LIBRARY
Michigan State
University

PLACE IN RETURN BOX
to remove this checkout from your record.
TO AVOID FINES return on or before date due.

DATE DUE	DATE DUE	DATE DUE

1/98 c:/CIRC/DateDue.p65-p.14

IDENTIFICATION AND CHARACTERIZATION OF PROMOTER REGIONS OF THE GENE FOR RAT TYPE I HEXOKINASE

By

Wenjing Liu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1997

ABSTRACT

IDENTIFICATION AND CHARACTERIZATION OF PROMOTER REGIONS OF THE GENE FOR RAT TYPE I HEXOKINASE

By

Wenjing Liu

The 5'-flanking region of the gene for rat type I hexokinase has previously been isolated from genomic clones and sequenced. 5'-RACE, RT-PCR and RNase protection assays indicated that there are multiple transcriptional start sites clustered in three regions at positions approximately -460, -300, and -100 relative to the translational start codon. These regions lack classical TATA sequence and are located in a GC-rich segment, a "CpG island". These characteristics are frequently associated with "housekeeping genes". The goal of this dissertation is to identify the promoter regions of type I hexokinase gene and to characterize important cis-elements and corresponding trans-factors regulating the promoter activity.

PC12 cells and H9c2 cells were transfected with luciferase reporter constructs containing genomic sequence between positions -3366 and -171. Marked (85%) decrease in promoter activity was associated with deletion of sequence between -742 and -516. In DNase I footprinting experiments, two regions, called P1 (-552 to -529) and P2 (-480 to -458) boxes, were protected by proteins present in nuclear extracts from PC12 cells. The P2 box overlaps with the most upstream cluster of transcriptional start sites, which is about 80 bp downstream from the P1 box.

Mutations or deletions in the P2 box had no effect on promoter activity. In contrast, mutations or deletions in the P1 box had markedly detrimental effects on promoter activity. A second Sp1 site (-570), just upstream from the P1 box, was also shown to be functionally important although not protected in footprinting experiments. Furthermore, the P1 box could be functionally replaced by the -570 Sp1 site.

Two DNA-protein complexes were observed in gel-shift experiments with P1 box sequence and PC12 nuclear extract. Maintenance of a consensus Sp1 binding site centrally located in the P1 box was critical for the formation of both complexes. Supershift experiments demonstrated the involvement of Sp1, Sp3, and Sp4 in formation of these complexes, and implicate these transcription factors in regulating promoter activity associated with this region.

Another series of reporter constructs, including sequence between -171 and -1, permitted detection of an additional promoter activity downstream from -364. While not yet extensively characterized, it is already evident that the cis-elements influencing the downstream promoter activity are distinct from the Sp factors determined to be important in expression from the upstream promoter region.

To my family

ACKNOWLEDGMENTS

I am grateful to my mentor, Dr. John E. Wilson for his guidance and encouragement over the years. I learned so much from his discipline in science and generosity in social life. I would like to express my deep appreciation to my committee members, Dr. Paul Coussens, Dr. Don Jump, Dr. Steve Triezenberg and Dr. John Wang for their criticisms and advice on my work. I also thank Dr. Zach Burton for his discussion and attendance at my dissertation defense.

I would like to thank my colleagues, past and present Wilson lab. members. They were always friendly and willing to help. Whether attending a scientific meeting or just a lunch out together, I enjoyed their company and friendship. Special thank goes to Dr. Joe White who taught me molecular biology techniques and cooperated with me at early stage of my research project.

I feel extremely lucky to make friends with some Chinese women here in the Department of Biochemistry at MSU. They are Yin Tang, Jie Qian, Yue Li and Tong Hao. Because of our common cultural and academic background, we shared many things personal and professional. I thank them for being always available when I need comfort and support in difficult times.

I want to thank my parents, sister and brother back in China. Their unconditional love and firm faith in me pushed me go this far. Last, but not least, I thank my husband, Yijun Guo, for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Glc-6-P	glucose-6-phosphate
KDa	kilodalton
M_r	molecular weight
HK	hexokinase
GK	glucokinase
RACE	rapid amplification of cDNA ends
PCR	polymerase chain reaction
RT-PCR	reverse transcription followed by PCR
bp	base pair(s)
Kb	kilobase pair(s)
nt	nucleotide(s)
h	hour(s)
min	minute(s)
Sp	specificity protein
SV40	simian virus 40
CMV	cytomegalovirus
μg	microgram(s)
μl	microlitre(s)

cAMP	cyclic AMP
UTR	untranslated region
P_i	orthophosphate
dig	digoxigenin
Hepes	N-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
PBS	phosphate buffered saline
EDTA	(ethylenedinitrilo) tetraacetic acid
PMSF	phenylmethylsulfonyl fluoride
DTT	dithiothreitol
cpm	count(s) per minute
V	volt(s)
SDS	sodium dodecyl sulfate
Fig.	figure
SD	standard deviation
mu	milliunit(s)
mg	milligram(s)
gal	ß-galactosidase
luc	luciferase

CHAPTER I

INTRODUCTION

1. Isozymes of Mammalian Hexokinase

Background and kinetic properties

Hexokinase (ATP: D-hexose 6 phosphotransferase, EC 2.7.1.1) catalyzes the phosphorylation of glucose, using Mg^{2+}ATP as phosphoryl donor. The product, glucose-6-phosphate, is a common substrate for phosphohexoisomerase, phosphoglucomutase, and Glc-6-P dehydrogenase, which introduce Glc-6-P into glycolysis, glycogen synthesis and the pentose monophosphate pathway, respectively.

Four distinct isozymes of hexokinase exist in mammalian tissues. They are designated as type I, II, III and IV isozymes, based on their order of electrophoretic mobility on starch gels (1). Alternatively, they can be separated by chromatographic technique, and named isozymes A-D according to their order of elution from DEAE-cellulose columns (2).

Although catalyzing the same reaction, the four isozymes can be distinguished from one another by their different molecular weights and kinetic properties. As summarized in Table 1, type I-III isozymes have molecular weights of about 100 kDa, low K_ms for glucose, in the submillimolar range, and are therefore often referred to as the "low K_m" isozymes. On the other hand, type IV isozyme (EC 2.7.1.2, commonly called glucokinase), has a molecular weight of about 50 kDa and a much higher K_m for glucose (4.5mM). All four isozymes have similar K_ms for ATP. The "low K_m" isozymes are sensitive to inhibition by their reaction product Glc-6-P at physiologically relevant concentrations, but

Table 1. Molecular weights and kinetic parameters of mammalian hexokinases

	HEXOKINASE ISOZYME			
PARAMETER	I	II	III	IV
M_r (kDa)	102.3	102.6	100.3	49
K_m Glc (mM)	0.04	0.13	0.02	4.5
K_m ATP (mM)	0.42	0.70	1.29	0.49
K_i Glc-6-P vs. ATP (mM)	0.026	0.021	0.074	15

This table was adapted from Ureta (3), and the references therein. Indicated molecular weights are based on amino acid sequences deduced from cloned cDNAs (4-9).

type IV is not (3).

cDNA sequences and evolution

In recent years, the cDNAs coding for all four isozymes have been cloned from rat (4-9), and the respective amino acid sequences have been deduced. Additionally, cDNAs coding for hexokinases and glucokinase from many other organisms have also been cloned. By comparing amino acid sequences deduced from cDNAs, rat hexokinase I, II and III are very similar. Table 2 shows the extensive sequence similarities of N- and C-terminal halves of type I, II and III hexokinases, to one another, and to glucokinase (9) and yeast hexokinase A (10).

Based on the fact that the molecular weight of mammalian hexokinase I-III is twice that of yeast hexokinase and glucokinase, it has been proposed by many researchers (3, 11-15) that mammalian 100 kDa hexokinases have evolved by gene duplication and fusion of an ancestral 50 kDa hexokinase which is similar to present-day yeast hexokinase and glucokinase. It has been suggested that one of the duplicated catalytic sites retained the catalytic function, while the other evolved to acquire a regulatory role.

This theory is strongly supported by the internal repetition of amino acid sequence between N- and C-terminal halves of 100 kDa hexokinases, as well as the sequence similarities among N- and C- terminal halves of hexokinases I-III, glucokinase and yeast hexokinase. This gene duplication and fusion theory is further supported by the same

Table 2. **Comparison of amino acid sequences of N- and C- terminal halves of rat Type I, II and III hexokinases, glucokinase, and yeast hexokinase A**

	NI (1-475)	NII (1-475)	NIII (1-488)	CI (476-918)	CII (476-917)	CIII (489-924)
NII	68 (14)					
NIII	39 (16)	44 (14)				
CI	46 (17)	54 (14)	38 (14)			
CII	49 (17)	55 (14)	41 (13)	76 (11)		
CIII	45 (15)	48 (15)	40 (14)	62 (11)	66 (9)	
IV	46 (18)	52 (15)	38 (15)	49 (15)	53 (14)	49 (15)
YHKA	27 (14)		22 (13)	27 (15)		27 (13)

a- This table was adapted from Schwab and Wilson (6), and Thelen and Wilson (7).
b- Abbreviations used: NI, N-terminal half of rat Type I isozyme; NII, N-terminal half of rat Type II isozyme; NIII, N-terminal half of rat Type III isozyme; CI, C-terminal half of rat Type I isozyme; CII, C-terminal half of rat Type II isozyme; CIII, C-terminal half of rat Type III isozyme; IV, rat Type IV hexokinase (glucokinase); YHKA, yeast hexokinase A.
c- Percentage of identical residues is shown without parenthesis; Percentage of conservative substitutions is shown in parenthesis.

intron-exon structure among hexokinases. The splicing sites and the exon sizes of type II (16) and type I (17) hexokinase genes repeat directly between the N- and C-terminal halves. The same splicing pattern is also observed in the gene for 50 kDa glucokinase (18), and thus suggested that an ancestral 50 kDa hexokinase gene similar to glucokinase underwent gene duplication and fusion to form the hexokinase I and II genes.

The gene duplication and fusion theory underwent some modification as more information became available. White and Wilson (19) were able to digest hexokinase I into a 52 kDa N-terminal fragment and a 48 kDa C-terminal domain. The N-terminal domain was selectively protected by Glc-6-P from denaturation in guanidine hydrochloride (19); the C-terminal half was protected by a glucose analog, N-acetylglucosamine (20). The isolated C-terminal half of the enzyme retained the catalytic activity (20). Thus, they concluded that the binding site for Glc-6-P resides in the N-terminal half of the intact enzyme and is separate from the catalytic site which is associated with the C-terminal half. Further study demonstrated, surprisingly, that the isolated C-terminal half of the enzyme is inhibited by Glc-6-P and that both halves of the enzyme possessed binding sites for the inhibitor Glc-6-P as well as the substrate glucose and ATP (19). This led to a modification of gene duplication and fusion theory such as that the ancestral 50 kDa hexokinase would have had both the glucose binding site and the Glc-6 P regulatory site before gene duplication and fusion occurred.

Direct measurement of ligand binding on the intact hexokinase I enzyme showed only one binding site for glucose (21, 22) and one for Glc-6-P (21, 23). Therefore, it was

suggested that the glucose site in the N-terminal half and the Glc-6-P site in the C-terminal half were masked in the intact enzyme molecule.

Studies using chimeric hexokinases consisting of the N-terminal half of one isozyme and the C-terminal half of another confirmed that catalytic activity is associated with the C-terminal halves and regulatory function is coupled to the N-terminal halves of the type I and type III isozymes (24, 25). However, in the type II isozyme, both halves possess catalytic activities and both halves are sensitive to Glc-6-P inhibition (26). Thus, type II hexokinase gene is suggested to be the immediate product of gene duplication and fusion, which subsequently evolved into type I and III, in which N- and C-terminal halves are functionally differentiated.

Tissue distribution

Each of the four isozymes has its distinct tissue distribution (reviewed in references 27, 28). Generally, more than one isozyme is found in most tissues. Type I isozyme is present in virtually all tissues examined to date, at relatively high levels in most tissues except for liver. In fact, in those tissues with a heavy reliance on blood-borne glucose, the type I isozyme is the predominant form. Brain is totally dependent on blood-borne glucose, through glycolysis, to provide energy and it contains exclusively the type I isozyme. It is also the case with erythrocytes. Therefore, type I isozyme has been referred to as the "basic" hexokinase and is suggested to play an important role in introducing glucose into glycolysis (27, 28) .

Type II isozyme, on the other hand, is the major form in insulin-sensitive tissues such as skeletal muscle, diaphragm, adipose tissue, and mammary gland (27, 28). There is an apparent relationship between the predominance of the type II hexokinase and insulin sensitivity of the tissue. It has been speculated that type II isozyme is responsible for directing glucose into energy storage forms such as lipids (in adipose or mammary tissue) or glycogen (in skeletal muscle, diaphragm) (27, 28).

Type III, the least studied isozyme, has not been found to be the predominant form in any tissue. The tissues which show the highest amount of activity attributable to type III hexokinase are liver, spleen, and lung (27, 29). Type IV, or glucokinase, is known to be present in the ß -cells of the pancreas and in liver (30-32). Because its K_m for glucose is in the range of normal blood glucose levels, fluctuation of blood glucose greatly regulates its activity. This in turn makes glucokinase a key enzyme to convert excess blood glucose into glycogen in liver and to serve as a "glucose sensor" governing the release of insulin in the pancreatic ß-islets (31).

Subcellular association

Intracellular distribution of hexokinases is not homogeneous. The association of hexokinases with particulate fractions of tissue homogenates has been well documented in a number of tissues (27, 28). A large portion of hexokinase I and II has been found to be associated with mitochondria in tissues such as brain, heart, diaphragm, skeletal muscle and mammary gland (27, 33, 34).

Type I hexokinase is believed to be bound to the outer mitochondrial membrane via both hydrophobic and electrostatic interactions between the N-terminal half of the enzyme and moieties on or in the membrane. The electrostatic interaction between the negative charge on the hexokinase surface and presumably mitochondrial membrane phospholipids is bridged by divalent cations such as Mg^{2+} (35). The hydrophobic interaction is dependent on a small hydrophobic N-terminal segment of type I isozyme which targets hexokinase I to mitochondria. The importance of this hydrophobic segment has been confirmed by several studies. First, cleavage of a 9 residue peptide from the N-terminus of type I hexokinase with chymotrypsin prevented this enzyme from binding to mitochondria (36). Second, this essential N-terminal hydrophobic region of the intact enzyme is inserted into the core of the lipid bilayer when hexokinase is bound to mitochondria (37). Recently, another approach was employed to show that the chimeric construct of the first 15 amino acid residues of hexokinase I coupled to reporter protein, chloramphenicol acetyltransferase (CAT), was able to bind to liver and hepatoma mitochondria, while the native CAT was not (38). Similar study with Green Fluorescence Protein (GFP) as reporter protein showed that N-terminal fragment of type I or II hexokinase was able to target the reporter protein to mitochondria (39).

The outer mitochondrial membrane protein to which type I hexokinase binds was originally isolated as the "hexokinase binding protein" (HBP) (40), but is now known to be identical with the pore forming protein (porin) (41, 42), through which molecules such as ADP and ATP flow. This association between hexokinase and porin gives the enzyme preferential access to mitochondrially generated ATP (43, 44). Together with the fact that

mitochondrially bound hexokinase has slightly greater affinity for ATP and considerably less sensitivity to G-6-P inhibition (45-47), the binding of hexokinase to mitochondria represents a mechanism for activation of the enzyme.

Hexokinase III was thought to be "soluble" and hence cytoplasmic in location (30), but recently it was found to be weakly associated with the nuclear periphery. This was demonstrated via confocal microscopy after staining the isozyme through the use of a monoclonal antibody (29).

It is generally accepted that hepatic type IV hexokinase is located in the cytosol (48). However, Miwa et al. showed both nuclear and cytoplasmic localization of glucokinase in liver (49) and translocation of glucokinase during fasting-refeeding (50) and postnatal development (51). Both the type III and type IV isozymes lack the hydrophobic N-terminal fragment critical for mitochondrial binding.

2. Regulation of Hexokinase Gene Expression and Promoters of Hexokinase Genes

Mammalian hexokinase is the key enzyme committing glucose into cellular metabolic pathways, and it is not surprising that this enzyme is under complex regulation. In short term, hexokinase activity can be regulated in response to altered metabolic status via product Glc-6-P inhibition (for type I-III), antagonism (for type I) or supplementation (for type II and III) of this inhibition by P_i, substrate glucose inhibition (for type III) and altered interaction with mitochondria (reviewed in 28). In the long term, levels of hexokinase protein change during development, in response to changes in hormone and

nutritional status, or under chronic alteration in metabolic status. In principle, these long-term changes in level of the enzyme could be the results of transcriptional regulation, posttranscriptional regulation (e.g., mRNA stability or translational rate) or posttranslational modification (e.g., phosphorylation or glycosylation) (28).

Type I hexokinase

Type I isozyme is ubiquitously expressed in mammalian tissues and appears to play a general role in mammalian glucose metabolism. Thus, type I hexokinase may be considered a "housekeeping enzyme". The type I isozyme is found at particularly high levels in brain (52), consistent with the importance of glycolytic metabolism of glucose for sustaining a highly active energy metabolism in this tissue (53). However, distribution of hexokinase activity shows marked variations in different brain regions (54) and in different layers of retina (55) and cerebellum (56). Subcellular fractionation suggested that major hexokinase activity is located in the nerve endings (57). The variation in hexokinase activity in different neural structural elements, measured histochemically, are correlated with the amount of the enzyme protein itself, measured by immunofluorescent procedure (58). This is consistent with the fact that no posttranslational modification has been found to affect the specific activity of hexokinase I (28). In situ hybridization with hexokinase I-specific oligonucleotide probe also demonstrated extensive neuronal distribution of hexokinase I mRNA with regional differences in the expression pattern (59).

Changes in hexokinase levels during development have been particularly well studied in brain. Rat brain hexokinase levels are relatively low prenatally, and increase several fold within the first three weeks postnatally to attain the adult levels (54, 60). Developmental increases in hexokinase activity in neural tissues are correlated with the amount of the protein found in those tissues, measured by immunofluorescent staining (61, 62). Griffin et al. examined the levels of mRNA for type I hexokinase (relative to that for phosphoglycerate kinase, PGK) in developing brain and other tissues of rat (63). They found that mRNA levels for type I hexokinase in brain were relatively high during all developmental stages, compared to those in other tissues. The relative level of hexokinase I mRNA increased to a maximum at about one week postnatally before declining to adult levels by about 3-4 weeks postnatally. However, close examination of the data revealed that the mRNA levels of PGK were not constant during development, so the absolute levels of HKI mRNA actually peaked at 2 to 4 weeks postnatally. Nevertheless, based on the lack of correlation between relative levels of mRNA and activity of hexokinase in brain and other tissues during development, it is proposed that both transcriptional and posttranscriptional regulation processes are involved in developmental stage- and tissue- specific regulation of hexokinase I (63).

Yokomori et al. reported that levels of mRNA for type I hexokinase were increased 2.5 fold in response to treatment of cultured rat thyroid FRTL5 cells with thyroid stimulating hormone (TSH) (64). This effect was due to an increase in the rate of gene transcription, shown in nuclear run-on transcriptional assays, but not due to the change in RNA stability. $(Bu)_2$cAMP and forskolin had a similar effect, suggesting that TSH stimulates hexokinase gene expression via the cAMP-dependent pathway (64). Thyroid hormone is known to have a major

influence on development of brain (65). Hypothyroidism delays the normally observed postnatal increase in hexokinase activity, whereas hyperthyroidism accelerates the increase (62).

Using quantitative immunofluorescence techniques, the levels of type I hexokinase (66, 67) in various regions of rat brain have been correlated with previously reported basal rates of glucose utilization in these regions (68, 69), except for several regions (referred to as group II) in which hexokinase content exceeded that expected from basal glucose utilization. It is suggested that group II regions may be adapted to sustain a large range of changes in glucose utilization rates. Hexokinase activity in various brain regions has been reported to be influenced by several physiological perturbations which cause persistently altered metabolic activity, including water deprivation (70), hypertension (71, 72), streptozotocin-induced diabetes (73), and surgically-induced heart failure (74).

The 5'-flanking region for the rat type I hexokinase gene has been studied in our laboratory (75, 76 and this thesis). My early work in cooperation with Dr. White was focused on isolation of the promoter region and identification of transcriptional start sites. A genomic clone containing sequence identical to the 5' region of the cDNA for rat type I hexokinase was isolated. A 5.4-kb EcoR I fragment from this clone, containing the matching sequence, was sequenced in its entirety (75). 5'-RACE, RT-PCR and RNase protection assays indicated that there are multiple transcriptional start sites clustered in three regions at positions approximately -460, -300, and -100 relative to the translational start codon (75). These regions lack classical TATA sequences and are located in a GC-rich segment, a "CpG island" (77, 78), approximately 1 kb in length. These characteristics are frequently associated with

"housekeeping genes". In this thesis, the promoter region for type I hexokinase is analyzed; important cis-elements and corresponding trans-factors regulating promoter activity are identified.

Type II hexokinase

Type II hexokinase is distributed in insulin-sensitive tissues such as skeletal muscle, diaphragm, adipose tissue, and mammary gland (27). It has long been known that the levels of type II hexokinase are regulated by insulin, with decreased type II hexokinase observed in insulin-sensitive tissues of diabetic animals (33, 79). Frank and Fromm reported that the rate of hexokinase II degradation increases by a factor of 3 in the skeletal muscle of diabetic rats as compared with that of normal animals (80). Furthermore, the relative rate of synthesis of hexokinase II is approximately 1.9 times higher in the normal than in the diabetic rat (81). Insulin treatment of diabetic animals restores the degradation and synthesis of hexokinase II to normal levels (80, 81). Printz et al. (16) reported that hexokinase II mRNA was decreased in adipose tissue from diabetic rats, but was restored by insulin treatment. Insulin also induced hexokinase II mRNA in adipose and skeletal muscle cell lines. In one of the skeletal muscle cell lines, the increase in mRNA is accounted for by a corresponding increase of gene transcription (16).

The activity of the type II hexokinase in muscle is regulated by contractile activity. Increases in HK activity have been demonstrated in exercising muscle (82) and muscles subjected to chronic, low-frequency stimulation (83, 84). Up to a 14-fold increase in total

hexokinase activity and the hexokinase II isoform was observed in rat fast-twitch muscle after 2 weeks of chronic, low-frequency stimulation. This increase in enzyme protein content was related to an approximately 30 fold increase in protein synthesis rate (85, 86). Cessation of stimulation resulted in a normalization of hexokinase activity associated with decreased rate of synthesis of type II hexokinase (86). The same stimulation also evoked an immediate increase in the ratio between structure (mitochondria)-bound and free hexokinase (87), and presumably represents an early response to increased energy demand. The observed transient increase in hexokinase II content represents an additional increase in glucose phosphorylation capacity under these stimulation conditions. After prolonged stimulation (3 weeks), hexokinase II activity declined, consistent with the previously observed switch from a carbohydrate-based to a fatty-acid-based energy metabolism (87). It was observed that under the same condition, hexokinase II mRNA was elevated significantly after one hour of stimulation and 30-fold after 12 hour, and the rate of HKII protein synthesis increased 20-fold after 24 hours (88). Another group also reported an increase in the levels of mRNA for type II hexokinase after a single brief period of exercise (89). Those experiments suggest that the increase in hexokinase II activity in response to contractile activity is at least partly at the transcriptional level.

The 5'-flanking region of hexokinase II has been isolated from a rat liver genomic library and analyzed in the rat skeletal muscle cell line, L6. The rate of hexokinase II gene transcription in L6 cells is increased by insulin, catecholamine, and cAMP, resulting in increased hexokinase II mRNA, protein synthesis, and glucose phosphorylation (16, 90, 91). The 5' untranslated region of hexokinase II mRNA is 462 bp long. The basal

promoter consists of about 160 base pairs of 5'-flanking sequence that includes a classical TATA box, an inverted CCAAT box (referred to as a Y box), a CCAAT box, and a cAMP response element (CRE). The CCAAT box and the CRE are both involved in cAMP responsiveness. The Y box contributes to basal promoter activity. Several known transcription factors bind to these sequences, notably CREB and ATF-1 to the CRE and NF-Y to both the Y and the CCAAT boxes (91).

Tumor cells exhibit increased glycolytic rates, and increased levels of key enzymes like hexokinase. The fast growing tumor cells have increased hexokinase activity (92) and higher percentage of hexokinase associated with mitochondria (93). It has been shown that various tumor cell lines have relatively high levels of mRNAs, particularly for the type II isozyme (7, 94, 95). The 4.3 kb proximal promoter region of the hexokinase II gene has been isolated and characterized in a rapidly growing hepatoma cell line, AS-30D (96). The DNA sequence of this promoter region is the same as that of the hexokinase II promoter in normal rat liver cells (91). However, in the AS-30D cells transfected with promoter-reporter construct, the promoter activity was enhanced by glucose, phorbol 12-myristate 13-acetate (a phorbol ester), insulin, cAMP, and glucagon, whereas these same agents produced little or no effect on promoter activity in transfected hepatocytes (96). The differences in the transcriptional regulation of HKII between normal cells and tumor cells suggested that transcription of the type II tumor gene may occur independent of metabolic state. The HKII promoter also contains two functional p53 response elements (96a). The highly abundant mutant form of p53, which lacks the ability to suppress or control cell cycle progression as wild type p53 does, activted the HKII promoter, thus providing a

linkage between the loss of cell cycle control and the high glycolytic rate in fast growing cancer cells (96a).

