP  84.6
    ------>                                          M00039 CREB   83.7
    --------->                                       M00109 C/EBPb 83.6
                                  <---------------   M00109 C/EBPb 83.6
                           ----------------->        M00099 S8     82.8
    >                                                M00075 GATA-1 82.4
                           --------->                M00199 AP-1   82.1
    -->                                              M00271 AML-1a 82.1
                           <---------------          M00162 Oct-1  81.6
                                           <-------  M00160 SRY    81.1
    >                                                M00076 GATA-2 81.0
    --------                                         M00260 HLF    80.9
          -------->                                  M00041 CRE-BP 80.8
                                            <------  M00137 Oct-1  80.7
      ------>                                        M00101 CdxA   80.7
                                               ---   M00101 CdxA   80.7
    ----------                                       M00116 C/EBPa 80.7
                     ---------->                     M00106 CDP CR 80.4
                                  <----------        M00260 HLF    80.1

151 TATGAATAAG TATACATATT TATTTTATTT TATTTTATTT TATTTTAGGA entry        score
                 ------------>                       M00131 HNF-3b 90.2
                     ------->                        M00101 CdxA   90.0
                     ------->                        M00100 CdxA   89.7
    <------                                          M00101 CdxA   87.9
                  <-------                           M00101 CdxA   87.1
    --                                               M00148 SRY    86.4
                       <---------------              M00081 Evi-1  85.2
                            <---------------         M00081 Evi-1  85.2
                                 <---------------    M00081 Evi-1  85.2
                 ------------>                       M00130 HFH-2  85.1
                  <-------                           M00101 CdxA   85.0
                      <-------                       M00148 SRY    84.5
                           <------                   M00148 SRY    84.5
                                <-------             M00148 SRY    84.5
                                     <------         M00148 SRY    84.5
                                          <-------   M00148 SRY    84.5
     -------->                                       M00199 AP-1   83.0
```

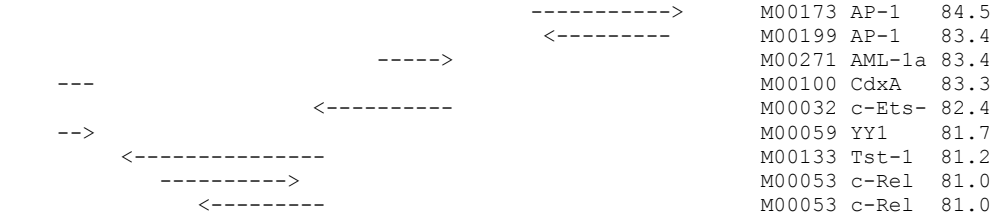
201 ACCAGCACTG TGGCTCAGCA GGTTAAGCCA CCATATGCGA AGATAGCATC entry score



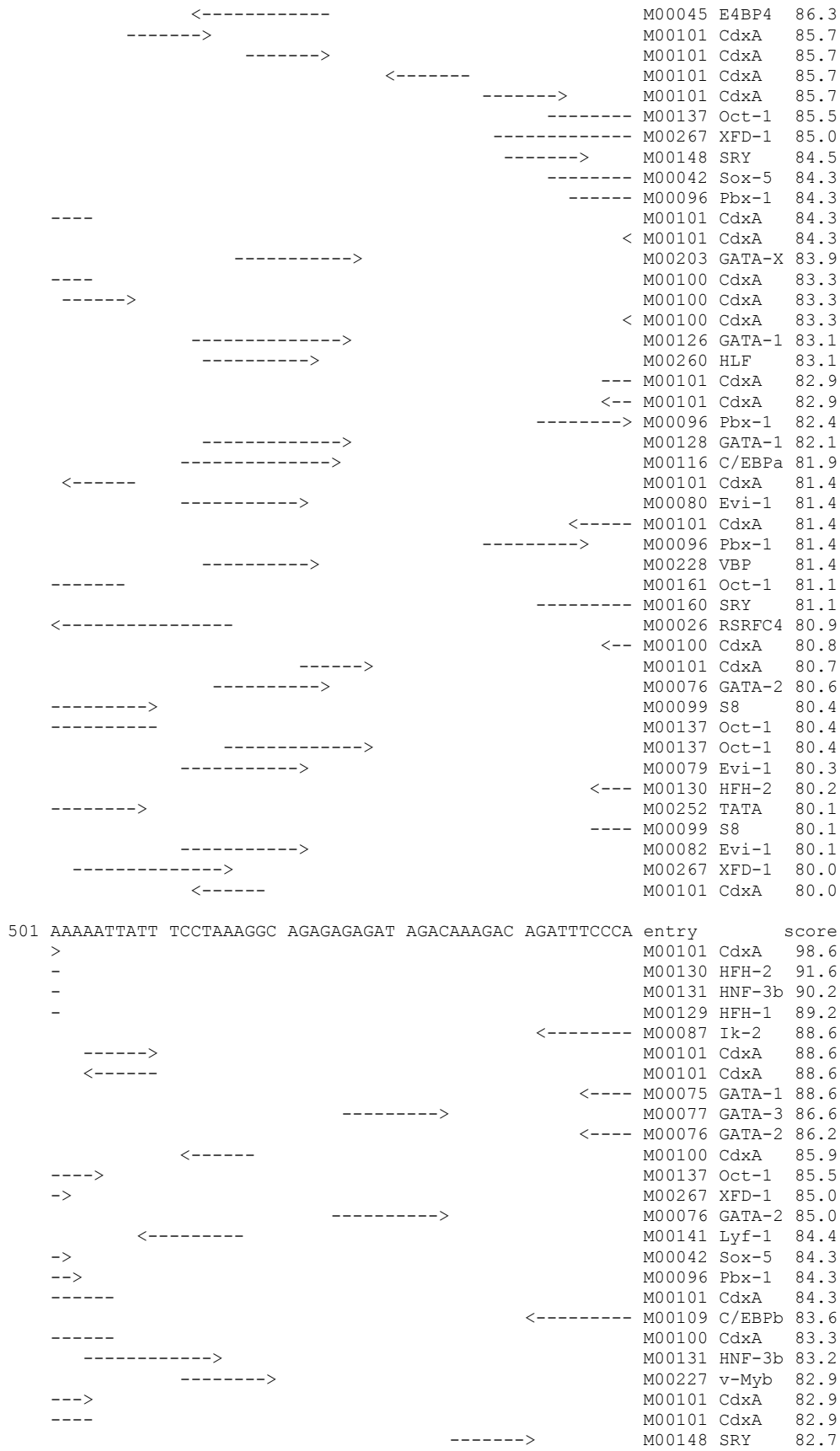
251 CCATAGGAGT GCTGGTTTGG GTCCCAGTAC ATGGGCCCAT GCCATCTACA entry score



301 AAGGAGACCT GGAATTCCAG GCTTCTGGTT TCAGCCTGAC CCAGTCCTAG entry score



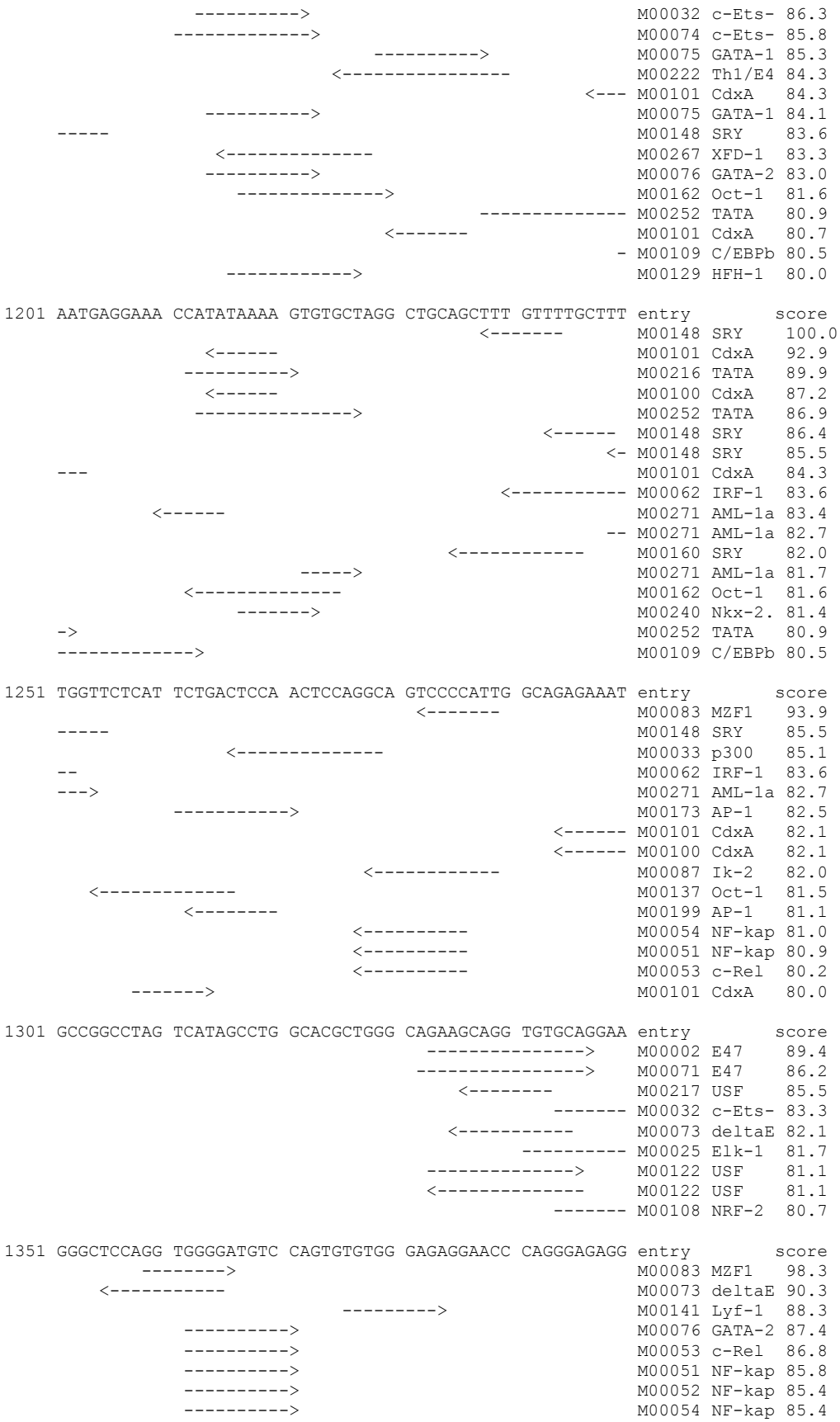
```
          <-------                                      M00101 CdxA   80.7
              ------>                                   M00101 CdxA   80.7
                     <-----------                       M00108 NRF-2  80.7
                                 --------->             M00199 AP-1   80.6
    ----------------->                                 M00222 Th1/E4 80.4
                     ------------------>               M00222 Th1/E4 80.0

351 CTATTATGGC CATTTGGAGA GTGGACCAGA GGATAGACTA TTTCTCTCCT entry       score
     ------>                                           M00101 CdxA   90.7
                               ---------->             M00076 GATA-2 90.5
    <-------                                           M00241 Nkx-2. 88.2
        ------------------->                           M00059 YY1    86.5
     <------                                           M00101 CdxA   86.4
              ---------->                              M00159 C/EBP  85.4
                               ---------->             M00075 GATA-1 84.5
                          <-----                       M00271 AML-1a 83.4
                                          ------->     M00101 CdxA   82.1
                                          ------->     M00100 CdxA   82.1
                     -------->                         M00083 MZF1   80.0
                                                <---   M00083 MZF1   80.0

401 CTCTCTCTCT CTCTCTCTGT AACTCTGACT TTCAAATAAA TAAATTATAA entry       score
                                               ------>  M00101 CdxA   98.6
                                                 <----  M00101 CdxA   92.9
                                       <-------         M00101 CdxA   90.0
                                           <-------     M00101 CdxA   90.0
                                       <-------         M00100 CdxA   89.7
                                           <-------     M00100 CdxA   89.7
                                      <------------     M00131 HNF-3b 89.6
                                     --------------->   M00267 XFD-1  88.3
                                      <------------     M00130 HFH-2  88.2
                                                 ------- M00216 TATA   87.4
                                        ---------------- M00099 S8     87.3
                                                   <---- M00100 CdxA   87.2
                                        ------->         M00101 CdxA   87.1
                                        --------->       M00096 Pbx-1  86.3
                                      ------>            M00148 SRY    84.5
                                          ------->       M00148 SRY    84.5
                                                    <--  M00101 CdxA   84.3
                                                ------>  M00100 CdxA   83.3
                                                   <--   M00100 CdxA   83.3
                                       <------------     M00129 HFH-1  81.5
                                             <------    M00101 CdxA   81.4
                                              <------   M00161 Oct-1  81.1
                                       <------------     M00130 HFH-2  80.9
                                         --------->     M00096 Pbx-1  80.4
                                                ------   M00099 S8     80.4
                                                   <--   M00137 Oct-1  80.4
                                                ------   M00252 TATA   80.1
    ----                                                M00083 MZF1   80.0

