This is to certify that the

dissertation entitled

THE CLONING OF cDNA'S CODING FOR TYPES I AND III RAT HEXOKINASES AND SEQUENCE COMPARISONS TO OTHER HEXOKINASES

presented by

David A. Schwab

has been accepted towards fulfillment
of the requirements for

Ph.D degree in Biochemistry

John E. Wilson
Major professor

Date April 9, 1994

THE CLONING OF cDNA'S CODING FOR TYPES I AND III RAT HEXOKINASES AND SEQUENCE COMPARISONS TO OTHER HEXOKINASES

By

David A. Schwab

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Biochemistry

1994

ABSTRACT

THE CLONING OF cDNA'S CODING FOR TYPES I AND III RAT HEXOKINASES AND SEQUENCE COMPARISONS TO OTHER HEXOKINASES

By

David A. Schwab

The cDNA's coding for types I and III mammalian hexokinases were cloned from rat brain and rat liver cDNA libraries, respectively. After sequencing, the respective amino acid sequences were deduced. Comparisons between the type I amino acid sequence and the deduced amino acid sequences of the yeast hexokinase isozymes demonstrated a sufficient degree of similarity that the crystallographic structure of the yeast isozymes was used to construct a model for the mammalian hexokinases. The model was shown to be consistent with a variety of experimental data derived directly from the type I enzyme.

The amino acid sequences of hexokinases and glucokinases (deduced from the respective cloned sequences) from various organism were aligned to determine which residues or regions were conserved. The alignment and the yeast crystallographic model were used to determine, at least to a first approximation, where these regions are located. Thus, the residues involved in the binding of glucose (previously determined from crystallographic studies of the yeast hexokinase isozymes) were determined to be conserved among the aligned sequences. Furthermore, regions utilized in the binding of ATP were proposed, based, in one

case, on conservation in the aligned sequences of previously determined sequences utilized in the binding of ATP, and in the other case, on proteins (HSC70, actin, and glycerol kinase) that have been shown to have structurally similar ATP binding sites.

Preliminary experiments were reported for the expression of type I hexokinase in E.coli.

DEDICATION

To my father, Don F. Schwab (who taught me everything I know, not everything he knows, but everything I know) and my mother, Edna J. Schwab (for always being there).

ACKNOWLEDGEMENTS

I would like to acknowledge the past and present members of the Wilson lab., my committee members - Dr. Jerry Dodgson, Dr. Shelagh Ferguson-Miller, Dr. Thomas B. Friedman, and Dr. Jon Kaguni. Additionally, I would like to acknowledge the patience, guidance, and support of Dr. John E. Wilson throughout this project.

On a more personal note, I would like to acknowledge the unrelenting faith and support of my wife, Deborah J. Schwab (who, at times, it seemed contributed more to the success of this degree than I did), and last (and also least) Amanda and Ashley (our latest cloning projects), who have already given me more pleasure than I thought was possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

bp	basepair
CAT	chloramphenicol acetyltransferase
DEAE	diethylaminoethyl
Glc	glucose
IPTG	isopropylthiogalactoside
kb	kilobase
kDa	kilodalton
NAD	nicotinamide adenine dinucleotide
OTG	*O*-toluoylglucosamine
PLP-AMP	pyridoxyl 5'-diphospho-5' adenosine monophosphate
R.T.	room temperature
SDS	sodium dodecyl sulfate
SSC	saline sodium citrate
Tris	tris[hydroxymethyl]aminomethane

CHAPTER I

Literature Review

Introduction

Hexokinase (ATP:D-hexose 6-phosphotransferase, EC 2.7.1.1) catalyzes the phosphorylation of glucose using $Mg^{+2}ATP$ as phosphoryl donor. There are four isozymes in mammalian tissues, designated as types I, II, III, and IV (type IV is commonly referred to as glucokinase). All four mammalian isozymes have been cloned as well as other hexokinases and glucokinases (see below). In this chapter the mammalian isozymes will be reviewed in terms of regulation of activity, tissue distribution, and subcellular associations. This is followed by discussion of the yeast hexokinase isozymes and yeast glucokinase. The chapter concludes with the current hypothesis for the evolution of the mammalian hexokinases.

Cloned Hexokinase and Glucokinase Sequences

The cDNA's coding for all four types of the mammalian isozymes from rat have been cloned and the respective amino acid sequences have been deduced. Additionally, cDNA's coding for hexokinases and glucokinases from different organisms have also been cloned. Accordingly, cDNA's for the type I isozyme have been cloned from rat (1,2 and this thesis), bovine (3), mouse (4), and human (5). The cDNA's for the types II (6) and III (7 and this thesis) isozymes have been cloned from rat and for the type IV isozyme from rat (8,9) and human (10). The genes from yeast coding for hexokinase isozymes A and B (11-13) and

glucokinase (14) as well as a hexokinase from *Schistosoma mansoni* (15) have also been cloned. An alignment of the above mentioned hexokinase and glucokinase deduced amino acid sequences will be shown in this thesis.

Mammalian Hexokinases

The four hexokinase isozymes present in mammalian tissues can be distinguished *via* different electrophoretic mobilities towards the anode during starch gel electrophoresis with mobility increasing with the designated number of the isozyme (16). Alternatively, the four isozymes have also been designated as types A through D as determined by their order of elution from a DEAE-cellulose column, with types A through D corresponding to types I through IV, respectively (17).

Regulation of Activity

Hexokinase catalyzes the conversion of glucose and $Mg^{+2}ATP$ to glucose-6-phosphate and $Mg^{+2}ADP$ (18). Three of the four hexokinases, types I, II and III, have low K_ms for glucose, in the submillimolar range, and are therefore often referred to as the "low K_m" isozymes (Table 1). The other isozyme, type IV or glucokinase, requires a much higher concentration of glucose to reach half saturation. One of the reaction products, glucose-6-phosphate, is a potent inhibitor of the reaction for all three "low K_m" isozymes, but does not inhibit the type IV isozyme at physiologically relevant levels (19). The "low K_m" isozymes are all similar in their specificity for ATP as substrate with ITP

being able to achieve less then 10% the activity relative to ATP while the other nucleoside triphosphates are even poorer substrates (18). All four types are composed of a single polypeptide chain with the "low K_m" isozymes all having a molecular weight of approximately 100 kDa while the type IV isozyme is only 50 kDa. Thus, the "low K_m" isozymes are easily distinguished from the type IV isozyme by size, inhibition by glucose-6-phosphate, and affinity for glucose.

Table 1. Kinetic Parameters of Mammalian Hexokinases

Parameter (mM)	Hexokinase			
	I	II	III	IV
K_m glucose	0.04	0.13	0.02	4.50
K_m ATP	0.42	0.70	1.29	0.49
K_i Glc-6-P vs ATP	0.026	0.021	0.074	15.0

This table was adapted from Ureta (19), and the references therein.

Most of the type I hexokinase in rat brain is bound reversibly to mitochondria (18). This binding is modulated by the inhibitory product glucose-6-phosphate, with increasing levels causing solubilization of the enzyme. Solubilization by glucose-6-phosphate is antagonized by inorganic phosphate while Mg^{+2} enhances binding. Inorganic phosphate alone has no effect on this isozyme.

Felgner *et al.* purified a protein from mitochondria which was shown to be necessary for the reversible binding of the type I isozyme (20). This protein was later determined to be the pore-forming protein porin (21,22)

through which ATP and ADP enter and exit the mitochondria. Consequently, it is suggested that the enzyme has preferential access to one of its substrates, namely ATP, due to the fact that the enzyme is bound to these pores through which ATP exits the mitochondria (23,24). The binding of the enzyme to mitochondria causes the K_m for ATP to decrease, while the K_i for the inhibitor glucose-6-phosphate increases (25-27). Therefore, the bound form of the active enzyme represents a more active form that is not as easily inhibited as the soluble form.

Studies on substrate specificity have led to the conclusion that the type I isozyme can tolerate quite a large variation in structure at the carbon 2 position of the glucose molecule (28). Accordingly, compounds such as mannose (C-2 epimer of glucose), 2-deoxyglucose, glucosamine, and N-acetylglucosamine are substrates for, or competitively inhibit, the reaction catalyzed by hexokinase.

In contrast to the type I isozyme, where inhibition by glucose-6-phosphate is instantaneous, the type II isozyme exhibits a pronounced delay in inhibition (29,30) with this delay becoming even more pronounced for the bound form (e.g. the half time for the response to glucose-6-phosphate inhibition is 12 seconds for the soluble form and 130 seconds for the mitochondrially bound form). Although this inhibition can be relieved by inorganic phosphate in the type I isozyme (18), the type II isozyme shows no such effect (30). Inorganic phosphate is actually an inhibitor of

the type II isozyme.

The K_i for inhibition of type III hexokinase by glucose-6-phosphate is much higher than for the type I and II isozymes. This isozyme is also inhibited by physiologically relevant levels of the substrate glucose (17). It is interesting to note that type III hexokinase from rat liver attains maximum substrate inhibition at approximately the same glucose concentration that glucokinase reaches half saturation (31). It has also been reported that the type III isozyme is affected by inorganic phosphate much the same way as the type II isozyme, *i.e.* inorganic phosphate does not reverse the glucose-6-phosphate induced inhibition of the type III isozyme of pig erythrocyte (32) or bovine liver (33) while inorganic phosphate alone has been shown to inhibit the type III isozyme isolated from bovine liver (33).

As previously stated, the type IV isozyme has an affinity[1] for glucose which is much higher than the other isozymes. Due to the lack of inhibition by glucose-6-phosphate, and a half saturation constant for glucose approximating normal blood glucose concentrations, this isozyme is well suited for its role in the homeostatic control of blood glucose levels (34).

---

[1] Actually, since the type IV isozyme exhibits cooperativity, K_m (which, strictly speaking, applies only to enzymes that adhere to Michaelis-Menten kinetics) is not really correct.

Tissue Distribution

Type I hexokinase has the distinction of being present in all tissues examined to date (18). In most tissues, except for muscle, this isozyme is present at relatively high levels. Due to its prevalence in such a wide diversity of tissues, it has been referred to as the "basic" hexokinase and suggested to be involved in a function basic to all these tissues: glycolysis. In fact, in those tissues with a substantial reliance on blood-borne glucose, the type I isozyme is the predominant form. Brain, being totally dependent on blood-borne glucose, contains virtually exclusively the type I isozyme (hence the designation of type I hexokinase as "brain hexokinase"), as is also the case with erythrocytes. Since in both cases high levels of metabolism are occurring through the glycolytic pathway, it is certainly reasonable to expect that this isozyme's physiological role is primarily glycolytic in nature (18).

Type II is the predominant form in insulin-sensitive tissues such as muscle[2], adipose tissue, and mammary gland (reviewed in 18). The predominance of the type II enzyme has been correlated with the degree to which a tissue is sensitive to insulin. For example, as the insulin sensitivity of rat mammary gland changed during lactation, the activity of type II hexokinase changed in parallel (35). Conversely, in skeletal muscle, which is highly

---

[2] Actually, and surprisingly, human muscle reportedly has type I levels that are much higher than type II levels.

insulin-sensitive, the type II isozyme predominates, but as the proportion of type II decreases, relative to type I, the insulin sensitivity decreases (36). Definite decreases of type II hexokinase have been noted in the insulin-sensitive tissues of diabetic animals (36,37); therefore the availability of insulin seems to be critical for the maintenance of type II hexokinase levels in insulin-sensitive tissues. The predominance of the type II isozyme in insulin-sensitive tissues, with episodic glucose availability, seems to suggest an anabolic role for this isozyme, such as would be required for glycogen synthesis in skeletal muscle (18). In support of this contention is the effect of inorganic phosphate on this isozyme. Glucose-6-phosphate inhibits the enzyme with inorganic phosphate not being able to reverse this inhibition. Muscle contraction is characteristically associated with increased levels of inorganic phosphate (due to increased hydrolysis of high energy phosphate compounds, ATP and creatine phosphate) and increased glycogenolysis leading to elevated levels of glucose-6-phosphate. Under these conditions, glucose-6-phosphate inhibition of hexokinase would not be relieved by inorganic phosphate, and hence, as the scenario goes (18), the type II isozyme would only be active during the anabolic phase of glycogen production. Additionally, since levels of inorganic phosphate increase and, unlike the type I isozyme, do not relieve the inhibition by glucose-6-phosphate, they may actually contribute to inhibition. Therefore, it appears

the type II isozyme would be inhibited during the catabolism of glycogen and active during the anabolic phase where glucose-6-phosphate levels return to much lower levels.

Type III, the least studied of the hexokinases, has not been found to be the predominant form in any tissue (18). This certainly does not preclude the possibility that it may still represent the dominant isozyme in a subpopulation of cells within a tissue (38). The tissues which show the highest amount of activity attributable to type III hexokinase are liver, spleen, and lung (18). This isozyme has also been detected in rat kidney and brain (38). Additionally, Preller and Wilson (38) have demonstrated a staining (using a monoclonal antibody) for type III hexokinase which locates the enzyme at the nuclear periphery in specific cell types in each of these tissues. They point out the prominence of transport functions in many of the cell types in which the nuclear staining for type III hexokinase occurred, although the possible relationship between transport activity and nuclear localization of type III hexokinase is unclear.

Type IV, or glucokinase, is known to be present in the ß-cells of the pancreas and in the liver (34 and ref. therein). Diet and fasting, insulin (39), and glucagon (40) all influence the levels of this isozyme in liver, although these factors do not seem to affect the glucokinase levels in the pancreatic islet ß-cells. Even though insulin does not affect the levels of glucokinase in the ß-cells,

the levels of blood glucose do seem to affect the levels of this isozyme (reviewed in 34).

In the scenario proposed by Magnuson (34), pancreatic ß-cells are stimulated to secrete insulin due to an elevation of activity of endogenous glucokinase brought on by elevated blood glucose levels. By increasing the rate of glycolysis, elevated glucokinase activity is thought to increase the ATP/ADP ratio and hence the ATP levels. This in turn inhibits the opening of ATP-sensitive K^+ channels, causing depolarization of the plasma membrane (41) which then triggers the voltage sensitive Ca^{+2} channels, thus leading to an increase in cytoplasmic Ca^{+2} levels. The release of insulin then occurs due to the increase in Ca^{+2} concentration (42). This insulin, in turn, increases glucokinase levels in the liver where glucose is taken up from the blood thereby decreasing blood glucose levels (glucose is also taken up by other insulin-sensitive tissues, *e.g.* muscle).

The glucokinase gene appears to be under differential regulation due to dual transcription control regions (reviewed in 34). In the liver and ß-cells, different transcription units give rise to tissue-specific mRNAs being altered only in their 5′ regions, with the resultant proteins differing solely in the first 15 amino acids. On the other hand, a cDNA has been isolated from an insulinoma library which has a deletion resulting in a ß-cell glucokinase which is missing 17 amino acids near the glucose

binding region (43). This deletion seems certain to have an impact on this isozyme, though exactly how it manifests itself is unknown.

Subcellular Association

The reversible binding of type I hexokinase to mitochondria is well documented (18). The binding is believed to have both hydrophobic and electrostatic components. The electrostatic component is due, in part, to divalent cations such as Mg^{+2}, presumably *via* the bridging of negative charges on both the enzyme and the mitochondrial membrane (44). Other electrostatic interactions may arise from the attraction of opposite charges contributed by the enzyme and those located on the mitochondrial membrane. Due to the variation in pIs of the isozymes, with these dissimilarities presumably a reflection of differences in surface charges between the isozymes, it is reasonable to expect that the electrostatic component of binding will be important in influencing the relative degrees to which the isozymes bind (18).

On the other hand, the hydrophobic interaction of hexokinase with mitochondria has also been determined to be extremely important. Cleavage of a small hydrophobic peptide (9 residues) from the N-terminus of the type I isozyme with chymotrypsin was shown, by Polakis and Wilson (45), to prevent mitochondrial binding of the enzyme. Xie and Wilson (46) determined later that this essential N-terminal hydrophobic region of the intact enzyme is inserted into the

outer mitochondrial membrane. In crosslinking studies, hexokinase bound to liver mitochondria was found to exist as a monomer or a tetramer with, curiously, no evidence found for intermediate dimers or trimers (47).

In another approach, Gelb et al. (48) generated a chimeric reporter construct which consisted of the first 15 amino acid residues of type I hexokinase coupled to chloramphenicol acetyltransferase (CAT). They demonstrated that these first 15 residues conferred on CAT the ability to bind to mitochondria, which otherwise does not occur. Furthermore, the native hexokinase isozyme was shown to compete with the chimeric CAT construct for binding to mitochondria. Additionally, N,N'-dicyclohexylcarbodiimide, which prevents hexokinase from binding mitochondria by covalently modifying porin, also prevented the chimeric CAT construct from binding. This certainly complements the work of Felgner *et al.* (49) who had previously shown that the protein (porin) they had purified from mitochondria was able to confer on lipid vesicles the ability to bind hexokinase and, most importantly, this binding was sensitive to glucose-6-phosphate.

Type II hexokinase has been shown to bind mitochondria in a competitive manner with the type I isozyme (50). The cDNA for the rat isozyme has been cloned and the amino acid sequence deduced (6). As with the type I isozyme, the type II isozyme has an N-terminal region which is hydrophobic, although less hydrophobic when compared to the type I

isozyme due to the presence of serine and histidine residues (6). Indeed, it is these hydrophilic residues which have been implicated (6), at least in part, in the decrease in avidity with which the type II isozyme binds mitochondria (50) (relative to type I hexokinase).

As previously stated, Type III hexokinase has been demonstrated to have a weak association with the external surface of nuclei by Preller and Wilson (38). In contrast to earlier findings labeling this isozyme as "soluble" and hence cytoplasmic in location, they were able to demonstrate this association *via* confocal microscopy after staining the isozyme through the use of a monoclonal antibody. The cDNA for the rat Type III enzyme has been cloned from liver (7) as part of the work described in this thesis.

Ontogenetic Studies

Ureta carried out a rather extensive study on the levels of each of the "low K_m " isozymes (Types I,II, and III) in rat liver as a function of time (51). The isozymes were isolated and separated (on DEAE-cellulose columns) starting five days before birth and terminating around 17 days after birth at which time the isozyme levels reach their adult levels.

Type I isozyme levels at 5 days before birth are approximately twice the adult level with a maximum level of 4.5 times the adult level attained at birth. Levels of this isozyme then fall to 2.5 times the adult level during the first week with the adult level being attained by the end of

the second week after birth.

The type II levels are very low at 5 days before birth with a maximum level of approximately 3.5 times that of the adult level being attained within the first few days after birth. The levels then decrease reaching adult levels midway into the second week after birth.

Isozyme III remains at low levels before and just after birth reaching a maximum level of approximately 2.5 times the adult level by the end of the first week after birth. Thereafter, the type III isozyme undergoes a rather precipitous decline to adult levels by midway through the second week after birth.

Although the data were not presented, type IV hexokinase was noted to be present at birth, albeit at very low levels. The levels of this isozyme begin to rise at the end of the second week after birth, reaching adult levels at the end of four weeks.

Yeast Hexokinases

Yeast contains two isozymes of hexokinase designated as A and B, or P-I and P-II, respectively (reviewed in ref. 52 and 53). Both isozymes have a molecular weight of approximately 50 kDa and are composed of a single polypeptide chain. Two separate groups have cloned both isozymes (11-13). The isozymes have 378 identical residues out of a total of 485, with the differences being scattered throughout the enzymes.

The yeast isozymes can be separated by chromatography

on DEAE-cellulose (52) using a pH gradient which results in the A isozyme eluting first. Alternatively, isozyme A migrates more anodically during electrophoresis using Tris buffer at pH 9.

During the isolation of the yeast hexokinase isozymes, due to endogenous protease action, alternative enzymatically active forms of these isozymes (i.e. S-I and S-II) (52) are detected in which the first 12 residues have been removed. Native forms of the yeast isozymes form dimers under conditions of high protein concentration and low pH. The first 24 amino acids of both isozymes are identical, and while removal of the first 12 residues has no effect on activity, they appear to be essential for the formation of the dimer. It is interesting to note how the first few residues of the N-terminal sequences of both the yeast isozymes and the mammalian type I isozyme play such an important role in binding.

Measurements of the dissociation constant for the binding of glucose to the dimer (52) indicate that the dimer binds glucose poorly at low glucose concentrations (K_{diss} *ca.* 10^{-3}) but shows positive cooperativity with binding improving at higher glucose concentrations (K_{diss} *ca.* 10^{-4}). In experiments with the proteolytically modified forms, S-I and S-II, which are unable to form dimers, the binding of glucose is much better (K_{diss} = $3x10^{-5}$ and $3x10^{-4}$, respectively). This led to the proposal that in the dimer the active site for glucose is largely buried, whereas, in

the dissociated monomer the active sites are readily accessible (52). The parallel between this behavior and the masking of sites in the intact type I hexokinase of mammals (see below) is intriguing.

The two yeast isozymes differ in their specific activities (52), with the B isozyme's specific activity being four times greater than that of the A isozyme. Additionally, the isozymes differ in their abilities to use fructose and glucose as substrates. The ratio of fructose maximum activity compared to the maximum activity with glucose is 3.0 for the A isozyme, while it is only 1.0 for the B isozyme (52).

The isozymes have been crystallized and the three dimensional structures have been determined. The B isozyme's structure has been determined after its crystallization as a dimer (54-56) without any substrates. Crystallization has also been carried out with the A isozyme and glucose (57,58) and the B isozyme with the glucose analog *O*-toluoylglucosamine (59). Comparisons between these structures have demonstrated that binding of the sugar causes extensive alterations in the structure of the enzyme (Figure 1)(60). That this conformational change has been brought about by the binding of glucose and is not due to a difference in the isozymes has been experimentally verified (61). The most convincing evidence is the fact that the change in the radius of gyration of the B isozyme in solution upon binding glucose is the same as that

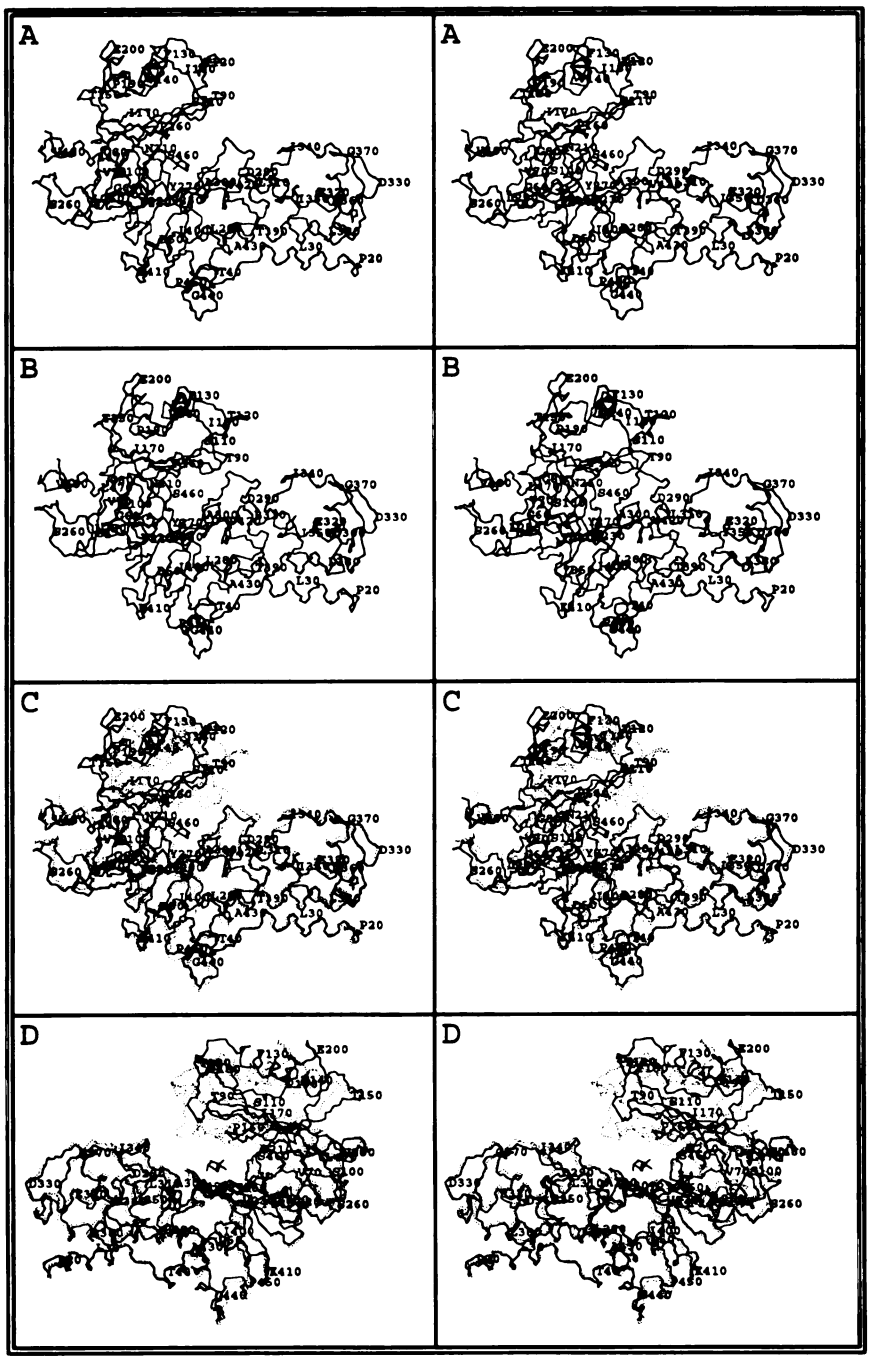

Figure 1. Crystallographic Structure of Yeast Hexokinase ("Open" vs. "Closed" Conformation). **A:** Yeast Hexokinase B in the "open" conformation with glucose (derived from OTG) in the active site. **B:** Yeast Hexokinase A in "closed" conformation. **C and D:** Both conformations superimposed. "Open" conformation is drawn with solid lines, "closed" conformation is drawn with dotted lines.

calculated from the crystallographic coordinates.

Yeast Glucokinase

In the yeast *Saccharomyces cerevisiae* there are three enzymes known to phosphorylate glucose; hexokinase isozymes A and B and yeast glucokinase (62). While the hexokinases are also able to utilize fructose, yeast glucokinase essentially does not. This was illustrated in a study carried out by Lobo and Maitra (62) where they measured the doubling time of yeast grown on glucose or fructose. The strains of yeast were altered such that each strain produced only one of the three enzymes that phosphorylate glucose. When grown on glucose, the strains containing only one of the three enzymes (hexokinase isozyme A, B or glucokinase) grew at a rate comparable to the wild type strain (contains all three enzymes), doubling in under three hours as opposed to under two hours in the wild type strain. On the other hand, the strains containing either hexokinase A or B, when grown on fructose, still doubled in under three hours (wild type still under two hours), while the strain containing only glucokinase took 16 hours to double when grown on fructose. Indeed, the enzyme is so specific for glucose that "trace quantities of glucose in fructose may be analyzed conveniently by using glucokinase" (63) (K_m Glucose = 0.03 mM, K_m Fructose = 31 mM).

Yeast glucokinase has a molecular weight of 51 kDa and can be isolated from a hexokinase deficient mutant principally using ammonium sulfate precipitation and DEAE-

cellulose chromatography (63). The amino acid sequence has been deduced from the cloned gene (14) and will be used in this thesis in comparisons with other hexokinases.

Evolution of Hexokinases

Rossman *et al.* (64) originally noted the similarity between regions of the two lobes of yeast hexokinase which border the substrate binding cleft. McLachlan (65) further pointed out that in comparisons of the two lobes, each of which possess a structural feature comprised of a five stranded ß-sheet and three α-helices (Figure 2), superposition of common regions resulted in 57 common pairs of α-carbons (32 from the ß-sheet and 25 from the α-helices). This led McLachlan (65) to propose that the yeast isozymes may have evolved, in part, by duplication and fusion of a smaller gene encoding the similar structural feature. Harrison (66) however, points out that the central three strands in the ß-sheet in the large lobe are shorter than their counterparts in the small lobe, which he concludes casts doubt on the theory of gene duplication in the evolution of the 50 kDa yeast isozymes.

Many researchers (19,50,67-70) have speculated that the 100 kDa mammalian hexokinases have evolved by duplication and fusion of an ancestral 50 kDa hexokinase not unlike the yeast isozymes. It was proposed that one of the catalytic sites was conserved while the other evolved to take on a regulatory role. This scenario has undergone some modifications as more information has

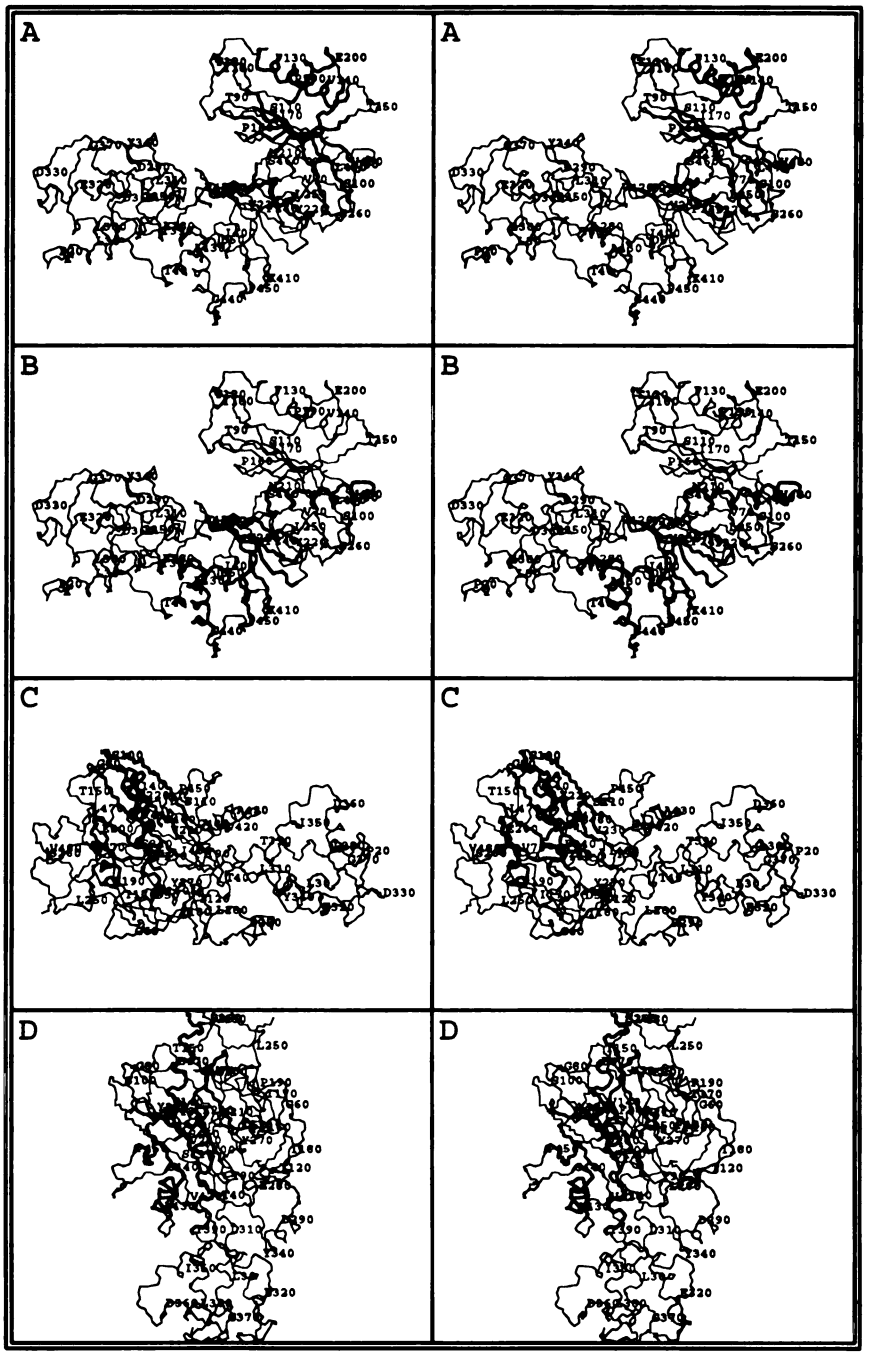

Figure 2. Internal Gene Duplication in Yeast Hexokinase. Stereo images of yeast hexokinase B highlighting the regions in each of the two lobes purported to have arisen through gene duplication. **A:** ß-sheet and α-helices of the small lobe. **B:** ß-sheet and α-helices of the large lobe. **C and D:** ß-sheets and α-helices of the small and large lobes, respectively, oriented to demonstrate similarity.

become available, as will be discussed below.