Type IV hexokinase (Glucokinase, GK)

Type IV hexokinase, or glucokinase, is expressed in the ß-cells of the pancreas and in liver, and is important in glucose metabolism and homeostasis (30-32).

In liver, glucokinase is involved in the utilization of excess circulatory glucose. The levels of glucokinase activity in rat liver vary with the nutritional status of the animal. Hepatic glucokinase activity falls during fasting and is restored by glucose refeeding (97, 98). The glucokinase mRNA is undetectable in liver from rats fasted for 24-72 h. Oral glucose administration causes a rapid, massive and transient accumulation of the hepatic glucokinase mRNA (9). The hepatic glucokinase is also up-regulated by insulin. Both enzyme protein and enzyme mRNA are absent from the livers of streptozotocin-induced diabetic rats. Insulin treatment causes a prompt transient build-up of mRNA and enzyme, resulting from a burst in the transcriptional activity of the glucokinase gene, as evidenced by run-on assays with isolated nuclei from liver (99) as well as from primary culture of rat hepatocytes (100). On the other hand, cyclic AMP exerts dominant negative control over glucokinase gene expression. Glucagon or derivatives of cAMP have suppressor effects on induction by insulin; this effect is primarily at the transcriptional level (100).

In contrast, the major function of glucokinase in ß-cells is to "sense" the circulating glucose level and to allow flux through glycolysis which controls the synthesis

and secretion of insulin. Levels of islet glucokinase mRNA and protein are relatively constant during the fasting-refeeding cycle (101). Hormones like insulin do not regulate glucokinase mRNA or protein in islet cells (92). The major determinant of glucokinase expression in islets is glucose (102). Transfection experiments suggested that glucose phosphorylation by glucokinase is the rate limiting step of glucose catabolism in ß-cells and the key step for activation of the insulin promoter by glucose. The glycolytic intermediates between fructose 1,6-diphosphate and phosphoenolpyruvate are essential for ß-cell glucose sensing (103).

Recent reports on the differences in glucokinase gene products in liver and pancreatic ß-cells provide a mechanism for tissue-specific regulation of this enzyme (101, 104; reviewed in 48, 105). Although glucokinase proteins in liver and pancreatic ß-cells display similar kinetic properties, they actually arise from alternative splicing of a single gene. The first exon encoding the 5' end of the hepatic mRNA is contiguous to the body of the structural gene. But the first exon for the 5' end of the insulinoma mRNA is more than 12 kb further upstream. The tissue specific splicing specifies not only the difference in 5' untranslated region of the islet and liver mRNA, but also their initial 15 amino acids (104). Also, the usage of alternative promoters presumably allows different regulation of glucokinase expression in two tissues, e.g., insulin regulation in liver and glucose regulation in ß-cells.

The hepatic glucokinase promoter, or downstream promoter, has been studied in primary culture of rat hepatocytes (106). The sequence between -123 to -34 (relative to the transcription start site) is the minimal promoter driving reporter expression in

hepatocytes, as well as in insulinoma and in hepatoma cells, which do not express the endogenous glucokinase gene. The fragment between -1003 to -707, however, is a hepatocyte-specific enhancer, which stimulates reporter expression in hepatocytes when linked to the SV40 promoter or the glucokinase promoter regardless of orientation or position. The same sequence is a silencer in hepatoma and is neutral in insolinoma cells (106).

The ß-cell glucokinase promoter, or upstream promoter, has no TATA box. Thus, transcription initiates over a region of 62 bases (104). Multiple cis-elements in the region between -280 to -1 (with the most proximal initiation sites designated as +1) contribute to transcription in insulinoma cells (107). The first element is three binding sites with a consensus sequence of CAT(T/C)A(C/G), designated as upstream promoter elements (UPEs). The factor binding to these sites is expressed preferentially in pancreatic islet ß-cells and is 50 kDa in size. The same factor also binds to similar elements, termed CT boxes, in the insulin promoter, suggesting a common control mechanism for pancreatic islet ß-cells specific gene expression (107). The second element is two copies of a pair of perfect palindromic repeats separated by a single base, TGGTCACCA, that have been termed Pal1 and Pal2 (107). A factor specific to neuroendocrine (NE) cell types, including the pancreatic ß-cell and pituitary corticotrope, binds to Pal elements, in addition to many other factors. The presence of the NE-specific factor in certain NE cell lines correlates with transcription of GK promoter-reporter constructs, suggesting a key role of this factor in determining NE-specific expression of GK (108).

Type III hexokinase

The regulation of expression of the gene for type III isozyme has not been studied.

3. Sp1 Family of Transcription Factors

Sp1

Sp1 (Specificity protein 1) was first identified as a factor required for the efficient transcription of the SV40 early promoter (109, reviewed in 110). In the SV40 promoter, Sp1 binds to proximal promoter elements, GC boxes, and contributes to the basal promoter activity. Alternatively, Sp1 can bind to a distal site (an enhancer) to activate gene expression. Moreover, when combined, the distal and proximal GC boxes act synergistically to give a strong Sp1 response through cooperative protein-protein interactions between Sp1 proteins, a phenomenon called superactivation (111). However, Sp1 molecules bound to adjacent sites are apparently unable to make such favorable protein-protein contacts. Sp1 is unable to bind simultaneously to adjacent two sites if the center-to-center distance between the two sites is less than 10 bp (110).

Sp1 contains three Cys-2His-2 zinc finger motifs at the carboxy-terminal end, serving as DNA binding domains. The activation domain is comprised of alternating serine/threonine-rich and glutamine-rich regions that constitute much of the amino-terminal two-thirds of the protein (112,113). The two domains can be functionally

uncoupled. The binding domain binds to DNA even when the activation domain is deleted (114). The activation domain, when attached to the transcriptionally inactive GAL-4 DNA binding domain, activates promoters containing GAL-4 binding sites. Even the fingerless Sp1, which lacks the DNA binding domain, could act with native Sp1 to reach superactivation, suggesting that superactivation is a result of protein-protein interaction between Sp1 factors (111). Sp1 also interacts and synergizes with other transcription factors, such as C/EBP beta, to activate gene expression (115).

Sp1 activates transcription by contacting components of the basal transcription machinery. For example, a glutamine-rich hydrophobic patch in Sp1 contacts the dTAFII110, a component of the Drosophila TFIID complex, and mediates transcription activation (116). A study with synthetic promoters containing TATA and/or Inr (initiator) showed that the Sp1 activation domain stimulates Inr-containing and TATA-containing core promoters equally well, while the VP16 activation domain activates the TATA-containing core promoter only (117). The lack of preference for TATA-containing core promoters might explain the frequent involvement of Sp1 in activation of TATA-less housekeeping genes.

Two posttranslational modifications, glycosylation and phosphorylation, regulate activity of Sp1 factor. A recent report showed that in cell culture under glucose deprivation, Sp1 protein becomes hypoglycosylated and more susceptible to proteasome degradation. This process could potentially reduce general transcription under conditions of inadequate nutrients (118).

Sp1 is a preferred substrate for a double-stranded DNA-dependent protein kinase. Infection of cells with SV40 virus results in a significant increase in the extent of Sp1 phosphorylation (119). A recent study showed the importance of Sp1 phosphorylation in the activation of HIV transcription, induced by okadaic acid (OKA), a selective inhibitor of the serine-threonine phosphatase (120). Another study showed that glucose-induced Sp1 dephosphorylation resulted in enhanced binding of Sp1 to promoter II of the acetyl-CoA carboxylase gene and transcriptional activation of this gene (121). The regulation of Sp1 by phosphorylation makes Sp1 a potential linkage between signal transduction pathways and transcriptional regulation of gene expression (110).

Sp3 and Sp4

Other members from the Sp1 family of transcription factors have been recently discovered. Sp2 and Sp3 have been cloned by screening a human HUT78 (αβ T cells) cDNA library using the Sp1 zinc finger domain as a probe (122). Sp4 and Sp3 (designated as SPR-1 and SPR-2 originally) have been cloned by screening an Ishikawa (a human endometrial cell line) cDNA expression library for proteins that bind to an Sp1 recognition site (123). Sp1, Sp3 and Sp4 are closely related members of a gene family encoding proteins with very similar structural features, including a zinc finger containing DNA binding domain as well as glutamine and serine/threonine-rich stretches. They bind to GC boxes and GT boxes with comparable specificity and affinity (123).

Sp1 is ubiquitously expressed, but expression levels vary considerably; how expression of Sp1 is regulated has not been determined (110). Sp3 is also ubiquitously expressed in various cell lines and organs, with the relative mRNA amount varying moderately. Sp4 transcripts, on the other hand, are abundant in brain and barely detectable in other organs (123).

In contrast to the structural similarities, Sp3 and Sp4 are functionally quite different from Sp1. Sp3 generally plays a repressive role in transcriptional regulation. It is suggested that Sp3 inhibits transcription by competing with Sp1 for binding sites in various viral and cellular promoters (124-126). Sp3 also acts as an activator on some promoters in some cell lines, depending on the presence of other factors in the same cells (127). A recent study showed that the glutamine-rich domain of Sp3 alone activates transcription; in intact Sp3 protein, an inhibitory domain can silence the glutamine-rich activation domain and completely suppress transcriptional activation (128). Sp4 is an activator like Sp1 (125), but is unable to act synergistically through multiple binding sites. However, Sp4 mediated activation can be enhanced in the presence of fingerless Sp1, suggesting the direct interaction between Sp1 and Sp4 (129).

Recently, other transcription factors are also identified as GT/GC box binding proteins. BTEB was isolated from rat liver by binding to a GC box found in the P-4501A1 gene promoter and was capable of activating other GC box-containing promoters (130). BTEB2 was isolated from human placenta by a similar procedure and had a similar activation function (131). These factors bind to the same DNA elements as the Sp family

of transcription factors, and potentially make the regulation through these cis-elements more complex.

Roles of Sp1 in transcriptional regulation

It is now known that Sp1 binds to a 9-bp consensus recognition sequence G/TG/AGGCG/TG/AG/AG/T in promoters of many viral and cellular genes; many of them are TATA-less housekeeping genes (110). The roles of Sp1 in transcriptional regulation are diverse.

Sp1 activates transcription of many viral genes, such as SV40 early gene (109). It mediates the activation of the hepatitis B virus pregenomic promoter by retinoblastoma susceptibility gene product (Rb) (132). In addition, Sp1 activates expression of many cellular genes by interacting with other transcription factors. For example, a distance-dependent cooperative interaction between transcription factors Sp1 and Oct-1 is critical for full activity of human U2 snRNA gene promoter (133).

Sp1 has been shown to be important in maintaining the basal, constitutive expression of many housekeeping genes, such as endothelial prostaglandin H synthase-1 gene (134). On the other hand, it is an essential element in directing tissue-specific expression of human CD14 in monocytes (135), as well as human insulin-like growth factor II in adult liver (136). It is also critical in start site selection for the TATA-less human Ha-ras promoter (137). Another recent study showed that a developmental

activation of an episomic hsp70 gene promoter in two-cell mouse embryos is mediated by Sp1 (138).

Roles of multiple Sp factors in transcriptional regulation

Since the discovery of other members of the Sp family of transcription factors, regulation of many promoters has been shown to involve multiple members bound to Sp1 binding sites. One example is the U5 repressive element of the long terminal repeat of human T cell leukemia virus type I, with a Sp1 binding core CACCC motif (139). Another report showed that expression of the SIS/PDGF-B gene in human osteosarcoma cells, U2-OS, requires both Sp1 and Sp3. Cotransfection of U2-OS cells with Sp expression plasmids and PDGF-B promoter/reporter constructs demonstrated that Sp1 and Sp3 can independently and additively activate the PDGF-B promoter (140). In a third study, transcription from the uteroglobin promoter was shown to be controlled by Sp1 and Sp3 through a non-classical Sp binding site. Gene transfer experiments into Drosophila SL2 cells that do not contain endogenous Sp factors revealed that expressed Sp1 activates the uteroglobin promoter, while Sp3 suppresses the activation by Sp1 (141). Finally, the pyruvate kinase M gene promoter is activated by expressed Sp1 in Drosophila SL2 cells. Sp3 has a synergistic effect on this Sp1 activation (142).

4. Thesis Overview

The goal of this thesis work was to identify the cis-elements and trans-factors governing the transcriptional regulation of the hexokinase I gene. This is the first step toward the understanding of transcriptional regulation of hexokinase I in different tissues, in response to hormone changes, such as that of cAMP, and ultimately, in response to changes in energy demand. Identification and characterization of cis-elements and trans-factors is the initial step that provides valuable information about how the promoter activity is maintained and regulated.

In this thesis, the PC12 cell line is used to study the promoter of hexokinase I gene. These cells are frequently used as a neuronal model (143) and, as with brain itself (144, 52), express a high level of the type I isozyme (145). Various techniques were employed, including transfection of PC12 cells with nested deletions of the 5'-flanking region of hexokinase I gene linked to a luciferase reporter gene, footprinting and gel-shift experiments, and mutagenesis made in important cis-element sequences. Most of the work in this thesis has been published in Archives of Biochemistry and Biophysics (75, 76).

CHAPTER II

MATERIALS AND METHODS

1. General methods

Standard methods (146) were used for routine procedures of molecular cloning. DNA sequencing was done by the method of Sanger (147) using Sequenase v.2.0 kits from U.S. Biochemicals. PCR methods were as described by Innis et al. (148). Oligonucleotides for PCR and electrophoretic mobility shift (gel shift) experiments were synthesized in the Macromolecular Structure Facility, Michigan State University. Spl and Ap2 consensus oligonucleotides were from Promega, with sequences of ATTCGATCGGGGCGGGGCGAGC and GATCGAACTGACCGCCCGCGGCCCGT, respectively. Protein was determined with the BCA Protein Assay reagent and bovine serum albumin standard (Pierce Chemical Co.). Human Sp1 expression vector was kindly provided by Dr. R. Tjian from Howard Hughes Medical Institute, Department of Molecular and Cell Biology, University of California, Berkeley; rat Sp3 and Sp4 expression vectors were by Dr. G. Suske from Institut für Molekularbiologie ünd Tumorforschung, Philipps-Universitat, Marburg, Germany.

2. Generation of promoter test constructs

Promoter test constructs were prepared in the vector pGL2-basic (Promega), which carries the coding sequence for firefly luciferase. Various fragments from previously described (75) genomic clone, pBS39.5, were inserted into pGL-2 basic, as described below:

The constructs pGL2 SS+ (inserted in correct orientation) and pGL2 SS- (inserted in reverse orientation) were obtained by subcloning a 572-bp Sac I fragment (positions -742 to -171, relative to translational start codon, ATG, where A= +1) into the Sac I site of pGL2-basic. pGL2 SM379 was constructed by subcloning SM fragment (position -742 to -516) into pGL2-basic. Another luciferase reporter construct, designated pGL2 HKII H2.6, contained 2.6 kb of sequence upstream from the transcriptional start site and included the promoter region of rat type II hexokinase; this was prepared by cloning a 2.9 kb Hind III fragment from the previously described genomic clone 5G3A (149) for Type II hexokinase into similar digested pGL2 basic.

The first set of luciferase reporter constructs was generated as follows. An Eco RI-Sac I fragment (-3366 to -171) from pBS39.5 was inserted into pGL2-basic to give the reporter construct designated pGL2 ES. A series of 5' deletions were then generated from pGL2 ES by conventional restriction digestion and religation. These constructs contained only the upstream transcriptional start sites (approximately -460 and -300) identified in the previous study (75).

A second set of constructs, with 3' end at -1 and thus including the downstream transcriptional start sites (75) at approximately -100, were generated from the set of constructs described above. A PCR fragment corresponding to sequence from an Mlu I site at -364 to position -1, with a Hind III site included in the downstream primer, was used to replace the Mlu I to Sac I region in the parent constructs.

All plasmids were purified using Qiagen plasmid kits, and checked for purity by agarose gel electrophoresis.

3. Transfection of PC12 and H9c2 cells with reporter constructs

Rat pheochromocytoma PC12 cells were cultured, as described by Tischler and Green (143), in RPMI medium (HyClone Laboratories) containing 10% horse serum (HyClone Laboratories), 5% Cosmic calf serum (HyClone Laboratories) and antibiotics (Sigma). Rat myoblast H9c2 cells (150) were grown in Dulbecco's modified Eagle's medium (DMEM, HyClone Laboratories) containing 10% Cosmic calf serum and antibiotics. Nearly confluent cultures of PC12 cells and H9c2 cells were replated at 1:3 dilution onto collagen coated and regular 6-well tissue culture plates, respectively. Twenty four hours later, at which time cells had attained 50-60% confluency, transfection was performed.

Initially, PC12 cells were transfected by incubation for 18 hours with 1.5 μg of test construct DNA plus 5 μg Lipofectin (Gibco-BRL), using the protocol described by the manufacturer. The medium was then exchanged for fresh medium and extracts were prepared after a further 48 h incubation. Cells were washed twice with phosphate-buffered saline (0.15 M NaCl, 0.015 M sodium phosphate, pH 7.4) and lysed in 200μl Lysis Buffer (Promega). The plates were placed at -80°C for 30 min, then at 23°C for 15 min before collecting the lysate. Lysates were centrifuged at 15,000g for 5 min at 4°C, and supernatants were either assayed immediately or frozen (-80°C) for later assay. In these experiments, luciferase activities were expressed on a per milligram protein basis.

In later experiments, cells were co-transfected with a control vector, pCMV ß-gal (Clontech), from which ß-galatosidase expression is driven by the CMV promoter and enhancer. Transfection conditions were as described above except that PC12 cells were

transfected with 8 μg Lipofectin, 1.5 μg test construct DNA, and 0.3 μg pCMV ß-gal DNA, while H9c2 cells were transfected with 10 μg Lipofectin, 2 μg test construct DNA, and 0.4 μg pCMV ß-gal DNA. Luciferase activities were then expressed relative to ß-galactosidase activities.

4. Reporter assay

Luciferase activities (arbitrary light units) in cell extract were assayed with the reagent kit and protocol from Promega, using a Turner Model 20 Luminometer. ß-galactosidase activities in the same extract were determined using the assay kit from Promega and following the manufacturer's instruction on microtiter plate format. Luciferase activities were then normalized to ß-galactosidase activities, and were expressed as percentage of the normalized luciferase activity seen with the pGL3-control vector (Promega) in which luciferase expression is driven by the SV 40 promoter. In mutant constructs, the activities were expressed as a percentage of that of the corresponding wild type construct.

5. Non-radioactive labeling of RNA probe and RNase protection assay

To synthesize an RNA probe corresponding to the 3' UTR region of HKI mRNA, a BamH I/ Pst I fragment (2855 to 3160) of HKI cDNA was subcloned into pBluescript II SK+ vector with same digestion. After digestion with BamH I, a 369 nt long anti-sense RNA probe was synthesized by MAXIscript In Vitro Transcription Kits (from Ambion), using T7 RNA

polymerase. This probe was labeled with dig-UTP (from BMB) in the reaction containing dig-UTP and normal UTP at the ratio of 1:5.6. Full length probe was excised and eluted after electrophoresis in 6% denaturing polyacrylamide gel.

RNase protection assays were done with RPA-II kits (from Ambion), following manufacturer's instruction, except that 10 units of RNase One (from Promega) was used to digest RNA at 37°C for 60 min. Protected probes (306 nt) were then separated in 6% denaturing gel, transferred onto Amersham's Hybond-N membrane, using Bio-Rad Trans-blot SD Semi-Dry Transfer Cell. Signals on the membrane were then detected with Genius System for Filter Hybridization (from BMB), following Chemiluminescent Detection protocol. The membrane was then exposed to X-ray film.

6. Preparation of nuclear extracts from PC12 cells

Preparation of nuclear extracts followed the procedure of Ausubel et al. (151). Specifically, PC12 cells from confluent cultures were collected by centrifugation, washed with PBS, and suspended in hypotonic buffer containing 10 mM Hepes, 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, and 0.5 mM DTT, pH 7.9. After incubation for 10 min on ice, the swollen cells were homogenized with a glass Dounce homogenizer and nuclei pelleted by centrifugation at 3300 x g for 15 min. Nuclei were resuspended in low salt buffer (20 mM Hepes, 25% (v/v) glycerol, 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, and 0.5 mM DTT, pH 7.9). Addition of an equal volume of "high salt" buffer, which had the same composition as above except with 1.2 M KCl instead of 20 mM KCl, resulted in

release of soluble protein from nuclei. After centrifugation at 25,000 x g for 30 min, the supernatant was dialyzed against 20 mM Hepes, 20% (v/v) glycerol, 100 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, and 0.5 mM DTT, pH 7.9. The nuclear extract was divided into aliquots and stored in liquid nitrogen.

7. Electrophoretic mobility shift ("gel shift") and "supershift" experiments

Nuclear extract (20μg protein) was incubated at room temperature for 30 min with 10-50 x 10^3 cpm of the ^{32}P-labeled DNA fragment, and 1.5 μg poly (dI-dC) as nonspecific competitor in a buffer containing 10mM Tris-HCl (pH7.5), 50mM NaCl, 2.5mM $MgCl_2$, 0.5mM DTT, 4% (v/v) glycerol and 0.05% (v/v) NP-40; the total volume was 15 μl. Samples were then applied to 5% native acrylamide gel and electrophoresis run at 150 V. For supershift experiments, 2 μl anti-Spl, anti-Sp3, or anti-Sp4 antibody (Santa Cruz Biotechnology) was added after the initial incubation, and incubation continued for another 30 min before loading onto the gel. Each anti-Sp antibody is specific and does not cross-react with other members in the Sp family. Gels were dried and exposed to Kodak Biomax MR film or analyzed on a phosphoimager (Molecular Dynamics Model 400B).

8. DNase I footprinting

DNase I footprinting was done with the SureTrack Footprinting Kit (Pharmacia), following the protocol provided by the manufacturer. A Sac I-Mlu I fragment (-742 to -364),

labeled with Klenow fragment (146) to a specific activity of approximately 200cpm/bp, was incubated at room temperature with indicated amounts of PC12 extract and poly(dI-dC) in the same buffer used for electrophoretic mobility shift assays (see above). After 30 min incubation, DNase I was added; after a further one min incubation, digestion was stopped by addition of the "DNase stop solution". Proteinase K (2 μg) was added, followed by incubation at 42° for 30 min. Samples were then extracted with phenol-chloroform, DNA collected by precipitation with ethanol, and samples loaded onto 5% sequencing gel. The same DNA fragment, treated with formic acid and piperidine to cleave at G and A residues, was loaded on this same gel to provide a sequencing ladder.

9. Mutation of promoter reporter constructs

Deletion or substitution mutations were made by the two step PCR method of Higuchi (148), using a Sac I fragment (-742 to -171) cloned in the pGL2-basic vector as template. Description of the exact nature of these modified constructs will become relevant only after presentation of results below. However, primers used for their generation are given in Table I. Outside primer JW86 (equivalent to the GLprimer1 sequencing primer from Promega) is located 10 nt upstream from the multiple cloning site in the pGL2-basic vector, while outside primer JW93 (equivalent to the GLprimer2 sequencing primer from Promega) corresponds to sequence just downstream from the multiple cloning site. Outside primer JW63 corresponds to hexokinase sequence at positions -361 to -332. Underlined regions in sequences for the inside primers correspond to hexokinase sequence located at positions appropriate for generating

specific mutations or deletions described below. The position of deleted sequence generated with primers JW115 and JW116 is indicated by double slashes within the primer sequences. Mutated sequences are shown in italics. All mutations and deletions were confirmed by direct sequencing of the entire region corresponding to the PCR fragment.

10. Hexokinase assay and western blot

PC12 cells were transfected with human Sp1 and rat Sp3, Sp4 expression plasmids, in which cDNA of Sp transcription factors is driven by CMV promoter. Transfection procedures were same as described for luciferase reporter constructs, except that 2μg of expression plasmid was used in each transfection. Cells were lysed 48 hours after transfection, in 200μl (per well) of ice-cold BTGE buffer (50 mM Bicine, 10 mM thioglycerol, 10mM Glc, 0.5 mM EDTA, 0.1% (v/v) TritonX-100, pH 8.2) and went through two freeze-thaw cycles. The lysate was then centrifuged at 800xg for 5 min at 4°C. Hexokinase activity in the supernatant was determined immediately, and then stored at -80°C for future use. Hexokinase activity was determined spectrophotometrically as described previously (152). Protein was determined with the BCA Protein Assay system. Samples were pretreated with excess iodoacetamide (153) to avoid interference by thioglycerol present in the PTGE buffer.