451 AAATTATAAA TATATTAGAT AATGTATAAT AGAATAGAAA AATAAATAAT entry       score
        <-------                                        M00101 CdxA   100.0
     ------>                                            M00101 CdxA   98.6
                                                ------   M00101 CdxA   98.6
        <-------                                        M00100 CdxA   96.2
    --                                                  M00101 CdxA   92.9
                                          <----------   M00130 HFH-2  91.6
                               ------->                 M00101 CdxA   91.4
                               <-------                 M00101 CdxA   91.4
                                          <----------   M00131 HNF-3b 90.2
                                            <------     M00101 CdxA   90.0
                                            <------     M00100 CdxA   89.7
                                          <----------   M00129 HFH-1  89.2
    -->                                                 M00216 TATA   87.4
    >                                                   M00099 S8     87.3
                     --------->                         M00077 GATA-3 87.2
    --                                                  M00100 CdxA   87.2
         ------->                                       M00101 CdxA   87.1
                                       <-------         M00101 CdxA   87.1
                                              ------>   M00101 CdxA   87.1
```

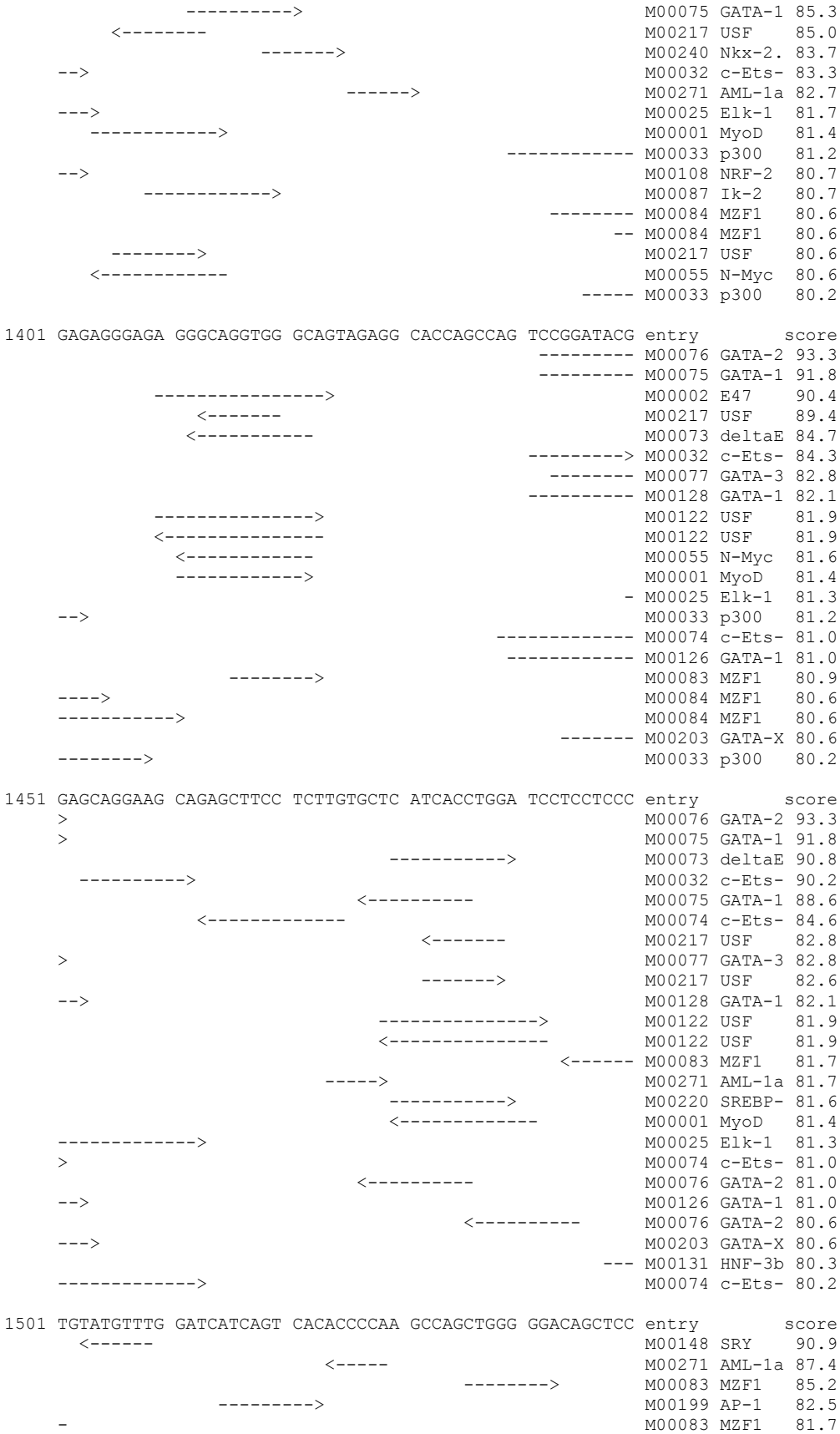

M00045 E4BP4 86.3
M00101 CdxA 85.7
M00101 CdxA 85.7
M00101 CdxA 85.7
M00101 CdxA 85.7
M00137 Oct-1 85.5
M00267 XFD-1 85.0
M00148 SRY 84.5
M00042 Sox-5 84.3
M00096 Pbx-1 84.3
M00101 CdxA 84.3
M00101 CdxA 84.3
M00203 GATA-X 83.9
M00100 CdxA 83.3
M00100 CdxA 83.3
M00100 CdxA 83.3
M00126 GATA-1 83.1
M00260 HLF 83.1
M00101 CdxA 82.9
M00101 CdxA 82.9
M00096 Pbx-1 82.4
M00128 GATA-1 82.1
M00116 C/EBPa 81.9
M00101 CdxA 81.4
M00080 Evi-1 81.4
M00101 CdxA 81.4
M00096 Pbx-1 81.4
M00228 VBP 81.4
M00161 Oct-1 81.1
M00160 SRY 81.1
M00026 RSRFC4 80.9
M00100 CdxA 80.8
M00101 CdxA 80.7
M00076 GATA-2 80.6
M00099 S8 80.4
M00137 Oct-1 80.4
M00137 Oct-1 80.4
M00079 Evi-1 80.3
M00130 HFH-2 80.2
M00252 TATA 80.1
M00099 S8 80.1
M00082 Evi-1 80.1
M00267 XFD-1 80.0
M00101 CdxA 80.0
501 AAAAATTATT TCCTAAAGGC AGAGAGAGAT AGACAAAGAC AGATTTCCCA entry score
M00101 CdxA 98.6
M00130 HFH-2 91.6
M00131 HNF-3b 90.2
M00129 HFH-1 89.2
M00087 Ik-2 88.6
M00101 CdxA 88.6
M00101 CdxA 88.6
M00075 GATA-1 88.6
M00077 GATA-3 86.6
M00076 GATA-2 86.2
M00100 CdxA 85.9
M00137 Oct-1 85.5
M00267 XFD-1 85.0
M00076 GATA-2 85.0
M00141 Lyf-1 84.4
M00042 Sox-5 84.3
M00096 Pbx-1 84.3
M00101 CdxA 84.3
M00109 C/EBPb 83.6
M00100 CdxA 83.3
M00131 HNF-3b 83.2
M00227 v-Myb 82.9
M00101 CdxA 82.9
M00101 CdxA 82.9
M00148 SRY 82.7

```
     <-------------                                    M00074 c-Ets- 82.6
                          ---------->                  M00075 GATA-1 82.4
-                                                     M00101 CdxA   81.4
      <------------                                   M00087 Ik-2   81.1
-->                                                   M00160 SRY    81.1
----                                                  M00100 CdxA   80.8
                            ----------->              M00203 GATA-X 80.6
                                                <---  M00032 c-Ets- 80.4
--------                                              M00130 HFH-2  80.2
------------>                                         M00099 S8     80.1
                                  ------>             M00148 SRY    80.0

551 TCCTGAGGTT TACTCCCCAA ATGGCCACAA CAGCTGAACC AGGAGCCTGA entry       score
   ----->                                             M00271 AML-1a 88.7
              <--------                               M00083 MZF1   88.7
---                                                   M00087 Ik-2   88.6
-----                                                 M00075 GATA-1 88.6
             <------------------                      M00059 YY1    86.5
-----                                                 M00076 GATA-2 86.2
      <--------------                                 M00033 p300   85.1
                         <-----                       M00271 AML-1a 83.7
                             <------------            M00001 MyoD   83.7
----                                                  M00109 C/EBPb 83.6
             <----------                              M00141 Lyf-1  83.1
                                                <---  M00100 CdxA   82.1
                            <------                   M00271 AML-1a 81.4
           <--------------                            M00159 C/EBP  80.8
               <-------                               M00100 CdxA   80.8
            <---------                                M00141 Lyf-1  80.5
------                                                M00032 c-Ets- 80.4

601 AATTCTATTC AGGTCTCCCA CATGGAGTCC TTTAGAATAA ATAAACCTAT entry       score
------>                                               M00101 CdxA   92.9
                                      <-------        M00101 CdxA   90.0
                                                <-    M00267 XFD-1  90.0
                                      <-------        M00100 CdxA   89.7
               <----------                            M00141 Lyf-1  88.3
                                                 -    M00131 HNF-3b 87.9
                                       ------->       M00101 CdxA   87.1
                                                 -    M00130 HFH-2  87.0
                   --------->                         M00217 USF    86.4
                                       ---------->    M00096 Pbx-1  86.3
                                     <------------    M00130 HFH-2  85.9
                                   <------            M00101 CdxA   85.7
                                     <------------    M00131 HNF-3b 85.5
                                         ------->     M00148 SRY    84.5
             <-------------                           M00087 Ik-2   83.8
                          <----------                 M00227 v-Myb  83.4
                 ------------>                        M00055 N-Myc  83.0
                                                 -    M00101 CdxA   82.9
                                                 <    M00101 CdxA   82.9
                <------                               M00271 AML-1a 82.7
                                                 -    M00129 HFH-1  82.3
                 <--------                            M00217 USF    82.1
---                                                   M00100 CdxA   82.1
                                 ------->             M00100 CdxA   82.1
                                           <------------ M00130 HFH-2  80.9
                                            <-------    M00100 CdxA   80.8
                                                 -    M00100 CdxA   80.8
                                               ------ M00131 HNF-3b 80.3
               <------------------                    M00059 YY1    80.2
                                              <---    M00082 Evi-1  80.1
                                  -------------->     M00267 XFD-1  80.0

651 TTTATTTATT TGAAAGGCAG AGTTACAGAG TGAGAGTGAC AGAGATGGAG entry       score
-------------                                         M00267 XFD-1  90.0
   ------>                                            M00101 CdxA   90.0
   ------>                                            M00100 CdxA   89.7
------------>                                         M00131 HNF-3b 87.9
                                    ------------>     M00173 AP-1   87.6
                                            --------> M00077 GATA-3 87.2
```

```
  <------                                              M00101 CdxA   87.1
----------->                                           M00130 HFH-2  87.0
<---------                                             M00096 Pbx-1  86.3
                                                ---------> M00075 GATA-1 85.7
<------                                                M00148 SRY    84.5
    <-------                                           M00148 SRY    84.5
----->                                                 M00101 CdxA   82.9
------                                                 M00101 CdxA   82.9
----------->                                           M00129 HFH-1  82.3
                                                ---------> M00076 GATA-2 82.2
----->                                                 M00100 CdxA   80.8
                      ---------->                      M00228 VBP    80.5
----->                                                 M00131 HNF-3b 80.3
-------                                                M00082 Evi-1  80.1

701 AGAGATTGCC CATCCTCAGG TCCACTCACC AAAAGCCTGT AACAACCAGG entry       score
       <----------                                     M00075 GATA-1 89.4
       <----------                                     M00076 GATA-2 89.3
  -------------->                                     M00116 C/EBPa 84.9
     <--------------                                   M00084 MZF1   84.7
  <--------------                                      M00109 C/EBPb 84.3
                        <--------------                M00159 C/EBP  83.8
                                                 --- M00008 Sp1    83.6
---------->                                            M00075 GATA-1 82.9
                                             ------->  M00148 SRY    82.7
                             <------                   M00271 AML-1a 82.7
---------->                                            M00076 GATA-2 82.6
                          ----------->                 M00073 deltaE 82.6
                         ----------->                  M00072 CP2    81.2
 -------->                                             M00077 GATA-3 80.9
  --------------->                                     M00117 C/EBPb 80.0

751 GCTGGGCCAG GCCAAAGTCA GGAGCTGGAA TCTCAATCCA CGTTGCCCGC entry       score
------>                                                M00008 Sp1    83.6
                                <----------            M00075 GATA-1 82.4
     ----------->                                      M00072 CP2    81.2

801 ATGGGTGTCA GGGATCGATG TACTTTAGTC AACACATATT ACCTCCCAGG entry       score
            <---------                                 M00106 CDP CR 99.5
            <---------                                 M00104 CDP CR 91.0
              ---------->                              M00106 CDP CR 87.2
                                             <-------- M00141 Lyf-1  84.4
                           --------------->             M00269 XFD-3  83.6
                            <--------------             M00159 C/EBP  82.3
                                  <-----               M00271 AML-1a 81.7
                         <-----------                  M00173 AP-1   81.4
                     <-------                           M00148 SRY    80.9
                          ---------->                  M00199 AP-1   80.7

851 GTGCACATTA GTGGGAAACT GATGTAGAGA GTGGATCTGG GACCCAAACC entry       score
       <-------                                        M00101 CdxA   86.4
                 ------->                              M00148 SRY    84.5
        ------->                                       M00101 CdxA   83.6
           -------->                                   M00083 MZF1   82.6
          ------------>                                M00087 Ik-2   82.5
                    ---------->                        M00075 GATA-1 80.8
                                     ------------>     M00087 Ik-2   80.3
                                 -------->             M00083 MZF1   80.0
                                               -----   M00148 SRY    80.0

901 CAGGCAAACT GATAGGGGAC ACAGCCATCC CAAGCAGTGG CTTAACCACT entry       score
          --------->                                   M00077 GATA-3 93.4
          ---------->                                  M00075 GATA-1 91.8
         --------------->                              M00126 GATA-1 91.5
          ------------->                               M00128 GATA-1 91.5
                        <----------                    M00076 GATA-2 90.9
          ---------->                                  M00076 GATA-2 90.1
                         <------------                 M00087 Ik-2   89.5
               ------->                                M00083 MZF1   87.8
                        <----------                    M00075 GATA-1 87.3
                                              <-----   M00271 AML-1a 85.4
```

```
          ------->                                          M00148 SRY    84.5
          -------------->                                   M00127 GATA-1 83.1
                                <---------                  M00141 Lyf-1  83.1
              ----------->                                  M00203 GATA-X 82.4
                                             <----------- M00159 C/EBP  80.8
     ->                                                   M00148 SRY    80.0

 951 GGGCCAAACA CCCACCCTTA TCTGAGCATA TTAACACAGC ATGTTGTACA entry        score
                  <------------                            M00128 GATA-1 95.1
                    <---------                             M00077 GATA-3 91.9
                  <--------------                          M00126 GATA-1 90.1
                 <-----------                              M00203 GATA-X 89.2
          ------->                                         M00148 SRY    89.1
                                 <-------                  M00101 CdxA   87.9
                                  ------->                 M00101 CdxA   87.9
                    <----------                            M00076 GATA-2 87.7
                  <--------------                          M00127 GATA-1 86.7
     <--------                                             M00050 E2F    86.2
                    <----------                            M00075 GATA-1 82.4
                                       <-----              M00271 AML-1a 81.7
                                             ----->        M00271 AML-1a 81.4
                         <---------------                  M00136 Oct-1  80.8
     --                                                    M00159 C/EBP  80.8
                               <-------------              M00159 C/EBP  80.8
                        <----------------                  M00081 Evi-1  80.2

1001 CCTCCTATAC AGTCAACTTT TTAAAGGTCT GTCTTAGCCC CTTGCAGGCC entry        score
                      <----------                          M00216 TATA   91.7
                                                         - M00073 deltaE 88.5
                        <-------                          M00100 CdxA   84.6
                                                      <-- M00002 E47    82.7
                                                        < M00001 MyoD   81.4
     <---------                                           M00141 Lyf-1  80.5

1051 TTCACCTGCC TCTGAAAACA CCACAGCTCA CGGTCCTGGA GGTCTCTAAC entry        score
                       <------                           M00271 AML-1a 100.0
                     ------->                            M00148 SRY    89.1
     --------->                                          M00073 deltaE 88.5
      ------->                                           M00217 USF    86.9
                                                  ------ M00227 v-Myb  86.1
     --------------                                      M00002 E47    82.7
     -------------                                       M00001 MyoD   81.4
                                ---------->              M00227 v-Myb  80.7

1101 AGCCCCAGGA AAGATCAACA CAGCAGATTA CCAGTGATTT TTGGAAAGCG entry        score
                                          -------------> M00159 C/EBP  89.2
                                         ------->        M00100 CdxA   87.2
     -->                                                 M00227 v-Myb  86.1
            ---------->                                  M00147 HSF2   84.6
                                         --------------> M00109 C/EBPb 84.3
                            <--------------              M00137 Oct-1  83.6
                                                    <-   M00148 SRY    83.6
                                         ------->        M00101 CdxA   83.6
                              <--------------------------M00250 Gfi-1  83.3
                                 ------->                M00101 CdxA   82.1
                                           --------->    M00141 Lyf-1  81.8
                                            --------->   M00141 Lyf-1  81.8
                       <------                           M00271 AML-1a 81.7
                                          ------------>  M00087 Ik-2   81.1
                               ---------->               M00076 GATA-2 81.0
                                          ---------->    M00075 GATA-1 80.8
                               ---------->               M00075 GATA-1 80.4