Polakis and Wilson (71) have shown that digestion of the native type I isozyme with trypsin results in the generation of three principal fragments. The smallest fragment, 10 kDa in size, represents the extreme N-terminal portion of the molecule. The other two fragments generated were of molecular weights 50 and 40 kDa with the 50 kDa fragment being the center fragment located between the N-terminal 10 kDa fragment and the C-terminal 40 kDa fragment (Figure 3). The 40 kDa fragment was subsequently shown by labeling experiments to contain binding sites for both substrates: ATP (72) and glucose (73). Thus the C-terminal portion of the molecule would be expected to contain the catalytic site.

White and Wilson (74), using a different approach, were able to derive a different pattern of digestion using trypsin. Incubating the enzyme in low concentrations of the denaturant guanidinium hydrochloride resulted in more extensive proteolysis with fragments of 52 and 48 kDa appearing as intermediate species (Figure 3). They determined that the enzyme is, in essence, comprised of two major domains of approximately the same size: a 52 kDa N-terminal portion and a 48 kDa C-terminal domain. By adding the inhibitor glucose-6-phosphate they were able to selectively protect the N-terminal portion from denaturation in guanidine hydrochloride and upon addition of trypsin the C-terminal portion was proteolytically removed (74). In the

converse experiment, this time using a glucose analog, N-acetylglucosamine, they were able to selectively protect the C-terminal half of the enzyme (75). Thus, they were able to conclude that the binding site for the allosteric effector glucose-6-phosphate resides in the N-terminal half of the intact enzyme and is separate from the catalytic site. Using a similar approach they were able to isolate the C-terminal portion of the enzyme and demonstrate that it does in fact possess catalytic activity (74). Further work demonstrated, surprisingly, that the isolated C-terminal portion of the enzyme was inhibited by glucose-6-phosphate and that both halves of the enzyme did, in fact, possess binding sites for the inhibitor glucose-6-phosphate as well as the substrates ATP and glucose (also inorganic phosphate) (74). This information lead to a modification of the gene duplication and fusion theory such that the ancestral 50 kDa hexokinase would have had both the glucose binding site as well as the glucose-6-phosphate regulatory site (Figure 4). That this was a reasonable postulation was further supported by the fact that starfish hexokinase has a molecular weight of 50 kDa and is, in fact, inhibited by glucose-6-phosphate.

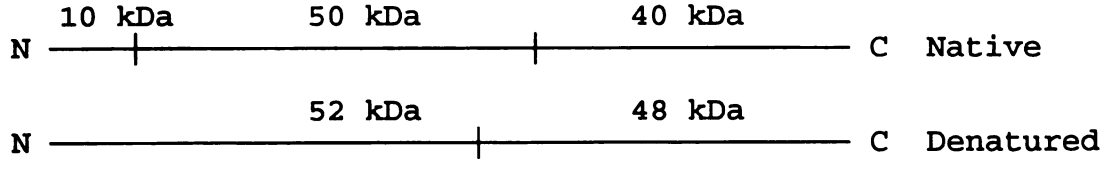

Figure 3. Tryptic Sites in Type I Hexokinase. The predominant sites at which trypsin cleaves the native enzyme and the enzyme under partially denatured conditions are shown with the resultant fragment sizes indicated.

Direct measurements of ligand binding on the intact enzyme have shown only one physiologically relevant binding site for glucose (76,77) and one for glucose-6-phosphate (76,78). Therefore, it was postulated that the glucose site is masked in the N-terminal portion of the intact enzyme with the glucose-6-phosphate site being masked in the C-terminal half.

In order to gauge the reactivity of sulfhydryl groups in the intact enzyme, Hutny and Wilson (79) used the sulfhydryl specific reagent 2-bromoacetamido-4-nitrophenol. Upon binding of glucose-6-phosphate to the high affinity site in the N-terminal portion of the molecule, some of the previously reactive sulfhydryls in the N-terminal portion were protected, as was expected. The fact that sulfhydryls present in the C-terminal portion were also partially protected supports the contention that the structure of the N-terminal half of the intact enzyme (and hence conformational changes occurring therein) impinge on the structure of the C-terminal half. Therefore, evolution of the 100 kDa mammalian hexokinases by gene duplication and fusion from an ancestral hexokinase similar to the 50 kDa starfish hexokinase (which contains sites for catalysis as well as inhibition), with the final 100 kDa enzyme having some of those sites altered or masked, is a reasonable postulation.

Figure 4. Proposed Evolution of Hexokinases. According to this scheme, a 50 kDa ancestral hexokinase evolved in two separate directions: one giving rise to present day yeast isozymes, the other giving rise to a Glc-6-P inhibited enzyme not unlike the starfish 50 kDa enzyme. The present day 100 kDa enzymes evolved from the duplication and fusion of a Glc-6-P inhibited form, except in one half the Glc-6-P regulatory site □ is masked (catalytic half) and in the other half the catalytic site o is masked (regulatory half).

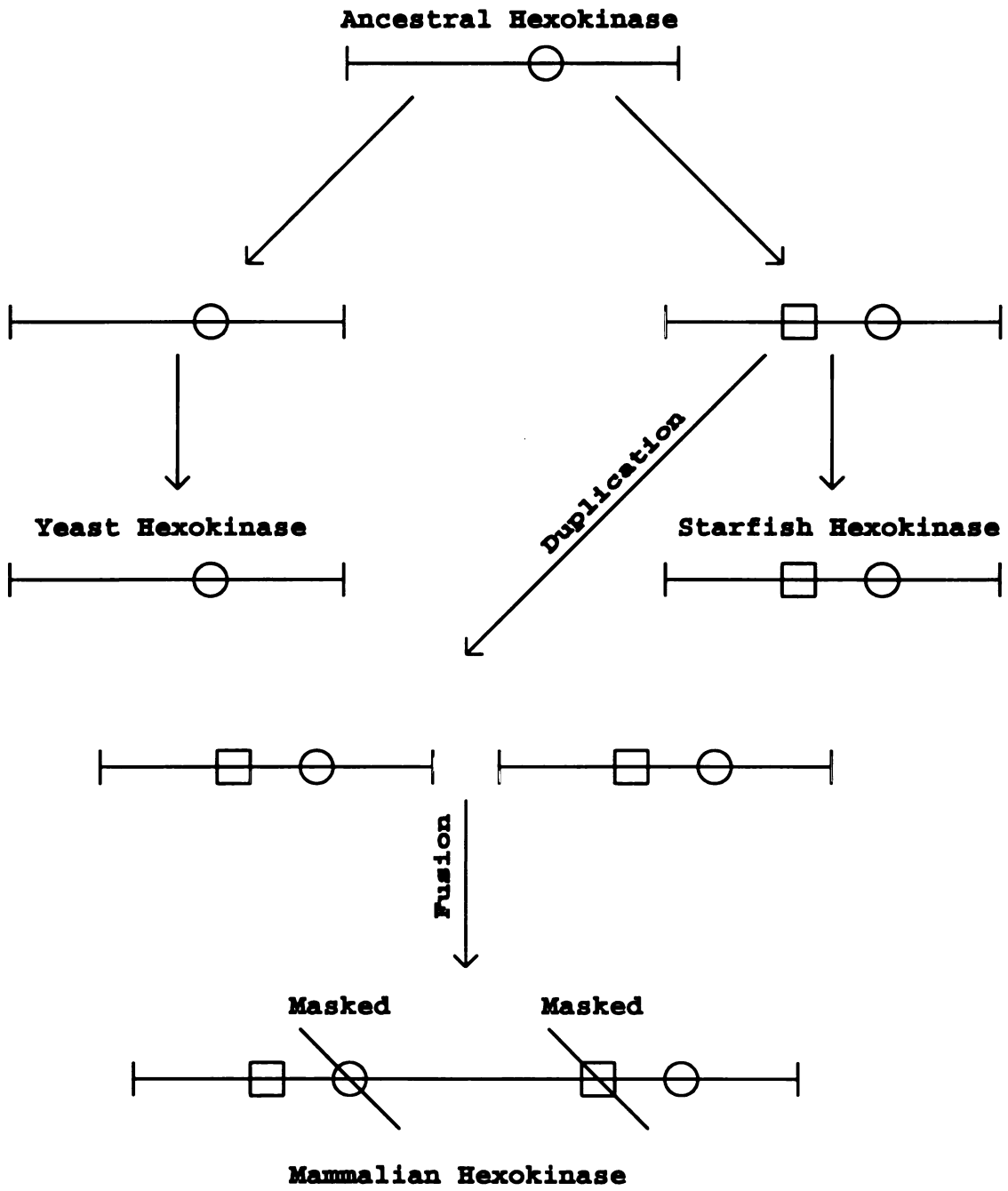
Ancestral Hexokinase
Yeast Hexokinase
Duplication
Starfish Hexokinase
Fusion
Masked
Masked
Mammalian Hexokinase

Figure 4

CHAPTER II
Materials and Methods

Materials

Enzymes used in the restriction or modification of DNA were obtained from a variety of sources, although most were purchased from New England Biolabs (Beverly, MA), Boehringer Mannheim Biochemicals (Indianapolis, IN), or BRL (Gaithersburg, MD). Other DNA modifying enzymes were purchased from Pharmacia (Piscataway, NJ), U.S. Biochemicals (Cleveland, OH), Life Sciences (St. Petersburg, FL), or Stratagene (La Jolla, CA). Radioisotopes were purchased from either NEN Dupont (Boston, MA) or Amersham (Arlington Heights, IL). Other reagents and materials were obtained from a variety of standard commercial suppliers.

A rat brain cDNA library constructed in λgt11, using mRNA from adult rat brains, was generously provided by Dr. Ronald L. Davis. Rat brain hexokinase (Type I) was prepared according to Chou and Wilson (80).

Methods

Preparation of Anti-hexokinase Antibodies

Preparation of anti-hexokinase antibodies was carried out as previously described (81).

Affinity Purification of Antibodies to Rat Brain Hexokinase

Preparation of Affigel-10 Hexokinase Column

1) 2 mg of HK were incubated overnight with 1 ml of Affigel-10 (Biorad) at 4°C in 50 mM $(Na)_3PO_4$ (pH 7.0). The Affigel-10 was then washed in 0.1 M ethanolamine at room temperature (R.T.) after loading it into a 3cc syringe which had been plugged with silanized glass wool. The column was washed sequentially with five column volumes of the following buffers.

TBS-NP40:	120 mM NaCl
	50 mM Tris pH 7.5
	0.5% Nonidet P-40
	1 M LiCl
Glycine buffer:	50 mM Glycine pH 2.5
	150 mM NaCl
PBS:	0.1 M NaCl
	0.01 M Na phosphate pH 7.5

The column was stored at 4°C in PBS + 0.1% NaN_3.

Purification of Anti-Rat Brain Hexokinase Antibodies

1.) 2 ml of antiserum were recycled over the column at R.T. a minimum of five times with a flow rate not exceeding 5 ml/hr.

2.) The column was washed sequentially with five column volumes each of PBS, TBS-NP40, and PBS.

3.) The affinity purified Ab's were eluted using the glycine buffer. The eluate (usually 10 ml) was collected and neutralized with 1 M Tris-HCl pH 9.0.

4.) The column was equilibrated with PBS and NaN_3 added to 0.1% for storage at 4°C.

Immunological Screening of λgt11 cDNA Library

1.) Grow an overnight culture of the bacterial strain Y1090 in L broth + ampicillin @ 50 ug/ml.

2.) 100 ul or less of the appropriate dilution of the λgt11 cDNA library is mixed with 100 ul of the bacterial strain Y1090 and incubated at 37°C for 20 min to allow for infection of the bacteria by the phage.

3.) The mixture is then plated on a warm (50°C) 100 mm diameter agar (L broth) plate using 3 ml of top agar (L broth) previously melted and kept at 45-50°C.

4.) Wet nitrocellulose filters with 10 mM IPTG (isopropylthiogalactoside) and air dry (20 min.).

5.) After the top agar has hardened (5 min. in cold room at 4°C or 15 min. at R.T.) the IPTG treated nitrocellulose filter is placed on the top agar while avoiding trapping any air bubbles between the filter and the

top agar.

6.) Incubate plates at 42°C for a minimum of 15 min. (time necessary for the entire plate to reach 42°C) followed by a minimum 3 hour incubation at 37°C. (It is common at this point to leave plates overnight @ 37°C.)

7.) The orientation of the nitrocellulose filter on the plate is then clearly marked by injecting an extremely small amount of black india ink into the agar plate after piercing the nitrocellulose filter and agar plate (three injections at the periphery of the filter in an unambiguous manner).

8.) 100 *ng* of purified rat brain hexokinase is spotted directly onto the top side of the filter (the side not in contact with the agar) as a positive control for the immunological screening. The filter is gently peeled off the agar plate using forceps while only touching the extreme edges (point indentations may "light up" as positives). NOTE: At this point the filter may be numbered using a pencil. The filter is placed face up (side in contact with agar up) in a petri dish containing 20 ml TBS (10 mM Tris pH 7.5, 0.15 M NaCl). The agar plate is stored at 4°C.

10.) Any agar sticking to the filter can be removed by swirling the petri dish. The filter is then incubated for a minimum of 5 min. with TBS containing 3% calf serum (or gelatin) to block any protein binding sites.

11.) Incubate filter with 20 ml of a 1/1000 dilution (using TBS) of the affinity purified rabbit anti-HK Ab's containing 0.5% Nonidet P-40 at R.T. with shaking for 1 hr

(1/1000 dilution with respect to the antiserum).

NOTE: Some Ab preparations give higher backgrounds than others. The background may be reduced by incubating the Ab soln. sequentially with two or three filters using non-recombinant λgt11 as phage (a preadsorption step) - if this is done, a 1/10 dilution of the Ab's can be used in the preadsorption step followed by dilution to 1/1000 for the screening. The resulting solution can be used numerous times.

12.) The Ab solution is poured off and kept at 4°C after addition of NaN_3 to 0.1%. The filter is washed 3 times with TBS for 5 min. (while shaking).

13.) Incubate filter with 20 ml TBS containing 3% goat serum for 5 min.

14.) Incubate filter with 20 ml of a 1/1000 dilution of affinity purified horseradish peroxidase conjugated goat anti-rabbit Ab's in TBS containing 0.5% Nonidet P-40 for 1 hr.

15.) The goat Ab's are poured off and saved at 4°C for future use (reuse up to three times without any difficulty in detection). Wash filter three times with TBS for 5 min. each (while shaking).

16.) The filter is then incubated with 10 ml of developing soln. for 5 to 15 min. in the dark. Developing soln.= 60 mg 4-chloro-naphthol in 20 ml cold CH_3OH added to 100 ml TBS containing 60 ul of 30% H_2O_2.

17.) After developing, the filters are rinsed with dH_2O

and stored in the dark to prevent yellowing.

18.) The filter is aligned with the agar plate using the black india ink spots and a plug of agar is cored from the region containing the suspected positive (the large end of a disposable pipette is ideal for this procedure).

19.) The plug is placed in a 1.5 ml Eppendorf tube containing 1 ml of SM (plus 50 ul $CHCl_3$ to prevent bacterial growth) and left overnight at 4°C to allow the phage to elute from the plug.

SM (per liter)	5.8g NaCl
	2gms $MgSO_4 \cdot 7H_2O$
	50 ml 1M Tris HCl pH 7.5
	5 ml 2% gelatin

NOTE: To expedite matters the plug can be broken up with a toothpick and the Eppendorf tube incubated at R.T. on a rocker for 1 hr. to elute the phage.

20.) The phage solns. from each positive are plated as above at a lower density and rescreened. A well resolved positive plaque is picked and used for subsequent phage DNA isolation.

cDNA Synthesis and Construction of λgt10 Libraries

For the rat brain cDNA library, total RNA was isolated from adult rat brains. Total RNA to be used in the construction of the rat liver cDNA library was isolated from the livers of 6-day old rats. (This is the time point at which the type III isozyme's activity is at a maximum, and therefore the mRNA levels for this isozyme were presumed to

be at a maximum.) Both libraries were constructed using the procedure of DeWitt and Smith (82) starting with 5 ug of mRNA which had been isolated from total RNA as described in (83).

Screening of λgt10 cDNA libraries

Plaque hybridization of the rat brain λgt10 cDNA library using the immunologically isolated clone HKI 12.4-4 (1) was carried out by procedures described in Maniatis (84).

The rat liver λgt10 cDNA library was screened *via* plaque hybridization (84) as in Maniatis, using the full length rat brain cDNA clone HKI 1.4-7 (2) as probe. The only procedural difference was that, after hybridization, the filters were washed only in 2 x SSC at 50°C.

Labeling of the cDNA clones used in the plaque hybridizations was carried out *via* random priming (85,86).

Sequencing of cDNA Clones

The dideoxy method (87) was used to sequence the cDNA clones after generating non-random deletions *via* the method of Henikoff (88,89). Non-random deletions were generated by digesting successively larger regions of DNA from one end of the pertinent cDNA clones. This was carried out (separately) on both ends of the cDNA clones such that the sequence of both strands could be determined.

Northern Blot

Preparation and hybridization of the northern blot was

carried out as in Maniatis using 10 ugs of rat brain mRNA and type I hexokinase cDNA clone HKI 12.4-4 as probe with the only difference being that after hybridization the blot was washed only in 2 x SSC at 48°C.

Construction of Plasmids for Expression of Rat Brain Hexokinase in E. coli.

pHB4 and pM1-7

Full length clone HKI 1.4-7 in pUC18 was digested with Bam H1 and religated. This resulted in the removal of the 3' untranslated region (Bam H1 cuts a few bases down stream of the stop codon and once in the multiple cloning site). The modified clone was designated pHKI 1.4-7-B and is the starting clone in Figure 5. The next step was removal of the 5' untranslated sequence which was carried out because this clone was initially going to be used in a different expression vector. Step 1: Digestion of pHKI 1.4-7-B with EcoR1 and Sma I to isolate the 256 bp fragment corresponding to the 5' untranslated region and the first 165 bps of the coding region. Step 2: Digestion of the isolated 256 bp EcoR1 - Sma I fragment with Nla III while removing aliquots throughout the digestion in order to isolate the fragment which is cleaved at only one of the Nla III sites (partial digestion), the site located at the starting Met codon. Step 3: The starting clone, pHKI 1.4-7-B, is digested with Sph I and Sma I in order to ligate it (step 4) to the partial digestion fragment of step 2. The Sph I and Nla III sites are compatible, although the Sph I site will be lost upon

Figure 5. Construction of Plasmids pHB4 and pM1-7 Used for Expression. See text for details.

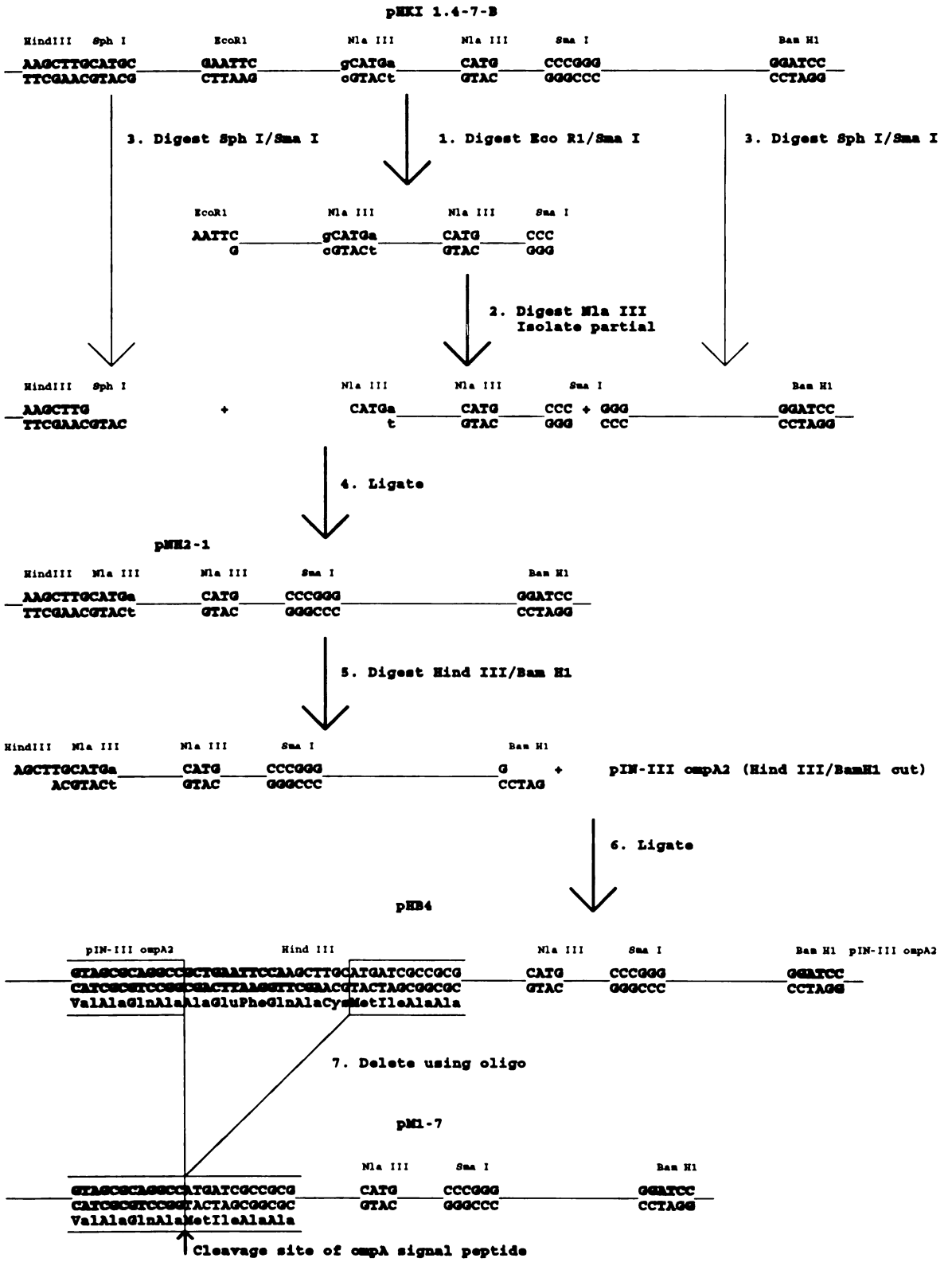
pHKI 1.4-7-B
HindIII Sph I EcoR1 Nla III Nla III Sma I Bam H1
AAGCTTGCATGC GAATTC gCATGa CATG CCCGGG GGATCC
TTCGAACGTACG CTTAAG cGTACt GTAC GGGCCC CCTAGG
3. Digest Sph I/Sma I
1. Digest Eco R1/Sma I
3. Digest Sph I/Sma I
EcoR1 Nla III Nla III Sma I
AATTC gCATGa CATG CCC
G cGTACt GTAC GGG
2. Digest Nla III
Isolate partial
HindIII Sph I Nla III Nla III Sma I Bam H1
AAGCTTG + CATGa CATG CCC + GGG GGATCC
TTCGAACGTAC t GTAC GGG CCC CCTAGG
4. Ligate
pHK2-1
HindIII Nla III Nla III Sma I Bam H1
AAGCTTGCATGa CATG CCCGGG GGATCC
TTCGAACGTACt GTAC GGGCCC CCTAGG
5. Digest Hind III/Bam H1
HindIII Nla III Nla III Sma I Bam H1
AGCTTGCATGa CATG CCCGGG G + pIN-III ompA2 (Hind III/BamH1 cut)
ACGTACt GTAC GGGCCC CCTAG
6. Ligate
pHB4
pIN-III ompA2 Hind III Nla III Sma I Bam H1 pIN-III ompA2
CATG CCCGGG GGATCC
GTAC GGGCCC CCTAGG
ValAlaGlnAlaAlaGluPheGlnAlaCysMetIleAlaAla
7. Delete using oligo
pM1-7
Nla III Sma I Bam H1
CATG CCCGGG GGATCC
GTAC GGGCCC CCTAGG
ValAlaGlnAlaMetIleAlaAla
Cleavage site of ompA signal peptide

Figure 5.

ligation (denoted by the small "a" next to the Nla III site arrived at in step 2). This clone was designated pNH2-1 and was digested with Hind III and Bam H1 (step 5) and ligated into the expression vector pIN-III ompA2 which had been cut similarly (step 6). The resulting clone, pHB4, was in frame with the ompA signal peptide and was used in the initial expression experiments aimed at determining if the expressed rat brain hexokinase was catalytically active. If the signal peptide was cleaved correctly, the expressed protein would still have 6 amino acid residues tacked onto the N-terminus which corresponds to the cloning site. Clone pM1-7 was constructed using the deletion mutagenesis procedure (step 7) outlined by Takahara *et al.* (90) and the 24 base oligomer designated J.E.W.4 (GTAGCGCAGGCCATGATCGCCGCG). The rat brain enzyme expressed from this clone, if correctly processed, should begin with the starting Met.

pXN1 and pNB6

Originally, the type III hexokinase cDNA clone was also to be used in experiments aimed at bacterial expression and clones were constructed using the same procedure as above (90). Unfortunately the oligonucleotide used to delete the 5' untranslated sequence contained an extra nucleotide which was inserted down stream from the start codon and hence prevented expression of this isozyme due to a frame shift error. Nevertheless, one of the constructs, designated pIII-1, still proved useful in that the start codon was conveniently located in an Nco I site. This site was

utilized in the construction of the plasmids used to express the two "halves" of rat brain hexokinase (pXN1 and pNB6, N-terminal and C-terminal halves, respectively) described below.

Clone pHKI 1.4-7 contains a unique Nco I site approximately midway through the coding region. This clone was digested with Nco I and Bam H1 and the 1403 bp fragment was cloned into pIII-1 which had similarly been cut. The resulting clone, pNB6, was constructed to express the C-terminal half of rat brain hexokinase.

Clone pM1-7, constructed above, was digested with Xba I and Nco I. The 1552 bp fragment corresponds to the coding region of the N-terminal half of rat brain hexokinase with an additional 100 bps on the 5′ end (up to the Xba I site) coming from the pIN-III ompA2 vector. This fragment was cloned into pIII-1 which had also been cut with Xba I and Nco I. The resulting clone, pXN1, should express the N-terminal half of rat brain hexokinase.

Expression of Rat Brain Hexokinase in E.coli

1.) Grow a 1.5 ml culture (strain JA221 harboring the appropriate plasmid) to be used for expression, overnight at 37°C in L broth with ampicillin @ 75 ug/ml.

2.) Add 100 ul to 10ml of media (TB broth + ampicillin @ 75 ug/ml) in a 50 ml screw cap tube (with appropriate amount of IPTG) and let grow on shaker for 16 hrs.

3.) Transfer 1 ml of culture to 1.5 ml eppendorf and spin down for 2 min. discarding supernatant.

4.) Resuspend pellet in 500 ul of 20% sucrose, 10 mM Glc, 10mM thioglycerol, 10 mM Tris, pH 7.5, and store on ice for 10 min.

5.) Spin down for 2 min. in cold room. Resuspend pellet in 200 ul of ice cold 10 mM Glc, 10 mM thioglycerol. Leave on ice for 15 min.

6.) Spin down 5 min. in cold room. Supernatant contains expressed enzyme. Hexokinase activity was measured spectrophotometrically as in (91).

SDS-gel Electrophoresis and Immunoblotting

Procedures for SDS-gel electrophoresis and immunoblotting were the same as those described in (71).

Alignment of Amino Acid Sequences

The alignments of amino acid sequences of hexokinase isozymes were determined by first matching regions with a high degree of similarity before aligning the remaining sequence while keeping gaps to a minimum. Amino acid residue changes that occurred within one of the following six categories were considered to be conservative changes:

a) Val,Met,Ile,Leu.

b) Gln,Asn.

c) His,Lys,Arg.

d) Ala,Thr,Ser.

e) Glu,Asp.

f) Phe,Tyr,Trp.

Generation of Stereo Images

Stereo images were generated using the Brookhaven

Protein Data Bank coordinates for the "open" conformation of yeast hexokinase B complexed with OTG (filename PDB2YHX.ENT), the "closed" conformation of yeast hexokinase A complexed with Glc (filename PDB1HKG.ENT), actin (filename PDB1ATC.ENT), and glycerol kinase (filename PDB1GLB.ENT). The program was written in Pascal on an IBM PC and designed for the generation of stereo images using the HP-GL/2 language of Hewlett-Packard LaserJet Printers (III or IV).

Secondary structural features of yeast hexokinase were determined at the computational chemistry facility of Upjohn (Kalamazoo, MI) by the algorithm intrinsic to the software package MOSAIC using x-ray crystallographic coordinates from the Brookhaven Protein Data Bank.

CHAPTER III

Cloning of cDNA'S Coding for Type I Rat Hexokinase; Comparison to Yeast Hexokinases; Proposed Model for Type I Hexokinase

This chapter begins with description of the type I hexokinase cDNA clones, isolated from rat brain cDNA libraries, followed by verification of the authenticity of the clones as coding for this enzyme. Comparisons between the deduced amino acid sequences of the N- and C-terminal halves of type I hexokinase and the yeast isozymes establish that type I hexokinase appears to have evolved via gene duplication and fusion of a 50 kDa ancestral hexokinase. Separate comparisons of the N- or C-terminal halves of rat brain hexokinase with yeast hexokinase isozymes A and B reveal that the yeast crystallographic structures provide a reasonable model (at least to a first approximation) for both halves of type I hexokinase. The chapter concludes with a proposed model of the mammalian enzyme constructed using the yeast crystal structures and relevant experimental data pertaining to type I hexokinase.

Initially, clones coding for the C-terminal half of rat brain hexokinase were isolated and sequenced. Subsequently, clones were isolated which contained the entire coding region. Therefore, discussion of the determination of the amino acid sequence for the C-terminal half of rat brain hexokinase occurs before the N-terminal half.

Cloning of the C-terminal Half of Rat Brain Hexokinase

A rat brain λgt11 cDNA library was screened immunologically for clones coding for rat brain hexokinase. The largest clone isolated was designated HKI 12.4-4 (1) and contained a 2.1 kb insert. Both strands were sequenced after

generating non-random deletions using the strategy depicted in Figure 6.

