This is to certify that the

dissertation entitled

TYPE II HEXOKINASE:
MOLECULAR CLONING, SEQUENCE AND PROMOTER ANALYSIS

presented by

Annette P. Thelen

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Biochemistry

Major professor

John E. Wilson
Professor and Chair.
Dept. of Biochemistry

Date April 7, 1992

TYPE II HEXOKINASE:

MOLECULAR CLONING, SEQUENCE AND PROMOTER ANALYSIS

By

Annette P. Thelen

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1992

ABSTRACT

TYPE II HEXOKINASE: MOLECULAR CLONING, SEQUENCE AND PROMOTER ANALYSIS

By

Annette P. Thelen

The 917-residue amino acid sequence of Type II hexokinase has been deduced from the overlapping nucleotide sequences of two clones, isolated from rat skeletal muscle cDNA libraries. The sequences of 197 nucleotides in the 5' untranslated region and 687 bases of the 3' untranslated region have also been determined from cloned cDNA. There is extensive similarity between the sequences of the N- and C-terminal halves of the Type II isozyme, as previously seen with the Type I isozyme; this is consistent with the view that these enzymes evolved by a process of gene duplication and fusion.

The region of overlap between the two discrete cDNA clones, was confirmed by isolation and sequencing of a genomic DNA clone that spanned the region. Within this region, the 634-nucleotide coding sequence was divided into three exons, each of 150-250 nucleotides. In addition, the predominant transcription start site was located in a region 465 nucleotides upstream from the translation initiation codon. The genomic sequence upstream from the transcription start site was found to contain several potential promoter elements. These include a TATA-like sequence, 2 CCAAT sequences, and 3 Sp1 binding sites. Alignment of genomic clones for Type

II hexokinase predict a minimum gene length of at least 35 kb. These results suggest that the gene encoding Type II hexokinase is likely to be quite complex.

A cDNA encoding the entire C-terminal half of a hexokinase from Novikoff ascites tumor cells was also isolated and found to be identical to a cDNA encoding the corresponding region of the Type II isozyme of skeletal muscle. Northern analysis indicated that a single mRNA, approximately 5200 nucleotides in length, encoded both the skeletal muscle and the tumor enzymes. These results do not support previous speculation that the hexokinase isozymes of normal tissue are distinct from those of tumor, and suggest the possibility that post-translational modifications of a single protein species might account for apparent differences between the isozymes of normal and tumor tissues.

To my family

ACKNOWLEDGEMENTS

I want to thank my family, especially my husband, Bill, without whose emotional support the completion of this degree would not have been possible. My children, Greg and Sarah, have been very patient and supportive through the easy and hard times. I want to thank John Wilson, for his guidance and intellectual support throughout this project. He has taught me much by his example, allowing me to develop and mature as a scientist. When things looked the darkest, he kept me looking for the dawn. I would like to acknowledge my committee whose advice has been most appreciated. They are: Susan Conrad, Steve Triezenberg, Clarence Suelter, and Jerry Dodgson. I want to give a special thanks to Patty Voss, Joe Leykam, Marion Healy, Linda Sherwood, Marcia Kieliszewski, and Al Smith for all their help and support. Lastly, my deepest appreciation to Dr. Dave McConnell, whose encouragement and assistance led to the fulfillment of a lifelong goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

LIST OF ABBREVIATIONS

BCA	bicinchoninic acid
bp	basepair
BSA	bovine serum albumin
DEAE	diethylaminoethyl
DMEM	Dulbecco's modified eagle's medium
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	disodium, ethylenediamine tetraacetate
EtBr	ethidium bromide
Glc-6-P	glucose 6-phosphate
Glc	glucose
GuHCl	guanidine hydrochloride
Hepes	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
kb	kilobase
kDa	kilodalton
Pipes	piperazine-N,N'-bis[2-ethane-sulfonic acid]
SDS	sodium dodecyl sulfate
SSC	saline sodium citrate
TG	monothioglycerol
Tris	tris[hydroxymethyl]aminomethane

CHAPTER 1

Introduction

Prologue

The phosphorylation of glucose by hexokinase yields glucose-6-phosphate. Because glucose-6-phosphate is a substrate in many metabolic pathways, hexokinase is one of the major points of regulation in the metabolism of glucose in mammals. The mammalian hexokinase family consists of several distinct isozymes, each with different biochemical properties. The major aim of this work was to determine the amino acid sequence of Type II hexokinase from rat muscle, permitting comparison with other rat isozymes and thereby contributing to our understanding of the structural differences between these isozymes that bring about their regulatory diversity. This chapter will provide background information on the localization, regulatory properties, expression, structure, and evolutionary aspects of the hexokinase isozyme family.

MAMMALIAN HEXOKINASES

The conversion of glucose to glucose-6-phosphate in mammalian tissues is catalyzed by hexokinase (ATP: D-hexose 6 phosphotransferase, EC 2.7.1.1.) using ATP as the phosphoryl donor. Since glucose-6-phosphate is the initial substrate in many metabolic pathways such as glycogen biosynthesis and glycolysis, it is not surprising that hexokinase is under complex regulation via product inhibition and mitochondrial binding. The study of the function, structure, and regulation of hexokinase is made more complex by the fact that there are at least four isozymes in mammalian tissues.

Background

The multiple isozymes, first observed in liver in 1963 by Viñuela (1) and Walker (2), can be separated from one another by either chromatographic (3) or electrophoretic (4) techniques. The isozymes are named A-D, or alternatively I-IV, based on their order of elution from a DEAE-cellulose column (3), or their order of increasing mobility during starch gel electrophoresis (5), respectively.

Although each of these isozymes has the same catalytic function, they differ significantly from one another in size, tissue distribution, and details of regulation. Several of these characteristics, discussed below, are summarized in Table I. Isozymes I-III are monomers with approximate molecular weights of 100 kDa and have low K_ms for glucose (0.02-0.13 mM). Isozyme IV (glucokinase, EC 2.7.1.2) is

Table I. Comparison of several parameters for the mammalian hexokinase isozymes[a].

	HEXOKINASE ISOZYME			
PARAMETER	I	II	III	IV
M_r,approx.	98,000	96,000	98,000	49,000
Source[c]	brain erythrocytes	muscle adipocytes	spleen lung	liver pancreas
K_m Glc	0.04[b]	0.13	0.02	4.5
K_m ATP	0.42	0.07	1.29	0.49
K_i Glc-6-P vs ATP	0.026	0.021	0.074	15

a-Table adapted from Ureta (57), and references therein.
b-All apparent kinetic constant values are expressed in mM.
c-Some examples are shown; this is not an all inclusive list.

a monomer of 50 kDa and has a much higher K_m for glucose (4.5 mM) (6-9). The apparent molecular weight of isozyme IV is similar to that of Types L1 and L2 hexokinase from wheat germ (10) and Types A and B hexokinase from yeast (23).

Tissue Distribution

A large variation in the hexokinase isozyme levels within various tissues has been observed (11). Generally, there is more than one hexokinase isozyme present in most tissues. Type I hexokinase is found in virtually all tissues and can be considered the "primary" hexokinase. Since Type I hexokinase is the most prevalent member of this family, much of the biochemical information available about the hexokinase reaction has been obtained from the study of this isozyme. This isoform is the predominant hexokinase in tissues with heavy reliance on blood-borne glucose for energy, as in brain and erythrocytes. In contrast, Type II hexokinase is the major isozyme in insulin-sensitive tissues such as muscle, adipose and mammary gland. In these tissues, Glc-6-P can be directed into energy storage forms such as lipids (in adipose or mammary tissue) or glycogen (in muscle tissue). The predominance of Type II hexokinase in these tissues leads one to speculate that this isozyme plays an integral part in such energy storage pathways. The Type III isoform can be found in several tissues (e.g. spleen, kidney, and lung) but in much lower amounts than the other three isozymes. For this reason, less is known about Type III hexokinase than the other isozymes.

Early work indicated that Type IV hexokinase, or glucokinase, was present only in the liver (5). However, glucokinase has since been found in the islets of

Langerhans in the pancreas (12). More recently, a glucokinase was detected in anterior pituitary cells and pituitary cell lines (13). Both Northern and Western blot analysis, using antibodies against the pancreatic glucokinase, detected glucokinase in the pituitary cell line AtT20. No glucokinase mRNA or enzyme activity was detectable in the pituitary tissue, even though the antibodies against the pancreatic enzyme did react with a protein band of the appropriate size in the Western analysis. Hughes *et al.* (13) postulated that an undetectably low amount of mRNA might be sufficient for expression of glucokinase since the protein half-life is as much as 30 h. However, Liang *et al.* (14) reported detecting glucokinase mRNA, but no enzyme activity, in both the pituitary cell line AtT20 and pituitary tissue. These investigators felt that alteration of the open reading frame for glucokinase may explain the lack of enzyme activity in the pituitary.

Influences on Hexokinase Levels

While the relative amount of each isozyme is influenced by age, diet, and hormones as well as physical activity, isozymes II and IV appear to be most affected. Bernstein and Kipnis investigated the effects of age and diet on Types I and II hexokinase activity in rat skeletal muscle and adipose tissues (15). They found that the hexokinase in both tissues decreased with age. Similar decreases in hexokinase activity were found in muscle and adipose tissues of young rats after a 3-day fast. These investigators determined that, since the levels of Type I hexokinase in both tissues were unaffected by age and diet, changes in the Type II isozyme were responsible for the alteration in hexokinase activity in muscle and adipose tissue due

to these conditions. These results were similar to those seen by Katzen and Schimke (5).

Under diabetic conditions, the level of Type II hexokinase decreases in fat pad, heart, and skeletal muscle as well as adipose tissue and mammary gland, and upon insulin treatment is returned to normal levels (16-18). Frank and Fromm used streptozotocin to induce diabetes in rats in order to investigate the effect of insulin on the synthesis (19) and degradation (20) of Type II hexokinase. These investigators monitored the incorporation of [^{3}H]leucine into Type II hexokinase in the skeletal muscle of diabetic and insulin treated diabetic rats. They found that the rate of synthesis of skeletal muscle Type II in normal rats was approximately 1.9 times greater than in diabetic animals. Insulin treatment of the diabetic animals brought the reduced synthesis rate of the Type II isozyme back to near normal levels (19). In a similar set of experiments (20), Frank and Fromm found that the degradation rate constant for Type II hexokinase was approximately 3 times greater in diabetic animals than in normal animals. In these experiments, the investigators did not determine if the changes in Type II hexokinase levels were related to the rate of transcription of its gene or the stability of its message. The half-life of this protein in diabetic muscle was 9 hr versus 28 hr in normal tissue. Insulin treatment of diabetic skeletal muscle restored both the rate of degradation and the half-life of the Type II isozyme to levels approximating those found in normal tissues.

The level of glucokinase in liver is influenced by several factors including insulin, diet, fasting (21) and glucagon (22). This is not the case for the glucokinase

found in the ß-cells of the pancreas, which appears to be influenced by changes in blood glucose concentration, and not insulin (reviewed in ref 24). Because insulin secretion from ß-cells is controlled through the glycolytic rate, the pancreatic glucokinase has been implicated in the control of insulin levels. These seemingly opposite controls can provide an effective regulation cycle for plasma glucose levels. Elevated glucose increases pancreatic glucokinase activity which in turn stimulates insulin release. This elevated insulin level stimulates hepatic glucokinase which will decrease blood glucose levels.

For a number of years, it has been known that exercise (25,26) and chronic stimulation (27,28) cause an increase in hexokinase activity in muscle. Weber and Pette, using [^{35}S]methionine and immunoprecipitation, monitored protein production in chronically stimulated rat skeletal muscle. They found that the increase in hexokinase activity in stimulated tissue was the result of increased Type II hexokinase synthesis (29). These observations were confirmed, and expanded upon, when Weber and Pette demonstrated that hexokinase activity reached a maximum peak after 14 days of chronic stimulation (30). When stimulation was withdrawn a decreased rate of synthesis and possibly an increased rate of degradation of the Type II isozyme (31) caused an immediate decline in both hexokinase activity and the levels of the Type II isozyme. These researchers also found that nearly 50% more total hexokinase activity was mitochondrially bound in stimulated muscle than in unstimulated tissue (30); the potential regulatory significance of the binding of hexokinase to mitochondria will be discussed below. The increases in both the Type II hexokinase protein levels and the

amount of mitochondrially bound hexokinase represent intracellular changes in response to the increased energy demand of chronic stimulation.

Regulation of Expression - the Glucokinase Gene

As previously discussed, several factors influence both the activity of glucokinase, and the level of the protein. Hormones, diet, age, and activity all appear to alter the amount of glucokinase, to differing degrees in different tissues (or cell types). Investigation into the regulation of glucokinase transcription became possible when the Granner laboratory isolated and characterized both the cDNA (32) and gene (33) for liver glucokinase. The gene encoding glucokinase is the only hexokinase gene studied in depth to date. The hepatic glucokinase gene was found to be 15.5 kb in length, and contained 10 exons ranging in size from 96 to 977 bp. Transcription initiation was localized over a 4 base range, with the strongest band 127 nucleotides upstream from the translation initiation codon. This translation initiation codon is located in exon 1. The message for hepatic glucokinase is 2.4 kb in length, encoding a protein of 465 residues.

Analysis of 5' flanking sequences located several promoter elements, including a TATA box and an Sp1 binding site. Also present were several elements found in other liver-specific genes. Using run-on transcription analysis, these investigators demonstrated that with insulin treatment, the rate of transcription of the hepatic glucokinase gene in diabetic rats increased approximately 20-fold in 2 hr, with a significant increase within 1 hr. Similar increases in the transcription of glucokinase mRNA in rat liver and hepatocytes have been reported by Sibrowski and Seitz (34),

and Iynedjian and coworkers (22), respectively. In both cases the half-life for glucokinase mRNA was determined to be 40-45 min.

The laboratory of Tanaka and coworkers (35) investigated the genomic region upstream from the transcription initiation site for liver glucokinase in an attempt to locate genomic regions important in the insulin regulation of this gene. Using the chloramphenicol acetytransferase (CAT) assay system connected to deletion constructs of the glucokinase gene (-5.5 kb to -48), these investigators determined that the genomic region, from immediately upstream of the transcription start site (+1), to nucleotide -87, was sufficient for promoter activity in rat hepatocytes. However, using the same experimental strategy, insulin treatment of the transformed hepatocytes resulted in no change in CAT activity of the cells. They concluded that the 5.5kb sequence 5' of the transcription initiation site for glucokinase did not contain insulin responsive elements. It is possible that such sequences may be further upstream or may be downstream contained within an intron in the glucokinase gene.

Magnuson and Shelton (36), in studies on the expression of glucokinase in pancreatic ß-cells, isolated and characterized the transcription unit of this isozyme from an insulinoma cell line. These investigators reported the size of the pancreatic glucokinase mRNA to be approximately 2600 nucleotides in length, about 250 longer that the message for liver glucokinase. Iynedjian and coworkers (37) reported a difference in glucokinase message size of 400 nucleotides, with the pancreatic message being longer (approx 2.8 kb). Similar message sizes have also been seen in both pituitary tissue and the AtT20 cell line (13).

Magnuson and Shelton (36) found that the promoter for this pancreatic enzyme was at least 12 kb upstream from the hepatic glucokinase promoter, making the transcription unit at least 27.5 kb. The 5' ends of the cDNAs for pancreatic and hepatic glucokinases were completely different, resulting in 15 different amino acids at their N-termini. The point at which the hepatic and pancreatic sequences diverged corresponded to a splice site between exons 1 and 2 in the hepatic transcript unit. However, all residues 3' of this splice site were identical between the hepatic and pancreatic glucokinases.

Several laboratories have detected a number of glucokinase variants in liver, pancreas, and pituitary cells and tissue. Magnuson and Shelton (36) found in pancreas a glucokinase cDNA with a 51 bp deletion which generated a 17 amino acid deletion (in frame) (14). Using PCR technology, Magnuson *et al.* (14) detected mRNAs, in both pancreas and liver, that contained the aforementioned 5' dissimilar regions and the 51 bp deletion. They also reported locating additional alternate splicing products in both pituitary tissue and the pituitary cell line, AtT20.

In recent work, Newgard and colleagues (13, 38) have identified several additional variant rat glucokinase transcripts in liver, pancreas, AtT20 cells and pituitary tissues. Expression of the pancreatic and liver variants containing a 52 bp deletion, at the 3' end of exon 2 of the glucokinase gene, resulted in no glucokinase activity (38). However, they could not detect the transcript containing the 51 bp deletion that Magnuson *et al.* had analyzed.

It is evident that there are at least two distinct promoters for the glucokinase gene that control tissue specific expression of unique rat glucokinases in the liver, pancreas, and pituitary. The significance of additional unique transcripts, resulting from several alternate splicing events, remains to be determined. As research into Types I-III hexokinase continues, similar complexity will undoubtedly be found in their gene structures. Preliminary characterization of the Type II gene, presented herein, supports this observation.

Intracellular Locations of Hexokinase

The association of hexokinase with the particulate fractions of tissue homogenates has been observed by many laboratories using a number of tissues (11). The actual subcellular structure, however, that bound the hexokinase(s) was not always determined. Several researchers have reported that the particulate hexokinase activity was associated with mitochondria, for most tissues tested. Significant portions of Type I and Type II hexokinase have been found associated with the mitochondria in tissues such as brain (11), as well as heart, diaphragm, skeletal muscle and mammary gland (39). Generally, Type III hexokinase has been found only in the soluble fraction of tissue homogenates such as lung (40) and hepatoma (41). One report of the association of Type III hexokinase with the mitochondria (42) may have incorrectly identified the Type II isozyme as Type III hexokinase (41). However, based on localization studies by Preller and Wilson (43), the Type III isozyme appears to have a weak association with nuclei of several different rat tissues. Many of the cell types exhibiting nuclear association of Type III were

endothelial or epithelial cells. One laboratory has reported the presence of glucokinase in the nuclei, as well as the cytoplasm, of parenchymal cells of rat liver (45).

The outer mitochondrial membrane protein to which Type I hexokinase binds was isolated by Felgner *et al.* (44). Subsequently, it was shown that this hexokinase binding protein was the pore-forming protein, porin (46,47). Hexokinase binding to porin, through which molecules such as ADP and ATP flow, creates a direct link between glycolysis in the cytosol and ATP production in the mitochondria. This association between hexokinase and porin gives the enzyme preferential access to mitochondrially generated ATP (48,49). Inui and Ishibashi (50) reported that the efficient utilization of mitochondrial ATP was dependent upon hexokinase being bound to the mitochondrial membrane. The bound form of the enzyme has a slightly greater affinity for ATP and is considerably less sensitive to Glc-6-P inhibition (40,51,52). These kinetic differences, coupled with preferential access to intramitochondrially generated ATP, have led to the suggestion that mitochondrial binding represents a mechanism for activation (relative to the unbound enzyme) of hexokinase.

Divalent cations and a hydrophobic "tail" at the N-terminus of the protein facilitate the association of hexokinase with the mitochondrial membrane. Divalent cations are known to enhance this protein-membrane association in brain (52), tumor (53), and skeletal muscle (49). It is likely that this occurs by bridging the negatively charged groups on both the protein and the membrane. Critical to the association of

Type I hexokinase with the mitochondria is the N-terminus of the enzyme. Polakis and Wilson (54) used limited chymotryptic digestion to demonstrate that the loss of the 9 hydrophobic residues at the N-terminus of Type I hexokinase resulted in the loss of binding to the mitochondrial membrane. Similar results were also reported by Rose and Warms (53). Further studies into the mitochondrial binding of Type I hexokinase, by Xie and Wilson (55), demonstrated that this hydrophobic N-terminus was actually inserted into a hydrophobic core in the mitochondrial membrane. Inspection of the N-terminal residues of the deduced amino acid sequence for Type III hexokinase (56) reveals that this region of Type III is much less hydrophobic than the N-terminus of Type I. The lack of such a hydrophobic region in Type III hexokinase is one factor that makes mitochondrial binding less likely, if not impossible.

Regulation of Hexokinase Activity

As with the regulation of the level of the hexokinases, the enzymatic activity of each hexokinase isozyme is regulated differently. Mammalian hexokinase Types I-III are sensitive to allosteric inhibition by Glc-6-P (11), whereas glucokinase shows sensitivity to this ligand only at concentrations higher than normal physiological levels (57). The inhibitory effect of Glc-6-P on the Type I isozyme is immediate which is in contrast to the delayed inhibition of Type II hexokinase by this ligand (58,59). The $t_{1/2}$ of the response of the Type II isozyme to Glc-6-P varied from 12 sec to 130 sec, for soluble or mitochondrially bound enzyme, respectively.

Other ligands affecting hexokinase activity are inorganic phosphate (P_i), Glc 1,6-P_2, and glucose. Inorganic phosphate reverses the inhibition of the Type I

isoform by Glc-6-P (11) but has little if any effect on the Type II isozyme (59). Glucose 1,6-bisphosphate is also an inhibitor of Type I hexokinase and has been shown to have an even greater affinity for the Type II isozyme (60,61). The substrate, glucose, is itself an inhibitor of Type III hexokinase at concentrations above 0.2mM (3), suggesting that the isozyme is likely to be active only when intracellular concentrations of glucose are low.

Tumor Hexokinase

For more than 60 years, it has been known that tumor cells exhibit increased glycolytic rates (62). Several changes in cellular metabolism in tumors are directly associated with increased glycolysis. Included in these changes are the increased levels of key enzymes such as hexokinase, and a change in the degree of hexokinase associated with the mitochondria. Bustamante and Pedersen showed that the transformation of liver cells into tumor cells increases the hexokinase activity by more than 20-fold (63). Using the hepatoma cell line H-91, these investigators found that at least half the total hexokinase activity was associated with the mitochondria, and that mitochondrially generated ATP was preferentially used to phosphorylate glucose. In a similar set of experiments utilizing the Ehrlich ascites tumor line, Bustamante *et al.* (64) found that approximately two-thirds of the hexokinase activity was mitochondrially associated. These authors also reported that fast growing tumor cells (those reaching maximum size within 1 month) have the highest levels of hexokinase activity, again with 70% of the activity associated with the mitochondria (64).

To gain further insight into the nature of tumor hexokinase, Nakashima and co-workers (65) purified and characterized a mitochondrially bound hexokinase from the AS-30D rat hepatoma cell line. They demonstrated that this cell line, as with those mentioned above, contained increased hexokinase with a significant portion bound to the mitochondria (when compared to normal liver). Anion exchange chromatography of liver and AS-30D hexokinases showed that two tumor isozymes coeluted with Types I and II from normal liver. These authors reported that the purified AS-30D hexokinase, which coeluted with the liver Type II hexokinase, had kinetic and chromatographic properties similar to those of the Type II isoform from normal tissues. A comparison of their amino acid compositions indicated that the tumor and liver isozymes were not identical. However, these researchers did note that the amino acid compositions were obtained from different laboratories using different techniques. Nakashima *et al.* concluded that, in the transformation from normal liver to hepatoma, there was a significant change in hexokinase content to an isozyme form with very different properties.

The complete amino acid sequence for a tumor hexokinase was first reported by Arora and coworkers (66). They isolated cDNA for a hexokinase from c37 mouse hepatoma cell line, using a partial cDNA for rat brain Type I hexokinase (67) as the probe. The deduced sequence for this mouse tumor hexokinase was the same length (918 residues) as the rat Type I isozyme (68), and the enzymes differed at only 32 positions in their amino acid sequences. The 12-amino acid hydrophobic stretches at the N-terminus of both hexokinases were identical. The presence of this hydrophobic

"tail" in the tumor hexokinase could account, at least in part, for the high level of mitochondrial binding seen with tumor hexokinases. The authors felt that their results could not conclusively identify the mouse hepatoma hexokinase as Types I, II or III. However, it seems likely that this hepatoma hexokinase is Type I from mouse with the 32-amino acid difference (only 3.5% of the total 918 residues) due to species variation (mouse *vs* rat) and not a reflection of the transformation process.

Evolution of the Mammalian Hexokinases

The fact that Types I-III hexokinases are twice the size of both glucokinase and the yeast isozymes has led several investigators to postulate that the mammalian hexokinases evolved through a process of gene duplication and fusion from an ancestral hexokinase similar to the yeast enzymes (23,57,69-72). The similarities in amino acid composition (57,73) and antigenic cross reactivities (41,74) among the various isozymes give support to this theory. This is not an unprecedented postulation since gene duplication and fusion are also suggested for the evolution of rabbit phosphofructokinase (75), glycogen phosphorylase (76), and the β-crystallin protein family (77).

However, in contrast to glucokinase and the yeast hexokinases (23), several organisms have 50 kDa hexokinases that are inhibited by Glc-6-P. These organisms include silkworm (78), locust (79) and starfish (80). Recent work (84) supports the current postulation that the starfish hexokinase, not the yeast monomer, may be a direct descendent of an ancestral hexokinase that contained both glucose and Glc-6-P binding sites (Fig. 1). It is possible that duplication and fusion of such an ancestral

Figure 1. Postulated evolution of mammalian hexokinases. The catalytic site is represented by a circle, and the Glc-6-P regulatory site by a square. In this scheme, a 50 kDa ancestral hexokinase, which was not inhibited by Glc-6-P, evolved in two directions, one leading to the present day yeast hexokinase (not inhibited by Glc-6-P). In the second pathway, a Glc-6-P site evolved on the 50 kDa protein, giving a protein with properties similar to those of the present-day starfish hexokinase. Gene duplication and fusion of the ancestral gene encoding the 50 kDa Glc-6-P sensitive hexokinase would give rise to the present-day 100 kDa mammalian hexokinases (84).