1151 TGTTTCTCTC CTGCAGGATG TTTACATGCC AGATATCCTA TAAACGAAGA entry        score
                                    ---------->          M00076 GATA-2 89.7
                                      <---------         M00076 GATA-2 89.7
                                                ------>  M00148 SRY    88.2
                                          <-------       M00101 CdxA   87.9
                       <-------                          M00148 SRY    87.3
                                          <-------       M00100 CdxA   87.2
                                         ---------->     M00216 TATA   86.8
```


M00032 c-Ets- 86.3
M00074 c-Ets- 85.8
M00075 GATA-1 85.3
M00222 Th1/E4 84.3
M00101 CdxA 84.3
M00075 GATA-1 84.1
M00148 SRY 83.6
M00267 XFD-1 83.3
M00076 GATA-2 83.0
M00162 Oct-1 81.6
M00252 TATA 80.9
M00101 CdxA 80.7
M00109 C/EBPb 80.5
M00129 HFH-1 80.0
1201 AATGAGGAAA CCATATAAAA GTGTGCTAGG CTGCAGCTTT GTTTTGCTTT entry score
M00148 SRY 100.0
M00101 CdxA 92.9
M00216 TATA 89.9
M00100 CdxA 87.2
M00252 TATA 86.9
M00148 SRY 86.4
M00148 SRY 85.5
M00101 CdxA 84.3
M00062 IRF-1 83.6
M00271 AML-1a 83.4
M00271 AML-1a 82.7
M00160 SRY 82.0
M00271 AML-1a 81.7
M00162 Oct-1 81.6
M00240 Nkx-2. 81.4
M00252 TATA 80.9
M00109 C/EBPb 80.5
1251 TGGTTCTCAT TCTGACTCCA ACTCCAGGCA GTCCCCATTG GCAGAGAAAT entry score
M00083 MZF1 93.9
M00148 SRY 85.5
M00033 p300 85.1
M00062 IRF-1 83.6
M00271 AML-1a 82.7
M00173 AP-1 82.5
M00101 CdxA 82.1
M00100 CdxA 82.1
M00087 Ik-2 82.0
M00137 Oct-1 81.5
M00199 AP-1 81.1
M00054 NF-kap 81.0
M00051 NF-kap 80.9
M00053 c-Rel 80.2
M00101 CdxA 80.0
1301 GCCGGCCTAG TCATAGCCTG GCACGCTGGG CAGAAGCAGG TGTGCAGGAA entry score
M00002 E47 89.4
M00071 E47 86.2
M00217 USF 85.5
M00032 c-Ets- 83.3
M00073 deltaE 82.1
M00025 Elk-1 81.7
M00122 USF 81.1
M00122 USF 81.1
M00108 NRF-2 80.7
1351 GGGCTCCAGG TGGGGATGTC CAGTGTGTGG GAGAGGAACC CAGGGAGAGG entry score
M00083 MZF1 98.3
M00073 deltaE 90.3
M00141 Lyf-1 88.3
M00076 GATA-2 87.4
M00053 c-Rel 86.8
M00051 NF-kap 85.8
M00052 NF-kap 85.4
M00054 NF-kap 85.4


M00075 GATA-1 85.3
M00217 USF 85.0
M00240 Nkx-2. 83.7
M00032 c-Ets- 83.3
M00271 AML-1a 82.7
M00025 Elk-1 81.7
M00001 MyoD 81.4
M00033 p300 81.2
M00108 NRF-2 80.7
M00087 Ik-2 80.7
M00084 MZF1 80.6
M00084 MZF1 80.6
M00217 USF 80.6
M00055 N-Myc 80.6
M00033 p300 80.2
1401 GAGAGGGAGA GGGCAGGTGG GCAGTAGAGG CACCAGCCAG TCCGGATACG entry score
M00076 GATA-2 93.3
M00075 GATA-1 91.8
M00002 E47 90.4
M00217 USF 89.4
M00073 deltaE 84.7
M00032 c-Ets- 84.3
M00077 GATA-3 82.8
M00128 GATA-1 82.1
M00122 USF 81.9
M00122 USF 81.9
M00055 N-Myc 81.6
M00001 MyoD 81.4
M00025 Elk-1 81.3
M00033 p300 81.2
M00074 c-Ets- 81.0
M00126 GATA-1 81.0
M00083 MZF1 80.9
M00084 MZF1 80.6
M00084 MZF1 80.6
M00203 GATA-X 80.6
M00033 p300 80.2
1451 GAGCAGGAAG CAGAGCTTCC TCTTGTGCTC ATCACCTGGA TCCTCCTCCC entry score
M00076 GATA-2 93.3
M00075 GATA-1 91.8
M00073 deltaE 90.8
M00032 c-Ets- 90.2
M00075 GATA-1 88.6
M00074 c-Ets- 84.6
M00217 USF 82.8
M00077 GATA-3 82.8
M00217 USF 82.6
M00128 GATA-1 82.1
M00122 USF 81.9
M00122 USF 81.9
M00083 MZF1 81.7
M00271 AML-1a 81.7
M00220 SREBP- 81.6
M00001 MyoD 81.4
M00025 Elk-1 81.3
M00074 c-Ets- 81.0
M00076 GATA-2 81.0
M00126 GATA-1 81.0
M00076 GATA-2 80.6
M00203 GATA-X 80.6
M00131 HNF-3b 80.3
M00074 c-Ets- 80.2
1501 TGTATGTTTG GATCATCAGT CACACCCCAA GCCAGCTGGG GGACAGCTCC entry score
M00148 SRY 90.9
M00271 AML-1a 87.4
M00083 MZF1 85.2
M00199 AP-1 82.5
M00083 MZF1 81.7

```
        ------------->                                      M00159 C/EBP  81.5
                      ----------->                          M00072 CP2    81.2
     -------->                                              M00131 HNF-3b 80.3
                   <--------------                          M00033 p300   80.2

1551 AAGAGGGCTC CCCCAACTAA AGCCCCACCT CAACGGCCCT TTTTGCCTCT entry       score
                               --------->                   M00227 v-Myb  91.4
                            ------------>                   M00073 deltaE 90.0
                                <------                     M00271 AML-1a 88.7
                                <--------------------       M00004 c-Myb  87.1
            <--------                                       M00083 MZF1   85.2
                         <-------                           M00083 MZF1   83.5
                                              ------->      M00100 CdxA   83.3
                                            ------------>   M00131 HNF-3b 82.1
                                                  <---- M00127 GATA-1 81.9
                         <-----------                       M00008 Sp1    80.8
       <-----------                                         M00051 NF-kap 80.2

1601 TGTCCACCTT CCTTGGAGGG TGCGAAGGGA TAGAAAAGTA TCTGAATAAT entry       score
                                                  ------ M00101 CdxA   98.6
                                                     <-- M00101 CdxA   90.0
                                                     <-- M00100 CdxA   89.7
                                        <-----------      M00076 GATA-2 87.4
                                     <------              M00101 CdxA   87.1
                                  ----------->            M00076 GATA-2 84.6
                                                  ------ M00096 Pbx-1  84.3
                                   --------->             M00077 GATA-3 83.8
                                                --------- M00137 Oct-1  83.6
                                               <--------- M00199 AP-1   83.5
      ------------>                                       M00073 deltaE 82.9
                                        <-----------      M00075 GATA-1 82.4
     ---------                                            M00127 GATA-1 81.9
                                                --------- M00042 Sox-5  81.7
                                                  <-----  M00101 CdxA   81.4
                <------------                             M00072 CP2    81.2
                                  ----------->            M00075 GATA-1 80.4
                                                     <-   M00074 c-Ets- 80.2

1651 AAATCCTGAA TTTGGAGATG GGGCTGTTCT CCCTGTCTTC TTCTGTATTT entry       score
     >                                                    M00101 CdxA   98.6
                  ----------->                            M00075 GATA-1 96.7
                   --------->                             M00077 GATA-3 92.2
                  ----------->                            M00076 GATA-2 90.9
     ----                                                 M00101 CdxA   90.0
     ----                                                 M00100 CdxA   89.7
             --------------->                             M00159 C/EBP  84.6
     -->                                                  M00096 Pbx-1  84.3
     ---->                                                M00137 Oct-1  83.6
                                                     <-   M00148 SRY    82.7
     ->                                                   M00042 Sox-5  81.7
     -                                                    M00101 CdxA   81.4
                                                  <-----  M00011 Evi-1  80.4
     -------------                                        M00074 c-Ets- 80.2
      ------>                                             M00101 CdxA   80.0

1701 TGTTGCTGTA ACAGAAAGCT AGAAACTGGG TGATTTATAA AGAAGATGGT entry       score
                                     ------>              M00101 CdxA   100.0
                                     ------>              M00100 CdxA   96.2
                                        <-------          M00100 CdxA   92.3
                                               ---------- M00075 GATA-1 89.8
                                                       <  M00100 CdxA   88.5
                                        <-------          M00101 CdxA   87.9
                                               ---------- M00076 GATA-2 87.0
                  <------                                M00100 CdxA   85.9
                                                       <  M00101 CdxA   84.3
                                  ---------->            M00075 GATA-1 83.7
           --------->                                    M00227 v-Myb  82.9
                             <---------------            M00252 TATA   82.7
     -----                                               M00148 SRY    82.7
                                <---------              M00096 Pbx-1  82.4
                                     ---------------->  M00252 TATA   82.2
```

```
  <----------                                          M00228 VBP    82.2
                                  ------------->       M00137 Oct-1  82.2
                      <-------                         M00101 CdxA   82.1
               <------                                 M00101 CdxA   81.4
                                        ------->       M00148 SRY    80.9
  --------->                                           M00223 STATx  80.8
                                  <----------          M00216 TATA   80.6
---                                                    M00011 Evi-1  80.4
                             <-------                  M00148 SRY    80.0

1751 TTAAATCGGC TCGTGGTCCT AGAGACTGGG AGATTTGAGG TTAAGGGGCA entry       score
                                                   <- M00032 c-Ets- 90.2
>                                                      M00075 GATA-1 89.8
                                     ------>           M00271 AML-1a 88.7
------                                                 M00100 CdxA   88.5
                                                   <- M00108 NRF-2  87.7
>                                                      M00076 GATA-2 87.0
                                                  <-- M00025 Elk-1  87.0
                                                  <-- M00074 c-Ets- 87.0
                                     -------->         M00041 CRE-BP 86.2
                                                 <--- M00054 NF-kap 86.0
                                                ----- M00053 c-Rel  86.0
                                                <---- M00007 Elk-1  85.4
                                                 ---- M00054 NF-kap 84.7
------                                                 M00101 CdxA   84.3
                                               <----- M00033 p300   84.2
           ----->                                      M00271 AML-1a 84.1
        <--------                                      M00217 USF    83.5
                                  ------>              M00100 CdxA   83.3
                                                ----- M00054 NF-kap 83.2
                             ---------->               M00075 GATA-1 82.9
                          --------->                   M00141 Lyf-1  81.8
                                                ----- M00052 NF-kap 81.5
                                           ------>     M00240 Nkx-2. 81.4
                                              <- M00240 Nkx-2. 81.4
                                                ----- M00208 NF-kap 80.7
                                --------->             M00077 GATA-3 80.3
      <------------                                    M00055 N-Myc  80.2
                                 <-----------          M00073 deltaE 80.0

1801 CTTCCCGTGC AGGTTTCTTG CCTGTGGGGA CTTTCTGCAG AGTCCCAAGG entry       score
                           ------->                    M00083 MZF1   95.7
--------                                               M00032 c-Ets- 90.2
--------                                               M00108 NRF-2  87.7
------------                                           M00025 Elk-1  87.0
----------                                             M00074 c-Ets- 87.0
  ------->                                             M00217 USF    86.3
                             ---------->               M00054 NF-kap 86.3
------                                                 M00054 NF-kap 86.0
---->                                                  M00053 c-Rel  86.0
                                  ------>              M00100 CdxA   85.9
------------                                           M00007 Elk-1  85.4
                            ------------>              M00208 NF-kap 85.1
----->                                                 M00054 NF-kap 84.7
--------                                               M00033 p300   84.2
                             ---------->               M00052 NF-kap 83.4
---->                                                  M00054 NF-kap 83.2
                           ----->                      M00271 AML-1a 82.7
                               ---------->             M00053 c-Rel  82.6
                                         <------------ M00208 NF-kap 82.5
                               <----------             M00146 HSF1   82.2
                                        <------------- M00087 Ik-2   82.0
                              ---------->              M00053 c-Rel  81.8
---->                                                  M00052 NF-kap 81.5
                                  ------>              M00101 CdxA   81.4
-----                                                  M00240 Nkx-2. 81.4
------>                                                M00208 NF-kap 80.7
                                     ----------->      M00037 NF-E2  80.2

1851 CAGTGCAGGG TAAGGGAAGG ATACATCCAA CAGAGTCAAG AGGACTTTTA entry       score
                                                  ----- M00101 CdxA   92.9
```

```
                                                  <-----  M00216 TATA   89.0
                                              <-----------  M00252 TATA   87.8
                                                   -----    M00100 CdxA   87.2
                                          ------->          M00240 Nkx-2. 83.7
                                    ---------->             M00227 v-Myb  83.4
                              ---------->                   M00076 GATA-2 82.2
                                ---------->                 M00203 GATA-X 81.8
                                        <-----------        M00173 AP-1   80.4
                        ------------->                      M00087 Ik-2   80.3

1901 TATTAGACCC GCTCTCTAGA TAATTCACTA AACCATTAAT CCAATAATGA entry        score
                                       <------              M00101 CdxA   99.3
                                                 ------>    M00101 CdxA   97.9
     ->                                                     M00101 CdxA   92.9
                                                ----------  M00137 Oct-1  91.3
                          <------------------               M00099 S8     90.0
     ----                                                   M00216 TATA   89.0
                                       <------              M00100 CdxA   88.5
                                        ------->            M00101 CdxA   87.9
     ---                                                    M00252 TATA   87.8
     ->                                                     M00100 CdxA   87.2
                           <----------                      M00147 HSF2   87.2
                                                 -----      M00101 CdxA   87.1
                            <-------                        M00101 CdxA   86.4
                            ----------->                    M00203 GATA-X 86.4
                                          <------------     M00131 HNF-3b 85.5
                                               <------     M00101 CdxA   83.6
                            -------->                       M00241 Nkx-2. 82.4
                                                <-------    M00241 Nkx-2. 82.4
                                    <-------                M00240 Nkx-2. 81.4
                                          ---------->       M00096 Pbx-1  81.4
                                ---------->                 M00075 GATA-1 80.8
                                          ------->          M00100 CdxA   80.8
                                                  ------    M00109 C/EBPb 80.5
     <------                                                M00101 CdxA   80.0

1951 TGAAAGTCAG AGTCCTGAGA ACTCATCAAC TATTAAAGTC CCTCTCCATG entry        score
                                     <------                M00101 CdxA   92.9
     -->                                                    M00137 Oct-1  91.3
     ->                                                     M00101 CdxA   87.1
                                 ---------->                M00216 TATA   85.9
                           <----------                      M00075 GATA-1 84.1
                                      ------>               M00101 CdxA   83.6
                                                <-----      M00008 Sp1    82.2
                                  ----------------->        M00252 TATA   81.5
                                     ------>                M00100 CdxA   80.8
     ------->                                               M00109 C/EBPb 80.5

2001 CCTCTACAAT GGGGATTACA TTTCAATGTG AGAGCTGCAG ACATTTGGCC entry        score
            -------->                                       M00083 MZF1   93.9
                   <--------------                          M00162 Oct-1  91.8
               <--------------                              M00249 CHOP-C 89.8
            ------------>                                   M00087 Ik-2   86.4
     ----                                                   M00008 Sp1    82.2
              ----------------->                            M00133 Tst-1  81.2
       ---------->                                          M00042 Sox-5  81.0
                ---------->                                 M00076 GATA-2 80.6