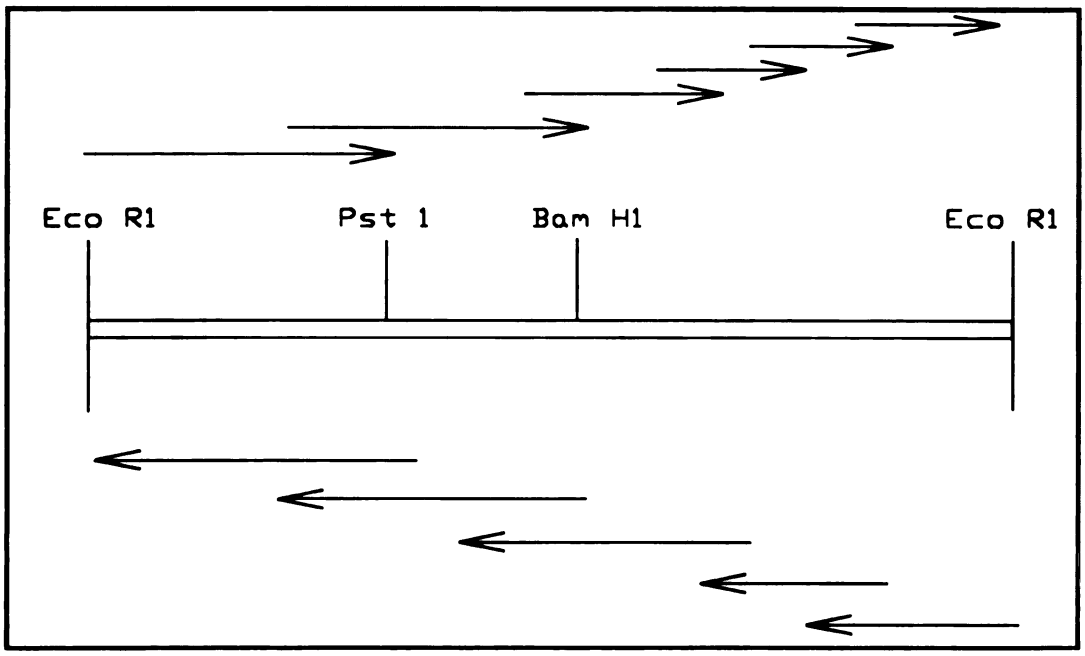

Figure 6. Sequencing Strategy for cDNA Clone HKI 12.4-4. The direction and extent of sequencing of subcloned fragments after generating nonrandom deletions are indicated by the arrows (relevant restriction sites are also indicated - the Eco R1 sites were derived from the vector).

Verification of cDNA Clone HKI 12.4-4 as Coding for the C-terminal Half of Type I Hexokinase

Figure 7 contains the nucleotide sequence of cDNA clone HKI 12.4-4 under which the deduced amino acid sequence is given. There exists a single open reading frame coding for approximately half of the type I isozyme which contains several regions where the deduced amino sequence matches that determined directly from the C-terminal half of the rat brain hexokinase enzyme (underlined in Figure 7). These

```
CGG CAG ATT GAG GAA ACC CTG GCC CAC TTC CGC CTC AGC AAG CAG ACG CTG ATG GAG 1546
Arg Gln Ile Glu Glu Thr Leu Ala His Phe Arg Leu Ser Lys Gln Thr Leu Met Glu  486

GTG AAG AAG AGG CTA CGG ACA GAG ATG GAA ATG GGG CTG AGG AAG GAG ACC AAC AGC 1603
Val Lys Lys Arg Leu Arg Thr Glu Met Glu Met Gly Leu Arg Lys Glu Thr Asn Ser  505

AAA GCT ACT GTC AAA ATG CTG CCT TCT TTT GTC CGG AGC ATC CCG GAT GGG ACT GAA 1660
Lys Ala Thr Val Lys Met Leu Pro Ser Phe Val Arg Ser Ile Pro Asp Gly Thr Glu  524

CAC GGT GAC TTC CTG GCC TTG GAT CTT GGA GGA ACG AAT TTC CGG GTT CTG CTG GTA 1717
His Gly Asp Phe Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg Val Leu Leu Val  543

AAG ATC CGC AGT GGG AAA AAG AGA ACA GTG GAA ATG CAC AAC AAG ATC TAC TCC ATT 1774
Lys Ile Arg Ser Gly Lys Lys Arg Thr Val Glu Met His Asn Lys Ile Tyr Ser Ile  562

CCC CTG GAA ATC ATG CAG GGC ACC GGG GAT GAG CTG TTT GAC CAC ATC GTC TCC TGC 1831
Pro Leu Glu Ile Met Gln Gly Thr Gly Asp Glu Leu Phe Asp His Ile Val Ser Cys  581

ATC TCT GAC TTC CTG GAC TAC ATG GGG ATC AAA GGC CCC CGG ATG CCT CTG GGC TTC 1888
Ile Ser Asp Phe Leu Asp Tyr Met Gly Ile Lys Gly Pro Arg Met Pro Leu Gly Phe  600

ACC TTC TCA TTT CCC TGC CAT CAG ACG AAC CTG GAC TGT GGA ATC TTG ATC TCA TGG 1945
Thr Phe Ser Phe Pro Cys His Gln Thr Asn Leu Asp Cys Gly Ile Leu Ile Ser Trp  619

ACA AAG GGT TTC AAA GCC ACT GAC TGT GAG GGC CAT GAT GTA GCC TCC TTA CTG AGG 2002
Thr Lys Gly Phe Lys Ala Thr Asp Cys Glu Gly His Asp Val Ala Ser Leu Leu Arg  638

GAT GCG GTG AAG AGG AGA GAG GAA TTT GAC TTG GAT GTG GTG GCT GTG GTC AAC GAC 2059
Asp Ala Val Lys Arg Arg Glu Glu Phe Asp Leu Asp Val Val Ala Val Val Asn Asp  657

ACC GTG GGC ACC ATG ATG ACC TGT GCG TAT GAA GAA CCC ACT TGC GAA ATT GGA CTC 2116
Thr Val Gly Thr Met Met Thr Cys Ala Tyr Glu Glu Pro Thr Cys Glu Ile Gly Leu  676

ATC GTG GGG ACG GGC ACC AAT GCC TGC TAC ATG GAG GAG ATG AAG AAT GTG GAG ATG 2173
Ile Val Gly Thr Gly Thr Asn Ala Cys Tyr Met Glu Glu Met Lys Asn Val Glu Met  695

GTG GAG GGG AAC CAG GGC CAG ATG TGC ATC AAC ATG GAG TGG GGC GCC TTC GGT GAC 2230
Val Glu Gly Asn Gln Gly Gln Met Cys Ile Asn Met Glu Trp Gly Ala Phe Gly Asp  714

AAT GGG TGT CTG GAT GAC ATC AGA ACA GAC TTT GAC AAA GTG GTG GAC GAA TAT TCT 2287
Asn Gly Cys Leu Asp Asp Ile Arg Thr Asp Phe Asp Lys Val Val Asp Glu Tyr Ser  733

CTA AAC TCT GGG AAA CAA AGG TTT GAG AAA ATG ATC AGT GGG ATG TAC CTG GGT GAG 2344
Leu Asn Ser Gly Lys Gln Arg Phe Glu Lys Met Ile Ser Gly Met Tyr Leu Gly Glu  752

ATC GTC CGT AAC ATC CTG ATT GAC TTC ACC AAG AAA GGC TTC CTC TTC CGG GGA CAG 2401
Ile Val Arg Asn Ile Leu Ile Asp Phe Thr Lys Lys Gly Phe Leu Phe Arg Gly Gln  771

ATC TCC GAA CCA CTC AAG ACC CGA GGC ATC TTT GAG ACC AAG TTT CTC TCT CAG ATT 2458
Ile Ser Glu Pro Leu Lys Thr Arg Gly Ile Phe Glu Thr Lys Phe Leu Ser Gln Ile  790

GAG AGT GAC CGG TTA GCG CTG CTC CAG GTG CGG GCC ATC CTT CAG CAG CTG GGT TTG 2515
Glu Ser Asp Arg Leu Ala Leu Leu Gln Val Arg Ala Ile Leu Gln Gln Leu Gly Leu  809

AAC AGC ACG TGT GAC GAC AGT ATC CTG GTC AAG ACC GTG TGT GGG GTG GTG TCC AAG 2572
Asn Ser Thr Cys Asp Asp Ser Ile Leu Val Lys Thr Val Cys Gly Val Val Ser Lys  828

AGG GCG GCT CAG CTG TGT GGT GCC GGC ATG GCC GCC GTG GTG GAA AAG ATC AGA GAG 2629
Arg Ala Ala Gln Leu Cys Gly Ala Gly Met Ala Ala Val Val Glu Lys Ile Arg Glu  847

AAC AGA GGC CTA GAC CAT CTG AAT GTA ACT GTG GGA GTG GAT GGG ACG CTC TAC AAA 2686
Asn Arg Gly Leu Asp His Leu Asn Val Thr Val Gly Val Asp Gly Thr Leu Tyr Lys  866

CTT CAT CCA CAC TTC TCC AGA ATC ATG CAC CAA ACT GTG AAG GAA CTG TCA CCA AAG 2743
Leu His Pro His Phe Ser Arg Ile Met His Gln Thr Val Lys Glu Leu Ser Pro Lys  885

TGT ACC GTG TCC TTC CTC CTG TCT GAA GAC GGC AGC GGC AAG GGG GCC GCC CTT ATC 2800
Cys Thr Val Ser Phe Leu Leu Ser Glu Asp Gly Ser Gly Lys Gly Ala Ala Leu Ile  904

ACA GCT GTG GGC GTG CGG CTC AGA GGA GAC CCT TCG ATC GCC TAA AAG CCA GGA TCC 2857
Thr Ala Val Gly Val Arg Leu Arg Gly Asp Pro Ser Ile Ala End                  918

TCC CAG CCC CCA GCC CGC CAC CCT TCC AGC ACT CCT CTC TAG AAC CGA CGA CCA CAC 2914
CCC CGT GTT CCA CCC AGC AAG CCC TGG GAG ACC CAG CCA GCG CCC ACT CCG CCG CAG 2971
CAG AGG GAG GAA GGG GAC CGC AGT AAC GGA GCA CCA CGT AGA ATA CCA CCC AGA GCG 3028
CGT GTG CTG TTG ATC TGA TCT CTC GCC TGG ACC CCT AAT CCC TGC CCT GCC ACT CTG 3085
CAT GAT TCA AGT TCG ACC TGG CCA TGC ATT GCC CAT GAG TGA ACG TAG CGG CAC CCC 3142
GGT GCG TCT ACT GCA GAT GTC CAG CTA GGA AAG AGT CCC CTC TCT TGG ACA GTC TTC 3199
TGG GCC CTT CCA AGC CCA TCC GTG GAG TCG GCC TCT CCT CCC CTC TCC CCC GTG TGA 3256
AGT GTG TTA TCA CCA GCA GAC ACT GCC GGA CTC CTG CCC ACA GGG GCG TGG CCT GAA 3313
GGC GGA GTG TGG ACA TGG CAC TGC TGT TCC GTT CCC TTC CCC TCC CAG CAC CCG CCG 3370
CAG CCT GCC ATC CCG TCT GGA TGT ATC GAT GCC ACA GAA TTG TGA ATT GTG TGT CCG 3427
TCC GTG GAG CCA GTC CTA GCC ACA TTA TTG ACA GTC TTG CAT TTT GTT TTG TCT CCT 3484
GGT GGT GGG GGT GGA GGT GGT AGG GGT GCG CTA AGG TGG GCA GTC CTG TGG GAG AAC 3541
ATC TTG CTA GAA GGA ACC AAC CCA CGA AAC AAC ACC ATC ACT GGA ATT TCC ATC GCC 3597
```

Figure 7. Nucleotide and Deduced Amino Acid Sequence of cDNA Clone HKI 12.4-4. The nucleotide sequence of clone HKI 12.4-4 is shown under which the deduced amino sequence is given. Pertinent regions establishing the identity of this clone as coding for type I hexokinase are underlined. These regions match amino acid sequences derived directly from the enzyme (see text).(NOTE: This figure contains amino acids 468-918 as is shown by the numbering. Figure 10 contains amino acids 1-467.)

regions are as follows. (a) A stretch of 35 amino acids determined directly from the 40 kDa C-terminal fragment of rat brain hexokinase, reported by Polakis and Wilson (71), matches residues 552-586. (b) The sequences of Peptides I and III, determined by Schirch and Wilson (92), match residues 625-636 and 597-616, respectively, with the only discrepancy being the N-terminal residue of Peptide III (previously identified as Ile and shown here to be Met). Additionally, the partial amino acid sequence reported for Peptide II matches residues 802-816. All of the peptides discussed so far were generated through the use of the protease trypsin, and, as Figure 7 shows, each peptide has either an Arg or a Lys immediately upstream from the N-terminal residue, as expected, due to the specificity of this protease. (c) The two terminal residues, 917-918, immediately preceding the stop codon, are identical to the C-terminal sequence reported by Polakis and Wilson (71) for the intact enzyme. Therefore, a single open reading frame spanning 451 residues coupled with the amino acid sequence identities discussed above clearly establish this clone as coding for type I hexokinase.

Cloning of Full Length Rat Brain Hexokinase cDNA

A rat brain λgt10 cDNA library was screened with the previously isolated cDNA clone, HKI 12.4-4. The largest clone isolated was designated as HKI 1.4-7 (2) and regions of this clone not contained in clone HKI 12.4-4 were sequenced (both strands) after generating non-random

deletions in order to complete the coding region for rat brain hexokinase (Figure 8).

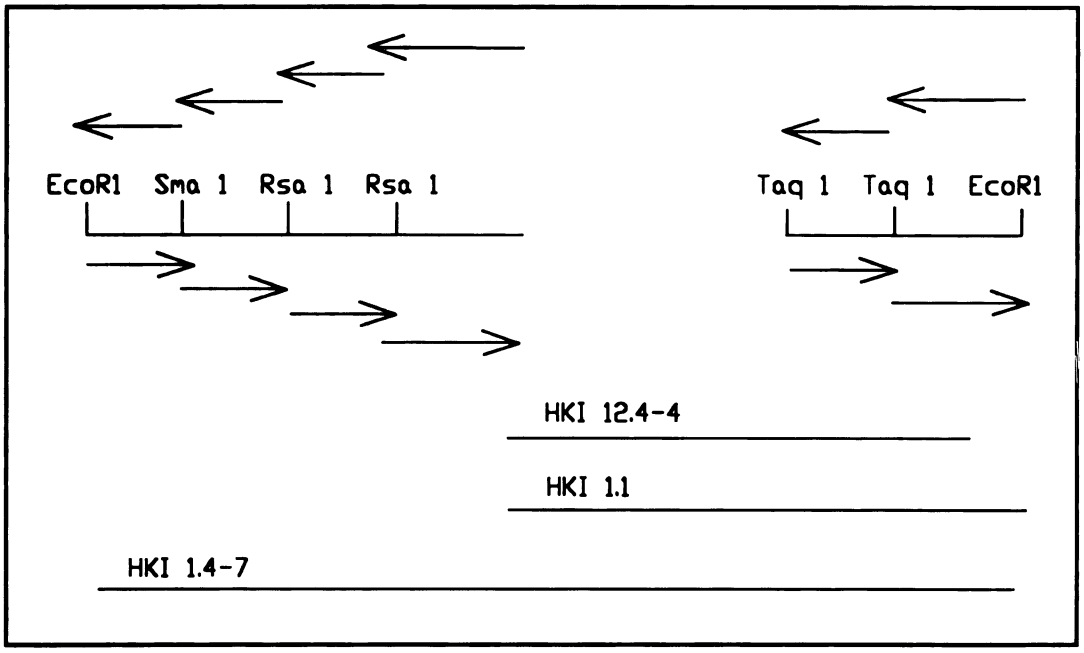

Figure 8. Sequencing Strategy for Type I cDNA Clones and Relevant Restriction Sites. Regions of cDNA clone HKI 1.4-7 not present in HKI 12.4-4 were sequenced after the generation of subclones (represented by arrows) *via* nonrandom deletions. The 3' end of HKI 1.1, not present in HKI 1.4-7, was also sequenced.

The cDNA clone HKI 1.4-7 contains a 3.7 kb insert which starts 91 bases upstream from the translation initiation codon. This clone extends 32 bases past the 3' end of the previously isolated and sequenced clone, HKI 12.4-4, and since clone HKI 12.4-4 contains the stop codon (and extensive 3' untranslated sequence > 700 bps), cDNA clone HKI 1.4-7, therefore, includes all of the coding region. A second cDNA clone, designated HKI 1.1 (2), contains a 2 kb

insert and extends an additional 13 bases beyond the 3′ end of HKI 1.4-7 concluding with 27 adenine residues, the beginning of a presumptive poly(A^+) tail. Sequencing of the 3′ end of clone HKI 1.1 provided the 40 bases not included in HKI 1.4-7 (Figure 8). (The sequence of greater than 500 bases upstream from the 40 bases at the 3′ end of HKI 1.1 was determined to be identical to the 3′ end of clone HKI 1.4-7.) In conclusion, these clones were determined to represent 3.7 of the 4.3 kb present in the mRNA detected in a Northern blot of rat brain mRNA (Figure 9).

Authenticity of Full Length Clone HKI 1.4-7

The composite nucleotide sequence determined from HKI 1.4-7 and HKI 1.1, and unique from clone HKI 12.4-4, is shown in Figure 10, under which the deduced amino acid sequence is given. Regions of the deduced amino acid sequence matching those previously determined directly from the enzyme are underlined and are as follows. (a) Starting at base 92, and establishing clone HKI 1.4-7 as containing the initiating Met (and hence the entire coding region), is a 9 residue amino acid sequence which agrees well with the N-terminal sequence of the enzyme determined by Polakis and Wilson (45). It should be pointed out that the deduced amino acid sequence Ala-Ala-Gln (residues 3 to 5), which could not be unambiguously identified (for technical reasons) by Polakis and Wilson (45), was reported as (Ala,Gln)-Ala. (b) Residues 102-121 match a 20 amino acid sequence determined directly from the N-terminus of the 90 kDa fragment produced

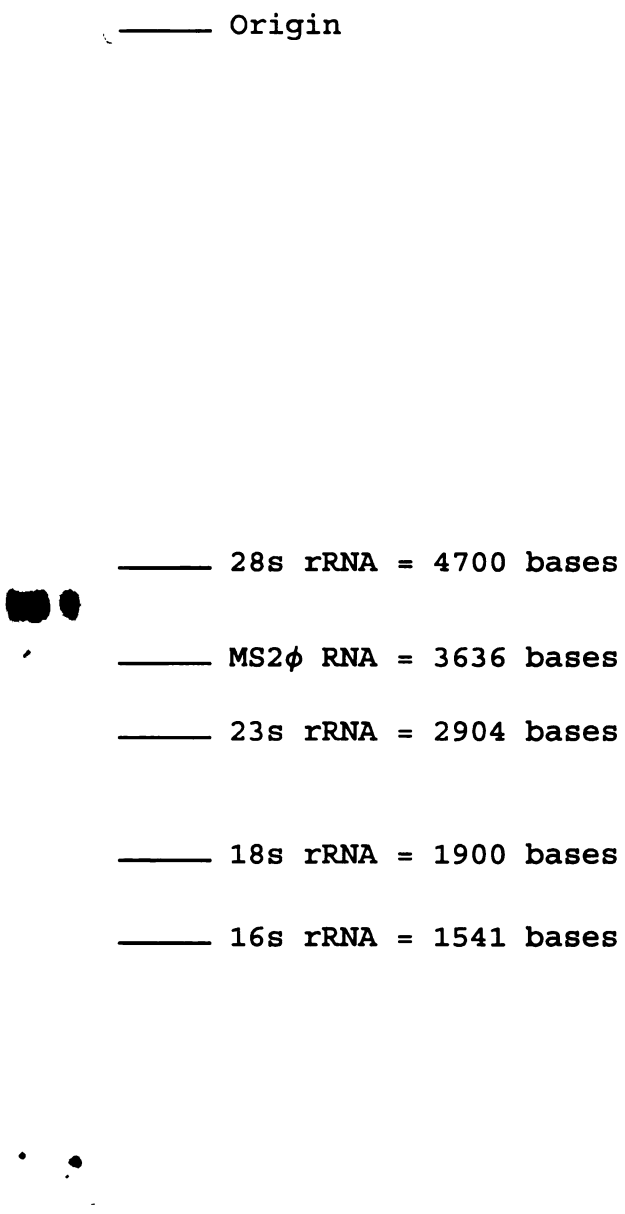

Figure 9. Northern Blot for Type I Hexokinase mRNA. Positions and sizes of control RNAs are as shown. 10 ugs of rat brain mRNA was probed with cDNA clone HKI 12.4-4 (see Methods).

```
  C GCC GAT CTG CCG CTG GAG GAC CAC TGC TCA CCA GGG CTA CTG AGG AGC CAC TGG    55
CCC CAC ACC TGC TTT CCC GCA TCC CCC ACC GTC AGC ATG ATC GCC GCG CAA CTA CTG   112
                                            Met Ile Ala Ala Gln Leu Leu      7

GCC TAT TAC TTC ACC GAG CTG AAG GAT GAC CAA GTC AAA AAG ATT GAC AAG TAT CTG   169
Ala Tyr Tyr Phe Thr Glu Leu Lys Asp Asp Gln Val Lys Lys Ile Asp Lys Tyr Leu    26

TAC GCC ATG CGG CTC TCT GAT GAG ATT CTG ATA GAT ATC CTG ACA CGA TTC AAG AAA   226
Tyr Ala Met Arg Leu Ser Asp Glu Ile Leu Ile Asp Ile Leu Thr Arg Phe Lys Lys    45

GAG ATG AAG AAT GGC CTC TCC CGG GAT TAT AAT CCA ACA GCC TCC GTC AAG ATG CTG   283
Glu Met Lys Asn Gly Leu Ser Arg Asp Tyr Asn Pro Thr Ala Ser Val Lys Met Leu    64

CCC ACC TTG CTC CGG TCC ATT CCG GAC GGC TCA GAA AAG GGG GAT TTC ATT GCC CTG   340
Pro Thr Leu Leu Arg Ser Ile Pro Asp Gly Ser Glu Lys Gly Asp Phe Ile Ala Leu    83

GAT CTC GGC GGG TCT TCC TTT CGA ATC CTG CGG GTG CAG GTG AAC CAC GAG AAG AAC   397
Asp Leu Gly Gly Ser Ser Phe Arg Ile Leu Arg Val Gln Val Asn His Glu Lys Asn   102

CAG AAC GTC AGC ATG GAG TCT GAG ATC TAC GAC ACC CCA GAG AAC ATC GTG CAT GGC   454
Gln Asn Val Ser Met Glu Ser Glu Ile Tyr Asp Thr Pro Glu Asn Ile Val His Gly   121

AGT GGA ACC CAG CTT TTC GAT CAT GTC GCT GAC TGC CTG GGA GAC TTC ATG GAG AAA   511
Ser Gly Thr Gln Leu Phe Asp His Val Ala Asp Cys Leu Gly Asp Phe Met Glu Lys   140

AAG AAG ATC AAG GAC AAG AAG TTA CCC GTG GGA TTC ACA TTT TCC TTC CCC TGC CGA   568
Lys Lys Ile Lys Asp Lys Lys Leu Pro Val Gly Phe Thr Phe Ser Phe Pro Cys Arg   159

CAA TCC AAG ATA GAT GAG GCT GTA CTG ATC ACG TGG ACA AAG CGG TTC AAA GCC AGT   625
Gln Ser Lys Ile Asp Glu Ala Val Leu Ile Thr Trp Thr Lys Arg Phe Lys Ala Ser   178

GGC GTG GAA GGA GCG GAT GTG GTC AAG TTG CTG AAT AAA GCC ATT AAG AAG CGA GGG   682
Gly Val Glu Gly Ala Asp Val Val Lys Leu Leu Asn Lys Ala Ile Lys Lys Arg Gly   197

GAC TAT GAT GCT AAC ATT GTC GCC GTG GTG AAT GAC ACA GTA GGG ACC ATG ATG ACC   739
Asp Tyr Asp Ala Asn Ile Val Ala Val Val Asn Asp Thr Val Gly Thr Met Met Thr   216

TGC GGT TAT GAT GAC CAA CAG TGT GAA GTC GGC CTG ATC ATT GGC ACA GGC ACC AAT   796
Cys Gly Tyr Asp Asp Gln Gln Cys Glu Val Gly Leu Ile Ile Gly Thr Gly Thr Asn   235

GCT TGC TAC ATG GAG GAA CTG CGA CAC ATC GAC CTG GTG GAA GGC GAC GAG GGG AGG   853
Ala Cys Tyr Met Glu Glu Leu Arg His Ile Asp Leu Val Glu Gly Asp Glu Gly Arg   254

ATG TGT ATT AAC ACG GAA TGG GGA GCC TTT GGG GAT GAT GGG TCC CTG GAA GAC ATC   910
Met Cys Ile Asn Thr Glu Trp Gly Ala Phe Gly Asp Asp Gly Ser Leu Glu Asp Ile   273

CGA ACC GAG TTT GAC AGA GAG TTA GAC CGT GGA TCT CTC AAC CCT GGG AAG CAG CTG   967
Arg Thr Glu Phe Asp Arg Glu Leu Asp Arg Gly Ser Leu Asn Pro Gly Lys Gln Leu   292

TTC GAG AAG ATG GTG AGC GGC ATG TAC ATG GGG GAG CTG GTC CGG CTA ATC CTG GTG 1024
Phe Glu Lys Met Val Ser Gly Met Tyr Met Gly Glu Leu Val Arg Leu Ile Leu Val   311

AAG ATG GCC AAG GAA GGC CTC TTA TTC GAA GGG CGC ATC ACT CCA GAG CTG CTC ACG 1081
Lys Met Ala Lys Glu Gly Leu Leu Phe Glu Gly Arg Ile Thr Pro Glu Leu Leu Thr   330

AGG GGA AAG TTC AAC ACT AGT GAC GTG TCC GCC ATT GAA AAG GAT AAG GAA GGC ATT 1138
Arg Gly Lys Phe Asn Thr Ser Asp Val Ser Ala Ile Glu Lys Asp Lys Glu Gly Ile   349

CAA AAT GCC AAG GAA ATC TTA ACC CGC TTG GGA GTG GAG CCG TCT GAT GTT GAC TGT 1195
Gln Asn Ala Lys Glu Ile Leu Thr Arg Leu Gly Val Glu Pro Ser Asp Val Asp Cys   368

GTG TCG GTC CAG CAC ATC TGC ACG ATC GTC TCC TTC CGA TCA GCC AAC CTG GTG GCC 1252
Val Ser Val Gln His Ile Cys Thr Ile Val Ser Phe Arg Ser Ala Asn Leu Val Ala   387

GCC ACG CTC GGT GCC ATC TTG AAC CGC CTG CGG GAC AAC AAG GGC ACA CCC AGC CTG 1309
Ala Thr Leu Gly Ala Ile Leu Asn Arg Leu Arg Asp Asn Lys Gly Thr Pro Ser Leu   406

CGG ACC ACG GTT GGC GTG GAC GGT TCT CTC TAC AAG ATG CAC CCA CAG TAC TCC CGG 1366
Arg Thr Thr Val Gly Val Asp Gly Ser Leu Tyr Lys Met His Pro Gln Tyr Ser Arg   425

CGG TTC CAC AAG ACC CTG AGG CGG GTG GTG CCT GAC TCC GAC GTC CGT TTC CTC CTC 1423
Arg Phe His Lys Thr Leu Arg Arg Val Val Pro Asp Ser Asp Val Arg Phe Leu Leu   444

TCA GAG AGT GGC ACG GGC AAG GGG GCC GCC ATG GTG ACG GCA GTA GCC TAC CGC CTG 1480
Ser Glu Ser Gly Thr Gly Lys Gly Ala Ala Met Val Thr Ala Val Ala Tyr Arg Leu   463

GCT GAG CAG CAC /1493-3597/ TTTAGTGAGCCATTGTTGTACGTCTAGTAAACTTTGTACTGATTCAA 3644
Ala Glu Gln His

AAAAAAAAAAAAAAAAAAAAAAAAA  3669
```

Figure 10. Composite Nucleotide Sequence Obtained from cDNA Clones HKI 1.4-7 and HKI 1.1. The last 40 bps are from HKI 1.1. Nucleotides 1490-3597 (not shown) correspond to cDNA clone HKI 12.4-4 (Figure 7). The deduced amino acid sequence corresponds essentially to the N-terminal half of the enzyme. Sequences derived directly from the enzyme and shown to be in the deduced sequence are underlined. Only part of the N-terminal amino acids corresponding to the 48 kDa fragment are shown (beginning at residue 463). The segment encoding the presumed polyadenylation signal (93) is also underlined; the consensus signal is $A_{98}A_{91}U_{100}A_{99}A_{99}A_{98}$, where the subscripts represent the percentage of 134 vertebrate mRNAs examined (93) that contained the designated base at the indicated position.

by tryptic digestion under the conditions of Polakis and Wilson (71). Summation of the molecular weights of the deduced 817 amino acid residues corresponding to the 90 kDa fragment gives a value of 90,719 Da, in agreement with the experimentally determined size. (c) A 9 residue amino acid sequence, corresponding to residues 463-471 in the deduced amino acid sequence, completely matches that determined by White and Wilson (74) directly from the N-terminus of the 48 kDa fragment (produced by tryptic cleavage at T_3 under partially denaturing conditions). Summation of the molecular weights of the 456 deduced amino acids corresponding to this fragment gives a value of 50,749 Da, which is similar to the experimentally determined size of 48 kDa. Additionally, immediately upstream from the N-terminus of each of the fragments discussed above is an Arg or Lys residue, as expected, due to the generation of these fragments *via* tryptic cleavage. In summary, the sequence identities (demonstrated between the deduced amino acid sequence and those derived directly from the enzyme), both discussed above and previously with clone HKI 12.4-4, are located throughout the deduced primary sequence of the enzyme beginning with the initiating Met, spanning an open reading frame coding for 918 amino acids, and concluding with the terminal Ala. Therefore, there is little doubt that cDNA clone HKI 1.4-7 contains the entire coding region of rat brain type I hexokinase.

Comparison of Hexokinase Type I Halves and Yeast Isozymes

The deduced amino acid sequences of the N- and C-terminal halves of rat brain hexokinase and yeast hexokinase isozymes A and B are aligned in Figure 11. It is evident that the similarity between the N- and C-terminal halves of the brain enzyme and between these and the yeast hexokinase isozymes is rather extensive. Indeed, when the N- and C-terminal halves are quantitatively compared, 47% of the amino acid residues are identical and an additional 17% represent conservative substitutions. This high degree of similarity, along with the similarity to the yeast isozymes, certainly supports the proposal (19,50,67-70) that this mammalian hexokinase evolved by duplication and fusion of a gene encoding an ancestral hexokinase of ≈ 50 kDa.

Comparison of the N-terminal half of rat brain hexokinase with the A isozyme of yeast hexokinase reveals that 27% of the residues are identical with an additional 15% being the result of conservative substitutions. Similarly, comparison of the C-terminal half of rat brain hexokinase with yeast hexokinase isozyme A shows that 28% of the residues are identical with an additional 15% classified as conservative substitutions. Furthermore, the alignment in Figure 11 shows that in comparisons of either the N- or C-terminal half of rat brain hexokinase with the yeast isozymes, the similar residues (identical + conservative substitutions) are located throughout the amino acid sequence of the yeast isozymes.