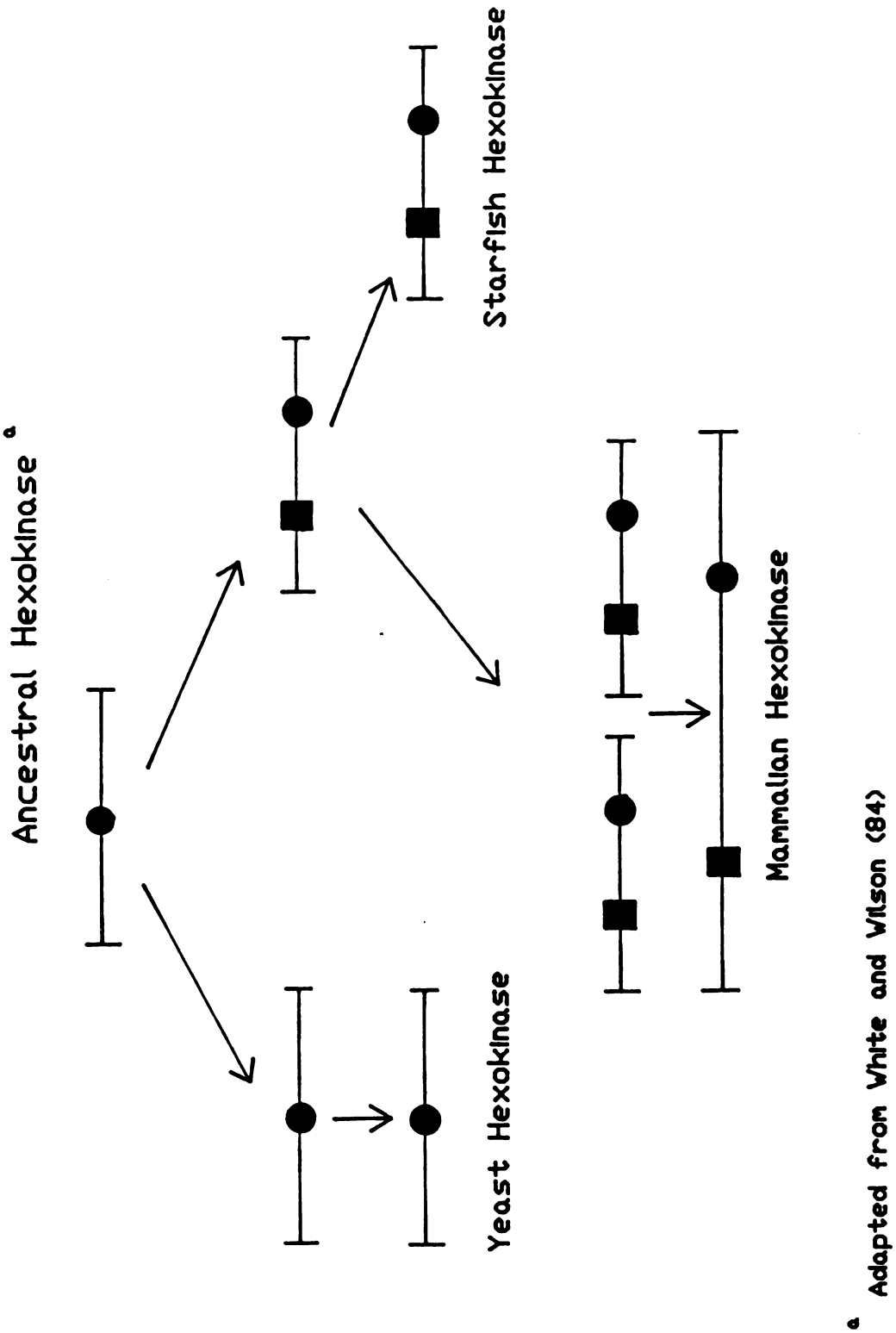

Ancestral Hexokinase a
Starfish Hexokinase
Yeast Hexokinase
Mammalian Hexokinase
a Adapted from White and Wilson (84)

FIGURE 1

gene gave rise to the genes encoding the current mammalian hexokinases. The extensive sequence similarities (Table II) of the N- and C-terminal halves of Types I and III hexokinase, to one another and to glucokinase, support this gene duplication and fusion theory. The degree of similarity among the deduced amino acid sequences of Type I hexokinase from human kidney (81), Type I isozyme from rat brain (68), Type III hexokinase (56) and glucokinase from rat liver (32), and a hexokinase from a mouse hepatoma cell line (66) lend additional support to this theory.

Structure-Function Relationship in Hexokinase

Since the work by Crane and Sols (82) which led to the view of a distinct Glc-6-P regulatory site in the hexokinase molecule, much research has been conducted to define and locate the catalytic and regulatory sites of hexokinase. The first substantive information about the location of the substrate binding sites was provided by Nemat-Gorgani and Wilson (83). Using a photoactivatable ATP analog (8-Azido-ATP) and limited tryptic proteolysis, these researchers demonstrated that the substrate nucleotide binding site was in the C-terminal portion of Type I hexokinase.

The location of the glucose binding site was also placed in the C-terminal domain of isozyme I. Schirch and Wilson (85) used a radiolabelled glucose analog (N-(bromoacetyl)-D-glucosamine (GlcNBrAc)) to modify sulfhydryl groups at the glucose binding site. Analysis of tryptically digested hexokinase, treated with [^{14}C]GlcNBrAc, showed that the C-terminal portion of the molecule had been labelled. A subsequent set of experiments (93) further defined the location of the regions that had been labelled by the glucose analog. Three tryptic fragments,

Table II. Comparison of Amino Acid Sequences of N- and C-Terminal Halves of Rat Type I and Rat Type III Hexokinases, and Glucokinase[a].

	NIII[b]	NI	CIII	CI
NI	39(16)[c]			
CIII	40(14)	45(15)		
CI	38(14)	46(17)	62(11)	
IV	38(15)	46(18)	49(15)	49(15)

a-Table adapted from Schwab and Wilson (56).

b-Abbreviations used: NIII, N-terminal half of rat Type III isozyme; NI, N-terminal of rat Type I isozyme; CIII, C-terminal half of rat Type III isozyme; CI, C-terminal half of rat Type I isozyme; IV, rat Type IV hexokinase (glucokinase).

c-Percent identical residues is shown without parenthesis; percent conservative substitutions is shown in parenthesis.

radiolabelled by GlcNBrAc treatment, were isolated and subjected to amino acid sequence analysis. Two of the fragments were quite similar to yeast sequences (86,87) that are located near the glucose binding site (88,89). The third peptide showed no significant similarity to other published hexokinase sequences. By comparing the two rat Type I peptides to the yeast sequences, Schirch and Wilson further localized at least a portion of the glucose binding site within a 5 kDa segment of the C-terminal half of the enzyme.

Under nondenaturing conditions, limited tryptic digestion of Type I hexokinase yields 3 specific fragments of M_r 10, 40 and 50 kDa (90). However, White and Wilson (91) found that the susceptibility of the Type I isozyme to trypsinization is increased when the enzyme is denatured with 0.6 M GuHCl. Furthermore, the ligand binding domains (N- and C-terminal halves) are selectively stabilized when the ligands, or their analogs, are present during denaturation and proteolysis. Using this selective protection from proteolysis as an indication of binding of the ligand, White and Wilson (84,91) demonstrated that the allosteric binding site for Glc-6-P was in the N-terminal of Type I hexokinase. The enzyme was denatured with GuHCl in the presence of analogs of either substrate or inhibitor. When the Glc-6-P analog, 1,5-anhydroglucitol-6-P, was used, the N-terminal half of Type I hexokinase was protected from digestion. Catalytic activity was lost under these conditions. In contrast, the substrate analog, N-acetylglucosamine, protected the C-terminal half from proteolysis, and catalytic function was retained. These results confirmed the previous findings that the catalytic domain was in the C-terminal portion of the

molecule, and provided direct evidence for the location of the allosteric regulatory site in the N-terminal half of the Type I isozyme.

Isolation and characterization of cDNA for the Type I isozyme from rat brain, by Schwab and Wilson (67,68), provided the first complete deduced amino acid sequence (918 residues in length) for a rat hexokinase. The sequences for yeast hexokinases A and B had previously been determined by two separate laboratories (86,87). Comparison of the rat Type I protein sequence with those of the yeast isozymes revealed extensive sequence similarities between these hexokinases, and internally between the N- and C-terminals halves of the rat isozyme. Schwab and Wilson (68) proposed a 3-D model for rat Type I hexokinase based on the yeast hexokinase x-ray structure of Steitz and coworkers (88,89,92). Each of the halves of the 100 kDa Type I isozyme were considered to be structurally similar to the yeast enzyme.

This model for mammalian hexokinase consists of two large domains, each composed of 2 lobes with a cleft between each lobe. The cleft in the C-terminal domain, shown to be responsible for catalytic function (84), contains the binding sites for the substrates glucose (85,93) and ATP-Mg^{2+} (the magnesium chelate of ATP) (83). Regulation of Type I hexokinase has been assigned to the N-terminal domain of the molecule which is not catalytically active (84) and contains the binding site for the allosteric effector, Glc-6-P (91,94). Considering the high degree of similarity between the deduced amino acid sequences for Types I and III hexokinase (56), it is

likely that this same structure-function arrangement is present in the Type III isozyme and, one may anticipate, in Type II hexokinase also.

The isolation and characterization of the Type II cDNA, as part of this project, is another step toward understanding how this protein's structure is related to its catalytic and regulatory properties. Comparison of the deduced amino acid sequences for the Types I, II and III isozymes demonstrates a high degree of similarity. Such sequence conservation clearly suggests that a structure-function arrangement similar to that found in the Type I enzyme is present in all of these isozymes. With the availability of the cDNAs for the hexokinase isozymes, it becomes possible to determine how the variation in the structures of these proteins lead to their distinctive catalytic/regulatory properties.

CHAPTER 2

Materials and Methods

MATERIALS

DNA modifying enzymes were obtained from Boehringer Mannheim Biochemicals (Indianapolis, IN), BRL (Gaithersburg, MD) or New England Biolabs (Beverly, MA). Radioisotopes were purchased from either NEN DuPont (Boston, MA) or Amersham (Arlington Heights, IL). Nitrocellulose from Schleicher and Schuell (Keene, NH) was used for library screenings and blot analyses. The primers used in cDNA synthesis and subsequent manipulations were either obtained from Boehringer Mannheim Biochemicals, Pharmacia (Piscataway, NJ), New England BioLabs, or U.S. Biochemicals (Cleveland, OH) or synthesized at the Macromolecular Structure, Sequencing, and Synthesis Facility (Michigan State University). DE-52 DEAE-cellulose marketed by Whatman (Clifton, NJ) and Affi-Gel Blue affinity chromatography gel from Bio-Rad Laboratories (Richmond, CA) were used in protein purification procedures. The BCA reagent for protein determination was obtained from Pierce Chemical Co. (Rockford, IL). Oligo (dT) cellulose and the Sequenase sequencing kit were purchased from Boehringer Mannheim Biochemicals and U.S. Biochemicals, respectively. AMV Reverse Transcriptase used in primer extension analysis was purchased from Life Sciences (St. Petersburg, FL). The Uni-ZAP cDNA synthesis kit was obtained from Stratagene (La Jolla, CA). The columns used to purify plasmid DNA for transfection were purchased from Qiagen, Inc. (Chatsworth, CA). U. S. Biochemicals was the source for T_1 RNase. All other reagents were obtained from standard commercial sources.

An amplified cDNA library, constructed in λgt10 using mRNA from adult rat soleus muscle was generously provided by Dr. F. H. Schachat, Duke University Medical Center. Two cDNA libraries, constructed in pUC 8 and pUC 9 from rat skeletal muscle mRNA, were kindly provided by Dr. D. M. Helfman, Cold Spring Harbor Laboratory. Total RNA isolated from electrically stimulated rat skeletal muscle which contains elevated levels of Type II hexokinase (29,31) was provided by Dr. D. Pette, University of Konstanz. A rat genomic library, constructed in λCharon 4A, was kindly made available by Dr. Tom Sargent, National Institutes of Health. The Novikoff ascites tumor cell line was obtained from the Fels Research Institute (Philadelphia, PA) with the assistance of Dr. Sidney Weinhouse, and propagated in female Sprague-Dawley rats obtained from Holtzman (Madison, WI).

METHODS

Standard procedures were used for library screening, DNA labeling, restriction enzyme mapping, subcloning, isolation of DNA, and Northern and Southern analyses. Unless otherwise noted, *in vitro* DNA manipulations were performed as previously described (95).

Purification and N-Terminal Sequencing of Type II Hexokinase from Rat Skeletal Muscle. Hexokinase activity was assayed spectrophotometrically, coupling the hexokinase reaction to NADPH formation using Glc-6-P dehydrogenase (96).

Hind limb skeletal muscle was obtained from adult Sprague-Dawley rats of either sex. The tissue was homogenized for 2 min in a Waring blender with 50 mM Tris, 1 mM EDTA, 50 mM Glc, 20 mM TG, pH 7.0 (2 ml buffer per g tissue). After

centrifugation for 30 min at 25,000 x g, the pH of the supernatant was adjusted to 7.0 with 0.1 M KOH, and extracted enzyme adsorbed batchwise onto DEAE-cellulose equilibrated with homogenization buffer; approx. one ml settled volume of DEAE cellulose per unit of enzyme was required for complete adsorption of activity. After exhaustive washing, the DEAE-cellulose was poured into a column. The column was washed with 2 column volumes of homogenization buffer, and hexokinase eluted with a linear gradient, 0-0.4 M in KCl. Fractions containing hexokinase activity eluting at 0.15 M-0.25 M KCl were combined, concentrated and dialyzed against the homogenization buffer. The enzyme was then loaded onto a 3.6 cm x 5.5 cm column of Affi-Gel Blue, equilibrated in the homogenization buffer. The column was sequentially washed with homogenization buffer of increasing pH, first 7.5 then 8.0 and finally pH 8.5; the pH 8.5 wash buffer included 20% (v/v) glycerol. In each case, washing was continued until the absorbance at 280 nm declined to a negligible value. Hexokinase was then eluted with 1.5 mM Glc-6-P in the pH 8.5 buffer, essentially as previously described for the Type I isozyme (96). At this point, the specific activity of the enzyme was about 10 units per mg protein, an approximately 50-fold increase over that in the initial extract.

SDS-gel electrophoresis on 6.5%-15% linear acrylamide gradient gels was performed as described previously (90,97). Type II hexokinase, migrating with an apparent mol. wt. of approx. 107 kDa (98), was well resolved from other components, only one of which - with apparent mol. wt. 66 kDa - was major. The Type II hexokinase band was excised from the gel, the enzyme electroeluted using a

CBS Scientific (Del Mar, CA) Model ELU-40 device, and prepared for sequencing as described by Hunkapillar *et al.* (99). Sequencing by automated Edman degradation was done by the Macromolecular Structure, Sequencing, and Synthesis Facility (Michigan State University).

cDNA Library Synthesis and Screenings. Two cDNA libraries were synthesized using the UniZAP II cDNA kit from Strategene, following manufacturer's directions. The sources of mRNA for these libraries were rat skeletal muscle and rat epididymal fat pad. Two cDNA libraries with λgt10 as the cloning vector were synthesized using mRNA from either normal or electrically stimulated rat skeletal muscle (29,31). The cDNAs for these libraries (one from normal tissue, the other from stimulated tissue) were synthesized using random hexanucleotide primers at a ratio of 0.5 μg primers per μg mRNA. Five micrograms of mRNA from normal tissue and 1 μg mRNA from stimulated tissue were used. The subsequent synthesis procedures were as described by DeWitt and Smith (100).

The prehybridization and hybridization solutions for all library screenings, consisted of deionized formamide (37% for medium stringency and 50% for high stringency requirements), 5x SSC, 5x Denhardt's reagent, 0.1% SDS, and 0.1 mg fragmented and denatured salmon sperm DNA per ml. Prehybridization was carried out at 42°C for at least 4 hours, and hybridization was performed at 42°C for at least 8 hours, and usually overnight. The hybridization solutions contained 10^5 cpm of radiolabelled probe per ml. The probes were radiolabelled using random hexanucleotide primers (101), to a specific activity greater than 10^8 cpm/ug. The

hybridized filters were washed at room temperature in 2x SSC/0.1% SDS, for two 15 minute periods. The filters were then washed either at 40°C (medium stringency) or 50°C (high stringency) in 0.1x SSC/0.1% SDS for 30 minutes, until no additional counts were removed from the filters. Positive recombinant clones were purified by multiple rescreening under the appropriate conditions.

Isolation of Genomic Clones Containing Type II Hexokinase Gene. The λCharon 4A library from Dr. T. Sargent was screened with several different probes from the cDNA for the Type II isozyme, under high stringency conditions. The procedure for these screenings was the same as mentioned in the above section on cDNA library screening. Southern blot analyses of genomic clones digested with *EcoR*I were used to determine relative placement of these clones with respect to cDNA sequences for Type II hexokinase. Five genomic clones, containing either pertinent sequence information or the largest inserts, were aligned as shown in Fig. 2. The radiolabelled cDNA probes used to isolate and characterize these genomic clones were obtained from the following regions of the Type II cDNAs (given by restriction sites and nucleotide position in the appropriate cDNA): *Eco*RI-*Xho*I (5' end of 12-1.3C - 152), *Eco*RI-*Bgl*II (5' end of 12-1.3C - 525), *Bgl*II-*Sty*I (525-1137), *Sty*I-*Eco*RI (1156 - 3' end of 12-1.3C), *Sph*I-*Eco*RI (307 - 3' end of RG2B), *Hae*III-*Hae*III (1312-1519) and *Hae*III-*Eco*RI (1698 - 3' end of RG2B). The first 4 cDNA fragments were isolated from cDNA clone 12-1.3C. The remaining 3 restriction fragments were isolated from cDNA clone RG2B. The *EcoR*I sites used in the above

Figure 2. Alignment of clones containing genomic DNA for rat Type II hexokinase. In the center is the composite figure which contains the coding region for Type II hexokinase, and the 5' and 3' untranslated regions of the cDNA. Shown above the composite figure are the overlapping cDNA clones, 12-1.3C and RG2B/NK3B. Below the composite drawing are the genomic clones isolated using portions of clones 12-1.3C and RG2B as the probes. The alignment of the genomic clones relative to the composite figure are the results of Southern blots probed with restriction fragments of cDNA clones, 12-1.3C and RG2B. The restriction sites used to generate the probes for these blots are designated by number below the cDNA clones. The restriction site positions in the Type II cDNAs are: 1, *EcoR*I at 5' end of 12-1.3C; 2, *Xho*I, 152; 3, *Bgl*II, 525; 4, *Sty*I, 1137/1156; 5, *EcoR*I at 3' end of 12-1.3C; 6, *Sph*I, 307; 7-9 *Hae*III, 1312, 1519, and 1698, respectively; and 10, *EcoR*I at 3' end of RG2B/NK3B. Sites 1-5 are in clone 12-1.3C; sites 6-10 are located in clones RG2B/NK3B. Slash marks (//) indicate that these clones are longer than represented, and are not drawn to scale with the composite figure for the enzyme. The dashed line at the 5' end of clone 5G3A indicates sequence upstream of the cDNA clone 12-1.3C, portions of which were used to determine the transcription initiation sites and promoter sequences.

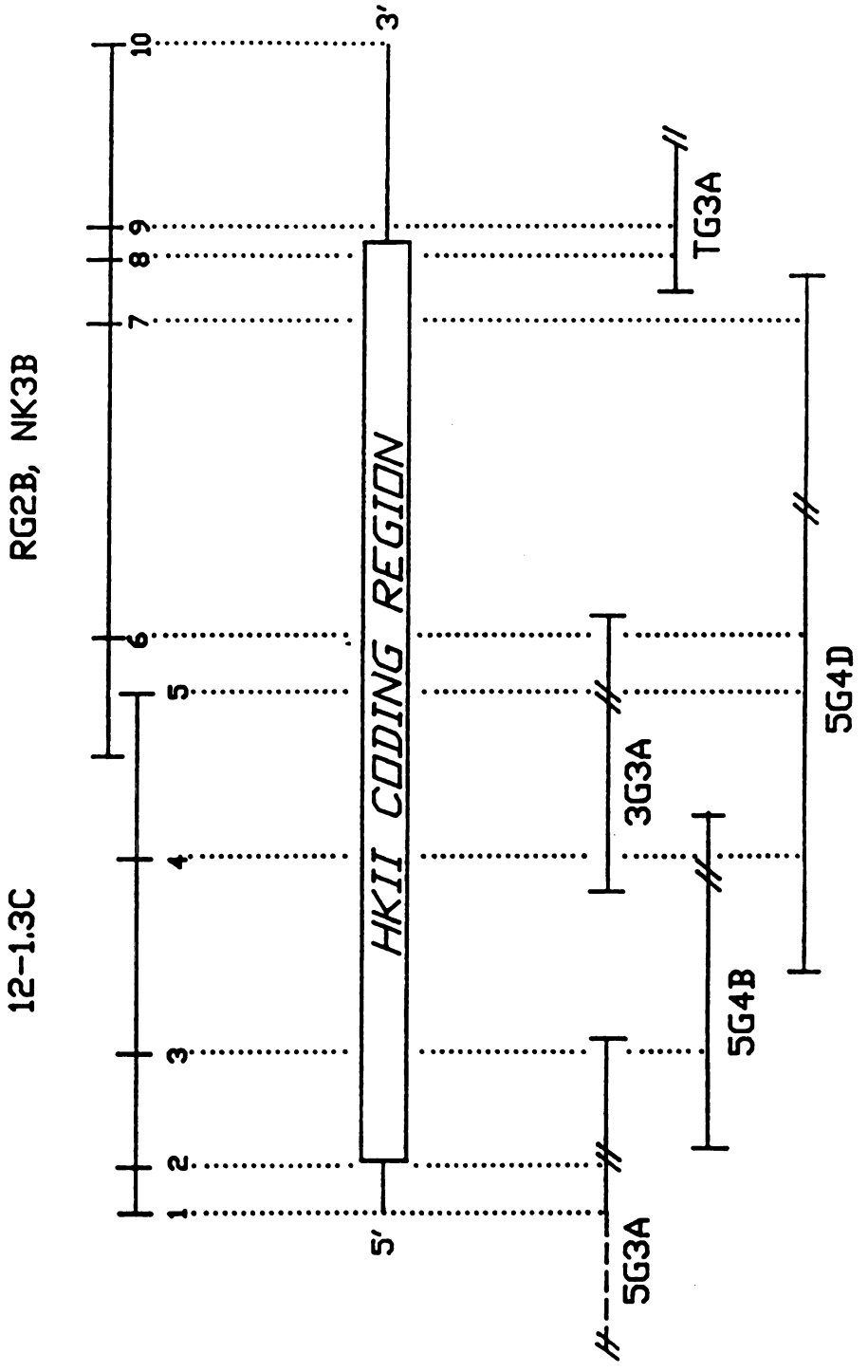

RG2B, NK3B
12-1.3C
HKII CODING REGION
3'
5'
1
2
3
4
5
6
7
8
9
10
TG3A
3G3A
5G4D
5G4B
5G3A

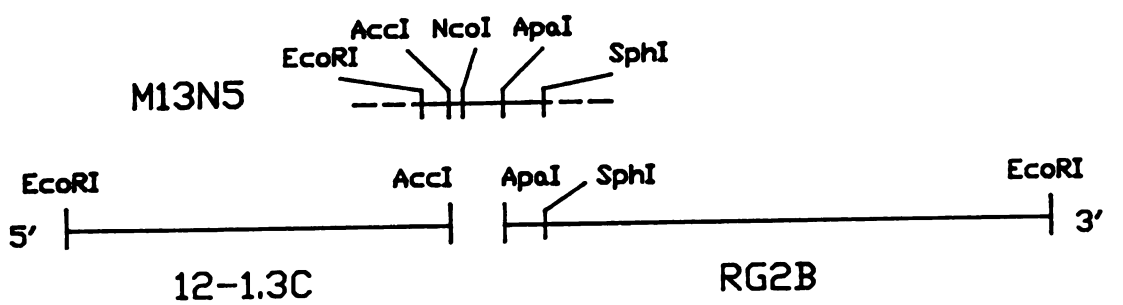
FIGURE 2

probe isolations were from the linker regions of the cDNA cloning vectors and not part of the cDNA for Type II hexokinase.

Site-directed Mutagenesis of Type II Hexokinase: Creation of *Nco*I Site. The mutagenesis procedure was that of Kunkel (102,103). The 306 bp *EcoR*I-*Sph*I fragment of clone RG2B (nucleotides 1-306) was directionally subcloned into M13mp19. The recombinant (designated M13N5) was grown in the *ung, dut* double mutant *E. coli* strain CJ236. This strain allows for the incorporation of some uracil residues in thymine positions. Single stranded phage DNA, containing uracil residues, was purified using PEG (20% polyethylene glycol 8000, 2 M NaCl) precipitation (95). The mutation primer, a 17-mer(5'-GGCTGCCATGGTGACGG-3') was identical to nucleotides 154-170 of clone M13N5. This corresponds to 1552-1568 of the composite cDNA for Type II (Fig. 5, Results Chapter) with a change from T to C at the underlined nucleotide 160 (nucleotide 1558, Fig. 5) generating an *Nco*I site. This primer was annealed to the template DNA at a ratio of 5 pm primer to 200 ng DNA in a 10 μl reaction that contained 20 mM Tris-HCl (pH 7.4), 2 mM $MgCl_2$, and 50 mM NaCl. The mutated strand was synthesized by the addition of 0.4 mM each dNTP, 0.75 mM ATP, 2 mM DTT, 1 unit T4 DNA ligase, and 1 unit T4 DNA polymerase. The reaction was incubated at 37°C for 1.5 hr. Approximately 30% of the reaction was used to transform *E. coli.* strain mv1190. This strain has a functional uracil N-glycosylase which inactivates the parental strand containing uracil residues. The non-uracil containing mutated strand remains intact. This becomes an efficient selection process as most of the recombinants will contain the desired

mutation. Several recombinants were analyzed for the desired mutation by sequencing the 306 bp fragment in M13mp19. Eighty percent of the recombinants chosen contained the *Nco*I site.

A full length cDNA for Type II hexokinase, containing the *Nco*I site, was constructed using restriction endonucleases *Acc*I and *Apa*I. *Acc*I restricts the N-terminal cDNA clone 12-1.3C at nucleotide 1442; the site for *Apa*I is at nucleotide 200 of the C-terminal Type II cDNA clone RG2B (nucleotide 1599, Fig. 5). The cDNA inserts were isolated from *EcoR*I digested λgt10 clones. The isolated cDNA fragments were subsequently restricted with the appropriate endonuclease (*Acc*I in clone 12-1.3C, *Apa*I in clone RG2B). Digestion of the mutated M13 clone (M13N5) with these two restriction enzymes generated a small (157 bp) fragment containing the *Nco*I site. These 3 fragments (shown below, schematically) were isolated and ligated together. The resulting 3.6 kb fragment, with *EcoR*I ends and the internal *Nco*I site, was subcloned into the *EcoR* site of pUC18. The resultant recombinant was designated pIIN-RI.