For western blot experiments, 100 μg extract was loaded onto 8% SDS acrylamide gel, then transferred onto nitrocellulose membrane. The Sp transcription factors were detected by specific anti-Sp antibodies (at 1:3000 dilution, Santa Cruz Biotechnology),

which do not cross-react with other members of Sp family. The secondary antibody was goat anti-rabbit IgG conjugated to horseradish peroxidase (from Bio-Rad), and detected with SuperSignal Chemiluminescent Substrate (from Pierce). The membrane was then either exposed to X-ray film or quantitatively analyzed on Molecular Imager System GS-505 (Bio-Rad). Protein contents in the samples were assessed by densitometric analysis of Coomassie Blue stained gels and used to adjust expression level calculated from western blot.

CHAPTER III

RESULTS

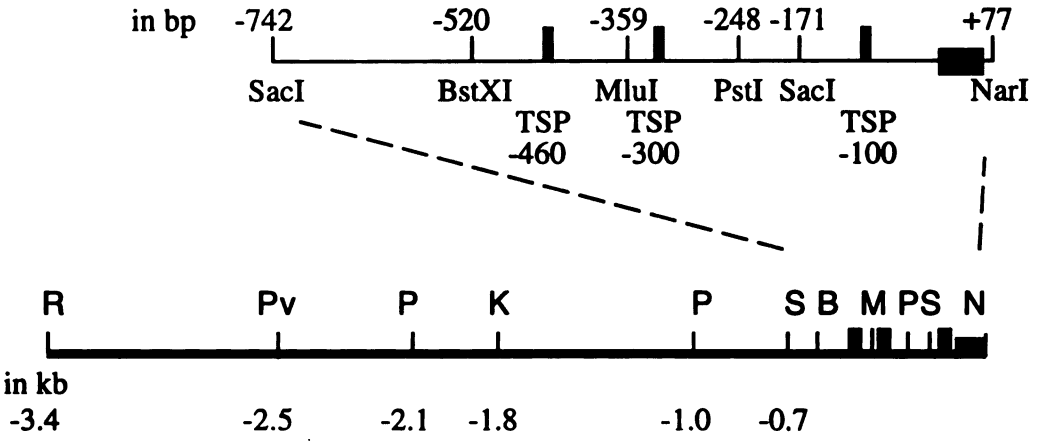
1. Promoter activity associated with Sac I fragment in HKI upstream region

Figure 1 shows a schematic representation of a partial restriction map of hexokinase I upstream region (-3366 to +77), which is related to this work. As previously identified, transcriptional start sites are clustered in three regions approximately located at -460, -300 and -100, with A in start codon ATG designated as +1 (75). Complete DNA sequence and restriction maps of the same region are shown in Appendices A-D.

A 572-bp Sac I fragment (nt -742 to -171) was inserted into the multiple cloning site upstream from the luciferase coding region of the reporter vector pGL-2 basic; plasmids with the insert in either the correct (pSS+) or reverse (pSS-) orientation were prepared. The pSS+ plasmid contains transcriptional start sites located near the -460 and -300 positions. Also constructed was pSM, containing only the upstream transcriptional start site (nt -737 to -359).

The various reporter constructs or the unmodified pGL-2 basic vector were used to transfect PC12 cells. These cells express a high level of the type I isozyme (145). Thus we anticipated that the promoter for type I hexokinase would be quite effective in PC12 cells and indeed, luciferase expression driven by the pSM and pSS+ constructs was well above that seen with the promoter-less pGL-2 basic vector (Table 3). Reversing the orientation of the inserted promoter region (pSS-) resulted in a marked reduction in expression of luciferase activity. The previous experiment (75) indicated that PC12 cells favored the upstream (-460) start sites; hence, deletion of the downstream (-300) transcriptional start site, in construct pSM, might have been expected to have only a marginal effect on expression. Thus, the substantial decrease (relative to pSS+) in expression from pSM, was somewhat surprising.

Figure 1. Partial restriction map and schematic organization of HKI promoter region. Genomic DNA of the hexokinase I gene from the Eco RI site (-3366) to the Nar I site (+77) is shown as a horizontal line. The positions of the restriction sites related to this work are represented by the vertical lines. The three vertical bars represent transcriptional start points (TSP) clustered at -460, -300 and -100. The horizontal bar represents the first exon (+1 to +66). Refer to Appendix B, C and D for complete DNA sequence and restriction map.



R = EcoR I
Pv = Pvu II
P = Pst I
K = Kpn I
S = Sac I
B = Bst XI
M = Mlu I
N = Nar I

Fig. 1

Table 3. Expression of luciferase in PC12 cells transfected with reporter constructs

Construct	Sequence in HKI upstream region	Luciferase activity[a]	Fold increase
pGL2-basic	---	4.3 ± 1.2	1
pGL2 SS+	-742 to -171	1100 ± 96	260
pGL2 SS-	-742 to -171 (inverse)	56 ± 23	13
pGL2 SM	-737 to -359	300 ± 150	70
pGL2 HKII H2.6	---	50 ± 16	12

[a] Luciferase activity expressed as units/mg protein (mean ± SD for four transfections).

Nonetheless, expression of luciferase activity remained 70-fold increased over the basal level. The type II isozyme of hexokinase is expressed at low level in PC12 cells (J.E. Wilson, unpublished observation). In agreement with this, low level of luciferase were expressed in PC12 cells transfected with a reporter construct (pHKII H2.6) containing the promoter and transcriptional start site for type II hexokinase.

2. Sequence between -742 and -516 is important for promoter activity

Results of experiments with reporter constructs containing Type I hexokinase genomic sequence between -3366 and -171 are shown in Fig. 2. Deletions of upstream sequence between -3366 and the *Sac* I site at -742 had no significant effect on expression of luciferase activity in PC12 cells (Fig. 2B). However, deletion of sequence between -742 and the *Bst* XI site at -516 resulted in 85% loss of promoter strength. Another significant decrease occurred after deletion of sequence between the *Mlu* I site at -364 and the *Pst* I site at -245.

Very similar results were seen with these reporter constructs transfected into H9c2 cells (Fig. 2C). The one qualitative difference between results with PC12 cells and those with H9c2 cells (Fig. 2B, C) was in expression driven by the PvS construct. In H9c2 cells, there was a modest but significant decrease in expression relative to that seen with longer and shorter constructs, ES and KS, respectively. This suggests the possible existence of positive and negative regulatory elements in sequence from -3366 to -2511 and from -2511 to -1833, respectively. If these indeed exist, they do not appear to be generally functional since their effect is not seen with PC12 cells. This observation has not been pursued further at this time.

Figure 2. Transfection of PC12 and H9c2 cells with promoter-reporter constructs. A, schematic representation of reporter constructs. All constructs included only sequence upstream from position -171 in the 5' flanking region of the gene encoding rat Type I hexokinase (4); see text for more detailed comments. B and C, normalized (to ß-galactosidase activity) luciferase activities after transfection of PC12 (B) or H9c2 (C) cells with indicated construct or basal control vector, pGL2-basic; results are the mean ± SD for at least eight independent transfections, and are given as percentage of normalized luciferase activity expressed from the pGL3-control plasmid. An asterisk means that the activity of a construct is significantly, according to statistic analysis, different from that of the next longer construct.

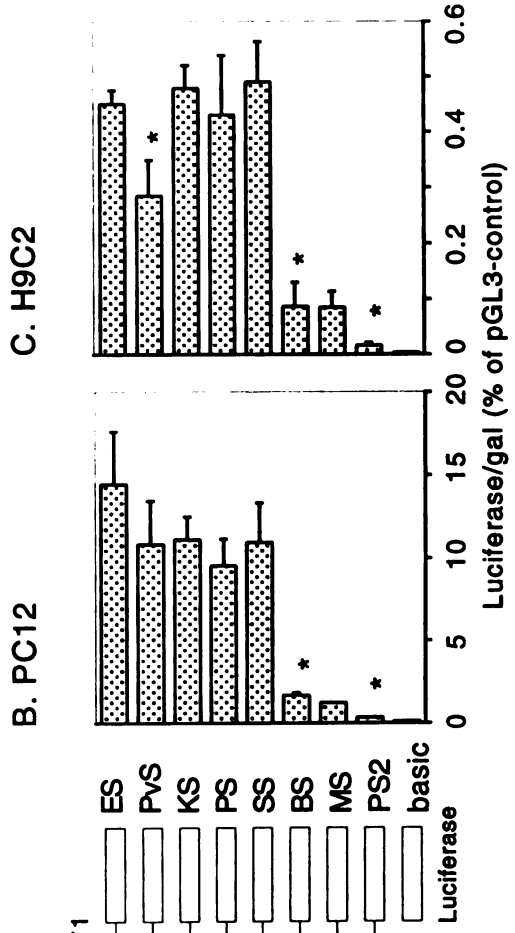

A. Constructs in pGL2-basic
-171
-3366
-2511
-1833
-1044
-742
-516
-364
-245
ES
PvS
KS
PS
SS
BS
MS
PS2
basic
Luciferase
B. PC12
C. H9C2
0
5
10
15
20
0
0.2
0.4
0.6
Luciferase/gal (% of pGL3-control)

Fig. 2

As noted previously (75), the overall expression was much lower in H9c2 cells than in PC12 cells. Thus, expression of luciferase from the longer constructs in PC12 cells was 10-15% of the pGL3-control, while in H9c2 cells, these same constructs were expressed at 0.4-0.5% of the pGL3-control, a level of more than 20 fold lower than that in PC12 cells.

The promoter activities are consistent with the observed levels of HKI mRNA by RNase protection assays (Fig. 3). The level of HKI mRNA seen with 10μg PC12 total RNA (lane 5) was comparable to that seen with 10μg rat brain total RNA (lane 4). The level of HKI mRNA in 100μg H9c2 total RNA (lane 7), on the other hand, was a little lower than those in 10μg total RNA from either PC12 cells or rat brain, indicating that the level of HKI mRNA in H9c2 is at least 10 fold lower than that in PC12 cells and rat brain.

The major point emerging from these transfection experiments was the importance of sequence from -742 to -516 for promoter activity. Thus, attention was focused on this segment.

3. DNase I footprinting reveals two protected regions

DNase I footprinting analysis was done using a *Sac* I-*Mlu* I fragment containing sequence between -742 and -359. Using nuclear extracts from PC12 cells, two protected regions were observed (Fig. 4). These were designated as the P1 box (24 nt in length, positions -552 to -529) and the P2 box (23 nt in length, positions -480 to -458). Densitometric analysis revealed that, with the maximum amount of nuclear extract used, densities in the P1 and P2 regions were reduced to approximately 10% and 25%, respectively, of that in the unprotected

Figure 3. RNase Protection Assay. The same amount of the dig-labeled anti-sense RNA probe corresponding to BamH I/ Pst I fragment (2855 to 3160) in the 3'-untranslated region of HKI cDNA, was incubated with total RNA from yeast (lanes 2,3) or various tissue or cell lines (lanes 4-7), then subjected to RNase I digestion (lanes 3-7). Lane 1 is RNA marker. In lane 2, one tenth of the sample was loaded onto the gel to give a signal comparable to that in lanes 3 to 7, where all the samples were loaded onto each lane.

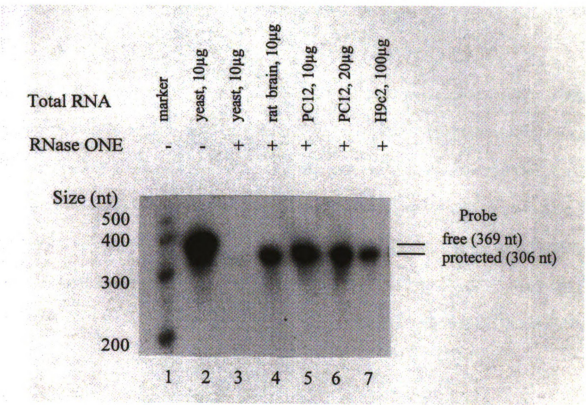

Total RNA
marker
yeast, 10µg
yeast, 10µg
rat brain, 10µg
PC12, 10µg
PC12, 20µg
H9c2, 100µg
RNase ONE
- - + + + + +
Size (nt)
500
400
300
200
1 2 3 4 5 6 7
Probe
free (369 nt)
protected (306 nt)

Fig. 3

Figure 4. DNase I footprinting analysis with a nuclear extract from PC12 cells. Protected regions in a ^{32}P-labeled *Sac* I-*Mlu* I fragment (positions -742 to -359), designated as the P1 and P2 boxes, are noted. The location of an additional Sp1 site upstream (position -570) from the P1 box (see text) is also indicated, although no protection was noted in this region. Lane 1 is a G+A chemical sequencing reaction of the same labeled fragment to provide a sequence ladder. The sequence of the P1 box is CTTTTTCCACGCCCACTTGCGTGC, and the P2 box is GAGGAAGGGGTGTGGCCCCGTTC.

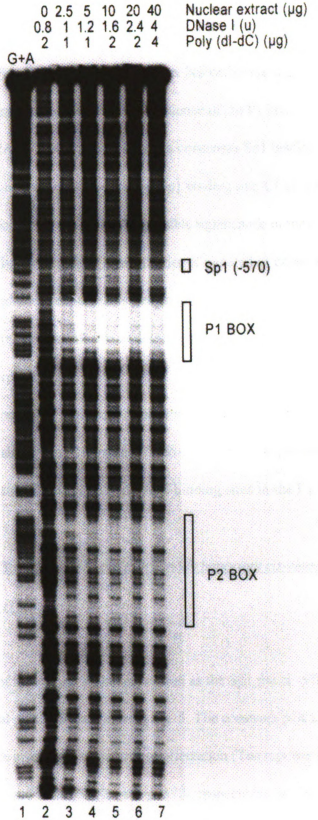

0 2.5 5 10 20 40
0.8 1 1.2 1.6 2.4 4
2 1 1 2 2 4
Nuclear extract (µg)
DNase I (u)
Poly (dI-dC) (µg)
G+A
Sp1 (-570)
P1 BOX
P2 BOX
1 2 3 4 5 6 7

Fig. 4

control (no nuclear extract). The P2 box is more weakly protected, includes the upstream transcriptional start sites identified in previous work (75), and does *not* appear to be important for promoter activity (Fig. 2). In contrast, the P1 box lies within the region identified as being of major importance for promoter activity. The sequence of the P1 box, CTTTTTCCACGCCCACTTGCGTGC, includes a consensus Sp1 binding site (154) in its central region. We also noticed another potential Sp1 binding site, CCGCCCA, just upstream from the P1 box (positions -574 to -568). The possible significance of these elements for promoter activity was further examined using a series of mutated or deletion constructs (results below). Although these experiments demonstrated the functional importance of the upstream (-570) Sp1 site, we were unable to detect significant protection of this region in DNase I footprinting experiments.

Compared to consensus sequences of binding sites for known transcription factors by searching the GCG database, the P1 box and P2 box contain some potential cis-elements shown in Table 4, notably GC-rich Sp1 and AP2 binding sites in the P1 box.

4. The P1 box and the Sp1 site at -570 are functionally important cis-elements for promoter activity but the P2 box is not

The locations of the P1 and P2 boxes, as well as the Sp1 site at -570, within the sequence of the parental pGL2SS are shown in Fig. 5. The construct pGL2SS- was identical to pGL2SS except that the insert was in the reverse orientation (The reporter constructs pGL2SS and pGL2SS- were referred to as pS572+ and pS572-, respectively, in Table 3). Several

Table 4. Potential cis-elements in P1 box and P2 box sequences

P1 BOX CTTTTTCCACGCCCACTTGCGTGC

cis-Elements	Consensus Sequence	Matching Sequence[a] in P1 box	Degree of Matching
XRE	TTGCGTG	CTTTTTCCACGCCCAC**TTGCGTG**C	7/7
Sp1	C/AC/TC/TC/AGCCC/TC/A	CTTTTT**CC*A*CGCCCA**CTTGCGTGC	8/9
Myc	CACGTG	CTTTTTC**CACG*CC*CAC*T*TG**CGTGC	4/6, 5/6
AP2	CCCA/CNG/CG/CG/C	CTTTTT**CC*A*CGCCC**ACTTGCGTGC	7/8

P2 BOX GAGGAAGGGGTGTGGCCCCGTTC

cis-Elements	Consensus Sequence	Matching Sequence[a] in P2 box	Degree of Matching
B-globin	GGGTGTGGC	GAGGAAG**GGGTGTGGC**CCCGTTC	9/9
PEA3-RS	AGGAAG	G**AGGAAG**GGGTGTGGCCCCGTTC	6/6
TEF-2-GT-I	GGGTGTGG	GAGGAAG**GGGTGTGG**CCCCGTTC	8/8
Pu box	GAGGAA	**GAGGAA**GGGGTGTGGCCCCGTTC	6/6

[a] Bold letters in underlined sequence represent nucleotides identical to consensus sequence of known cis-element; italic letters represent mismatching nucleotides.

Table 5. Primers used for generation of substitution or deletion mutants

Outside Primers		Sequence
	JW86	TGTATCTTATGGTACTGTAACTG
	JW63	CGTCCCCCACGTAAGGAGCGCACGGTCACG
	JW93	CTTTATGTTTTTGGCGTCTTCCA

Inside Primers

Mutant Construct	Primer	Sequence
SSdelP1(24)	JW116	5' GGCTTCCGCT//TTCTCCAACAGTGTGGATG
	JW115	3' GTCAAAGCCGAAGGCGA//AAGAGGTTGT
SSsubP1	JW128	5' *GGAACATAAACAGGTATGTA*TTCTCCAACAGTGTG
	JW129	3' AAAGCCGAAGGCGA*TCCCCCTTGTATTTGTCCAT*
SS(NdeI)	JW153	5' TGAA*CATATG*AGTTTCGGCTTCCGCTCTTT
	JW154	3' TGGGTCACGCTCTAGTCACTT*GTATAC*TCAAAG
SSsubP2	JW157	5' *CCTGTCGTGTTACCATGG*CGTGTTCTCCAGTTTGTG
	JW158	3' GCTCGGGATTAGAGG*GGAAGGACAGCACAATGGTAC*

Note: In inside primers, underlined sequences will anneal to template in PCR reaction. Italic letters represent mutated sequences introduced into PCR products. Double slashes represent the sequence being deleted.

modified constructs were prepared by the two step PCR method of Higuchi (148), using primers shown in Table 5. These were: pGLSSdelP1(24), in which a 24 nt sequence corresponding to the P1 box was deleted; pGLSSsubP1, in which the entire 24 nt sequence was modified, converting C to A, T to G, A to C, and G to T; pGL2SS(Nde I), in which the -570 Sp1 site CCGCCC was substituted with an *Nde* I restriction site, CATATG; pGL2SSsubP2, in which the P2 box (-480 to -458), GAGGAAGGGGTGTGGCCCCGTTC, was replaced by the permuted sequence, GTTTCCTGTCGTGTTACCATGGC. Additional constructs were then generated from those just described: pGL2NdeS was prepared by digestion of pGLSS(Nde I) with *Kpn* I (cutting upstream within the multiple cloning site) and *Nde* I, and religation; pGL2SSdelP1(55), in which a 55 nt segment (-572 to -518) containing the P1 box and the upstream Sp1 site was deleted, was obtained by digestion of pGL2SS(Nde I) with *Nde* I and *Bst* XI, followed by polishing the ends and religation; pGL2BSsubP2 was generated by deleting sequence between -742 and -561 as a result of digesting pGLSSsubP2 with *Kpn* I and *Bst* XI, followed by religation.

Results of transfection experiments with several of these constructs are compared in Fig. 5. Deletion of the P1 box and upstream Sp1 site (in pGL2delP1(55)) resulted in expression of normalized luciferase activity that was only about 16% of that seen with the parental pGL2SS. This reduced activity was virtually identical to that seen with pGL2SS-, in which the promoter region had been inserted in reverse orientation. Similar reduction in promoter activity was seen with pGL2SSsubP1, in which the sequence of the P1 box had been permuted as described above. Thus, the P1 box was clearly a critical *cis*-element for promoter activity. The adjacent Sp1 site, at -570, also had a significant effect on promoter activity, with either

Figure 5. Effect of mutations in the P1 box and upstream (-570) Sp1 site on promoter activity. Normalized luciferase expression from the parental pGL2-SS (top of figure), containing sequence between -742 and -171, was taken as 100%. pGL2-SS- (bottom of figure) corresponds to the same sequence, but inserted in reverse orientation. Other constructs contain mutations or deletions as described in the text. Filled boxes represent wild type sequence while open boxes represent mutated sequence; bent lines represent deletions. The positions of the P1 box (-552 to -529) and upstream (-570) Sp1 site are indicated. The angled arrow indicates the location of P2 box (-480 to -458), which encompasses the multiple transcriptional start sites identified in the -460 region (4). Results given are mean ± SD for eight transfections.

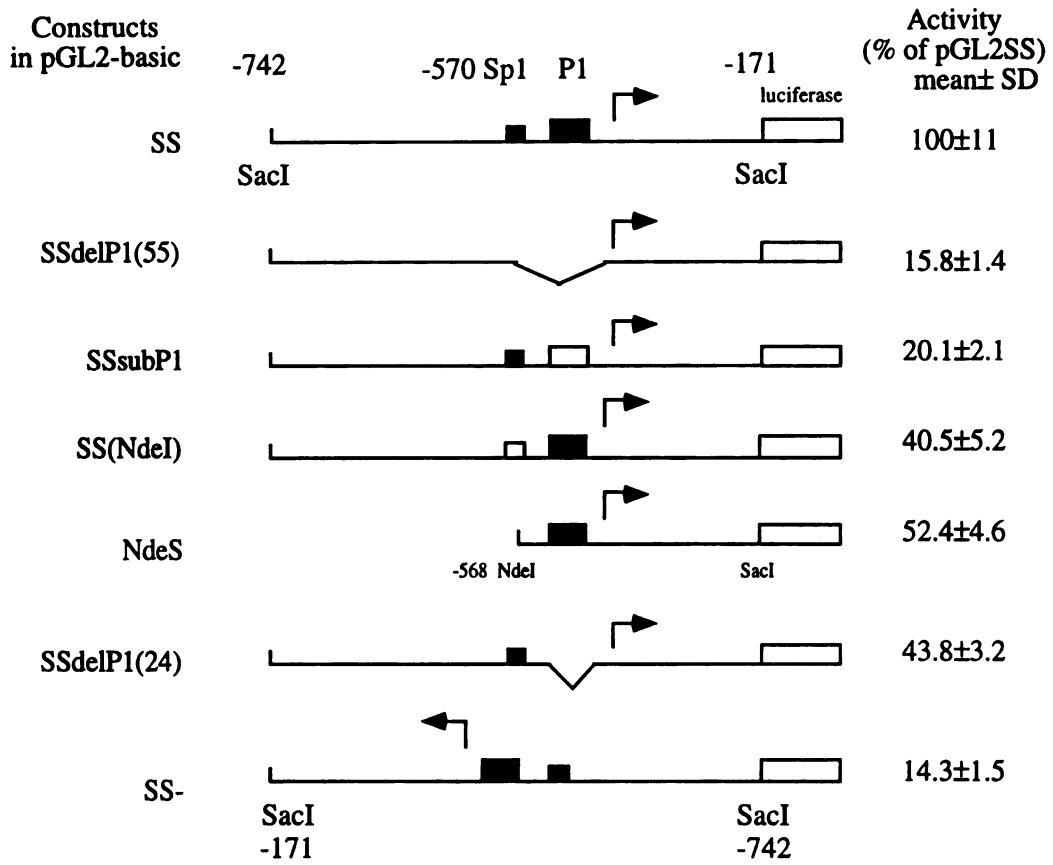

Constructs in pGL2-basic
Activity (% of pGL2SS) mean± SD
-742
-570 Sp1
P1
-171
luciferase
SS
SacI
SacI
100±11
SSdelP1(55)
15.8±1.4
SSsubP1
20.1±2.1
SS(NdeI)
40.5±5.2
NdeS
-568 NdeI
SacI
52.4±4.6
SSdelP1(24)
43.8±3.2
SS-
SacI
-171
SacI
-742
14.3±1.5

Fig. 5

Table 6. P2 box mutation data

P2 BOX sequence [b]	GAGGAAGGGGTGTGGCCCCGTTC
P2 BOX mutant sequence	GTTTCCTGTCGTGTTACCATGGC

Construct	Wild type construct	Activity [a]
pGL2 SSsubP2	pGL2 SS	142 ± 14
pGL2 BSsubP2	pGL2 BS	108 ± 16

[a] Activities were expressed as percentage of the activities of corresponding wild type constructs. Data were mean ± SD of six transfections.

[b] Underlined G and T in P2 box sequence represents the most upstream transcriptional start sites for HKI gene.

mutation of the sequence in pGL2SS(Nde I) or deletion of this site in pGL2NdeS resulting in 50-60% decrease in expressed luciferase activity. Interestingly, when the P1 box was deleted in pGL2SSdelP1(24), bringing the upstream -570 Sp1 site into the position previously occupied by the P1 box, expressed luciferase was about 2-fold higher than that seen with the mutated P1 box of pGLSSsubP1 and was similar to that seen with the mutated Sp1 site of pGLSS(NdeI), suggesting that P1 box can be functionally replaces by the Sp1 site.