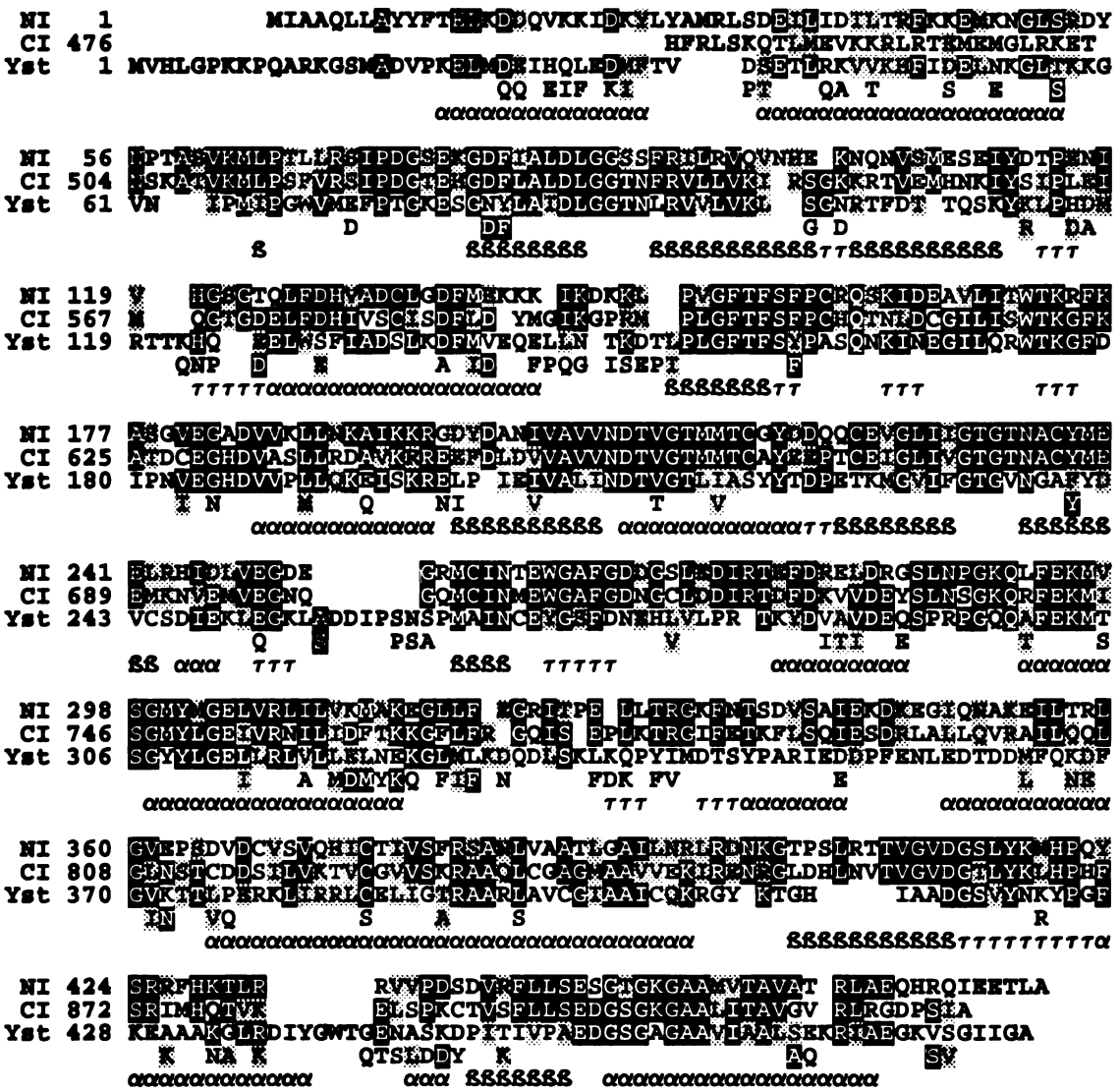

Figure 11. Aligned Amino Acid Sequences of Rat Brain and Yeast Hexokinases. The N- and C-terminal halves of rat brain hexokinase (NI and CI, respectively) are aligned above the yeast A (Yst) and B isozymes (only those residues in the B isozyme which differ from the A isozyme are shown). Blackened regions correspond to identical residues and stippled regions correspond to conserved residues. Secondary structural features are designated beneath the aligned sequences, with α denoting α-helices, ß denoting ß-strands, and τ denoting ß-turns.

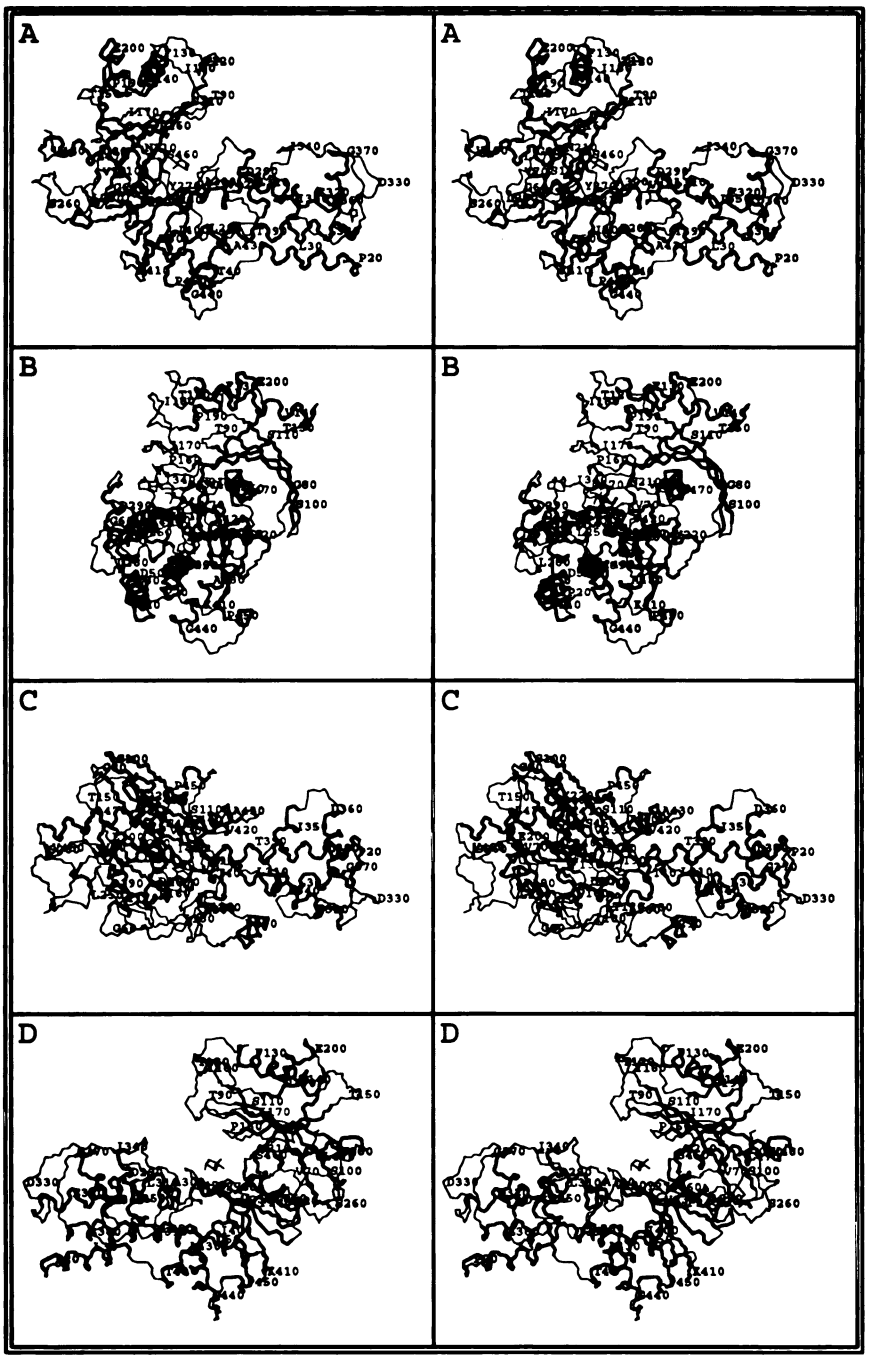

Figure 12. Stereo Images of Yeast Hexokinase Highlighting Secondary Structural Features. Alternate views of the yeast hexokinase crystal structure containing bound glucose are shown with the darkened regions corresponding to either α-helices or ß-sheets.

Secondary structural features are designated below the sequences in Figure 11. These features were determined from the crystal structure of the "open" conformation of yeast hexokinase (see Methods, chapter II) (54,55) and are highlighted in the stereo images in Figure 12. Using the alignment in Figure 11, the residues conserved in both the N-terminal half of rat brain hexokinase and yeast hexokinase A were mapped to the yeast hexokinase crystal structure. These residues are highlighted in the stereo images in Figure 13, parts A and B. This has also been carried out with the conserved residues in the C-terminal half of rat brain hexokinase and the A isozyme of yeast hexokinase and is shown in parts C and D of Figure 13. The stereo images demonstrate that, although the conserved residues are located throughout the respective structures, there is a high degree of conservation in regions that comprise the cleft and secondary structural features. Conversely, the least conserved regions map to the surface of the enzyme structure, as expected. Therefore, extensive similarity in the secondary and tertiary structures of these enzymes seems to be a reasonable expectation (94-96). Consequently, the yeast crystal structures provide a reasonable model which can be used to establish the location within the tertiary structure (at least to a first approximation) of conserved residues present in either half of rat brain hexokinase. Harrison (66), working in Steitz's laboratory, refined the crystal coordinates well enough to

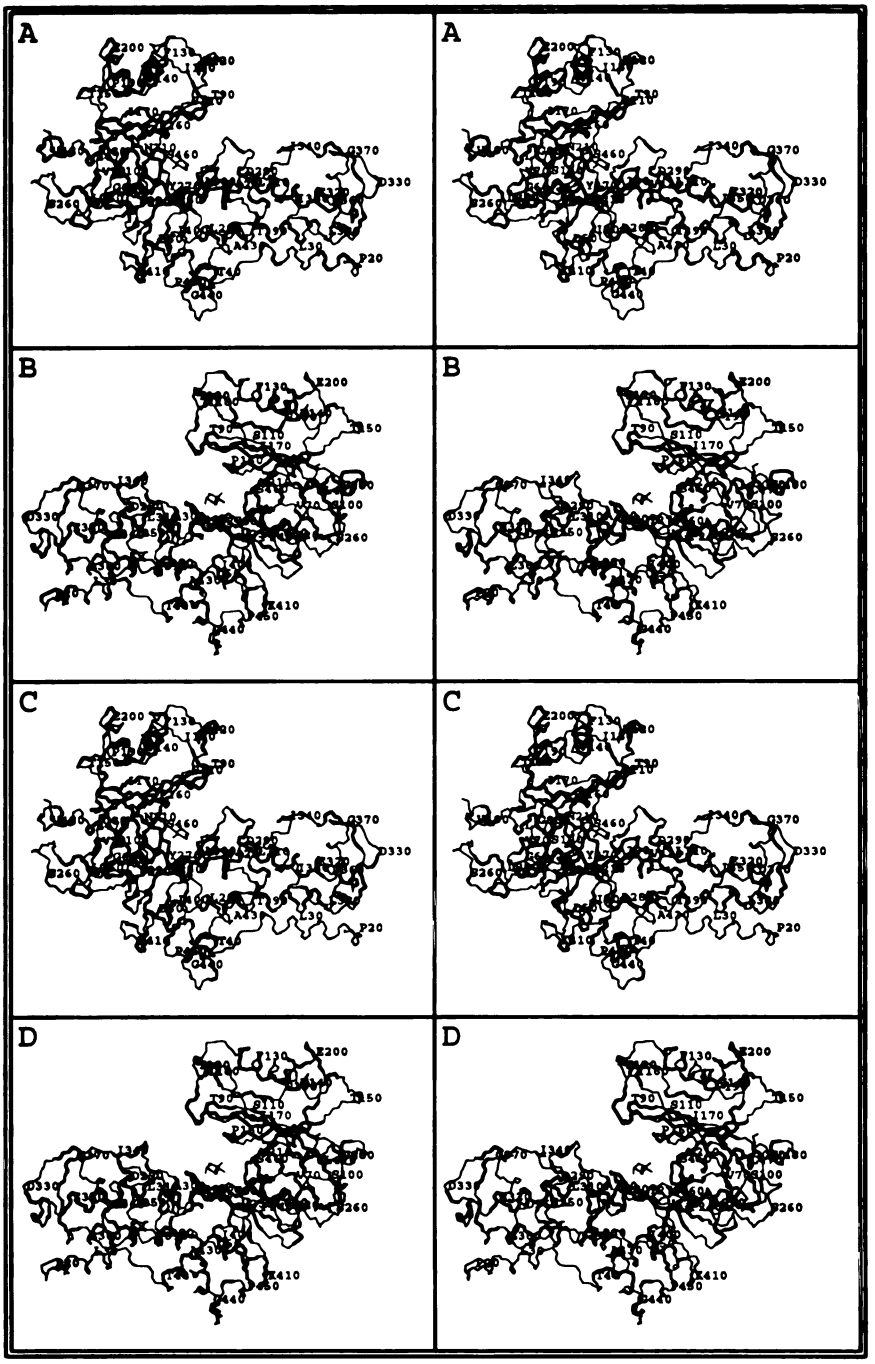

Figure 13. Stereo Images Highlighting Conserved Residues of Type I Hexokinase. Alternate views of yeast hexokinase with darkened residues being conserved between yeast hexokinase A and the N-terminal half of type I hexokinase (**A and B**) or the C-terminal half of type I hexokinase (**C and D**).

identify residues that hydrogen bond to the hydroxyls of the bound glucose molecule. These residues include: Ser-158, Asn-210, Asp-211, Gly-235, Asn-237, Glu-269, and Glu-302. If the yeast crystal structures are reasonable approximations to the two halves of rat brain hexokinase, conservation of residues providing as crucial a role as the binding of the substrate glucose would be a fair expectation (certainly in the C-terminal half of rat brain hexokinase which has been shown to be catalytically active (74)). Conservation of Ser-158 had previously been demonstrated due to its presence in the sequence of Peptide III isolated by Schirch and Wilson (92 and discussed below). Conservation of the other residues could not be confirmed with the limited sequence information that existed before the cloning of the cDNA for rat brain hexokinase. The alignment in Figure 11 now demonstrates that each of these residues has been conserved in the C-terminal half of rat brain hexokinase. Surprisingly, these residues are also conserved in the N-terminal half of the molecule. Conservation of <u>all</u> these residues in the N-terminal half of the molecule was not expected since this half appears to no longer possess catalytic activity.

Studies have been conducted by Schirch and Wilson (92) on the glucose binding site of hexokinase. During the course of their work, three key peptides were isolated, designated Peptides I, II, and III (mentioned above). Peptides I and III were identified as being located at the glucose binding site of brain hexokinase, based on their reactivity with a

glucose analog and protection by competing ligands, and are highly similar to sequences found at the glucose binding region of the yeast enzymes. Although Peptide II was also labeled with the reactive glucose analog, unlike Peptides I and III, competitive ligands did not prevent the labeling of this peptide. (The reason Peptide II was labeled is presently unclear.) Due to the fact that there exists no significant homology between this peptide and the sequences of the yeast isozymes, Schirch and Wilson (92) were unable to locate this peptide within the yeast structure. Now, with the determination of the entire amino acid sequence of rat brain hexokinase and the ability to use the yeast crystal structures to map the location of Peptide II, the inability of competitive ligands (vs. glucose) to prevent the labeling of this peptide is readily apparent. Figure 14 shows that this peptide is located in the large lobe far from the cleft containing the active site. Peptides I and III are also highlighted and their proximity to the active site is easily seen, in accord with the proposal of Schirch and Wilson (92). Additionally, examination of the sequence for the N-terminal portion of brain hexokinase (Figure 11) shows that these peptides are sufficiently unique in sequence that their location within the overall sequence could be established. Accordingly, all three peptides were derived from the C-terminal half of the enzyme, which is certainly in support of the C-terminal half as possessing the catalytic site, as was concluded by Schirch and Wilson (92).

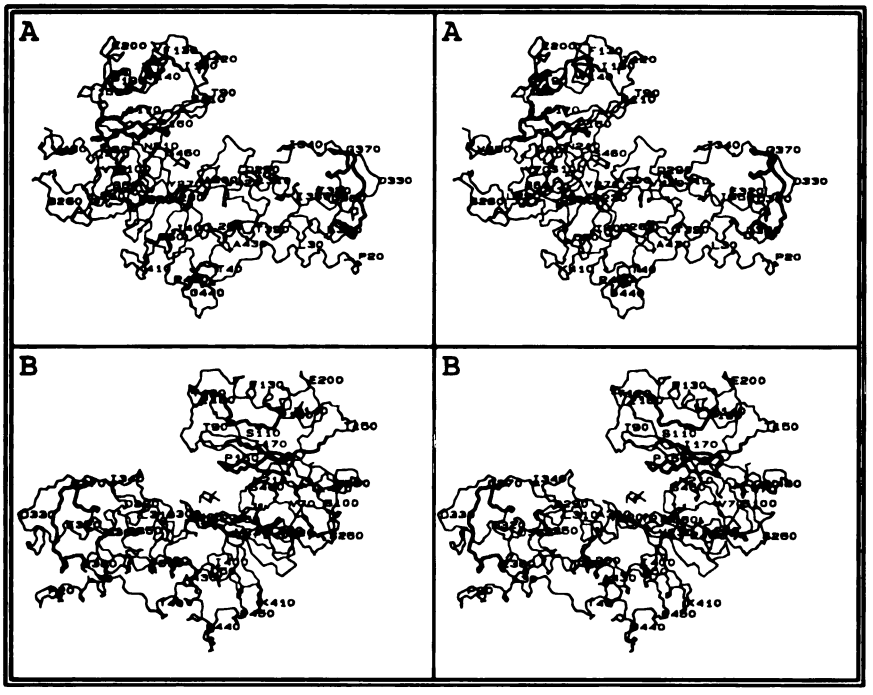

Figure 14. Stereo Images Showing the Locations of Peptides I, II, and III. **A and B**: Alternate views depicting the location of Peptides I (180-191), II (364-378), and III (152-171) which were labeled with a glucose analog by Schirch and Wilson (92).

Figure 11 reveals that there are several insertions and deletions that have occurred during the evolution of yeast and mammalian hexokinases. Most of these differences map to surface regions in the yeast crystal structure or are located near the ends of secondary structural features. Frequently, these differences seem unlikely to result in radical changes to the overall structure (94,97). However, there are some insertions and deletions in the rat brain enzyme which, due to their magnitude, seem likely to significantly alter the yeast crystal structure. They are: the two deletions in the mammalian enzyme corresponding to residues 255-261 and 437-443 of the yeast hexokinases; and a 5 residue segment in both the N- and C-terminal halves (residues 405-409 and 853-857, respectively) of the mammalian enzyme, which would be inserted between residues 413 and 414 in the yeast hexokinases (Figure 15). All of these changes occur in the hinge region which links the small and large lobes in the yeast hexokinase structure. Although exactly how these differences manifest themselves is unknown, they are located in a region where they seem certain to impact on the structure.

Proposed Structure for Mammalian Hexokinase Type I

Using the yeast hexokinase crystal structures as reasonable approximations to the structures of the two halves of rat brain hexokinase, a model for the entire rat brain hexokinase enzyme was constructed (2). The alignment in Figure 11 indicates that the C-terminal half of rat brain

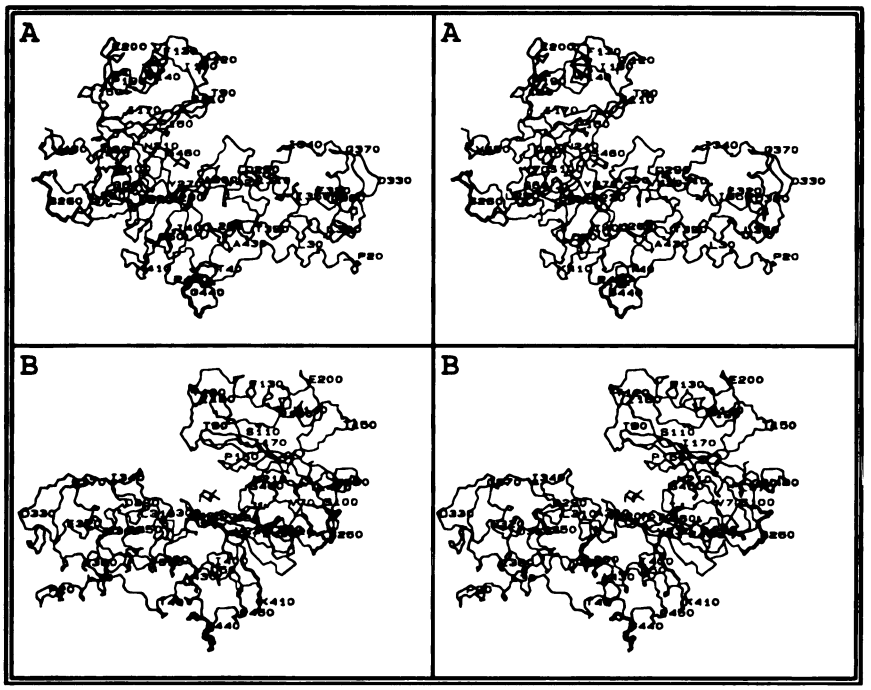

Figure 15. Stereo Images Depicting Structural Differences Between Yeast and Type I Hexokinases. **A and B**: Alternate views of insertion (darkened region) and deletions (dotted region) that have occurred in the evolution of type I hexokinase.

hexokinase lacks the region spanning from the N-terminus, up through, and including the first α-helix of yeast hexokinase. After deleting this region from one of the yeast structures, the resulting N-terminus of this molecule was fused to the C-terminus of a complete yeast crystal structure. The C-terminal half was then rotated, keeping the N-terminal half stationary, in order to eliminate any steric conflicts. The final structure arrived at (Figure 16, parts A and B) was one in which the two "halves" were allowed close enough approach such that noncovalent interactions between them were possible. Although the model was constructed in a rather subjective manner, it seems far from arbitrary due to the fact that the possible structural alignments were limited.

It should be pointed out that the structure in Figure 16 is missing those amino acids that comprise the extreme N-terminal sequence up to the beginning of the first α-helix. Steitz and colleagues were unable to determine a structure for this region due to localized disorder (55,56,58,66); therefore, not shown is, presumably, a flexible peptide attached to the N-terminus of the first α-helix. Many structural features of this model agree with previously determined experimental results, as will be discussed below.

It has been well established that rat brain hexokinase binds to the outer membrane of mitochondria. One of the crucial features of this binding is the presence of the hydrophobic N-terminal amino acid segment (45) which is

inserted into the membrane (46). Protrusion of this segment from the enzyme's surface would be consistent with its role in tethering the enzyme to the mitochondrial membrane as well as its noted susceptibility to proteolysis (45). This segment corresponds to the flexible peptide, referred to above, which is attached to the first α-helix in the model presented above.

Digestion of native rat brain hexokinase with trypsin results in cleavage at two very susceptible sites, T_1 and T_2 (71), which correspond to Lys-101 (Asn-102 in yeast hexokinase) in the N-terminal half, and Arg-551 (Thr-104 in yeast hexokinase) in the C-terminal half of the rat brain enzyme. These tryptic sites both map to virtually the same structural region of the yeast crystal structure; at the end of one of the ß-strands that comprise the ß-sheet of the small lobe. This region is at the surface of the yeast structure as would be expected due to its marked susceptibility to trypsin. Manifestation of both of these sites, one in each of the two halves, is consistent with an enzyme structure that is composed of two conformationally similar halves. This is precisely the case for the constructed model.

Cleavage of the native rat brain enzyme with trypsin has revealed that the N- and C-terminal halves of the enzyme interact strongly by noncovalent forces. In fact, the interactions are so strong that the proteolyzed enzyme, under native conditions behaves, in many respects, as the

intact enzyme (71). Alternatively, if tryptic digestion is carried out in 0.6 M guanidine hydrochloride, the interactions between the two halves of the enzyme are weakened and a new cleavage site is revealed (74). This tryptic cleavage site, designated T_3, has been determined to be located at Arg-462 *via* direct sequencing of the protein and is only a few residues from the site at which both halves of the model were fused together. In the proposed model (Figure 16, parts A and B), this site would be inaccessible to trypsin under native conditions due to the juxtaposition of the strongly interacting halves. However, if the interactions between the two halves were weakened by denaturant, this site would become susceptible to proteolysis. Therefore, the location of T_3 in the model is consistent with the behavior of this site in the enzyme.

Yeast hexokinase B has been crystallized as a dimer and its structure determined (ref. 98 and Figure 16, parts C and D). Due to the similarities between the two halves of rat brain hexokinase and the yeast enzymes, the possibility of the yeast dimer structure as representing that of rat brain hexokinase should be considered in that the yeast dimer may provide a model for the mammalian isozyme with respect to the relative disposition of the two monomers. This will now be discussed with the "yeast dimer" model referring to a model which would be based on the dimer structure, and the "mammalian" model referring to the model proposed above.

The absence of the region leading up to and including the first α-helix in the C-terminal half of the mammalian enzyme seems contradictory to a "yeast dimer" model (Figure 16, parts C and D). If this region is removed from either of the yeast monomers, the newly generated N-terminal end (FP - fusion point in Figure 16) of the C-terminal half (of rat brain hexokinase), would be located quite a distance from the C-terminal end (T_3 in Figure 16) of the other monomer (the N-terminal half of rat brain hexokinase). These two ends would have to be fused to create the single polypeptide of rat brain hexokinase. Additionally, manifestation of T_3 does not support this model in that this region is totally exposed in both of the monomers and hence would be susceptible to proteolysis in the native structure.

Although manifestation of T_3 and problems with the fusion of the two halves indicate the yeast dimer is unsatisfactory as a model for rat brain hexokinase, more recent experimental evidence totally eliminates this structure from consideration, and moreover, is consistent with the "mammalian" model proposed above. This evidence comes from the work of Smith and Wilson (99,100) in which they defined the epitopic regions recognized by monoclonal antibodies raised to native rat brain hexokinase and will be discussed below.

Although the yeast dimer is composed of two identical subunits (with respect to primary sequence), Steitz *et al.* (98) concluded that due to heterologous interactions between

the two monomers, they are not structurally equivalent. Hence the designation of one of the monomers as being the "up subunit" and the other as being the "down subunit". Therefore, the "yeast dimer" model presents two possibilities in terms of modeling rat brain hexokinase. The first possibility, in which the "down subunit" (monomer on the right in Figure 16, part C) corresponds to the N-terminal half of rat brain hexokinase (with the "up subunit" corresponding to the C-terminal half of rat brain hexokinase), can be eliminated due to the fact that monoclonal antibody 3A2 (99) binds to residues in the N-terminal half of rat brain hexokinase which correspond to yeast hexokinase residues 36-60 (highlighted in Figure 16, part C). This region is occupied by the other monomer in the yeast dimer, which of course would preclude the binding of this antibody. The second possibility is that the "up subunit" (monomer on the left in Figure 16, part D) corresponds to the N-terminal half of rat brain hexokinase (with the "down subunit" now corresponding to the C-terminal half of rat brain hexokinase). This possibility does not seem reasonable again due to the epitope of monoclonal antibody 3A2 which appears to be somewhat occluded by the other monomer. Furthermore, Smith and Wilson (100) were able to successfully represent the epitopic regions recognized by a battery of monoclonals using the "mammalian" model presented above. Consequently, in mapping these epitopes, they accounted for the entire surface area of the N-terminal

half of rat brain hexokinase using the "mammalian" model, exclusive of the region that would be in contact with the C-terminal half. Not only does this potentially eliminate any variation of the "yeast dimer" model, but this strongly supports the relative disposition of the two halves of rat brain hexokinase in the "mammalian" model.

The structure proposed in Figure 16 (parts A and B) is certainly not meant to represent rat brain hexokinase in detail. The insertions and deletions mentioned previously have not been taken into account in the construction of this model nor is such an undertaking feasible at this time. More refined coordinates (66) are not available through the Brookhaven data base, although major structural changes are certainly not expected due to the resolution to which the present coordinates have been refined (55,56,58).

In conclusion, the proposed model has been shown to agree with a variety of experimental data, and despite its limitations, should prove useful in the future design and interpretation of experiments aimed at the elucidation of function to structure relationships in rat brain hexokinase.

Figure 16. Stereo Images Showing the Proposed Model of Type I Hexokinase / Yeast Hexokinase Dimer. **A and B**: Alternate views of a model of type I hexokinase constructed from two yeast hexokinase structures with the darkened half corresponding to the C-terminal half of type I hexokinase. T_1, T_2, and T_3 are tryptic cleavage sites (see text for details). **C and D**: Yeast hexokinase dimer. The darkened region is the segment corresponding to the location of the epitope for monoclonal antibody 3A2 in the type I hexokinase sequence. T_3 is located at the carboxy end of the N-terminal half and this is the point at which the N-terminal half is fused to the beginning of the C-terminal half (FP for fusion point).

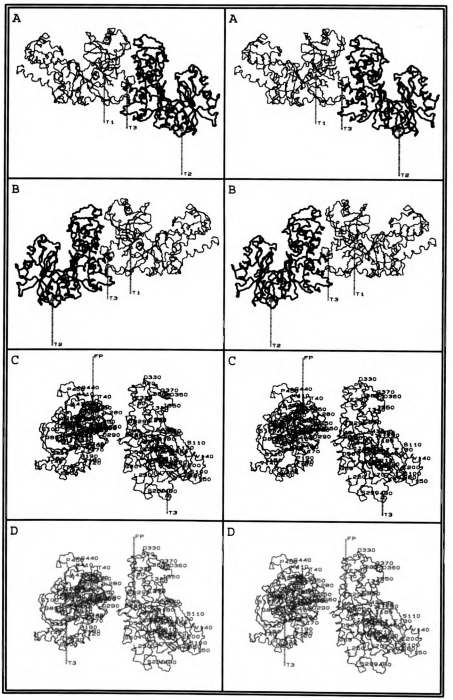

Figure 16.

CHAPTER IV

Cloning of cDNA'S Coding for Type III Hexokinase from Rat Liver and Quantitative Comparisons of Sequence Similarities Between Hexokinases

This chapter covers the cloning of cDNA's coding for type III hexokinase, after which the amino acid sequences of hexokinases from different organisms are aligned. Subsequent quantitative comparisons, using this alignment, support the duplication and fusion proposal for the evolution of the "low K_m" mammalian hexokinases as well as providing further insight into the evolution of glucokinase.

Cloning of cDNA's Coding for Type III Hexokinase

Type III hexokinase cDNA clones (7) were isolated from a rat liver cDNA library using the type I hexokinase cDNA clone HKI 1.4-7 (2). Three of the positive clones were determined to overlap and furthermore, their combined length was sufficient to provide the entire coding sequence for the 100 kDa type III isozyme (Figure 17). A 2.5 kb clone, designated L4.1-h, contained approximately 85% of the coding region and 180 bases of 3' untranslated sequence. A second clone, designated L7.1-1, included L4.1-h and additional 3' noncoding sequence which contained a presumptive polyadenylation signal and concluded with 19 adenine residues. The third clone, L7.1-2, overlapped with L4.1-h and extended in the 5' direction giving the remaining 15% of the coding region and 80 bps of 5' untranslated sequence.

Both strands of L4.1-h were completely sequenced, as were the unique regions of L7.1-1 and L7.1-2. The sequencing of L7.1-1 and L7.1-2 was extended such that at least 200 bp of overlapping sequence with the corresponding region of L4.1-h was obtained. Restriction sites relevant to

sequencing, and sequencing strategy, are depicted in Figure 17.