Expression of Type II Hexokinase in COS-1 Cells. The expression vector, pSVT7, and COS-1 cells were provided by Dr. W. L. Smith of this department, with the permission of Dr. J. Sambrook (104). A 3 Kb fragment of Type II hexokinase cDNA (from clone pIIN-RI, containing an *Nco*I site), from the *EcoR*I site at the 5' end to the K*pn*I site at nucleotide 3082 at the 3' end, was directionally subcloned into the *EcoR*I-K*pn*I sites of pUC18. This 3 Kb fragment included the entire coding region for the Type II isozyme, along with 196 bases upstream of the start codon, and 146 nucleotides downstream of the end of the coding region. The cDNA was excised from the pUC18 recombinant using restriction enzymes *EcoR*I and P*st*I, sites within the polylinker of pUC18. This fragment was directionally subcloned into pSVT7, previously digested with the same enzymes. This aligned the coding region for Type II hexokinase properly with respect to the SV40 origin of replication and early promoter in pSVT7. The recombinant plasmid (designated pSVT7-HKII), containing Type II hexokinase as determined by restriction analysis, was grown in DH5α cells and purified over QIAGEN columns according to manufacturer's directions. Quantities of the plasmid, pSVT7, were also purified in the same manner.

COS-1 cells were grown in Dulbecco's modified Eagle's medium (high glucose) supplemented with 8% bovine calf serum, 2% fetal calf serum, and 2 mM glutamine. Cell cultures used for transfection were 80-90% confluent. The COS-1 cells were transfected, as previously described (105), using 36 μg DNA and 750 μg DEAE Dextran per plate of cells. Sham (no DNA added) and pSVT7 (vector only) transfected cultures served as controls. Cells, harvested 42 hrs after transfection,

were resuspended in PBS (0.045 M potassium phosphate, 0.15 M NaCl pH7.3) containing 10 mM glucose and 10 mM TG. The resuspended cells were sonicated and centrifuged at 15,000xg for 15 min. Hexokinase activity in the supernatants was determined spectrophotometrically, as previously described (96). Protein concentration was determined with the BCA reagent using BSA as the standard.

Aliquots of the homogenates were electrophoresed on 6.5-20% SDS-polyacrylamide gels. Purified Type I hexokinase was used as a positive control for the anti-Type I antibody reactions. Also included on the gel was a sample from a crude homogenate of rat skeletal muscle. The crude sample was prepared by initially grinding 2 gm of tissue under liquid N_2, followed by homogenization in 50 mM sodium phosphate (pH 8.0), 1 mM Glc, 10 mM TG, and 1% Triton X100 (4 ml per gm of tissue). The crude homogenate was centrifuged at 30,000 rpm for 1 hr. The clear supernatant was stored at -20°C after the protein concentration was determined as above. The separated proteins were transferred to nitrocellulose in a carbonate buffer system (106). The protein blots were reacted with either pre-immune sera or anti-hexokinase I polyclonal antibodies, at a dilution of 1:500 in TTBS (50 mM Tris-HCl, 154 mM NaCl, 0.05% (v/v) Tween 20, pH7.5), as previously described (90), using 5% nonfat dry milk as the blocking reagent. The blots were developed with the tetrazolium method of Taketa *et al.* (107).

RNA Isolation and Northern Analysis. Skeletal muscle RNA was prepared from hind limb muscles of 4-7 week old Sprague-Dawley rats using the method of Chomczynski and Sacchi (108) except that, prior to homogenization, the tissue was

ground to a fine powder under liquid N_2. Novikoff ascites tumor RNA was prepared using the method of Chirgwin *et al.* (109). RNA preparations were redissolved in water to give approx 1 mg/ml. Polyadenylated mRNA was isolated from total RNA by chromatography on oligo (dT) cellulose as described by Maniatis *et al.* (95).

Northern blotting was done by standard procedures, essentially as described by Maniatis *et al.* (95). Duplicate blots containing 3 μg of Novikoff mRNA and 20 μg of rat skeletal muscle mRNA were air dried, and vacuum baked for 1 hr. After prehybridization in 5x SSC, 0.1% SDS, 5x Denhardt's, and 50% formamide at 42°C for at least 1 hr, the blots were hybridized overnight using cDNA probes radiolabelled with [α-^{32}P]dCTP by random primer synthesis (101) to a specific activity greater than 10^8cpm/μg. Two non-overlapping cDNA probes from Type II hexokinase were used, representing N-terminal (nucleotides 1-1562, Fig. 5) and C-terminal (nucleotides 1705-3082, Fig. 5) regions of the enzyme. After hybridization the blots were washed sequentially in: 2x SSC, 0.1% SDS at room temperature for 15 minutes; 2x SSC, 0.1% SDS at 60°C for 15 minutes (twice); and 2x SSC, at room temperature, 15 minutes. Hybridized bands were visualized autoradiographically. The size of the mRNA bands was estimated by extrapolation between the 28s (4.72 kb) and 18s (1.87 kb) ribosomal RNA bands.

Primer Extension. A 24-base oligonucleotide (5'-GCTTAACCACGATGGCTCACCAGC-3'), complementary to nucleotides 34-58 of Type II hexokinase mRNA, was synthesized and 5'-end-labelled with [γ-^{32}P]ATP (6000 Ci/mm) and T4 polynucleotide kinase. The primer ($1x10^6$ cpm) was annealed in a

total of 15 ul to either 3 μg Novikoff tumor poly(A)$^+$ RNA or 20 μg rat skeletal muscle poly(A)$^+$ RNA. The annealing buffer contained 0.15 M KCl, 10 mM Tris-HCl (pH 8.3), and 1 mM EDTA. After hybridization for 1.5 hr at 62°C, the elongation reagents were added to the reaction mixtures containing the RNA and annealed primer. The 40 μl elongation reaction contained 20 mM Tris-HCl (pH 8.3), 10 mM $MgCl_2$, 6 mM DTT, 0.3 mM each dNTP, actinomycin D (150 ug/ml) and 5 units of AMV reverse transcriptase. The reaction was incubated at 42°C for 1 hr. and then treated with RNase A (15 μg/ml) for 15 min. The products were fractionated on 5% or 9% polyacrylamide denaturing gels and visualized by autoradiography.

S1 Nuclease Protection Assay. Single-stranded DNA, generated from a 630 bp (*Sma*I-*Sma*I) fragment of genomic clone 5G3A that had been subcloned into M13mp19, was annealed to the same 24-base primer used in the primer-extension analysis that had been 5'-end-labelled with [γ-^{32}P]ATP (6000 Ci/mm) to a specific activity of approximately 10^9 cpm per μg. The primer was extended in the presence of 0.4 mM each dNTP and the Klenow fragment of DNA polymerase I. The double-stranded DNA was digested with *Sal*I, and the strands were separated on a 3.5% polyacrylamide 7M urea gel. The isolated 649-nucleotide fragment (5 x 10^5 cpm) was hybridized to either 3 μg Novikoff tumor poly(A)$^+$ RNA or 20 μg rat muscle poly(A)$^+$ RNA in a 10 μl reaction containing 0.4M NaCl and 10 mM Pipes (pH 6.4). The reaction was heated at 70°C for 5 min and incubated at 65°C for at least 6 hrs. After hybridization, 300 μl of a solution containing 0.2 M NaCl, 2 mM $ZnSO_4$,

20 mM sodium acetate (pH 5.0), and 200 units of S1 nuclease were added and the mixture was incubated at 37°C for 1 hr. The digestion by S1 nuclease was terminated with the addition of 80 μl of a solution containing 4M ammonium acetate, 50 mM EDTA. Following phenol/chloroform extract and ethanol precipitation, the resultant products were fractionated on 5% or 9% polyacrylamide 7M urea gels and visualized by autoradiography.

Cell-Free *in vitro* Transcription Assay. The "G-free cassette" plasmids, $p(C_2AT)_{19}$ (110) and pML-GFC2 (111), and the rat liver nuclear proteins (prepared according to Gorski *et al.* (112)) were generously provided by Dr. D. Jump (Physiology Dept., Michigan State University). The plasmids were used with the permission of Dr. R. G. Roeder (Rockefeller University) (110). These plasmids contain a synthetic DNA fragment subcloned into the *Sac*I-*Sma*I sites of pUC13 that generate a discrete G-free RNA product when transcribed in the absence of GTP. The length of the synthetic fragment in $p(C_2AT)_{19}$ is approx. 380 nucleotides. The plasmid, pML-GFC2, contains the same synthetic fragment shortened to approx 280 nucleotides, which is under the control of the adenovirus-2 major late promoter. Inclusion of 3'-O-methyl-GTP hinders transcription of any promoter-like sequences in the vector and T_1 RNase degrades spurious transcripts containing G residues. The Type II hexokinase promoter is numbered relative to a transcription start site of +1. The plasmid $p(C_2AT)$-HKII contains the Type II hexokinase promoter from positions -260 to +12 upstream of the "G-free cassette" of $p(C_2AT)_{19}$. This section of the promoter region was isolated using endonucleases *Sma*I and *Dde*I (5'→3') and blunt

ends were generated using T4 polymerase. After the addition of *EcoRI* linkers, the promoter fragment was subcloned into the *EcoRI* site at the 5' end of the "G-free cassette" of $p(C_2AT)_{19}$ and grown in *E. coli* strain JM105. Restriction analysis was used to determine orientation of the promoter fragment in several recombinants. Recombinants with this promoter fragment in both orientations were purified using QAIGEN columns. These purified templates were used in the *in vitro* transcription assay system, essentially as described by MacDougald and Jump (111).

Using the $p(C_2AT)$-HKII constructs, the ability of a portion of the Type II hexokinase promoter to direct transcription was determined. The transcription reactions (40μl) contained 25 mM Hepes (pH 7.5), 5% glycerol, 50 mM KCl, 6 mM $MgCl_2$, 0.6 mM ATP and CTP, 0.03 mM UTP, 15 μCi [α-^{32}P]UTP, 0.1 mM 3'-O-methyl GTP, 15 units of T_1-RNase, and 60 μg of rat liver nuclear proteins. Two micrograms of each $p(C_2AT)$-HKII construct was included in a reaction along with 50 ng of pML-GFC2. As a negative control, one transcription assay included 2 μg of the original plasmid, $p(C_2AT)_{19}$ containing no promoter, along with 50 ng of pML-GFC2. The plasmid, pML-GFC2, was included in each assay as an internal positive control. The transcription reactions were incubated at 30°C for 90 min. The reactions were stopped with the addition of 380 μl of 50 mM Tris-Cl (pH 7.5), 1% SDS and 5 mM EDTA. To aid in the precipitation of the transcription products, 40 μg of yeast t-RNA were added. The reactions were extracted 3 times with phenol/$CHCl_3$ (equilibrated with 10 mM NaOAc (pH 5.0), 100 mM NaCl,

1 mM EDTA). After ethanol precipitation the transcription products were fractionated on a 6% denaturing and visualized by autoradiography.

CHAPTER 3

Results

Isolation of Clones Containing cDNA for Rat Type II Hexokinase.

Approximately 250,000 recombinants of an amplified λgt10 rat soleus muscle cDNA library (Library 1, Table III) were screened under medium stringency conditions with a partial cDNA for rat Type I hexokinase, previously described and designated HKI 12.4-4 (67), as the radiolabelled probe. Two positive recombinants, termed 12-1.3C and 15-1, were isolated. Both contained 1.6 kb inserts, and partial sequence analysis and *Hae*III digestion patterns (Fig. 3, lanes 3 and 4) indicated that these inserts were identical. Thus, further analysis was restricted to clone 12-1.3C.

The restriction map and sequencing strategy for 12-1.3C are shown in Fig. 4. The nucleotide sequence for 12-1.3C represents nucleotides at positions 1-1562 of the sequence shown in Fig. 5. There is an open reading frame that extends from nucleotide 197 to the 3' end of the clone at nucleotide 1562. The deduced amino acid sequence shows extensive similarity to that of the rat Type I (68), Type III (56), and Type IV (32) isozymes, yet is clearly distinct from these.

An 8-residue segment of the deduced sequence (underlined in Fig. 5) is identical to the sequence determined by direct N-terminal sequencing of the Type II isozyme, prepared as described in Methods: Phe-Thr-Glu-Leu-Asn-Gln-Asn-Gln. The N-terminal sequence for Type II hexokinase, purified from rat skeletal muscle by a different procedure, is Glu-Leu-Asn-Gln-Asn-Gln-Val-Gln-Lys-Val-Asp-Gln-Phe-Leu-Tyr-X-Met-Arg-Val (P. Fischer, F.E.Weber, K. Beyreuther, and D. Pette, personal communication). It is evident that residues 3-8 in the sequence determined from

Table III. Summary of cDNA library synthesis/screening strategies and results.

Library No.	Vector tissue	Primer	Probe (stringency[a])	Results[b]
1	λgt10 muscle	dT	12-4.4 (HKI-C)(1) 12-1.3C(2) NK3B(2)	2-1.6kb clones (12-1.3, 15) 39 ≡ 12-1.3C 6 ≡ 12-1.3C
2	λgt10 liver	dT	12-1.3C (2)	6 ≡ 12-1.3C
3	pUC8/9 muscle	dT directional	12-1.3C (2)	3 clones, all within 12-1.3C
4	UniZAPII muscle	dT-*Xho*I	12-1.3C (2)	no +
5	UniZAPII fat pad	dT-*Xho*I	12-1.3C (2)	no +
6	UniZAPII kidney	dT-*Xho*I	12-1.3C (2)	11 ≡ 2.3 kb[e] 6 ≡ 0.4 kb[c,d]
7	UniZAPII spleen	dT-*Xho*I	12-1.3C (2)	4 ≡ 2.3 kb[e] 1 ≡ 0.4 kb[d,e]
8	λgt10 muscle[c]	random	12-1.3C(3') (2)	3 clones, largest RG2B(2.3 kb);other two contained within RG2B
9	λgt10 muscle	random	12-1.3C (3') (2)	1 ≡ RG2B
10	λgt10 tumor	dT	12-1.3C (2)	2 clones:largest 2.3 kb, NK3B ≡ RG2B

a-Stringency determined by temperature and percent formamide: (1) medium-42°C and 37%; (2) high-42°C and 50%

b- ≡ identical as determined by *Hae*III restriction digest patterns

c-RNA from stimulated muscle tissue, a generous gift of Dr. Pette

d-Identical to sequence at 5' end of 12-1.3C

e-Each clone also contained an identical 3.4 kb *EcoR*I-*Xho*I fragment

Figure 3. *Hae*III restriction patterns of cDNA clones. Shown is an EtBr-stained 5% polyacrylamide gel containing cDNA restricted with the endonuclease *Hae*III. Lane 1 contains pBR digested with *Hae*III which was used as size markers. The sizes (in bp) of several bands in lane 1 are marked at the left. Lanes 2 shows the *Hae*III restriction pattern of the insert of cDNA clone HKI-12.4-4. Lanes 3 and 4 contain cDNA from clones 12-1.3C and 15-1 (respectively), restricted with *Hae*III. Clone HKI-12.4-4 was the cDNA probed used to isolate clones 12-1.3C and 15-1.

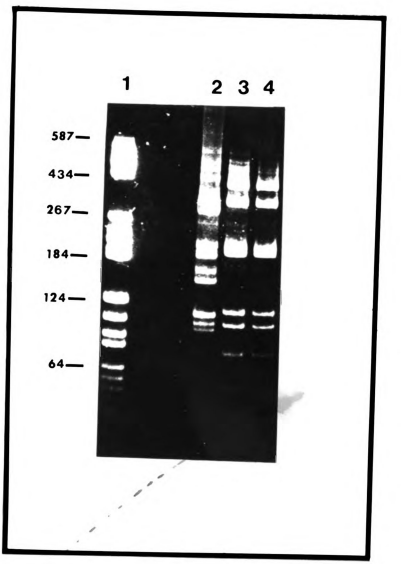


FIGURE 3

Figure 4. Relevant restriction sites and sequencing strategy for clones containing either cDNA or genomic DNA for rat Type II hexokinase. The composite sequence, which includes the coding region and 5' and 3' untranslated regions, is shown near the top of the figure. Depicted above it are the portions of genomic clone 3G3A that were sequenced for verification of the overlapping cDNA sequences. The solid lines in the genomic clone represent exons and the dashed lines are introns. The dotted lines from the genomic clone to the composite figure show the positions of the three exons in the composite sequence. Shown below the composite figure are the cDNA clones 12-1.3C, RG2B and NK3B (see text). Restriction sites used in sequencing and the direction of sequencing are shown for each clone. Restriction site abbreviations are: A, A*va*II; B, B*gl*II; C, H*inc*II; D, H*ind*III; E, H*ae*III; H, H*ph*I; M, S*ma*I; P, P*st*I; R, E*coR*I; S, S*ph*I; V, P*vu*II.