2051 TAGGAGCTAA GATGCCAGTT CCATTACCCT GTCCCATATG GGATTGCCTG entry        score
                                          ------------>     M00087 Ik-2   89.5
                                             ---------->    M00076 GATA-2 87.4
                                             ---------->    M00053 c-Rel  84.3
                                          ------------->    M00086 Ik-1   83.9
                                             ---------->    M00052 NF-kap 83.4
                                             ---------->    M00075 GATA-1 82.9
                        <-------------                      M00137 Oct-1  81.5
            ---------->                                     M00076 GATA-2 81.0
                                <----------                 M00008 Sp1    80.8

2101 GATTCAATTC TCATCTCCAG TTACTGATTC CTGCCAGTGC AGGCTCTGGG entry        score
                                                  ------    M00141 Lyf-1  93.5
```

```
          <----------                                    M00075 GATA-1 84.5
                  <----------                            M00003 v-Myb  82.8
                           ---------->                   M00075 GATA-1 81.6
                       ------->                          M00101 CdxA   81.4
                              ------->                   M00101 CdxA   81.4
          <---------                                     M00077 GATA-3 81.2
        ------->                                         M00101 CdxA   80.7
                            <----------                  M00032 c-Ets- 80.4
        <-----------                                     M00203 GATA-X 80.2

2151 AGGCAGCAGG GACGGCTCAA GAAGTTGAGT CTCTGCCACC CAATAAGGGG entry       score
     -->                                                 M00141 Lyf-1  93.5
                                                  ------ M00083 MZF1   90.4
                                               ------>   M00101 CdxA   82.9

2201 GACCTGGATG GAGTTCCCAG CTCCCAGCCC CACCAACAGC CCAGACGTGG entry       score
     ->                                                  M00083 MZF1   90.4
                          <----------                    M00008 Sp1    90.4
             <------------                               M00087 Ik-2   86.8
        ---------->                                      M00075 GATA-1 86.1
                                  <--------------------  M00004 c-Myb  85.2
           <--------------                               M00088 Ik-3   83.8
                               <--------                 M00083 MZF1   83.5
           ---------->                                   M00052 NF-kap 83.4
        ---------->                                      M00076 GATA-2 81.8
                                               <------- M00217 USF    81.5
                                                      - M00032 c-Ets- 81.4
           <--------------                               M00086 Ik-1   81.3
           ---------->                                   M00054 NF-kap 80.7
                                     -------->           M00227 v-Myb  80.2

2251 CAGGAATTCT AGGTATGAGC TAGTGAATGG GAGCTCTCGT TCTCTATCAT entry       score
         ------->                                        M00101 CdxA   92.9
                                               <-------- M00075 GATA-1 90.6
                                               <-------- M00076 GATA-2 89.7
                                               <-------- M00077 GATA-3 89.1
                                             <---------- M00128 GATA-1 83.6
                                            ----------- M00085 ZID    82.2
                                                     <- M00077 GATA-3 81.9
                                 ------------>          M00087 Ik-2   81.6
     <------                                             M00101 CdxA   81.4
     -------->                                           M00032 c-Ets- 81.4
                                                  <--- M00127 GATA-1 80.7
                                                  <--- M00126 GATA-1 80.3
                                             <----------  M00203 GATA-X 80.0

2301 CTATCTATCT ATCTATCTAT CTATCTATCT ATCTACCTAC CTACCTACCT entry       score
                                                    <--- M00077 GATA-3 91.6
     -                                                   M00075 GATA-1 90.6
     -                                                   M00076 GATA-2 89.7
                                                  <--- M00076 GATA-2 89.3
                                                  <--- M00075 GATA-1 89.0
                                                <-----  M00128 GATA-1 86.0
                                                <-----  M00126 GATA-1 85.9
                                                   ---  M00145 Brn-2  84.2
                                 <----------            M00075 GATA-1 84.1
                                 <----------            M00076 GATA-2 83.8
     --                                                  M00128 GATA-1 83.6
     >                                                   M00085 ZID    82.2
                               <--------------           M00127 GATA-1 81.9
     -------                                             M00077 GATA-3 81.9
       <----------                                       M00077 GATA-3 81.9
           <----------                                   M00077 GATA-3 81.9
                <---------                               M00077 GATA-3 81.9
                    <----------                          M00077 GATA-3 81.9
                         <----------                     M00077 GATA-3 81.9
                              <----------                M00077 GATA-3 81.9
                                 <----------             M00077 GATA-3 81.9
                               <--------------           M00128 GATA-1 81.2
                                               <------  M00203 GATA-X 80.9
     ----------                                          M00127 GATA-1 80.7
```

```
     <-------------                                          M00128 GATA-1 80.5
         <-------------                                     M00128 GATA-1 80.5
             <---------------                               M00128 GATA-1 80.5
                  <-------------                            M00128 GATA-1 80.5
                      <-------------                        M00128 GATA-1 80.5
                           <--------------                  M00128 GATA-1 80.5
     ----------                                             M00126 GATA-1 80.3

2351 ATCTCTAATA AACAATTTTA ACATGAGATT TGGTGGGGAC ATTCAAACCA entry        score
                                    ------->                M00083 MZF1   98.3
             ---------->                                    M00042 Sox-5  94.1
     -----                                                  M00077 GATA-3 91.6
              ------->                                      M00148 SRY    90.0
     ------                                                 M00076 GATA-2 89.3
          <------------                                     M00130 HFH-2  89.3
          <------------                                     M00129 HFH-1  89.2
     ------                                                 M00075 GATA-1 89.0
                               ---------->                  M00077 GATA-3 86.2
     -------                                                M00128 GATA-1 86.0
     --------                                               M00126 GATA-1 85.9
                       <----------------                    M00162 Oct-1  85.7
            ------------>                                   M00160 SRY    85.4
     -------------->                                        M00145 Brn-2  84.2
          <------------                                     M00131 HNF-3b 83.8
                                                  <---      M00271 AML-1a 83.4
                    ------>                                 M00101 CdxA   82.9
                                 ------>                    M00271 AML-1a 82.7
                               ---------->                  M00075 GATA-1 82.0
                                                 -------    M00042 Sox-5  81.7
         -------------->                                    M00267 XFD-1  81.7
                                 <-----------               M00072 CP2    81.2
                                        ------------>       M00208 NF-kap 81.2
     ----                                                   M00203 GATA-X 80.9
           <-------                                         M00100 CdxA   80.8
                       ------->                             M00101 CdxA   80.7
                             ---------->                    M00076 GATA-2 80.6
        ------>                                             M00101 CdxA   80.0
                 <---------------                           M00267 XFD-1  80.0

2401 TAGCGGTCCC CCCAAGCAAT CTCTTTCCCT TAATTTCTTC CAGCACTTAC entry        score
            <--------                                       M00083 MZF1   92.2
                                 -------->                  M00241 Nkx-2. 91.2
                               --------->                   M00223 STATx  90.4
                                                 <------    M00240 Nkx-2. 88.4
                     <----------                            M00075 GATA-1 84.5
                                   <-------                 M00101 CdxA   84.3
                                         ------>            M00101 CdxA   84.3
                                  <------------------       M00099 S8     84.1
     --                                                     M00271 AML-1a 83.4
                     <----------                            M00076 GATA-2 83.0
            <------------                                   M00244 NGFI-C 82.8
                                   <-------                 M00101 CdxA   82.1
     -->                                                    M00042 Sox-5  81.7
                                  <-------                  M00240 Nkx-2. 81.4
                                          <----------       M00032 c-Ets- 81.4
                     <---------                             M00077 GATA-3 80.9
            <------------                                   M00246 Egr-2  80.8
            <------------                                   M00243 Egr-1  80.7
                         <-------                           M00148 SRY    80.0

2451 AGCCTGATAA TGACAACTGC ACACTGTATT ATTTCTTACC CTGCCTTATT entry        score
                                  <-------                  M00101 CdxA   98.6
          ------------->                                    M00137 Oct-1  93.5
        ---------->                                         M00075 GATA-1 89.8
       ------------->                                       M00128 GATA-1 89.4
     --------------->                                       M00127 GATA-1 89.2
         ----------->                                       M00203 GATA-X 87.5
       ---------->                                          M00076 GATA-2 87.0
                                            <----------     M00008 Sp1    86.3
          ------->                                          M00101 CdxA   85.7
         --------->                                         M00077 GATA-3 85.3
```


M00101 CdxA 84.3
M00101 CdxA 82.9
M00042 Sox-5 81.7
M00101 CdxA 81.4
M00199 AP-1 80.9
M00081 Evi-1 80.2
M00267 XFD-1 80.0
2501 CACTAGTGCT GCTTGGTTAA CTTGGCCTCT CCAACTCCAG CCCAACTTCC entry score
M00032 c-Ets- 88.2
M00074 c-Ets- 88.1
M00209 NF-Y 84.7
M00159 C/EBP 83.1
M00033 p300 82.2
M00025 Elk-1 82.2
M00199 AP-1 80.9
M00117 C/EBPb 80.8
M00223 STATx 80.8
M00033 p300 80.2
M00267 XFD-1 80.0
M00159 C/EBP 80.0
2551 TGGAGGACAG AGAATAGCAT AGTTTTTGGT ACCATTTCTC CCCACCACCC entry score
M00083 MZF1 98.3
M00032 c-Ets- 88.2
M00074 c-Ets- 88.1
M00148 SRY 84.5
M00271 AML-1a 84.1
M00083 MZF1 83.5
M00221 SREBP- 83.1
M00271 AML-1a 82.7
M00025 Elk-1 82.2
M00101 CdxA 82.1
M00100 CdxA 82.1
M00257 RREB-1 81.0
M00008 Sp1 80.8
M00223 STATx 80.8
M00033 p300 80.2
M00101 CdxA 80.0
M00159 C/EBP 80.0
2601 CCACCACCCC GACCCATTAA CCTCAAGTTG AGCTCCATGA GTTGATGAAA entry score
M00271 AML-1a 88.7
M00101 CdxA 87.1
M00041 CRE-BP 86.2
M00240 Nkx-2. 86.0
M00008 Sp1 84.9
M00271 AML-1a 84.1
M00083 MZF1 83.5
M00073 deltaE 81.5
M00257 RREB-1 81.0
M00008 Sp1 80.8
M00109 C/EBPb 80.5
2651 ATGAACTAGG AGCAGTCACA CACATGATGA CAGGCTGGCC CAGATATGTA entry score
M00075 GATA-1 88.6
M00076 GATA-2 88.1
M00203 GATA-X 85.7
M00217 USF 84.8
M00240 Nkx-2. 83.7
M00128 GATA-1 83.0
M00122 USF 82.5
M00122 USF 82.5
M00075 GATA-1 82.4
M00045 E4BP4 82.4
M00217 USF 82.1
M00123 c-Myc/ 80.9
M00101 CdxA 80.7
M00101 CdxA 80.7
M00077 GATA-3 80.6
M00157 RORalp 80.5

```
     -->                                                  M00109 C/EBPb 80.5

2701 AAATACTTGC CCAAAAGAAC CACAGTAAAA GAGCCTGGCA TGAGTAGGTT entry       score
                     <------                              M00271 AML-1a 100.0
                                                       <- M00148 SRY    90.0
                           <------                        M00101 CdxA   86.4
                ------>                                   M00148 SRY    86.4
     ->                                                   M00203 GATA-X 85.7
     >                                                    M00128 GATA-1 83.0
      ------>                                             M00101 CdxA   82.9
        -------------->                                   M00117 C/EBPb 82.4
     -->                                                  M00045 E4BP4  82.4
        -------------->                                   M00116 C/EBPa 81.1
     ---                                                  M00101 CdxA   80.7
           <---------                                     M00141 Lyf-1  80.5
                           ---------->                    M00216 TATA   80.2

2751 TCTTTTGACA GACAGTAAAA TTCAGCCTCA AGCGGAAGCT GCTGCTAGAG entry       score
     -----                                                M00148 SRY    90.0
                                 -------------->          M00074 c-Ets- 87.0
                <------                                   M00101 CdxA   86.4
                                 --------------->         M00025 Elk-1  85.2
                                     --------->           M00108 NRF-2  84.2
                                 ------->                 M00240 Nkx-2. 83.7
                <----------------                         M00133 Tst-1  83.3
                                       ---------->        M00147 HSF2   83.3
                   <-------                               M00101 CdxA   82.9
                                     --------->           M00032 c-Ets- 81.4

2801 TGTGTGGGCT GGGGAGGTGT CGCCAGTTGC AGCCTGGCTT CCTCTGTGGC entry       score
            -------->                                     M00083 MZF1   90.4
                                                  -----> M00271 AML-1a 83.7
                                                -------- M00159 C/EBP  83.1
       ----->                                             M00271 AML-1a 82.7
                                       <-------------     M00074 c-Ets- 81.8
              ------------>                               M00001 MyoD   81.4
          ---------->                                     M00008 Sp1    80.8
                 --------->                               M00008 Sp1    80.8
               ----------------->                         M00002 E47    80.8

2851 CAGTCTTAAG AGAGGCATTC AAAATAGGTG GAGTGCAAGG TTTGTTCCCT entry       score
                   ---------------->                      M00145 Brn-2  91.1
                              --------------->            M00033 p300   84.2
     ---->                                                M00159 C/EBP  83.1
                                           <------        M00148 SRY    82.7
                            <------------                 M00073 deltaE 81.7
          ------->                                        M00240 Nkx-2. 81.4
                                                  -----   M00223 STATx  80.8
                                        ------------>     M00130 HFH-2  80.2

2901 GGAACAGGGT AGGGGATGTC CCCTGTGGCC AAATGAAGTT GAATCAGTGT entry       score
                  <----------                             M00054 NF-kap 90.7
                                              --------->  M00199 AP-1   89.9
                 ---------->                              M00076 GATA-2 87.4
                        <--------                         M00083 MZF1   87.0
                 ---------->                              M00053 c-Rel  86.8
                  ---------->                             M00054 NF-kap 86.6
               -------->                                  M00083 MZF1   86.1
                 ---------->                              M00051 NF-kap 85.8
                 ---------->                              M00052 NF-kap 85.4
                 ---------->                              M00054 NF-kap 85.4
                 ---------->                              M00075 GATA-1 85.3
                              ----->                      M00271 AML-1a 83.7
                                           <---------     M00199 AP-1   83.5
                 <-------------                           M00208 NF-kap 83.0
     --->                                                 M00223 STATx  80.8
                                        ------------------ M00250 Gfi-1  80.5
                                            ----------->  M00173 AP-1   80.4
                   <----------                            M00053 c-Rel  80.2
                              <------------------         M00059 YY1    80.2
```

```
2951 CTGAAAGGAA ACCCAGATTC TTCCCTCCAG GTGAGGACTA CTCCAGGTGC entry        score
                              <-----------                 M00073 deltaE 93.6
                  <----------                              M00147 HSF2   91.0
                 ---------->                               M00075 GATA-1 86.1
                                                 -------> M00217 USF    85.0
                                               <--------  M00073 deltaE 84.9
                  ---------->                              M00147 HSF2   84.0
                                    -------->              M00083 MZF1   83.5
                                             ------------- M00071 E47    83.0
                                 -------->                 M00217 USF    82.8
        <----------                                        M00053 c-Rel  82.6
                                 <--------                 M00217 USF    82.6
                                           ------------    M00002 E47    81.7
    ------>                                                M00250 Gfi-1  80.5
                ---------->                                M00076 GATA-2 80.2
          ------->                                         M00148 SRY    80.0