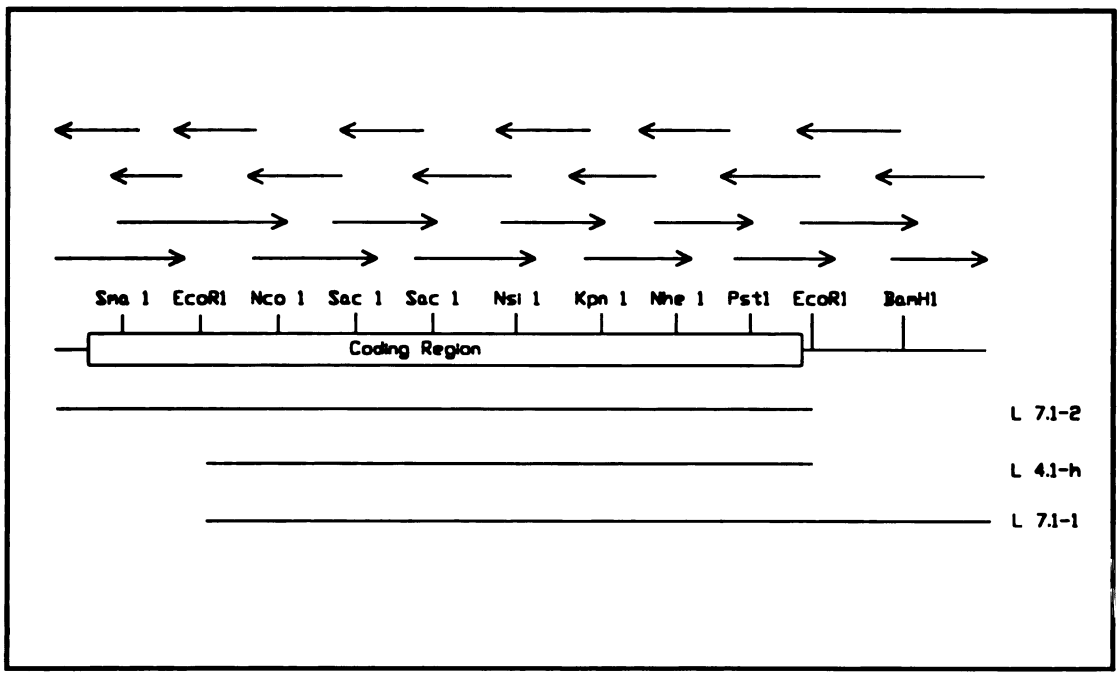

Figure 17. Sequencing Strategy for cDNA Clones Coding for Type III Hexokinase and Relevant Restriction Sites. The regions contained within clones L4.1-h, L7.1-1, and L7.1-2 are shown beneath the composite sequence. Direction and extent of sequencing of subclones (generated *via* nonrandom deletions) is indicated by the arrows.

Authenticity of Type III Hexokinase cDNA Clones

Figure 18 contains the nucleotide sequence determined from the overlapping clones coding for type III hexokinase under which the deduced amino acid sequence is given. Marcus and Ureta (101) have previously isolated tryptic peptides, designated Peptides 1 through 7, from the type III isozyme. The amino acid sequences determined from these peptides are underlined in Figure 18 (five of which are

distinct from the type I isozyme) and the presence of these sequences throughout the deduced sequence confirms that these cDNA clones code for the type III isozyme of hexokinase. Although the overlapping sequence of clone L7.1-2 with clone L4.1-h indicates it as coding for type III hexokinase, further verification was provided by the presence of Peptide 7 (unique to type III) in the deduced amino acid sequence of the 5′ region of L7.1-2 not contained in L4.1-h. In the deduced amino acid sequence immediately preceding the N-terminus of each peptide is a Lys or Arg residue which is consistent with the generation of these peptides by trypsin. There is one discrepancy, Cys-171, which was reported by Marcus and Ureta (101) to be a Ser.

Comparisons of Deduced Amino Acid Sequences of Hexokinases

The cloning of hexokinases and glucokinases from different organisms has been carried out by various researchers (see chapter I, page 2). The deduced amino acid sequences of these clones are aligned in Figure 19. (Note: The sequence of *Z. mobilis* glucokinase (102) was not included in Figure 19. The degree of similarity was very low and regions that were highly conserved in all of the other sequences were not conserved in this glucokinase. Upon translating the nucleotide sequence, some of the highly conserved regions were found to exist in the alternate reading frames; therefore this sequence was not included since it seems likely to contain sequencing errors.)

Using the alignment in Figure 19, quantitative

```
GT CGT CTT ATT TGG GAG CTG AGA CTT GAG GAA GGT GAT AAC TTC TGA ATC CCC CCA GGT AGT CAA TAC CAT TGT GGA AAC ATG GCC GCC   89
                                                                                                  Met Ala Ala    3

ATT GAG CCT TCT GGT CTG CAC CCG GGA GAA AGA GAC TCA AGC TGC CCC CAG GAG GGC ATT CCA AGG CCC TCA GGT AGC TTA GAA CTG GCA  179
Ile Glu Pro Ser Gly Leu His Pro Gly Glu Arg Asp Ser Ser Cys Pro Gln Glu Gly Ile Pro Arg Pro Ser Gly Ser Leu Glu Leu Ala   33

CAG GAA TAC TTG CAA CAA TTC AAG GTG ACC ATG ACA CAG CTG CAG CAG ATC CAA GCC AGT CTT CTG TGT TCC ATG GAG CAG GCG CTG AAG  269
Gln Glu Tyr Leu Gln Gln Phe Lys Val Thr Met Thr Gln Leu Gln Gln Ile Gln Ala Ser Leu Leu Cys Ser Met Glu Gln Ala Leu Lys   63

GGA CAG GAC AGT CCC GCT CCT TCT GTC CGG ATG TTG CCC ACA TAC GTG AGG TCC ACA CCA CAT GGC ACC GAG CAA GGA GAC TTC CTG GTG  359
Gly Gln Asp Ser Pro Ala Pro Ser Val Arg Met Leu Pro Thr Tyr Val Arg Ser Thr Pro His Gly Thr Glu Gln Gly Asp Phe Leu Val   93

CTG GAG CTG GGG GCC ACA GGA GCC TCA CTA CGT GTG TTG TGG GTA ACA CTG ACG GGC ACC AAG GAA CAC AGC GTG GAG ACC AGG AGC CAG  449
Leu Glu Leu Gly Ala Thr Gly Ala Ser Leu Arg Val Leu Trp Val Thr Leu Thr Gly Thr Lys Glu His Ser Val Glu Thr Arg Ser Gln  123

GAG TTT GTG ATC CCT CAA GAG GTG ATC CTA GGT GCT GGC CAG CAG CTC TTT GAC TTT GCT GCC CGC TGC CTC TCT GAA TTC CTG GAT GCA  539
Glu Phe Val Ile Pro Gln Glu Val Ile Leu Gly Ala Gly Gln Gln Leu Phe Asp Phe Ala Ala Arg Cys Leu Ser Glu Phe Leu Asp Ala  153

TAC CCC GTG GAG AAT CAG GGT CTG AAG CTT GGG TTT AAT TTC TCT TTT CCT TGT CAC CAG ACA GGC TTG GAC AAG AGC ACC CTC ATT TCC  629
Tyr Pro Val Glu Asn Gln Gly Leu Lys Leu Gly Phe Asn Phe Ser Phe Pro Cys His Gln Thr Gly Leu Asp Lys Ser Thr Leu Ile Ser  183

TGG ACA AAA GGT TTT AGG TGC AGT GGT GTG GAA GGC CAG GAT GTG GTC CAG TTG CTA AGG GAT GCC ATT CAG AGG CAG GGG ACC TAC AAT  719
Trp Thr Lys Gly Phe Arg Cys Ser Gly Val Glu Gly Gln Asp Val Val Gln Leu Leu Arg Asp Ala Ile Gln Arg Gln Gly Thr Tyr Asn  213

ATT GAT GTG GTA GCC ATG GTG AAT GAC ACA GTG GGT ACC ATG ATG GGC TGT GAG CTG GGC ACC AGG CCA TGT GAA GTC GGG CTT ATT GTA  809
Ile Asp Val Val Ala Met Val Asn Asp Thr Val Gly Thr Met Met Gly Cys Glu Leu Gly Thr Arg Pro Cys Glu Val Gly Leu Ile Val  243

GAC ACT GGT ACC AAT GCC TGT TAT ATG GAG GAA GCG AGG CAC GTG GCA GCT CTG GAT GAG GAC CGC GGC CGT ACC TGT GTC AGC ATC GAG  899
Asp Thr Gly Thr Asn Ala Cys Tyr Met Glu Glu Ala Arg His Val Ala Ala Leu Asp Glu Asp Arg Gly Arg Thr Cys Val Ser Ile Glu  273

TGG GGC TCC TTC TAT GAC GAA GAG GCC CTA GGG CCA GTA CTG ACC ACC TTC GAC GAT GCC CTG GAC CAC GAG TCC CTG GTT CCT GGT GCT  989
Trp Gly Ser Phe Tyr Asp Glu Glu Ala Leu Gly Pro Val Leu Thr Thr Phe Asp Asp Ala Leu Asp His Glu Ser Leu Val Pro Gly Ala  303

CAG AGG TTT GAG AAG ATG ATT GGT GGC CTT TAC TTG GGT GAG CTG GTA AGG CTG GTG CTG GTC CAC TTG TCC CAG CAT GGG GTC CTC TTT 1079
Gln Arg Phe Glu Lys Met Ile Gly Gly Leu Tyr Leu Gly Glu Leu Val Arg Leu Val Leu Val His Leu Ser Gln His Gly Val Leu Phe  333

GGT GGC TGC GCC TCT CCT GCG TTG CTG AGT CAA AAC AGC ATC CTC CTG GAA CAT GTG GCC AAA ATG GAG GAC CCT GCC ACT GGG ATA GCC 1169
Gly Gly Cys Ala Ser Pro Ala Leu Leu Ser Gln Asn Ser Ile Leu Leu Glu His Val Ala Lys Met Glu Asp Pro Ala Thr Gly Ile Ala  363

CAC GTC CAC ACA GTC CTG CAG GGC TTG GGT CTG AGC CCT CAG GCC TCA GAT GCT GAG CTC GTG CAG CGC GTG TGC ATG GCT GTG TGC ACG 1259
His Val His Thr Val Leu Gln Gly Leu Gly Leu Ser Pro Gln Ala Ser Asp Ala Glu Leu Val Gln Arg Val Cys Met Ala Val Cys Thr  393

CGA GCT GCC CAG CTC TGT GCC TCT GCC CTG GCT GCA GTC CTA TCC CGC CTC CAG CAC AGC AGG GAG CAG CAG ACA CTG CAC GTG GCC GTG 1349
Arg Ala Ala Gln Leu Cys Ala Ser Ala Leu Ala Ala Val Leu Ser Arg Leu Gln His Ser Arg Glu Gln Gln Thr Leu His Val Ala Val  423

GCC ACT GGA GGG CGA GTG TTC GAA TGG CAC CCC AGG TTC CTC TGC ATC CTA AAG GAG ACG GTA ATG CTC TTG GCC CCA GAG TGT GAT GTC 1439
Ala Thr Gly Gly Arg Val Phe Glu Trp His Pro Arg Phe Leu Cys Ile Leu Lys Glu Thr Val Met Leu Leu Ala Pro Glu Cys Asp Val  453

TCC TTC ATC CCC TCT GTG GAT GGT GGC CGG GGT GTG GCA ATG GTG ACT GCT GTG GCA GCC CGC CTG GCT ACC CAC AGG CGC ATC CTG 1529
Ser Phe Ile Pro Ser Val Asp Gly Gly Gly Arg Gly Val Ala Met Val Thr Ala Val Ala Ala Arg Leu Ala Thr His Arg Arg Ile Leu  483

GAA GAG ACC CTG GCA CCA TTT CAG CTG AGC TTG GAG CAG CTG ACA GCG GTG CAG GCA CAA ATG CGG GAA GCC ATG ATC AGG GGG CTT CAA 1619
Glu Glu Thr Leu Ala Pro Phe Gln Leu Ser Leu Glu Gln Leu Thr Ala Val Gln Ala Gln Met Arg Glu Ala Met Ile Arg Gly Leu Gln  513

GGA GAG AGC TCC TCC CTC CGC ATG CTG CCC ACT TAC GTC CGA GCA ACG CCC GAT GGC AGC GAA CGA GGT GAC TTC CTG GCT TTG GAC CTA 1709
Gly Glu Ser Ser Ser Leu Arg Met Leu Pro Thr Tyr Val Arg Ala Thr Pro Asp Gly Ser Glu Arg Gly Asp Phe Leu Ala Leu Asp Leu  543

GGG GGC ACC AAC TTC CGT GTC CTG TTG GTA CGC GTG GCC GAG GGC AGT GTT CAG ATC ACC AAC CAG GTC TAC TCT ATT CCT GAG TAT GTA 1799
Gly Gly Thr Asn Phe Arg Val Leu Leu Val Arg Val Ala Glu Gly Ser Val Gln Ile Thr Asn Gln Val Tyr Ser Ile Pro Glu Tyr Val  573

GCC CAG GGC TCT GGA CAG AAG CTC TTT GAT CAT ATT GTG GAC TGC ATC GTG GAC TTC CAG AAG AGG CAA GGC CTT AGC GGA CAG AGC CTA 1889
Ala Gln Gly Ser Gly Gln Lys Leu Phe Asp His Ile Val Asp Cys Ile Val Asp Phe Gln Lys Arg Gln Gly Leu Ser Gly Gln Ser Leu  603

CCC CTG GGT TTC ACC TTC TCT TTT CCT TGC AAG CAG CTT GGC CTG GAC CAG GGC ATC CTC CTC AAC TGG ACT AAG GGG TTC AAT GCA TCA 1979
Pro Leu Gly Phe Thr Phe Ser Phe Pro Cys Lys Gln Leu Gly Leu Asp Gln Gly Ile Leu Leu Asn Trp Thr Lys Gly Phe Asn Ala Ser  633

GGC TGC GAG GGC CAA GAT GTT GTG TAT TTA TTA CGG GAA GCC ATT AGG CGC AGA CAG GCA GTG GAG CTG AAT GTG GTT GGC ATT GTC AAT 2069
Gly Cys Glu Gly Gln Asp Val Val Tyr Leu Leu Arg Glu Ala Ile Arg Arg Arg Gln Ala Val Glu Leu Asn Val Val Ala Ile Val Asn  663

GAC ACG GTG GGG ACC ATG ATG TCC TGT GGC TAT GAT GAT CCC TGT TGT GAG ATG GGC CTC ATT GTC GGA ACC GGT ACC AAC GCC TGC TAT 2159
Asp Thr Val Gly Thr Met Met Ser Cys Gly Tyr Asp Asp Pro Cys Cys Glu Met Gly Leu Ile Val Gly Thr Gly Thr Asn Ala Cys Tyr  693

ATG GAA GAA CTC CGG AAT GTG GCG AGT GTG CCC GGG GAC TCA GGC CAC ATG TGT ATC AAC ATG GAG TGG GGT GCC TTT GGG GAT GAC GGC 2249
Met Glu Glu Leu Arg Asn Val Ala Ser Val Pro Gly Asp Ser Gly His Met Cys Ile Asn Met Glu Trp Gly Ala Phe Gly Asp Asp Gly  723

TCA CTG AGC ATG CTC GGC ACC TGC TTT GAT GCT AGC GTG GAC CAG GCA TCC ATC AAC CCA GGC AAA CAG AGG TTT GAG AAA ATG ATC AGC 2339
Ser Leu Ser Met Leu Gly Thr Cys Phe Asp Ala Ser Val Asp Gln Ala Ser Ile Asn Pro Gly Lys Gln Arg Phe Glu Lys Met Ile Ser  753

GGA ATG TAC CTG GGG GAG ATC GTC CGC CAT ATC CTC CTG CAC TTA ACC AGT CTT GGA GTT CTC TTC CGG GGC CAG AAG ACG CAA TGC CTT 2429
Gly Met Tyr Leu Gly Glu Ile Val Arg His Ile Leu Leu His Leu Thr Ser Leu Gly Val Leu Phe Arg Gly Gln Lys Thr Gln Cys Leu  783

CAG ACC AGG GAC ATC TTT AAG ACC AAG TTT CTC TCC GAG ATT GAG AGC GAC AGC CTG GCC CTG CGT CAG GTC CGA GCC ATC CTG GAG GAC 2519
Gln Thr Arg Asp Ile Phe Lys Thr Lys Phe Leu Ser Glu Ile Glu Ser Asp Ser Leu Ala Leu Arg Gln Val Arg Ala Ile Leu Glu Asp  813

CTG GGG CTG ACT CTG ACG TCT GAT GAT GCC TTG ATG GTC CTA GAG GTG TGC CAG GCT GTG TCC CGC AGG GCC GCC CAA CTC TGC GGG GCA 2609
Leu Gly Leu Thr Leu Thr Ser Asp Asp Ala Leu Met Val Leu Glu Val Cys Gln Ala Val Ser Arg Arg Ala Ala Gln Leu Cys Gly Ala  843

GGT GTG GCT GCA GTG GTG GAA AAG ATA CGG GAG AAC GGG CTG CAG GAG CTG ACA GTG TCT GTG GGA GTG GAT GGG ACG CTC TAC AAG 2699
Gly Val Ala Ala Val Val Glu Lys Ile Arg Glu Asn Arg Gly Leu Gln Glu Leu Thr Val Ser Val Gly Val Asp Gly Thr Leu Tyr Lys  873

CTA CAT CCC CAC TTC TCC AGG CTG GTG TCA GTG ACA GTT CGG AAG CTA GCC CCT CAG TGC ACA GTC ACC TTT TTG CAA TCG GAG GAT GGG 2789
Leu His Pro His Phe Ser Arg Leu Val Ser Val Thr Val Arg Lys Leu Ala Pro Gln Cys Thr Val Thr Phe Leu Gln Ser Glu Asp Gly  903

TCT GGG AAA GGG GCA GCG TTG GTC ACT CGT GTC GCT TGC CGC CTG ACC CAG ATG GCC TGC GTT TGA GGA AAA TCT CCA AAG AGC AAT TGG 2879
Ser Gly Lys Gly Ala Ala Leu Val Thr Arg Val Ala Cys Arg Leu Thr Gln Met Ala Cys Val End

GAT CTC CCT GGA CTA AGG ATT CAG CCC ACA CGG ACC CTC TGG CGG GCC CAG CCA CCC ACA ACT CCC AAA GAA ACC CAC GGG TGG TGC CCT 2969
CAT AAC TGT CCG CCC ACA CCC AAT GGC TTT CTG AGA GAA GCA CCA CTC TGG TTA GCA AGG AAT TCC CCG AAG ATG GCG GAG GGG CAG GTT 3059
CTA GTA TTG GAT GGC CGG GGC CAT CTT CTG GGC CGC CTG GCC GAT TGT GGC CAA GCA GGT ACT GCT GGG CCG AAA GGT GGT GGT TGT ACG 3149
CTG TGA GGG CAT CAA CAT TTC TGG AAA TTT CTA CAG AAA CAA GTT AAA GTA TCT GGC CTT CTC CGA AAG CGG ATG AAC ACC AAC CCG TCT 3239
CGA GGC CCC TAC CAC TTC CGA GCC CCA AGC CGC ATT TTT TGG CGC ACT GTG CGA GGC ATG CTG CCG CAC AAG ACC AAA AGA GGC CAG GCT 3329
GCC CTG GAA CGC CTC AAG GTG TTG GAT GGG ATC CCT CCA CCC TAT GAC AAG AAA AAG CGG ATG GTG GTC CCT GCT GCC CTC AAG GTT GTG 3419
CGG CTG AAG CTA CCA GAA AGT TTG CTT ACC TGG GGC GTC TGG CTC ATG AGG TCG GGT GGA AGT ACC AGG CAG TGA CAG CTA CTC TGG AGG 3509
AGA AAC GGA AGG AAA AGG CAA AGA TCC ATT ACC GGA AGA AGA AGC AGC TCT TGA GGC TAA GGA AAC AGG CAG AAA AGA ATG TGG AGA AGA 3599
AAA TCT GCA AGT TCA CAG AGG TCC TCA AGA CCA ATG GAC TCT TGG TGT GAA CCA ATA AAG ACT GTT TGT GCC TCA AAA AAA AAA AAA AAA 3689
AAA 3692
```

Figure 18. Composite Nucleotide Sequence and Deduced Amino Acid Sequence of Rat Type III Hexokinase. Amino acid sequences identical to the sequence of tryptic peptides determined by Marcus and Ureta (101) are underlined. (NOTE: within the peptide comprised of residues 163-178, the Cys-171 has not been underlined due to a discrepancy with the sequence reported by Marcus and Ureta (101), which had a Ser at this position.) Underlined in the 3′ untranslated region is the presumed polyadenylation signal (93).

comparisons between the N- and C-terminal halves of the deduced amino acid sequences of the 100 kDa enzymes (Table 2) reveal that within each isozyme the C-terminal half is quite similar to the respective N-terminal half. This supports the proposal (19,50,67-70) that the 100 kDa hexokinases arose by duplication and fusion of a gene coding for a 50 kDa enzyme.

Comparisons between the C-terminal halves of the 100 kDa isozymes show that all three "low K_m" isozymes have very similar C-terminal halves (over 60% of the residues are identical). The N-terminal halves of types I, II and III are also similar, although this similarity is not as pronounced with the type III isozyme. Nevertheless, the N-terminal half of type III is more similar to the N-terminal halves of types I or II than to the corresponding C-terminal halves. Therefore, the similarity among the N-terminal halves along with the similarity among the C-terminal halves gives support to the concept that the original 100 kDa "fused" protein was (at least) subsequently triplicated resulting in the three "low K_m" isozymes.

An indication of the evolutionary relationship of the type IV isozyme to the other hexokinases is given by the quantitative comparisons in Table 2 coupled with the evolutionary scheme described by Ureta (19). Two distinct possibilities were presented by Ureta (19) for the evolution of the mammalian isozymes. In one of the possibilities, the type IV isozyme and the other present-day 50 kDa isozymes

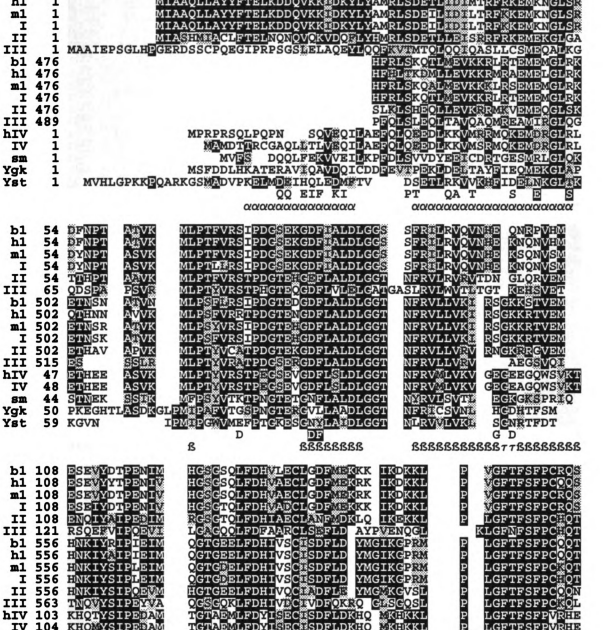

Figure 19. Alignment of Known Hexokinase and Glucokinase Sequences. The aligned sequences from top to bottom are the N-terminal halves of hexokinases from: bovine type I (b1), human type I (h1), mouse type I (m1), rat type I (I), rat type II (II), rat type III (III) followed by the respective C-terminal halves. Next are the sequences of human liver glucokinase (hIV), rat liver glucokinase (IV), schistosoma mansoni hexokinase (sm), yeast glucokinase (Ygk), yeast hexokinase A (Yst), yeast hexokinase B (only residues that differ from isozyme A are shown), and secondary structural features determined from the yeast crystal structures (α = α-helices, ß = ß-sheet, τ = ß-turns). Blackened regions correspond to identical residues and stippled correspond to conserved.

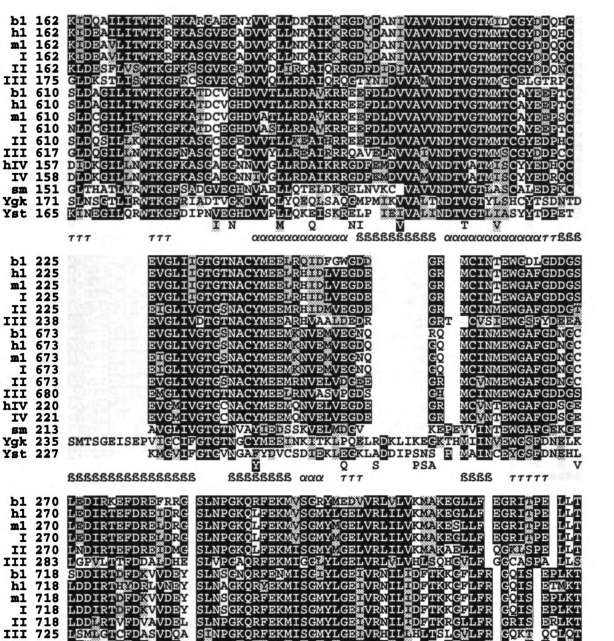

Figure 19. (Cont.)

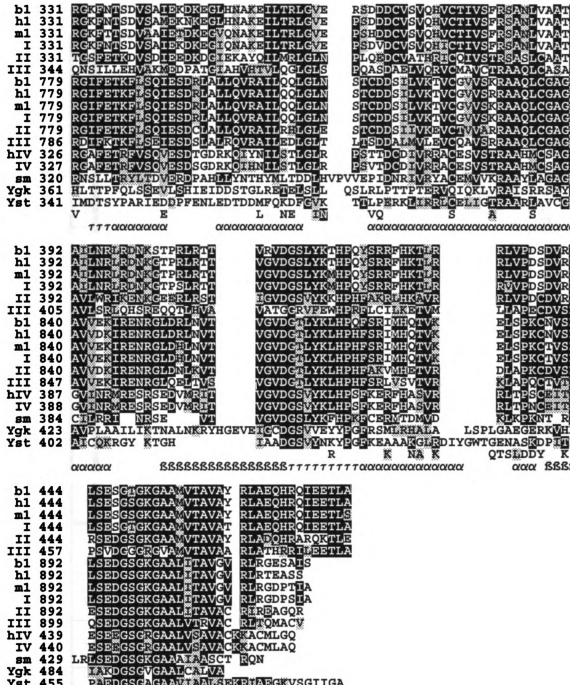

Figure 19. (Cont.)

Table 2. Quantitative Comparison of Hexokinase and Glucokinase Sequences

	Nb1	Nh1	Nm1	NI	NII	NIII	Cb1	Ch1	Cm1	CI	CII	CIII	hIV	IV	sm	Ygk
Nh1	91/3															
Nm1	89/3	95/2														
NI	89/4	94/4	96/2													
NII	63/14	68/13	68/14	67/13												
NIII	37/15	40/15	40/15	39/16	44/14											
Cb1	45/16	48/16	46/17	46/17	52/14	37/14										
Ch1	45/16	47/16	46/17	46/17	53/14	38/14	90/4									
Cm1	45/17	48/16	47/18	46/18	53/14	38/14	93/3	89/6								
CI	46/17	48/16	47/17	47/17	54/14	38/14	93/3	89/5	97/2							
CII	48/17	50/16	49/17	49/17	55/14	40/13	75/10	77/10	76/10	76/11						
CIII	44/15	47/14	46/15	45/15	48/15	40/14	60/11	61/11	62/11	62/11	66/9					
hIV	45/16	46/16	46/16	46/17	51/15	38/15	49/15	51/14	49/16	50/16	54/13	49/15				
IV	45/17	46/17	47/18	46/18	51/15	38/15	49/16	51/14	49/16	50/16	53/14	49/15	93/2			
sm	36/13	39/13	39/13	38/14	41/14	31/13	42/11	42/12	43/11	44/11	44/11	40/11	41/14	41/14		
Ygk	24/13	26/13	26/14	26/14	31/12	25/11	28/12	28/12	29/12	29/11	29/12	28/12	27/14	27/14	25/12	
Yst	26/14	28/15	27/15	27/15	29/15	23/14	28/15	28/15	28/15	28/15	29/16	27/13	28/15	28/16	27/13	31/10

Sequences are: the N-terminal halves of hexokinases from bovine type I (Nb1), human type I (Nh1), mouse type I (Nm1), rat type I (NI), rat type II (NII), rat type III (NIII) and the respective C-terminal halves (Cb1, Ch1, Cm1, CI, CII, CIII). Next are the sequences of human liver glucokinase (hIV), rat liver glucokinase (IV), schistosoma mansoni hexokinase (sm), yeast glucokinase (Ygk), yeast hexokinase A (Yst). Numbers given are: % identical/% conservative changes (91/3 = 91% of the amino acid residues in the two sequences compared are identical and an additional 3% are conservative changes).

(not inhibited by glucose-6-phosphate) diverged from the ancestral 50 kDa enzyme (also not inhibited by glucose-6-phosphate) **before** the initial gene duplication and fusion event giving rise to the 100 kDa isozymes (types I-III, which are inhibited by glucose-6-phosphate). Therefore, the type IV isozyme would be expected to be more similar to the other 50 kDa isozymes than to the mammalian types I-III. The other possibility proposed by Ureta (19) was that the type IV isozyme arose **after** the duplication and fusion event which gave rise to the 100 kDa hexokinases. The type IV isozyme would then be a product of the subsequent *resplitting* of one of the genes to restore a 50 kDa form. In this case, the type IV isozyme would be expected to be more similar to the types I-III isozymes than to the other 50 kDa hexokinases. The results in Table 2 indicate the latter to be the case. Indeed, comparisons of type IV with the 100 kDa mammalian isozymes results in similarities where approximately 50% of the residues are identical as opposed to the 50 kDa yeast isozymes where only 27% are identical.

The alignment in Figure 19 demonstrates that the insertions and deletions that have previously been noted between the type I isozyme and yeast hexokinase, which are likely to impact on structure (chapter III, page 59), are also present in the other mammalian enzymes. Extensive conservation of sequence among the enzymes of Figure 19 make it reasonable to expect an overall conservation of structure in these enzymes (94,103-105). The secondary

structural features of yeast hexokinase are also indicated in Figure 19 below the sequences of the yeast hexokinases. As previously shown for the type I isozyme (Figure 11), most of the insertions or deletions evident in the sequence alignment are located near the ends of secondary structural features such as α-helices and ß-strands. This would be expected in the case of homologous proteins (94). Yeast glucokinase (14), however, appears to contain a region that is an exception. An insertion of 11 residues is present in yeast glucokinase which would occur between yeast hexokinase residues Thr-226 and Lys-227. This insertion is not present in any of the other enzymes. It appears this insertion would increase the length of a ß-strand located in the "hinge" region joining the two lobes of yeast hexokinase which seems certain to have a major impact on this region of the molecule (Figure 20).

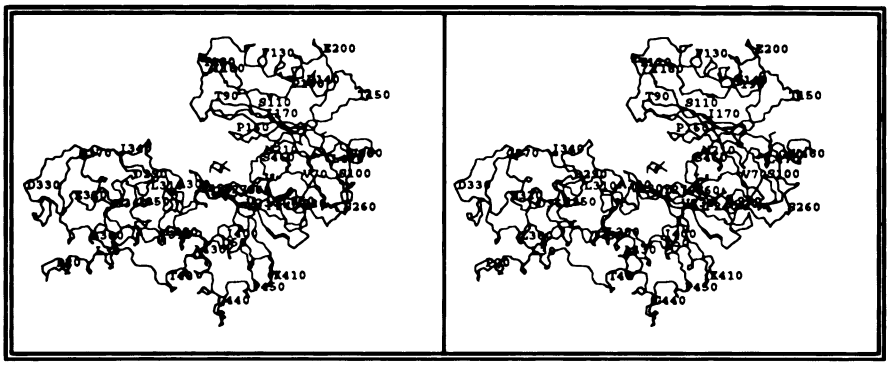

Figure 20. Stereo Image Showing Insertion in Yeast Glucokinase. Residues 226 and 227 are shown by thickened regions of the backbone. This is the location of an apparent 11 residue insertion in yeast glucokinase.