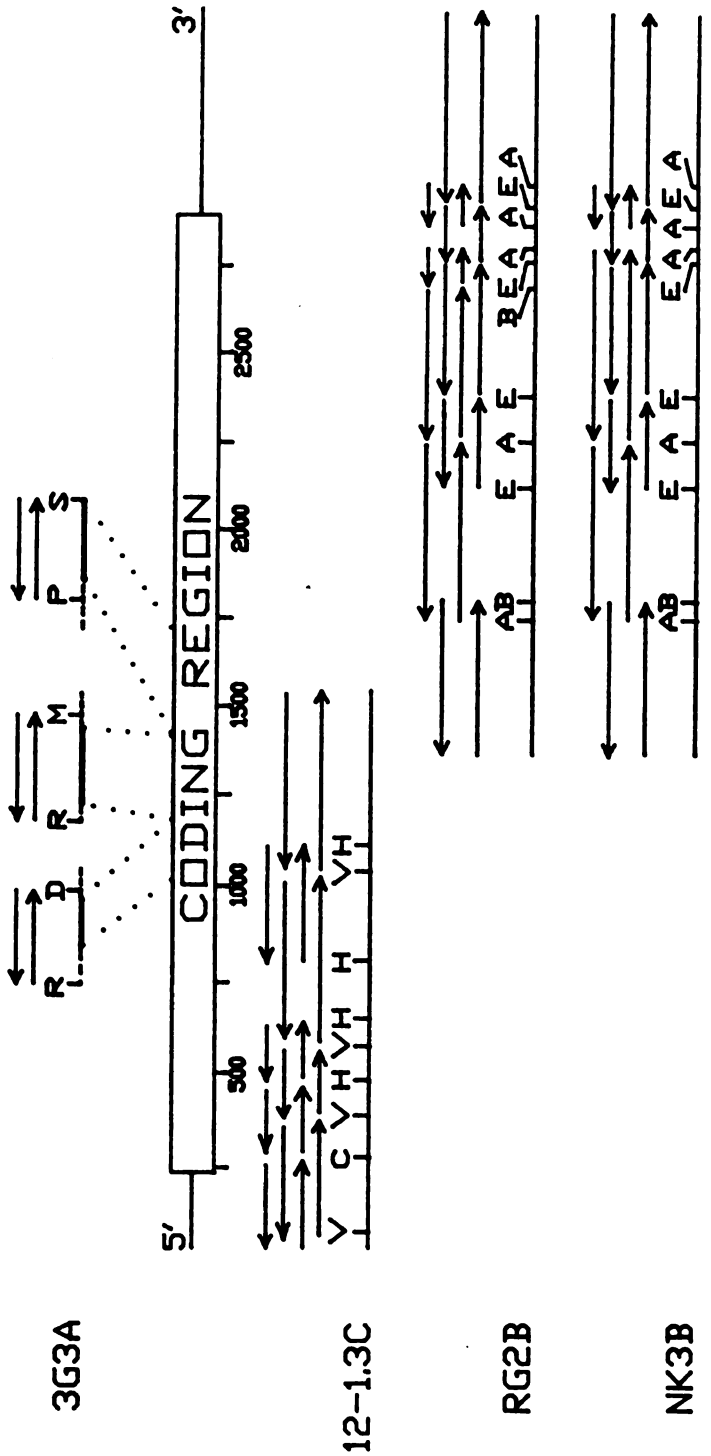

3'
5'
CODING REGION
500
1000
1500
2000
2500
R
D
J
R
M
P
S
V
C
VH
VH
H
VH
E
A
B
3G3A
12-1.3C
RG2B
NK3B

Figure 5. Nucleotide and deduced amino acid sequences for rat Type II hexokinase. The underlined region in the amino acid sequence corresponds to the sequence obtained by direct N-terminal sequencing of the enzyme (see text for further comments in this regard). The dashed line *above* the nucleotide sequence indicates the region of overlap between the cDNA clone 12-1.3C, coding for the N-terminal half of the enzyme, and clone RG2B which encoded the C-terminal half.

```
                                                                                GGAAAACCTGTTTCTGGAAACGCGAGGCCCTCAGCTGGTGAGCCATCGTGGTTAAGCTTCTTTGTGTGGCTCCTGGA    77
GTCTCCGATCCCAGCCGGACACCCGGGCCTGGTTTCAAAGCGGTCGAACTGCTCTGCCCGCTCCACCGGTAGCGCTCGAGCCTCGGTTTCTCTACTCGACCCCGACTCGCCGCAGCAGG   196

ATG ATC GCC TCG CAT ATG ATC GCC TGC TTA TTC ACG GAG CTC AAC CAA AAC CAA GTG CAG AAG GTT GAC CAA TTT CTC TAC CAC ATG CGT   286
Met Ile Ala Ser His Met Ile Ala Cys Leu Phe Thr Glu Leu Asn Gln Asn Gln Val Gln Lys Val Asp Gln Phe Leu Tyr His Met Arg    30

CTC TCA GAT GAG ACC CTT CTG GAG ATT TCT AGG CGG TTC CGG AAG GAG ATG GAG AAA GGG CTA GGA GCT ACC ACG CAC CCT ACA GCA GCT   376
Leu Ser Asp Glu Thr Leu Leu Glu Ile Ser Arg Arg Phe Arg Lys Glu Met Glu Lys Gly Leu Gly Ala Thr Thr His Pro Thr Ala Ala    60

GTG AAA ATG TTG CCT ACC TTT GTG AGG TCA ACT CCG GAT GGG ACA GAA CAT GGG GAG TTC CTG GCT CTG GAT CTT GGA GGA ACC AAC TTC   466
Val Lys Met Leu Pro Thr Phe Val Arg Ser Thr Pro Asp Gly Thr Glu His Gly Glu Phe Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe    90

CGT GTG CTC CGA GTA AGG GTG ACG GAC AAT GGC CTC CAG AGA GTG GAG ATG GAG AAC CAG ATC TAC GCC ATC CCT GAG GAC ATC ATG CGG   556
Arg Val Leu Arg Val Arg Val Thr Asp Asn Gly Leu Gln Arg Val Glu Met Glu Asn Gln Ile Tyr Ala Ile Pro Glu Asp Ile Met Arg   120

GGC AGT GGA ACC CAG CTG TTT GAC CAC ATC GCC GAA TGC CTG GCC AAC TTC ATG GAC AAG CTA CAA ATC AAA GAG AAG AAG CTC CCT CTG   646
Gly Ser Gly Thr Gln Leu Phe Asp His Ile Ala Glu Cys Leu Ala Asn Phe Met Asp Lys Leu Gln Ile Lys Glu Lys Lys Leu Pro Leu   150

GGT TTC ACC TTC TCG TTC CCC TGC CAC CAG ACA AAA CTG GAT GAG AGT TTT TTG GTC TCG TGG ACT AAG GGG TTC AAG TCC AGT GGC GTG   736
Gly Phe Thr Phe Ser Phe Pro Cys His Gln Thr Lys Leu Asp Glu Ser Phe Leu Val Ser Trp Thr Lys Gly Phe Lys Ser Ser Gly Val   180

GAA GGC AGA GAT GTG GTG GAC CTG ATC CGG AAG GCT ATC CAG CGC AGA GGG GAC TTT GAC ATT GAC ATT GTG GCC GTG GTG AAT GAC ACA   826
Glu Gly Arg Asp Val Val Asp Leu Ile Arg Lys Ala Ile Gln Arg Arg Gly Asp Phe Asp Ile Asp Ile Val Ala Val Val Asn Asp Thr   210

GTT GGG ACC ATG ATG ACT TGT GGC TAT GAT GAT CAG AAC TGC GAG ATT GGT CTC ATT GTG GGC ACT GGC AGC AAC GCC TGC TAC ATG GAG   916
Val Gly Thr Met Met Thr Cys Gly Tyr Asp Asp Gln Asn Cys Glu Ile Gly Leu Ile Val Gly Thr Gly Ser Asn Ala Cys Tyr Met Glu   240

GAA ATG CGT CAT ATT GAC ATG GTG GAG GGA GAT GAG GGG CGC ATG TGC ATC AAC ATG GAG TGG GGA GCC TTT GGG GAC GAC GGT ACA CTC  1006
Glu Met Arg His Ile Asp Met Val Glu Gly Asp Glu Gly Arg Met Cys Ile Asn Met Glu Trp Gly Ala Phe Gly Asp Asp Gly Thr Leu   270

AAT GAC ATC CGA ACC GAG TTT GAC CGA GAG ATC GAC ATG GGC TCG CTG AAC CCT GGG AAG CAG CTG TTT GAG AAG ATG ATT AGC GGG ATG  1096
Asn Asp Ile Arg Thr Glu Phe Asp Arg Glu Ile Asp Met Gly Ser Leu Asn Pro Gly Lys Gln Leu Phe Glu Lys Met Ile Ser Gly Met   300

TAC ATG GGG GAG CTG GTC AGG CTC ATC CTG GTG AAG ATG GCC AAG GCA GAG CTG TTG TTC CAA GGG AAA CTC AGC CCA GAA CTC CTT ACC  1186
Tyr Met Gly Glu Leu Val Arg Leu Ile Leu Val Lys Met Ala Lys Ala Glu Leu Leu Phe Gln Gly Lys Leu Ser Pro Glu Leu Leu Thr   330

ACT GGC TCC TTC GAG ACC AAA GAT GTC TCG GAT ATT GAA GAG GAT AAG GAT GGA ATC GAG AAG GCC TAC CAA ATC CTG ATG CGC CTG GGT  1276
Thr Gly Ser Phe Glu Thr Lys Asp Val Ser Asp Ile Glu Glu Asp Lys Asp Gly Ile Glu Lys Ala Tyr Gln Ile Leu Met Arg Leu Gly   360

CTG AAT CCA TTG CAG GAG GAT TGT GTG GCC ACG CAC CGA ATC TGC CAG ATT GTG TCC ACG CGC TCG GCC AGT CTG TGC GCA GCC ACC CTG  1366
Leu Asn Pro Leu Gln Glu Asp Cys Val Ala Thr His Arg Ile Cys Gln Ile Val Ser Thr Arg Ser Ala Ser Leu Cys Ala Ala Thr Leu   390
                                                                                ----------------------------------------
GCC GCG GTG CTG TGG CGA ATC AAA GAG AAC AAG GGC GAG GAG CGA CTT CGC TCC ACC ATC GGT GTC GAT GGC TCC GTC TAC AAG AAA CAT  1456
Ala Ala Val Leu Trp Arg Ile Lys Glu Asn Lys Gly Glu Glu Arg Leu Arg Ser Thr Ile Gly Val Asp Gly Ser Val Tyr Lys Lys His   420
----------------------------------------------------------------------------------------------------------------------
CCC CAT TTT GCC AAG CGT CTC CAT AAG GCA GTG AGG AGG CTG GTG CCC GAC TGT GAT GTC CGC TTC CTC CGC TCT GAG GAT GGC AGC GGC  1546
Pro His Phe Ala Lys Arg Leu His Lys Ala Val Arg Arg Leu Val Pro Asp Cys Asp Val Arg Phe Leu Arg Ser Glu Asp Gly Ser Gly   450
----------
AAG GGG GCT GCT ATG GTG ACG GCG GTG GCT TAC CGT CTG GCT GAC CAA CAC CGG GCC CGC CAG AAG ACC CTG GAG TCT CTG AAG CTG AGC  1636
Lys Gly Ala Ala Met Val Thr Ala Val Ala Tyr Arg Leu Ala Asp Gln His Arg Ala Arg Gln Lys Thr Leu Glu Ser Leu Lys Leu Ser   480

CAC GAG CAG CTT CTG GAG GTT AAG AGA AGA ATG AAG GTG GAA ATG GAG CAG GGT CTG AGC AAG GAG ACG CAT GCG GTC GCC CCT GTG AAG  1726
His Glu Gln Leu Leu Glu Val Lys Arg Arg Met Lys Val Glu Met Glu Gln Gly Leu Ser Lys Glu Thr His Ala Val Ala Pro Val Lys   510

ATG CTG CCC ACT TAC GTG TGT GCC ACT CCA GAT GGC ACA GAG AAA GGA GAC TTC TTG GCC TTG GAT CTT GGA GGA ACA AAC TTC CGG GTC  1816
Met Leu Pro Thr Tyr Val Cys Ala Thr Pro Asp Gly Thr Glu Lys Gly Asp Phe Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg Val   540

CTG CTG GTG CGT GTG CGT AAT GGC AAG CGG AGG GGC GTG GAG ATG CAT AAC AAG ATC TAC TCC ATC CCA CAG GAG GTT ATG CAT GGC ACT  1906
Leu Leu Val Arg Val Arg Asn Gly Lys Arg Arg Gly Val Glu Met His Asn Lys Ile Tyr Ser Ile Pro Gln Glu Val Met His Gly Thr   570

GGG GAA GAG CTC TTC GAC CAC ATT GTC CAG TGC ATT GCG GAC TTC CTG GAG TAC ATG GGC ATG AAG GGC GTG TCC CTG CCT TTG GGT TTC  1996
Gly Glu Glu Leu Phe Asp His Ile Val Gln Cys Ile Ala Asp Phe Leu Glu Tyr Met Gly Met Lys Gly Val Ser Leu Pro Leu Gly Phe   600

ACA TTC TCC TTC CCT TGC CAG CAG AAC AGC CTA GAC CAG AGC ATC CTC CTC AAG TGG ACA AAG GGA TTC AAG GCA TCT GGC TGC GAG GGT  2086
Thr Phe Ser Phe Pro Cys Gln Gln Asn Ser Leu Asp Gln Ser Ile Leu Leu Lys Trp Thr Lys Gly Phe Lys Ala Ser Gly Cys Glu Gly   630

GAG GAT GTG GTC ACC TTG CTG AAG GAA GCG ATT CAC CGG CGA GAG GAG TTT GAC CTG GAT GTG GTT GCC GTG GTG AAT GAC ACA GTT GGG  2176
Glu Asp Val Val Thr Leu Leu Lys Glu Ala Ile His Arg Arg Glu Glu Phe Asp Leu Asp Val Val Ala Val Val Asn Asp Thr Val Gly   660

ACT ATG ATG ACT TGT GGC TAC GAA GAC CCT CAC TGT GAA GTT GGC CTC ATT GTT GGC ACC GGA AGC AAC GCC TGC TAC ATG GAA GAG ATG  2266
Thr Met Met Thr Cys Gly Tyr Glu Asp Pro His Cys Glu Val Gly Leu Ile Val Gly Thr Gly Ser Asn Ala Cys Tyr Met Glu Glu Met   690

CGT AAT GTG GAG CTG GTG GAC GGA GAG GAG GGA CGG ATG TGT GTC AAC ATG GAG TGG GGA GCA TTT GGG GAC AAT GGC TGC CTG GAT GAC  2356
Arg Asn Val Glu Leu Val Asp Gly Glu Glu Gly Arg Met Cys Val Asn Met Glu Trp Gly Ala Phe Gly Asp Asn Gly Cys Leu Asp Asp   720

TTG CGG ACC GTG TTT GAT GTT GCT GTG GAT GAG CTT TCT CTC AAC CCT GGC AAA CAG AGG TTC GAG AAG ATG ATC AGC GGC ATG TAC TTG  2446
Leu Arg Thr Val Phe Asp Val Ala Val Asp Glu Leu Ser Leu Asn Pro Gly Lys Gln Arg Phe Glu Lys Met Ile Ser Gly Met Tyr Leu   750

GGA GAG ATT GTG CGC AAC ATT CTC ATC GAT TTC ACG AAG CGG GGG CTG CTC TTC CGA GGC CGC ATC TCA GAG CGC CTC AAG ACA AGG GGA  2536
Gly Glu Ile Val Arg Asn Ile Leu Ile Asp Phe Thr Lys Arg Gly Leu Leu Phe Arg Gly Arg Ile Ser Glu Arg Leu Lys Thr Arg Gly   780

ATC TTT GAA ACT AAG TTC CTG TCT CAG ATA GAG AGC GAC TGC CTA GCC CTG CTA CAG GTT CGT GCC ATC CTG CGC CAC CTA GGG CTG GAG  2626
Ile Phe Glu Thr Lys Phe Leu Ser Gln Ile Glu Ser Asp Cys Leu Ala Leu Leu Gln Val Arg Ala Ile Leu Arg His Leu Gly Leu Glu   810

AGC ACG TGC GAT GAC AGC ATC ATC GTG AAG GAG GTG TGC ACT GTG GTT GCC CGG CGC GCT GCA CAG CTC TGT GGC GCA GGC ATG GCC GCC  2716
Ser Thr Cys Asp Asp Ser Ile Ile Val Lys Glu Val Cys Thr Val Val Ala Arg Arg Ala Ala Gln Leu Cys Gly Ala Gly Met Ala Ala   840

GTA GTG GAC AAG ATA AGA GAG AAC CGT GGG CTG GAC AAC CTC AAA GTG ACA GTG GGC GTG GAC GGG ACT CTG TAT AAG CTT CAT CCT CAC  2806
Val Val Asp Lys Ile Arg Glu Asn Arg Gly Leu Asp Asn Leu Lys Val Thr Val Gly Val Asp Gly Thr Leu Tyr Lys Leu His Pro His   870

TTT GCC AAG GTC ATG CAT GAG ACG GTG AGA GAT CTG GCT CCG AAA TGT GAC GTG TCC TTC CTG GAA TCC GAG GAC GGC AGT GGG AAG GGA  2896
Phe Ala Lys Val Met His Glu Thr Val Arg Asp Leu Ala Pro Lys Cys Asp Val Ser Phe Leu Glu Ser Glu Asp Gly Ser Gly Lys Gly   900

GCA GCT CTC ATC ACT GCC GTG GCC TGC CGC ATC CGG GAG GCT GGG CAG AGA TAGAAGCTTGGGGACCAGACGGGGCTTCTCTGCTTCCTCACTTCCCCGTTT  2998
Ala Ala Leu Ile Thr Ala Val Ala Cys Arg Ile Arg Glu Ala Gly Gln Arg End

TAAATTATGAGAAGATGGACCCCTTGGCAGAGAGGACCCTGTGAGACTGGGACTTTTGTCTCTGTATATTCACTGTAGAGTTTGGTACCCAATCCTTGGCCTTCCTGAGAATCTGAAGT  3117
AAGGAGTTGTTCGCAGTTCCTGCTGGTCACCTCCTGGGATGGTCCTCCCCCACTCTACTTCCTGTACAATCCTATGGGATGGGGACAGTCTTGAGAAAATGGTTGCCTCAGCTTGGAGC  3236
TGAACCTGACATTTGTAGGTGGAGAACATCTTGAAATGTCGATGCTTCTCTCATCAGAGGTAGATGTTGGGTGGGAGGGAGGAGGAGAGATAGGACGTCCCCTGCCCTACGTGGCTGTC  3355
CCTCTTCAACTCAACCCATAAGCATCACCCTGCCGGTTCTAAAACATTTGCTGTCCTGGGATAGAGGAATCTGAACCAAAGCTAGAGTGGACGTCTTCTTCAAGTGTTTGTGGCTTTAC  3474
CTCACATGGCTGAATGTGTCATCCACGAGCTACAGACCTTGGGGACAGGAGTGGCTTTGGTCGCTTTTTAGTTTTTTGTTTTGTTTTGTTTTTTTCTTTTTTAATTTATGTTTGAAATT  3593
TACTTAGTTTAAAGCAGGTCCTCTCCATTGTGTAATGGTTG  3634
```

FIGURE 5

enzyme purified in our laboratory (underlined in Fig. 5) correspond to the initial portion of the latter sequence. Furthermore, the sequence found by Fischer and colleagues matches the deduced sequence (Fig. 5) exactly, with the exception that the C-terminal Val in the sequence of Fischer *et al.* is a Leu in the deduced sequence. The obvious heterogeneity at the N-terminus of the enzyme purified in these two different laboratories, together with the finding of Fischer *et al.* (personal communication) that approx. 80% of the enzyme in their preparation lacked the first two residues (Glu-Leu), undoubtedly reflects artifactual proteolytic modification during purification of the enzyme, previously shown to occur with the Type I isozyme (54,113,114).

The above observations made it evident that clone 12-1.3C represented only the N-terminal region of the rat Type II isozyme. Exhaustive rescreening of the cDNA library (Library 1, Table III), from which 12-1.3C had been derived, failed to provide additional clones containing sequence for the C-terminal half of the enzyme. Several other libraries, No. 2-9 of Table III, were obtained from other investigators or were synthesized using RNA from tissues known to contain the Type II enzyme. Each library was screened with the Type II hexokinase cDNA clone 12-1.3C in order to identify clones containing sequence for the remaining portion of the Type II isozyme.

Extensive screening of a rat liver cDNA library (No. 2, Table III) generated six clones identical to clone 12-1.3C as determined by size, *Hae*III digestion pattern, and partial sequence analysis. The plasmid libraries (No. 3, Table III) produced only 3

clones whose inserts (500-800 bases in length) were completely contained within the sequence of 12-1.3C. The muscle and epididymal fat pad libraries (No. 4-5, Table III), using UniZAPII as the cloning vector, produced no positive recombinants which placed the quality of these libraries in question. Screening of the other two libraries constructed in UniZAPII (No. 6 and 7, Table III) yielded clones containing inserts of either 3.8 kb or 5.7 kb in length. Both sets of clones contained a 3.4 kb fragment upon double digestion with restriction enzymes *EcoR*I and *Xho*I, and also either a 0.4 kb or a 2.3 kb fragment. The 0.4 kb fragment was identical to the 400 bases at the 5' end of 12-1.3C, as determined by sequence analysis. The 2.3 kb fragment was distinct (by size and *Hae*III digestion pattern) from any other clones investigated to date. However, Southern blot and partial sequence analyses indicated no homology between the 3.4 kb fragment and any known rat hexokinase cDNA. Since the recombinants from both the kidney and spleen libraries appeared to contain at least some cloning artifact(s), i.e. the unidentified 3.4 kb fragment, the search for the cDNA encoding the C-terminal portion of Type II hexokinase focused on other libraries.

Therefore, two random-primed cDNA libraries were constructed in λgt10, following the procedure of DeWitt and Smith (100) and using mRNA isolated from either chronically-stimulated rat skeletal muscle in which the level of Type II hexokinase is elevated more than 10-fold (29,31), or normal muscle tissue. If the level of the Type II isozyme in stimulated muscle tissue reflected transcriptional regulation, elevated levels of Type II hexokinase mRNA would also be expected.

These two unamplified libraries were screened under high stringency conditions using a 407 bp fragment from the 3' end of clone 12-1.3C as probe. One recombinant (2.3 kb in length and designated RM5) was isolated from the library synthesized using mRNA from normal tissue (No. 8, Table III). Three recombinants, designated RG2A, RG2B and RG2C, were isolated from the library constructed using mRNA from stimulated tissue (No. 9, Table III). Restriction mapping indicated that clones RG2A and RG2C were contained entirely within clone RG2B. *Hae*III digestion patterns (Fig. 6, lane 2 and 3) and partial sequence analysis indicated that clones RM5 and RG2B were identical. Therefore, further studies concentrated on clone RG2B.

Clone RG2B contained an insert of 2236 bp. The restriction map and sequencing strategy are indicated in Fig. 4; the sequence of RG2B corresponds to nucleotides at positions 1399 to 3634 in Fig. 5. The first 164 nucleotides at the 5' end of RG2B were identical in sequence to the 164 residues at the 3' end of clone 12-1.3C. A single open reading frame extended for 1548 nucleotides, followed by a TAG termination codon and a 3' untranslated region of 687 bases. Overlapping of the sequence from clones 12-1.3C and RG2B produced the composite nucleotide sequence shown in Fig. 5, from which the complete amino acid sequence (917 residues) of rat Type II hexokinase was deduced.

Isolation of Genomic Clones for Rat Type II Hexokinase. In order to: (1) confirm the overlap of the cDNA clones, (2) identify the 5' end of the message for Type II hexokinase, and (3) characterize its gene, a λCharon 4A rat genomic library was probed with three unique restriction fragments of the cDNA for this isozyme. A

Figure 6. *Hae*III restriction patterns of cDNA clones for Type II hexokinase (C-terminus). Shown is an EtBr-stained 5% polyacrylamide gel containing cDNA restricted with the endonuclease *Hae*III. Lane 1 contains pBR, digested with *Hae*III, which was used as size markers. The lengths (in bp) of several fragments in lane 1 are shown to the left. Lanes 2 and 3 contain *Hae*III restricted cDNA from clones RM5 and RG2B. Lane 4 shows the *Hae*III restriction pattern for the cDNA from clone 12-1.3C, which was used as the probe in the isolation of clones RM5 and RG2B.

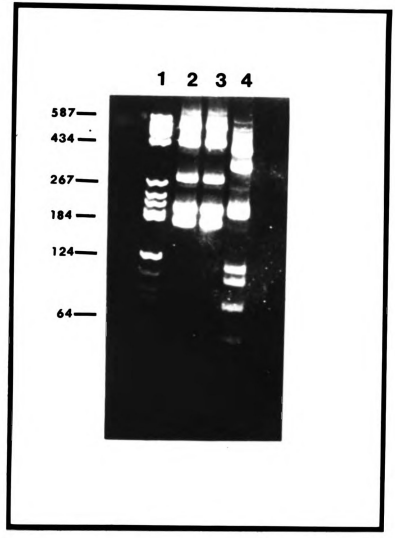


FIGURE 6

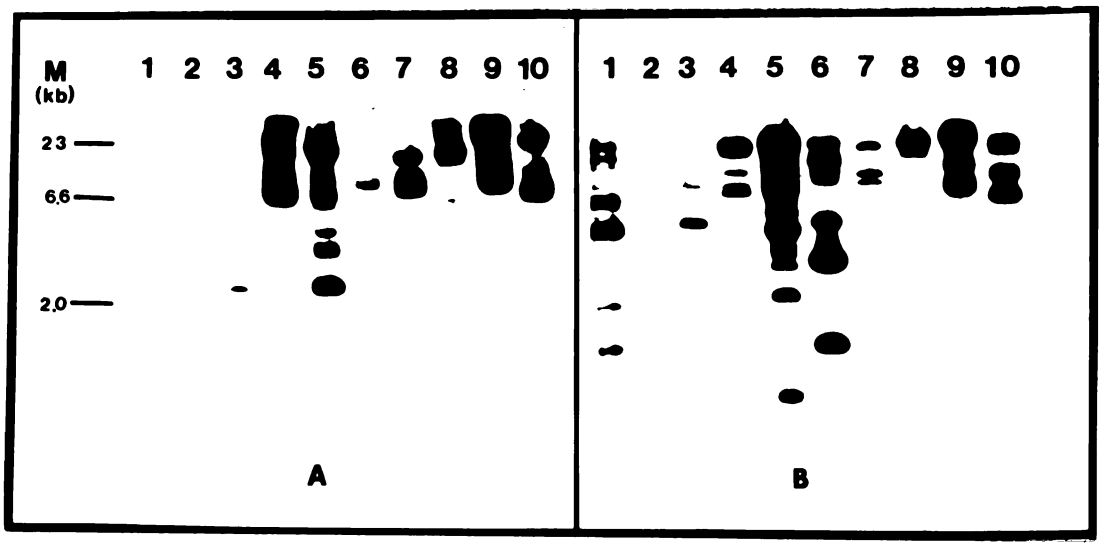
diagram of Type II cDNA clones 12-1.3C and RG2B, with relevant restriction sites indicated, is shown below. The restriction sites are: 1, 5, 10, *Eco*RI; 2, *Xho*I; 3, *Bgl*III; 4, *Sty*I; 6, *Sph*I; 7-9, *Hae*III. In the following results, these restriction sites are referenced in parentheses.

When the genomic library was probed with the 407 bp 3' region *Sty*I-*Eco*RI fragment (4-5) of clone 12-1.3C, two recombinants designated 3G3A and 3G5A were isolated. Further analysis was performed only on clone 3G3A, since the two clones appeared to be identical by restriction analysis. Eight recombinants, whose insert sizes ranged from 9 kb to 20 kb, were isolated when this same genomic library was probed with a 526 bp *Eco*RI-*Bgl*II (1-3) fragment from the 5' end of clone 12-1.3C. These clones were designated 5G2A, 5G3A, 5G4B, 5G4D, 5G5A, 5G5B, 5G6A, and 5G6C. Screening of the genomic library with the 3' 1928 bp *Sph*I-*Eco*RI fragment (6-10) of cDNA clone RG2B as the probe resulted in one additional recombinant, designated TG3A.

Southern blots of all 10 recombinants isolated with the three cDNA probes were used to align the genomic clones within the cDNA for Type II hexokinase. The probes used in the Southern analyses were restriction fragments of cDNA clones

12-1.3C and RG2B. Autoradiographs of these Southern blots are shown in Fig. 7 and 8. Each blot in Fig. 7 contained DNA from the 10 genomic clones restricted with *EcoRI*, fractionated on 0.8% agarose gels and transferred to nitrocellulose. Panel A, Fig. 7, was hybridized to the 5' 526 bp *EcoRI*-*BglII* (1-3) fragment of cDNA clone 12-1.3C. Eight of the ten genomic clones reacted with this probe to varying degrees. Only clone TG3A, in Fig. 7B, did not react with the *BglII*-*StyI* (3-4) fragment of clone 12-1.3C (nucleotide 525-1137, Fig. 5). Clones 3G3A, 5G4B and 5G4D (Fig. 7C) reacted most intensely with the 3' 407 bp *StyI*-*EcoRI* fragment (4-5) of clone 12-1.3C, while clones 5G2A and 5G5A demonstrated some homology to the probe. As shown in Fig. 7D, clones TG3A and 5G4D gave the strongest signal when hybridized to the 3' 1928 bp (6-10) of cDNA clone RG2B. Genomic clone 3G3A also showed some crossreactivity to this probe. The results of additional Southern blots used to further define the results of Fig. 7 are shown in Fig. 8. As shown in panel A (Fig. 8), reacted with the 5' 153 bp *EcoRI*-*XhoI* portion of clone 12-1.3C, genomic clone 5G3A (lane 3) hybridized strongly to the probe; clone 5G4B (lane 2) gave a slight signal indicating little sequence in common with the probe. Genomic clone 5G6C (lane 1) gave no signal. Panel B (Fig. 8), containing clones TG3A, 3G3A, and 5G4D (lanes 4-6, respectively) was reacted with a 207 nucleotide *HaeIII*-*HaeIII* fragment (7-8) of clone RG2B (near the 3' end of the open reading frame in the Type II cDNA). Clone TG3A gave the strongest signal, with clone 5G4D showing some crossreactivity and clone 3G3A showing no sequence similarity. Only clone TG3A (Fig. 8, Panel C,lane 9) gave a signal when probed with the 3' 543 bp *HaeIII*-*EcoRI* fragment (9-10)

Figure 7. Southern blot analyses of ten clones containing genomic DNA for Type II hexokinase. Blots A-D are identical, each containing *EcoR*I restricted DNA from 10 genomic clones isolated using Type II cDNA as the probe. The genomic clones, with their respective lane designations, are: 3G3A (1), TG3A (2), 5G2A (3), 5G3A (4), 5G4B (5), 5G4D (6), 5G5A (7), 5G5B (8), 5G6A (9), and 5G6C (10). The radiolabelled probes used for each high stringency hybridization were from the following regions of two cDNA clones: Blot A, 5' 526 bp *EcoR*I-B*glI*I (1-3) fragment of clone 12-1.3C; Blot B, 612 bp B*glI*I-S*tyI* (3-4) fragment of clone 12-1.3C; Blot C, 3' 407 bp S*tyI*-E*coR*I (4-5) fragment of clone 12-1.3C; Blot D, 3' 1928 S*phI*-E*coR*I (6-10) fragment of clone RG2B. The numbers in parentheses refer to restriction sites in the schematic diagram of clones 12-1.3C and RG2B shown on pg 56. Shown to the left of Blots A and C are the sizes (in kb) of several fragments of λbacteriophage digested with H*ind*III, used as size markers.


M
(kb)
1 2 3 4 5 6 7 8 9 10
23
6.6
2.0
A
1 2 3 4 5 6 7 8 9 10
B

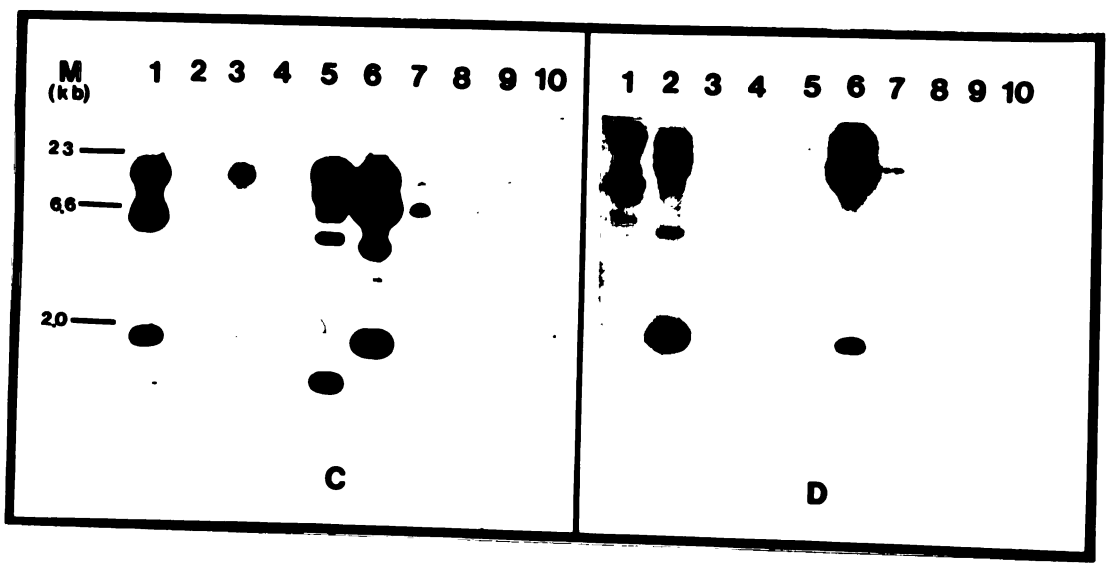

M
(kb)
1 2 3 4 5 6 7 8 9 10
23
6.6
2.0
C
1 2 3 4 5 6 7 8 9 10
D

FIGURE 7

Figure 8. Southern analysis of clones containing 5' and 3' genomic sequences for Type II hexokinase. Blots A-C contain DNA from genomic clones restricted with *EcoR*I, electrophoresed in 0.8% agarose gels, and transferred to nitrocellulose. Each blot was probed under high stringency with radiolabelled cDNA fragments from clones 12-1.3C and RG2B. Blot A, containing DNA from clones 5G6C, 5G4B and 5G3A in lanes 1-3, respectively, was probed with the 5' 153 bp *EcoR*I-X*ho*I (1-2) fragment of clone 12-1.3C. The restricted DNA from clones TG3A, 3G3A, and 5G4D (lanes 4-6 of Blot B, respectively) were hybridized with a 207 H*ae*III (7-8) fragment of clone RG2B. Restricted DNA from clones 3G3A, 5G4D, and TG3A (Blot C, lanes 7-9 respectively) were probed with the 3' 543 bp H*ae*III-E*coR*I (9-10) fragment of clone RG2B. The numbers in parentheses refer to restriction sites in the schematic diagram of clones 12-1.3C and RG2B shown on pg 56. Indicated to the left of Blot A are the sizes (in kb) of several fragments of λbacteriophage DNA digested with H*ind*III.