3001 TCCCATCACA GACAGCCAGC AGCCACTCAG GCACAACAGG ACATCCAGGC entry        score
    <---------                                             M00075 GATA-1 88.6
    --                                                     M00073 deltaE 84.9
                   <----------------                       M00209 NF-Y   84.6
                                        ---------->        M00032 c-Ets- 83.3
                                            <----------    M00032 c-Ets- 83.3
    --->                                                   M00071 E47    83.0
    <---------                                             M00076 GATA-2 83.0
    <--------                                              M00077 GATA-3 82.8
        <-----                                             M00271 AML-1a 82.1
                                     ------------->        M00074 c-Ets- 81.8
    --->                                                   M00002 E47    81.7
                                    <-----                 M00271 AML-1a 81.4
                                          <----------      M00076 GATA-2 80.2
                                          <----------      M00075 GATA-1 80.0

3051 GACTAGAGGC AGCAGGGGGC TGGCCCCACG TTCCCTCTTC AGTACAGGTC entry        score
        ------------------->                              M00005 AP-4   85.0
                             -------->                     M00217 USF    83.5
                           ------------>                   M00055 N-Myc  82.7
                  ---------->                              M00008 Sp1    82.2
                                                ---------- M00156 RORalp 81.5
                           <-------                        M00083 MZF1   80.9
                         <-----------------                M00236 Arnt   80.0

3101 AACGCTCCGG GACACCTGAA GACACCAGGG ACTGCCAAAG GCCATGGCAG entry        score
             ----------->                                  M00073 deltaE 87.7
                  ---------->                              M00108 NRF-2  84.2
                <------                                    M00240 Nkx-2. 83.7
           <-----------------                              M00002 E47    83.7
                ------->                                   M00217 USF    83.3
       ------------>                                       M00087 Ik-2   81.6
    ->                                                     M00156 RORalp 81.5
            <-------------                                 M00001 MyoD   81.4
               <-------                                    M00217 USF    80.6
                                    <--------              M00141 Lyf-1  80.5
            ---------------->                              M00122 USF    80.3
            <----------------                              M00122 USF    80.3
                                       --------->          M00227 v-Myb  80.2

3151 CAGCAGCAGC AGCAGCTAAC ACTGATCCCT GGCAAAAGCT GGAAACTGTA entry        score
                       ---------->                         M00106 CDP CR 84.5
                     ---------->                           M00042 Sox-5  80.4

3201 CAAGGCCAGA GAACGTGACA ATAGAAACTC TTTTGTCCAG TTTGCTGGTA entry        score
           <-----------------                              M00236 Arnt   83.0
                    ---------->                            M00042 Sox-5  82.4
                           <------                         M00101 CdxA   82.1
                 ------------>                             M00160 SRY    82.0
              <-------                                     M00217 USF    81.1
             ------------>                                 M00123 c-Myc/ 80.9
            ---------------->                              M00122 USF    80.7
            <----------------                              M00122 USF    80.7
```

```
3251 AGTCTACTTC TTCACACTTG ACACCAGCAG AGATAAGCAG CCCACTTCCC entry        score
                  <-------                                 M00240 Nkx-2. 100.0
                                  ----------->            M00203 GATA-X 89.5
                                                  --      M00240 Nkx-2. 88.4
                                ---------->               M00127 GATA-1 86.7
                                   --------->             M00077 GATA-3 86.6
                                 ---------->              M00128 GATA-1 86.0
                                                <-----    M00141 Lyf-1  85.7
                                                  ----    M00055 N-Myc  84.5
                                               <-------   M00087 Ik-2   84.2
                                                     --   M00217 USF    83.0
                                  ---------->             M00076 GATA-2 82.6
                                ---------->               M00126 GATA-1 82.4
                  -------->                               M00217 USF    81.9
                                               <------    M00240 Nkx-2. 81.4
                                              --------    M00072 CP2    81.2
                                          <-----------    M00033 p300   81.2
             ---------->                                  M00122 USF    80.9
             <----------                                  M00122 USF    80.9
                <--------                                 M00217 USF    80.6
                                  ---------->             M00075 GATA-1 80.0

3301 AAGTGTAGTT GTAATTCACC AATTTTTAAA ATTGCACCAT CACTTCTCCA entry        score
                                        <----------       M00075 GATA-1 91.4
                                  ---------->             M00116 C/EBPa 90.0
                                  <----------             M00109 C/EBPb 89.9
          <----------                                     M00133 Tst-1  89.6
     ---->                                                M00240 Nkx-2. 88.4
            <-------                                      M00101 CdxA   86.4
     ---                                                 M00141 Lyf-1  85.7
                                        <----------       M00077 GATA-3 85.6
                                  ---------->             M00117 C/EBPb 84.8
     ------->                                            M00055 N-Myc  84.5
                           ------>                        M00101 CdxA   84.3
     ----                                                M00087 Ik-2   84.2
                                        <----------       M00076 GATA-2 84.2
               <----------                               M00159 C/EBP  83.8
                           ------>                       M00100 CdxA   83.3
     ----->                                              M00217 USF    83.0
                              <-------                   M00101 CdxA   82.9
             <----------                                 M00116 C/EBPa 81.9
        ----->                                           M00271 AML-1a 81.4
                                              <------    M00240 Nkx-2. 81.4
                           <----------                   M00216 TATA   81.3
     ->                                                  M00072 CP2    81.2
     --                                                  M00033 p300   81.2
          <----------                                    M00099 S8     81.1
                     <----------                         M00042 Sox-5  81.0
                         ---------->                     M00216 TATA   80.4

3351 TGAGTCCCCT CCCTCCACAC CTGCCAACAC ACTACATCCA CCCCCACCTC entry        score
         <--------                                       M00083 MZF1   93.0
                     -------->                           M00217 USF    85.5
                   ---------->                           M00073 deltaE 84.6
                   <----------                           M00001 MyoD   83.7
                  <--------                              M00083 MZF1   83.5
                                              <-------    M00083 MZF1   83.5
                  <-----                                 M00271 AML-1a 83.1
                 <----------                             M00002 E47    81.7
                                 <------                 M00271 AML-1a 81.7
                                      <-----             M00271 AML-1a 81.4
                                                ---------M00073 deltaE 80.8
                        <---------                       M00141 Lyf-1  80.5
                                           <--------     M00083 MZF1   80.0

3401 CCTATATTCC ATCACTAAAG CTTGAATGGG GCTTCACTCT GGGCACTCTG entry        score
           <----------                                   M00075 GATA-1 83.3
        <----------                                      M00109 C/EBPb 83.0
                                 ---------->             M00054 NF-kap 82.2
                <------                                  M00240 Nkx-2. 81.4
     ->                                                  M00073 deltaE 80.8
```

```
             -------->                                  M00096 Pbx-1  80.4
       -------------->                                  M00117 C/EBPb 80.0

3451 GGCACTACTT CTCTGTGCCA CACCTGGCCC AAAAGTTGGA ACAAGGCTAG entry       score
                     ---------->                        M00073 deltaE 86.1
                     <----------                        M00001 MyoD   86.0
                                                      < M00073 deltaE 85.6
                    <------                             M00271 AML-1a 83.7
                       <--------                        M00217 USF    83.0
                                                  <-- M00101 CdxA   82.1
                                                  <-- M00100 CdxA   82.1
                   <----------------                    M00002 E47    81.7
                       -------->                        M00217 USF    81.1

3501 AAAGGTGGAA AGTGAGCAGC AGCCACCTGT ACAACGACTC TCACCAGAGG entry       score
                          <-----------                  M00001 MyoD   88.4
                            ------->                    M00217 USF    87.4
     ----------                                         M00073 deltaE 85.6
                          ---------->                   M00073 deltaE 85.5
                                                <---- M00147 HSF2   84.6
                                              <-----   M00271 AML-1a 83.4
                                                   <- M00032 c-Ets- 83.3
                                           ----------> M00073 deltaE 83.1
     ----                                               M00101 CdxA   82.1
     ----                                               M00100 CdxA   82.1
           ------->                                     M00240 Nkx-2. 81.4
      -------------->                                   M00159 C/EBP  80.8
                        ---------------->               M00122 USF    80.1
                        <----------------               M00122 USF    80.1

3551 TTTCCAGTAA ACAGTGAACT GTCAGTTTCA TAAAATTGTG TATTTCGTTA entry       score
                                 <-------               M00101 CdxA   92.1
                                 <-------               M00100 CdxA   91.0
      -------->                                         M00223 STATx  89.4
         <-------                                       M00101 CdxA   86.4
                                       ------------>    M00129 HFH-1  86.2
         <------------                                  M00131 HNF-3b 86.1
                                       ------------>    M00130 HFH-2  85.1
     -----                                              M00147 HSF2   84.6
      <---------                                        M00223 STATx  84.6
                           <------                      M00148 SRY    84.5
                                <--------------         M00162 Oct-1  83.7
     --------                                           M00032 c-Ets- 83.3
                                    <------             M00101 CdxA   82.9
                                <--------------         M00109 C/EBPb 82.4
                                       ------>          M00101 CdxA   82.1
                                              <-------- M00260 HLF    81.6
                                 ---------->            M00228 VBP    81.4
                                                  ---- M00228 VBP    81.4
                                             <---------- M00109 C/EBPb 81.1
                                                <-------- M00137 Oct-1  81.1
           ---------->                                  M00042 Sox-5  81.0
                                 ------->               M00101 CdxA   80.7
                                       <----------      M00042 Sox-5  80.4
                                  ---------->           M00216 TATA   80.2
            ------->                                    M00148 SRY    80.0

3601 CCTCTTTCTG CAAATCTTTC TTTGGAGAAA AGATACAAAG CAGAGCTCCT entry       score
                     <-------                           M00148 SRY    90.0
                                 ---------->            M00082 Evi-1  88.1
                               -------------->          M00159 C/EBP  86.9
                                 ---------->            M00079 Evi-1  86.4
        ------>                                         M00100 CdxA   85.9
                                 ---------->            M00080 Evi-1  84.6
                  ------------>                         M00130 HFH-2  84.4
                                       <------          M00100 CdxA   83.3
     -                                                  M00260 HLF    81.6
        ------>                                         M00101 CdxA   81.4
     ----->                                             M00228 VBP    81.4
     ---                                                M00109 C/EBPb 81.1
     ----                                               M00137 Oct-1  81.1
```

```
                                -------->                M00148 SRY    80.9
                                     ------>             M00148 SRY    80.9
     ------------------->                                 M00135 Oct-1  80.8
                                                  <---    M00108 NRF-2  80.7
                                ---------->              M00011 Evi-1  80.4
                                ------------->           M00128 GATA-1 80.2

3651 TCCGATGACC CTGCTGCTTC AGTTTAGACT AGAATCTACT CTCCCCTCCA entry        score
                                                <-------  M00083 MZF1   93.0
                                   --------->            M00147 HSF2   87.2
                        <-------                          M00148 SRY    84.5
     --------->                                          M00076 GATA-2 81.8
                                     ------>             M00148 SRY    81.8
                                           <-----------  M00084 MZF1   81.5
                                   <---------            M00147 HSF2   80.8
                                 <-------                M00101 CdxA   80.7
     ------                                              M00108 NRF-2  80.7
         <-------------                                  M00156 RORalp 80.6
                                   --------->            M00146 HSF1   80.1
                                            <--------    M00083 MZF1   80.0

3701 ACTCTGAAGG ACCTGTGATG TGTGATCTCT GCACAAGCTG TCAATGCCAT entry        score
                                                 <-----  M00076 GATA-2 87.0
                                                 <-----  M00075 GATA-1 86.5
                  ---------->                            M00075 GATA-1 85.7
                  ----->                                 M00271 AML-1a 82.1
                         ----->                          M00271 AML-1a 82.1
     --                                                  M00084 MZF1   81.5
                  ---------->                            M00076 GATA-2 80.6

3751 CTTCTTGTCC CTTAAGAGTT AATGAACGGC CGGCGCTGCG GCTCACTAGG entry        score
                       ------->                          M00101 CdxA   87.1
     ----                                                M00076 GATA-2 87.0
     ----                                                M00075 GATA-1 86.5
                  ----------------->                     M00099 S8     82.1
                           -------->                     M00227 v-Myb  81.8
             <-------                                    M00240 Nkx-2. 81.4
                 ------>                                 M00240 Nkx-2. 81.4
                    ------------->                       M00137 Oct-1  80.7

3801 CTAATCCTCC GCCTTGCAGC GCCGGCACAC TGGGTTCTAG TCCTGGTCAG entry        score
                           ------------------->          M00192 GR     81.8
                             ----------------->          M00205 GR     81.5

3851 GGCACTGGAT TCTGTCCCGG TTGCCCCTCT TCCAGGCCAG CTCTCTGCTG entry        score
                           <-------------                M00084 MZF1   91.1
       <------                                           M00240 Nkx-2. 83.7
                                                  --     M00271 AML-1a 83.7
                                <----------              M00108 NRF-2  82.5
         ---------->                                     M00075 GATA-1 82.4
            ------->                                     M00101 CdxA   80.0
                             <-------                    M00083 MZF1   80.0

3901 TGGCCAGGGA GTGCAGTGGA GGATGGCCCA AGCACTTGGG AGACCAGGAT entry        score
                                                  ----   M00203 GATA-X 94.5
                        ---------->                      M00076 GATA-2 91.3
                                      --------->         M00141 Lyf-1  89.6
                                    <------              M00240 Nkx-2. 88.4
                        ---------->                      M00075 GATA-1 87.8
                                               ------    M00076 GATA-2 87.7
                                             --------    M00127 GATA-1 86.7
        --------------->                                 M00033 p300   86.1
                                     <-------            M00217 USF    85.0
                                       ------------->    M00087 Ik-2   84.6
                                               ------    M00075 GATA-1 84.5
     --->                                                M00271 AML-1a 83.7
                                   ------------->        M00123 c-Myc/ 83.2
                                   <-------------        M00055 N-Myc  80.6
                                              -------    M00128 GATA-1 80.2

3951 AAGTACCTGG CTCCTGCCAT CCTATCAGCA CGGTGCGCTG GCCGCAGCAC entry        score
```

```
     ------>                                                M00203 GATA-X 94.5
                       <----------                          M00075 GATA-1 93.1
                       <----------                          M00076 GATA-2 90.1
     --->                                                   M00076 GATA-2 87.7
                       <---------                           M00077 GATA-3 87.5
                   <-----------                             M00076 GATA-2 87.0
                      <--------------                       M00128 GATA-1 86.9
     ----->                                                 M00127 GATA-1 86.7
     --->                                                   M00075 GATA-1 84.5
                                                  <-----    M00235 AhR/Ar 83.3
                      <----------------                     M00126 GATA-1 83.1
                      <----------------                     M00127 GATA-1 80.7
                   <-----------                             M00075 GATA-1 80.4
     ----->                                                 M00128 GATA-1 80.2

4001 GCCAGCCATG GCGGCCATTG GAGGGTGAAC CAACGGCAAA GGAAGACCTT entry        score
                                 --------->                M00227 v-Myb  94.1
                                               <------     M00147 HSF2   85.9
                                  ---------->              M00003 v-Myb  85.3
                                               <------     M00146 HSF1   85.1
     ----------                                            M00235 AhR/Ar 83.3
                                        ------------->      M00074 c-Ets- 83.0
                   <---------                               M00227 v-Myb  82.9
                                 ------->                   M00148 SRY    81.8
                                               -------      M00147 HSF2   80.8
                                          ---------->       M00108 NRF-2  80.7