CHAPTER V

Glucose and ATP Binding Sites

In this chapter, a closer look is taken at the residues involved in the binding of glucose and the conservation of these residues in the known hexokinase sequences using the sequence alignment in the previous chapter (Figure 19). The chapter concludes with discussion of the region (and the residues therein) proposed to be involved in the binding of the other substrate, $Mg^{+2}ATP$. (Note: This chapter contains stereo images of the yeast hexokinase isozymes, actin, and glycerol kinase. In the cases of actin and glycerol kinase, amino acid residues determined in the crystal structures agree with those deduced from the respective cDNA sequences. However, the crystal structures for the yeast hexokinase isozymes were determined prior to the availability of the amino acid sequences. As a result many of the side chains were misidentified. Therefore, in the stereo images of the yeast hexokinase isozymes, if the side chains do not match the amino acid label, the amino acid label is correct.)

The Glucose Binding Site

Figure 19 shows many regions where the sequences of the enzymes are well conserved. Not surprisingly, some of these regions comprise the glucose binding site. Residues, determined by Harrison (66), which appear to hydrogen bond with the hydroxyls of glucose in the open conformation of yeast hexokinase include the side chains of Asn-210, Asn-237, Glu-269, and Glu-302 as well as the carbonyl oxygens from the peptide bonds of residues Gly-235 and Val-236 (Figure 21). In the closed conformation (Figure 22), yeast

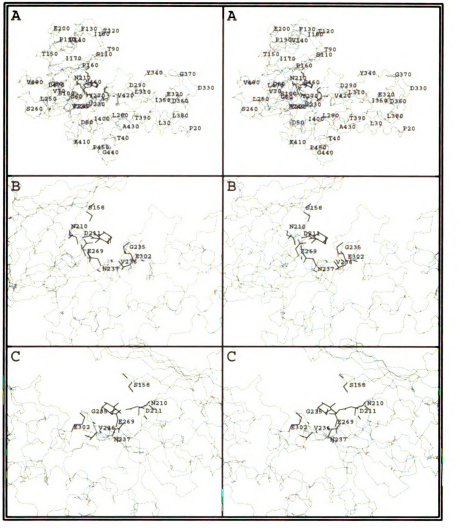

Figure 21. Stereo Images of Residues Involved in the Binding of Glucose in the "Open" Conformation of Yeast Hexokinase. The side chains as well as carbonyl oxygen bonds of residues involved in the binding of glucose to yeast hexokinase are darkened. **A:** "Open" conformation of hexokinase with bound glucose. **B: and C:** Alternate close up views. Residues utilized to bind glucose in the "open" conformation are Asn-210, Gly-235, Val-236, Asn-237, Glu-269, and Glu-302.

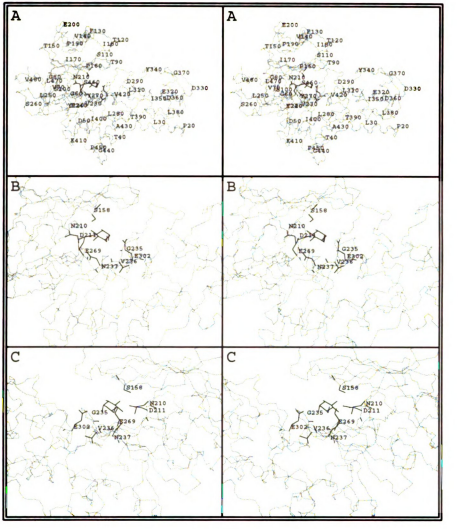

Figure 22. Stereo Images of Residues Involved in the Binding of Glucose in the "Closed" Conformation of Yeast Hexokinase. The side chains as well as carbonyl oxygen bonds of residues involved in the binding of glucose to yeast hexokinase are darkened. **A:** "Closed" conformation of hexokinase with bound glucose. **B: and C:** Alternate close up views. Residues utilized to bind glucose in the "closed" conformation are Ser-158, Asp-211, Glu-269, and Glu-302.

hexokinase has fewer contacts with glucose than the open conformation. Side chains of residues that hydrogen bond with glucose in the closed conformation include: Asp-211, Glu-302, and Glu-269. The carbonyl oxygen from the peptide bond corresponding to residue Ser-158 also makes contact with glucose in the closed conformation. All of the residues that have side chains that participate in the binding of glucose in either the closed conformation or the open conformation are totally conserved in all the sequences of Figure 19.

Two of the residues that hydrogen bond *via* the carbonyl oxygens of their peptide bonds, Ser-158 (closed conformation) and Gly-235 (open conformation), are also totally conserved in the sequences in Figure 19. The conservation of Gly-235 is not surprising due to its juxtaposition to the terminal phosphate of ATP (discussed below). Conservation of Ser-158 may also be expected because the side chain of this residue is located in the cleft, above glucose, and is juxtaposed to the side chains of Asn-210 and Asp-211 (see Figure 22, part C). Both of the later residues are utilized in binding glucose (discussed above) and it appears that they are also interacting with Ser-158. The total lack of conservation of Val-236 can also be explained. The side chain of this residue is oriented such that it is not pointed into the cleft, but is pointed in the opposite direction where it is buried in the large lobe. Since its side chain is not in the cleft and its "essential"

feature is its carbonyl oxygen (from the peptide bond), minimal changes to this residue may not impact on the enzyme's ability to bind glucose. Indeed, the corresponding position in the other hexokinases contains serine, threonine, or cysteine residues, and comparison of the sizes of these side chains relative to valine shows that they all would fit reasonably well in the region occupied by valine. Therefore, while the character (hydrophobic vs hydrophilic) of this residue has not been conserved, the dominant structural feature, size, appears to have been.

It is interesting to note that just as with the type I isozyme discussed in chapter III (page 56), the residues involved in the binding of glucose (as determined for yeast hexokinase) are totally conserved in the N-terminal halves of all the "low Km" isozymes presented in Figure 19 (as well as the C-terminal halves). Again, this is surprising since, based on the type I isozyme as precedent, only one half of these enzymes is expected to be catalytically competent.

The ATP binding site

The rest of this chapter covers the location of the ATP binding site and is divided into two parts. In the first section, the proposed ATP binding site is based on the common sequence characteristics of previously known nucleotide binding sites of various proteins. The second section is based on ATP binding proteins that have recently been found to be structurally similar to yeast hexokinase, although this was not evident from comparisons of the amino

acid sequences of these proteins. The stereo images in the first section correspond to the "open" conformation of yeast hexokinase. This conformation is used to depict features of the enzyme thought to be important in the initial stages of binding of ATP. In the second section, the "closed" conformation of yeast hexokinase is used. In this section, the features of the enzyme discussed are those that are apparent from structural comparisons with other ATP binding proteins that have nucleotides already bound in the crystals used to determine the respective structures. It should be noted that the predictions in the first section were made before the structural similarities to the proteins of the second section were known.

Prediction of the ATP Binding Site Based on Sequence

Unlike analyses carried out by Steitz and colleagues (55,56,58,59,66) on the binding of glucose to yeast hexokinase, the interactions that occur in the binding of ATP have not been defined. However, Steitz *et al.* (59) were able to determine the crystal structure of a complex of AMP with yeast hexokinase. They rationalized that the AMP molecule is binding to the same site as ATP due to the fact that AMP is a competitive inhibitor of ATP (106,107), although a poor one as indicated by the concentrations used (5-20 mM) in kinetic studies. Furthermore, they were unable to locate the phosphate group of AMP due to localized disorder, and subsequently modeled a triphosphate side chain using the position of a bound sulfate molecule to represent

the γ-phosphate of ATP. Consequently, the ATP site proposed by Steitz *et al.* (59) is not firmly established, certainly not by a crystallographic analysis of an ATP-enzyme complex. Additionally, the region where the ATP is suggested to bind involves three helical regions that form a shallow depression. One of these helices (yeast residues 346-352) is not well conserved throughout the alignments in Figure 19. It is not unreasonable to expect that a region providing such a crucial role would be well conserved. Therefore, an alternate site is proposed below.

The amino acid sequences of several enzymes that bind nucleotides have been determined along with their respective crystal structures (108 and ref. therein). Analyses have shown that most of the amino acids in these proteins are not conserved, and in some cases, although the nucleotide may be bound in identical positions, the amino acids participating in the binding may be different. In fact, in comparisons of liver alcohol dehydrogenase isolated from different species, the residues involved in binding the nucleotide substrate were found to vary at almost the same rate as surface residues.

Fortunately, in these structurally conserved regions, there are a few residues whose conservation appears to be essential, thus permitting the determination of diagnostic binding motifs. One such motif is defined by the amino acid sequence: Gly-X-Gly-X-X-(Gly/Ala) (109), with the amino acid corresponding to X being variable. This sequence motif

is commonly located in a region proceeding from a ß-strand, through a loop, and into an α-helix, with this motif occurring in the loop and the first few residues of the α-helix (108,110). The loop is generally very short. The invariant glycines are necessary to permit the close approach of the phosphate side chain allowing the positive dipole of the N-terminus of the α-helix to interact with the negative phosphates.

In yeast hexokinase (66), residues 459-464 are located in such a structure, and furthermore, this region is highly conserved in the hexokinases in Figure 19. This region, highlighted in Figure 23 (part A), is next to a ß-sheet (Figure 23, part B) that is composed of five strands: residues 80-87, 92-100, 103-111, 151-157, and 201-209. All of these strands, except the strand consisting of residues 103-111, are well conserved and are predominantly hydrophobic. The lack of conservation in the strand comprised of residues 103-111 is not surprising due to its location at the surface of the enzyme. Nevertheless, the predominantly hydrophobic ß-sheet should provide a suitable surface for the binding of the adenine base.

In support of the proposed role of the ß-sheet, a 50-residue peptide which includes part of the ß-sheet (residues 78-127 in yeast hexokinase, see Figure 23, part C) has been shown to bind adenine nucleotides (or an analog) (111). An indication that this hydrophobic region is interacting only with the adenine moiety of ATP is that the binding of

the nucleotide to this peptide is not affected by the chelation status (with Mg^{+2}) of the triphosphate side chain.

If the adenine ring is bound to the hydrophobic ß-sheet and the ribose and α-phosphate are interacting with the N-terminus of the α-helix, the phosphate side chain may be properly oriented into the cleft towards the 6-hydroxyl of glucose *via* hydrogen bonding to the conserved Thr-215 (Figure 24). Additionally, conserved acidic residues at positions 457 or 458 may aid in the proper orientation of the phosphate side chain through repulsive forces. The resulting orientation would place the γ-phosphate of the bound ATP in close proximity to the conserved Asp-211 (Figure 24). This acidic residue may function as a general base, facilitating the nucleophilic attack of the 6-hydroxyl of glucose on the terminal phosphate of ATP (55,112).

While residues 459-464, which comprise the Gly-X-Gly-X-X-(Gly/Ala) motif, are the only residues that appear to be in the proper structural orientation, this amino acid sequence occurs in one other region, residues 233-238. This region (Figure 25), which is located in the cleft, is comprised of two ß-strands separated by a looping region containing the motif (the last residues being part of the second ß-strand). The ß-strands are located in the large lobe (although they originate from the small lobe), directly across from the previously discussed region (located in the small lobe). While it may not play as prominent a role in the initial binding of the nucleotide, it does appear that

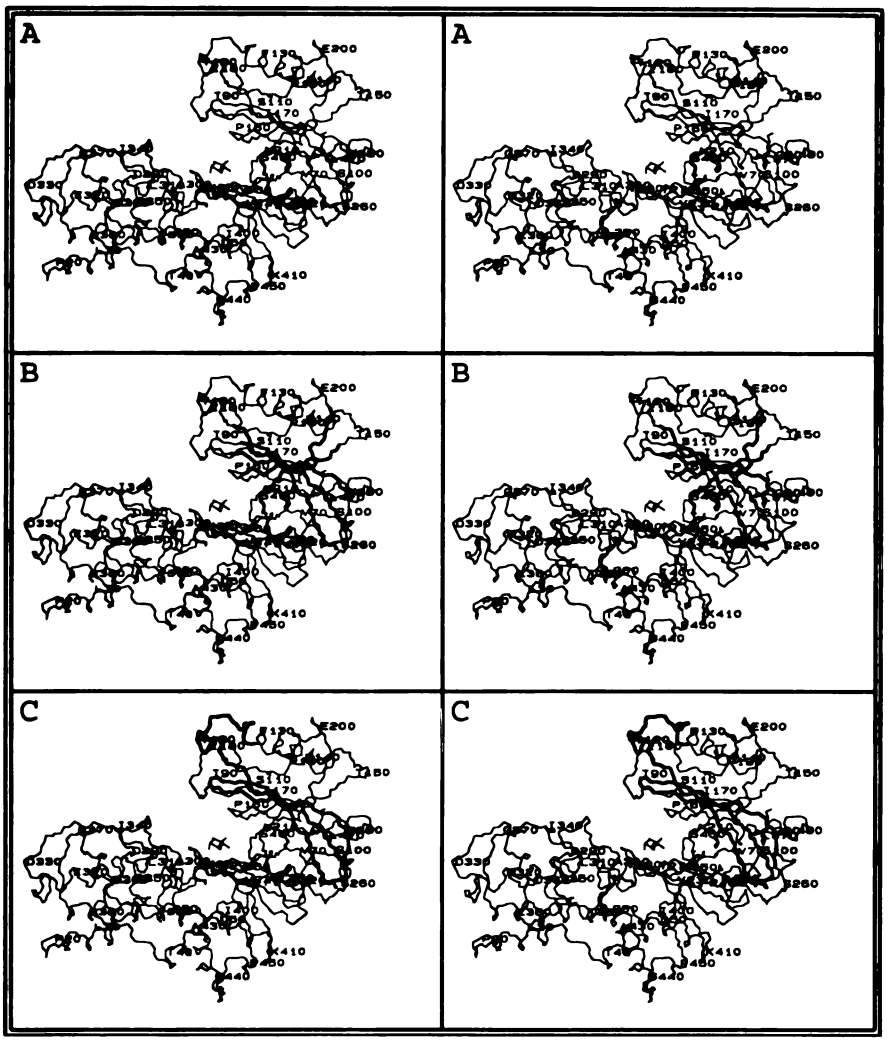

Figure 23. ATP Site Based on the Sequence Gly-X-Gly-X-X-(Gly/Ala). **A:** Darkened region corresponds to Gly-X-Gly-X-X-(Gly/Ala) sequence. **B:** A and ß-sheet of small lobe. **C:** Darkened region corresponds to residues 78-127. This peptide has been shown to bind ATP.

the glycines are in a good position to accommodate the triphosphate side chain upon closure of the cleft. The sequence in this region is also well conserved in the hexokinases compared in Figure 19.

It should be pointed out that Tamura *et al.* (113), using the ATP analog pyridoxal 5'-diphospho-5'adenosine (PLP-AMP), have reported the labeling of Lys-111 (Figure 26) in yeast hexokinase. They propose that this residue is involved in the binding of ATP *via* electrostatic interactions with the phosphodiester side chain. Lys-111 is part of the ß-strand located at the surface of the yeast enzyme corresponding to residues 103-111. In order to bring the terminal phosphate close to the 6-hydroxyl of the bound glucose, extensive movement of this ß-strand from the surface in the "open" conformation of yeast hexokinase, deep into the cleft, would have to take place. Due to the location of this strand and the conformational change that occurs upon binding glucose, further movement of this ß-strand into the cleft seems unlikely (the cleft is closing and this strand is still outside, see Figure 26). In fact, if the adenine moiety of PLP-AMP were binding at the site normally occupied by the adenine moiety of ATP, the labeling of Lys-111 by the reactive pyridoxal group appears to orient the phosphate side chain *away* from the cleft as opposed to into the cleft towards the bound glucose. Therefore, the modified phosphate side chain of the PLP-AMP analog appears to result in binding that does not accurately reflect the

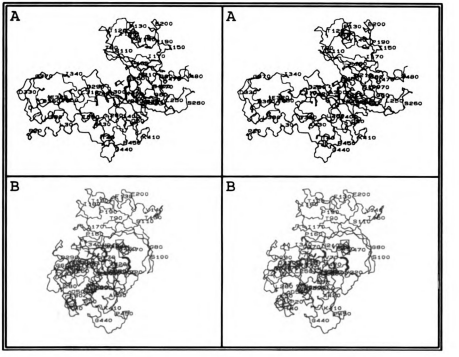

Figure 24. Residues Proposed to be Used in Orienting ATP into the Active Site. **A: and B:** Alternate views with darkened residues corresponding to Asp-211, Thr-215, and acidic residues at positions 457 and 458.

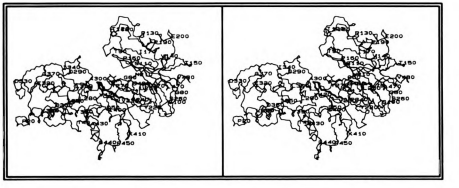

Figure 25. Location of the Additional Gly-X-Gly-X-X-(Gly/Ala) Sequence Purported to be Utilized in the Binding of ATP.

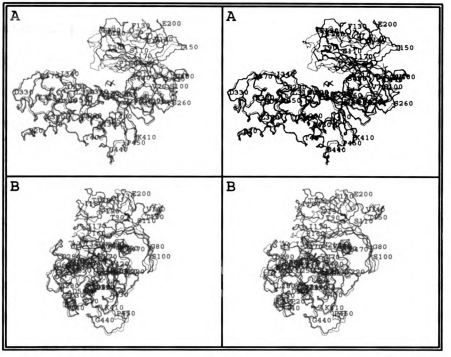

Figure 26. Location of Lys-111 Suggested, by Tamura *et al.* (113), to be Involved in the Binding of ATP. **A and B:** Alternate views of superimposed conformations of hexokinase. "Open" conformation = solid line, "closed" conformation = dotted line. The location of Lys-111 is shown by thicker lines.

orientation of the phosphate side chain of ATP.

ATP Binding Site Based on Structurally Similar Proteins

Standard pairwise alignment algorithms do not detect any significant similarities between yeast hexokinase, HSC70 (70 kDa bovine heat-shock cognate protein), and actin (114) amino acid sequences. This is not surprising due to the vastly different functions of these proteins, with actin being involved in the formation of cytoskeletal filaments and muscular contraction (reviewed in 115,116,117), heat shock proteins being involved in chaperoning functions and the refolding of denatured proteins (118-121) and yeast hexokinase phosphorylating glucose. All three proteins bind and hydrolyze ATP; nevertheless, it was still surprising to find that they have similar three dimensional structures in the region utilized in the binding of ATP. This is a situation not unlike the NAD binding domains of liver alcohol dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, and lactate dehydrogenase which also have similar three dimensional structures even though they lack extensive amino acid sequence homology (108 and ref. therein).

The HSP70-related proteins comprise a family of proteins in which amino acid sequence is highly conserved, with most of this conservation being in the N-terminal ATPase domain while the C-terminal substrate recognition domain is more variable (123). Initially Flaherty *et al.* (122) determined the three dimensional structure of the

N-terminal 44 kDa ATPase fragment of HSC70 and noted the similarity (123) to the tertiary structure of yeast hexokinase. The HSC70 ATPase fragment was to be crystallized with bound ADP. Although the crystals were grown in 1 mM Mg^{+2}ADP, upon building the model Flaherty *et al.* (122) determined that the actual nucleotide bound was ATP. This was verified by thin layer chromatography after redissolving the crystals. The source of the ATP was surmised to have been from the last step in the purification of the 44 kDa fragment: chromatographic elution from ATP-agarose. Hydrolysis of the bound ATP was inhibited by the high concentration of monovalent cation (1M NaCl) present during the crystallization. Subsequently, the crystals were adapted to low ionic strength conditions after which the bound nucleotide was determined to be ADP (122). The hydrolysis of the bound ATP to ADP in the crystals (under low ionic strength conditions) demonstrated that the site of binding was, in fact, an active site.

In the structural comparison of HSC70 to yeast hexokinase by Flaherty *et al.* (122), two different acidic residues were suggested to be candidates for the catalytic proton acceptor: a Glu residue which has Asp-211 at an equivalent position in yeast hexokinase; and an Asp which is located within the sequence Gly-Ile-Asp-Leu-Gly-Thr-Thr. A very similar sequence, Ala-(Ile/Leu)-Asp-Leu-Gly-Gly-(Thr/Ser), is found at positions 84-90 in yeast hexokinase; structurally, this is located in the same relative (to

HSC70) position in the yeast enzyme. This sequence is also highly conserved in the sequences in Figure 19.

Kabsch *et al.* (124) have determined the structure of the actin:DNase I complex with either ATP or ADP separately bound to the crystals. They noted the similarity between this nucleotide binding structure and an analogous structure present in yeast hexokinase. Specific residues involved in the binding of ATP were determined for the actin:DNase complex, but these were not discussed in terms of specific residues present in the yeast hexokinase structure. In a more recent publication, Flaherty *et al.* (123) discussed the specific interactions involved in the binding of ATP to HSC70, comparing this binding in a residue by residue fashion to the ATP binding site of actin. Subsequently, Bork *et al.* (114) examined the structural similarities between the common ATP binding core of actin, HSC70, and yeast hexokinase and related this binding to specific regions of the yeast hexokinase structure and sequence. With this information, the previously determined interactions of specific residues with ATP in either actin (124) or HSC70 (123) could now be related to the yeast hexokinase structure (66). Additionally, the crystal structure of E.coli glycerol kinase complexed with ADP was recently reported by Hurley *et al.* (125) and shown to have a nucleotide binding site structurally similar to yeast hexokinase. This structure was also used to identify residues involved in the binding of the nucleotide substrate. (In the stereo images to follow,

the 44 kDa ATPase fragment of HSC70 is not shown because complete coordinates were not available.)

Alignment of the crystallographic structures of HSC70, yeast hexokinase, actin, and glycerol kinase revealed five regions where the polypeptide backbones of the structures could be superimposed (Figure 28) and, upon further inspection, specific residues were found within these regions that were well conserved (see Table 3). A modified version of the alignment given by Bork *et al.* (114) is shown in Figure 27. The superimposed regions are (in terms of yeast hexokinase) PHOSPHATE 1: residues Tyr-82 to Leu-96, CONNECT 1: residues Ala-207 to Ser-219, PHOSPHATE 2: residues Lys-227 to Phe-240, ADENOSINE: residues Ile-414 to Leu-435, CONNECT 2: residues Asp-458 to Ala-469. Figure 19 shows that all five of these regions are well conserved in the aligned hexokinase and glucokinase sequences. These regions comprise all of the area in contact with the bound ATP (see Figure 28) except one additional loop which was suggested by Bork *et al.* (114) to be related to sugar binding. This is precisely the case as has been demonstrated by Schirch and Wilson (92) with labeling of this region in rat brain hexokinase (peptide III which corresponds to yeast residues 157-188) with a reactive glucose analogue.

In all four proteins, HSC70 (122), yeast hexokinase (66), actin (124), and glycerol kinase (125), there exists a common structural feature which is a deep cleft formed by two lobes. Each of the lobes is comprised of a ß-sheet which

```
                PHOSPHATE 1              CONNECT 1              PHOSPHATE 2
Yst     82 YLAIDLGGTNLRVVL  207 ALINDTVGTLIAS  227  KMGVIFGTG VNGAF
Actin    7 ALVCDNGSGLVKAGF  133 YVAIQAVLSLYAS  150  GIVLDSGDGVTHNVP
Gk       6 IVALDQGTTSSRAVV  241 GIAGDQQAALFGQ  260  MAKNTYGTG CFMLM

                    ADENOSINE                  CONNECT 2
Yst     414 IAADGSVNKYPGFKEAAAKGL  458 DG   SGAGAAVIAA
Actin   297 NVMSGGTTMYPGIADRMQKEI  334 ERKYSVWIGGSILAS
Gk      406 LRVDGGAVANNFLMQFQSDIL  437 EV   TALGAAYLAG
```

Figure 27. Sequences of Structurally Similar Regions in Yeast Hexokinase (Yst), Actin, and Glycerol Kinase (Gk). Modified version of the alignment originally proposed by Bork *et al.* (114).(In PHOSPHATE 2, actin appears to have a single amino acid insertion: Val-159. In CONNECT 2, actin appears to have another insertion: Lys-336 to Ser-338.)

Table 3. Structurally Equivalent Residues in HSC70, Yeast Hexokinase (Yst), Actin, and Glycerol Kinase (Gk).

HSC70	Yst	Actin	Gk
Asp-10	Asp-86	Asp-11	Asp-10
Gly-12	Gly-88	Gly-13	Gly-12
Cys-7	Arg-93	Lys-18	Arg-17
Glu-175	Asp-211	Gln-137	Asp-245
Asp-199	Ile-231	Asp-154	Thr-264
Gly-201	Gly-233	Gly-156	Gly-266
Gly-203	Gly-235	Gly-158	Gly-268
Gly-338	Gly-418	Gly-301	Gly-410
Gly-372	Gly-463	Gly-342	Gly-442

has helices on both sides. As mentioned previously, the three dimensional structures of actin, HSC70, and glycerol kinase were determined with crystals containing bound nucleotide. In all three cases the nucleotide is bound in the cleft formed by the two lobes with the phosphate side chain being bound by two ß-hairpins, one from each of the two lobes. These ß-hairpins, designated PHOSPHATE 1 and PHOSPHATE 2, are highlighted in the close up views in Figure 29.

The ATP binding core contains invariant glycines (Table 3, Figures 29 & 30) located in three separate loops (two of the loops are the ß-hairpins mentioned above) common to all four structures (HSC70, yeast hexokinase, actin, and glycerol kinase) that are necessary for the close approach of the phosphate side chain. Accordingly, the strict conservation in actin of Gly-301 (123,124), which is structurally equivalent to Gly-418 in yeast hexokinase (Figure 30), is due to the close approach of the α-phosphate of the bound ATP, while conservation of Gly-13 (124), structurally equivalent to Gly-88 in yeast hexokinase (Figure 29), is necessary due to the juxtaposition of the ß-phosphate, and conservation of Gly-158 (124), equivalent to Gly-235 in yeast hexokinase (Figure 29), is due to the γ-phosphate. Similarly, conservation of Gly-266 in glycerol kinase (125), equivalent to Gly-233 in yeast hexokinase (Figure 29), is due to the close proximity of the pyrophosphate moiety of the bound ADP. (Glycerol kinase

residue Gly-266 is located midway between the α- and ß-phosphates of ADP, more specifically, between one of the oxygens contributed by the ß-phosphate and two of the oxygens from the α-phosphate.) All of these glycine residues are strictly conserved in all of the hexokinase sequences in Figure 19 including both halves of the 100 kDa enzymes with the exception of the equivalent of yeast hexokinase residue Gly-233 in the N-terminal half of type III hexokinase.

The actin structure was determined with the Ca^{+2} chelate of ATP (124) (as opposed to the Mg^{+2} chelate of ATP used by hexokinase) and this metal ion appears to be able to interact with Asp-11 (Asp-86 in yeast hexokinase (114)), Gln-137 (Asp-211 in yeast hexokinase), and Asp-154 (Ile-231 in yeast hexokinase). All three residues, Asp-86, Asp-211, and Ile-231 are conserved in the sequences of Figure 19 except for the N-terminal half of the type III isozyme which has a glutamate residue at the position equivalent in yeast hexokinase to residue 86. Although the side chain of Ile-231 is not likely to participate in the binding of the divalent cation, it is located in the cleft and may be critical for function (see Figure 29, part A). Thus, it is interesting to note that in the protein structures that do not phosphorylate a substrate (actin and HSC70 and other heat shock related proteins), but hydrolyze ATP, the residue equivalent in position to yeast hexokinase residue Ile-231 is Asp-199 (HSC70) and Asp-154 (actin), whereas in the sugar kinases (114) this residue is an isoleucine (Figure 19),

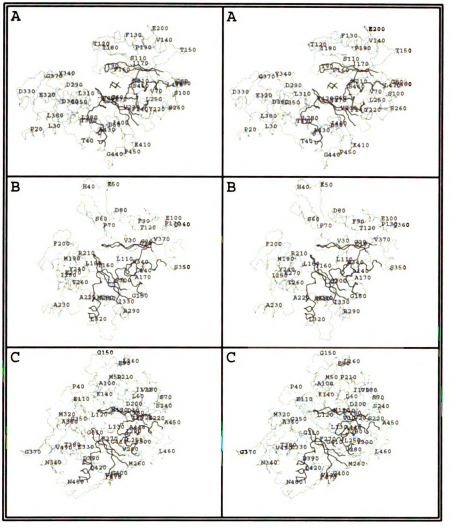

Figure 28. Stereo Images Showing Structurally Similar Regions in ATP Binding Proteins (see Figure 27). **A:** Yeast Hexokinase, **B:** Actin, **C:** Glycerol Kinase.

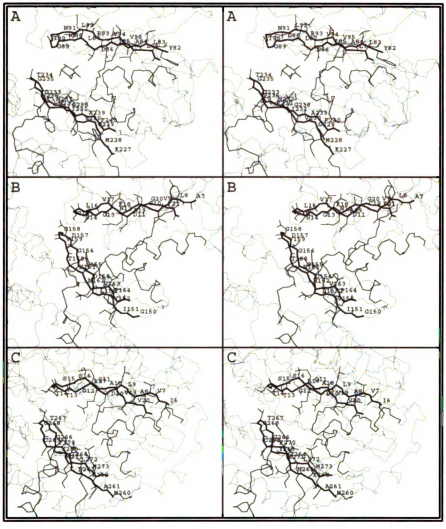

Figure 29. Stereo Images Showing Close Up Views of PHOSPHATE 1 and PHOSPHATE 2 (see text). Thick lines are ß-hairpins involved in the binding of the phosphate side chain of ATP. Thin lines are the rest of the regions that are structurally similar. **A:** Yeast Hexokinase, **B:** Actin, **C:** Glycerol kinase.

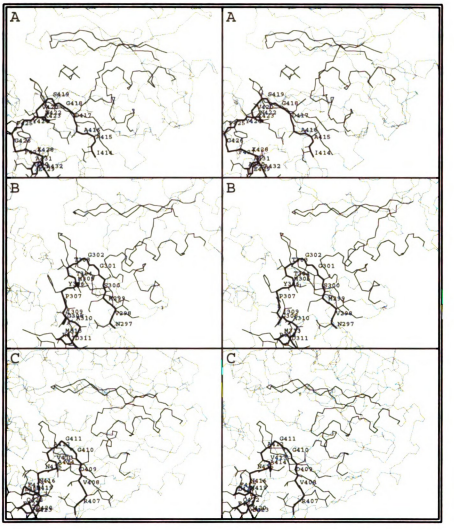

Figure 30. Stereo Images Highlighting Structurally Similar Region (ADENOSINE) Utilized in Binding the Adenine Base of ATP. Thick lines represent region used to bind adenine moiety. Thin lines are the rest of the regions that are structurally similar. **A:** Yeast hexokinase, **B:** Actin, **C:** Glycerol kinase.

serine (fucokinase, xylulokinase), or threonine (gluconokinase, glycerol kinase). Due to this residue's location, it is certain to have a major impact on the environment of the terminal phosphate of the bound ATP.

In the actin structure with ADP bound (124), the side chain of Lys-18 forms hydrogen bonds with oxygens from both the α- and ß-phosphates. This is also the case with the bound ADP in the glycerol kinase structure (125). In yeast hexokinase the analogous residue would be Arg-93 (114) which is also conserved in all the sequences in Figure 19. The side chain of yeast hexokinase residue Asp-211 has previously been shown to extend into the cleft (Figure 21). Although Asp-86 and Arg-93 were not correctly identified in the crystal structures of the yeast isozymes, sufficient coordinates were available to determine the direction in which the side chains were extended. Both residues are in ß-strands and, as Figure 29 (part A) shows, both side chains are oriented into the cleft as opposed to the opposite direction were they would have been buried in the small lobe. Therefore, it is very likely that these residues serve the same function in the hexokinases as in actin and glycerol kinase.