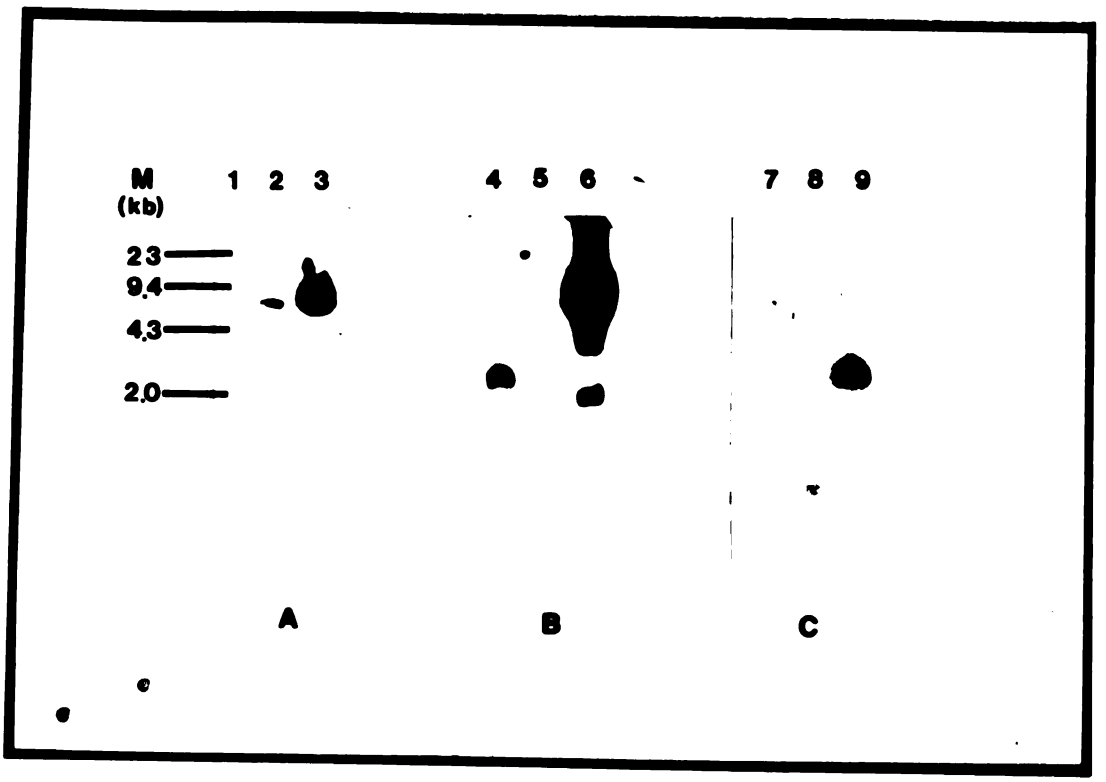


FIGURE 8

of cDNA clone RG2B. Genomic clones 3G3A and 5G4D (Panel C, lanes 7 and 8, respectively) demonstrated no sequence similarity with the probe. The signals seen in the 23 kb region of each blot in Figs. 7 and 8 are likely due to incomplete digestion of the genomic DNA, as the cloning vector Charon 4A accepts inserts totally 20 kb or less.

The results from the aforementioned Southern analyses were used to align the five clones containing pertinent sequence information or the largest unique inserts, as shown in Fig. 9. Each clone spans those portions of cDNA that gave positive results on the Southern blots. The 5' and 3' endpoints of these clones have been positioned within the regions of cDNA that hybridized to these clones. These positions are only approximate and have not been further defined by sequence analysis. The approximate insert sizes for the clones represented in Fig. 9 are: 5G3A, 15 kb; 5G4B, 16 kb; 5G4D, 20 kb; 3G3A, 15 kb, TG3A, 10 kb. From inspection of this alignment, it is likely that the gene encoding Type II hexokinase is at least 35 kb in length.

Of the 10 genomic clones isolated, Southern blots of only clone 5G3A reacted to a significant degree with the 5' *EcoRI*-*XhoI* 153 bp fragment (1-2) of cDNA clone 12-1.3C (Fig. 8A, lane 3). Thus, this clone, containing a 15 kb insert, was further analyzed for additional 5' message and promoter sequences for Type II hexokinase (discussed below). Sequence and restriction map analyses indicate that clone 5G3A contains approx 8 kb of nucleotide information upstream from the 5' end of the cDNA clone 12-1.3C. A partial restriction map of clone 5G3A is shown in the

Figure 9. Alignment of relevant clones containing genomic DNA for rat Type II hexokinase. Shown is the composite figure which contains the coding region for Type II hexokinase, and the 5' and 3' untranslated regions of the cDNA. Below this composite drawing are five genomic clones isolated and aligned using portions of clones 12-1.3C and RG2B as the probes (see Fig. 2, Material and Methods Chapter). The approximate size (in kb) is indicated above each genomic clone. The slash marks (//) indicate that the genomic clones are longer than represented, and are not drawn to scale with the composite figure for Type II hexokinase. The dashed line at the 5' end of clone 5G3A indicated sequence upstream from the 5' end of cDNA clone 12-1.3C. Portions of this region of clone 5G3A were analyzed for the transcription initiation sites and promoter sequences for Type II hexokinase.

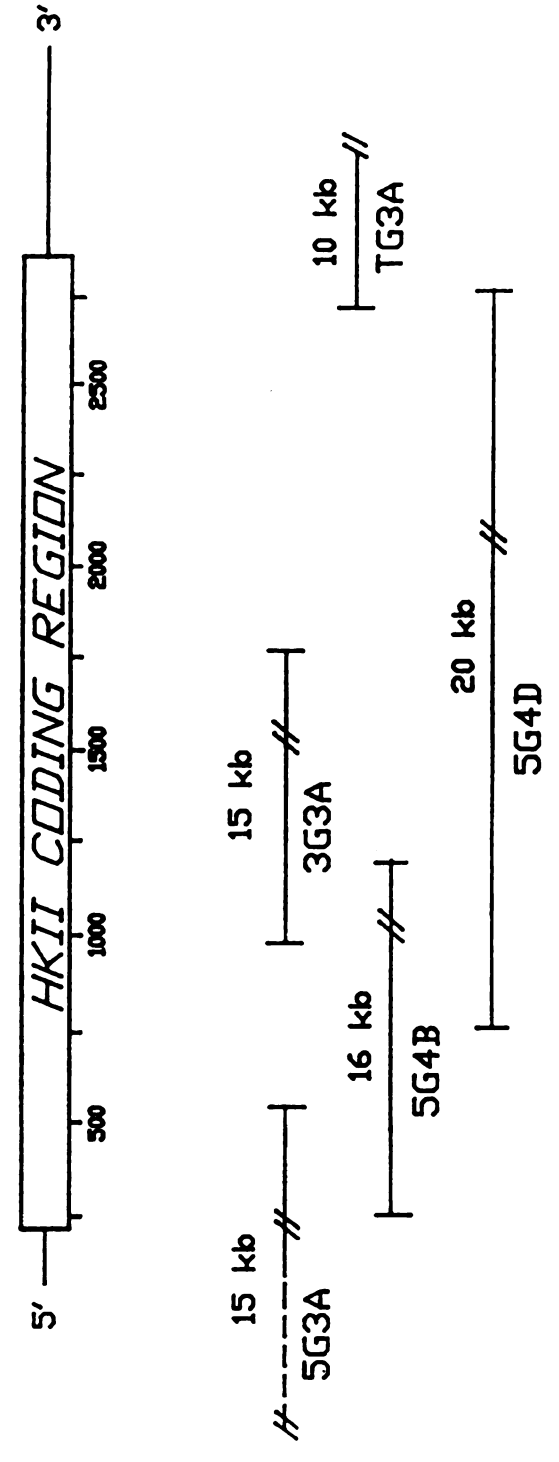
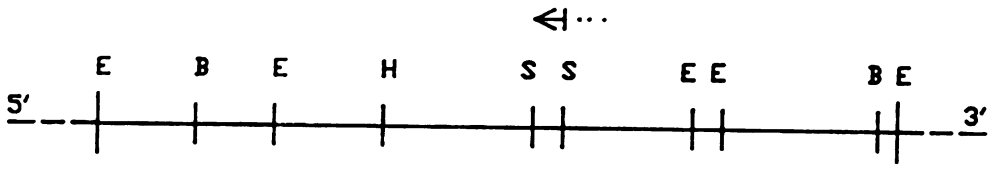

3'
HKII CODING REGION
500
1000
1500
2000
2500
5'
10 kb
TG3A
15 kb
3G3A
20 kb
5G4D
16 kb
5G4B
15 kb
5G3A

FIGURE 9

schematic below. The restriction site abbreviations are: B, *BamHI*; E, *EcoRI*; H, *HindIII*; S, *SmaI*. The arrow above the *SmaI* fragment delineates the portion of this clone that was analyzed for promoter sequences. The arrow points in the direction upstream from the 5' end of the cDNA for Type II hexokinase. The dotted line indicates cDNA sequences. The dashed lines represent λCharon 4A vector arms.



Confirmation of the overlap of the cDNA clones was obtained by isolation of a genomic DNA clone (3G3A) containing segments that span the region of overlap. After restriction with various endonucleases, Southern blots of genomic clone 3G3A were probed under high stringency conditions (as above) to identify fragments including sequences found in the 3' region of clone 12-1.3C. The results of these Southern blots are shown in Fig. 10. Panel A contains *EcoRI* digested DNA from clone 3G3A. Hybridization with the 407 bp fragment from the 3' end of cDNA clone 12-1.3C resulted in signals from the 1.6 kb and 6.2 kb fragments of the clone. The 6.2 kb fragment was isolated and restricted with several endonucleases. A Southern blot of these digests, probed with the 3' end of clone 12-1.3C, gave results shown Fig. 8B. The 1.6 kb *EcoRI* fragment of clone 3G3A and the 2 kb *SphI* fragment

Figure 10. Southern analysis of genomic clone 3G3A. Genomic clone 3G3A (or portions thereof) was restricted with endonucleases, electrophoresed in an 0.8% agarose gel, and transferred to nitrocellulose. Blot A, containing clone 3G3A digested with *EcoR*I, was probed with the 3' 407 bp *Sty*I-*EcoR*I fragment of cDNA clone 12-1.3C. Lanes 1-10 of Blot B contain DNA from the 6.6 kb fragment of genomic clone 3G3A, restricted with various endonucleases. The restriction enzymes used to digest DNA from the 6.6 kb fragment in each lane were: 1, *Sac*II; 2, *Sac*I; 3, *Sma*I; 4, *Hind*III; 5, *BamH*I; 6, *Bgl*II; 7, *Hinc*II; 8, *Sph*I; 9, *Pvu*II; and 10, *Pst*I. Blot B was also hybridized to the 3' 407 bp fragment of 12-1.3C, mentioned above. Indicated to the left of both blots are the sizes (in kb) of several fragments of λbacteriophage digested with *Hind*III.

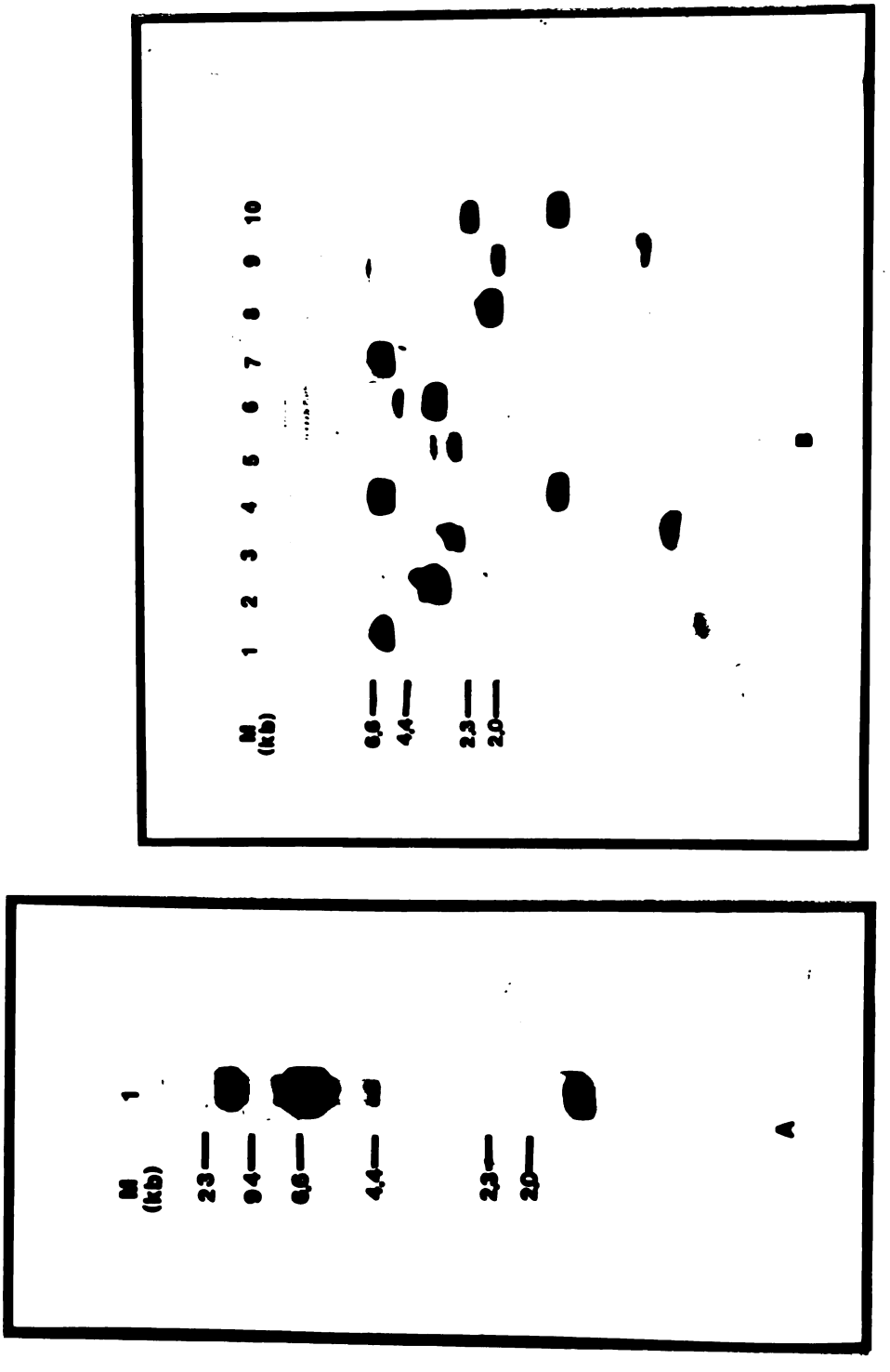


FIGURE 10

from the 6.2 *EcoR*I fragment of clone 3G3A (lane 8, Fig. 10B) were subcloned into M13 (mp18 and mp19). Initial sequence analysis located 3 regions (one at the 3' end of the 1.6 kb fragment, and the 5' and 3' ends of the 2 kb *Sph*I fragment) containing sequences identical to clone 12-1.3C. Those 3 regions were sequenced as indicated in Fig. 4. These included coding regions spanning the section from nucleotides 1072 to 1705 (Fig. 5). Two introns divided this sequence into segments comprised of nucleotides 1072-1227, 1228-1461, and 1462-1705.

Isolation of a Partial cDNA Clone for Novikoff Tumor Hexokinase. An unamplified cDNA library (No. 10, Table III) was prepared in λgt10, using oligo-dT as primer and mRNA isolated from Novikoff ascites tumor cells. Screening of this library under high stringency conditions, using the 407 bp fragment representing the 3' region of clone 12-1.3C as probe, produced three positive recombinants. These were designated NK3B, having an insert of approx. 2300 bp, and NK3A and NK3C, each with inserts of approx. 500 bp. *Hae*III restriction analysis and partial sequencing indicated that NK3A and NK3C were contained entirely within NK3B, which was further characterized and found to be *identical* to the previously isolated RG2B (Fig. 5).

Northern Blot Analysis of mRNA from Rat Skeletal Muscle and Novikoff Ascites Tumor Cells. Polyadenylated mRNA from rat skeletal muscle and Novikoff ascites tumor cells was examined by Northern blotting (Fig. 11). Duplicate blots were probed with cDNA representing either the N-terminal or C-terminal half of Type II hexokinase; these were the insert from clone 12-1.3C (nucleotides 1-1562 in

Figure 11. Northern blot analyses of rat Type II hexokinase mRNA. Blots A and B are duplicates, and each contains approx. 3 μg of Novikoff rat hepatoma mRNA and 20 μg of rat skeletal muscle mRNA in lanes 1 and 2, respectively. Blot A was probed with radiolabelled cDNA corresponding to nucleotides 1705-3082, while the probe for blot B was cDNA corresponding to nucleotides 1-1562 (nucleotide sequences in Fig. 3). The relative positions of the 18S and 28S bands of ribosomal RNA are indicated at the left.

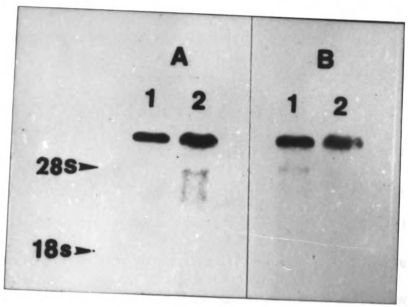


FIGURE 11

Fig. 5) and a *Sph*I/*Kpn*I fragment from clone RG2B (nucleotides 1705-3082, Fig. 5), respectively. Both probes hybridized to a 5.2 kb message, found in both skeletal muscle and tumor mRNA.

Identification of the Transcription Initiation Sites. The S1 nuclease protection assay was used to locate the transcription initiation site of the Type II hexokinase gene. The S1 probe was synthesized from a 630 bp *Sma*I fragment of the genomic clone 5G3A that had hybridized to the 5' 153 bp of the cDNA for Type II hexokinase in Southern blots. After hybridization of the S1 probe to poly(A)$^+$ RNA from either tumor or normal muscle tissue, the products were treated with S1 nuclease and the protected fragments separated on a 9% polyacrylamide 7M urea gel. A portion of an autoradiograph from one such experiment is shown in Fig. 12. Lanes 1-7 show pBR size markers and dideoxynucleotide sequencing reactions from clone 5G3A used to define the region protected from S1 digestion. The sequencing reactions were generated from the same *Sma*I fragment and primer used to produce the S1 nuclease probe. Lanes 6 and 7 contain the products from tumor and skeletal muscle mRNA, respectively. There were two protected fragments in each lane, the predominant fragments having an approximate size of 315 nucleotides and the minor fragments slightly longer.

A primer-extension assay was also used to identify the transcription initiation site. A portion of an autoradiograph, from a 9% denaturing gel, obtained from one of these experiments is shown in Fig. 13. Lane 1 shows pBR size markers with the fragment lengths indicated at the left. Lanes 2-5 are dideoxynucleotide sequence

Figure 12. S1 nuclease protection assay results. Shown is a portion of an autoradiograph from an S1 nuclease protection assay, whose products were separated in a 9% denaturing polyacrylamide gel. Lane 1 contains pBR digested with M*sp*I as size markers, with fragment lengths (bp) indicated to the left. The S1 nuclease protection experiment was performed using approx 3 μg of poly(A)$^+$ RNA from Novikoff rat hepatoma cells (lane 6) or 20 μg of poly(A)$^+$ RNA from rat skeletal muscle (lane 7). Lanes 2-5 show A, C, G, and T dideoxynucleotide sequencing reactions, respectively, used to define the relative position of the protected products.

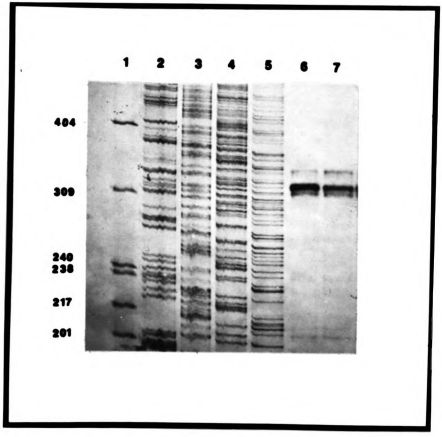


FIGURE 12

Figure 13. Primer-extension results. Shown is a portion of an autoradiograph of primer-extension products electrophoresed in a 9% denaturing gel. Lane 1 contains pBR (digested with M*spI*) as markers, with the fragment lengths (bp) indicated to the left. Lanes 2-5 show A, C, G, and T dideoxynucleotide sequencing reactions (respectively) that define the relative nucleotide positions of the extension products. Lanes 6 and 7 show the extension products from approx 3 μg of Novikoff hepatoma mRNA and 20 μg of rat skeletal muscle mRNA, respectively. Lanes 6 and 7 were exposed approx twice as long as the other lanes.

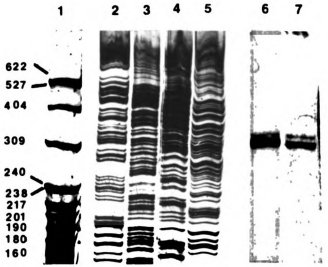


FIGURE 13

reactions synthesized from the 630 bp *Sma*I fragment of genomic clone 5G3A and the same primer used in the extension assays. Lanes 6 and 7 show the DNA products of an extension reaction using poly(A)$^+$ RNA from Novikoff tumor cells and normal rat skeletal muscle tissue, respectively. Again, there were two synthesized fragments in each extension lane, and the predominant bands were approximately 315 bases in length. Shown in Fig. 14 and 15 are the results of two primer-extension experiments, similar to the aforementioned work. The products of these extension experiments were electrophoresed in 5% denaturing polyacrylamide gels. In each figure, lane 1 contains pBR size markers with the 309 bp fragment indicated at the left. Lanes 2-5 are the dideoxynucleotide sequences of the *Sma*I fragment of genomic clone 5G3A. Lanes 6 and 7 of Fig. 14 contain extension products from muscle and tumor mRNA, respectively. Lane 6 of Fig. 15 contains the extension products from approx 1 μg Novikoff hepatoma mRNA; indicated to the right of this lane is the nucleotide sequence surrounding the transcription initiation site, determined from lanes 2-5, Fig. 13. It is clear from Fig. 14 (lanes 6 and 7) that, in both muscle and tumor, extension products were 343-348 nucleotides long; however, the majority of the extension products were in the range of 320-328 nucleotides long. In this region, the strongest bands (as seen in Fig. 15) were 324 and 325 nucleotides in length including the primer and correspond to initiation at either an adenine or guanine residue. The adenine, the first intense base near the center of the primary region, has been designated as position +1 for numbering the nucleotides in this portion of the gene for Type II hexokinase.

Figure 14. Identification of the transcription initiation region of the Type II hexokinase gene. Shown is a portion of an autoradiograph from a primer extension experiment. The products were electrophoresed in a 5% denaturing gel, for 15 hrs. Lane 1 is M*sp*I restricted pBR used as the size marker, with the 309 bp fragment indicated at the left. Lanes 2-5 show A, C, G, and T dideoxynucleotide sequencing reactions (respectively) used to define the nucleotide positions of the extension products. Lanes 6 and 7 contain the extension products from approx 20 μg of rat skeletal muscle mRNA and 3 μg of Novikoff hepatoma mRNA, respectively.

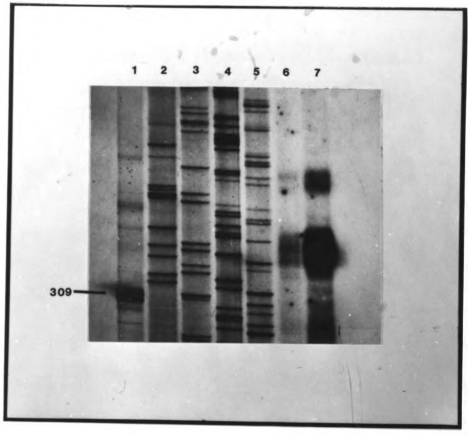


FIGURE 14

Figure 15. Identification of the transcription initiation site of the Type II hexokinase gene. A portion of an autoradiograph from a primer extension experiment is shown. The extension products were separated on a 5% denaturing gel, electrophoresed for 15 hrs. Lane 1 is M*sp*I restricted pBR used as the size marker, with the 309 bp fragment indicated at the left. Lanes 2-5 show A, C, G, and T dideoxynucleotide sequencing reactions (respectively) used to define the nucleotide positions of the extension products. Lane 6 shows the primer extension products from approx 1 μg Novikoff hepatoma mRNA. The nucleotides surrounding the transcription initiation site are shown to the right. An asterisk (*) marks the two nucleotides giving the strongest signal, with +1 designating the adenine residue as the first predominant site.

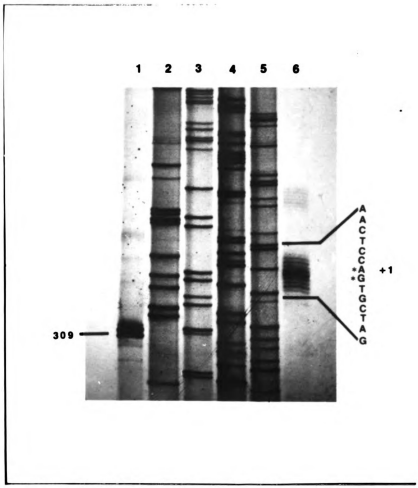


FIGURE 15

Type II Hexokinase Promoter. Genomic clone 5G3A was the only recombinant to demonstrate a high degree of cross-reactivity with the extreme 5' portion of the cDNA for Type II hexokinase. After restriction with several endonucleases, Southern blots of this clone were probed to locate fragments that contained the 5' sequence of the Type II isozyme. Figure 16 shows the results of this hybridization. Fragments hybridizing to the 5' 153 bp of Type II cDNA varied in length from approx 600 bp to 20 kb. A *Sma*I fragment of genomic clone 5G3A (near the bottom of lane 8, Fig. 16), 630 bp in length, was isolated and subcloned into M13 (mp18 and mp19) for complete sequence analysis. This sequence is shown in Fig. 17. Examination of this sequence, along with the 5' sequence from the previously isolated cDNA (Fig. 5), places the predominant transcription start site 465 nucleotides upstream from the ATG initiation codon. There is a TATA-like box centered 25 nucleotides upstream from the adenine at position +1; a CCAAT sequence starts at position -79, with an inverted CCAAT sequence starting at -135. There are also 3 Sp1 sequences starting at nucleotides -55, -123 and -212, and a potential cAMP reponse element at -64. The 260 nucleotides upstream from the transcription start site is GC rich (70% G+C) with a ratio of CpG to GpC of 0.63. The 3' end of this genomic fragment corresponds identically to the 101 nucleotides at the 5' end of Type II hexokinase cDNA clone 12-1.3C.

To verify that this genomic fragment did indeed possess promoter capability, a portion of this fragment (272 base *Sma*I-D*de*I fragment, -260 to + 12 of Fig. 17) was subcloned into the vector $p(C_2AT)_{19}$. Recombinants containing the Type II promoter

Figure 16. Southern analysis of genomic clone 5G3A. Shown is an autoradiograph of genomic DNA from clone 5G3A probed with the 5' 153bp *EcoR*I-*Xho*I fragment of cDNA clone 12-1.3C. The genomic DNA from clone 5G3A was restricted with various endonucleases, electrophoresed in an 0.8% agarose gel and transferred to nitrocellulose. The restriction enzymes used were: 1, *EcoR*I; 2, *EcoR*I + *BamH*I; 3, *BamH*I; 4, *EcoR*I + *Hind*III; 5, *Hind*III; 6, *EcoR*I + *Xho*I; 7, *Xho*I; 8, *Sma*I. The blot in lane 8 was exposed approximately twice as long as the other blot. The sizes (in kb) of the λbacteriophage/*Hind*III fragments used as markers are indicated to the left of the blot.