4051 TCTCTCTGTC TCTCTCTCTC ACTGTCCTCT CTGCCTGTCA AAAAAAAAAA entry        score
     ---                                                   M00147 HSF2   85.9
     ---                                                   M00146 HSF1   85.1
                                             ------->      M00148 SRY    84.5
                                               ------>     M00148 SRY    84.5
                                                ------>    M00148 SRY    84.5
                                                 ------>   M00148 SRY    84.5
                                                  ------>  M00148 SRY    84.5
                                                   ------> M00148 SRY    84.5
                                                    ------ M00148 SRY    84.5
                                                     ----- M00148 SRY    84.5
                                                      ---- M00148 SRY    84.5
                                                       --- M00148 SRY    84.5
                                                        -- M00148 SRY    84.5
                                                         - M00148 SRY    84.5
                                             <----------- M00130 HFH-2  82.1
                                               <--------- M00130 HFH-2  82.1
                                                <-------- M00130 HFH-2  82.1
                                                 <------- M00130 HFH-2  82.1
                                                  <------ M00130 HFH-2  82.1
                                                   <----- M00130 HFH-2  82.1
                                                    <---- M00130 HFH-2  82.1
                                                     <--- M00130 HFH-2  82.1
                                                      <-- M00130 HFH-2  82.1
                                                       <- M00130 HFH-2  82.1
                                                        < M00130 HFH-2  82.1
     -->                                                   M00147 HSF2   80.8
                            <--------                      M00083 MZF1   80.0

4101 AAAAAAAAAA AAAAAAAAGA GTTAATGAAG AAGTAACTCA CTGGCGTCAT entry        score
                                                <--------- M00113 CREB   92.6
                            ------>                       M00101 CdxA   87.1
                                                <-------  M00039 CREB   87.1
                                                ------->  M00041 CRE-BP 86.2
                                                <-------  M00041 CRE-BP 86.2
                         --------------->                 M00137 Oct-1  85.8
                                                ------->  M00039 CREB   85.6
     >                                                     M00148 SRY    84.5
     ->                                                    M00148 SRY    84.5
     -->                                                   M00148 SRY    84.5
     --->                                                  M00148 SRY    84.5
     ---->                                                 M00148 SRY    84.5
     ----->                                                M00148 SRY    84.5
     ------>                                               M00148 SRY    84.5
      ------>                                              M00148 SRY    84.5
```

```
       ------>                                                M00148 SRY    84.5
        ------>                                               M00148 SRY    84.5
         ------->                                             M00148 SRY    84.5
          ------->                                            M00148 SRY    84.5
           ------->                                           M00148 SRY    84.5
            ------->                                          M00148 SRY    84.5
             ------->                                         M00148 SRY    84.5
              ------->                                        M00148 SRY    84.5
                ------>                                       M00148 SRY    84.5
                 ------>                                      M00148 SRY    84.5
                                                   <-----     M00076 GATA-2 83.8
                                                    <----     M00087 Ik-2   82.9
     -                                                        M00130 HFH-2  82.1
     --                                                       M00130 HFH-2  82.1
     ---                                                      M00130 HFH-2  82.1
     ----                                                     M00130 HFH-2  82.1
     -----                                                    M00130 HFH-2  82.1
     ------                                                   M00130 HFH-2  82.1
     -------                                                  M00130 HFH-2  82.1
     --------                                                 M00130 HFH-2  82.1
     ---------                                                M00130 HFH-2  82.1
     ----------                                               M00130 HFH-2  82.1
     ------------                                             M00130 HFH-2  82.1
     <------------                                            M00130 HFH-2  82.1
      <------------                                           M00130 HFH-2  82.1
       <------------                                          M00130 HFH-2  82.1
        <------------                                         M00130 HFH-2  82.1
         <------------                                        M00130 HFH-2  82.1
          <------------                                       M00130 HFH-2  82.1
           <------------                                      M00130 HFH-2  82.1
                 <------------                               M00130 HFH-2  81.7
                        ------------------>                  M00206 HNF-1  81.6
                                 <----------                 M00228 VBP    81.4
                                                   <-----    M00075 GATA-1 81.2
           <------------                                      M00130 HFH-2  80.9
                     ------------------>                     M00099 S8     80.4
                                         --------------      M00036 v-Jun  80.3

4151 CCCTGTCCCA TAGCCAGAGA ATCCCAGACT TAAATTCTGC TCTTTGGATT entry        score
     ---                                                     M00113 CREB   92.6
                                  <-------                   M00100 CdxA   92.3
                                       ------>              M00101 CdxA   92.1
                                                    ---     M00100 CdxA   89.7
                        <-----------------                   M00222 Th1/E4 85.9
                              --------------->               M00162 Oct-1  85.7
                         <------------                       M00087 Ik-2   85.1
                                                    ---     M00101 CdxA   84.3
     ----                                                    M00076 GATA-2 83.8
                               <-------                     M00240 Nkx-2. 83.7
                                 <-------                   M00101 CdxA   83.6
     -------                                                 M00087 Ik-2   82.9
                                                 -------     M00133 Tst-1  81.2
                                                 <------     M00072 CP2    81.2
     ----                                                    M00075 GATA-1 81.2
                    <----------                              M00076 GATA-2 80.6
     -->                                                     M00036 v-Jun  80.3

4201 TAGGGTGTTT CTAATGGATT TTCTTTTTGT AACTTCAAGT ATTTATCTAT entry        score
                                             <-----------   M00203 GATA-X 92.8
                                                 ------>    M00100 CdxA   92.3
                                                      -     M00101 CdxA   92.1
                        <-------                             M00148 SRY    90.0
     --->                                                    M00100 CdxA   89.7
        <------                                              M00148 SRY    89.1
                                       ------->              M00240 Nkx-2. 88.4
                                              ------>        M00101 CdxA   87.9
                                           <-------          M00101 CdxA   87.1
                                  <--------                 M00041 CRE-BP 86.2
                                             <------------  M00128 GATA-1 84.8
                             <----------                     M00228 VBP    84.7
     --->                                                    M00101 CdxA   84.3
```

```
                        <---------------                M00116 C/EBPa 84.2
           --------->                                   M00227 v-Myb  83.4
                      ------>                           M00100 CdxA   83.3
                       <----------                      M00260 HLF    83.1
                                  <-------              M00101 CdxA   82.9
                                        ---------      M00099 S8     82.4
       ------->                                         M00101 CdxA   82.1
                                                  --    M00100 CdxA   82.1
                                    <------------       M00127 GATA-1 81.9
                      ------------>                     M00045 E4BP4  81.4
     ------->                                           M00133 Tst-1  81.2
     ----                                               M00072 CP2    81.2
        <------------------                             M00099 S8     81.1
            -------------->                             M00074 c-Ets- 81.0
                                     <---------------   M00267 XFD-1  80.0
                                                  <--   M00101 CdxA   80.0

4251 TTAATGTATT TGAAAGACAA AGACACAGAG ATGTCTTCCC CCTTACTGGT entry        score
     ----->                                             M00101 CdxA   92.1
                                       <--------        M00083 MZF1   90.4
     -                                                  M00128 GATA-1 84.8
                                     <-------------     M00084 MZF1   83.9
               ------>                                  M00148 SRY    82.7
                     ------->                           M00148 SRY    82.7
     ------>                                            M00099 S8     82.4
     ---->                                              M00100 CdxA   82.1
     --                                                 M00127 GATA-1 81.9
                                  <-------------        M00087 Ik-2   81.6
                                           ------>      M00101 CdxA   81.4
                            ---------->                 M00075 GATA-1 81.2
                            ---------->                 M00076 GATA-2 81.0
     ----                                               M00101 CdxA   80.0
                  ------->                              M00148 SRY    80.0

4301 TCACTCTCCA AATGTCCCCA AAGCAGGGGC TGGGCCAAGA GGGAGCCAGG entry        score
                           ---------->                  M00008 Sp1    89.0
                 <--------                              M00083 MZF1   88.7
      <--------------                                   M00159 C/EBP  85.4
                                                  <-    M00106 CDP CR 81.7

4351 GATCCAAGGG CTGGGAGCTC AACCTAGGTG TCCCATGTGG GAAGCCAGGA entry        score
                                     ------------>      M00087 Ik-2   89.5
                                   <-------             M00217 USF    86.4
           ---------->                                  M00008 Sp1    84.9
                                      --------->        M00141 Lyf-1  84.4
                                                ------  M00072 CP2    83.3
                                    ------------->      M00159 C/EBP  83.1
       --------->                                       M00227 v-Myb  82.9
                                      ------>           M00271 AML-1a 82.7
                                    ------->            M00217 USF    82.1
     --------                                           M00106 CDP CR 81.7
                                 <------------          M00123 c-Myc/ 81.4
                            <------------               M00087 Ik-2   81.1
                                 <------------          M00055 N-Myc  80.9
                                     ------------->     M00086 Ik-1   80.8
            ------------>                               M00087 Ik-2   80.7
                                 ------------>          M00055 N-Myc  80.2

4401 CCCAAATACT TGAGCAGTCC CTGCTGCCTC CCAGGGCATG TATTAGCAGG entry        score
                                                  -----  M00032 c-Ets- 91.2
           <-------                                      M00240 Nkx-2. 88.4
                                               --------  M00025 Elk-1  87.0
                                               --------  M00007 Elk-1  85.4
                            <---------                  M00141 Lyf-1  84.4
                                              -------   M00074 c-Ets- 83.4
     ---->                                              M00072 CP2    83.3
         ------->                                       M00101 CdxA   82.9
     <------                                            M00100 CdxA   80.8

4451 AAGTCAGAAT TGGGAGGAGG GGCGGTGCTT GAATCCAGGT ACTCTGATAA entry        score
         <-------                                       M00101 CdxA   92.1
```

```
     ---->                                                 M00032 c-Ets- 91.2
                                                    ------ M00203 GATA-X 90.9
             --------->                                    M00141 Lyf-1  89.6
                                                 --------- M00128 GATA-1 89.4
     ----->                                                M00025 Elk-1  87.0
                       ---------->                         M00008 Sp1    86.3
     ------->                                              M00007 Elk-1  85.4
                                                    ------ M00079 Evi-1  84.1
                                                    ------ M00082 Evi-1  83.5
     ----->                                                M00074 c-Ets- 83.4
                                                   ------- M00077 GATA-3 82.5
         --------->                                        M00241 Nkx-2. 82.4
                                                  -------- M00076 GATA-2 81.8
            ------------>                                  M00087 Ik-2   81.6
                                                ---------- M00127 GATA-1 80.7
                                                 -------- M00075 GATA-1 80.0

4501 GACAGGCAGA CGTCTCCAGC GCTGCACCAA ATGGTCACCC CAGATCCTGG entry       score
     ---->                                                 M00203 GATA-X 90.9
     --->                                                  M00128 GATA-1 89.4
                                       <------             M00271 AML-1a 87.4
                               <--------                   M00050 E2F    86.2
                                        ------------>      M00221 SREBP- 85.9
                                               ---------->  M00075 GATA-1 85.7
     ---->                                                 M00079 Evi-1  84.1
            <------------                                  M00113 CREB   83.6
     ---->                                                 M00082 Evi-1  83.5
                             ------------>                 M00072 CP2    83.3
                                  <------                  M00271 AML-1a 82.7
                                               <----------  M00076 GATA-2 82.6
     ->                                                    M00077 GATA-3 82.5
                                               <----------  M00075 GATA-1 82.4
                               ----------------->          M00066 Tal-1a 82.3
                                               ---------->  M00076 GATA-2 82.2
     ->                                                    M00076 GATA-2 81.8
     --->                                                  M00127 GATA-1 80.7
            -------->                                      M00039 CREB   80.7
            <--------                                      M00039 CREB   80.7
                             <------------------           M00059 YY1    80.2
     ->                                                    M00075 GATA-1 80.0

4551 CCTTGAGGAG CCAAGGGAAG GGGCGAAAAC TTGTGTGTGC AGTACTGTGG entry       score
                           <-------                        M00050 E2F    86.2
                                                    -- M00008 Sp1    84.9
                                                 ----- M00271 AML-1a 82.7
                                               <-------  M00101 CdxA   82.1
                                                 ------>  M00101 CdxA   82.1
                      ----------------->                   M00024 E2F    81.4
                       -------->                           M00083 MZF1   80.0

4601 GTCGGGGTCT CCATCGATCC GGAGTCTGGG GGTGTCATCC TAAATTCTGC entry       score
                  ---------->                              M00106 CDP CR 99.1
                                                    ------>  M00101 CdxA   92.1
              <----------                                  M00106 CDP CR 90.0
                                            <-------       M00100 CdxA   89.7
              <----------                                  M00104 CDP CR 89.2
                 ---------->                               M00104 CDP CR 89.2
                                      ----------------->   M00145 Brn-2  86.1
     ------->                                              M00008 Sp1    84.9
                 <----------                               M00075 GATA-1 84.5
                                            <-------       M00101 CdxA   84.3
                                          <---------       M00141 Lyf-1  83.1
                 <----------                               M00076 GATA-2 83.0
     >                                                     M00271 AML-1a 82.7
                                       <------------       M00087 Ik-2   82.5
       ---------->                                         M00053 c-Rel  81.8
              <----------                                  M00075 GATA-1 81.2
                                     <----------           M00076 GATA-2 80.6

4651 TAACAACTGG GCGCAGCACC ACGCAGCCCG GGACCCCAGA GCCCTCCACG entry       score
                                     <-----                M00271 AML-1a 87.4
```

```
                   <------                              M00271 AML-1a 84.1
     ------>                                            M00148 SRY    82.7
                                                <----- M00083 MZF1   80.0

4701 CTCACTGACC CAGACCCTCC ATCCCCAACC CCTCTCTCCT CCAAAATGGC entry        score
                    <----------                         M00075 GATA-1 91.4
                         <-------                       M00083 MZF1   91.3
        ----------->                                    M00173 AP-1   89.7
                    <----------                         M00076 GATA-2 88.9
                                          <---------    M00141 Lyf-1  88.3
                                          <------------ M00059 YY1    85.7
         <---------                                     M00199 AP-1   83.4
                                  <---------            M00077 GATA-3 82.5
                    <---------                          M00077 GATA-3 81.9
                              <---------------          M00126 GATA-1 81.0
         --------->                                     M00199 AP-1   80.6
                       <---------                       M00141 Lyf-1  80.5
                     <------------                      M00087 Ik-2   80.3
                           -------------->              M00257 RREB-1 80.2
     --                                                 M00083 MZF1   80.0

4751 TTCAGAATTC CCAAATTCTG ATGTAGATGC TGCACAGGGT ATACCAGCGT entry        score
         <------------                                  M00087 Ik-2   96.5
       <------                                          M00101 CdxA   92.1
                  ------>                               M00101 CdxA   92.1
        <-------------                                  M00086 Ik-1   90.7
                 --------------->                       M00162 Oct-1  87.8
           <---------                                   M00141 Lyf-1  87.0
         <-------------                                 M00159 C/EBP  86.9
     ----                                               M00059 YY1    85.7
                             ---------->                M00075 GATA-1 84.9
         <--------------                                M00109 C/EBPb 82.4
        <-------------                                  M00088 Ik-3   82.1
                     ---------->                        M00075 GATA-1 81.2
              <-------                                   M00100 CdxA   80.8

4801 CTCCTTAGAG TCTCTCGCGG CTGGGCCCAC AGTGGCGGTT AACCCAGATC entry        score
                                                <------ M00076 GATA-2 84.6
                                                ------- M00075 GATA-1 84.1
                                                <------ M00075 GATA-1 84.1
                                <------                 M00271 AML-1a 82.7
                     ---------->                        M00008 Sp1    82.2
                                                ------- M00076 GATA-2 80.6
                                                <------ M00074 c-Ets- 80.2

4851 CGCTCCCCAA GCGACTTGAC CTTCACTCTG AGAGTGCAGC TGCTACTGGA entry        score
        <--------                                       M00083 MZF1   88.7
                   <-------------                       M00156 RORalp 87.0
                <------                                 M00240 Nkx-2. 86.0
     ---                                                M00076 GATA-2 84.6
                                                -------- M00074 c-Ets- 84.6
                                                       - M00249 CHOP-C 84.3
     -->                                                M00075 GATA-1 84.1
     ---                                                M00075 GATA-1 84.1
                                                 ------ M00032 c-Ets- 82.4
                                                     <- M00133 Tst-1  81.2
     -->                                                M00076 GATA-2 80.6
      <--------                                         M00141 Lyf-1  80.5
     ------                                             M00074 c-Ets- 80.2