In contrast to the importance of sequence within the P1 box for promoter activity, extensive modification of the sequence in the P2 box did not have any detrimental effect on promoter activity (Table 6). Thus, normalized luciferase activity expressed from pGL2SSsubP2 was 142 ± 14% (mean ± SD for 8 transfections) of that seen with the parental pGL2SS construct. Similarly, luciferase expression from pGL2BSsubP2 was 108 ± 16% of that seen with the corresponding construct, pGL2BS, containing the wild type P2 box sequence.

5. Gel shift and supershift experiments demonstrate binding of Sp family members to the P1 box.

Gel shift experiments were done with a 33 bp oligonucleotide having the sequence GAT- [P1 box]-TAGATC, where "[P1 box]" represents the 24 nt sequence (positions -552 to -529) given above, annealed with its complement. Sequences for other DNA fragments in gel shift experiments are shown in Table 7. After incubation of the ^{32}P-labeled oligonucleotide with nuclear extracts from PC12 cells, two major complexes (Fig. 6), referred to here as the "upper"

Table 7. DNA Fragments Used in Gel shift Experiments

Name	Primers	DNA sequence
P1 BOX SEQUENCE		CTTTTTCCACGCCCACTTGCGTGC
Probe	127/130	GATCTTTTTCCACGCCCACTTGCGTGCTA AAAAAGGTGCGGGTGAACGCACGATCTAG
P1 BOX	113/114	TCTTTTTCCACGCCCACTTGCGTGCT AGAAAAAGGTGCGGGTGAACGCACGA
P1 #1	182/183	ATAGGGGGCCACGCCCACTTGCGTGC TCCCCCGGTGCGGGTGAACGCACGTA
P1 #2	184/185	ATCTTTTTAACATACCACTTGCGTGC GAAAAATTGTATGGTGAACGCACGTA
P1 #3	186/187	ATCTTTTTCCACGCAACAGGGCGTGC GAAAAAGGTGCGTTGTCCCGCACGTA
P1 #4	188/189	ATCTTTTTCCACGCCCACTTTATGTA GAAAAAGGTGCGGGTGAAATACATTA
P1 (7-9)	195/196	CTTTTTAACCGCCCACTTGCGTGC GAAAAATTGGCGGGTGAACGCACG
P1 (10-11)	201/202	CTTTTTCCAATCCCACTTGCGTGC GAAAAAGGTTAGGGTGAACGCACG
P1 (12-13)	191/192	CTTTTTCCACGAACACTTGCGTGC GAAAAAGGTGCTTGTGAACGCACG
P1 (14-15)	203/204	CTTTTTCCACGCCACCTTGCGTGC GAAAAAGGTGCGGTGGAACGCACG
P1 (16-18)	197/198	CTTTTTCCACGCCCAAGGGCGTGC GAAAAAGGTGCGGGTTCCCGCACG

Note: Underlined sequences represent mutations in the P1 box.

Figure 6. Gel shift experiment showing complexes formed by proteins in nuclear extracts from PC12 cells and an oligonucleotide including the P1 box sequence. In Lane 1, free oligonucleotide and no nuclear extract were added. In Lane 2, nuclear extract from PC12 cells and no competing oligonucleotide were present. In Lanes 3 and 4, same as in Lane 2, but the indicated molar excess of unlabeled P1 box oligonucleotide was added as competitor. In Lanes 5-12, same as in Lane 2, but a molar excess of unlabeled mutant oligonucleotides containing the indicated modifications within the P1 box was used as competitor; the 24 nt P1 box sequence was changed progressively, 6 nt at a time, as indicated schematically at top of figure (see text for details). Lanes 13 and 14 show the lack of competition by added molar excess of AP2 consensus oligonucleotide. Lanes 15 and 16 show the competition by added molar excess of Sp1 consensus oligonucleotide.

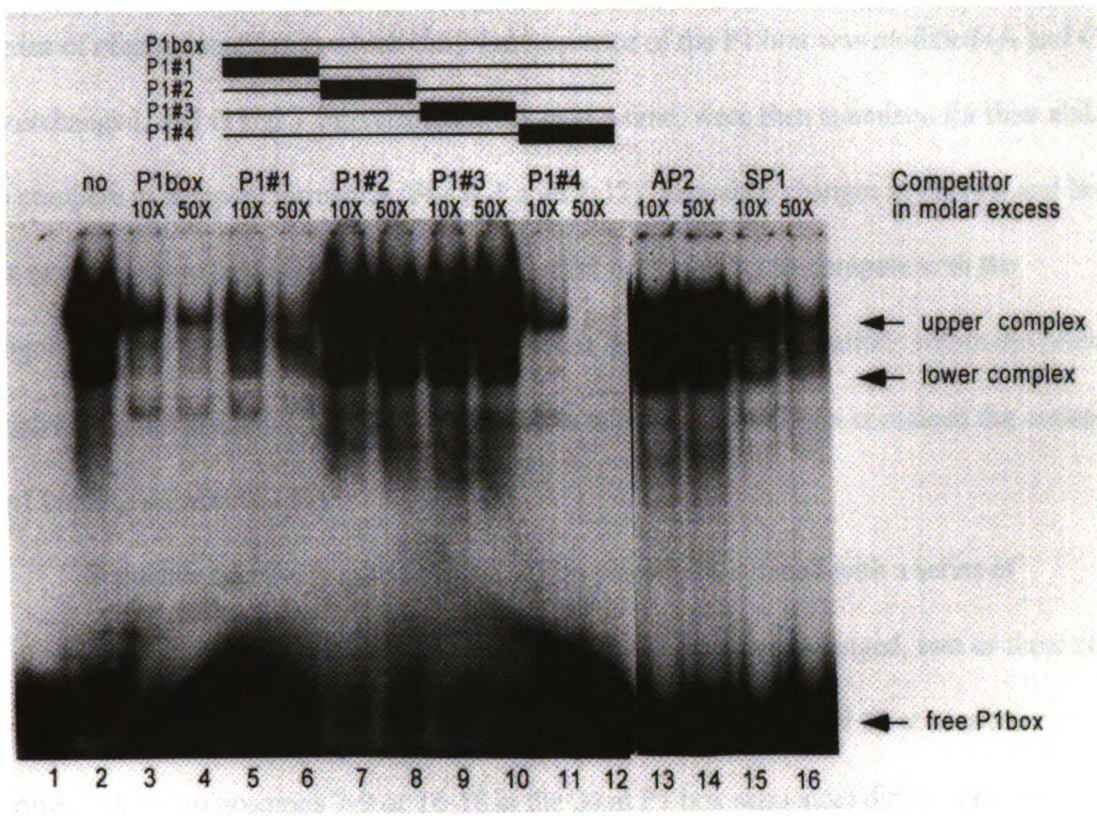

P1box
P1#1
P1#2
P1#3
P1#4
no
P1box
10X 50X
P1#1
10X 50X
P1#2
10X 50X
P1#3
10X 50X
P1#4
10X 50X
AP2
10X 50X
SP1
10X 50X
Competitor
in molar excess
upper complex
lower complex
free P1box
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Fig. 6

and "lower" complexes, were observed. As expected, the unlabeled oligonucleotide was competitive with the labeled nucleotide (Fig. 6, Lanes 3 and 4) for formation of both complexes. An oligonucleotide containing a consensus Sp1 binding sequence was also competitive with the ^{32}P-labeled P1 box oligonucleotide, but an oligonucleotide containing the somewhat similar (GC-rich) consensus AP2 binding sequence was not (Fig. 6, Lanes 13-16). A series of oligonucleotides in which the 24 nt sequence of the P1 box was modified (A and C interchanged, and G and T interchanged), six nt at a time, were then examined for their ability to compete in complex formation (Fig. 6, Lanes 5-12). Sequence changes in the first and last six nt of this 24 nt sequence had little, if any, effect on the ability to compete with the oligonucleotide having the intact P1 box sequence. However, modifications within the central twelve nt in this sequence abolished the ability to compete. This region contained the consensus Sp1 binding sequence (154).

Sequence specificity in this central region was then examined with a series of oligonucleotides in which the central twelve nt of the P1 box were changed, two or three at a time (Fig. 7). Modification of the first or last three nt within this central 12 nt sequence (corresponding to positions 7-9 or 16-18 in the 24 nt P1 box sequence) did not prevent competition for complex formation (Fig. 7, Lanes 5 and 6, Lanes 13 and 14) *but*, at comparable concentrations, the modified oligonucleotides were much less effective than the intact P1 box sequence (Lanes 3 and 4). Changes within the central six nt, which represents the core of the consensus Sp1 binding site (154), virtually abolished the ability of the oligonucleotides to compete in complex formation (Fig. 7, Lanes 7-12). These results clearly

Figure 7. Gel shift experiment showing the effect of mutations in the Sp1 binding site within the P1 box sequence. Conditions were same as in Fig. 4, except for oligonucleotides added as competitors. In Lane 1, free oligonucleotide and no nuclear extract were added. In Lane 2, nuclear extract from PC12 cells and no competitor oligonucleotide were present. In Lanes 3 and 4, as in Lane 2, but indicated molar excess of unlabeled P1 box oligonucleotide was added as competitor. Lanes 5-14, as in Lane 2, but the indicated molar excess of unlabeled oligonucleotides containing modifications within the central 12 nt segment of the P1 box was added as the competitor; mutations were as shown at top of figure.

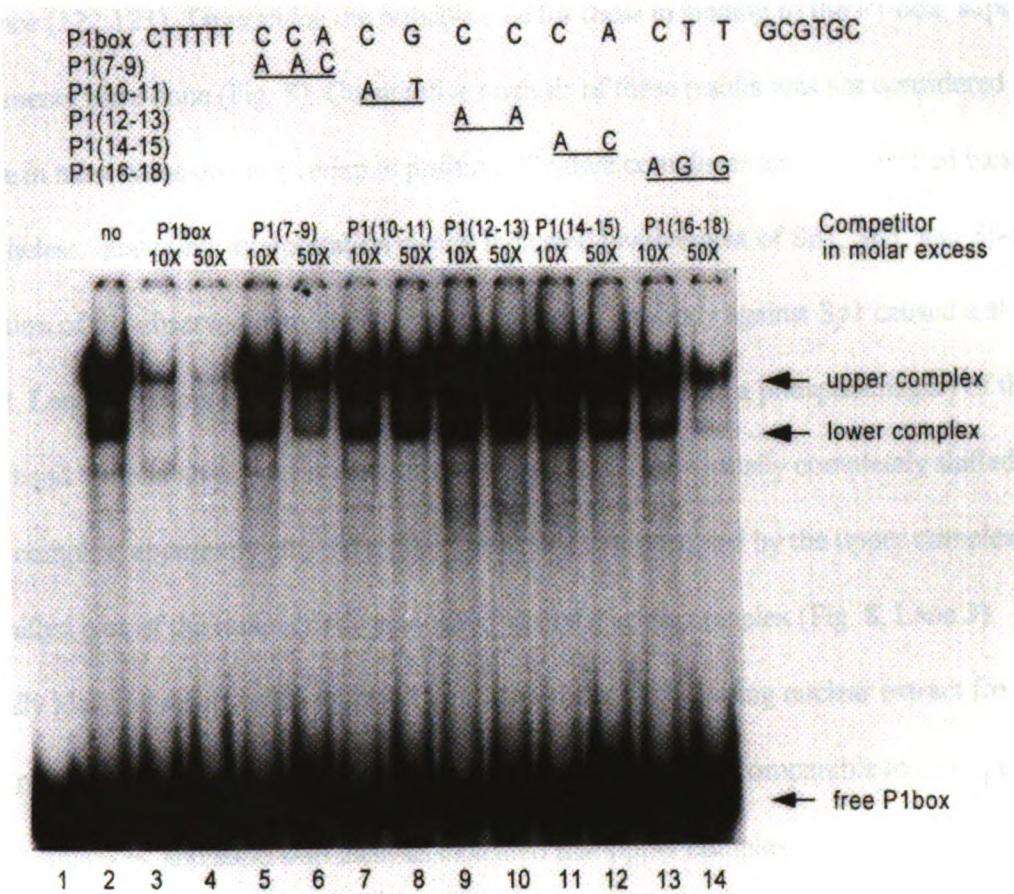

P1box CTTTTT C C A C G C C C A C T T GCGTGC
P1(7-9) A A C
P1(10-11) A T
P1(12-13) A A
P1(14-15) A C
P1(16-18) A G G
no
P1box 10X 50X
P1(7-9) 10X 50X
P1(10-11) 10X 50X
P1(12-13) 10X 50X
P1(14-15) 10X 50X
P1(16-18) 10X 50X
Competitor in molar excess
upper complex
lower complex
free P1box
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Fig. 7

demonstrate that formation of both upper and lower complexes require the presence of an intact Sp1 binding site within the P1 box.

Other members of the Sp family of transcription factors are also known to bind to this sequence (122,123). To examine the possible role for these in binding to the P1 box, supershift experiments were done (Fig. 8). Quantitative analysis of these results was not considered reliable in most cases due to overlap in position of native complexes and supershifted bands. Nonetheless, qualitative interpretation clearly indicated involvement of Sp1, Sp3, and Sp4 in formation of the observed complexes. Thus, addition of antibody against Sp1 caused a shift (Fig. 8, Lane 2) of approximately 50% (determined by analysis on a phosphoimager) of the upper band but had no effect on the lower complex. Anti-Sp3 virtually completely shifted the lower complex, apparently into the region just below that occupied by the upper complex, and also shifted part of the radioactivity associated with the upper complex (Fig. 8, Lane 3). Virtually identical results were reported by Majello *et al.* (125), using nuclear extract from HeLa cells, i.e., two complexes were seen, with relative positions comparable to our upper and lower complexes, and supershift analysis indicated that upper complex involved both Sp1 and Sp3 while the lower complex was due solely to binding of Sp3. Addition of antibody against Sp4 (Fig. 6, Lane 4) had no effect on the lower complex, but shifted approximately 10% of the radioactivity associated with the upper complex into a clearly defined region well above the residual upper complex. Combinations of these antibodies (Fig. 8, Lanes 5-8) had effects consistent with those predicted from observations with the individual antibodies, including a virtually complete shift of both complexes when all three antibodies were present.

Figure 8. Supershift experiment showing involvement of Sp1, Sp3, and Sp4 in complexes formed with nuclear extracts from PC12 cells. Conditions were as described for Fig. 4 except with addition of antibody or antibodies indicated at top of figure. Lane 1, no added antibody. The positions of the original "upper" and "lower" complexes are indicated by arrows at right of figure.

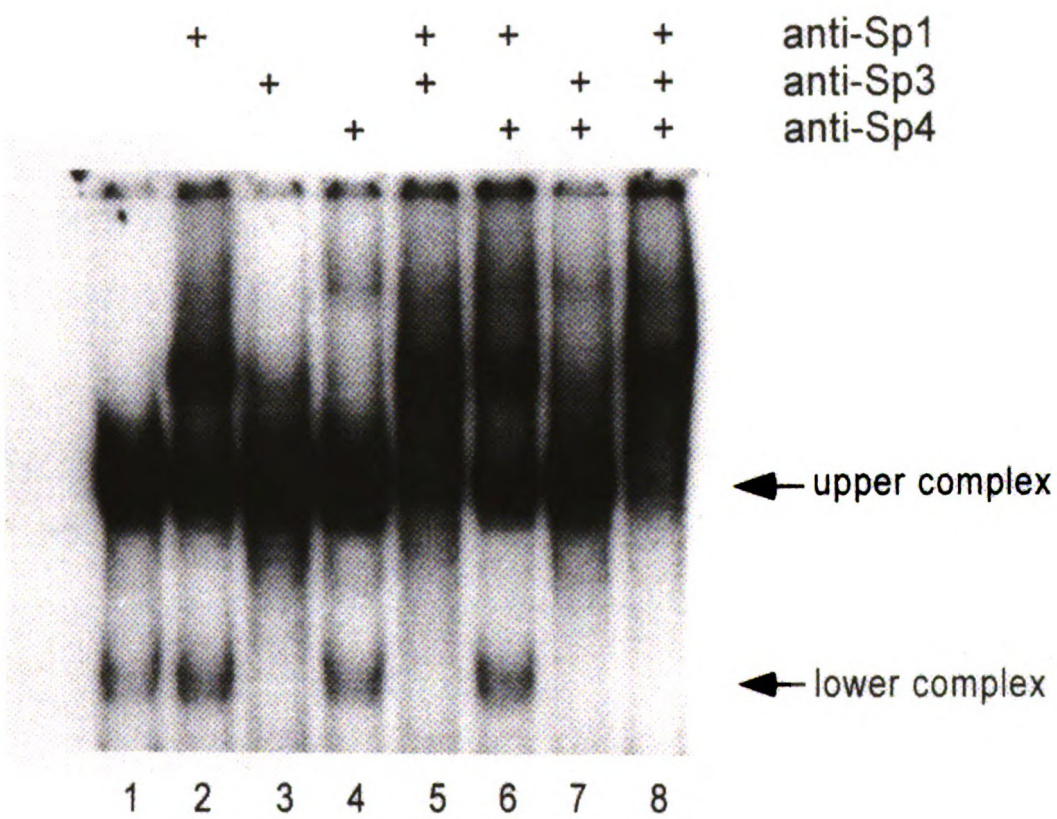
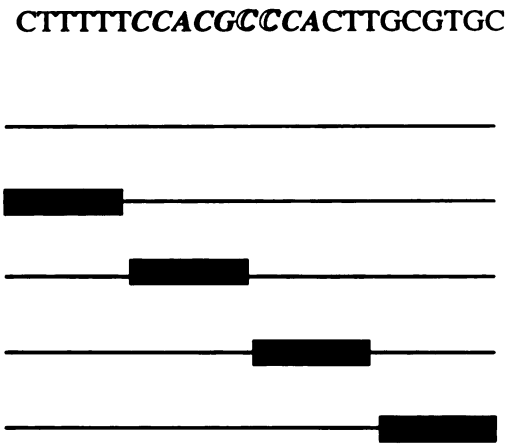

+ + + + anti-Sp1
+ + + + anti-Sp3
+ + + + anti-Sp4
upper complex
lower complex
1 2 3 4 5 6 7 8

Fig. 8

Figure 9. Summary of gel-shift and super-shift experiments. Shown on the top of the figure is the wild type sequence of the P1 box, in which the sequence matching to the Sp1 consensus is shown in bold and italic. All the subsequent lines also represent wild type sequences in competitor primers. Black bars represent block mutations in which six of the 24 nt sequence in P1 box are mutated. The mutant sequence in the central 12 nt are shown by the letters on the lines. The stars represent the unchanged sequences in the central 12 nt.



	competitor for upper complex		competitor for lower complex
	YES		YES
	YES		YES
	NO		NO
	NO		NO
	YES		YES
SP1 OLIGO	YES		YES
AP2 OLIGO	NO		NO
AAC*********	YES	(LA)	YES
AT****	NO		NO
*****AA*****	NO		NO
*******AC***	NO		NO
*********AGG	YES	(LA)	YES
ANTI SP1	SUPER		NO
ANTI SP3	SUPER		SUPER
ANTI SP4	SUPER		NO

SUPER = supershift
LA = lower affinity

Fig. 9

Figure 9 summarizes the results of gel shift and supershift experiments discussed above. It clearly shows that the central region of P1 box, which contains a Sp1 binding site, is critical for the formation of both complexes. The Sp1 consensus sequence competes for complexformation; the upper complex contains Sp1, Sp3 and Sp4, while the lower complex contains Sp3 only.

6. Effect of overexpression of Sp transcription factors on HKI expression in PC 12 cells

Human Sp1 and rat Sp3, Sp4 were overexpressed in PC12 cells by transfecting cells with plasmids containing cDNAs for Sp factors driven by CMV promoter. Western blot experiment (Fig.10) showed that protein levels of Sp1 (lane 2), Sp3 (lane 4) and Sp4 (lane 6) in extracts of transfected PC12 cells were higher than those in control PC12 cells (lanes 1, 3, 5), which were transfected with pGL3-control. Although some non-specific bands were also detected, the ones indicated by arrows were the major ones in the expected positions according to molecular weights; thus they were analyzed on Molecular Imager System, and the intensities relative to controls were shown in Table 8. Sp1, Sp3 and Sp4 in transfected cells were overexpressed 2.9, 6.7 and 2.0 fold, respectively, over the control. However, hexokinase activities in transfected PC12 cells were not significantly changed in transfected PC12 cells. This might be due to a saturating amount of endogenous Sp factors in PC12 cells.

7. Reporter constructs including sequence between positions -171 and +1 reveal another region with promoter activity

Figure 10. Overexpression of Sp1, Sp3 and Sp4 in PC12 cells. 100 μg protein was loaded into each lane. In lanes 1, 3 and 5, PC12 cells were transfected with pGL3-control plasmid; in lanes 2, 4 and 6, PC 12 cells were transfected with Sp1, Sp3 and Sp4 expression plasmids, respectively. Membranes were probed with corresponding antibodies shown on top of the figure. The arrows indicate the major bands for Sp factors.

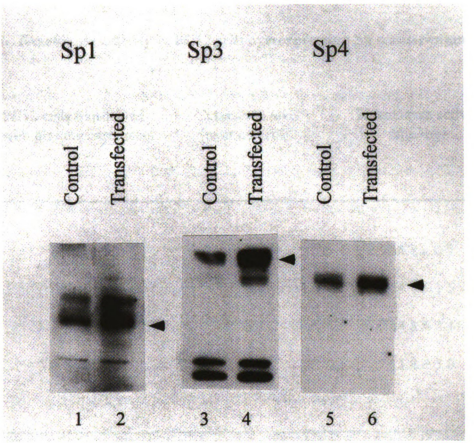

Sp1
Sp3
Sp4
Control
Transfected
Control
Transfected
Control
Transfected
1
2
3
4
5
6

Fig. 10

Table 8. Hexokinase activity in PC12 cells overexpressing Sp transcription factors

PC12 cells transfected with plasmid expressing	Expression level[b] (over control)	Hexokinase activity[a] (mu/mg)
control	1	14.1 ± 0.7
Sp1	2.9	14.5 ± 1.5
Sp3	6.7	14.8 ± 2.4
Sp4	2.0	11.8 ± 3.0

[a] Data for hexokinase activity were expressed as mean ± SD of six independent transfection experiments.

[b] Data for expression level were obtained from a representative experiment shown in Figure 10.

Transfection studies were also done with a series of reporter constructs that included the most downstream transcriptional start sites, in the -100 region (75). The results are shown in Fig. 11. Comparison with data in Fig. 2 immediately reveals very notable differences. In contrast to the virtual lack of effect of deletions in the -2511 to -742 region, seen with the previous set of reporter constructs (Fig. 2), significant decreases in luciferase expression were observed with the second set of constructs which included the -171 to -1 sequence (Fig. 11). While an 85% decrease in promoter activity was seen with deletion of the -742 to -516 region when the shorter constructs were used (Fig. 2), the corresponding deletion had only a modest effect with the longer reporter constructs (Fig. 11). Progressive decreases in luciferase expression were seen as the sequence between -364 and -1 were deleted. The decrease with loss of sequence between -364 and -245 was also seen with the shorter reporter constructs (Fig. 2), but this was, quantitatively, much less striking than the effects of deletions in the -742 to -516 region; this was, of course, why we focused on this latter region for the detailed analysis presented above.

These results indicate the presence of an additional region of promoter activity, located in the -364 to -1 region. The strength of this promoter is comparable to that identified in the -742 to -516 region. Thus, luciferase expression driven by pGL2SS (which includes sequence from -742 to -171) was 10.9 ± 2.4% of that seen with the pGL3-Control (Fig. 2); this appears to be largely associated with the upstream promoter, being markedly decreased by deletion of this region (-742 to -516). Very similar expression, 8.8 ± 1.5% of that with the pGL3-Control (Fig. 11), was seen with the construct including only the sequence between -364 and -1, and

Figure 11. Transfection of PC12 cells with promoter-reporter constructs *including* sequence between -171 and -1. A, schematic representation of reporter constructs. All constructs included sequence upstream from position -1 in the 5' flanking region of the gene encoding rat Type I hexokinase (75); see text for more detailed comments. B, normalized (to ß-galactosidase activity) luciferase activities after transfection of PC12 cells with indicated construct or basal control vector, pGL2-basic; results are the mean ± SD for at least six independent transfections, and are given as percentage of normalized luciferase activity expressed from the pGL3-control plasmid. An asterisk means that the activity of a construct is significantly, according to statistic analysis, different from that of the next longer construct.