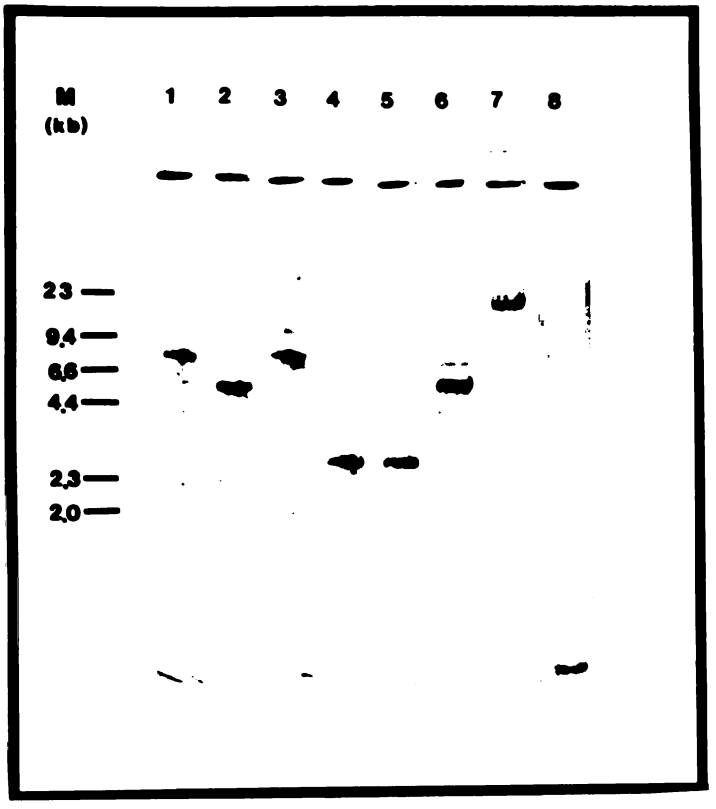
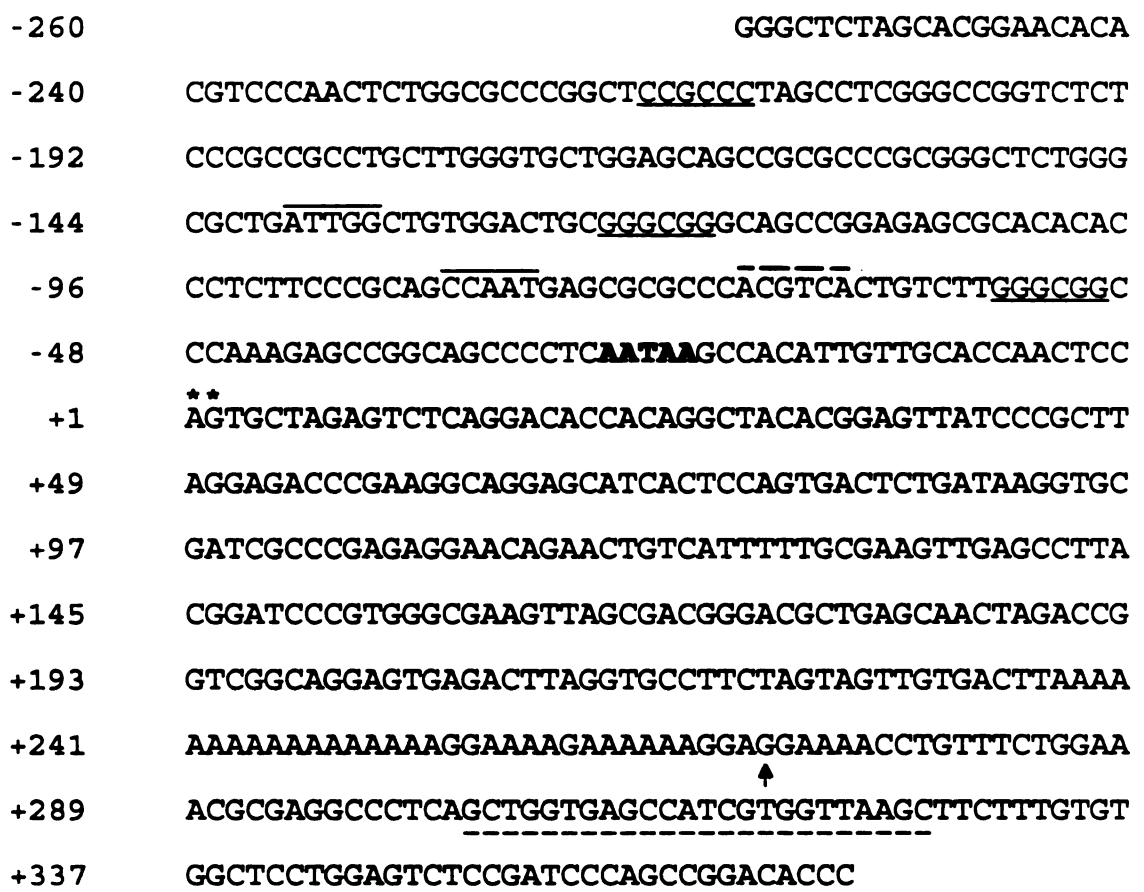


FIGURE 16

Figure 17. Sequence of the Type II hexokinase promoter region. This figure shows the sequence of the 630 nucleotide S*ma*I fragment of genomic clone 5G3A. It contains the 260 bases of DNA that are on the 5' side of the adenine residue designated as +1 (primary transcription initiation site) and the 369 bases to the 3' side of this initiation site. The overlined nucleotides are the 2 CCAAT sequences; a TATA-like box is marked in bold print; three potential Sp1 binding sites are underlined. A potential cAMP response element is marked by a dashed overline; the A and G residues at the transcription initiation site are marked with an asterisk (*). The primer used for the S1 nuclease protection and primer-extension assays was synthesized complementary to nucleotides +303-326, which are marked by a dashed underline. The arrow (↑) delineates the 5' end of the cDNA for Type II hexokinase (Fig 3). The D*de*I site, which is the 3' end of the promoter fragment in $p(C_2AT)_{19}$, is marked by a closed circle (●) under the nucleotide.


-260 GGGCTCTAGCACGGAACACA
-240 CGTCCCAACTCTGGCGCCCGGCTCCGCCCTAGCCTCGGGCCGGTCTCT
-192 CCCGCCGCCTGCTTGGGTGCTGGAGCAGCCGCGCCCGCGGGCTCTGGG
-144 CGCTGATTGGCTGTGGACTGCGGGCGGGCAGCCGGAGAGCGCACACAC
-96 CCTCTTCCCGCAGCCAATGAGCGCGCCCACGTCACTGTCTTGGGCGGC
-48 CCAAAGAGCCGGCAGCCCCTCAATAAGCCACATTGTTGCACCAACTCC
+1 AGTGCTAGAGTCTCAGGACACCACAGGCTACACGGAGTTATCCCGCTT
+49 AGGAGACCCGAAGGCAGGAGCATCACTCCAGTGACTCTGATAAGGTGC
+97 GATCGCCCGAGAGGAACAGAACTGTCATTTTTGCGAAGTTGAGCCTTA
+145 CGGATCCCGTGGGCGAAGTTAGCGACGGGACGCTGAGCAACTAGACCG
+193 GTCGGCAGGAGTGAGACTTAGGTGCCTTCTAGTAGTTGTGACTTAAAA
+241 AAAAAAAAAAAAAGGAAAAGAAAAAAGGAGGAAAACCTGTTTCTGGAA
+289 ACGCGAGGCCCTCAGCTGGTGAGCCATCGTGGTTAAGCTTCTTTGTGT
+337 GGCTCCTGGAGTCTCCGATCCCAGCCGGACACCC

region were used as the template in a cell-free *in vitro* assay system. The results of the transcription assays are shown in Fig. 18. Lanes 1-5 are the results of transcription assays which included pML-GFC2 as an internal positive control. Two transcripts, approximately 280 and 380 bases in length, are seen in lanes 1-3 which contain the constructs with the Type II hexokinase promoter in the correct orientation (designated $p(C_2AT)_{19}$-HKII+7, +11, and +17, respectively). The designations for the recombinants used in the assay include a (+) denoting correct promoter orientation or a (-) for the inverse orientation; the numbers are for clone identification and are not related to sequence location. There are few, if any, transcription products 380 bases in length, in lane 4 (with the Type II promoter in the wrong orientation, designated $p(C_2AT)_{19}$-HKII-6) and lane 5 (containing $p(C_2AT)_{19}$ with no promoter). These results indicate that the region of genomic DNA laying immediately upstream of the putative transcription start site for Type II hexokinase is capable of directing transcription, at least in a cell-free system.

Expression of Type II Hexokinase in COS-1 Cells. The plasmid, pSVT7-HKII, was used to transfect COS-1 cells. As controls, COS-1 cells were also transfected with either the plasmid pSVT7 (with no insert) or with no DNA (Sham). Shown in Table IV are the results from two sets of transfections. The level of hexokinase activity in the COS-1 cells transfected with pSVT7-HKII was approximately 14 times greater than that found in cells transfected with pSVT7 or in the Sham transfection.

Figure 18. Cell-free *in vitro* transcription assay results. Shown is a section of an autoradiograph of *in vitro* transcription products from Type II hexokinase promoter constructs and pML-GFC2, using rat liver nuclear proteins. The transcription products were electrophoresed in a 6% denaturing polyacrylamide gel. In lanes 1-5, 50 ng of pML-GFC2 and 2 μg of $p(C_2AT)_{19}$ constructs were used as the template DNA. The constructs $p(C_2AT)_{19}$-HKII+7, +11, and +17 were the template DNA in lanes 1-3, respectively. The construct $p(C_2AT)_{19}$-HKII-6 was included in lane 4. The original plasmid, $p(C_2AT)_{19}$ with no promoter sequence included, was used as a negative control in lane 5. Lane 6 contains pBR digested with M*sp*I, with the size (bp) of the fragments indicated to the right.

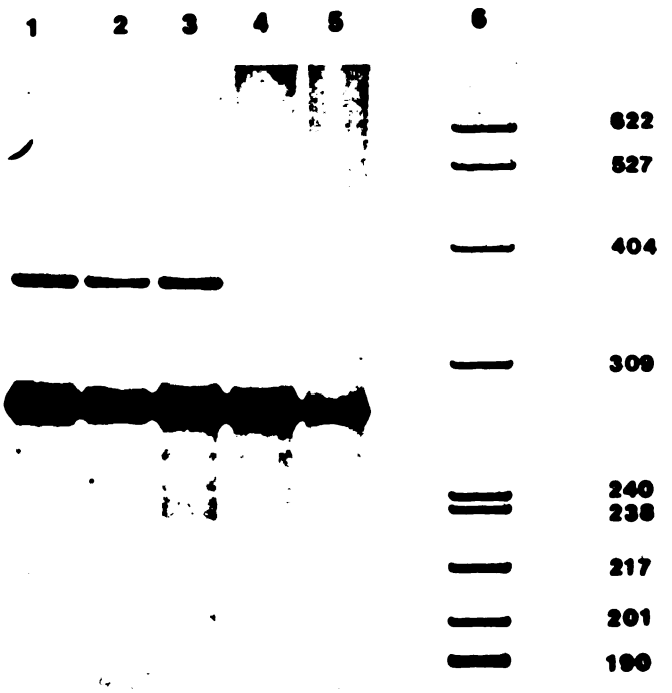


FIGURE 18

Table IV. Summary of Transfection Results

Transfection	Specific Activity (u/mg)
Sham	0.044, 0.035
pSVT7	0.047, 0.035
pSVT7-HKII	0.63, 0.52

The results of the Western blot analysis, using anti-Type I hexokinase polyclonal antibodies, are shown in Fig. 19. Lane 1 contained purified Type I hexokinase. Lanes 2 and 3 contained protein from the pSVT7-HKII and Sham transfected cells, respectively. It is readily apparent that, in the cells transfected with pSVT7-HKII (lane 2), there is a definite increase in protein that reacts with the anti-Type I hexokinase polyclonal antibodies. The band in lane 2 that reacted most intensely with these antibodies migrated slightly faster than purified Type I hexokinase. This was as expected since the reported M_r for Type II hexokinase (98) is somewhat smaller than that of the Type I isozyme. This same mobility pattern has been observed when purified Types I and II hexokinases are electrophoresed in SDS polyacrylamide gels and stained for protein. Bands corresponding to hexokinase isozymes I and II were also seen at the appropriate positions in the crude muscle extract (data not shown). There is also evidence of crossreactivity between the endogenous hexokinase (lane 3) and the Type I polyclonal antibodies, which was not unexpected. The bands below the Type II band in lane 2 are also present in the Sham extract (lane 3). It has not been determined whether the peptides reacting with the anti-Type I hexokinase antibodies are proteolytic fragments of the endogenous hexokinase or unrelated proteins. The lower band seen in lane 1 is most likely a proteolytic fragment (approx 90 kDa) of Type I hexokinase (54,96).

Figure 19. Western blot analysis of Type II hexokinase expressed in COS-1 cells. Proteins isolated from transfected COS-1 cells were electrophoresed, blotted onto nitrocellulose and reacted with antibodies, as described in Methods. Lane 1 contains 0.5 μg of purified Type I hexokinase. Lanes 2 and 3 contain 45 μg of protein from COS-1 cells transfected with pSVT7-HKII and no DNA (Sham), respectively. The blot was reacted with anti-hexokinase Type I polyclonal antibodies.

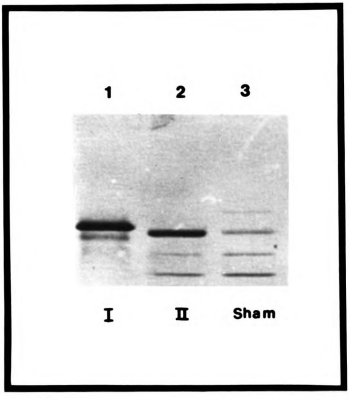


FIGURE 19

CHAPTER 4

Discussion

Amino Acid Sequence of Type II Hexokinase

The deduced amino acid sequence of rat Type II hexokinase is compared with previously determined sequences of the Type I (68), Type III (56), and Type IV (32) isozymes in Fig. 20. The striking similarity between sequences in the N- and C-terminal halves of the Type II hexokinase has been previously seen for the Type I (68) and III (56) isozymes. The similarities between these sequences and that of yeast hexokinase (86) confirm the view that all of the 100 kDa enzymes evolved by a mechanism of duplication and fusion of a gene encoding an ancestral 50 kDa hexokinase related to the yeast enzyme. A quantitative comparison of the sequence similarities among the rat isozymes is presented in Table V.

The 9 amino acid residues at the N-terminus of Type II hexokinase (Met-Ile-Ala-Ser-His-Met-Asn-Ile-Ala-Cys) are nearly as hydrophobic as the corresponding sequence in the Type I isozyme (Met-Ile-Ala-Ala-Gln-Leu-Leu-Ala-Tyr). Polakis and Wilson (54) have shown the necessity of this hydrophobic "tail" for binding of the Type I isozyme to mitochondria. Thus, these results are consistent with reports that the Type II isozyme also binds to mitochondria (39,41), although apparently with somewhat less affinity than does the Type I isozyme (116). The less hydrophobic nature of the N-terminus of Type III hexokinase may, at least partially, explain the lack of binding of this isozyme to mitochondria (41).

The kinetic differences that presumably underlie differences in physiological function of the various isozymes, e.g., the ability of P_i to antagonize inhibition of the Type I isozyme, but not the Type II isozyme, by Glc-6-P (11), clearly must depend

Figure 20. Comparison of aligned amino acid sequences of N- and C-terminal halves of rat Types I-III hexokinases and rat glucokinase (Type IV). Three or more aligned residues identical to the corresponding amino acid in either the N- or C-terminal region of Type II hexokinase are considered identities (blackened residues). Shaded residues are conservative substitutions when compared to identical aligned amino acids at a given position. In alignments of two sets of 3 identical amino acid residues, all residues are viewed as conservative substitutions.

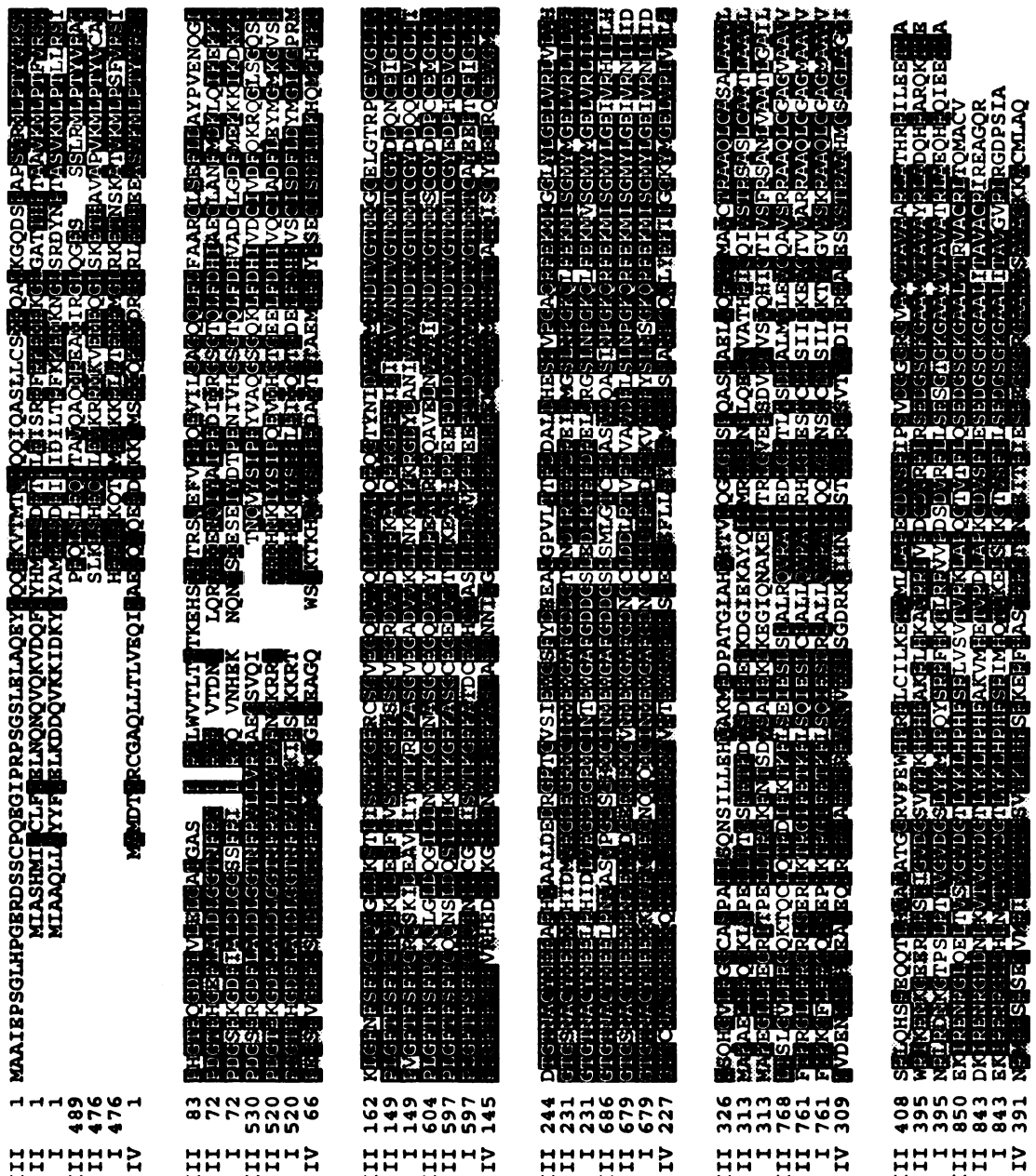


FIGURE 20

Table V. Comparison of Deduced Amino Acid Sequences of the N- and C-Terminal Halves of Rat Type II Hexokinase with Sequence of the Type IV Isozyme and Sequences of the N- and C-Terminal Halves of the Type I and III Isozymes.[a]

	N-II[b]	C-II
N-II	100	55,14
N-I	68,14	49,17
C-I	54,14	76,11
N-III	44,14	41,13
C-III	48,15	66,9
IV	52,15	53,14

a-Based on the alignments shown in Fig. 20. The first number is the percent identity, and the second number is the percent of conservative substitutions. Thus the sum of the two numbers reflects the overall similarity of the compared sequences.

b-The abbreviations used are: N-II, C-II: N-terminal half (residues 1-475) and C-terminal half (residues 476-917), respectively, of the Type II isozyme; N-I, C-I: N-terminal half (residues 1-475) and C-terminal half (residues 476-918), respectively, of the Type I isozyme (68); N-III, C-III: N-terminal half (residues 1-488) and C-terminal half (residues 489-924), respectively, of the Type III isozyme (56); IV, Type IV isozyme (glucokinase), sequenced by Andreone *et al.* (32).

on structural differences resulting from changes in amino acid sequence. Since the conservation of sequence is so extensive, the sequence comparisons in Fig. 20 focus attention on relatively limited regions which might be responsible for the observed alterations in functional properties. This is facilitated by the demonstration that catalytic activity is associated with the C-terminal half of the Type I isozyme (84,85,93) while regulatory function is associated with the N-terminal half (91,94). Presumably this same functional organization is found in the other 100 kDa enzymes.

Although differences in the N-terminal halves of the molecules can obviously not be excluded, the identity - at both the nucleotide and amino acid level - between the sequences of the C-terminal halves of Type II hexokinase and the enzyme from Novikoff ascites tumor cells suggests that there is a single Type II isozyme in both normal tissue and tumors. Further support for this view is provided by the observation that these enzymes are encoded by polyadenylated mRNAs which are indistinguishable in size. This also appears to be the case with the rat Type III isozyme since the deduced amino acid sequence of this enzyme (56) includes several segments that are identical to those of tryptic peptides (115) derived from the Type III isozyme isolated from Novikoff tumor cells (41). These observations do not support speculation (65) that the tumor isozymes are distinct in amino acid composition, and hence sequence, from the isozymes of normal tissues.

However, the Type II hexokinase in the Novikoff cell line used in the present work *was* distinct from the Type II enzyme of normal skeletal muscle when compared by ion exchange chromatography, nondenaturing gel electrophoresis, or isoelectric

focusing (T. Ureta and J.E. Wilson, unpublished results). It seems appropriate to consider possible posttranslational modifications that might account for these differences, *and* which might lead to the apparent altered function of hexokinase in highly glycolytic tumor cell lines (63,64).

Structure of Type II Hexokinase mRNA

A message size of approx 5.2 kb was seen for both tumor and skeletal muscle Type II hexokinase (Fig. 11). An open reading frame of 2751 nucleotides was contained in two cDNA clones. Included in these clones were 197 and 687 nucleotides in the 5' and 3' untranslated regions, respectively.

The identification of the transcription start sites, through primer extension and S1 mapping, indicate that the 5' untranslated region for Type II hexokinase is approximately 467 nucleotides. Transcription initiation spans a 30 base region. Diffuse transcription initiation has also been seen from both the hepatic and pancreatic glucokinase promoters (33,36). The lack of a single discrete transcription initiation site may be the result of a "weak" TATA sequence (discussed below). While the function of such long 5' leader sequences is not known, they have been identified in other proteins. For instance, the 5' untranslated region for HMG CoA reductase encompasses as many as 670 nucleotides (117). The 5' leader sequences for the ATPases (Na^+K^+) from rat and sheep kidney contain 460 and 528 nucleotides, respectively (118,119).

The 33 nucleotides immediately upstream from the 5' end of the Type II cDNA (Fig. 17) contain 28 A residues. This region is an excellent candidate for

hybridization to the oligo-dT primer used in the synthesis of several of the libraries that were screened for Type II cDNA. This observation would explain why the cDNA terminates approximately 270 downstream from the actual 5' end of the message.

The designation of the first ATG in the cDNA as the translation start site is confirmed by alignment of the N-terminal amino acid sequence of Type II with the corresponding sequence of Type I. Three nucleotides upstream from the starting "ATG" (GCAGGATGATC) is a purine (in this case an A), which is highly conserved at that position. In 97% of 699 mRNAs analyzed by M. Kozak (120), there was a purine, usually an A, at this -3 position. However, other features of the translation initiation consensus sequence (CCA/GCCATGG) (120) are not conserved.

Both tumor and skeletal muscle mRNAs were used to determine the transcription start region. The primer extension and S1 digestion results obtained from both mRNAs were identical. These results clearly indicate that the message for Type II hexokinase is the same, at least for the 5' untranslated region and the C-terminal half of the protein, in both normal and tumor tissues. These facts, in conjunction with an identical message size of 5.2 kb, make it highly unlikely that the mRNA encoding the N-terminus of this tumor hexokinase is different from the Type II message found in skeletal muscle. However, this complete message identity has yet to be verified.

The length of the message for Type II hexokinase, characterized thus far, is approximately 1300 nucleotides shorter than suggested by Northern analysis. Since

poly A^+ RNA was used for both Northern analysis and library syntheses, it seems likely that the polyadenylation signal and the polyA^+ region will be found in the 1300 nucleotides not yet isolated.

Although the role of the unusual stretch of 'Ts' near the 3' end of the cDNA (Fig. 5) is not known, it may function in several ways. It could serve as a signal for termination of transcription as proposed by Resnekov *et al* (121). However, if this is the purpose, the termination signal is approximately 1300 nucleotides upstream from the end of the transcript. It has also been suggested that such a stretch of 'Ts' may facilitate the release of the transcript unit from the genomic DNA template (122), due to the instability of (dA:rU) regions. Alternatively, another proposed function for such a sequence involves message stability. Wilson and Treisman (123) found that the shortening of the poly A^+ region of c-*fos* mRNA was much slower when the 3' AU-rich sequences were deleted. These researchers suggested that such 3' AU-rich sequences may destabilize mRNA by causing rapid removal of the poly A^+ region. However, until further research is conducted the function of this T-rich region remains unknown.