4901 ACTGCAATTT CCTCTCCTCT GCTTACATAT CTGTATAAAC CCCTTTATGG entry        score
                                                ------>  M00100 CdxA   96.2
                                        <------          M00101 CdxA   87.9
                                        <------          M00100 CdxA   87.2
                                                ------>  M00101 CdxA   87.1
         <-------------                                 M00074 c-Ets- 86.6
                              <-------------            M00128 GATA-1 85.4
                                                   <- M00159 C/EBP  85.4
     ---->                                              M00074 c-Ets- 84.6
     ------------>                                      M00249 CHOP-C 84.3
                               <----------              M00076 GATA-2 84.2
```

```
  ---------->                                          M00052 NF-kap 84.1
                      <------------                    M00045 E4BP4  82.8
                       <-----------                    M00203 GATA-X 82.8
--->                                                   M00032 c-Ets- 82.4
                          <----------                  M00075 GATA-1 81.6
--------------                                         M00133 Tst-1  81.2
  ---------->                                          M00053 c-Rel  81.0
                            ------->                   M00101 CdxA   80.7
                          <---------                   M00077 GATA-3 80.6
                                              <--------  M00216 TATA   80.5

4951 GTTAGCAAAT GAAATTTTAT AAAGACAAGT GTGTAGGGGT TCCCACGAAA entry        score
                   ------->                            M00101 CdxA   92.9
                      <-------                         M00100 CdxA   92.3
                              ------->                 M00240 Nkx-2. 90.7
                      <-------                         M00101 CdxA   87.9
                   ------->                            M00100 CdxA   87.2
                                        <-------------   M00086 Ik-1   86.5
                <----------                            M00216 TATA   86.4
------------                                           M00159 C/EBP  85.4
                                         <------------   M00087 Ik-2   84.6
                                        <-------------   M00088 Ik-3   84.3
     <--------------                                   M00162 Oct-1  83.7
                  ----------------->                   M00252 TATA   83.5
              ------>                                  M00101 CdxA   82.9
                        ------>                        M00148 SRY    82.7
                   ------------->                      M00137 Oct-1  82.2
                          <---------                   M00217 USF    82.0
       ------->                                        M00148 SRY    80.9
                 <----------                           M00216 TATA   80.6
-                                                      M00216 TATA   80.5
                       --------------->                M00122 USF    80.5
                       <---------------                M00122 USF    80.5

5001 GCTTGAACAG GGAGTGGGAG CACCCGGAGC GCGGAGCCTC AGCAGCCCCG entry        score
          -------------->                              M00033 p300   97.0
               ------------>                           M00087 Ik-2   84.6
                                    <-----------       M00037 NF-E2  81.5
          ---------->                                  M00032 c-Ets- 80.4
                                                 <-    M00257 RREB-1 80.2

5051 GGGCGCTTGG GGACTTGGGG GCTCCGGATC CTGGGGCTCC GGGGTGGGGG entry        score
          -------->                                    M00083 MZF1   88.7
                           ---------->                 M00075 GATA-1 85.3
                           <----------                 M00075 GATA-1 84.5
                           ---------->                 M00076 GATA-2 84.2
                           <----------                 M00076 GATA-2 83.8
                                             ------->  M00083 MZF1   83.5
                         ------------->                M00074 c-Ets- 82.2
                                           --------->  M00008 Sp1    82.2
-------------                                          M00257 RREB-1 80.2

5101 TGCTGAGCAC AGAGGGCTAC TGCGGAGCTG AAGGCGTTGT TCCAAGCGCC entry        score
                                                <----  M00050 E2F    86.2
     <------                                           M00271 AML-1a 81.7
<--------------                                        M00159 C/EBP  80.8

5151 AAGGATTTGG GACCCGGCCC GGAGACGCCC CACGCCGCTG TGTTCGGCTC entry        score
---                                                    M00050 E2F    86.2
            <----------                                M00008 Sp1    84.9
     --------->                                        M00141 Lyf-1  83.1
    ------------>                                      M00087 Ik-2   82.9
                                          ------>      M00271 AML-1a 81.7
                              <--------                M00083 MZF1   80.9
    ------->                                           M00100 CdxA   80.8
                                                 ---   M00223 STATx  80.8
                    ------------>                      M00113 CREB   80.3

5201 CTGGAAGGAA TTGGGTCCCC AGCCCCGGAC TCTCCCTGCC TCTTGCCATA entry        score
                  <--------                            M00083 MZF1   90.4
                                <----------            M00008 Sp1    87.7
```

```
                  <----------                          M00008 Sp1    86.3
                                                    < M00008 Sp1    80.8
     ----->                                           M00223 STATx  80.8
                            <--------------           M00033 p300   80.2
                            ---------->               M00053 c-Rel  80.2

5251 GCCAGCCCGG TCCCGGACTG CGCATCCTCG GTTTCCCAGC CCCCTGGGGT entry        score
                                     <----------        M00008 Sp1    90.4
                                  <-------------        M00087 Ik-2   89.9
                                                 -----> M00271 AML-1a 87.4
                        <----------                      M00076 GATA-2 85.4
                        <----------                      M00075 GATA-1 85.3
     ---------                                           M00008 Sp1    80.8

5301 GTCTGCAGGC CGGGCTACTT GCACAGCAGC AGGTGCGTAG GCGGGGCGCG entry        score
                                     ------------>      M00246 Egr-2  96.0
                                     ------------>      M00245 Egr-3  93.1
                                     ------------>      M00243 Egr-1  90.7
                                     ------------>      M00244 NGFI-C 90.7
                                          ---------->   M00008 Sp1    84.9
                                   <-----------         M00073 deltaE 84.9
                                    <--------           M00217 USF    84.3
                                   ---------------->    M00002 E47    81.7
                                    ------------>       M00001 MyoD   81.4
                                                 -------- M00008 Sp1    80.8

5351 CAGCATTTAA GGCGGACACC ACCTCCCCTG GGCAGCGGCT GGCGATCGGC entry        score
         ------->                                       M00100 CdxA   92.3
                                                 ---------> M00075 GATA-1 90.2
                                                 <--------- M00075 GATA-1 87.3
                                                 <--------- M00076 GATA-2 87.0
        <--------                                       M00241 Nkx-2. 85.3
                                                 ---------> M00076 GATA-2 85.0
                     <------                            M00271 AML-1a 84.1
                    <------------                       M00001 MyoD   83.7
         ------->                                       M00101 CdxA   83.6
                              <--------                 M00083 MZF1   82.6
     <------                                            M00101 CdxA   81.4
        <------                                         M00240 Nkx-2. 81.4
     ->                                                 M00008 Sp1    80.8

5401 TGCGGAGGTG CGCGCAGGGC CCGCGTGGCT GTGGGTACCT CCTTCGCCAG entry        score
                                 ------>                M00271 AML-1a 82.7
                       ------------>                    M00055 N-Myc  82.7
                          ------->                      M00217 USF    81.7

5451 CACCGTCGCC ACTACCAACG CCGCCACCGC GGGACCCTAC CCCGCATCGG entry        score
                                               <--------- M00075 GATA-1 86.9
                                        <----------       M00008 Sp1    86.3
                                                   ------ M00104 CDP CR 84.9
                                                   ------ M00106 CDP CR 84.0
          <--------                                       M00083 MZF1   83.5
                      <----------                         M00008 Sp1    82.2
                                                 <------- M00106 CDP CR 82.2
                                                <--------- M00076 GATA-2 81.8
                                             <----------  M00008 Sp1    80.8

5501 TCGCCGCCGC CACCGCAGGT CCCACGACCC CTCCTGCCCT CCGGCGCCCC entry        score
                <-----                                  M00271 AML-1a 92.0
     -                                                  M00075 GATA-1 86.9
     --->                                               M00104 CDP CR 84.9
                                                  <------ M00002 E47    84.6
     --->                                               M00106 CDP CR 84.0
     --                                                 M00106 CDP CR 82.2
     -                                                  M00076 GATA-2 81.8
                                                    <---- M00083 MZF1   81.7
                                                       <- M00008 Sp1    80.8
           <------------                                M00243 Egr-1  80.7
           <------------                                M00246 Egr-2  80.3

5551 CCTGCCCGCC GACCCGCGGC GCCTCCAGCG CGACATGCGC CAGGTGTGTG entry        score
```

```
                                        <-----------  M00073 deltaE 87.6
                                 <--------            M00050 E2F    86.2
     --------                                        M00002 E47    84.6
                                          -------->    M00217 USF    83.0
                                       ---------------  M00002 E47    82.7
     ---                                             M00083 MZF1   81.7
                                        ------------>  M00001 MyoD   81.4
                                                   --  M00033 p300   81.2
                                          <--------    M00217 USF    81.1
                                                  ----  M00083 MZF1   80.9
     --------                                        M00008 Sp1    80.8
                                        -------------  M00033 p300   80.2

5601 AGGAAGTGAC GCGGTGCGGA CTGAAGAGAA GTGCGGGAAA GGGTGAAGGG entry        score
                                      <-------          M00050 E2F    90.8
         ----------->                                   M00173 AP-1   84.5
                                    ------------>       M00087 Ik-2   82.9
     >                                               M00002 E47    82.7
                 ------->                            M00083 MZF1   82.6
      ------>                                        M00240 Nkx-2. 81.4
                               ------->               M00240 Nkx-2. 81.4
     ------------>                                   M00033 p300   81.2
                             --------------->          M00033 p300   81.2
     --->                                            M00083 MZF1   80.9
                                        <------------ M00155 ARP-1  80.8
     ->                                              M00033 p300   80.2

5651 CTCCGTCCGG GGGTCTTTAC TCTGCAACCC TGTTCCAGCC GCCGAGCACC entry        score
                                                   ---  M00217 USF    82.9
      <-----------                                  M00032 c-Ets- 81.4
                ---------->                         M00053 c-Rel  81.0
     ----                                            M00155 ARP-1  80.8

5701 CGTGTGTCAC TCGGGAACTG GCTGGGTAAA GAGGTCAATC CAGACACGCG entry        score
                                                  ----  M00083 MZF1   90.4
                                 ------------->         M00156 RORalp 88.9
                                           <--------    M00055 N-Myc  86.9
                               ------------->           M00137 Oct-1  86.9
                                             <------    M00217 USF    84.2
                                 ------------->         M00157 RORalp 83.9
     ---->                                           M00217 USF    82.9
                                    <--------           M00039 CREB   81.7
                                          <----------------  M00222 Th1/E4 81.6
     <------------                                   M00173 AP-1   81.4
                       ---------->                   M00008 Sp1    80.8
                                  <---------         M00106 CDP CR 80.4
              ------------>                          M00087 Ik-2   80.3
                                     <----------     M00075 GATA-1 80.0

5751 GGGAAGGAGT TCCAGGGGTC AGCTCCGCCC TCGCACCTGC GGGCTCGGAT entry        score
     --->                                            M00083 MZF1   90.4
                               <------------          M00246 Egr-2  88.3
                                      ------------>    M00073 deltaE 87.1
     ---                                             M00055 N-Myc  86.9
                               <------------          M00244 NGFI-C 86.8
                               <------------          M00245 Egr-3  86.6
                               <------------          M00243 Egr-1  86.2
                                                 <  M00223 STATx  84.6
                                       ------->       M00217 USF    84.3
     -                                               M00217 USF    84.2
                                                ------  M00076 GATA-2 83.8
                                  <----------------    M00002 E47    83.7
      --------------->                               M00033 p300   83.2
                                                ------  M00075 GATA-1 82.4
     -                                               M00222 Th1/E4 81.6
                               <---------            M00083 MZF1   80.9

5801 TCGGAGAAAA GTGCTAGACT GGAGCTACAC GTATGCGTAG CGGTCTGGAA entry        score
                                                  <----  M00208 NF-kap 86.2
     --------                                        M00223 STATx  84.6
     --->                                            M00076 GATA-2 83.8
```

```
                                                  <--- M00054 NF-kap 82.9
                                                  <--- M00052 NF-kap 82.8
                                                   <-- M00053 c-Rel  82.6
     --->                                              M00075 GATA-1 82.4
                                                  <--- M00053 c-Rel  81.8
                                               --------- M00159 C/EBP  81.5
          ------->                                     M00240 Nkx-2. 81.4
                                               --------- M00074 c-Ets- 81.0
                                                     <-- M00052 NF-kap 80.8
         ----------->                                 M00203 GATA-X 80.8
                                                   <-- M00054 NF-kap 80.7
                                          ---------- M00222 Th1/E4 80.0

5851 AATGCCCCAG GCTCGGGTCT GAGGGGCCCA AGTCTATGCA CCGCTGGTGT entry       score
     -------                                              M00208 NF-kap 86.2
     ------                                               M00054 NF-kap 82.9
     ------                                               M00052 NF-kap 82.8
     -------                                              M00053 c-Rel  82.6
     ------                                               M00053 c-Rel  81.8
                                             ------------ M00002 E47    81.7
     --->                                                 M00159 C/EBP  81.5
     --->                                                 M00074 c-Ets- 81.0
     -------                                              M00052 NF-kap 80.8
     -------                                              M00054 NF-kap 80.7
     ->                                                   M00222 Th1/E4 80.0

5901 GACCCCGCAG GGCAACCCCG CGGTTAACTT CTCTCCTGCC CACCCCTAGA entry       score
               <----------                                M00051 NF-kap 85.2
                           <-----------                   M00079 Evi-1  84.1
                 <---------                               M00053 c-Rel  82.6
                           <-----------                   M00082 Evi-1  81.8
     --->                                                 M00002 E47    81.7
                           <-----------                   M00080 Evi-1  81.4
                                       <--------          M00083 MZF1   80.9
                        <----------------                 M00127 GATA-1 80.7

5951 GGTGTCTTCC TGGGAAGACG ATGGCAGGCG GTGCCCACCG AGCCGACCGT entry       score
                     ---------->                          M00075 GATA-1 88.2
                     ---------->                          M00076 GATA-2 86.6
            ------------>                                 M00087 Ik-2   86.0
            ------------->                                M00088 Ik-3   85.1
            <---------                                    M00223 STATx  83.7
          <----------                                     M00032 c-Ets- 83.3
            ------------->                                M00086 Ik-1   82.4
           --------->                                     M00223 STATx  81.7
                                    <-------              M00083 MZF1   80.9
          <----------                                     M00108 NRF-2  80.7
         <-------------                                   M00074 c-Ets- 80.2

6001 GCAACAGGGG AAGAGAGGAA GGAGGGAGGT GGGAGGTGGC GCGCTCCCCA entry       score
                                                  <----- M00083 MZF1   95.7
        -------->                                         M00083 MZF1   87.0
                                                   <--- M00271 AML-1a 82.7
          ---------->                                     M00108 NRF-2  82.5
                           <-----------                   M00073 deltaE 82.3
                               --------->                 M00141 Lyf-1  81.8
                -------->                                M00083 MZF1   80.0

6051 CAGCCCTTCC CCTCCTGGCC CGCGAGGGTG TCCGGTCCCA CTCAAGGCAG entry       score
     --                                                   M00083 MZF1   95.7
           <--------                                      M00083 MZF1   93.0
                              <----------                 M00032 c-Ets- 85.3
     --                                                   M00271 AML-1a 82.7
                                          <-------        M00240 Nkx-2. 81.4