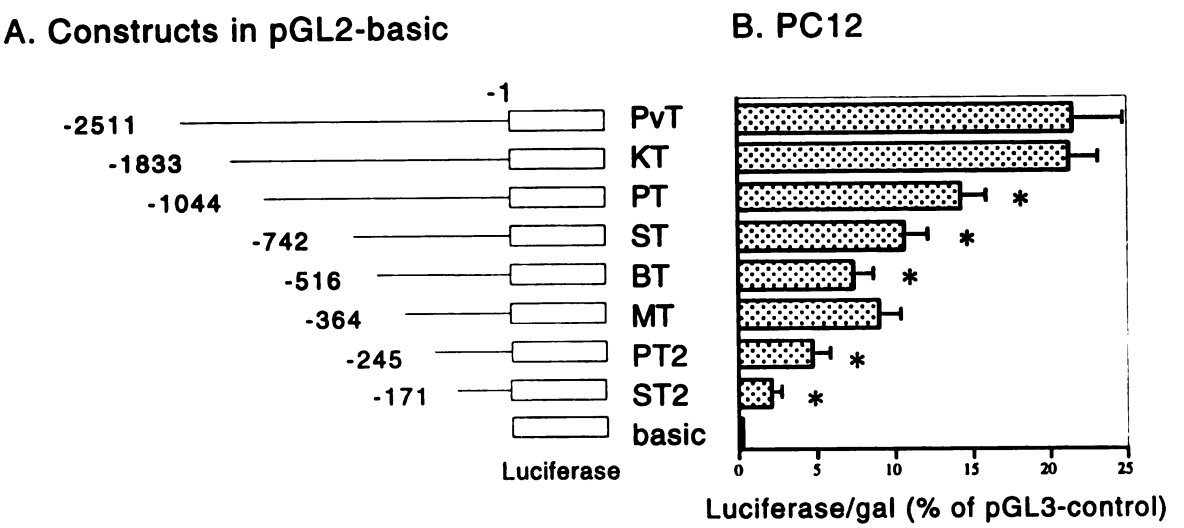

A. Constructs in pGL2-basic
B. PC12
-1
-2511
-1833
-1044
-742
-516
-364
-245
-171
PvT
KT
PT
ST
BT
MT
PT2
ST2
basic
Luciferase
0
5
10
15
20
25
Luciferase/gal (% of pGL3-control)

Fig. 11

thus completely lacking any promoter activity associated with the upstream region. Moreover, the activity of these promoter regions was not additive since luciferase expression from the reporter construct pGL2ST (Fig. 11), which includes sequence from -742 to -1 and thus contains both promoter regions, was 10.5 ± 1.6% of the pGL3-Control; this is virtually identical to the activity seen with the constructs containing either one of the promoter regions alone.

8. Effect of downstream sequence (+1 to +77) on promoter activity

To investigate the possible influence of downstream sequence on the hexokinase I promoter activity, two more promoter-reporter constructs were made. The first one, pGL2-SN, containing hexokinase I sequence from -742 to +77, was made by replacing the Mlu I (-364) / Hind III (in multiple cloning site of vector) fragment in pGL2-SS construct with Mlu I (-364) / Nar I (+77) fragment of hexokinase I genomic DNA (pbs5). pGL2-SN2 construct (-171 to +77) was made by deleting Sac I (-742) / Sac I (-171) fragment from the pGL2-SN construct. Transfection data are shown in Table 9. The promoter activity of ST construct, as shown previously, was about 10.5% of that of pGL3-control. When sequence from +1 to +77 was included, as in SN construct, promoter activity was increased by about 60%, to 16.1% of that of pGL3-control. On the other hand, when the same region (+1 to +77) was added to ST2 construct resulting in SN2 construct, the promoter activity remained at basal level, e.g., around 2% of that of pGL3-control. Taken together, these data suggest that sequence from +1 to +77

Table 9. Effect of downstream sequence on promoter activity

Constructs in pGL2-basic	Position of promoter fragment	Activity[a] (% pGL3-control)
ST	-742 to -1	10.5 ± 1.6
SN	-742 to +77	16.1 ± 4.1
ST2	-171 to -1	2.0 ± 0.7
SN2	-171 to +77	2.2 ± 0.3

[a] Luciferase activities are normalized to ß-galactosidase activities and expressed as mean ± SD of six independent transfection experiments.

increase the activity of either the upstream promoter (P1 box and -570 Sp1 site) or the downstream promoter (-364 to -1) or both, but it does not possess promoter activity of its own to drive transcription from -100 start sites.

CHAPTER IV

DISCUSSION

Two regions of promoter activity have been identified in the 5' flanking region of the gene for rat Type I hexokinase. The upstream promoter is associated with fragment -742 and -516, the downstream one is with fragment -364 to -1, based on the results of experiments with promoter-reporter constructs. The major focus of the present study has been on the upstream promoter. Footprinting analysis identified two protected regions, the P1 and P2 boxes. Functional analysis revealed that P1 box and -570 Sp1 site are important cis-elements for hexokinase I promoter activity. It is evident that members of the Sp family play a prominent role in regulation of the upstream promoter.

1. The upstream promoter: P1 box and -570 Sp1 site

Promoter activity was markedly dependent on maintenance of sequence at *both* the Sp1 binding site located within the P1 box (approximate position -540) as well as an additional Sp1 site just upstream (approximate position -570) from the P1 box. These sites are located more than the minimal 10 bp required to permit simultaneous binding of Sp1 at both sites (110) and the results with the reporter constructs clearly suggest that both sites must be occupied for expression of full promoter activity. The inability to detect protection of the upstream Sp1 site by DNase I footprinting presumably reflects a lower affinity for binding to this site, at least under the conditions of the footprinting experiments. Multiple Sp1 sites are often associated with regulation of TATA-less promoters in GC-rich regions, characteristic of housekeeping enzymes (155, 156) such as Type I hexokinase.

It is interesting to note that, although intactness of both Sp1 sites (-540 and -570) is required for full activity of the promoter in this region (see results in Fig. 5), the presence of a single Sp1 site at the position (-540) normally occupied by the Sp1 site intrinsic to the P1 box is sufficient for maintaining promoter activity at approximately 50% of its full activity. Thus, either deletion of the upstream (-570) Sp1 site *or* deletion of the P1 box in construct pGL2SSdelP1(24), which brings the upstream site into about the same position as that normally occupied by the -540 Sp1 site, results in approximately 50% of the full promoter activity.

2. Multiple members of Sp family of transcription factors regulate the activity of upstream promoter

Gel shift and supershift experiments suggested that several members of the Sp1 family of factors (122, 123) can bind to P1 box region and are likely to participate in regulating expression of the Type I isozyme of hexokinase. Both Sp1 and Sp4 serve as activators of transcription (124, 125). In contrast, Sp3 generally seems to play a repressive role in transcriptional regulation (124, 125), although this may depend on both promoter type as well as presence or absence of other transcription factors (127). The activity of the upstream promoter region can be expected to depend on the relative levels Sp factors present in various tissues.

Based on the fact that the sequence between -742 and -516 accounts for 85% of the promoter activity in both PC12 cells and H9c2 cells (Fig 2), we think that the activities of Sp1

family of transcription factors in a given tissue determine the activity of hexokinase I upstream promoter in that tissue. The ubiquitous expression of Sp1 and Sp3 (123) as well as the frequent involvement of the Sp family in regulating expression of housekeeping genes (155, 156) suggests that the upstream promoter may be involved in basal expression of the ubiquitous Type I isozyme. The transcriptional activator, Sp4, is expressed at particularly high levels in brain (124), and this may well be related to the very high levels of Type I hexokinase in brain when compared to other tissues (52).

Although the supershift experiment showed that with Sp1, Sp3 and Sp4 antibodies added to the reaction, almost all the complexes were supershifted, it is still possible that the remaining radioactivity is the result of some other transcription factor(s) bound to the Sp1 site. Thus, there might be other factor(s) regulate hexokinase I promoter activity through binding to P1 box sequence.

3. The P2 box and transcription initiation

The P2 box includes the most upstream (approximately -460) transcriptional start sites (75). We think it is likely that protection of the P2 box results from binding of RNA polymerase II or associated factors. Thus, the lack of specific sequence requirement (indicated by the lack of effect of mutations on expression from pGL2SSsubP2 and pGL2BSsubP2) is consistent with the existence of multiple transcriptional start sites throughout this region (75), implying lack of sequence specificity in binding of the polymerase.

There are two Inr (initiator) elements flanking 5'- and 3'- end of the P2 box sequence (Appendix A). Inrs are involved in transcriptional start site selection in many TATA-less promoters (157). They usually overlap transcriptional start sites, with consensus sequence of T/CT/CANT/AT/CT/C (158). The two Inrs around P2 box might specify the transcriptional start sites at -460.

Recently, another element called MED-1 (Multiple start site Element Downstream) is identified as an element regulating a window of multiple start sites in TATA-less promoter (159). The MED-1 sequence is usually positioned at 20-45 bp downstream from the 3'- end of multiple start site window, or up to 110 bp downstream from the 5'- end of the window. The consensus sequence is GCTCCC/G. Thus, the MED-1 element located at -170 in hexokinase I promoter region (see Appendix A) might be the determinant for transcriptional initiation at the multiple start sites around -300 to -200.

4. The downstream promoter

The upstream promoter activity is clearly functional with the upstream transcriptional start sites (75), located in the -460 and -300 regions, since the downstream transcriptional start sites in the -100 region were not included in the shorter reporter constructs (results in Fig. 2). In view of the distance between the P1 box and the downstream start sites (-100), it seems unlikely that this upstream promoter can drive transcription from the downstream start sites, although this remains to be directly determined. It is more probable that the promoter activity located in the -364 to -1 region is associated with transcription from the downstream start sites.

More extensive analysis of the downstream promoter is currently underway. However, the results presented here (compare Figs. 2 and 11) already make it clear that distinct *cis*-elements will be involved.

Recent report showed that myogenesis and MyoD down-regulated Sp1, and that it is the mechanism for the repression of GLUT1 during muscle differentiation (160). In contrast, level of hexokinase I did not change under the same condition (personal communication with Dr. A. Zorzano from Departament de Bioquimica i Biologia Molecular, Universitat de Barcelona, Spain). This might be due to the second (downstream) promoter of hexokinase I, which is not regulated by Sp1 factor.

5. Effect of +1 to +77 on the promoter activity

Transfection data showed that sequence between +1 and +77 increased promoter activity by about 60%. But it is not an independent promoter, since it did not drive the transcription from -100 start sites. It remains to be tested whether this region is acting at the upstream or downstream promoter.

CHAPTER V

FUTURE WORK

1. To investigate roles of Sp factors in regulation of the upstream promoter

- Transfection into Drosophila SL2 cells

The overexpression of Sp factors in PC12 cells did not increase hexokinase level in the cells. It is perhaps due to the saturating amount of endogenous Sp factors. Drosophila cell SL2 lacks endogenous Sp factors; thus it is a cell line frequently used to study the role of Sp family of transcription factors. Co-transfection of SL2 cells with Sp expression plasmids and hexokinase I promoter/reporter constructs should reveal how Sp factors regulate activity of the upstream promoter, e.g., which factor is up-regulating or down-regulating the promoter activity, and which factor has stimulating or suppressing effect on other factors. These experiments will show the function of Sp factors on activity of the upstream promoter.

2. To test downstream promoter activity in H9c2 cells

Downstream promoter activity associated with -364 to -1 region has been shown in PC12 cells (Fig. 11). The same set of promoter-reporter constructs can be used to transfect H9c2 cells to see if the same region is also active in H9c2 cell. If the profile of promoter activity is different in H9c2 cells from that in PC12 cells (Fig. 11), it might suggest a mechanism for the different expression level of hexokinase I in different tissue and cell lines.

3. To identify cis-elements in the downstream promoter

More extensive analysis of the downstream promoter is currently underway. The fragment associated with downstream promoter, MT (-364 to -1) is being used in DNase I footprinting experiment with PC12 nuclear extract. This experiment should reveal the cis-elements in this promoter region. The protected fragment in the footprinting experiment will be used in gel-shift and supershift experiments to identify the trans-factors bound to the region. Finally, the functional significance of the cis-elements will be tested by mutating the elements in promoter-reporter constructs which will be transfected into PC12 cells.

4. To study transcriptional regulation of hexokinase I expression

By cAMP

It has been shown that hexokinase I mRNA is increased in cultured rat thyroid FRTL5 cells by thyroid stimulating hormone (TSH) through cAMP pathway (64). With the availability of various promoter-reporter constructs, cAMP response element can be identified by stimulating FRTL5 cells which are transfected with nested deletion of promoter-reporter constructs. When the element is deleted, the cAMP response will be lost.

By energy demand

The levels of type I hexokinase in various rat brain regions have been correlated with basal rates of glucose utilization. We want to study the regulation of hexokinase I in response to changes in energy demand in PC12 cells. The energy demand and glucose utilization of PC12 cells will be stimulated by increasing the extracellular KCl concentration which constantly depolarizes PC12 cells. Once the condition is set up for hexokinase level in response to changes in energy demand, it will be asked whether the changes are due to transcriptional regulation. If that is the case, various promoter-reporter constructs can be transfected to PC12 cell under stimulating condition, and see which fragment keeps HKI promoter responsive to the change of energy demand, and what cis-element(s) and trans-factor(s) are involved in the regulation.

5. To define transcriptional start site and promoter usage in different tissues and cell lines

Levels of hexokinase I mRNA vary in different tissues and cell lines. It is of great interest to study whether this variation is determined by different usage of transcriptional start sites and /or promoters. RNase protection assay with mRNA from different sources should reveal the preferential usage of transcriptional start sites. With the availability of HKI promoter-reporter constructs, the activities of the upstream promoter and downstream promoter can be tested in different cell lines. In addition, similar promoter

constructs linked to different reporter system in an in-vitro assay allow us to test promoter activity driven by nuclear extract from different tissues. Comparison of start site and promoter usage should give us valuable information about how HKI levels are regulated in different tissues at the transcription level.

LIST OF REFERENCES

LIST OF REFERENCES

1. Katzen, H.M., Soderman, D.D., and Nitowsky, H.M. (1965) Biochem. Biophys. Res. Commun., 19, 377-382.

2. Gonzalec, C., Ureta, T., Babul, J., Rabajille, E., and Niemeyer, H. (1967) Biochemistry, 6, 460-468.

3. Ureta, T. (1982) Comp. Biochem. Physiol. 71B, 549-555.

4. Schwab, D.A. and Wilson, J.E. (1988) J. Biol. Chem. 263, 3220-3224.

5. Schwab, D.A. and Wilson, J.E. (1989) Proc. Natl. Acad. Sci. USA 86, 2563-2567.

6. Schwab, D.A. and Wilson, J.E. (1991) Arch. Biochem.Biophys. 285, 365-370.

7. Thelen, A.P. and Wilson, J.E. (1991) Arch. Biochem.Biophys. 286, 645-651.

8. Andreone, T.L., Printz, R.L., Pilkis, S.J., Magnuson, M.A. and Granner, D.K. (1989) J.Biol.Chem. 264, 363-369.

9. Iynedjian, P.B., Ucla, C., and Mach, B. (1987) J.Biol.Chem. 262, 6032-6038.

10. Stachelek, C., Stachelek, J., Swan, J., Botstein, D., and Konigsberg, W. (1986) Nucleic Acid Research 14, 945-963.

11. Kurokawa, M., Oda, S., Tsubotani, E., Fujiwara, H., Yokoyama, K., and Ishibashi, S. (1982) Mol. Cell. Biochem. 45, 151-157.

12. Easterby, J.S. and O'Brien, M.J. (1973) Eur. J. Biochem. 38, 201-211.

13. Rose, I.A., Warms, J.V.B., and Kosow, D.P. (1974) Arch. Biochem. Biophys. 164, 729-735.

14. Holroyde, M.J., Trayer, I.P., and Cornish-Bowden, A. (1976) FEBS Lett. 62, 215-219.

15. Gregoriou, J., Trayer., I.P., and Cornish-Bowden, A. (1983) Eur. J. Biochem. 134, 283-288.

16. Printz, R.L., Koch, S., Potter, L.R., O'Doherty, R.M., Tiesinga, J.J., Moritz, S., and Granner, D.K. (1993) J. Biol. Chem. 268, 5209-5219.

17. White, J.A. and Wilson, J.E. (1997) Arch. Biochem. Biophys. 343, 207-214.

18. Magnuson, M.A., Andreone, T.L., Printz, R.L., Koch, S., and Granner, D.K. (1989) Proc. Natl. Acad. Sci. USA 86, 4838-4842.

19. White, T.K., and Wilson, J.E. (1987) Arch. Biochem. Biophys. 259, 402-411.

20. White, T.K., and Wilson, J.E. (1989) Arch. Biochem. Biophys. 274, 375-393.

21. Chou, A.C., and Wilson, J.E. (1974) Arch. Biochem. Biophys. 165, 628-633.

22. Ellison, W.R., Lueck, J.D., and Fromm, H.J. (1974) Biochem. Biophys. Res. Comm. 57, 1214-1220.

23. Ellison, W.R., Lueck, J.D., and Fromm, H.J. (1975) J.Biol.Chem. 250, 1864-1871.

24. Tsai, H.J., and Wilson, J.E. (1995) Arch. Biochem. Biophys. 316, 206-214.

25. Tsai, H.J., and Wilson, J.E. (1997) Arch. Biochem. Biophys. 338, 183-92.

26. Tsai, H.J., and Wilson, J.E. (1996) Arch. Biochem. Biophys. 329, 17-23.

27. Wilson, J.E. (1985) in Regulation of Carbohydrate Metabolism (Beitner, R., Ed.), Vol.I, pp.45-85, CRC Press, Inc., Boca Raton, FL.

28. Wilson, J. E. (1995) Rev. Physiol. Biochem. Pharmacol. 126, 65-198.

29. Preller, A., ans Wilson J.E. (1992) Arch. Biochem. Biophys. 294, 482-492.

30. Katzen, H.M., and Schimke, R.T. (1965) Proc. Natl. Acad. Sci. USA 54, 1218-1225.

31. Printz, R.L., Magnuson, M.A., and Granner, D.K. (1993) Annu. Rev. Nutr. 13, 463-496.

32. Magnuson, M.A. (1990) Diabetes 39, 523-527.

33. Katzen, H.M., Soderman, D.D., and Wiley, C.E. (1970) J. Biol. Chem. 245, 4081-

4096.

34. Bartley, J.C., Barber, S., and Abraham, S. (1975) Cancer Res. 35, 1649-1653.

35. Felgner, P.L., and Wilson, J.E. (1977) Arch. Biochem. Biophys. 182, 282-294

36. Polakis, P.G., and Wilson. J.E. (1985) Arch. Biochem. Biophys. 236, 328-337.

37. Xie, G., ans Wilson J.E. (1988) Arch. Biochem. Biophys. 267, 803-810.

38. Gelb, B.D., Adams, V., Jones, S.N., Griffin, L.D., MacGregor, G.R., and McCabe, R.B. (1992) Proc. Natl. Acad. Sci. USA 89, 202-206.

39. Sui, D., and Wilson J.E. Arch. Biochem. Biophys (in press).

40. Felgner, P.L., Messer, J.L., and Wilson, J.E. (1979) J. Biol. Chem. 254, 4946-4949.

41. Linden, M., Gellerfors, P., and Nelson, B.D. (1982) FEBS Lett. 141, 189-192.

42. Fiek, C., Benz, R., Roos, N., and Brdiczka, D. (1982) Biochim. Biophys. Acta 688, 429-440.

43. BeltrandelRio, H., ans Wilson J.E. (1991) Arch. Biochem. Biophys. 286, 183-194.

44. Viitanen, P.V., Geiger, P.J., Erickson,-Viitanen, S. and Bessman, S.P. (1984) J. Biol. Chem. 259, 9679-9686.

45. Salotra, P.T., and Singh, V.N. (1982) Arch. Biochem. Biophys. 216, 758-764.

46. Kosow, D.P., and Rose, I.A. (1968) J. Biol. Chem. 243, 3623-3630.

47. Wilson, J.E. (1968) J. Biol. Chem. 243, 3640-3647.

48. Iynedjian, P.B. (1993) Biochem. J. 293, 1-13.

49. Miwa, I., Mitsuyama, S., Toyoda, Y., Nonogaki, T., Aoki, S., and Okuda, J. (1990) Biochem. Int. 22, 759-767.

50. Toyoda, Y., Miwa, I., Kamiya, M., Ogiso, S., Nonogaki, T., Aoki, S., and Okuda, J. (1995) Histochem. Cell Biol. 103, 31-38.

51. Toyoda, Y., Miwa, I., Kamiya, M., Ogiso, S., Okuda, J., and Nonogaki, T (1995) FEBS-Lett. 359, 81-84.

52. Wilson, J.E., and Felgner, P.L. (1977) Mol. Cell. Biochem. 18, 39-47.

53. Hawkins, R. (1985) in Cerebral Energy Metabolism and Metabolic Encephalopathy (McCandless, D.W., Ed.), pp. 3-23, Plenum Press, New York.

54. Bennett, E.L., Drori, J.B., Krech, D., Rosenzweig, M.,R., and Abraham, S. (1962) J. Biol. Chem. 237, 1758-1763.

55. Lowry, O.H., Roberts, N.R., Schulz, D.W., Clow, J.E., and Clark, J.R. (1961) J. Biol. Chem. 236, 2813-2820.

56. Buell, M.V., Lowry, O.H., Roberts, N.R., Chang, M.W., and Kapphahn, J.I. (1958) J. Biol. Chem. 232, 979-993.

57. Wilson, J.E. (1972) Arch. Biochem. Biophys. 150, 96-104.

58. Wilkin, G.P, and Wilson, J.E. (1977) J. Neurochem 29, 1039-1051.

59. Jacobsson, G., Meister, B. (1994) Mol. Cell. Neurosci. 5, 658-677.

60. Wilson, J.E. (1972) J. Neurochem 19, 223-227.

61. Simurda, M.,and Wilson, J.E. (1980) J. Neurochem 35, 58-66.

62. Gutekunst, D.I., and Wilson, J.E. (1981) J. Neurochem 36, 1781-1785.

63. Griffin, L.D., Gelb, B.D., Adams, V., and McCabe, E.R.B. (1992) Biochim Biophys Acta 1129, 309-317.

64. Yokomori, N., Tawata, M., Hosaka, Y., and Onaya, T. (1992) Life Sciences 51, 1613-1619.

65. Balazs, R., Lewis, P.D., Patel, A.J. (1975) In: Brazier MAB (ed) Growth and Development of the Brain, Raven Press, New York pp83-115.

66. Turek, T.J., Hawkins, R.A., and Wilson, J.E. (1986) J. Neurochem 46, 983-985.

67. Burgess, R.W., and Wilson, J.E. (1991) J. Neurochem 57, 441-446.

69. Batipps, M., Miyaoka, M., Shinohara, M., Sokoloff, L., and Kennedy, C. (1981) Neurology 31, 58-62.

69. Hawkins, R.A., Mans, A.M., Davis, D.W., Vina, J.R., and Hibbard, L.S. (1985) Am. J. Physiol. 248, C170-C176.

70. Krukoff, T.L., and Vincent, D.H. (1989) Am. J. Physiol. 256, R1050-R1055.

71. Simon, J.K., Zhang, T-X, Ciriello, J. (1989) Am. J. Physiol. 256, R930-R938.

72. Krukoff, T.L., and Weigel, M.A. (1989) Brain Res. 499, 1-6.

73. Krukoff, T.L., and Patel, K.P. (1990) Brain Res. 522, 157-160.

74. Patel, K.P., Zhang, P.L., and Krukoff, T.L. (1993) Am. J. Physiol. 265, R923-R928.

75. White, J.A., Liu, W., and Wilson, J.E. (1996) Arch. Biochem. Biophys. 335, 161-172.

76. Liu, W., and Wilson, J.E. (1997) Arch. Biochem. Biophys. (in press).

77. Gardiner-Garden, M., and Frommer, M. (1987) J. Mol. Biol. 196, 261-282.

78. Bird, A.P. (1986) Nature 321, 209-213.

79. Katzen, H.M. (1967) Adv.Enzyme Regul. 5, 335-356.

80. Frank, S.K., and Fromm, H.J. (1986) Biochem. Biophys. Res. Comm. 138, 374-380.

81. Frank, S.K., and Fromm, H.J. (1986) Arch. Biochem. Biophys.249, 61-69.

82. Peter, J.B., Jeffres, R.N., and Lamb, D.R. (1968) Science 160, 200-201.

83. Pette, D., Smith, M.E., Staudte, H.W., and Vrbova, G. (1973) Pflugers Arch. 338, 257-272.

84. Simoneau J-A, and Pette, D. (1988) Pflugers Arch Eur J Physiol 412, 86-92.

85. Weber, F.E., and Pette, D. (1988) FEBS Lett. 238, 71-73.

86. Weber, F.E., and Pette, D. (1990) FEBS Lett. 261, 291-293.

87. Weber, F.E., and Pette, D. (1990) Eur. J. Biochem. 191, 85-90.

88. Hofmann, S., and Pette, D. (1994) Eur. J. Biochem. 219, 307-315.

89. O'Doherty, R.M., Bracy, D.P., Osawa, H, Wasserman, D.H., and Granner, D.K. (1994) Am. J. Physiol. 266, E171-E178.

90. Osawa, H., Printz, R.L., Whitesell, R.R., and Granner, D.K. (1995) Diabetes 44, 1426-1432.

91. Osawa, H., Brooks Robey, R., Printz, R.L., and Granner, D.K. (1996) J. Biol. Chem. 271, 17296-17303.

92. Bustamante, E., and Pedersen, P.L. (1977) Proc. Natl. Acad. Sci. USA 74, 3735-3739.

93. Bustamante, E., Morris, H.P., and Pedersen, P.L. (1981) J. Biol. Chem. 256, 8699-8704.

94. Johansson, T, Berres, J-M, Nelson, B.D. (1985) Biochem. Biophys. Res. Comm. 133, 608-613.

95. Shinohara, Y., Ichihara, J., and Terada, H. (1991) FEBS Lett. 291, 55-57.

96. Mathupala, S.P., Rempel, A., and Pedersen, P.L. (1995) J. Biol. Chem. 270, 16918-16925.

96a. Mathupala, S.P., Heese C., and Pedersen, P.L. (1997) J. Biol. Chem. 272, 22776-22780.