Inspection of the nucleotide sequences surrounding the overlapping regions of cDNA clones 12-1.3C and RG2B has given no clues as to why full length Type II cDNAs have yet to be isolated. As seen in Table III, a total of 10 libraries were prepared and screened in the search for the cDNA for the Type II isozyme. These libraries were synthesized by various methods, using mRNA isolated from several rat tissues known to contain the Type II isozyme. The libraries were screened with

portions of cDNA clones 12-1.3C or RG2B/NK3B. The *only* clones isolated from any of these libraries contained inserts identical to those described above for clones 12-1.3C and RG2B. Each library gave *only* the "12-1.3C-type" insert or the "RG2B-type" insert, or portions thereof; *no* library gave both, nor did any library yield a full length cDNA. Also curious is the fact that clones RG2B and NK3B were identical, with both lacking a 3' polyadenylated region, even though the Novikoff cDNA library yielding clone NK3B was synthesized using oligo-dT as primer while RG2B was isolated from a random primed library. There are no *EcoR*I restriction sites in this area that may have affected the cloning of full length cDNA, nor are there any multiple adenine residues that could serve as an additional binding site for the oligo-dT primer used in library synthesis. Clearly there is something most unusual occurring, possibly resulting from some exceptional secondary structural features in these mRNAs.

Type II Hexokinase Gene

Alignment of 5 genomic clones (Fig. 9) containing sequences complementary to Type II hexokinase cDNA indicates that the gene for this enzyme is at least 35 kb in length. Sequence analysis of portions of genomic clone 3G3A, used to verify the overlap of the two cDNA clones, generated 3 exons of length 150-250 nucleotides. Similar exon sizes have been seen in the hepatic glucokinase gene, with an average length of nine exons being 153 nucleotides (33).

Several potential regulatory sequences have been found in the promoter region, upstream from the transcription start region (Fig. 17). The sequence AATAA,

located at positions -27 to -23, may be a "TATA box," although it is not a good consensus sequence (124). Conservation of the TATA box sequence could play a critical role in the precise initiation of transcription (124,125). The lack of complete homology between the AATAA sequence in Fig. 17 and the consensus TATA sequence may explain the existence of two regions, rather than a unique site, of transcription initiation for this gene. Two CCAAT sequences, at nucleotides -79 and -135, and 3 potential Sp1 binding sites (starting at nucleotides -55, -123, and -212) have also been identified. The positions of the AATAA and CCAAT sequences are in agreement with the TATA (-25 to -30) and CCAAT box (-70 to -80) locations found in many eucaryotic genes (124). However, if the AATAA sequence is not a TATA element, transcription may be directed by an initiator (Inr) element located at or near the transcription initiation site. Even though a number of such Inr elements have been defined and investigated (reviewed in 126) no concensus sequences for these elements have been proposed. It remains to be seen whether the gene for Type II hexokinase is under the control of such Inr elements.

The 5' untranslated region (725 nucleotides) may be, at least a portion of, a CpG island as defined by Gardiner-Garden and Frommer (127). The %GC (G+C) of this genomic region is 60% with a CpG/GpC ratio of 0.77. Such CpG islands may serve an important function in either transcriptional or post-transcriptional regulation of the Type II hexokinase gene. Post-transcriptional regulation has been found for a number of genes including c-*fos* (128) and c-*myc* (129).

The nucleotides from -63 through -70 (CCACGTCA) are 75% identical to the cAMP response element (TGACGT/CC/AA/A) (130). Whether Type II hexokinase is influenced by cAMP has yet to be determined. However, it is known that transcription of glucokinase is negatively affected by cAMP (acting through glucagon (22)); the Type II isozyme may be regulated in a similar fashion.

Even though the importance of each of the aforementioned promoter elements has yet to be determined, the 260 bp genomic fragment containing these elements has the ability to direct transcription as demonstrated in a cell-free *in vitro* system. The low activity level of this promoter, relative to that of the adenovirus major late promoter (40 fold more Type II promoter was used in the transcription assays seen in Fig. 18), may indicate that elements important to promoter activity were not present in the genomic fragment used in the transcription assay. Alternatively, a negative regulatory element may be present within this 260 bp fragment. Such negative elements have been found in the S14 promoter from rat liver (111) and in the human c-*fos* promoter (131).

There also exists the possibility that nuclear factors in rat liver, important for the efficient transcription of the Type II isozyme, are present at higher concentrations immediately following birth and decline shortly thereafter. Ureta *et al.* (132) has shown that the level of Type II hexokinase in rat liver peaks during the first week after birth. The liver nuclear extracts, used for the *in vitro* transcription assay, were prepared from 6-7 week old rats. The nuclear factors necessary for efficient transcription may be deficient, or in low concentrations, in the liver of older rats.

Nuclear extracts, from tissues in which Type II hexokinase is the predominant isozyme (e.g., muscle, and adipose), may contain increased levels of factors necessary for efficient transcription of the isozyme.

Recently Alexander *et al.* proposed a 25 nucleotide sequence (AACTTTCCCGCCTCTCAGCCGAAAG) as a minimum core of a putative insulin response element (IRE) (133). No significant homology is seen between this possible insulin element and the promoter region of the Type II gene (Fig. 17). However, the IREs located by Alexander *et al.* were in a region from 270 to 490 residues upstream from the transcription start site. Since the promoter region for the Type II isozyme, characterized thus far, extends only 260 nucleotides upstream from the transcript site, it is possible that insulin responsive elements for this gene are in upstream regions not yet isolated. However, Tanaka *et al.* (35) have found no enhancer sequences in the 5' flanking region of the glucokinase gene that appear to be responsible for insulin regulation. They had investigated 5.5 kb upstream from the transcript start site. Further studies are needed to determine the existence and location of any insulin response sequences, and to confirm the role of other promoter/enhancer elements, in the Type II hexokinase gene.

CHAPTER 5

Future Work

With the availability of the cDNA (134) and promoter region for Type II hexokinase, several directions for future research have become feasible. Determination of factors controlling the transcription of the Type II hexokinase in normal and diabetic rats, as well as tumor cells, can now be undertaken. Also, with cDNAs for all the hexokinase isozymes now available, future work can focus on elucidating the structural differences that underlie the diversity of their catalytic and regulatory properties.

Several potential regulatory elements in the promoter region for Type II hexokinase have been located upstream from the transcription initiation site for Type II hexokinase, as defined in this project. Ordered deletions of this region can be tested, in *in vitro* transcription assays with the "G-free cassette" vector $p(C_2AT)_{19}$, to identify segments important in the regulation of this gene. Nuclear extracts, prepared from various tissues (e.g., liver, muscle and mammary) from normal, diabetic and insulin-treated diabetic rats, should be used in these transcription assays. The use of such nuclear extracts, with the deleted "G-free" constructs, will define promoter regions that either enhance or suppress the rate of initiation of this gene. Through this approach, it may be possible to locate insulin response elements (IREs) in this gene. However, considering the findings of Tanaka *et al* (35), it may be necessary to include regions of the Type II hexokinase gene further downstream (than investigated herein) in the search for such IREs.

Using a similar approach with fast growing tumor cell lines (e.g., Novikoff ascites hepatoma) or stimulated muscle tissue, it may be possible to identify

cis-elements, and *trans*-acting factors, responsible for increased levels of Type II hexokinase seen in both tumor cells (64) and stimulated muscle (29,31). Techniques such as "gel shift" assays (135,136) and DNase I footprint analysis (95,133) can further define the location of important regulatory elements.

The availability of the Type II cDNA makes site-directed mutagenesis a very useful tool to study the importance of specific amino acid residues implicated in catalytic or regulatory functions of hexokinase. Residues in the Type I isozyme thought to be critical in the binding of glucose are Ser 603, Asp 657, Glu 708, and Glu 742 (68,92,93); Thr 661 and the sequence Gly-Ser-Gly-Lys-Gly-Ala (896-901, Fig. 20, Discussion Chapter) may be involved in the binding of ATP in Type I hexokinase (56,68). These residues are completely conserved between the Types I, II and III isozymes (Fig. 20), which may imply conservation of function as well. Curiously, these residues are also well conserved at the corresponding positions in the N-termini of the three hexokinase isozymes, even though the N-terminal domain of Type I is not catalytically active (84,91,94). Site-directed mutagenesis of these residues, in both halves of Type II hexokinase, will help determine their function.

The 100 kDa hexokinases differ in both kinetic and regulatory properties (reviewed in ref. 11). However, because of the extensive similarities of their amino acid sequences, it is reasonable to assume that structure/function relationships found in the Type I isozyme (84) exist in the other hexokinases as well. Construction and expression of chimeric hexokinase molecules may be a very useful method to

substantiate this supposition, and to investigate the interactions between the N- and C-terminal domains of each isozyme.

The exchange of the N-terminal half of Type II hexokinase for that of the Type I isozyme will test the proposition that structural regions affecting regulatory properties (such as reversal of Glc-6-P inhibition by P_i) reside in the N-terminal domains of these isozymes. In order to facilitate this exchange, an N*co*I site was created in the Type II cDNA at nucleotide 1558 (Fig. 5, Results Chapter). This new site coincides with an N*co*I site at the same position in the Type I cDNA. This mutation does not alter the amino acid sequence, and expression of this cDNA in COS-1 cells produced active protein. Hence, the stage is now set for construction of chimeric hexokinases in which the N- and C-terminal domains of the Type I and Type II isozymes can be interchanged.

CHAPTER 6

References

1. Viñuela, E., Salas, M., and Sols, A. (1963) *J. Biol. Chem.* **238**, PC1175-PC1177.

2. Walker, D.G. (1963) *Biochim. Biophys. Acta* **77**, 209-226.

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

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

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

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

7. Easterby, J.S. (1971) *FEBS Lett.* **18**, 23-26.

8. Chou, A.C. and Wilson, J.E. (1972) *Arch. Biochem. Biophys.* **151**, 48-55.

9. Neumann, S., Falkenburg, F., and Pfleiderer, G. (1974) *Biochim. Biophys. Acta* **334**, 328-342.

10. Meunier, J.C., Buc, J., and Richard, J. (1971) *FEBS Lett.* **14**, 25-28.

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

12. Meglasson, M.D., Buch, P.T., Berner, D.K., Najafi, H., Vogin, A.P., and Matschinsky, F.M. (1983) *Proc. Natl. Acad. Sci.* USA **80**, 85-89.

13. Hughes, S.D., Quaade, D., Milburn, J.L., Cassidy, L., and Newgard, C.B. (1991) *J. Biol. Chem.* **266**, 4521-4530.

14. Liang, Y., Jeffon, T.L., Zimmerman, E.C., Najafi, H., Matschinsky, F.M. and Magnuson, M.A. (1991) *J. Biol. Chem.* **266**, 6999-7007.

15. Bernstein, R.S. and Kipnis, D.M. (1973) *Diabetes* **22**, 913-922.

16. Katzen H.M. (1967) *in* Advances in Enzyme Regulation (Weber, G., Ed.) Vol. 5, pp. 335-356, Pergamon Press, New York.

17. McLean, P., Brown, J., Walters, E., and Greenslade, K. (1967) *Biochem J.* **105**, 1301-1305.

18. Walters, E. and McLean, P. (1968) *Biochem J.* **109**, 737-741.

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

20. Frank, S.K. and Fromm, H.J. (1986) *Biochem. Biophys. Res. Commun.* **138**, 374-380.

21. Sharma, D., Manjeshwar, R., and Weinhouse, S. (1963) *J. Biol. Chem.* **238**, 3840-3845.

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

23. Colowick, S.P. (1973) *in* The Enzymes (Boyer, P.D., Ed.) 3rd Ed., Vol. 9, pp. 1-48, Academic Press, New York.

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

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

26. Green, H.J., Reichmann, H., and Pette, D. (1983) *Pflügers Arch.* **399**, 216-222.

27. Pette, D., Smith, M.E., Staudte, H.W., and Vrbová, G. (1973) *Pflügers Arch.* **338**, 257-272.

28. Simoneau, J.-A. and Pette, D. (1988) *Pflügers Arch.* **412** 86-92.

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

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

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

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

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

34. Sibrowski, W. and Seitz, H.J. (1984) *J. Biol. Chem.* **259**, 343-346.

35. Noguchi, T., Takenaka, M., Yamada, K., Matsuda, T., Hashimoto, M., and Tanaka, T. (1989) *Biochem. Biophys. Res. Comm.* **164**, 1247-1252.

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

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

38. Quaade, D., Hughes, S.D., Coats, W.S., Sestabk, A.L., Iynedjian, P.B., and Newgard, C.B. (1991) *FEBS Lett.* **280**, 47-52.

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

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

41. Radojković, J. and Ureta, T. (1987) *Biochem. J.* **242**, 895-903.

42. Parry, D.M. and Pedersen, P.L. (1983) *J. Biol. Chem.* **258**, 10904-10912.

43. Preller, A. and Wilson, J. (1992) *Arch. Biochem. Biophys.* (in press)

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

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

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

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

48. BeltrandelRio, H. and Wilson, J.E. (1991) *Arch. Biochem. Biophys.* **286**, 183-194.

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

50. Inui, M. and Ishibashi, S. (1979) *J. Biochem.* **85**, 1151-1156.

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

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

53. Rose, I.A. and Warms, J.V.B. (1967) *J. Biol. Chem.* **242**, 1635-1645.

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

55. Xie, G. and Wilson, J. (1988) *Arch. Biochem. Biophys.* **267**, 803-810.

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

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

58. Kosow, D.P. and Rose, I.A. (1972) *Biochem. Biophys. Res. Commun.* **48**, 376-383.

59. Kosow, D.P., Oski, F.A., Warms, J.V.B., and Rose, I.A. (1973) *Arch. Biochem. Biophys.* **157**, 114-124.

60. Beitner, R., Haberman, S., and Livni, L. (1975) *Biochim. Biophys. Acta* **397**, 355-369.

61. Rose, I.A. and Warms J.V.B. (1975) *Arch. Biochem. Biophys.* **171**, 678-681.

62. Warburg, O., Posener, K., and Negelein, F. (1924) *Biochem. Z.* **152**, 309-344.

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

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

65. Nakashima, R.A., Paggi, M.G., Scott, L.J., and Pedersen, P.L. (1988) *Cancer Res.* **48**, 913-919.

66. Arora, K.K., Fanciulli, M., and Pedersen, P.L. (1990) *J. Biol. Chem.* **265**, 6481-6488.

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

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

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

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

71. Holroyde, M.J., Trayer, I.P., and Cornish-Bowden, A. (1976) *FEBS Lett.* **62**, 213-219.

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

73. Trayer, I.P. and Darby, M.K. (1981) *Biochem. Soc. Trans.* **9**, 62.

74. Lawrence, G.M. and Trayer, I.P. (1984) *Comp. Biochem. Physiol.* **79B**, 233-238.

75. Poorman, R.A., Randolph, A., Kemp., R.G., and Heinrikson, R.L. (1984) *Nature* **309**, 467-469.

76. Palm, D., Goerl, R., and Burger, K.J. (1985) *Nature* **313**, 500-502.

77. Wistow, G. (1990) *J. Mol. Evol.* **30**, 140-145.

78. Yanagawa, H.A. (1978) *Insect Biochem.* **8**, 293-305.

79. Storey, K.B. (1980) *Insect Biochem.* **10**, 637-645.

80. Mochizuki, Y. and Hori, S.H. (1977) *J. Biochem.* **81**, 1849-1855.

81. Nishi, S., Seino, S., and Bell, G.I. (1988) *Biochem. Biophys. Res. Commun.* **157**, 937-943.

82. Crane, R.K. and Sols, A. (1954) *J. Biol. Chem.* **210**, 597-606.

83. Nemat-Gorgani, M. and Wilson, J.E. (1986) *Arch. Biochem. Biophys.* **251**, 97-103.

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

85. Schirch, D.M. and Wilson, J.E. (1987) *Arch. Biochem. Biophys.* **254**, 385-396.

86. Stachelek, C., Stachelek, J., Swan, J., Botstein, D., and Konigsberg, S. (1986) *Nuc. Acids Res.* **14**, 945-963.

87. Kopetzki, E., Entian K-D., and Mecke, D. (1985) *Gene (Amst.)* **39**, 95-102.

88. Anderson, C.M., Stenkamp, R.E., McDonald, R.C., and Steitz, T.A. (1978) *J. Mol. Biol.* **123**, 15-33.

89. Anderson, C.M., Stenkamp, R.E., McDonald, R.C., and Steitz, T.A. (1978) *J. Mol. Biol.* **123**, 207-219.

90. Polakis, P.G. and Wilson, J.E. (1984) *Arch. Biochem. Biophys.* **234**, 341-352.

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

92. Bennett, W.S. Jr. and Steitz, T.A. (1980) *J. Mol. Biol.* **140**, 211-230.

93. Schirch, D.M. and Wilson, J.E. (1987) *Arch. Biochem. Biophys.* **257**, 1-12.

94. White, T.K. and Wilson, J.E. (1990) *Arch. Biochem. Biophys.* **277**, 26-34.

95. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

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

97. Laemmli, U.K. (1970) *Nature* **227**, 680-685.

98. Qadri, S.S., and Easterby, J.S. (1980) *Anal. Biochem.* **105**, 299-303.

99. Hunkapillar, M.W., Lujan, E., Ostrander, F., and Hood, L.E. (1983) *in* Methods in Enzymology (Hirs, C.H.W., and Timasheff, S.N., Eds.), Vol. 91, pp. 227-236, Academic Press, New York.

100. DeWitt, D.L. and Smith, W.L. (1988) *Proc. Natl. Acad. Sci.* USA **85**, 1412-1416.

101. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6-13.

102. Kunkel, T.A. (1985) *Proc. Natl. Acad. Sci.* USA **82**, 488-492.

103. Kunkel, T.A., Roberts, J.D. and Zabour, R.A. *in* Methods in Enzymology (Wu, R. and Grossman, L., Eds.), Vol 154, pp. 367-382, Academic Press, New York.

104. Bird, P., Gething, M.-J., and Sambrook, J. (1987) *J. Cell Biol.* **105**, 2905-2914.

105. Gluzman, Y. (1981) *Cell* **23**, 175-182.

106. Dunn, S.D. (1986) *Anal. Biochem.* **157**, 144-153.

107. Taketa, K., Ichikawa, E., and Hanada, T. (1986) *J. Immunol. Meth.* **95**, 71-77.

108. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156-159.

109. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., and Rutter, W.J. (1979) *Biochem.* **18**, 5294-5299.

110. Sawadogo, M. and Roeder, R.G. (1985) *Proc. Natl. Acad. Sci.* USA **82**, 4394-4398.

111. MacDougald, O.A. and Jump, D.B. (1991) *Biochem. J.* **280**, 761-767.

112. Gorski, K., Carneiro, M. and Schibler, U. (1986) *Cell* **47**, 767-776.

113. Felgner, P.L., and Wilson, J.E. (1976) *Biochem. Biophys. Res. Commun.* **68**, 592-597.

114. Polakis, P.G., and Wilson, J.E. (1982) *Biochem. Biophys. Res. Commun.* **107**, 937-943.

115. Marcus, F. and Ureta, T. (1986) *Biochem. Biophys. Res. Commun.* **139**, 714-719.

116. Imai, J., Akimoto, H., Oda, M., Okazaki, H., Ishibashi, S., and Kurokawa, M. (1988) *Mol. Cell. Biochem.* **81**, 37-41.

117. Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G., Brown, M.S., Goldstein, J.L., and Luskey, K.L. (1984) *Cell* **38**, 275-285.

118. Young, R.M., Shull, G.E., and Lingrel J.B. (1987) *J. Biol. Chem.* **262**, 4905-4910.

119. Shull, G.E., Lane, L.K., and Kingrel, J.B. (1986) *Nature* **321**, 429-431.

120. Kozak, M. (1987) *Nuc. Acids Res.* **15**, 8125-8148.

121. Resnekov. L., Ben-Asher, E., Bengal, E., Choder, M., Hay, N., Kessler, M., Ragimov, N., Seiberg, M., Skolnik-David, H., and Aloni, Y. (1988) *Gene (Amst)* **72**, 91-104.

122. Martin, F.H. and Tinoco, I. (1980) *Nuc. Acids Res.* **8**, 2295-2299.

123. Wilson, T. and Treisman, R. (1988) *Nature* **336**, 396-399.

124. Breathnach, R. and Chambon, P. (1981) *in* Annual Review of Biochemistry (Snell, E.E., Boyer, P.D., Meister, A., and Richardson, C.C., Eds.), Vol. 50, pp. 349-383, Annual Reviews, Inc., Palo Alto, CA.

125. Myers, R.M., Tilly, K., and Maniatis, T. (1986) *Science* **232**, 613-618.

126. O'Shea-Greenfield, A. and Smale, S.T. (1992) *J. Biol. Chem.* **267**, 1391-1402.

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

128. Treisman, R. (1985) *Cell* **42**, 889-902.

129. Knight, E., Jr., Anton, E.D., Fahey, D., Friedland, B.K., and Jonak, G.J. (1985) *Proc. Natl. Acad. Sci.* USA **82**, 151-154.

130. Faisst, S. and Meyer, S. (1992) *Nuc. Acids Res.* **20**, 3-26.

131. Hipskind, R.A. Nordheim, A. (1991) *J. Biol. Chem.* **266**, 19583-19592.

132. Ureta, T., Radojković, J., Lagos, R., Guixé, V., and Nuñez, L. (1979) *Arch. Biol. Med. Exper.* **12**, 587-604.

133. Nasrin, N., Ercolani, L., Denaro, M., Kong, X.F., Kang, I., and Alexander, M. (1990) *Proc. Natl. Acad. Sci.* USA **87**, 5273-5277.

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

135. Garner, M.M. and Revzin, A.R. (1981) *Nuc. Acids Res.* **5**, 3157-3170.

136. Ceglarek, J.A. and Revzin, A.R. (1989) *Electrophoresis* **10**, 360-365.

APPENDIX

APPENDIX A

LIST OF RESTRICTION SITES IN HEXOKINASE TYPE II cDNA
(Composite, includes engineered *Nco*I Site at 1557)

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
AAT 2 (GACGTC)						
	2					
		3330	3330	(91.6)	1	3330
		3445	189	(5.2)	3445	3634
			115	(3.2)	3330	3445
ACC 1 (GTVWAC)						
	1					
		1442	2192	(60.3)	1442	3634
			1442	(39.7)	1	1442
ACY 1 (GPCGQC)						
	2					
		3330	3330	(91.6)	1	3330
		3445	189	(5.2)	3445	3634
			115	(3.2)	3330	3445
AFL 3 (ACPQGT)						
	2					
		1740	1740	(47.9)	1	1740
		2856	1116	(30.7)	1740	2856
			778	(21.4)	2856	3634
AHA 2 (GPCGQC)						
	2					
		3330	3330	(91.6)	1	3330
		3445	189	(5.2)	3445	3634
			115	(3.2)	3330	3445
AHA 3 (TTTAAA)						
	2					
		2997	2997	(82.5)	1	2997
		3601	604	(16.6)	2997	3601
			33	(0.9)	3601	3634
ALU 1 (AGCT)						
	24					
		33	483	(13.3)	1146	1629
		55	432	(11.9)	636	1068
		234	363	(10.0)	1914	2277
		352	303	(8.3)	2388	2691
		373	275	(7.6)	2952	3227
		570	270	(7.4)	1644	1914
		615	201	(5.5)	3234	3435
		636	197	(5.4)	373	570
		1068	179	(4.9)	55	234
		1107	132	(3.6)	3502	3634
		1146	118	(3.2)	234	352
		1629	111	(3.1)	2277	2388
		1644	106	(2.9)	2793	2899

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1914	102	(2.8)	2691	2793
		2277	67	(1.8)	3435	3502
		2388	53	(1.5)	2899	2952
		2691	45	(1.2)	570	615
		2793	39	(1.1)	1107	1146
		2899	39	(1.1)	1068	1107
		2952	33	(0.9)	1	33
		3227	22	(0.6)	33	55
		3234	21	(0.6)	615	636
		3435	21	(0.6)	352	373
		3502	15	(0.4)	1629	1644
			7	(0.2)	3227	3234
APA 1 (GGGCCC)						
	1					
		1599	2035	(56.0)	1599	3634
			1599	(44.0)	1	1599
APA L1 (GTGCAC)						
	1					
		2662	2662	(73.3)	1	2662
			972	(26.7)	2662	3634
AVA 1 (CQCGPG)						
	2					
		99	3482	(95.8)	152	3634
		152	99	(2.7)	1	99
			53	(1.5)	99	152
AVA 2 (GGRCC)						
	9					
		754	982	(27.0)	831	1813
		831	754	(20.7)	1	754
		1813	598	(16.5)	2361	2959
		2361	548	(15.1)	1813	2361
		2959	452	(12.4)	3158	3610
		3015	126	(3.5)	3032	3158
		3032	77	(2.1)	754	831
		3158	56	(1.5)	2959	3015
		3610	24	(0.7)	3610	3634
			17	(0.5)	3015	3032
AVA 3 (ATGCAT)						
	3					
		1859	1859	(51.2)	1	1859
		1895	924	(25.4)	1895	2819
		2819	815	(22.4)	2819	3634
			36	(1.0)	1859	1895
AVR 2 (CCTAGG)						
	1					
		2614	2614	(71.9)	1	2614
			1020	(28.1)	2614	3634

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
BAL 1 (TGGCCA)						
	3					
		597	2332	(64.2)	1302	3634
		1134	597	(16.4)	1	597
		1302	537	(14.8)	597	1134
			168	(4.6)	1134	1302
BAN 1 (GGQPCC)						
	3					
		1498	1498	(41.2)	1	1498
		2231	851	(23.4)	2231	3082
		3082	733	(20.2)	1498	2231
			552	(15.2)	3082	3634
BAN 2 (GPGCQC)						
	4					
		233	1721	(47.4)	1913	3634
		1045	812	(22.3)	233	1045
		1599	554	(15.2)	1045	1599
		1913	314	(8.6)	1599	1913
			233	(6.4)	1	233
BBV 1 (GCTGC)						
	14					
		189	737	(20.3)	2897	3634
		371	523	(14.4)	371	894
		894	347	(9.5)	1729	2076
		1066	289	(8.0)	1066	1355
		1355	267	(7.3)	2076	2343
		1539	213	(5.9)	2684	2897
		1553	193	(5.3)	2491	2684
		1642	189	(5.2)	1	189
		1729	184	(5.1)	1355	1539
		2076	182	(5.0)	189	371
		2343	172	(4.7)	894	1066
		2491	148	(4.1)	2343	2491
		2684	89	(2.4)	1553	1642
		2897	87	(2.4)	1642	1729
			14	(0.4)	1539	1553
BCL 1 (TGATCA)						
	2					
		856	1571	(43.2)	856	2427
		2427	1207	(33.2)	2427	3634
			856	(23.6)	1	856
BGL 1 (GCCNNNNNGGC)						
	1					
		1358	2276	(62.6)	1358	3634
			1358	(37.4)	1	1358
BGL 2 (AGATCT)						
	3					
		525	1344	(37.0)	525	1869
		1869	967	(26.6)	1869	2836
		2836	798	(22.0)	2836	3634
			525	(14.4)	1	525