In the actin structure (123) the adenine base fits into a hydrophobic pocket formed by part of an α-helix (Arg-210 to Glu-214; no equivalent in the yeast hexokinase structure), a 3_{10}-helix containing residues Gly-302 to Tyr-306, and by Lys-336 (see Figure 31). In glycerol kinase, the

adenine base of ADP is also in a hydrophobic pocket that is made up of an apparent 3_{10}-helix, residues Gly-411 to Asn-415 (125). Using the modified alignment, the yeast hexokinase equivalent of the 3_{10}-helix would be residues Ser-419 to Tyr-424 (ADENOSINE (114)) contributed by the large lobe on one side of the cleft. Figure 31, part A, shows that although the region corresponding to the 3_{10}-helix was not determined in yeast hexokinase to be within the limits of a "standard" α-helix, it does appear to be helical in nature and if this region is of the tighter 3_{10}-helix variety, this would have been missed by the secondary structural prediction algorithm of MOSAIC.

It is interesting to note that the region utilized in binding the adenine base (ADENOSINE) is located in a crevice created at the carboxy ends of two parallel ß-strands. This feature of open α/ß sheet structures (α-helices on both sides of a ß-sheet) of α/ß proteins (126) has proven to be a highly accurate predictive element in the determination of active site regions (127). It occurs in a ß-sheet where two adjacent parallel ß-strands have, at the carboxy end, connections that in one strand go above the ß-sheet (usually to an α-helix) and in the other strand go below the ß-sheet (again, usually to an α-helix). The looping regions originating from the carboxy ends of these ß-strands form a crevice where active site residues are located. (This type of structure was first noticed by Branden (127) when he compared the structures of 20 different α/ß proteins

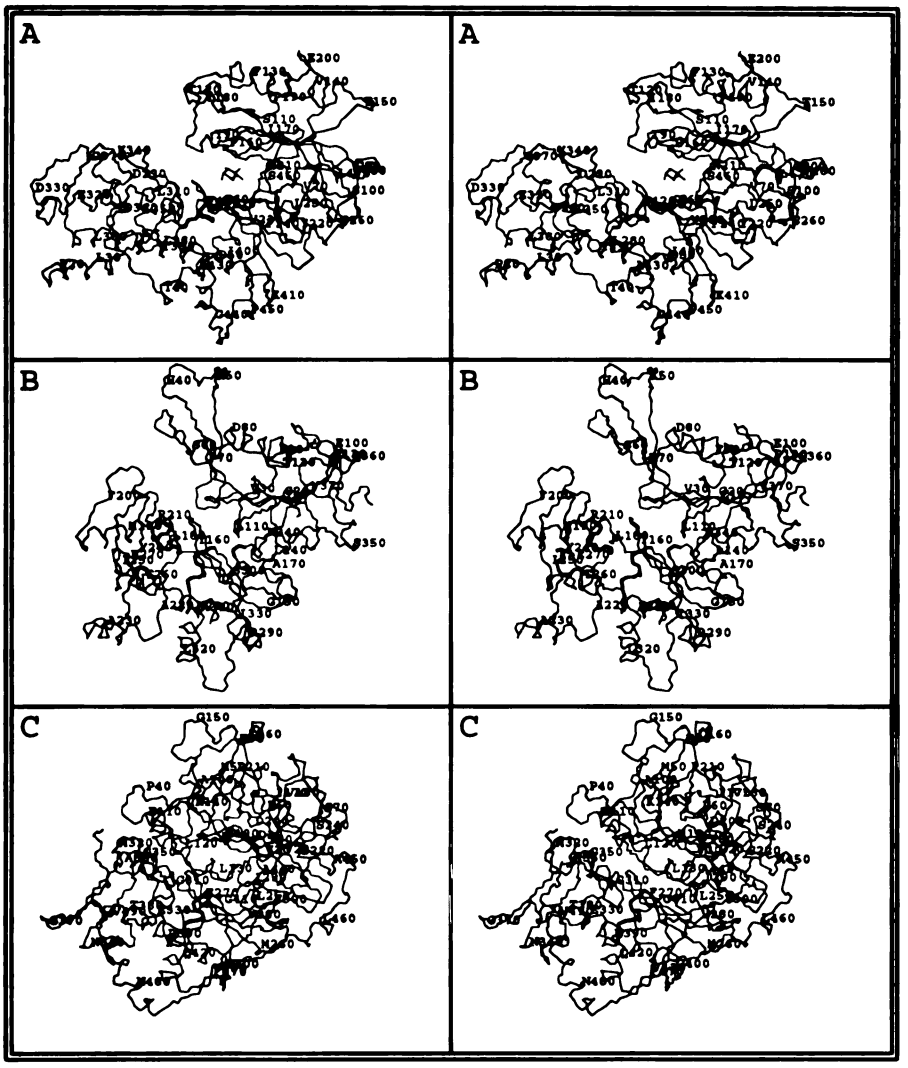

Figure 31. Stereo Images Showing Adenine Base Binding Regions of Actin and Glycerol Kinase that are Structurally Similar to Yeast Hexokinase. Darkened regions correspond to **A:** Yeast hexokinase Ser-419 to Tyr-424, **B:** Actin Gly-302 to Tyr-306 and Lys-336, **C:** Glycerol kinase Gly-411 to Asn-415.

containing open α/ß sheet structures and initially determined that functional residues were located in the loop region which connects the carboxy end of a ß-strand with the amino end of the following α-helix.) In fact, the yeast hexokinase structure contains two of these crevices (Figure 32), one in each of the two ß-sheets. One crevice occurs in the ß-sheet of the large lobe. This region, previously discussed, is utilized in binding the adenine base. The other crevice, located in the ß-sheet of the small lobe, contributes residues utilized in binding glucose: Ser-158 and Asn-210 (66).

Located in the hinge region are two helices that come into contact with one another (CONNECT 1 and CONNECT 2) (66) (see Figure 33). This region appears to involve another strictly conserved glycine, Gly-463, on one helix and the well conserved residues on the other helix from Asp-211 to Ala-218. Bork *et al.* (114) propose that this region, which contains this helix-helix contact, appears to be an interdomain hinge due to the fact that in comparisons between the "closed" actin and HSC70 structures with the "open" yeast hexokinase structure this region is maintained even though nearby regions undergo considerable movements.

It is interesting to note that in one of the helices involved in the interdomain hinge in the actin structure, CONNECT 2, there appears to be an insertion of approximately one turn (Lys-336 to Ser-338).

In the analysis of the structures of actin and the

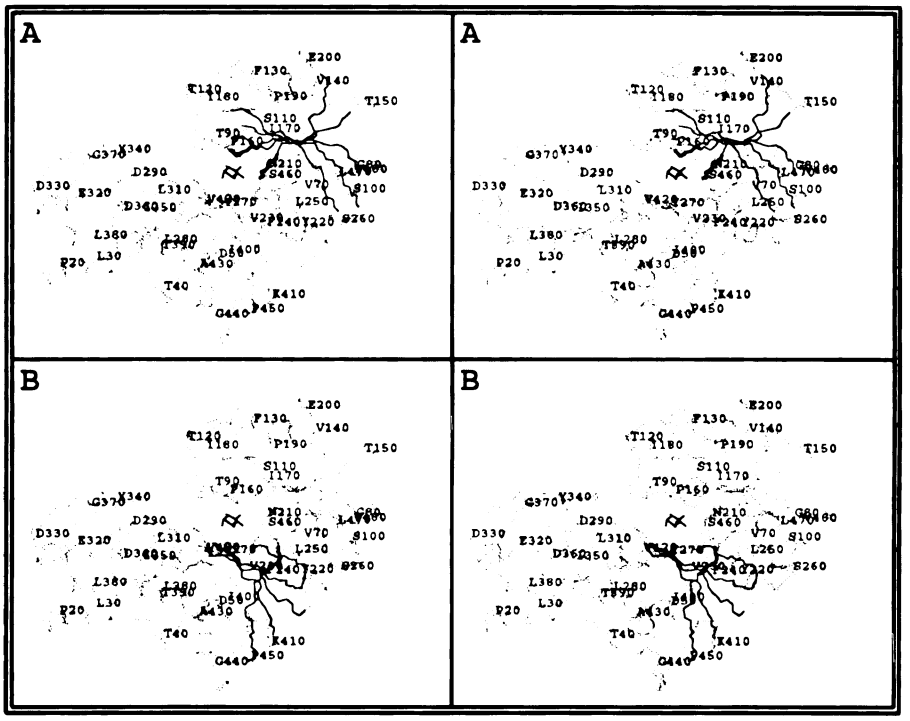

Figure 32. Crevices in ß-sheets in Yeast Hexokinase that Contribute Active Site Residues. Thin lines denote the ß-sheets and thickened lines denote the region of the crevice where active site residues are located (see text). **A:** ß-sheet of small lobe, **B:** ß-sheet of large lobe.

HSC70 proteins, Flaherty *et al.* (123) noted that the structure of HSC70 was sufficiently refined that some of the water molecules could be located. One such water molecule, referred to as Wat 546, is oriented such that its oxygen atom is situated 3.5 Å from the terminal phosphate of ATP. The O3' γ-phosphate bond of the bound ATP is aligned with a line from Wat 546's oxygen to the γ-phosphate. (The O3' oxygen is the oxygen in the phosphodiester bond "linking" the ß and γ phosphorus atoms). They suggest, therefore, that this water molecule is a good candidate for an in-line attack on the γ-phosphate of ATP. Additionally, they noted that in the yeast hexokinase crystal structure containing N-*o*-toluolylglucosamine, the position of the 6-hydroxyl of the bound glucose molecule approximates that of Wat 546. An in-line attack of the 6-hydroxyl of glucose on the γ-phosphate of ATP would be consistent with stereochemical studies of yeast hexokinase which show the reaction proceeds with inversion of configuration at phosphorus (128,129).

The previous two sections have discussed the ATP binding site, based in the first section on predictions derived from proteins of similar sequence, and based in the second section on comparisons to proteins that utilize structurally equivalent regions to bind the nucleotide substrate. Both sections have pointed out the importance of conserved glycine residues in the ATP binding site that are necessary for the binding (close approach) of the phosphate side chain. On the other hand, the region used in binding

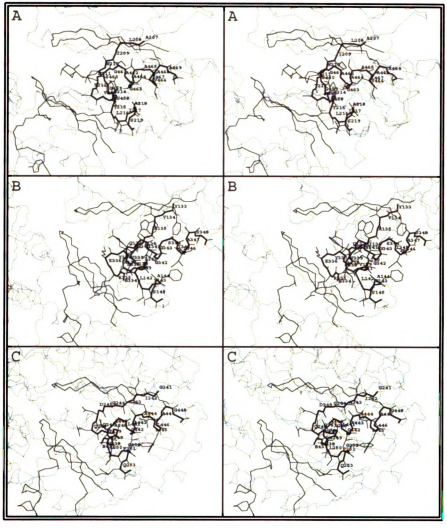

Figure 33. Stereo Images Depicting the Interdomain Hinge. Thick lines correspond to CONNECT 1 and CONNECT 2 which are helices proposed by Bork *et al.* (114) to form an interdomain hinge. Thin lines are the rest of the regions that are structurally similar. **A:** Yeast hexokinase, **B:** Actin, **C:** Glycerol kinase.

the adenine moiety may not be as clear. In the first section, the adenine moiety is predicted to be bound by the hydrophobic ß-sheet of the small lobe. This was supported by studies showing that the binding of ATP to a peptide containing a portion of this ß-sheet, and equivalent to yeast hexokinase residues 78-127 (Figure 23, part C), was independent of the chelation status (by Mg^{++}) of the phosphate side chain of ATP. Therefore, the region was surmised to be binding the adenine moiety of ATP. In the second section, the importance of the 3_{10}-helix (yeast hexokinase residues 419-424) in the binding of the adenine moiety was discussed. Additionally, based on the actin structure, analogous residues that appear to interact with the divalent cation in yeast hexokinase were determined to be Asp-86 and Asp-211. Only one of these residues is present in the peptide containing yeast hexokinase residues 78-127. Both residues may be necessary to bind the divalent cation. This may explain why binding of ATP by this peptide was independent of the chelation status of ATP. Nevertheless, this does not eliminate the ß-sheet of the small lobe from involvement in the binding of the adenine moiety of ATP. The adenine moiety could be initially bound by the hydrophobic surface of the ß-sheet, after which extensive movement of this region would occur with the result being that the adenine moiety would be "clamped" into place between the ß-sheet and the 3_{10}-helix upon closure of the cleft. Therefore, the hydrophobic ß-sheet of the small lobe would

be important in the initial binding of ATP while the 3_{10}-helix would become important upon closure of the cleft. This is similar to the binding of glucose in yeast hexokinase where the majority of the residues interacting with glucose in the "open" conformation are from the large lobe, whereas residues from both lobes are interacting with glucose in the "closed" conformation.

This chapter has described the glucose binding site and residues proposed to be involved in the binding of $Mg^{+2}ATP$. It is readily apparent from Figure 19 that the residues involved in binding glucose as well as most of those proposed to be involved in binding ATP are well conserved in both halves of the "low K_m" mammalian isozymes. The functional significance of these residues in the presumably noncatalytic N-terminal halves of the mammalian hexokinases remains unknown. A possible reason for the conservation of these residues might be that the N-terminal half is catalytically active only under specific conditions, such as when the enzyme is bound to mitochondria (type I hexokinase). Through the use of site directed mutagenesis, these residues, and the role they play, may be further investigated, an undertaking already occurring in this laboratory.

CHAPTER VI

Heterologous Expression of Type I Hexokinase

This chapter covers preliminary results on the bacterial expression of type I hexokinase using the plasmid pIN-III *ompA* (130,131). Parameters such as media, temperature, and concentration of inducer were varied in order to achieve the highest yield (in terms of activity).

Background

In pIN-III *ompA* expression is under control of the *lac* promoter that can be induced by the gratuitous inducer IPTG (isopropylthiogalactoside). Additionally, this vector is designed such that the type I hexokinase protein produced will contain the signal peptide of the OmpA protein which targets the protein for secretion into the periplasmic space. Upon translocation across the cytoplasmic membrane the signal peptide is cleaved. The expressed protein accumulates in the periplasmic space where protease activity is greatly reduced (130,132) compared to intracellular expression. Additionally, the periplasmic space provides an oxidizing environment (133) (as opposed to the intracellular reducing environment (134)) which may help prevent the formation of insoluble aggregates of the expressed protein (reviewed in 135). After expression into the periplasmic space the expressed protein is isolated *via* osmotic shock (136). Expression into the periplasmic space should provide for a simplified isolation of the expressed protein.

Plasmid Constructs

Construction of the plasmids that were used to express type I hexokinase is given in the methods section. The four plasmids designated pHB4, pM1-7, pXN1, and pNB6 are described below.

Plasmid pHB4 contains the entire coding region of type I hexokinase with an additional sequence corresponding to the multiple cloning site of pIN-III *ompA* located at the 5′ end of the cDNA insert. Upon induction, the expressed hexokinase protein will have additional amino acids "tacked" on the N-terminus even after the signal peptide has been cleaved (during translocation into the periplasmic space). pHB4 was used in the initial experiments to determine whether or not the expressed type I hexokinase would be catalytically active.

The next step in the constructions was deletion of the region corresponding to the multiple cloning site located between the 5′ end of the type I hexokinase cDNA and the region coding for the signal sequence. The resulting plasmid, pM1-7, codes for a protein that, after translocation into the periplasmic space and cleavage of the signal peptide, should be full length type I hexokinase beginning with the native N-terminal starting Met.

Sequence comparisons shown earlier have clearly established the mammalian hexokinases as being comprised of two similar halves. Plasmids pXN1 and pNB6 were constructed for expression of the N-terminal and the C-terminal halves,

respectively, of the type I isozyme.

Expression Results

Replica nitrocellulose filters containing *E.coli.* (strain JA221) harboring either pIN-III *ompA* (negative control) or pHB4 (should produce rat brain hexokinase) were induced with IPTG (2mM) and grown overnight at room temperature. Colonies on the filters were osmotically shocked (to release the contents of the periplasmic space) and the filters were then screened with affinity purified polyclonal antibodies (rabbit) raised against type I hexokinase. Colonies harboring pHB4 were immunoreactive to the anti-hexokinase antibodies while those harboring only the vector pIN-III *ompA* were not. Additionally, replica filters were incubated (after osmotic shock) in hexokinase activity stain. The filter from colonies harboring pHB4 developed positive signals much faster and more intense than the filter derived from colonies harboring only pIN-III *ompA*. Therefore, polyclonal antibodies indicated hexokinase was being produced in cells harboring pHB4 and activity staining indicated the enzyme was catalytically active.

Initial experiments aimed at expressing type I hexokinase in culture (strain JA221), using pM1-7, did not result in the detection of any significant hexokinase activity (after osmotic shock) relative to the negative control, pIN-III *ompA*. These experiments were carried out at 37°C with induction by 2 mM IPTG. The growth temperature was reduced to room temperature (137,138) and the levels

of IPTG used to induce expression were varied as well as the media (L broth *vs.* TB broth)(138). The results are shown in Table 4.

Table 4. Expression of Type I Hexokinase

Clone	Inducer conc.	mU/ml* culture L broth	TB broth
pM1-7	2.0 mM IPTG	8	7
pM1-7	0.20 mM IPTG	8	7
pM1-7	0.020 mM IPTG	8	41
pM1-7	0.002 mM IPTG	38	76
pM1-7	0 mM IPTG	36	89
pIN-III ompA	0.002 mM IPTG	8	1
pIN-III ompA	0 mM IPTG	8	1

* Values are expressed as milliunits/ml of culture. 1 unit is equivalent to the reduction of 1 umole $NADP^+$/min.

The results in Table 4 indicate that the highest levels of activity were achieved using TB broth as media with no added IPTG. Yeast extract, a component of both L broth (yeast extract @ 5 g/l) and TB broth (yeast extract @ 24 g/l) apparently contains an activator of the lac operon and is therefore able to induce expression (135).

The time course of expression (Figure 34) of type I hexokinase in *E. coli* (strain JA221) was determined using pM1-7 in TB broth with no added inducer (IPTG). It is apparent that maximum levels of expression are reached shortly after the culture reaches saturation.

The two "halves" of rat brain hexokinase were expressed at room temperature using TB broth and varying the level of inducer - IPTG. The results are listed in Table 5.

Surprisingly, the N-terminal half (pXN1) appears to possess some activity, although this is not certain since the protein was not purified. As expected the C-terminal half (pNB6) appears to be catalytically active although maximum activity levels appear to be at an IPTG concentration of 2mM as opposed to the full length expressed enzyme (pM1-7) where maximum activity levels occurred with no addition of IPTG.

This chapter has described preliminary results of expression of rat brain hexokinase in *E. coli*. The expressed enzyme appears to be catalytically active and based on a specific activity of 60 u/mg for the enzyme isolated from rat brains, the amount of active enzyme (pM1-7) at maximum expression is calculated to be 1.8 mg/l (assuming all the activity measured is ascribable to the expressed enzyme). Unfortunately, it appears that the expressed enzyme is unable to bind mitochondria (data not shown). Whether this is because the signal peptide was not cleaved from the enzyme, or cleavage of this peptide results in a charged N-terminus (which can not insert into the mitochondrial membrane as is required for binding) has not been determined.

The expression levels of type I hexokinase (and the N- and C-terminal halves) were maintained for well over a year. Unfortunately, in recent experiments the expression levels of all clones (as determined by activity measurements) have dropped to negligible levels for reasons which, regrettably, remain undetermined.

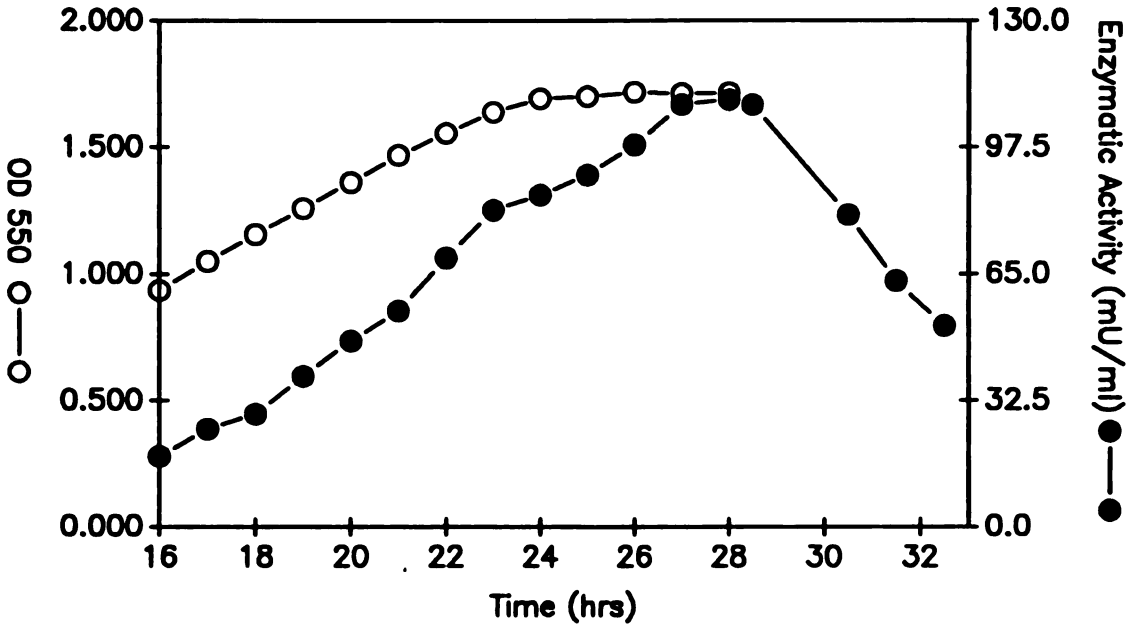

Figure 34. Time Course for Heterologous Expression of Type I Hexokinase. 500 mls of T.B. broth were inoculated with 5 mls of culture (in T.B. broth, previously grown to saturation at 37°C). The resultant culture was grown at room temperature with aliquots taken at 1 hour intervals for analysis. Total glucose phosphorylating activity (•) isolated from the periplasmic space and optical density of the culture at 550 nm (∘) were determined for each time point.

Table 5. Expression of N- and C-terminal Halves of Type I Hexokinase

Clone	Inducer conc.	mU/ml culture
pXN1	2.0 mM IPTG	3
pXN1	0.20 mM IPTG	1
pXN1	0.020 mM IPTG	13
pXN1	0.002 mM IPTG	16
pXN1	0 mM IPTG	26
pNB6	2.0 mM IPTG	206
pNB6	0.20 mM IPTG	202
pNB6	0.020 mM IPTG	61
pNB6	0.002 mM IPTG	25
pNB6	0 mM IPTG	17
pIN-III ompA	0.002 mM IPTG	1
pIN-III ompA	0 mM IPTG	1

CHAPTER VII

Future Research

The cDNA's coding for types I and III hexokinases make possible mutagenesis experiments aimed at investigating the structure to function relationships in the hexokinases. The importance of residues involved in binding glucose as well as those proposed to be involved in binding ATP, discussed in chapter V, can be explored with mutagenesis of the regions of interest. Specifically, the importance of the hydrophobic ß-sheet of the small lobe and the 3_{10}-helix (yeast hexokinase residues 419-424) in the binding of the adenine moiety of ATP could be investigated with site-directed mutagenesis. Mutagenesis of hydrophobic portions of the ß-sheet should prevent binding of ATP and catalysis, whereas, mutagenesis of the 3_{10}-helix should not prevent binding of ATP since the adenine moiety, and hence ATP, can still initially bind. The enzyme should still be catalytically inactive since the conformational changes necessary for catalysis cannot occur because the adenine moiety cannot be properly "clamped" between the ß-sheet and the mutated 3_{10}-helix.

In the sequence alignment in Figure 19, there are 50 residues that are identical and an additional 43 residues that are conserved in all of the sequences (including both halves of the 100 kDa isozymes). These residues are highlighted in Figure 35. It is not surprising that these residues are located almost exclusively in either the cleft or are buried in the enzyme. In the close up views of Figure 35 (parts C and D), two strictly conserved residues (in

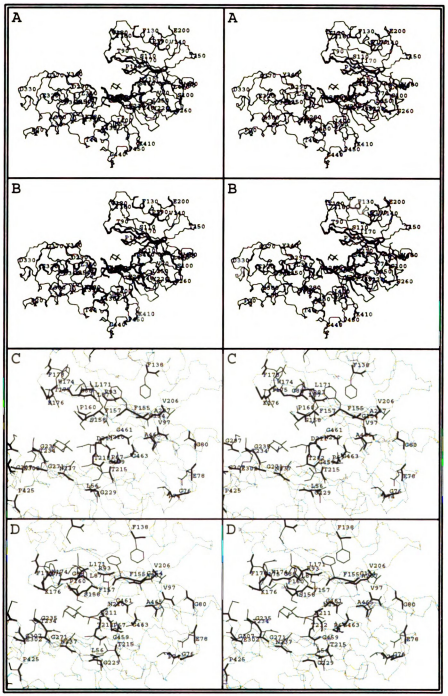

Figure 35. Conserved Residues in Hexokinases. Darkened regions are (**A:**) identical, or (**B:**) identical + conserved residues in all the sequences of Figure 19. Close up views of the region containing Lys-176 and Thr-212 using the (**C:**) "open" or (**D:**) "closed" conformation of yeast hexokinase.

addition to those previously suggested to be important in the binding of substrates) appear to be reasonable candidates for site directed mutagenesis. They are Thr-212 which is located deep in the cleft with its side chain oriented towards glucose, and Lys-176 which is located in the small lobe at the "lip" of the cleft. Figure 35 shows that Thr-212 is in a position to affect the interactions between the conserved residues Asn-210, Asp-211, Asn-237 and glucose. Similarly, the side chain of Lys-176 is oriented into the cleft directly above the bound glucose and seems certain to affect the bound glucose.

As discussed in chapter IV, comparisons between the "low K_m" isozymes demonstrate that the C-terminal halves are similar, as are the N-terminal halves. Figure 36 highlights the residues that are **identical** in comparisons of the amino acid sequences of either all the N-terminal halves of the "low K_m" isozymes (part A), or all the C-terminal halves of the "low K_m" isozymes (part B), or the catalytic "halves" (part C) (the C-terminal halves of the "low K_m" isozymes + all the 50 kDa enzymes of Figure 19) of all the sequences of Figure 19. As expected, in all three cases, the majority of the residues that are identical are located either in the cleft or are buried in the enzyme. There are, however, two helices located on the surface (Figure 36, part B, yeast residues 346-352 and 359-369) that are comprised of residues that are strictly conserved only in the C-terminal halves of the "low K_m" isozymes (which are the only "enzymes"

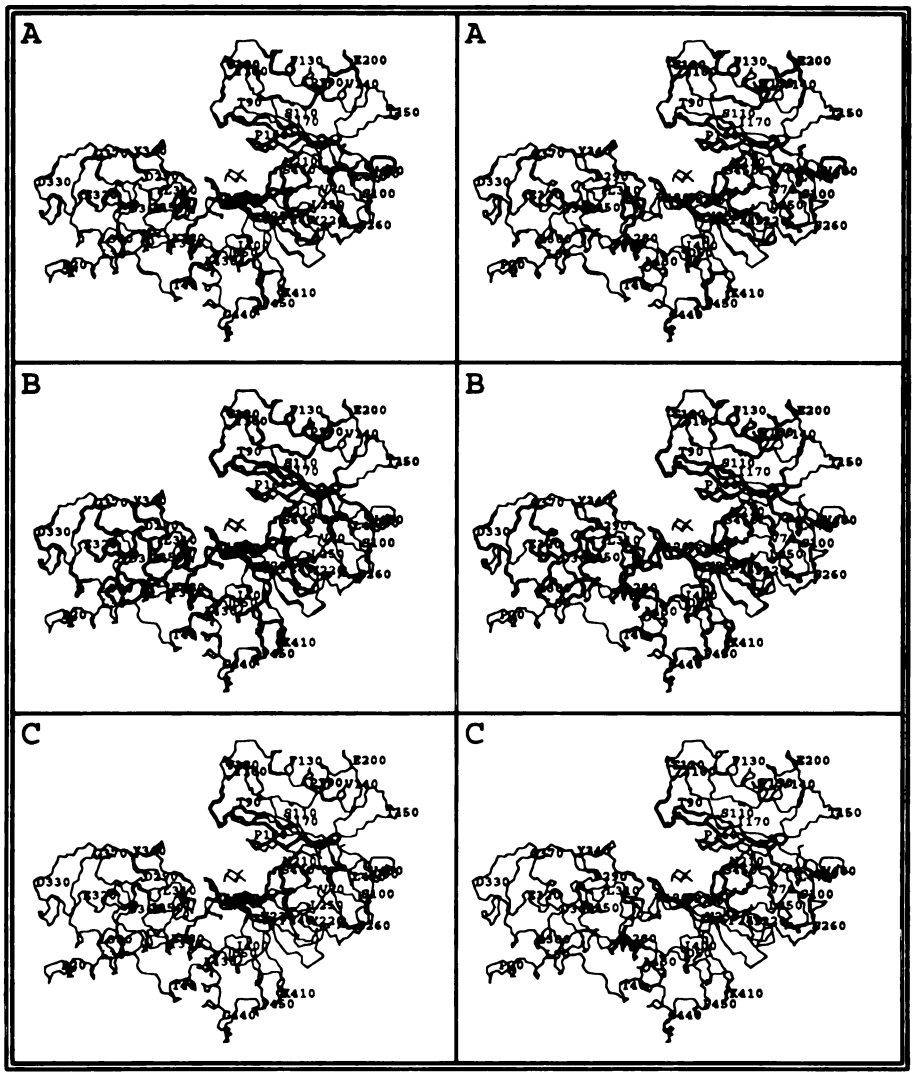

Figure 36. Stereo Images Highlighting Conserved Residues in Comparisons of Groups of Hexokinases. Darkened residues correspond to identical residues in the sequences of **A:** N-terminal halves or **B:** C-terminal halves of the "low K_m" isozymes, or **C:** catalytic "halves" of all the sequences in Figure 19.

inhibited by physiologically relevant levels of glucose-6-phosphate). Glucose-6-phosphate is a competitive inhibitor of ATP (139) and these conserved helices are close to the 3_{10}-helix implicated, in chapter V, in the binding of ATP. This leads to speculation as to whether this region is involved in the glucose-6-phosphate inhibitory site. Mutagenesis of these residues may reveal the reason for the strict conservation of so many of the residues in this region of the C-terminal halves of the "low K_m" isozymes.

In chapter V, the residues that are utilized in binding the substrates glucose and ATP were shown to be conserved in both halves of the mammalian hexokinases questioning whether or not the N-terminal halves possess catalytic activity. In the case of type I hexokinase, one possibility is that the N-terminal half is only active when the enzyme (type I) is bound to mitochondria. This could be investigated by mutating the C-terminal half so that the soluble (unbound) enzyme is no longer active (*via* a mutation in the C-terminal half) and then binding this mutant to mitochondria and assaying for activity. Any detected activity would be an indication that the N-terminal half possesses catalytic activity.

The N-terminal sequence is critical for binding type I hexokinase to mitochondria (18). This sequence also appears to be sufficient to effect the binding of other proteins to mitochondria (48). Type III hexokinase has been shown to be associated (weakly bound) with the nuclear envelope (38). By

manipulating the cDNA's of types I and III, the type I N-terminal sequence could be changed to the type III sequence. It would be interesting to see if the expressed enzyme is now associated with the nuclear envelope. If this were the case, the kinetic properties of the type III isozyme would have been exchanged for the type I isozyme. A major difference is that the type III isozyme is inhibited by the substrate glucose (17), and since type I is not, the intracellular effects of eliminating this inhibition may possibly be investigated.