97. DiPietro, D. L., and Weinhouse, S. (1960) J. Biol. Chem. 235, 2542-2545.

98. Sharma, C., Manjeshwar, R., and Weinhouse, S. (1963) J. Biol. Chem. 238, 3840-3845.

99. Iynedjian, P. B., Gjinovci, A., and Renold, A.E. (1988) J. Biol. Chem. 263, 740-744.

100. Iynedjian, P. B., Jotterand, D., Nouspikel, T., Asfari, M., an Pilot. P.-R. (1989) J. Biol. Chem. 264, 21824-21829.

101. Iynedjian, P. B., Pilot. P.-R., Nouspikel, T., Milburn, J.L., Quaade, C., Huges, S., Ucla, C., and Newgard, C.B. (1989) Proc. Natl. Acad. Sci. USA 86, 7838-7842.

102. Bedoya, F.J., Matschinsky, F.M., Shimizu, T., O'Neil, J.J., and Appel, M.C. (1986) J. Biol. Chem. 261, 10760-10764.

103. German, M.S. (1993) Proc. Natl. Acad. Sci. USA 90, 1781-1785.

104. Magnuson, M.A., and Shelton, K.D. (1989) J. Biol. Chem. 264, 15936-15942.

105. Magnuson, M.A. (1992) J. Cell. Biochem. 48, 115-121.

106. Iynedjian, P. B., Marie, S., Wang, H., Gjinovci, A., and Nazaryan, K. (1996) J. Biol. Chem. 271, 29113-29120.

107. Shelton, K.D., Franklin, A.J., Khoor, A., Beechem, J., and Magnuson, M.A. (1992) Mol. Cell. Biol. 12, 4578-4589.

108. Moates, J.M., Shelton, K.D., and Magnuson, M.A. (1996) Mol. Endocinol. 10, 723-731.

109. Dynan, W.S., and Tjian, R. (1983) Cell 32, 669-680.

110. Courey, A.J., and Tjian, R. (1992) in Transcriptional Regulation (McKnight, S.L., and Yamamoto, K.R., Eds.), pp. 743-769, Cold Spring Harbor University Press, Cold Spring Harbor, NY.

111. Courey, A.J., Holzman, D.A., Jackson, S.P., and Tjian, R. (1989) Cell 59, 827-836.

112. Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987) Cell 51,1079-90.

113. Kadonaga, J.T., Courey, A.J., Ladika, J., and Tjian, R. (1988) Science 242, 1566-1570.

114. Courey, A.J., and Tjian, R. (1988) Cell 55, 887-898.

115. Lee, Y.H., Williams, S.C., Baer, M., Sterneck, E., Gonzalez, F.J., and Johnson, P.F. (1997) Mol. Cell. Biol. 17, 2038-2047.

116. Gill, G., Pascal, E., Tseng, Z.H., and Tjian, R. (1994) Proc. Natl. Acad. Sci. USA 91, 192-196.

117. Emami, K.H., Navarre, W.W., and Smale, S.T. (1995) Mol. Cell. Biol. 15, 5906-5916.

118. Han, I, and Kudlow, J.E. (1997) Mol. Cell. Biol. 17, 2550-2558.

119. Jackson, S.P., MacDonald, J.J., Lees-Miller, S., and Tjian, R. (1990) Cell 63, 155-165.

120. Vlach, J., Garcia, A., Jacque, J. M., Rodriguez, M.S., Michelson, S., and Virelizier, J.L. (1995) Virology 208, 753-761.

121. Daniel, S., Zhang, S., DePaoli, Roach, A..A., and Kim, K.H. (1996) J. Biol. Chem. 271, 14692-14697.

122. Kingsley, C., and Winoto, A. (1992) Mol. Cell. Biol. 12, 4251-4261.

123. Hagen, G., Müller, S., Beato, M., and Suske, G. (1992) Nucl. Acids Res. 20, 5519-5525.

124. Hagen, G., Müller, S., Beato, M., and Suske, G. (1994) EMBO J. 13, 3843-3851.

125. Majello, B., De Luca, P., Hagen, G., Suske, G., and Lania, L. (1994) Nucl. Acids Res. 22, 4914-4921.

126. Birnbaum, M.J., van Wijnen, A.J., Odgren, P.R., Last, T.J., Suske, G., Stein, G.S., and Stein, J.L. (1995) Biochemistry 34, 16503-16508.

127. Majello, B., De Luca, P., and Lania, L. (1997) J. Biol. Chem. 272, 4021-4026.

128. Dennig, J., Beato, M. and Suske, G. (1996) EMBO J. 15, 5659-5667.

129. Hagen, G., Dennig, J., Preiss, A. Beato, M. and Suske, G. (1995) J-Biol-Chem., 270(42), 24989-24994.

130. Imataka, H., Sogawa, K., Yasumoto, K-I, Kikuchi, Y, Sasano, K., Kobayashi, A., Hayami, M., and Fujii-Kuriyama, Y. (1992) EMBO J. 11, 3663-3671.

131. Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H., and Fujii-Kuriyama, Y. (1993) Nucl. Acids Res. 21, 1527-1532.

132. Park, G.T., Choi, C.Y., and Rho, H.M. (1996) J. Gen. Virol. 77, 2775-2780.

133. Strom, A.-C., Forsberg, M., Lillhager, P., and Westin, G. (1996) Nucl. Acids Res. 24, 1981-1986.

134. Xu, X.-M., Tang, J.-L., Chen, X, Wang, L.-H., and Wu, K.K. (1997) J. Biol. Chem. 272, 6943-6950.

135. Zhang, D.-E., Hetherington, C.J., Tan, S., Dziennis, S.E.,,Gonzalez, D.A., Chen, H.M., and Tenen, D. G. (1994) J. Biol. Chem. 269, 11425-11434.

136. Rodenberg, R. J., Holthuizen, P.E., and Sussenbach, J. S. (1997) Mol. Endocrinol. 11, 237-250.

137. Lu, J., Lee, W., Jiang, C., and Keller, E.B. (1994) J. Biol. Chem. 269, 5391-5402.

138. Bevilacqua, A., Fiorenza, M. T., and Mangia, F. (1997) Nucl. Acids Res. 25, 1333-1338.

139. Okumura, K., Sakaguchi, G., Takagi, S., Naito, K., Memori, T., and Igarashi, H., (1996) J. Biol. Chem. 271, 12944-12950.

140. Liang, Y., Robinson, D.F., Dennig, J., Suske, G., and Fahl, W. E. (1996) J. Biol. Chem. 271, 11792-11797.

141. Dennig, J., Hagen, G., Beato M., and Suske, G. (1995) J. Biol. Chem. 270, 12737-12744.

142. Netzker, R., Weigert, C., and Brand, K. (1997) Eur. J. Biochem. 245, 174-181.

143. Greene, L.A., and Tischler, A.S. (1982) *Adv. Cell. Neurobiol.* 3, 373-414.

144. Grossbard, L., and Schimke, R., T. (1966) J. Biol. Chem. 241, 3546-3560.

145. Whallon, J. H., Preller, A., and Wilson, J.E. (1994) Scanning 16, 111-117.

146. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

147. Sanger, F., Nicklen, S., and Coulson, A.R. (9177) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.

148. Higuchi, R. (1990) *in* PCR Protocols. A Guide to Methods and Applications (Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J., Eds.), pp. 177-183, Academic Press, New York.

149. Thelen, A. P. (1992) Ph. D. Thesis, Michigan State University, East Lansing.

150. Kimes, B.W., and Brandt, B.L. (1976) *Exp. Cell Res.* 98, 367-381.

151. Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1993) Current Protocols in Molecular Biology, pp. 12.1.1-12.1.9, Wiley, New York.

152. Wilson, J. E. (1989) Prep. Biochem. 19, 13-21.

153. Hill, H. D., and Straka, J. G. (1988) Anal. Biochem. 170, 203-208.

154. Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986) *Science* 234, 47-52.

155. Lu, J., Lee, W., Jiang, C., and Keller, E.B. (1994) *J. Biol. Chem.* 269, 5391-5402.

156. Xu, X.-M., Tang, J.-L., Chen, X., Wang, L.-H., and Wu, K.K. (1997) *J. Biol. Chem.* 272, 6943-6950.

157. Smale, S.T., and Baltimore, D. (1989) Cell 57, 103-113.

158. Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., and Smale, S.T. (1994) *Mol. Cell. Biol.* 14, 116-127.

159. Ince, T.A., and Scotto, K.W. (1995) *J. Biol. Chem.* 270, 30249-30252.

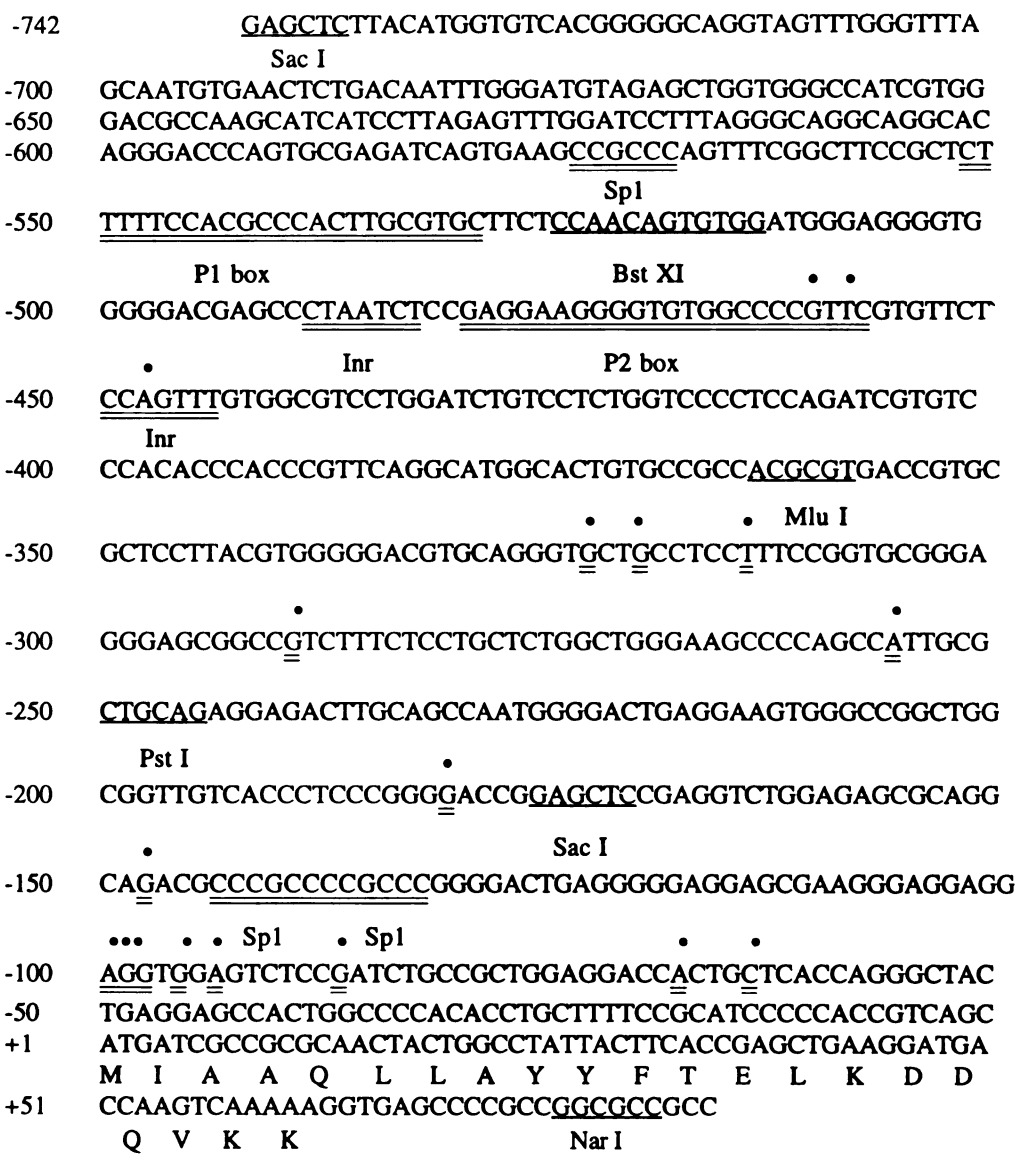
160. Vinals, F., Fandos, C., Santalucia, T., Ferre, J., Testar, X., Palacin, M., and Zorzano, A. (1997) *J. Biol. Chem.* 272, 12913-12921.

APPENDICES

APPENDIX A

APPENDIX A

DNA SEQUENCE AROUND HKI PROMOTER REGION



Dots above DNA sequence represent transcriptional start sites.
Underlined sequences represent restriction sites.
Double underlined sequences represent transcriptional start sites, Inrs, important promoter elements and potential Sp1 binding sites.
A in start codon ATG is designated as +1.

APPENDIX B

APPENDIX B

DNA SEQUENCE OF HKI UPSTREAM REGION

pBS39.5 (1 TO 3450)
(-3366 to +84, relative to start codon)

```
   1 GAATTCCAAG ACAGCCAGGA GTACACAGAG GGACCGTGTC TTGAAAAAAC  -3317
     ⇒
  51 AACCAACCAA CCAATCAACA TAACCCATGC ATGAAGCCTG AAGAGATGAC  -3267

 101 TCAGTGGTAA AGAGTTTTTG ACTGCTCTAG CAGAGGACCT AAGTTCAATT  -3217

 151 CCCAGCACCC ACTGGTGTCT CACAAACATC TTTAACTCCA GTTCCAGGGG  -3167

 201 ACCTGATGTT CTCTTCTGAC CTCCTCAAGC ATCAGGAACA TGTATGGTAC  -3117

 251 GTACACACAC ACACACACAC ACACACACAC ACACACACAC ACACACACTA  -3067

 301 TACATACATA TATACATGTA AGCAAAATCT TCATATAAAA AAACCTAAAA  -3017

 351 AAATTAAAAA TAAATCCAGG CATAGTAGCG ATCACCTTGG GCCTTGCGCT  -2967

 401 TCCCAGGCAA GCGCTCTACC ACTGAGCTAA ATCCCCAACC CCATAGTAGC  -2917

 451 GATAACCTTT AAGCCCAGCA CTTGGGAGGC AGAGGTAGGC AGATCTCTGT  -2867

 501 GAGTTCAAGG CCAGCCTGGT CTATATAGTT CCAGGACTGC CAAAGCTATA  -2817

 551 CAGTAAGACC CTGTTTCAAA AACTAATAAA AAACCAACAA GAAACATTCA  -2767

 601 TGAGGTAGCC TAGGGCTTAA CCAAGAGAAG ATACAACACG GCCATAGTGC  -2717

 651 TGCCTCCAGG GAATGGCCTG ATGGAAGTGA TAAGGCCTGC TCCGGCAGAC  -2667

 701 CGTCTCCCCT ATCAGGGAGG CAGTCTTGAC TCAGTTGCTT ATCTGGCCCA  -2617

 751 GTCTGCCTGA GGTGGTGGGG TGGGATGGAA AGAGTGCTGG CTAAGCCTCA  -2567

 801 GAGTCCCGGC TCGAAATTCT ACGGCAACCT GCCGCTGGGC AAGATGGCCC  -2517

 851 ACCAGCTGCG CACCCTACAC CTTGTTTGTA GAGAAGAGCC CTGGTCTCCA  -2467
           ⇒
 901 TTCGCATGGA CTCTGAGCAC TGTCTGAGTG TGCCCTGTGG AGGTGCTGGG  -2417

 951 GCCTGGGAGG CCAGGGTGAG TGGCTGAGGT CAGGGGAGGC CAGCTCGGGA  -2367
```

1001 AGGAGGTCAG ATGTGCCTTC CTGTGGTGAG GTCGGAACCT GGACAGGTGT -2317

1051 AGGAGAATCG GGACGGAAGG GTGGTATCGA GGGGAGCCAC ACAGAGTGGA -2267

1101 TGTGTGGGGT TAAGGACAGT GGGACACACT AGTGCTGTCG GTCAGGGGCT -2217

1151 GGGTAAGAGA TTCCGCGCTC AGAGTCCGGT AACAGGGCCA GTGTGTTTAA -2167

1201 ATTGGCTTCC ACCAGGATGC AGGGCCCAGA AGGGACTTTG AGCAGGGGCA -2117

1251 TAACGTAATT AAGGTTGCTT TTCAGTGAGA CTGTCGCGGC CTGCAGCACA -2067

1301 GAATGGCTTG GATGGCGGGT AGCAACCAAC CAGAAAGCCA GCTCAAATCG -2017

1351 TTCTTGTCAT ATTTCTAGCT AGACAAGAGA AAAACTGGGG TGGAGGTGGT -1967

1401 GTTGGGAGTA GCAGTGAGTC TGACTTAGAG AGATATGCAT TAGGAAACAA -1917

1451 CACTAAGCCT CGGTGCCTGG GTGTATGTGT GTGTGTCTGT GTCCTGAGTC -1867

1501 ACCTTGGTAC AGGGGTTGAG GGAAGGGACA GCAGAGGTAC CTACCTGCCT -1817

⇒

1551 ACAAGCTGGT CCAGCTTTTC CCCAGCTCTG GTTTGTTGGC TAAGGGTTGT -1767

1601 ATCCCCATAT AATCCCACAG ATTATAGTGA TGGCTGAAGG TTGGGGGAGG -1717

1651 GGTCCATGAG CCCGACACTG GGGTTGGAGA ATGGGGTGGC CCATTCAACC -1667

1701 TACAGACCTC TGAGAAAACA AAACAACCCT CCATCCCCAA AGACTTGCCA -1617

1751 GCCTTGAGAG CTGGCCGAAG TACATTTGCA GATTCAAATG CAGCTCAGAA -1567

1801 GAACCATTTT TGTCTTGACG CTGTCAGGCA ATATTGGTAG CGTGTGTGTG -1517

1851 TGTGTGTGTG TTAGTTTGAG ATTTACGTGG GACTCAATGC TAATCTTTTT -1467

1901 AAGGGAAGGG CAACTATAAC AGGCCCTCCC TGGGCTGGAA ATGGTGACAT -1417

1951 TCAGGTGACA GGTCCCCCAA ACTTTAACCT GTAGGTTCCC GAGTCCGTTC -1367

2001 TATGAGAGTA CCTCCCATCT AATCTATTAG ATACCAGTGT TTATGCTCTT -1317

2051 TCAGAAAAAT TACAGGGTGT AGGCATCAGC TCCATCTTAT GGGTAAAGAA -1267

2101 AGAAAGCCTG AGAGGTGACT CAGGACACAT CTGCACCAGC CATTAGGCAG -1217

2151 GGTGGTGGCT GGAAGGGCCC TGGGATACTT GAAGGGCACC CTGGGGTATC -1167

2201 TTTGCCCCGA TGGATGCTGG GCCTAGCCAG GGGCTCCCAG TCCCCAGGCG -1117

2251 TGGGGTAGAA GTTGGACTCT ATAGTCACCT AAGGGCCTAT GTTGCAGTCC -1067

2301 TGGTCTCAGG CACACGGCCT GCAGGAGAGG CTTAAAAGAA GAGAAACTGC -1017

⇒

2351 ACACAATGGC TAGGTCACCG GCGTTAAAGC TAAGCAAACC CAGCGCTACT -967

2401 CCTGGGCAGC AACTGCAAAG CGTTTCTTCA GGTCCCTCAC CTGTAGAATC -917

2451 AGAGCGGTAG TCGCCTCTGC ATGTCTGAGT TCTTACATGT CGTAATGTAC -867

2501 AAACACGATT TCCCCTCAAT CACCGCCCGG AACAGTACCT CCAACTTCCC -817

2551 AGACCCGGAT GCCCCAAGAG CCAGAGTAGG GTGGGAAAAT CGGGACAGGC -767

2601 CCCCAAATTC CACTCGGGGG CCTTGAGCTC TTACATGGTG TCACGGGGGC -717

⇒

2651 AGGTAGTTTG GGTTTAGCAA TGTGAACTCT GACAATTTGG GATGTAGAGC -667

2701 TGGTGGGCCA TCGTGGGACG CCAAGCATCA TCCTTAGAGT TTGGATCCTT -617

2751 TAGGGCAGGC AGGCACAGGG ACCCAGTGCG AGATCAGTGA AGCCGCCCAG -567

2801 TTTCGGCTTC CGCTCTTTTT CCACGCCCAC TTGCGTGCTT CTCCAACAGT -517

2851 GTGGATGGGA GGGGTGGGGG ACGAGCCCTA ATCTCCGAGG AAGGGGTGTG -467

⇒

2901 GCCCCGTTCG TGTTCTCCAG TTTGTGGCGT CCTGGATCTG TCCTCTGGTC -417

2951 CCCTCCAGAT CGTGTCCCAC ACCCACCCGT TCAGGCATGG CACTGTGCCG -367

3001 CCACGCGTGA CCGTGCGCTC CTTACGTGGG GGACGTGCAG GGTGCTGCCT -317

⇒

3051 CCTTTCCGGT GCGGGAGGGA GCGGCCGTCT TTCTCCTGCT CTGGCTGGGA -267

3101 AGCCCCAGCC ATTGCGCTGC AGAGGAGACT TGCAGCCAAT GGGGACTGAG -217

⇒

3151 GAAGTGGGCC GGCTGGCGGT TGTCACCCTC CCGGGGACCG GAGCTCCGAG -167

⇒

3201 GTCTGGAGAG CGCAGGCAGA CGCCCGCCCC GCCCGGGGAC TGAGGGGGAG -117

3251 GAGCGAAGGG AGGAGGAGGT GGAGTCTCCG ATCTGCCGCT GGAGGACCAC -67

3301 TGCTCACCAG GGCTACTGAG GAGCCACTGG CCCCACACCT GCTTTTCCGC -17

+1

3351 ATCCCCCACC GTCAGCATGA TCGCCGCGCA ACTACTGGCC TATTACTTCA +34

M I A A Q L L A Y Y F T

3401 CCGAGCTGAA GGATGACCAA GTCAAAAAGG TGAGCCCCGC CGGCGCCGCC +84

E L K D D Q V K K

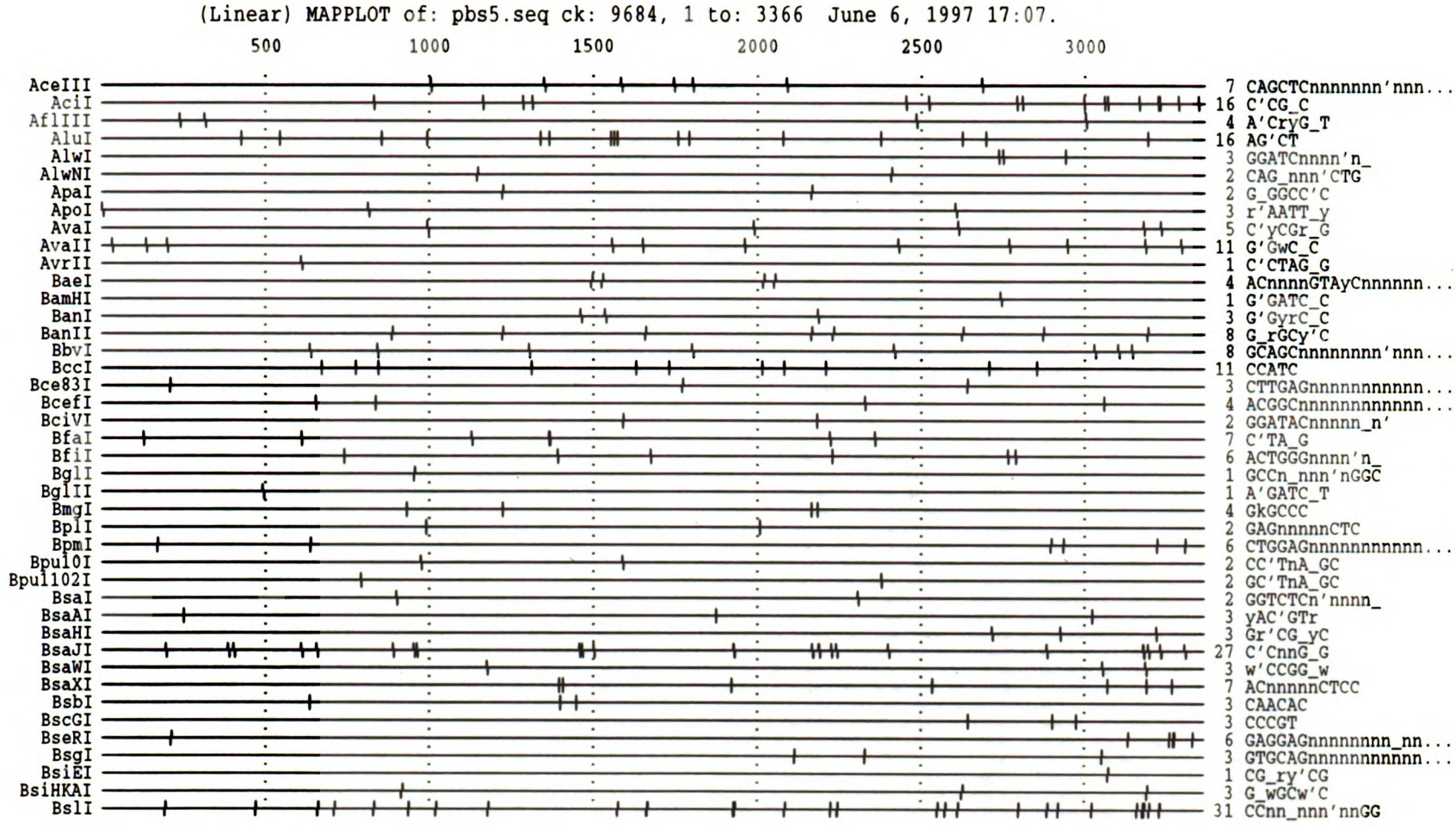
⇒ Represent 5' end points of HKI upstream sequence in promoter-reporter constructs.