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
BIN 1 (GGATC)						
	4					
		84	1845	(50.8)	1789	3634
		445	1029	(28.3)	760	1789
		760	361	(9.9)	84	445
		1789	315	(8.7)	445	760
			84	(2.3)	1	84
BSM 1 (GAATGC)						
	1					
		590	3044	(83.8)	590	3634
			590	(16.2)	1	590
BSP 1286 (G2GC3C)						
	9					
		233	972	(26.7)	2662	3634
		470	713	(19.6)	1913	2626
		886	454	(12.5)	1045	1499
		1045	416	(11.4)	470	886
		1499	314	(8.6)	1599	1913
		1599	237	(6.5)	233	470
		1913	233	(6.4)	1	233
		2626	159	(4.4)	886	1045
		2662	100	(2.8)	1499	1599
			36	(1.0)	2626	2662
BSP M1 (ACCTGC)						
	1					
		3607	3607	(99.3)	1	3607
			27	(0.7)	3607	3634
BSP M2 (TCCGGA)						
	3					
		324	2872	(79.0)	762	3634
		409	353	(9.7)	409	762
		762	324	(8.9)	1	324
			85	(2.3)	324	409
BSS H2 (GCGCGC)						
	1					
		2680	2680	(73.7)	1	2680
			954	(26.3)	2680	3634
BST E2 (GGTNACC)						
	2					
		2095	2095	(57.6)	1	2095
		3142	1047	(28.8)	2095	3142
			492	(13.5)	3142	3634
BST N1 (CCRGG)						
	16					
		72	464	(12.8)	2402	2866
		105	463	(12.7)	595	1058
		436	336	(9.2)	1615	1951
		595	331	(9.1)	105	436
		1058	284	(7.8)	2866	3150
		1123	260	(7.2)	3150	3410

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1270	252	(6.9)	1363	1615
		1363	224	(6.2)	3410	3634
		1615	207	(5.7)	2140	2347
		1951	189	(5.2)	1951	2140
		2140	159	(4.4)	436	595
		2347	147	(4.0)	1123	1270
		2402	93	(2.6)	1270	1363
		2866	72	(2.0)	1	72
		3150	65	(1.8)	1058	1123
		3410	55	(1.5)	2347	2402
			33	(0.9)	72	105
BST X1 (CCANNNNNNTGG)						
	2					
		600	2908	(80.0)	726	3634
		726	600	(16.5)	1	600
			126	(3.5)	600	726
CFR 1 (QGGCCP)						
	7					
		597	1344	(37.0)	1365	2709
		807	925	(25.5)	2709	3634
		1134	597	(16.4)	1	597
		1302	327	(9.0)	807	1134
		1341	210	(5.8)	597	807
		1365	168	(4.6)	1134	1302
		2709	39	(1.1)	1302	1341
			24	(0.7)	1341	1365
CLA 1 (ATCGAT)						
	1					
		2471	2471	(68.0)	1	2471
			1163	(32.0)	2471	3634
DDE 1 (CTNAG)						
	14					
		30	821	(22.6)	1691	2512
		289	544	(15.0)	2559	3103
		540	455	(12.5)	711	1166
		711	372	(10.2)	3224	3596
		1166	364	(10.0)	1166	1530
		1530	259	(7.1)	30	289
		1631	251	(6.9)	289	540
		1691	171	(4.7)	540	711
		2512	121	(3.3)	3103	3224
		2547	101	(2.8)	1530	1631
		2559	60	(1.7)	1631	1691
		3103	38	(1.0)	3596	3634
		3224	35	(1.0)	2512	2547
		3596	30	(0.8)	1	30
			12	(0.3)	2547	2559
EAE 1 (QGGCCP)						
	7					
		597	1344	(37.0)	1365	2709
		807	925	(25.5)	2709	3634
		1134	597	(16.4)	1	597
		1302	327	(9.0)	807	1134
		1341	210	(5.8)	597	807

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1365	168	(4.6)	1134	1302
		2709	39	(1.1)	1302	1341
			24	(0.7)	1341	1365
ECO O109 (PGGNCCQ)						
	4					
		25	1787	(49.2)	25	1812
		1812	1219	(33.5)	1812	3031
		3031	578	(15.9)	3031	3609
		3609	25	(0.7)	3609	3634
			25	(0.7)	1	25
FNU 4H1 (GCNGC)						
	21					
		186	712	(19.6)	2922	3634
		189	523	(14.4)	371	894
		371	347	(9.5)	1729	2076
		894	289	(8.0)	1066	1355
		1066	267	(7.3)	2076	2343
		1355	186	(5.1)	2711	2897
		1367	186	(5.1)	1	186
		1539	182	(5.0)	189	371
		1542	179	(4.9)	2505	2684
		1553	172	(4.7)	1367	1539
		1642	172	(4.7)	894	1066
		1729	90	(2.5)	2343	2433
		2076	89	(2.4)	1553	1642
		2343	87	(2.4)	1642	1729
		2433	58	(1.6)	2433	2491
		2491	27	(0.7)	2684	2711
		2505	25	(0.7)	2897	2922
		2684	14	(0.4)	2491	2505
		2711	12	(0.3)	1355	1367
		2897	11	(0.3)	1542	1553
		2922	3	(0.1)	1539	1542
			3	(0.1)	186	189
FNU D2 (CGCG)						
	4					
		21	1314	(36.2)	21	1335
		1335	1312	(36.1)	1369	2681
		1369	953	(26.2)	2681	3634
		2681	34	(0.9)	1335	1369
			21	(0.6)	1	21
FOK 1 (GGATG)						
	23					
		195	345	(9.5)	1534	1879
		412	327	(9.0)	685	1012
		535	300	(8.3)	3194	3494
		685	228	(6.3)	2926	3154
		1012	220	(6.1)	1234	1454
		1092	219	(6.0)	2383	2602
		1120	217	(6.0)	195	412
		1234	196	(5.4)	2602	2798
		1454	195	(5.4)	1	195
		1534	159	(4.4)	1879	2038
		1879	158	(4.3)	2143	2301
		2038	150	(4.1)	535	685

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2089	140	(3.9)	3494	3634
		2143	128	(3.5)	2798	2926
		2301	123	(3.4)	412	535
		2350	114	(3.1)	1120	1234
		2383	80	(2.2)	1454	1534
		2602	80	(2.2)	1012	1092
		2798	54	(1.5)	2089	2143
		2926	51	(1.4)	2038	2089
		3154	49	(1.3)	2301	2350
		3194	40	(1.1)	3154	3194
		3494	33	(0.9)	2350	2383
			28	(0.8)	1092	1120
GDI 2 (QGGCCG)						
	4					
		807	1344	(37.0)	1365	2709
		1341	925	(25.5)	2709	3634
		1365	807	(22.2)	1	807
		2709	534	(14.7)	807	1341
			24	(0.7)	1341	1365
HAE 1 (RGGCCR)						
	9					
		496	698	(19.2)	2218	2916
		597	539	(14.8)	3095	3634
		1134	537	(14.8)	597	1134
		1248	496	(13.6)	1	496
		1302	480	(13.2)	1302	1782
		1782	436	(12.0)	1782	2218
		2218	179	(4.9)	2916	3095
		2916	114	(3.1)	1134	1248
		3095	101	(2.8)	496	597
			54	(1.5)	1248	1302
HAE 2 (PGCGCQ)						
	2					
		148	2369	(65.2)	148	2517
		2517	1117	(30.7)	2517	3634
			148	(4.1)	1	148
HAE 3 (GGCC)						
	17					
		26	538	(14.8)	3096	3634
		103	436	(12.0)	1783	2219
		497	394	(10.8)	103	497
		598	327	(9.0)	808	1135
		808	285	(7.8)	2219	2504
		1135	234	(6.4)	1366	1600
		1249	210	(5.8)	598	808
		1303	207	(5.7)	2710	2917
		1342	206	(5.7)	2504	2710
		1366	183	(5.0)	1600	1783
		1600	179	(4.9)	2917	3096
		1783	114	(3.1)	1135	1249
		2219	101	(2.8)	497	598
		2504	77	(2.1)	26	103
		2710	54	(1.5)	1249	1303

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2917	39	(1.1)	1303	1342
		3096	26	(0.7)	1	26
			24	(0.7)	1342	1366
HGA 1 (GACGC)						
	4					
		283	1932	(53.2)	1702	3634
		922	639	(17.6)	283	922
		1471	549	(15.1)	922	1471
		1702	283	(7.8)	1	283
			231	(6.4)	1471	1702
HGI A1 (GRGCRC)						
	5					
		233	1443	(39.7)	470	1913
		470	972	(26.7)	2662	3634
		1913	713	(19.6)	1913	2626
		2626	237	(6.5)	233	470
		2662	233	(6.4)	1	233
			36	(1.0)	2626	2662
HGI C1 (GGQPCC)						
	3					
		1498	1498	(41.2)	1	1498
		2231	851	(23.4)	2231	3082
		3082	733	(20.2)	1498	2231
			552	(15.2)	3082	3634
HGI J2 (GPGCQC)						
	4					
		233	1721	(47.4)	1913	3634
		1045	812	(22.3)	233	1045
		1599	554	(15.2)	1045	1599
		1913	314	(8.6)	1599	1913
			233	(6.4)	1	233
HHA 1 (GCGC)						
	12					
		149	1105	(30.4)	1353	2458
		778	934	(25.7)	2700	3634
		955	629	(17.3)	149	778
		1267	312	(8.6)	955	1267
		1336	177	(4.9)	778	955
		1353	149	(4.1)	1	149
		2458	90	(2.5)	2518	2608
		2518	72	(2.0)	2608	2680
		2608	69	(1.9)	1267	1336
		2680	60	(1.7)	2458	2518
		2682	18	(0.5)	2682	2700
		2700	17	(0.5)	1336	1353
			2	(0.1)	2680	2682

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
HINC 2 (GTQPAC)						
	3					
		260	1906	(52.4)	403	2309
		403	1325	(36.5)	2309	3634
		2309	260	(7.2)	1	260
			143	(3.9)	260	403
HIND 3 (AAGCTT)						
	3					
		54	2738	(75.3)	54	2792
		2792	683	(18.8)	2951	3634
		2951	159	(4.4)	2792	2951
			54	(1.5)	1	54
HINF 1 (GANTC)						
	14					
		76	1058	(29.1)	181	1239
		181	442	(12.2)	1619	2061
		1239	419	(11.5)	2116	2535
		1279	315	(8.7)	3107	3422
		1314	247	(6.8)	2535	2782
		1383	237	(6.5)	2870	3107
		1619	236	(6.5)	1383	1619
		2061	212	(5.8)	3422	3634
		2116	105	(2.9)	76	181
		2535	88	(2.4)	2782	2870
		2782	76	(2.1)	1	76
		2870	69	(1.9)	1314	1383
		3107	55	(1.5)	2061	2116
		3422	40	(1.1)	1239	1279
			35	(1.0)	1279	1314
HPA 2 (CCGG)						
	13					
		92	834	(22.9)	763	1597
		100	459	(12.6)	2929	3388
		143	442	(12.2)	2235	2677
		325	353	(9.7)	410	763
		410	312	(8.6)	1810	2122
		763	252	(6.9)	2677	2929
		1597	246	(6.8)	3388	3634
		1810	213	(5.9)	1597	1810
		2122	182	(5.0)	143	325
		2235	113	(3.1)	2122	2235
		2677	92	(2.5)	1	92
		2929	85	(2.3)	325	410
		3388	43	(1.2)	100	143
			8	(0.2)	92	100
HPH 1 (GGTGA)						
	13					
		37	672	(18.5)	2158	2830
		484	523	(14.4)	1561	2084
		651	447	(12.3)	37	484
		814	435	(12.0)	1126	1561
		1126	314	(8.6)	2830	3144
		1561	312	(8.6)	814	1126
		2084	254	(7.0)	3380	3634
		2097	236	(6.5)	3144	3380

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2119	167	(4.6)	484	651
		2158	163	(4.5)	651	814
		2830	39	(1.1)	2119	2158
		3144	37	(1.0)	1	37
		3380	22	(0.6)	2097	2119
			13	(0.4)	2084	2097
KPN 1 (GGTACC)						
	1					
		3082	3082	(84.8)	1	3082
			552	(15.2)	3082	3634
MBO 2 (GAAGA)						
	17					
		631	631	(17.4)	1	631
		1078	513	(14.1)	2496	3009
		1129	447	(12.3)	631	1078
		1223	386	(10.6)	1223	1609
		1609	349	(9.6)	3009	3358
		1662	281	(7.7)	1917	2198
		1723	187	(5.1)	1723	1910
		1910	182	(5.0)	3452	3634
		1917	164	(4.5)	2258	2422
		2198	94	(2.6)	1129	1223
		2258	91	(2.5)	3358	3449
		2422	74	(2.0)	2422	2496
		2496	61	(1.7)	1662	1723
		3009	60	(1.7)	2198	2258
		3358	53	(1.5)	1609	1662
		3449	51	(1.4)	1078	1129
		3452	7	(0.2)	1910	1917
			3	(0.1)	3449	3452
MNL 1 (CCTC)						
	54					
		24	276	(7.6)	950	1226
		29	196	(5.4)	204	400
		158	152	(4.2)	1889	2041
		204	145	(4.0)	1652	1797
		400	142	(3.9)	641	783
		453	139	(3.8)	3474	3613
		499	136	(3.7)	2521	2657
		542	131	(3.6)	783	914
		641	129	(3.5)	29	158
		783	120	(3.3)	1532	1652
		914	119	(3.3)	2294	2413
		941	117	(3.2)	3030	3147
		950	111	(3.1)	1292	1403
		1226	99	(2.7)	542	641
		1292	98	(2.7)	2657	2755
		1403	89	(2.4)	2413	2502
		1489	86	(2.4)	1403	1489
		1492	75	(2.1)	2801	2876
		1522	75	(2.1)	2129	2204
		1532	70	(1.9)	3223	3293
		1652	70	(1.9)	2221	2291
		1797	66	(1.8)	1226	1292
		1846	63	(1.7)	3356	3419
		1889	62	(1.7)	3161	3223

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2041	57	(1.6)	2876	2933
		2044	55	(1.5)	3419	3474
		2081	53	(1.5)	400	453
		2087	50	(1.4)	2933	2983
		2129	49	(1.3)	1797	1846
		2204	47	(1.3)	2983	3030
		2221	46	(1.3)	2755	2801
		2291	46	(1.3)	453	499
		2294	46	(1.3)	158	204
		2413	43	(1.2)	1846	1889
		2502	43	(1.2)	499	542
		2521	42	(1.2)	2087	2129
		2657	38	(1.0)	3318	3356
		2755	37	(1.0)	2044	2081
		2801	30	(0.8)	1492	1522
		2876	27	(0.7)	914	941
		2933	24	(0.7)	1	24
		2983	21	(0.6)	3613	3634
		3030	19	(0.5)	2502	2521
		3147	18	(0.5)	3293	3311
		3161	17	(0.5)	2204	2221
		3223	14	(0.4)	3147	3161
		3293	10	(0.3)	1522	1532
		3311	9	(0.2)	941	950
		3315	6	(0.2)	2081	2087
		3318	5	(0.1)	24	29
		3356	4	(0.1)	3311	3315
		3419	3	(0.1)	3315	3318
		3474	3	(0.1)	2291	2294
		3613	3	(0.1)	2041	2044
			3	(0.1)	1489	1492
MST 1 (TGCGCA)						
	2					
		1352	1352	(37.2)	1	1352
		2457	1177	(32.4)	2457	3634
			1105	(30.4)	1352	2457
MST 2 (CCTNAGG)						
	1					
		539	3095	(85.2)	539	3634
			539	(14.8)	1	539
NCI 1 (CCSGG)						
	6					
		99	1497	(41.2)	100	1597
		100	866	(23.8)	1810	2676
		1597	705	(19.4)	2929	3634
		1810	253	(7.0)	2676	2929
		2676	213	(5.9)	1597	1810
		2929	99	(2.7)	1	99
			1	(0.0)	99	100
(NCO 1, CCATGG)	1	1557	2077	(57.2)	1558	3634
			1557	(42.8)	1	1557

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
NDE 1 (CATATG)						
	1					
		209	3425	(94.2)	209	3634
			209	(5.8)	1	209
NLA 3 (CATG)						
	22					
		280	657	(18.1)	2822	3479
		425	607	(16.7)	1099	1706
		550	288	(7.9)	1966	2254
		607	280	(7.7)	1	280
		835	270	(7.4)	2437	2707
		910	228	(6.3)	607	835
		934	192	(5.3)	1706	1898
		958	155	(4.3)	3479	3634
		970	145	(4.0)	280	425
		1042	125	(3.4)	425	550
		1099	123	(3.4)	2314	2437
		1706	111	(3.1)	2707	2818
		1898	75	(2.1)	835	910
		1960	72	(2.0)	970	1042
		1966	62	(1.7)	1898	1960
		2254	60	(1.7)	2254	2314
		2314	57	(1.6)	1042	1099
		2437	57	(1.6)	550	607
		2707	24	(0.7)	934	958
		2818	24	(0.7)	910	934
		2822	12	(0.3)	958	970
		3479	6	(0.2)	1960	1966
			4	(0.1)	2818	2822
NLA 4 (GGNNCC)						
	17					
		68	611	(16.8)	2231	2842
		321	552	(15.2)	3082	3634
		455	419	(11.5)	1812	2231
		563	267	(7.3)	563	830
		830	253	(7.0)	68	321
		980	246	(6.8)	1190	1436
		1190	213	(5.9)	1599	1812
		1436	210	(5.8)	980	1190
		1498	150	(4.1)	830	980
		1599	134	(3.7)	321	455
		1812	116	(3.2)	2842	2958
		2231	108	(3.0)	455	563
		2842	101	(2.8)	1498	1599
		2958	68	(1.9)	1	68
		3015	62	(1.7)	1436	1498
		3032	57	(1.6)	2958	3015
		3082	50	(1.4)	3032	3082
			17	(0.5)	3015	3032
NSI 1 (ATGCAT)						
	3					
		1859	1859	(51.2)	1	1859
		1895	924	(25.4)	1895	2819
		2819	815	(22.4)	2819	3634
			36	(1.0)	1859	1895

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
NSP B2 (CVGCWG)						
	7					
		32	1203	(33.1)	2431	3634
		372	891	(24.5)	1540	2431
		569	498	(13.7)	569	1067
		1067	340	(9.4)	32	372
		1368	301	(8.3)	1067	1368
		1540	197	(5.4)	372	569
		2431	172	(4.7)	1368	1540
			32	(0.9)	1	32
NSP C1 (PCATGQ)						
	4					
		279	1198	(33.0)	2436	3634
		957	748	(20.6)	957	1705
		1705	731	(20.1)	1705	2436
		2436	678	(18.7)	279	957
			279	(7.7)	1	279
PFL M1 (CCANNNNNTGG)						
	2					
		502	2585	(71.1)	502	3087
		3087	547	(15.1)	3087	3634
			502	(13.8)	1	502
PPU M1 (PGGRCCQ)						
	3					
		1812	1812	(49.9)	1	1812
		3031	1219	(33.5)	1812	3031
		3609	578	(15.9)	3031	3609
			25	(0.7)	3609	3634
PVU 2 (CAGCTG)						
	4					
		32	2567	(70.6)	1067	3634
		372	498	(13.7)	569	1067
		569	340	(9.4)	32	372
		1067	197	(5.4)	372	569
			32	(0.9)	1	32
RSA 1 (GTAC)						
	6					
		999	999	(27.5)	1	999
		1096	861	(23.7)	1096	1957
		1957	643	(17.7)	2440	3083
		2440	483	(13.3)	1957	2440
		3083	453	(12.5)	3181	3634
		3181	98	(2.7)	3083	3181
			97	(2.7)	999	1096
RSR 2 (CGGRCCG)						
	1					
		2360	2360	(64.9)	1	2360
			1274	(35.1)	2360	3634

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
SAC 1 (GAGCTC)						
	2					
		233	1721	(47.4)	1913	3634
		1913	1680	(46.2)	233	1913
			233	(6.4)	1	233
SAC 2 (CCGCGG)						
	1					
		1368	2266	(62.4)	1368	3634
			1368	(37.6)	1	1368
SAU 1 (CCTNAGG)						
	1					
		539	3095	(85.2)	539	3634
			539	(14.8)	1	539
SAU 3A (GATC)						
	12					
		84	797	(21.9)	2837	3634
		199	754	(20.7)	1036	1790
		214	558	(15.4)	1870	2428
		446	409	(11.3)	2428	2837
		526	234	(6.4)	526	760
		760	232	(6.4)	214	446
		857	179	(4.9)	857	1036
		1036	115	(3.2)	84	199
		1790	97	(2.7)	760	857
		1870	84	(2.3)	1	84
		2428	80	(2.2)	1790	1870
		2837	80	(2.2)	446	526
			15	(0.4)	199	214
SAU 96 (GGNCC)						
	13					
		26	768	(21.1)	831	1599
		102	652	(17.9)	102	754
		754	598	(16.5)	2361	2959
		831	548	(15.1)	1813	2361
		1599	452	(12.4)	3158	3610
		1600	213	(5.9)	1600	1813
		1813	126	(3.5)	3032	3158
		2361	77	(2.1)	754	831
		2959	76	(2.1)	26	102
		3015	56	(1.5)	2959	3015
		3032	26	(0.7)	1	26
		3158	24	(0.7)	3610	3634
		3610	17	(0.5)	3015	3032
			1	(0.0)	1599	1600
SCR F1 (CCNGG)						
	22					
		72	463	(12.7)	595	1058
		99	331	(9.1)	105	436
		100	274	(7.5)	2402	2676
		105	260	(7.2)	3150	3410
		436	234	(6.4)	1363	1597
		595	224	(6.2)	3410	3634
		1058	221	(6.1)	2929	3150
		1123	207	(5.7)	2140	2347

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1270	195	(5.4)	1615	1810
		1363	190	(5.2)	2676	2866
		1597	189	(5.2)	1951	2140
		1615	159	(4.4)	436	595
		1810	147	(4.0)	1123	1270
		1951	141	(3.9)	1810	1951
		2140	93	(2.6)	1270	1363
		2347	72	(2.0)	1	72
		2402	65	(1.8)	1058	1123
		2676	63	(1.7)	2866	2929
		2866	55	(1.5)	2347	2402
		2929	27	(0.7)	72	99
		3150	18	(0.5)	1597	1615
		3410	5	(0.1)	100	105
			1	(0.0)	99	100
SDU 1 (G2GC3C)						
	9					
		233	972	(26.7)	2662	3634
		470	713	(19.6)	1913	2626
		886	454	(12.5)	1045	1499
		1045	416	(11.4)	470	886
		1499	314	(8.6)	1599	1913
		1599	237	(6.5)	233	470
		1913	233	(6.4)	1	233
		2626	159	(4.4)	886	1045
		2662	100	(2.8)	1499	1599
			36	(1.0)	2626	2662
SFA N1 (GATGC)						
	12					
		963	963	(26.5)	1	963
		1264	462	(12.7)	1264	1726
		1726	352	(9.7)	2925	3277
		1858	301	(8.3)	963	1264
		2037	282	(7.8)	2643	2925
		2069	257	(7.1)	3377	3634
		2263	245	(6.7)	2263	2508
		2508	194	(5.3)	2069	2263
		2643	179	(4.9)	1858	2037
		2925	135	(3.7)	2508	2643
		3277	132	(3.6)	1726	1858
		3377	100	(2.8)	3277	3377
			32	(0.9)	2037	2069
SMA 1 (CCCGGG)						
	1					
		99	3535	(97.3)	99	3634
			99	(2.7)	1	99
SPH 1 (GCATGC)						
	1					
		1705	1929	(53.1)	1705	3634
			1705	(46.9)	1	1705
STU 1 (AGGCCT)						
	1					
		1248	2386	(65.7)	1248	3634
			1248	(34.3)	1	1248

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
STY 1 (CCRRGG)						
	8					
		1137	1137	(31.3)	1	1137
		1156	829	(22.8)	1785	2614
		1785	629	(17.3)	1156	1785
		2614	419	(11.5)	3092	3511
		2811	209	(5.8)	2811	3020
		3020	197	(5.4)	2614	2811
		3092	123	(3.4)	3511	3634
		3511	72	(2.0)	3020	3092
			19	(0.5)	1137	1156
TAQ 1 (TCGA)						
	11					
		121	865	(23.8)	173	1038
		153	803	(22.1)	2472	3275
		173	498	(13.7)	1920	2418
		1038	489	(13.5)	1431	1920
		1197	359	(9.9)	3275	3634
		1242	189	(5.2)	1242	1431
		1431	159	(4.4)	1038	1197
		1920	121	(3.3)	1	121
		2418	54	(1.5)	2418	2472
		2472	45	(1.2)	1197	1242
		3275	32	(0.9)	121	153
			20	(0.6)	153	173
TTH111 2 (CCAPCA)						
	6					
		1591	1591	(43.8)	1	1591
		1818	911	(25.1)	2227	3138
		2014	333	(9.2)	3301	3634
		2227	227	(6.2)	1591	1818
		3138	213	(5.9)	2014	2227
		3301	196	(5.4)	1818	2014
			163	(4.5)	3138	3301
XHO 1 (CTCGAG)						
	1					
		152	3482	(95.8)	152	3634
			152	(4.2)	1	152
XHO 2 (PGATCQ)						
	5					
		445	1264	(34.8)	525	1789
		525	967	(26.6)	1869	2836
		1789	798	(22.0)	2836	3634
		1869	445	(12.2)	1	445
		2836	80	(2.2)	1789	1869
			80	(2.2)	445	525
XMN 1 (GAANNNNTTC)						
	1					
		2111	2111	(58.1)	1	2111
			1523	(41.9)	2111	3634

The following do not appear:

AFL 2	ASU 2	BAM H1	DRA 3
ECO R1	ECO R5	HPA 1	MLU 1
NAE 1	NAR 1	NCO 1	NHE 1
NOT 1	NRU 1	PST 1	PVU 1
RRU 1	SAL 1	SCA 1	SFI 1
SNA 1	SNA B1	SPE 1	SSP 1
TTH111 1	XBA 1	XMA 3	