6101 CTGCGCAGAG CCTGTGCAGA AAACCCACCT GGGGCCGGTA TTGCACTCTG entry       score
                            ----------->                  M00073 deltaE 90.0
                                       <-------------     M00249 CHOP-C 89.8
                     <-------                             M00101 CdxA   86.4
                              --------->                  M00217 USF    85.0
                          ------------------>             M00005 AP-4   81.7
```

```
                     <----------                          M00053 c-Rel  81.0
                    <-------                              M00100 CdxA   80.8
                            <---------                   M00217 USF    80.6
                         ------>                         M00148 SRY    80.0
                                        <-------          M00101 CdxA   80.0

6151 CTTCTCTTTC AGAGAAAGCT GGAAATTTAC TCCGTGGAGC ACCATGCAGC entry        score
                    ------------->                       M00074 c-Ets- 84.2
            <---------                                   M00223 STATx  82.7
                        <-------                         M00101 CdxA   82.1
                              ------->                   M00100 CdxA   80.8
       <------                                           M00148 SRY    80.0

6201 TACAGATATC AAGAAGAAGG AGGGGCGAGA TGGCAAGAAA GACAATGACT entry        score
                                ---------->              M00075 GATA-1 91.0
                                ---------->              M00076 GATA-2 87.4
                                                 ------- M00173 AP-1   84.5
        <----------                                      M00075 GATA-1 84.5
      ---------->                                        M00076 GATA-2 83.4
     -------------->                                     M00128 GATA-1 83.0
                                           ------->      M00148 SRY    82.7
        <----------                                      M00076 GATA-2 82.6
                     ------------->                      M00084 MZF1   82.3
                             ---------->                 M00008 Sp1    82.2
                                  ---------->            M00077 GATA-3 81.6
       <------                                           M00101 CdxA   80.7
                            ------->                    M00083 MZF1   80.0

6251 TGGAACTCAA AAGGAATCAG CAGAAAGAGG AGCTTAAGAA AGAACTTGAT entry        score
                <----------                              M00075 GATA-1 87.8
                                           ------->      M00148 SRY    86.4
                                                <------  M00240 Nkx-2. 86.0
                            <------                      M00100 CdxA   85.9
     --->                                                M00173 AP-1   84.5
                                      --------------->   M00109 C/EBPb 84.3
                <-----------                             M00037 NF-E2  84.0
                <----------                              M00076 GATA-2 82.6
                           <------                       M00101 CdxA   81.4


 Total 1576 high-scoring sites found.
 Max score: 100.0 point, Min score: 80.0 point
```

LIST OF REFERENCES

1. **Ahn, KY, KY Park, KK Kim, and BC Kone.** 1996. Chronic hypokalemia enhances expression of the H(+)-K(+)-ATPase alpha 2-subunit gene in renal medulla. Am J Physiol **271:**F314-21.

2. **Arend, LJ, JS Handler, JS Rhim, F Gusovsky, and WS Spielman.** 1989. Adenosine-sensitive phosphoinositide turnover in a newly established renal cell line. Am J Physiol **256:**F1067-74.

3. **Barri, YM, and CS Wingo.** 1997. The effects of potassium depletion and supplementation on blood pressure: a clinical review. Am J Med Sci **314:**37-40.

4. **Campbell, NA.** 1996. Biology, 4th ed. Benjamin/Cummings, Menlo Park, CA

5. **Campbell, WG.** 1998. Characterization of the rabbit renal H,K-ATPases. Dissertation. Unversity of Florida, Gainesville

6. **Campbell, WG, ID Weiner, CS Wingo, and BD Cain.** 1999. H-K-ATPase in the RCCT-28A rabbit cortical collecting duct cell line. Am J Physiol **276:**F237-45.

7. **Carey, M, and ST Smale.** 2000. Transcriptional regulation in eukaryotes : concepts, strategies, and techniques. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

8. **Caviston, TL, WG Campbell, CS Wingo, and BD Cain.** 1999. Molecular identification of the renal H+,K+-ATPases. Semin Nephrol **19:**431-7.

9. **Crowson, MS, and GE Shull.** 1992. Isolation and characterization of a cDNA encoding the putative distal colon H+,K(+)-ATPase. Similarity of deduced amino acid sequence to gastric H+,K(+)-ATPase and Na+,K(+)-ATPase and mRNA expression in distal colon, kidney, and uterus. J Biol Chem **267:**13740-8.

10. **DuBose, TD, Jr., J Codina, A Burges, and TA Pressley.** 1995. Regulation of H(+)-K(+)-ATPase expression in kidney. Am J Physiol **269:**F500-7.

11. **Eiam-Ong, S, NA Kurtzman, and S Sabatini.** 1993. Regulation of collecting tubule adenosine triphosphatases by aldosterone and potassium. J Clin Invest **91:**2385-92.

12. **Fejes-Toth, G, A Naray-Fejes-Toth, and H Velazquez.** 1999. Intrarenal distribution of the colonic H,K-ATPase mRNA in rabbit. Kidney Int **56:**1029-36.

13. **Fejes-Toth, G., E Rusvai, KA Longo, and A Naray-Fejes-Toth.** 1995. Expression of colonic H-K-ATPase mRNA in cortical collecting duct: regulation by acid/base balance. Am J Physiol **269:**F551-7.

14. **Garg, LC, and Y Komatsu.** 1994. In vitro and in vivo effects of ammonium chloride on H-K-ATPase activity in the outer medullary collecting duct. Contrib Nephrol **110:**67-74.

15. **Georgescu, HI, D Mendelow, and CH Evans.** 1988. HIG-82: an established cell line from rabbit periarticular soft tissue, which retains the "activatable" phenotype. In Vitro Cell Dev Biol **24:**1015-22.

16. **Grishin, AV, MO Bevensee, NN Modyanov, V Rajendran, WF Boron, and MJ. Caplan.** 1996. Functional expression of the cDNA encoded by the human ATP1AL1 gene. Am J Physiol **271:**F539-51.

17. **Grishin, AV, VE Sverdlov, MB Kostina, and NN Modyanov.** 1994. Cloning and characterization of the entire cDNA encoded by ATP1AL1--a member of the human Na,K/H,K-ATPase gene family. FEBS Lett **349:**144-50.

18. **Gumz, ML, D Duda, R McKenna, CS Wingo, and BD Cain.** 2003. Molecular modeling of the rabbit colonic (HKalpha2a) H(+), K(+) ATPase. J Mol Model (Online).

19. **Guntupalli, J, M Onuigbo, S Wall, RJ Alpern, and TD DuBose, Jr.** 1997. Adaptation to low-K+ media increases H(+)-K(+)-ATPase but not H(+)-ATPase-mediated pHi recovery in OMCD1 cells. Am J Physiol **273:**C558-71.

20. **Heinemeyer, T, E Wingender, I Reuter, H Hermjakob, AE Kel, OV Kel, EV Ignatieva, EA Ananko, OA Podkolodnaya, FA Kolpakov, NL Podkolodny, and NA Kolchanov.** 1998. Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic Acids Res **26:**362-7.

21. **Israel, DI.** 1993. A PCR-based method for high stringency screening of DNA libraries. Nucleic Acids Res **21:**2627-31.

22. **Jaisser, F, N Coutry, N Farman, HJ Binder, and BC Rossier.** 1993. A putative H(+)-K(+)-ATPase is selectively expressed in surface epithelial cells of rat distal colon. Am J Physiol **265:**C1080-9.

23. **Jaisser, F, B Escoubet, N Coutry, E Eugene, JP Bonvalet, and N Farman.** 1996. Differential regulation of putative K(+)-ATPase by low-K+ diet and corticosteroids in rat distal colon and kidney. Am J Physiol **270:**C679-87.

24. **Kim, JK, SN Summer, and T Berl.** 1984. The cyclic AMP system in the inner medullary collecting duct of the potassium-depleted rat. Kidney Int **26:**384-91.

25. **Koeppen, BM, and BA Stanton.** 2000. Renal physiology, 3rd ed. Mosby, St. Louis.

26. **Kone, BC, and SC Higham.** 1998. A novel N-terminal splice variant of the rat H+-K+-ATPase alpha2 subunit. Cloning, functional expression, and renal adaptive response to chronic hypokalemia. J Biol Chem **273:**2543-52.

27. **Kraut, JA, KG Helander, HF Helander, ND Iroezi, EA Marcus, and G Sachs.** 2001. Detection and localization of H+-K+-ATPase isoforms in human kidney. Am J Physiol Renal Physiol **281:**F763-8.

28. **Kuwahara, M, WJ Fu, and F Marumo.** 1996. Functional activity of H-K-ATPase in individual cells of OMCD: localization and effect of K+ depletion. Am J Physiol **270:**F116-22.

29. **Lagrange, T, AN Kapanidis, H Tang, D Reinberg, and RH Ebright.** 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. Genes Dev **12:**34-44.

30. **Laroche-Joubert, N, and A Doucet.** 1999. Collecting duct adaptation to potassium depletion. Semin Nephrol **19:**390-8.

31. **Lodish, HF, and JE Darnell.** 1995. Molecular cell biology, 3rd ed. Scientific American Books, New York.

32. **Lutsenko, S, and JH Kaplan.** 1995. Organization of P-type ATPases: significance of structural diversity. Biochemistry **34:**15607-13.

33. **Marsy, S, JM Elalouf, and A Doucet.** 1996. Quantitative RT-PCR analysis of mRNAs encoding a colonic putative H, K- ATPase alpha subunit along the rat nephron: effect of K+ depletion. Pflugers Arch **432:**494-500.

34. **Meneton, P, PJ Schultheis, J Greeb, ML Nieman, LH Liu, LL Clarke, JJ Duffy, T Doetschman, JN Lorenz, and GE Shull.** 1998. Increased sensitivity to K+ deprivation in colonic H,K-ATPase-deficient mice. J Clin Invest **101:**536-42.

35. **Modyanov, NN, KE Petrukhin, VE Sverdlov, AV Grishin, MY Orlova, MB Kostina, OI Makaravich, NE Broude, GS Monastyrskaya, and ED Sverdlov.** 1991. The Family of Na,K-ATPase Genes: ATP1AL1 gene is transcriptionally competent and probably encodes the related ion transport ATPase. FEBS **278:**91-94.

36. **Nakamura, S, H Amlal, JH Galla, and M Soleimani.** 1998. Colonic H+-K+-ATPase is induced and mediates increased HCO3- reabsorption in inner medullary collecting duct in potassium depletion. Kidney Int **54:**1233-9.

37. **Otto, TC.** 2001. Molecular analysis of F1F0 adenosine triphosphate synthase and renal H+,K+-adenosine triphosphatases. Dissertation. University of Florida, Gainesville.

38. **Ryan, MJ, G Johnson, J Kirk, SM Fuerstenberg, RA Zager, and B Torok-Storb.** 1994. HK-2: an immortalized proximal tubule epithelial cell line from normal adult human kidney. Kidney Int **45:**48-57.

39. **Sambrook, J, EF Fritsch, and T Maniatis.** 1989. Molecular cloning : a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

40. **Sangan, P, VM Rajendran, AS Mann, M Kashgarian, and HJ Binder.** 1997. Regulation of colonic H-K-ATPase in large intestine and kidney by dietary Na depletion and dietary K depletion. Am J Physiol **272:**C685-96.

41. **Silver, RB, H Choe, and G Frindt.** 1998. Low-NaCl diet increases H-K-ATPase in intercalated cells from rat cortical collecting duct. Am J Physiol **275:**F94-102.

42. **Simmons, NL.** 1990. A cultured human renal epithelioid cell line responsive to vasoactive intestinal peptide. Exp Physiol **75:**309-19.

43. **Stevens, MS, and RW Dunlay.** 2000. Hyperkalemia in hospitalized patients. Int Urol Nephrol **32:**177-80.

44. **Stokes, DL, and NM Green.** 2003. Structure and function of the calcium pump. Annu Rev Biophys Biomol Struct.

45. **Sverdlov, VE, MB Kostina, and NN Modyanov.** 1996. Genomic organization of the human ATP1AL1 gene encoding a ouabain-sensitive H,K-ATPase. Genomics **32:**317-27.

46. **Toyoshima, C, M Nakasako, H Nomura, and H Ogawa.** 2000. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. Nature **405:**647-55.

47. **Toyoshima, C, and H Nomura.** 2002. Structural changes in the calcium pump accompanying the dissociation of calcium. Nature **418:**605-11.

48. **Verlander, J, R Moudy, WG Campbell, BD Cain, and CS Wingo.** 2001. Immunohistochemical localization of H-K-ATPase alpha 2c in rabbit kidney. Am J Physiol Renal Physiol **272:**F357-65.

49. **Weiner, ID, and CS Wingo.** 1998. Hyperkalemia: a potential silent killer. J Am Soc Nephrol **9:**1535-43.

50. **Weiner, ID, and CS Wingo.** 1997. Hypokalemia--consequences, causes, and correction. J Am Soc Nephrol **8:**1179-88.

51. **Wingo, CS, and BD Cain.** 1993. The renal H-K-ATPase: physiological significance and role in potassium homeostasis. Annu Rev Physiol **55:**323-47.

52. **Zhang, W, T Kuncewicz, SC Higham, and BC Kone.** 2001. Structure, promoter analysis, and chromosomal localization of the murine H(+)/K(+)-ATPase alpha 2 subunit gene. J Am Soc Nephrol **12:**2554-64.

53. **Zies, DL, CS Wingo, and BD Cain.** 2002. Molecular regulation of the HKalpha2 subunit of the H+,K(+)-ATPases. J Nephrol **15 Suppl 5:**S54-60.

BIOGRAPHICAL SKETCH

Deborah Milon Zies was born on December 6, 1964, to Frank and Barbara Milon. She grew up in New York with her older sister Patricia and younger brother Frank. She graduated from Pelham Memorial High School in May of 1982 and moved to Florida where she attended Rollins College. Deborah graduated from Rollins College in May of 1986 with a major in biology and a minor in teacher certification. Upon graduation, she remained in Florida where she taught biology at Oviedo High School. In May of 1988, Deborah married Peter Zies and moved to New Orleans, Louisiana, where she began graduate school at Tulane University. There, she worked under the supervision of Dr. David Mullin and completed a master's thesis project entitled "The Genetic Characterization of Insertion Element IS*511* from *Caulobacter cresentus.*" Deborah graduated from Tulane University with a Master of Science degree in December of 1990. She then returned to Orlando, Florida, and worked as a laboratory technician at the United States Department of Agriculture Horticultural Research Laboratory. During her four year appointment, she worked with Dr. Stephen Garnsey on the development of molecular techniques for the detection of citrus diseases. It was also during this time that she gave birth to her daughter Lee Ann and became divorced from her husband. At the end of the appointment, Deborah decided to return to teaching. Over the next three years she taught biology and chemistry at Apopka High School, Seminole Community College and Valencia Community College, all in the Orlando area. While teaching, Deborah came to realize that although teaching was rewarding, she missed research. She therefore

decided to return to graduate school for a Doctor of Philosophy degree. With this degree she would be able to obtain a teaching position that had a research component. In the fall of 1998, Deborah entered the Interdisciplinary Program at the University of Florida. She has carried out the work described in this dissertation in the laboratory of Dr. Brian Cain. She received a predoctoral fellowship from the American Heart Association which supported her for two years. Additionally, she received an outstanding graduate student award from the Interdisciplinary Program and the Boyce Award for outstanding research from the Department of Biochemistry and Molecular Biology. Upon graduation, Deborah will begin a postdoctoral position at the Mayo Clinic in Jacksonville, Florida. She will be studying changes in gene expression associated with colon cancer. From there, Deborah hopes to obtain a teaching position at a small college and combine her love for teaching and research into one career.