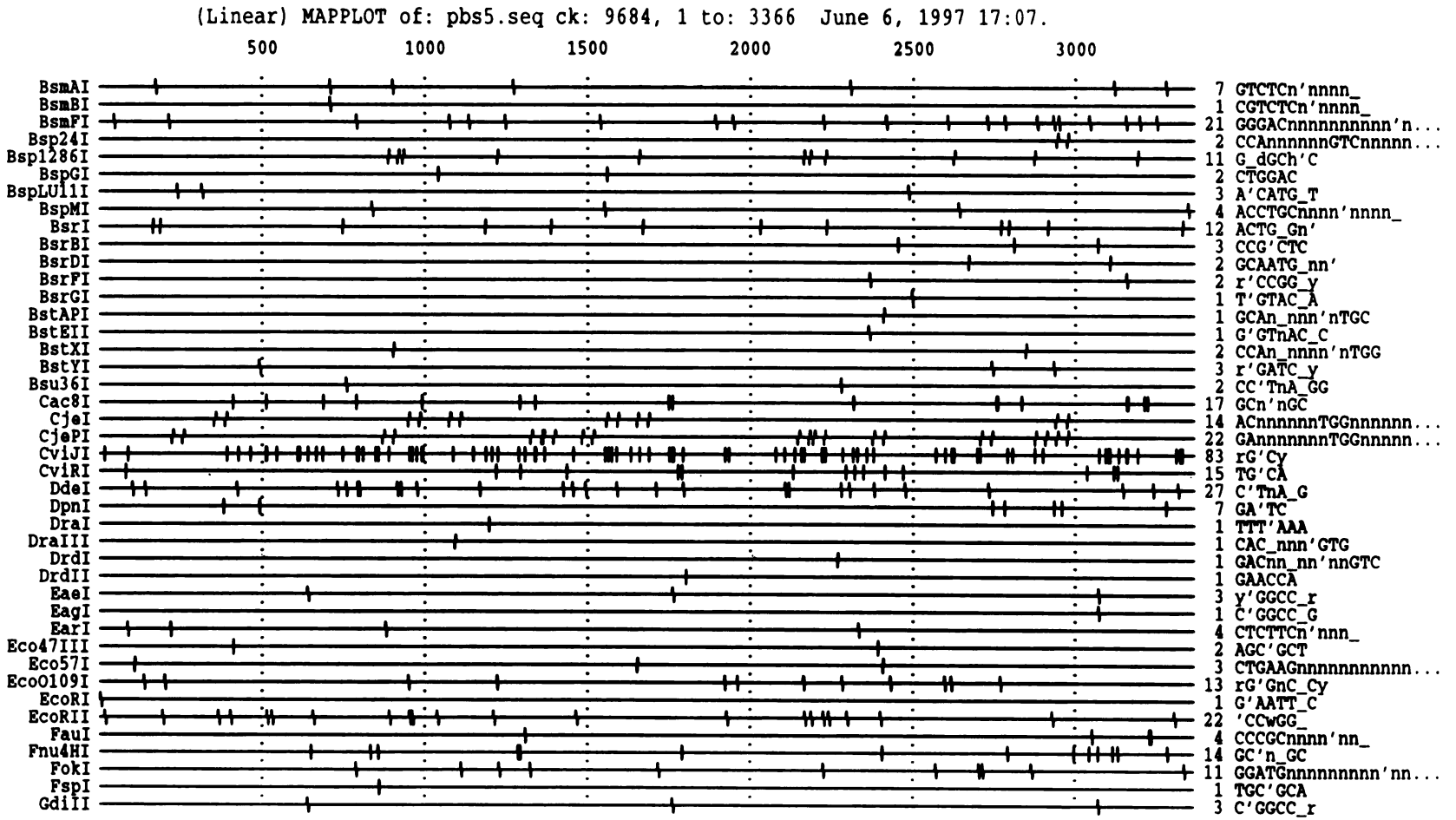
APPENDIX C

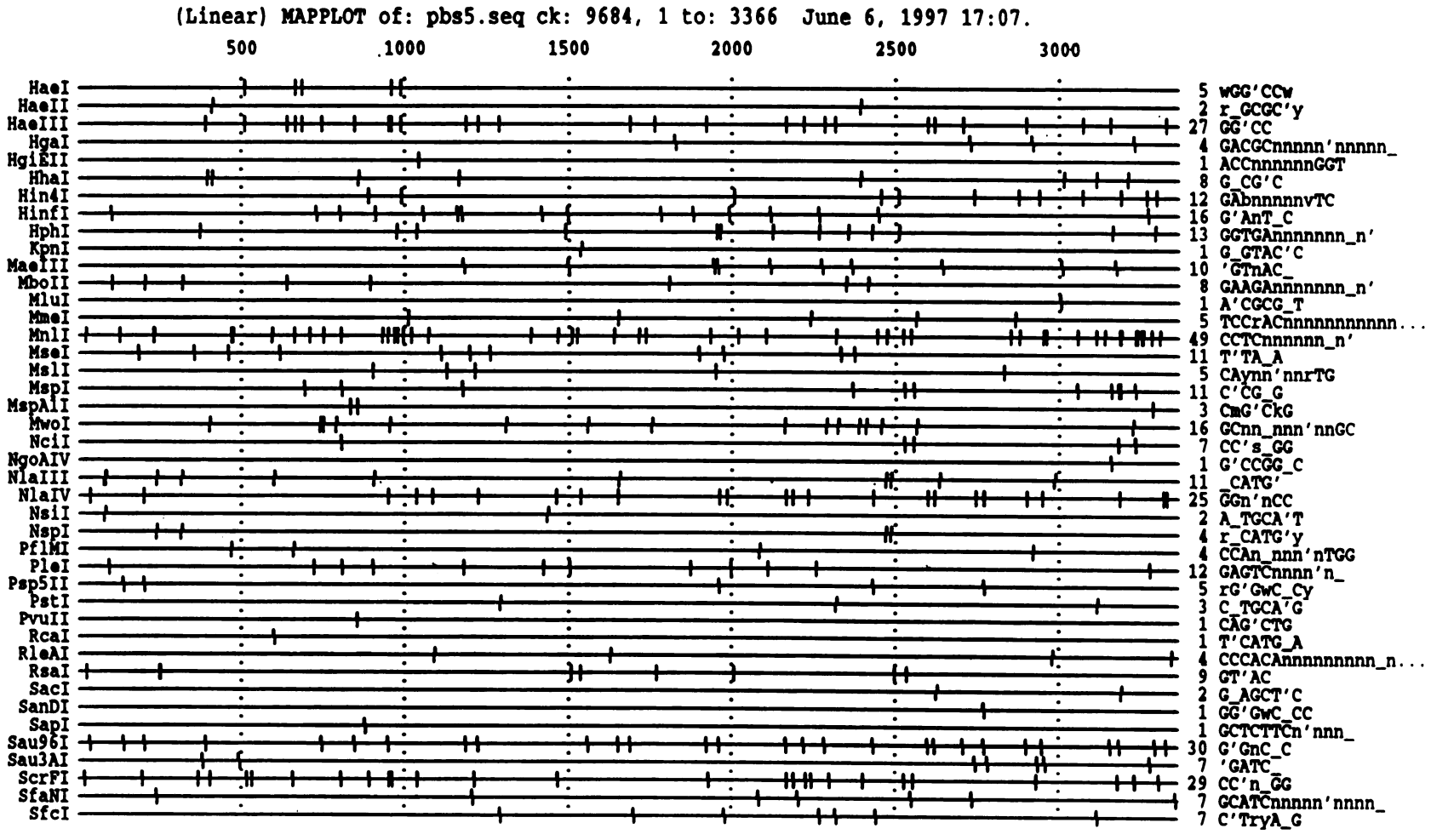
APPENDIX C

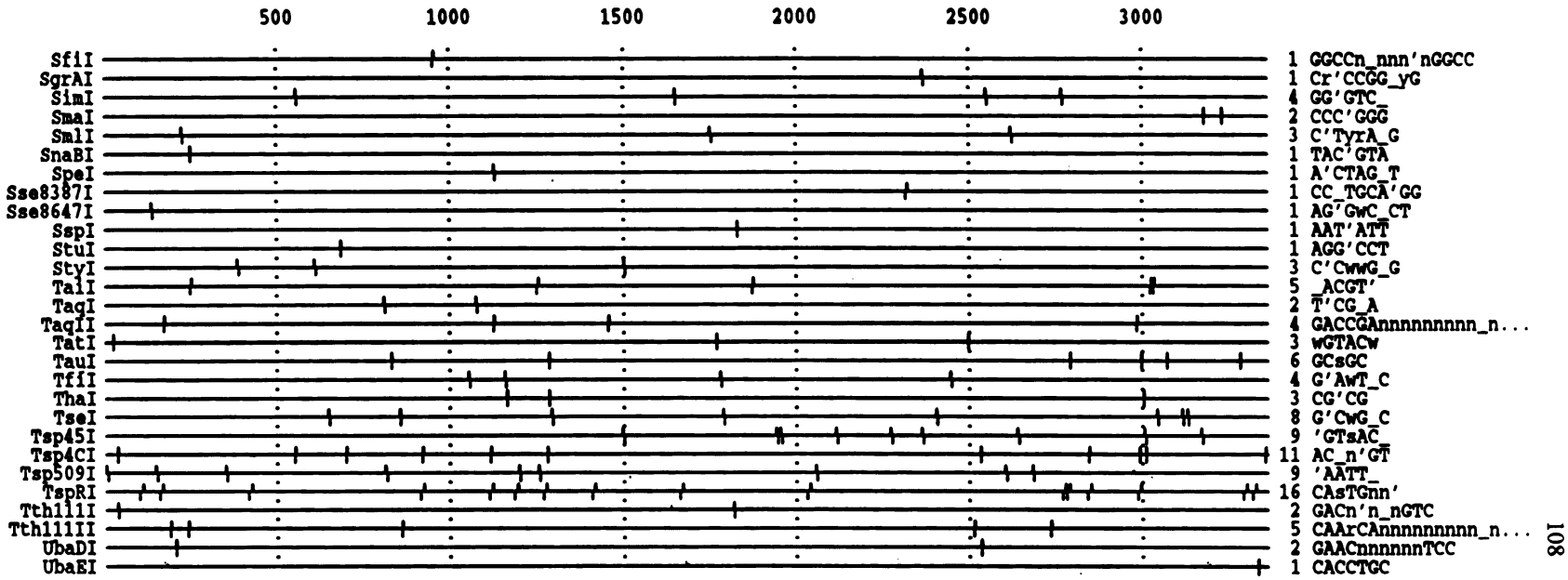
SCHEMATIC REPRESENTATION OF RESTRICTION SITES IN HKI UPSTREAM REGION

pBS39.5 (1 TO 3366)


(Linear) MAPPLOT of: pbs5.seq ck: 9684, 1 to: 3366 June 6, 1997 17:07.
500
1000
1500
2000
2500
3000
AceIII 7 CAGCTCnnnnnnn'nnn...
AciI 16 C'CG_C
AflIII 4 A'CryG_T
AluI 16 AG'CT
AlwI 3 GGATCnnnn'n_
AlwNI 2 CAG_nnn'CTG
ApaI 2 G_GGCC'C
ApoI 3 r'AATT_y
AvaI 5 C'yCGr_G
AvaII 11 G'GwC_C
AvrII 1 C'CTAG_G
BaeI 4 ACnnnnGTAyCnnnnnn...
BamHI 1 G'GATC_C
BanI 3 G'GyrC_C
BanII 8 G_rGCy'C
BbvI 8 GCAGCnnnnnnnn'nnn...
BccI 11 CCATC
Bce83I 3 CTTGAGnnnnnnnnnnn...
BcefI 4 ACGGCnnnnnnnnnnn...
BciVI 2 GGATACnnnnn_n'
BfaI 7 C'TA_G
BfiI 6 ACTGGGnnnn'n_
BglI 1 GCCn_nnn'nGGC
BglII 1 A'GATC_T
BmgI 4 GkGCCC
BplI 2 GAGnnnnnCTC
BpmI 6 CTGGAGnnnnnnnnnn...
Bpu10I 2 CC'TnA_GC
Bpu1102I 2 GC'TnA_GC
BsaI 2 GGTCTCn'nnnn_
BsaAI 3 yAC'GTr
BsaHI 3 Gr'CG_yC
BsaJI 27 C'CnnG_G
BsaWI 3 w'CCGG_w
BsaXI 7 ACnnnnnCTCC
BsbI 3 CAACAC
BscGI 3 CCCGT
BseRI 6 GAGGAGnnnnnnnn_nn...
BsgI 3 GTGCAGnnnnnnnnnn...
BsiEI 1 CG_ry'CG
BsiHKAI 3 G_wGCw'C
BslI 31 CCnn_nnn'nnGG


(Linear) MAPPLOT of: pbs5.seq ck: 9684, 1 to: 3366 June 6, 1997 17:07.
500
1000
1500
2000
2500
3000
BsmAI 7 GTCTCn'nnnn_
BsmBI 1 CGTCTCn'nnnn_
BsmFI 21 GGGACnnnnnnnnnn'n...
Bsp24I 2 CCAnnnnnnGTCnnnnn...
Bsp1286I 11 G_dGCh'C
BspGI 2 CTGGAC
BspLU11I 3 A'CATG_T
BspMI 4 ACCTGCnnnn'nnnn_
BsrI 12 ACTG_Gn'
BsrBI 3 CCG'CTC
BsrDI 2 GCAATG_nn'
BsrFI 2 r'CCGG_y
BsrGI 1 T'GTAC_A
BstAPI 1 GCAn_nnn'nTGC
BstEII 1 G'GTnAC_C
BstXI 2 CCAn_nnnn'nTGG
BstYI 3 r'GATC_y
Bsu36I 2 CC'TnA_GG
Cac8I 17 GCn'nGC
CjeI 14 ACnnnnnnTGGnnnnnn...
CjePI 22 GAnnnnnnnTGGnnnnn...
CviJI 83 rG'Cy
CviRI 15 TG'CA
DdeI 27 C'TnA_G
DpnI 7 GA'TC
DraI 1 TTT'AAA
DraIII 1 CAC_nnn'GTG
DrdI 1 GACnn_nn'nnGTC
DrdII 1 GAACCA
EaeI 3 y'GGCC_r
EagI 1 C'GGCC_G
EarI 4 CTCTTCn'nnn_
Eco47III 2 AGC'GCT
Eco57I 3 CTGAAGnnnnnnnnnnn...
Eco0109I 13 rG'GnC_Cy
EcoRI 1 G'AATT_C
EcoRII 22 'CCwGG_
FauI 4 CCCGCnnnn'nn_
Fnu4HI 14 GC'n_GC
FokI 11 GGATGnnnnnnnnn'nn...
FspI 1 TGC'GCA
GdiII 3 C'GGCC_r


(Linear) MAPPLOT of: pbs5.seq ck: 9684, 1 to: 3366 June 6, 1997 17:07.
500
1000
1500
2000
2500
3000
HaeI 5 wGG'CCw
HaeII 2 r_GCGC'y
HaeIII 27 GG'CC
HgaI 4 GACGCnnnnn'nnnnn_
HgiEII 1 ACCnnnnnnGGT
HhaI 8 G_CG'C
Hin4I 12 GAbnnnnnnvTC
HinfI 16 G'AnT_C
HphI 13 GGTGAnnnnnnn_n'
KpnI 1 G_GTAC'C
MaeIII 10 'GTnAC_
MboII 8 GAAGAnnnnnnn_n'
MluI 1 A'CGCG_T
MmeI 5 TCCrACnnnnnnnnnnn...
MnlI 49 CCTCnnnnnn_n'
MseI 11 T'TA_A
MslI 5 CAynn'nnrTG
MspI 11 C'CG_G
MspAlI 3 CmG'CkG
MwoI 16 GCnn_nnn'nnGC
NciI 7 CC's_GG
NgoAIV 1 G'CCGG_C
NlaIII 11 _CATG'
NlaIV 25 GGn'nCC
NsiI 2 A_TGCA'T
NspI 4 r_CATG'y
PflMI 4 CCAn_nnn'nTGG
PleI 12 GAGTCnnnn'n_
Psp5II 5 rG'GwC_Cy
PstI 3 C_TGCA'G
PvuII 1 CAG'CTG
RcaI 1 T'CATG_A
RleAI 4 CCCACAnnnnnnnnn_n...
RsaI 9 GT'AC
SacI 2 G_AGCT'C
SanDI 1 GG'GwC_CC
SapI 1 GCTCTTCn'nnn_
Sau96I 30 G'GnC_C
Sau3AI 7 'GATC_
ScrFI 29 CC'n_GG
SfaNI 7 GCATCnnnnn'nnnn_
SfcI 7 C'TryA_G



Enzymes that do not cut:

AatII	AccI	AflII	AhdI	ApaLI	AscI	BbsI	BcgI	BclI	BsaBI	BsmI	BspEI
BssHII	BssSI	BstDSI	BstZ17I	ClaI	EciI	EcoNI	EcoRV	FseI	HincII	HindIII	HpaI
MscI	MunI	NarI	NcoI	NdeI	NheI	NotI	NruI	NspV	PacI	Pfl1108I	PinAI
PmeI	PmlI	PshAI	Psp1406I	PvuI	RsrII	SacII	SalI	ScaI	SexAI	SgfI	SmiI
		SphI	SrfI	SunI	VspI	XbaI	XcmI	XhoI	XmnI		

APPENDIX D

APPENDIX D

LIST OF RESTRICTION SITES IN HKI UPSTREAM REGION

pBS39.5 (1 TO 3366)

AceIII CAGCTCnnnnnnn'nnnn_

Cuts at: 0 1003 1351 1585 1746 1803 2089 2685 3366
Size: 1003 348 234 161 57 286 596 681

AciI C'CG_C

Cuts at: 0 832 1163 1286 1315 2454 2523 2793 2810
Size: 832 331 123 29 1139 69 270 17

Cuts at: 2810 2998 3061 3071 3166 3224 3229 3286 3347
Size: 188 63 10 95 58 5 57 61

Cuts at: 3347 3366
Size: 19

AflIII A'CryG_T

Cuts at: 0 238 314 2485 3003 3366
Size: 238 76 2171 518 363

AluI AG'CT

Cuts at: 0 426 545 855 993 1341 1368 1555 1564
Size: 426 119 310 138 348 27 187 9

Cuts at: 1564 1575 1760 1793 2079 2379 2627 2699 3193
Size: 11 185 33 286 300 248 72 494

Cuts at: 3193 3366
Size: 173

AlwI GGATCnnnn'n_

Cuts at: 0 2738 2751 2942 3366
Size: 2738 13 191 424

AlwNI CAG_nnn'CTG

Cuts at: 0 1148 2412 3366
Size: 1148 1264 954

ApaI G_GGCC'C

Cuts at: 0 1226 2169 3366
Size: 1226 943 1197

ApoI r'AATT_y

Cuts at: 0 1 814 2605 3366
Size: 1 813 1791 761

AvaI C'yCGr_G

Cuts at: 0 994 1988 2613 3180 3232 3366
Size: 994 994 625 567 52 134

AvaII G'GwC_C

Cuts at: 0 31 135 199 1558 1651 1961 2431 2769
Size: 31 104 64 1359 93 310 470 338

Cuts at: 2769 2947 3185 3294 3366
Size: 178 238 109 72

AvrII C'CTAG_G

Cuts at: 0 609 3366
Size: 609 2757

BaeI GrTACnnnnGTnnnnnnnnnn_nnnnn'

Cuts at: 0 1498 1531 2022 2055 3366
Size: 1498 33 491 33 1311

BamHI G'GATC_C

Cuts at: 0 2743 3366
Size: 2743 623

BanI G'GyrC_C

Cuts at: 0 1462 1536 2185 3366
Size: 1462 74 649 1181

BanII G_rGCy'C

Cuts at: 0 890 1226 1662 2169 2235 2629 2877 3195
Size: 890 336 436 507 66 394 248 318

Cuts at: 3195 3366
Size: 171

BbvI GCAGCnnnnnnnn'nnnn_

Cuts at: 0 636 842 1305 1802 2418 3031 3103 3144
Size: 636 206 463 497 616 613 72 41

Cuts at: 3144 3366
Size: 222

BccI CCATC

Cuts at: 0 672 776 845 1313 1631 1732 2016 2083
Size: 672 104 69 468 318 101 284 67

Cuts at: 2083 2211 2709 2856 3366
Size: 128 498 147 510

Bce83I CTTGAGnnnnnnnnnnnnnn_nn'

Cuts at: 0 209 1774 2643 3366
Size: 209 1565 869 723

BcefI ACGGCnnnnnnnnnnnn'n_

Cuts at: 0 654 837 2330 3060 3366
Size: 654 183 1493 730 306

BciVI GGATACnnnnn_n'

Cuts at: 0 1593 2184 3366
Size: 1593 591 1182

BfaI C'TA_G

Cuts at: 0 127 610 1129 1365 1369 2223 2360 3366
Size: 127 483 519 236 4 854 137 1006

BfiI ACTGGGnnnn'n_

Cuts at: 0 741 1393 1676 2230 2766 2790 3366
Size: 741 652 283 554 536 24 576

BglI GCCn_nnn'nGGC

Cuts at: 0 957 3366
Size: 957 2409

BglII A'GATC_T

Cuts at: 0 491 3366
Size: 491 2875

BmgI GkGCCC

Cuts at: 0 932 1224 2167 2186 3366
Size: 932 292 943 19 1180

BplI GAGnnnnnCTC

Cuts at: 0 990 2008 3366
Size: 990 1018 1358

BpmI CTGGAGnnnnnnnnnnnnnn_nn'

Cuts at: 0 171 639 2900 2938 3224 3310 3366
Size: 171 468 2261 38 286 86 56

Bpu10I CC'TnA_GC

Cuts at: 0 974 1590 3366
Size: 974 616 1776

Bpu1102I GC'TnA_GC

Cuts at: 0 791 2380 3366
Size: 791 1589 986

BsaI GGTCTCn'nnnn_

Cuts at: 0 899 2308 3366
Size: 899 1409 1058

BsaAI yAC'GTr

Cuts at: 0 250 1876 3025 3366
Size: 250 1626 1149 341

BsaHI Gr'CG_yC

Cuts at: 0 2718 2927 3220 3366
Size: 2718 209 293 146

BsaJI C'CnnG_G

Cuts at: 0 194 385 402 609 656 889 952 961
Size: 194 191 17 207 47 233 63 9

Cuts at: 961 1458 1466 1502 1928 1929 2168 2169 2189
Size: 497 8 36 426 1 239 1 20

Cuts at: 2189 2190 2227 2243 2401 2885 3180 3181 3196
Size: 1 37 16 158 484 295 1 15

Cuts at: 3196 3232 3233 3307 3366
Size: 36 1 74 59

BsaWI w'CCGG_w

Cuts at: 0 1175 3055 3187 3366
Size: 1175 1880 132 179

BsaXI ACnnnnnCTCC

Cuts at: 0 1397 1410 1923 2536 3073 3191 3270 3366
Size: 1397 13 513 613 537 118 79 96

BsbI CAACAC

Cuts at: 0 636 1401 1450 3366
Size: 636 765 49 1916

BscGI CCCGT

Cuts at: 0 2645 2904 2977 3366
Size: 2645 259 73 389

BseRI GAGGAGnnnnnnnn_nn'

Cuts at: 0 212 3137 3263 3275 3278 3333 3366
Size: 212 2925 126 12 3 55 33

BsgI GTGCAGnnnnnnnnnnnnnn_nn'

Cuts at: 0 2116 2332 3056 3366
Size: 2116 216 724 310

BsiEI CG_ry'CG

Cuts at: 0 3075 3366
Size: 3075 291

BsiHKAI G_wGCw'C

Cuts at: 0 919 2629 3195 3366
Size: 919 1710 566 171

BslI CCnn_nnn'nnGG

Cuts at: 0 194 471 662 712 713 834 939 1022
Size: 194 277 191 50 1 121 105 83

Cuts at: 1022 1182 1577 1667 1668 1930 1934 2088 2228
Size: 160 395 90 1 262 4 154 140

Cuts at: 2228 2249 2250 2554 2577 2616 2802 2891 2923
Size: 21 1 304 23 39 186 89 32

Cuts at: 2923 3026 3165 3182 3187 3202 3233 3234 3366
Size: 103 139 17 5 15 31 1 132

BsmAI GTCTCn'nnnn_

Cuts at: 0 172 707 899 1271 2308 3119 3279 3366
Size: 172 535 192 372 1037 811 160 87

BsmBI CGTCTCn'nnnn_

Cuts at: 0 707 3366
Size: 707 2659

BsmFI GGGACnnnnnnnnnn'nnnn_

Cuts at: 0 44 212 788 1074 1135 1246 1539 1893
Size: 44 168 576 286 61 111 293 354

Cuts at: 1893 1947 2225 2417 2606 2729 2782 2882 2933
Size: 54 278 192 189 123 53 100 51

Cuts at: 2933 2949 3044 3156 3198 3250 3366
Size: 16 95 112 42 52 116

Bsp24I CCAnnnnnnGTCnnnnnnnn_nnnnn'

Cuts at: 0 2947 2979 3366
Size: 2947 32 387

Bsp1286I G_dGCh'C

Cuts at: 0 890 919 934 1226 1662 2169 2188 2235
Size: 890 29 15 292 436 507 19 47

Cuts at: 2235 2629 2877 3195 3366
Size: 394 248 318 171

BspGI CTGGAC

Cuts at: 0 1041 1561 3366
Size: 1041 520 1805

BspLU11I A'CATG_T

Cuts at: 0 238 314 2485 3366
Size: 238 76 2171 881

BspMI ACCTGCnnnn'nnnn_

Cuts at: 0 836 1552 2640 3346 3366
Size: 836 716 1088 706 20

BsrI ACTG_Gn'

Cuts at: 0 166 188 748 1188 1389 1672 2034 2237
Size: 166 22 560 440 201 283 362 203

Cuts at: 2237 2773 2797 2917 3331 3366
Size: 536 24 120 414 35

BsrBI CCG'CTC

Cuts at: 0 2454 2812 3071 3366
Size: 2454 358 259 295

BsrDI GCAATG_nn'

Cuts at: 0 2674 3109 3366
Size: 2674 435 257

BsrFI r'CCGG_y

Cuts at: 0 2367 3158 3366
Size: 2367 791 208

BsrGI T'GTAC_A

Cuts at: 0 2496 3366
Size: 2496 870

BstAPI GCAn_nnn'nTGC

Cuts at: 0 2412 3366
Size: 2412 954

BstEII G'GTnAC_C

Cuts at: 0 2363 3366
Size: 2363 1003

BstXI CCAn_nnnn'nTGG

Cuts at: 0 905 2850 3366
Size: 905 1945 516

BstYI r'GATC_y

Cuts at: 0 491 2743 2934 3366
Size: 491 2252 191 432

Bsu36I CC'TnA_GG

Cuts at: 0 757 2279 3366
Size: 757 1522 1087

Cac8I GCn'nGC

Cuts at: 0 409 512 687 788 991 1291 1339 1749
Size: 409 103 175 101 203 300 48 410

Cuts at: 1749 1762 2319 2757 2761 2835 3160 3164 3214
Size: 13 557 438 4 74 325 4 50

Cuts at: 3214 3224 3366
Size: 10 142

CjeI CCAnnnnnnGTnnnnnnnnn_nnnnnn'

Cuts at: 0 357 391 952 986 1078 1079 1112 1113
Size: 357 34 561 34 92 1 33 1

Cuts at: 1113 1563 1597 1657 1691 2946 2980 3366
Size: 450 34 60 34 1255 34 386

CjePI CCAnnnnnnnTCnnnnnnnn_nnnnnn'

Cuts at: 0 227 260 874 907 1330 1363 1368 1401
Size: 227 33 614 33 423 33 5 33

Cuts at: 1401 1487 1520 2153 2186 2200 2233 2382 2415
Size: 86 33 633 33 14 33 149 33

Cuts at: 2415 2713 2746 2881 2914 2947 2980 3366
Size: 298 33 135 33 33 33 386

CviJI rG'Cy

Cuts at: 0 14 86 391 426 463 510 514 545
Size: 14 72 305 35 37 47 4 31

Cuts at: 545 608 615 641 666 685 746 790 795
Size: 63 7 26 25 19 61 44 5

Cuts at: 795 809 847 855 888 951 960 973 989
Size: 14 38 8 33 63 9 13 16

Cuts at: 989 993 1086 1148 1187 1205 1224 1289 1306
Size: 4 93 62 39 18 19 65 17

Cuts at: 1306 1337 1341 1368 1457 1555 1564 1575 1589
Size: 31 4 27 89 98 9 11 14

Cuts at: 1589 1633 1660 1689 1751 1760 1764 1793 1923

Size: 44 27 29 62 9 4 29 130

Cuts at: 1923 1934 2079 2106 2139 2158 2167 2221 2226
Size: 11 145 27 33 19 9 54 5

Cuts at: 2226 2233 2285 2317 2330 2359 2379 2570 2599
Size: 7 52 32 13 29 20 191 29

Cuts at: 2599 2620 2627 2699 2707 2792 2806 2875 2901
Size: 21 7 72 8 85 14 69 26

Cuts at: 2901 3074 3094 3102 3108 3135 3158 3162 3193
Size: 173 20 8 6 27 23 4 31

Cuts at: 3193 3312 3323 3330 3366
Size: 119 11 7 36

CviRI TG'CA

Cuts at: 0 79 1219 1293 1437 1778 1790 2133 2294
Size: 79 1140 74 144 341 12 343 161

Cuts at: 2294 2321 2349 2415 2469 3037 3119 3132 3366
Size: 27 28 66 54 568 82 13 234

DdeI C'TnA_G

Cuts at: 0 100 139 422 730 757 791 797 913
Size: 100 39 283 308 27 34 6 116

Cuts at: 913 924 974 1168 1424 1453 1494 1590 1710
Size: 11 50 194 256 29 41 96 120

Cuts at: 1710 1794 2108 2119 2279 2305 2380 2475 2733
Size: 84 314 11 160 26 75 95 258

Cuts at: 2733 3146 3240 3316 3366
Size: 413 94 76 50

DpnI GA'TC

Cuts at: 0 381 493 2745 2783 2936 2959 3281 3366
Size: 381 112 2252 38 153 23 322 85

DraI TTT'AAA

Cuts at: 0 1198 3366
Size: 1198 2168

DraIII CAC_nnn'GTG

Cuts at: 0 1095 3366

Size: 1095 2271

DrdI GACnn_nn'nnGTC

Cuts at: 0 2271 3366
Size: 2271 1095

DrdII GAACCA

Cuts at: 0 1803 3366
Size: 1803 1563

EaeI y'GGCC_r

Cuts at: 0 639 1762 3072 3366
Size: 639 1123 1310 294

EagI C'GGCC_G

Cuts at: 0 3072 3366
Size: 3072 294

EarI CTCTTCn'nnn_

Cuts at: 0 85 217 878 2333 3366
Size: 85 132 661 1455 1033

Eco47III AGC'GCT

Cuts at: 0 412 2394 3366
Size: 412 1982 972

Eco57I CTGAAGnnnnnnnnnnnnnn_nn'

Cuts at: 0 109 1655 2411 3366
Size: 109 1546 756 955

EcoO109I rG'GnC_Cy

Cuts at: 0 135 199 949 1222 1922 1961 2165 2166
Size: 135 64 750 273 700 39 204 1

Cuts at: 2166 2283 2431 2598 2618 2769 3366
Size: 117 148 167 20 151 597

EcoRI G'AATT_C

Cuts at: 0 1 3366
Size: 1 3365

EcoRII 'CCwGG_

Cuts at: 0 14 193 365 402 514 530 655 889
Size: 14 179 172 37 112 16 125 234

Cuts at: 889 951 960 1037 1211 1465 1928 2168 2189
Size: 62 9 77 174 254 463 240 21

Cuts at: 2189 2226 2243 2298 2400 2930 3306 3366
Size: 37 17 55 102 530 376 60

FauI CCCGCnnnn'nn_

Cuts at: 0 1308 3054 3231 3236 3366
Size: 1308 1746 177 5 130

Fnu4HI GC'n_GC

Cuts at: 0 650 832 856 1287 1294 1791 2407 2793
Size: 650 182 24 431 7 497 616 386

Cuts at: 2793 2998 3045 3072 3117 3133 3286 3366
Size: 205 47 27 45 16 153 80

FokI GGATGnnnnnnnnn'nnnn_

Cuts at: 0 786 1111 1228 1323 1718 2225 2570 2703
Size: 786 325 117 95 395 507 345 133

Cuts at: 2703 2715 2866 3336 3366
Size: 12 151 470 30

FspI TGC'GCA

Cuts at: 0 859 3366
Size: 859 2507

GdiII C'GGCC_r

Cuts at: 0 639 1762 3072 3366
Size: 639 1123 1310 294

HaeI wGG'CCw

Cuts at: 0 510 666 685 960 989 3366
Size: 510 156 19 275 29 2377

HaeII r_GCGC'y

Cuts at: 0 414 2396 3366
Size: 414 1982 970

HaeIII GG'CC

Cuts at: 0 391 510 641 666 685 746 847 951
Size: 391 119 131 25 19 61 101 104

Cuts at: 951 960 989 1187 1224 1289 1689 1764 1923
Size: 9 29 198 37 65 400 75 159

Cuts at: 1923 2167 2221 2285 2317 2599 2620 2707 2901
Size: 244 54 64 32 282 21 87 194

Cuts at: 2901 3074 3158 3330 3366
Size: 173 84 172 36

HgaI GACGCnnnnn'nnnnn_

Cuts at: 0 1826 2726 2916 3228 3366
Size: 1826 900 190 312 138

HgiEII ACCnnnnnnGGT

Cuts at: 0 1042 3366
Size: 1042 2324

HhaI G_CG'C

Cuts at: 0 398 413 860 1167 2395 3017 3116 3212
Size: 398 15 447 307 1228 622 99 96

Cuts at: 3212 3366
Size: 154

Hin4I GAbnnnnnvTC

Cuts at: 0 890 990 2008 2456 2511 2741 2877 2939
Size: 890 100 1018 448 55 230 136 62

Cuts at: 2939 3073 3190 3270 3299 3366
Size: 134 117 80 29 67

HinfI G'AnT_C

Cuts at: 0 98 728 801 909 1055 1159 1172 1416
Size: 98 630 73 108 146 104 13 244

Cuts at: 1416 1496 1781 1881 1991 2117 2265 2446 3272
Size: 80 285 100 110 126 148 181 826

Cuts at: 3272 3366
Size: 94

HphI GGTGAnnnnnnn_n'

Cuts at: 0 374 977 1037 1491 1955 1966 2126 2267
Size: 374 603 60 454 464 11 160 141

Cuts at: 2267 2357 2429 2512 3165 3296 3366
Size: 90 72 83 653 131 70

KpnI G_GTAC'C

Cuts at: 0 1540 3366
Size: 1540 1826

MaeIII 'GTnAC_

Cuts at: 0 1178 1497 1943 1954 2114 2273 2363 2639
Size: 1178 319 446 11 160 159 90 276

Cuts at: 2639 3006 3171 3366
Size: 367 165 195

MboII GAAGAnnnnnnn_n'

Cuts at: 0 102 204 320 639 895 1810 2350 2417
Size: 102 102 116 319 256 915 540 67

Cuts at: 2417 3366
Size: 949

MluI A'CGCG_T

Cuts at: 0 3003 3366
Size: 3003 363

MmeI TCCrACnnnnnnnnnnnnnnnnn_nn'

Cuts at: 0 1012 1654 2242 2565 2867 3366
Size: 1012 642 588 323 302 499

MnlI CCTCnnnnnn_n'

Cuts at: 0 21 126 230 233 469 475 595 663
Size: 21 105 104 3 236 6 120 68

Cuts at: 663 710 752 806 933 950 969 979 996
Size: 47 42 54 127 17 19 10 17

Cuts at: 996 1021 1072 1386 1468 1511 1527 1640 1717
Size: 25 51 314 82 43 16 113 77

Cuts at: 1717 1738 1935 2021 2105 2320 2445 2474 2524
Size: 21 197 86 84 215 125 29 50

Cuts at: 2524 2548 2852 2880 2952 2962 3058 3115 3141

Size: 24 304 28 72 10 96 57 26

Cuts at: 3141 3187 3191 3235 3241 3253 3256 3259 3285
Size: 46 4 44 6 12 3 3 26

Cuts at: 3285 3311 3366
Size: 26 55

MseI TTA_A

Cuts at: 0 182 354 459 617 1110 1197 1259 1899
Size: 182 172 105 158 493 87 62 640

Cuts at: 1899 1974 2332 2374 3366
Size: 75 358 42 992

MslI CAynn'nnrTG

Cuts at: 0 903 1129 1214 1952 2832 3366
Size: 903 226 85 738 880 534

MspI C'CG_G

Cuts at: 0 692 806 1176 2368 2527 2555 3056 3159
Size: 692 114 370 1192 159 28 501 103

Cuts at: 3159 3181 3188 3233 3366
Size: 22 7 45 133

MspA1I CmG'CkG

Cuts at: 0 834 855 3288 3366
Size: 834 21 2433 78

MwoI GCnn_nnn'nnGC

Cuts at: 0 404 743 752 792 957 1312 1561 1757
Size: 404 339 9 40 165 355 249 196

Cuts at: 1757 2164 2291 2327 2390 2412 2460 2567 3228
Size: 407 127 36 63 22 48 107 661

Cuts at: 3228 3366
Size: 138

NciI CC's_GG

Cuts at: 0 806 2527 2555 3181 3182 3233 3234 3366
Size: 806 1721 28 626 1 51 1 132

NgoAIV G'CCGG_C

Cuts at: 0 3158 3366
Size: 3158 208

NlaIII _CATG'

Cuts at: 0 79 83 242 318 602 908 1658 2473
Size: 79 4 159 76 284 306 750 815

Cuts at: 2473 2489 2637 2989 3366
Size: 16 148 352 377

NlaIV GGn'nCC

Cuts at: 0 32 200 950 1036 1085 1224 1464 1538
Size: 32 168 750 86 49 139 240 74

Cuts at: 1538 1652 1963 1986 2167 2187 2234 2433 2600
Size: 114 311 23 181 20 47 199 167

Cuts at: 2600 2619 2745 2770 2771 2902 2949 3186 3322
Size: 19 126 25 1 131 47 237 136

Cuts at: 3322 3331 3366
Size: 9 35

NsiI A_TGCA'T

Cuts at: 0 81 1439 3366
Size: 81 1358 1927

NspI r_CATG'y

Cuts at: 0 242 318 2473 2489 3366
Size: 242 76 2155 16 877

PflMI CCAn_nnn'nTGG

Cuts at: 0 471 662 2088 2923 3366
Size: 471 191 1426 835 443

PleI GAGTCnnnn'n_

Cuts at: 0 92 722 809 903 1180 1424 1504 1875
Size: 92 630 87 94 277 244 80 371

Cuts at: 1875 1999 2111 2259 3280 3366
Size: 124 112 148 1021 86

Psp5II rG'GwC_Cy

Cuts at: 0 135 199 1961 2431 2769 3366
Size: 135 64 1762 470 338 597

PstI C_TGCA'G

Cuts at: 0 1295 2323 3121 3366
Size: 1295 1028 798 245

PvuII CAG'CTG

Cuts at: 0 855 3366
Size: 855 2511

RcaI T'CATG_A

Cuts at: 0 598 3366
Size: 598 2768

RleAI CCCACAnnnnnnnnn_nnn'

Cuts at: 0 1093 1631 2983 3349 3366
Size: 1093 538 1352 366 17

RsaI GT'AC

Cuts at: 0 22 248 252 1508 1538 1771 2009 2498
Size: 22 226 4 1256 30 233 238 489

Cuts at: 2498 2536 3366
Size: 38 830

SacI G_AGCT'C

Cuts at: 0 2629 3195 3366
Size: 2629 566 171

SanDI GG'GwC_CC

Cuts at: 0 2769 3366
Size: 2769 597

SapI GCTCTTCn'nnn_

Cuts at: 0 878 3366
Size: 878 2488

Sau96I G'GnC_C

Cuts at: 0 31 135 199 389 745 846 949 1185
Size: 31 104 64 190 356 101 103 236

Cuts at: 1185 1222 1223 1558 1651 1688 1922 1961 2165
Size: 37 1 335 93 37 234 39 204

Cuts at: 2165 2166 2219 2283 2431 2598 2618 2705 2769
Size: 1 53 64 148 167 20 87 64

Cuts at: 2769 2900 2947 3156 3185 3294 3329 3366
Size: 131 47 209 29 109 35 37

Sau3AI 'GATC_

Cuts at: 0 379 491 2743 2781 2934 2957 3279 3366
Size: 379 112 2252 38 153 23 322 87

ScrFI CC'n_GG

Cuts at: 0 16 195 367 404 516 532 657 806
Size: 16 179 172 37 112 16 125 149

Cuts at: 806 891 953 962 1039 1213 1467 1930 2170
Size: 85 62 9 77 174 254 463 240

Cuts at: 2170 2191 2228 2245 2300 2402 2527 2555 2932
Size: 21 37 17 55 102 125 28 377

Cuts at: 2932 3181 3182 3233 3234 3308 3366
Size: 249 1 51 1 74 58

SfaNI GCATCnnnnn'nnnn_

Cuts at: 0 238 1206 2082 2203 2548 2734 3358 3366
Size: 238 968 876 121 345 186 624 8

SfcI C'TryA_G

Cuts at: 0 1291 1700 1979 2269 2319 2441 3117 3366
Size: 1291 409 279 290 50 122 676 249

SfiI GGCCn_nnn'nGGCC

Cuts at: 0 957 3366
Size: 957 2409

SgrAI Cr'CCGG_yG

Cuts at: 0 2367 3366
Size: 2367 999

SimI GG'GTC_

Cuts at: 0 556 1651 2551 2769 3366
Size: 556 1095 900 218 597

SmaI CCC'GGG

Cuts at: 0 3182 3234 3366
Size: 3182 52 132

SmlI C'TyrA_G

Cuts at: 0 224 1753 2622 3366
Size: 224 1529 869 744

SnaBI TAC'GTA

Cuts at: 0 250 3366
Size: 250 3116

SpeI A'CTAG_T

Cuts at: 0 1128 3366
Size: 1128 2238

Sse8387I CC_TGCA'GG

Cuts at: 0 2323 3366
Size: 2323 1043

Sse8647I AG'GwC_CT

Cuts at: 0 135 3366
Size: 135 3231

SspI AAT'ATT

Cuts at: 0 1832 3366
Size: 1832 1534

StuI AGG'CCT

Cuts at: 0 685 3366
Size: 685 2681

StyI C'CwwG_G

Cuts at: 0 385 609 1502 3366
Size: 385 224 893 1864

TaiI _ACGT'

Cuts at: 0 252 1256 1878 3027 3036 3366
Size: 252 1004 622 1149 9 330

TaqI T'CG_A

Cuts at: 0 811 1077 3366
Size: 811 266 2289

TaqII GACCGAnnnnnnnnn_nn'

Cuts at: 0 172 1128 1458 2986 3366
Size: 172 956 330 1528 380

TatI wGTACw

Cuts at: 0 22 1771 2498 3366
Size: 22 1749 727 868

TauI GCsGC

Cuts at: 0 832 1287 2793 2998 3072 3286 3366
Size: 832 455 1506 205 74 214 80

TfiI G'AwT_C

Cuts at: 0 1055 1159 1781 2446 3366
Size: 1055 104 622 665 920

ThaI CG'CG

Cuts at: 0 1165 1286 3005 3366
Size: 1165 121 1719 361

TseI G'CwG_C

Cuts at: 0 649 855 1293 1790 2406 3044 3116 3132
Size: 649 206 438 497 616 638 72 16

Cuts at: 3132 3366
Size: 234

Tsp45I 'GTsAC_

Cuts at: 0 1497 1943 1954 2114 2273 2363 2639 3006
Size: 1497 446 11 160 159 90 276 367

Cuts at: 3006 3171 3366
Size: 165 195

Tsp4CI AC_n'GT

Cuts at: 0 35 552 701 921 1118 1282 2534 2848
Size: 35 517 149 220 197 164 1252 314

Cuts at: 2848 2994 3012 3360 3366
Size: 146 18 348 6

Tsp509I 'AATT_

Cuts at: 0 1 146 351 814 1199 1256 2057 2605
Size: 1 145 205 463 385 57 801 548

Cuts at: 2605 2683 3366
Size: 78 683

TspRI CAsTGnn'

Cuts at: 0 108 166 426 924 1123 1195 1279 1418
Size: 108 58 260 498 199 72 84 139

Cuts at: 1418 1672 2041 2780 2791 2853 2997 3304 3331
Size: 254 369 739 11 62 144 307 27

Cuts at: 3331 3366
Size: 35

Tth111I GACn'n_nGTC

Cuts at: 0 35 1820 3366
Size: 35 1785 1546

Tth111II CAArCAnnnnnnnnn_nn'

Cuts at: 0 189 242 863 2516 2738 3366
Size: 189 53 621 1653 222 628

UbaDI GAACnnnnnnTCC

Cuts at: 0 205 2535 3366
Size: 205 2330 831

UbaEI CACCTGC

Cuts at: 0 3338 3366
Size: 3338 28

Enzymes that do cut:

AceIII	AciI	AflIII	AluI	AlwI	AlwNI	ApaI	ApoI
AvaI	AvaII	AvrII	BaeI	BamHI	BanI	BanII	BbvI
BccI	Bce83I	BcefI	BciVI	BfaI	BfiI	BglI	BglII
BmgI	BplI	BpmI	Bu10I	Bpu1102I	BsaI	BsaAI	BsaHI
BsaJI	BsaWI	BsaXI	BsbI	BscGI	BseRI	BsgI	BsiEI
BsiHKAI	BslI	BsmAI	BsmBI	BsmFI	Bsp24	Bsp1286I	BspGI
BspLU11I	BspMI	BsrI	BsrBI	BsrDI	BsrFI	BsrGI	BstAPI
BstEII	BstXI	BstYI	Bsu36I	Cac8I	CjeI	CjePI	CviJI
CviRI	DdeI	DpnI	DraI	DraIII	DrdI	DrdII	EaeI
EagI	EarI	Eco47III	Eco57I	EcoO109I	EcoRI	EcoRII	FauI
Fnu4HI	FokI	FspI	GdiII	HaeI	HaeII	HaeIII	HgaI
HgiEII	HhaI	Hin4I	HinfI	HphI	KpnI	MaeIII	MboII
MluI	MmeI	MnlI	MseI	MslI	MspI	MspA1I	MwoI
NciI	NgoAIV	NlaIII	NlaIV	NsiI	NspI	PflMI	PleI
Psp5II	PstI	PvuII	RcaI	RleAI	RsaI	SacI	SanDI
SapI	Sau96I	Sau3AI	ScrFI	SfaNI	SfcI	SfiI	SgrAI
SimI	SmaI	SmlI	SnaBI	SpeI	Sse8387I	Sse8647I	SspI
StuI	StyI	TaiI	TaqI	TaqII	TatI	TauI	TfiI
ThaI	TseI	Tsp45I	Tsp4CI	Tsp509I	TspRI	Tth111I	Tth111II
UbaDI	UbaEI						

Enzymes that do not cut:

AatII	AccI	AflII	AhdI	ApaLI	AscI	BbsI	BcgI
BclI	BsaBI	BsmI	BspEI	BssHII	BssSI	BstDSI	BstZ17I
ClaI	EciI	EcoNI	EcoRV	FseI	HincII	HindIII	HpaI
MscI	MunI	NarI	NcoI	NdeI	NheI	NotI	NruI
NspV	PacI	Pfl1108I	PinAI	PmeI	PmlI	PshAI	Psp1406I
PvuI	RsrII	SacII	SalI	ScaI	SexAI	SgfI	SmiI
SphI	SrfI	SunI	VspI	XbaI	XcmI	XhoI	XmnI


MICHIGAN STATE UNIV. LIBRARIES
31293017164181