APPENDICES

APPENDIX A

RESTRICTION SITES FOR HEXOKINASE TYPE I cDNA

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
AAT 1 (AGGCCT)						
	2					
		1039	1599	(43.5)	1039	2638
		2638	1039	(28.3)	1	1039
			1035	(28.2)	2638	3673
AAT 2 (GACGTC)						
	1					
		1406	2267	(61.7)	1406	3673
			1406	(38.3)	1	1406
ACC 1 (GTVWAC)						
	1					
		3151	3151	(85.8)	1	3151
			522	(14.2)	3151	3673
ACC 2 (CGCG)						
	2					
		100	2930	(79.8)	100	3030
		3030	643	(17.5)	3030	3673
			100	(2.7)	1	100
ACC 3 (TCCGGA)						
	2					
		304	2035	(55.4)	1638	3673
		1638	1334	(36.3)	304	1638
			304	(8.3)	1	304
ACY 1 (GPCGQC)						
	2					
		1406	1454	(39.6)	2219	3673
		2219	1406	(38.3)	1	1406
			813	(22.1)	1406	2219
AFL 3 (ACPQGT)						
	2					
		1104	1421	(38.7)	1104	2525
		2525	1148	(31.3)	2525	3673
			1104	(30.1)	1	1104
AHA 2 (GPCGQC)						
	2					
		1406	1454	(39.6)	2219	3673
		2219	1406	(38.3)	1	1406
			813	(22.1)	1406	2219
ALU 1 (AGCT)						
	11					
		129	699	(19.0)	1809	2508
		465	538	(14.6)	1071	1609
		963	505	(13.7)	3168	3673
		1002	498	(13.6)	465	963
		1071	362	(9.9)	2806	3168
		1609	336	(9.1)	129	465

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1809	220	(6.0)	2586	2806
		2508	200	(5.4)	1609	1809
		2586	129	(3.5)	1	129
		2806	78	(2.1)	2508	2586
		3168	69	(1.9)	1002	1071
			39	(1.1)	963	1002
APA 1 (GGGCCC)						
	1					
		3204	3204	(87.2)	1	3204
			469	(12.8)	3204	3673
ASU 2 (TTCGAA)						
	2					
		360	2624	(71.4)	1049	3673
		1049	689	(18.8)	360	1049
			360	(9.8)	1	360
AVA 1 (CQCGPG)						
	2					
		246	2178	(59.3)	246	2424
		2424	1249	(34.0)	2424	3673
			246	(6.7)	1	246
AVA 2 (GGRCC)						
	9					
		19	1678	(45.7)	1311	2989
		297	613	(16.7)	3060	3673
		726	429	(11.7)	297	726
		894	278	(7.6)	19	297
		1006	195	(5.3)	1006	1201
		1201	168	(4.6)	726	894
		1311	112	(3.0)	894	1006
		2989	110	(3.0)	1201	1311
		3060	71	(1.9)	2989	3060
			19	(0.5)	1	19
AVA 3 (ATGCAT)						
	1					
		3112	3112	(84.7)	1	3112
			561	(15.3)	3112	3673
BAL 1 (TGGCCA)						
	2					
		1029	2078	(56.6)	1029	3107
		3107	1029	(28.0)	1	1029
			566	(15.4)	3107	3673
BAM H1 (GGATCC)						
	1					
		2855	2855	(77.7)	1	2855
			818	(22.3)	2855	3673
BAN 1 (GGQPCC)						
	9					
		788	788	(21.5)	1	788
		1262	544	(14.8)	2594	3138

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1393	535	(14.6)	3138	3673
		1796	474	(12.9)	788	1262
		2069	403	(11.0)	1393	1796
		2132	375	(10.2)	2219	2594
		2219	273	(7.4)	1796	2069
		2594	131	(3.6)	1262	1393
		3138	87	(2.4)	2132	2219
			63	(1.7)	2069	2132
BAN 2 (GPGCQC)						
	1					
		3204	3204	(87.2)	1	3204
			469	(12.8)	3204	3673
BBV 1 (GCTGC)						
	10					
		280	855	(23.3)	1624	2479
		961	681	(18.5)	280	961
		1072	414	(11.3)	1072	1486
		1486	402	(10.9)	2971	3373
		1624	300	(8.2)	3373	3673
		2479	280	(7.6)	1	280
		2506	272	(7.4)	2506	2778
		2778	193	(5.3)	2778	2971
		2971	138	(3.8)	1486	1624
		3373	111	(3.0)	961	1072
			27	(0.7)	2479	2506
BCL 1 (TGATCA)						
	3					
		594	1548	(42.1)	774	2322
		774	1351	(36.8)	2322	3673
		2322	594	(16.2)	1	594
			180	(4.9)	594	774
BGL 1 (GCCNNNNNGG						
	1					
		2597	2597	(70.7)	1	2597
			1076	(29.3)	2597	3673
BGL 2 (AGATCT)						
	3					
		420	1344	(36.6)	420	1764
		1764	1270	(34.6)	2403	3673
		2403	639	(17.4)	1764	2403
			420	(11.4)	1	420
BIN 1 (GGATC)						
	7					
		340	995	(27.1)	1860	2855
		941	817	(22.2)	2856	3673
		1684	743	(20.2)	941	1684
		1723	601	(16.4)	340	941
		1860	340	(9.3)	1	340
		2855	137	(3.7)	1723	1860
		2856	39	(1.1)	1684	1723
			1	(0.0)	2855	2856

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
BSM 1 (GAATGC)						
	1					
		1134	2539	(69.1)	1134	3673
			1134	(30.9)	1	1134
BSP 1286 (G2GC3C)						
	6					
		1294	1294	(35.2)	1	1294
		1795	872	(23.7)	2131	3003
		2068	501	(13.6)	1294	1795
		2131	469	(12.8)	3204	3673
		3003	273	(7.4)	1795	2068
		3204	201	(5.5)	3003	3204
			63	(1.7)	2068	2131
BSP M1 (ACCTGC)						
	3					
		62	2936	(79.9)	737	3673
		376	361	(9.8)	376	737
		737	314	(8.5)	62	376
			62	(1.7)	1	62
BSP M2 (TCCGGA)						
	2					
		304	2035	(55.4)	1638	3673
		1638	1334	(36.3)	304	1638
			304	(8.3)	1	304
BST N1 (CCRGG)						
	23					
		32	380	(10.3)	3105	3485
		337	339	(9.2)	490	829
		490	310	(8.4)	2542	2852
		829	305	(8.3)	32	337
		898	267	(7.3)	1921	2188
		953	234	(6.4)	1243	1477
		1018	225	(6.1)	1018	1243
		1243	188	(5.1)	3485	3673
		1477	165	(4.5)	1510	1675
		1510	153	(4.2)	337	490
		1675	150	(4.1)	2188	2338
		1780	147	(4.0)	2338	2485
		1846	117	(3.2)	2940	3057
		1921	105	(2.9)	1675	1780
		2188	88	(2.4)	2852	2940
		2338	75	(2.0)	1846	1921
		2485	69	(1.9)	829	898
		2542	66	(1.8)	1780	1846
		2852	65	(1.8)	953	1018
		2940	57	(1.6)	2485	2542
		3057	55	(1.5)	898	953
		3105	48	(1.3)	3057	3105
		3485	33	(0.9)	1477	1510
			32	(0.9)	1	32
BST X1 (CCANNNNNN						
	3					
		621	1900	(51.7)	1773	3673
		1239	621	(16.9)	1	621
		1773	618	(16.8)	621	1239

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
			534	(14.5)	1239	1773
CFR 1 (QGGCCP)						
	4					
		1029	1356	(36.9)	1248	2604
		1248	1029	(28.0)	1	1029
		2604	566	(15.4)	3107	3673
		3107	503	(13.7)	2604	3107
			219	(6.0)	1029	1248
CLA 1 (ATCGAT)						
	1					
		3398	3398	(92.5)	1	3398
			275	(7.5)	3398	3673
DDE 1 (CTNAG)						
	13					
		41	965	(26.3)	417	1382
		313	696	(18.9)	2822	3518
		417	454	(12.4)	2000	2454
		1382	414	(11.3)	1586	2000
		1423	272	(7.4)	41	313
		1482	239	(6.5)	2583	2822
		1526	155	(4.2)	3518	3673
		1586	129	(3.5)	2454	2583
		2000	104	(2.8)	313	417
		2454	60	(1.6)	1526	1586
		2583	59	(1.6)	1423	1482
		2822	44	(1.2)	1482	1526
		3518	41	(1.1)	1382	1423
			41	(1.1)	1	41
EAE 1 (QGGCCP)						
	4					
		1029	1356	(36.9)	1248	2604
		1248	1029	(28.0)	1	1029
		2604	566	(15.4)	3107	3673
		3107	503	(13.7)	2604	3107
			219	(6.0)	1029	1248
ECO O109 (PGGNCCQ						
	3					
		893	1337	(36.4)	1867	3204
		1867	974	(26.5)	893	1867
		3204	893	(24.3)	1	893
			469	(12.8)	3204	3673
ECO R5 (GATATC)						
	1					
		203	3470	(94.5)	203	3673
			203	(5.5)	1	203
FNU 4H1 (GCNGC)						
	24					
		10	855	(23.3)	1624	2479
		98	681	(18.5)	280	961
		178	300	(8.2)	3373	3673
		280	234	(6.4)	3136	3370
		961	198	(5.4)	1250	1448
		984	178	(4.8)	1072	1250

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1072	172	(4.7)	2606	2778
		1250	165	(4.5)	2971	3136
		1448	150	(4.1)	2818	2968
		1486	138	(3.8)	1486	1624
		1624	102	(2.8)	178	280
		2479	88	(2.4)	984	1072
		2506	88	(2.4)	10	98
		2579	80	(2.2)	98	178
		2606	73	(2.0)	2506	2579
		2778	38	(1.0)	1448	1486
		2781	27	(0.7)	2579	2606
		2792	27	(0.7)	2479	2506
		2818	26	(0.7)	2792	2818
		2968	23	(0.6)	961	984
		2971	11	(0.3)	2781	2792
		3136	10	(0.3)	1	10
		3370	3	(0.1)	3370	3373
		3373	3	(0.1)	2968	2971
			3	(0.1)	2778	2781
FNU D2 (CGCG)						
	2					
		100	2930	(79.8)	100	3030
		3030	643	(17.5)	3030	3673
			100	(2.7)	1	100
FOK 1 (GGATG)						
	21					
		75	738	(20.1)	907	1645
		136	526	(14.3)	2693	3219
		640	504	(13.7)	136	640
		852	281	(7.7)	3392	3673
		886	212	(5.8)	640	852
		907	207	(5.6)	2038	2245
		1645	174	(4.7)	2497	2671
		1651	163	(4.4)	3219	3382
		1804	153	(4.2)	1651	1804
		1875	138	(3.8)	2359	2497
		2005	130	(3.5)	1875	2005
		2038	86	(2.3)	2245	2331
		2245	75	(2.0)	1	75
		2331	71	(1.9)	1804	1875
		2359	61	(1.7)	75	136
		2497	34	(0.9)	852	886
		2671	33	(0.9)	2005	2038
		2693	28	(0.8)	2331	2359
		3219	22	(0.6)	2671	2693
		3382	21	(0.6)	886	907
		3392	10	(0.3)	3382	3392
			6	(0.2)	1645	1651
GDI 2 (QGGCCG)						
	2					
		1248	1356	(36.9)	1248	2604
		2604	1248	(34.0)	1	1248
			1069	(29.1)	2604	3673
HAE 1 (RGGCCR)						
	8					
		111	961	(26.2)	1677	2638
		238	791	(21.5)	238	1029

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1029	638	(17.4)	1039	1677
		1039	469	(12.8)	2638	3107
		1677	365	(9.9)	3308	3673
		2638	201	(5.5)	3107	3308
		3107	127	(3.5)	111	238
		3308	111	(3.0)	1	111
			10	(0.3)	1029	1039
HAE 2 (PGCGCQ)						
	3					
		2219	2219	(60.4)	1	2219
		2476	717	(19.5)	2956	3673
		2956	480	(13.1)	2476	2956
			257	(7.0)	2219	2476
HAE 3 (GGCC)						
	21					
		54	531	(14.5)	239	770
		112	364	(9.9)	3309	3673
		239	317	(8.6)	2791	3108
		770	302	(8.2)	2192	2494
		1030	260	(7.1)	770	1030
		1040	213	(5.8)	1979	2192
		1249	209	(5.7)	1040	1249
		1447	198	(5.4)	1249	1447
		1513	190	(5.2)	1678	1868
		1678	165	(4.5)	1513	1678
		1868	152	(4.1)	2639	2791
		1979	127	(3.5)	112	239
		2192	111	(3.0)	2494	2605
		2494	111	(3.0)	1868	1979
		2605	97	(2.6)	3108	3205
		2639	77	(2.1)	3232	3309
		2791	66	(1.8)	1447	1513
		3108	58	(1.6)	54	112
		3205	54	(1.5)	1	54
		3232	34	(0.9)	2605	2639
		3309	27	(0.7)	3205	3232
			10	(0.3)	1030	1040
HGA 1 (GACGC)						
	3					
		1537	1537	(41.8)	1	1537
		2677	1140	(31.0)	1537	2677
		3149	524	(14.3)	3149	3673
			472	(12.9)	2677	3149
HGI A1 (GRGCRC)						
	1					
		3003	3003	(81.8)	1	3003
			670	(18.2)	3003	3673
HGI C1 (GGQPCC)						
	9					
		788	788	(21.5)	1	788
		1262	544	(14.8)	2594	3138
		1393	535	(14.6)	3138	3673
		1796	474	(12.9)	788	1262
		2069	403	(11.0)	1393	1796
		2132	375	(10.2)	2219	2594

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2219	273	(7.4)	1796	2069
		2594	131	(3.6)	1262	1393
		3138	87	(2.4)	2132	2219
			63	(1.7)	2069	2132
HGI J2 (GPGCQC)						
	1					
		3204	3204	(87.2)	1	3204
			469	(12.8)	3204	3673
HHA 1 (GCGC)						
	7					
		101	1163	(31.7)	1057	2220
		1057	956	(26.0)	101	1057
		2220	486	(13.2)	3029	3515
		2477	480	(13.1)	2477	2957
		2957	257	(7.0)	2220	2477
		3029	158	(4.3)	3515	3673
		3515	101	(2.7)	1	101
			72	(2.0)	2957	3029
HINC 2 (GTQPAC)						
	2					
		1187	1619	(44.1)	2054	3673
		2054	1187	(32.3)	1	1187
			867	(23.6)	1187	2054
HINF 1 (GANTC)						
	14					
		193	857	(23.3)	543	1400
		216	594	(16.2)	2115	2709
		363	532	(14.5)	1400	1932
		413	383	(10.4)	2709	3092
		543	354	(9.6)	3288	3642
		1400	193	(5.3)	1	193
		1932	183	(5.0)	1932	2115
		2115	147	(4.0)	216	363
		2709	130	(3.5)	413	543
		3092	86	(2.3)	3092	3178
		3178	61	(1.7)	3227	3288
		3227	50	(1.4)	363	413
		3288	49	(1.3)	3178	3227
		3642	31	(0.8)	3642	3673
			23	(0.6)	193	216
HPA 2 (CCGG)						
	16					
		247	704	(19.2)	305	1009
		295	546	(14.9)	2598	3144
		305	522	(14.2)	1873	2395
		1009	388	(10.6)	3285	3673
		1363	354	(9.6)	1009	1363
		1492	247	(6.7)	1	247
		1639	147	(4.0)	1492	1639
		1649	141	(3.8)	3144	3285
		1705	129	(3.5)	1363	1492
		1800	128	(3.5)	2470	2598
		1873	95	(2.6)	1705	1800
		2395	75	(2.0)	2395	2470
		2470	73	(2.0)	1800	1873
		2598	56	(1.5)	1649	1705

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		3144	48	(1.3)	247	295
		3285	10	(0.3)	1639	1649
			10	(0.3)	295	305
HPH 1 (GGTGA)						
	16					
		29	531	(14.5)	2738	3269
		123	435	(11.8)	1021	1456
		379	404	(11.0)	3269	3673
		709	365	(9.9)	2373	2738
		979	330	(9.0)	379	709
		1021	270	(7.4)	709	979
		1456	256	(7.0)	123	379
		1549	223	(6.1)	1667	1890
		1667	217	(5.9)	2011	2228
		1890	121	(3.3)	1890	2011
		2011	118	(3.2)	1549	1667
		2228	114	(3.1)	2228	2342
		2342	94	(2.6)	29	123
		2373	93	(2.5)	1456	1549
		2738	42	(1.1)	979	1021
		3269	31	(0.8)	2342	2373
			29	(0.8)	1	29
MBO 2 (GAAGA)						
	15					
		232	528	(14.4)	1024	1552
		353	475	(12.9)	3198	3673
		391	459	(12.5)	1555	2014
		514	427	(11.6)	2771	3198
		902	388	(10.6)	514	902
		973	380	(10.3)	2391	2771
		1024	232	(6.3)	1	232
		1552	230	(6.3)	2161	2391
		1555	123	(3.3)	391	514
		2014	121	(3.3)	232	353
		2093	79	(2.2)	2014	2093
		2161	71	(1.9)	902	973
		2391	68	(1.9)	2093	2161
		2771	51	(1.4)	973	1024
		3198	38	(1.0)	353	391
			3	(0.1)	1552	1555
MNL 1 (CCTC)						
	44					
		17	317	(8.6)	267	584
		43	303	(8.2)	1081	1384
		241	209	(5.7)	2180	2389
		267	203	(5.5)	2981	3184
		584	198	(5.4)	43	241
		678	192	(5.2)	850	1042
		809	188	(5.1)	1692	1880
		845	172	(4.7)	3501	3673
		850	148	(4.0)	2427	2575
		1042	144	(3.9)	3357	3501
		1081	131	(3.6)	678	809
		1384	129	(3.5)	2024	2153
		1417	124	(3.4)	2637	2761
		1420	113	(3.1)	3244	3357
		1502	104	(2.8)	1588	1692
		1525	96	(2.6)	1880	1976

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1547	94	(2.6)	584	678
		1558	83	(2.3)	2894	2977
		1588	82	(2.2)	1420	1502
		1692	65	(1.8)	2761	2826
		1880	62	(1.7)	2575	2637
		1976	50	(1.4)	3184	3234
		1992	39	(1.1)	1042	1081
		2002	38	(1.0)	2389	2427
		2017	36	(1.0)	809	845
		2024	35	(1.0)	2859	2894
		2153	33	(0.9)	2826	2859
		2180	33	(0.9)	1384	1417
		2389	30	(0.8)	1558	1588
		2427	27	(0.7)	2153	2180
		2575	26	(0.7)	241	267
		2637	26	(0.7)	17	43
		2761	23	(0.6)	1502	1525
		2826	22	(0.6)	1525	1547
		2859	17	(0.5)	1	17
		2894	16	(0.4)	1976	1992
		2977	15	(0.4)	2002	2017
		2981	11	(0.3)	1547	1558
		3184	10	(0.3)	1992	2002
		3234	7	(0.2)	2017	2024
		3239	5	(0.1)	3239	3244
		3244	5	(0.1)	3234	3239
		3357	5	(0.1)	845	850
		3501	4	(0.1)	2977	2981
			3	(0.1)	1417	1420
MST 2 (CCTNAGG)						
	1					
		1381	2292	(62.4)	1381	3673
			1381	(37.6)	1	1381
NAE 1 (GCCGGC)						
	1					
		2597	2597	(70.7)	1	2597
			1076	(29.3)	2597	3673
NAR 1 (GGCGCC)						
	1					
		2219	2219	(60.4)	1	2219
			1454	(39.6)	2219	3673
NCI 1 (CCSGG)						
	9					
		246	1115	(30.4)	247	1362
		247	748	(20.4)	2395	3143
		1362	530	(14.4)	3143	3673
		1648	523	(14.2)	1872	2395
		1705	286	(7.8)	1362	1648
		1800	246	(6.7)	1	246
		1872	95	(2.6)	1705	1800
		2395	72	(2.0)	1800	1872
		3143	57	(1.6)	1648	1705
			1	(0.0)	246	247

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
NCO 1 (CCATGG)						
	1					
		1452	2221	(60.5)	1452	3673
			1452	(39.5)	1	1452
NLA 3 (CATG)						
	24					
		91	459	(12.5)	994	1453
		175	393	(10.7)	2209	2602
		409	376	(10.2)	2713	3089
		449	343	(9.3)	3330	3673
		476	336	(9.1)	1453	1789
		502	234	(6.4)	175	409
		730	228	(6.2)	502	730
		805	208	(5.7)	3122	3330
		988	183	(5.0)	805	988
		994	111	(3.0)	2602	2713
		1453	92	(2.5)	1982	2074
		1789	91	(2.5)	1	91
		1855	89	(2.4)	1855	1944
		1944	84	(2.3)	91	175
		1982	75	(2.0)	2074	2149
		2074	75	(2.0)	730	805
		2149	66	(1.8)	1789	1855
		2209	60	(1.6)	2149	2209
		2602	40	(1.1)	409	449
		2713	38	(1.0)	1944	1982
		3089	27	(0.7)	449	476
		3111	26	(0.7)	476	502
		3122	22	(0.6)	3089	3111
		3330	11	(0.3)	3111	3122
			6	(0.2)	988	994
NLA 4 (GGNNCC)						
	28					
		45	404	(11.0)	54	458
		54	375	(10.2)	2219	2594
		458	351	(9.6)	1445	1796
		725	280	(7.6)	894	1174
		788	267	(7.3)	458	725
		875	232	(6.3)	3204	3436
		893	201	(5.5)	1868	2069
		894	195	(5.3)	2594	2789
		1174	133	(3.6)	2855	2988
		1262	121	(3.3)	3436	3557
		1368	116	(3.2)	3557	3673
		1393	106	(2.9)	1262	1368
		1445	88	(2.4)	1174	1262
		1796	87	(2.4)	788	875
		1868	78	(2.1)	3060	3138
		2069	72	(2.0)	2988	3060
		2132	72	(2.0)	1796	1868
		2184	66	(1.8)	3138	3204
		2219	66	(1.8)	2789	2855
		2594	63	(1.7)	2069	2132
		2789	63	(1.7)	725	788
		2855	52	(1.4)	2132	2184
		2988	52	(1.4)	1393	1445
		3060	45	(1.2)	1	45
		3138	35	(1.0)	2184	2219

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		3204	25	(0.7)	1368	1393
		3436	18	(0.5)	875	893
		3557	9	(0.2)	45	54
			1	(0.0)	893	894
NSI 1 (ATGCAT)						
	1					
		3112	3112	(84.7)	1	3112
			561	(15.3)	3112	3673
NSP B2 (CVGCWG)						
	6					
		11	1545	(42.1)	962	2507
		962	951	(25.9)	11	962
		2507	868	(23.6)	2805	3673
		2585	194	(5.3)	2585	2779
		2779	78	(2.1)	2507	2585
		2805	26	(0.7)	2779	2805
			11	(0.3)	1	11
NSP C1 (PCATGQ)						
	1					
		987	2686	(73.1)	987	3673
			987	(26.9)	1	987
PPU M1 (PGGRCCQ)						
	1					
		893	2780	(75.7)	893	3673
			893	(24.3)	1	893
PST 1 (CTGCAG)						
	1					
		3156	3156	(85.9)	1	3156
			517	(14.1)	3156	3673
PVU 1 (CGATCG)						
	2					
		1218	1620	(44.1)	1218	2838
		2838	1218	(33.2)	1	1218
			835	(22.7)	2838	3673
PVU 2 (CAGCTG)						
	4					
		962	1545	(42.1)	962	2507
		2507	962	(26.2)	1	962
		2585	868	(23.6)	2805	3673
		2805	220	(6.0)	2585	2805
			78	(2.1)	2507	2585
RRU 1 (AGTACT)						
	1					
		1356	2317	(63.1)	1356	3673
			1356	(36.9)	1	1356
RSA 1 (GTAC)						
	8					
		169	978	(26.6)	1357	2335
		590	871	(23.7)	2748	3619
		991	421	(11.5)	169	590
		1357	413	(11.2)	2335	2748
		2335	401	(10.9)	590	991

#	SITES	FRAGMENTS		FRAGMENT ENDS	
	2748	366	(10.0)	991	1357
	3619	169	(4.6)	1	169
	3637	36	(1.0)	3637	3673
		18	(0.5)	3619	3637
SAU 1 (CCTNAGG)					
1					
	1381	2292	(62.4)	1381	3673
		1381	(37.6)	1	1381
SAU 3A (GATC)					
24					
	5	625	(17.0)	3048	3673
	94	452	(12.3)	1233	1685
	341	384	(10.5)	1939	2323
	421	277	(7.5)	942	1219
	473	247	(6.7)	94	341
	517	219	(6.0)	2404	2623
	595	216	(5.9)	2623	2839
	775	187	(5.1)	2856	3043
	942	180	(4.9)	595	775
	1219	167	(4.5)	775	942
	1233	96	(2.6)	1765	1861
	1685	89	(2.4)	5	94
	1723	80	(2.2)	341	421
	1765	78	(2.1)	1861	1939
	1861	78	(2.1)	517	595
	1939	57	(1.6)	2347	2404
	2323	52	(1.4)	421	473
	2347	44	(1.2)	473	517
	2404	42	(1.1)	1723	1765
	2623	38	(1.0)	1685	1723
	2839	24	(0.7)	2323	2347
	2856	17	(0.5)	2839	2856
	3043	14	(0.4)	1219	1233
	3048	5	(0.1)	3043	3048
		5	(0.1)	1	5
SAU 96 (GGNCC)					
19					
	19	468	(12.7)	3205	3673
	54	429	(11.7)	297	726
	297	355	(9.7)	1513	1868
	726	302	(8.2)	2191	2493
	894	297	(8.1)	2493	2790
	1006	243	(6.6)	54	297
	1201	213	(5.8)	1978	2191
	1311	199	(5.4)	2790	2989
	1446	195	(5.3)	1006	1201
	1513	168	(4.6)	726	894
	1868	144	(3.9)	3060	3204
	1978	135	(3.7)	1311	1446
	2191	112	(3.0)	894	1006
	2493	110	(3.0)	1868	1978
	2790	110	(3.0)	1201	1311
	2989	71	(1.9)	2989	3060
	3060	67	(1.8)	1446	1513
	3204	35	(1.0)	19	54

#	SITES	FRAGMENTS		FRAGMENT ENDS	
	3205	19	(0.5)	1	19
		1	(0.0)	3204	3205
SCA 1 (AGTACT)					
1					
	1356	2317	(63.1)	1356	3673
		1356	(36.9)	1	1356
SCR F1 (CCNGG)					
32					
	32	342	(9.3)	3143	3485
	246	339	(9.2)	490	829
	247	310	(8.4)	2542	2852
	337	267	(7.3)	1921	2188
	490	225	(6.1)	1018	1243
	829	214	(5.8)	32	246
	898	188	(5.1)	3485	3673
	953	153	(4.2)	337	490
	1018	150	(4.1)	2188	2338
	1243	138	(3.8)	1510	1648
	1362	119	(3.2)	1243	1362
	1477	117	(3.2)	2940	3057
	1510	115	(3.1)	1362	1477
	1648	90	(2.5)	2395	2485
	1675	90	(2.5)	247	337
	1705	88	(2.4)	2852	2940
	1780	75	(2.0)	1705	1780
	1800	69	(1.9)	829	898
	1846	65	(1.8)	953	1018
	1872	57	(1.6)	2485	2542
	1921	57	(1.6)	2338	2395
	2188	55	(1.5)	898	953
	2338	49	(1.3)	1872	1921
	2395	48	(1.3)	3057	3105
	2485	46	(1.3)	1800	1846
	2542	38	(1.0)	3105	3143
	2852	33	(0.9)	1477	1510
	2940	32	(0.9)	1	32
	3057	30	(0.8)	1675	1705
	3105	27	(0.7)	1648	1675
	3143	26	(0.7)	1846	1872
	3485	20	(0.5)	1780	1800
		1	(0.0)	246	247
SDU 1 (G2GC3C)					
6					
	1294	1294	(35.2)	1	1294
	1795	872	(23.7)	2131	3003
	2068	501	(13.6)	1294	1795
	2131	469	(12.8)	3204	3673
	3003	273	(7.4)	1795	2068
	3204	201	(5.5)	3003	3204
		63	(1.7)	2068	2131
SFA N1 (GATGC)					
12					
	74	971	(26.4)	2430	3401
	277	412	(11.2)	277	689
	689	370	(10.1)	689	1059
	1059	299	(8.1)	1345	1644
	1345	286	(7.8)	1059	1345

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1644	272	(7.4)	3401	3673
		1833	228	(6.2)	2202	2430
		1876	203	(5.5)	74	277
		2006	196	(5.3)	2006	2202
		2202	189	(5.1)	1644	1833
		2430	130	(3.5)	1876	2006
		3401	74	(2.0)	1	74
			43	(1.2)	1833	1876
SMA 1 (CCCGGG)						
	1					
		246	3427	(93.3)	246	3673
			246	(6.7)	1	246
SPE 1 (ACTAGT)						
	1					
		1097	2576	(70.1)	1097	3673
			1097	(29.9)	1	1097
SSP 1 (AATATT)						
	1					
		2283	2283	(62.2)	1	2283
			1390	(37.8)	2283	3673
STU 1 (AGGCCT)						
	2					
		1039	1599	(43.5)	1039	2638
		2638	1039	(28.3)	1	1039
			1035	(28.2)	2638	3673
STY 1 (CCRRGG)						
	4					
		1032	1993	(54.3)	1680	3673
		1146	1032	(28.1)	1	1032
		1452	306	(8.3)	1146	1452
		1680	228	(6.2)	1452	1680
			114	(3.1)	1032	1146
TAQ 1 (TCGA)						
	8					
		361	1787	(48.7)	1050	2837
		471	361	(9.8)	1	361
		825	354	(9.6)	471	825
		969	298	(8.1)	3101	3399
		1050	274	(7.5)	3399	3673
		2837	264	(7.2)	2837	3101
		3101	144	(3.9)	825	969
		3399	110	(3.0)	361	471
			81	(2.2)	969	1050
TTH111 1 (GACNNNG						
	1					
		140	3533	(96.2)	140	3673
			140	(3.8)	1	140
TTH111 2 (CCAPCA)						
	9					
		260	1173	(31.9)	1713	2886

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		754	494	(13.4)	260	754
		1204	450	(12.3)	754	1204
		1599	395	(10.8)	1204	1599
		1713	341	(9.3)	2931	3272
		2886	312	(8.5)	3361	3673
		2931	260	(7.1)	1	260
		3272	114	(3.1)	1599	1713
		3361	89	(2.4)	3272	3361
			45	(1.2)	2886	2931
XBA 1 (TCTAGA)						
	1					
		2898	2898	(78.9)	1	2898
			775	(21.1)	2898	3673
XHO 2 (PGATCQ)						
	8					
		340	818	(22.3)	2855	3673
		420	743	(20.2)	941	1684
		941	639	(17.4)	1764	2403
		1684	521	(14.2)	420	941
		1722	452	(12.3)	2403	2855
		1764	340	(9.3)	1	340
		2403	80	(2.2)	340	420
		2855	42	(1.1)	1722	1764
			38	(1.0)	1684	1722
XMN 1 (GAANNNNTTC						
	2					
		1130	1293	(35.2)	2380	3673
		2380	1250	(34.0)	1130	2380
			1130	(30.8)	1	1130

The following do not appear:

AFL 2	AHA 3	AOS 1	APA L1
ASP718 1	AVR 2	BSPH 1	BSS H2
BST E2	DRA 3	ECO R1	HIND 3
HPA 1	KPN 1	MLU 1	MST 1
NDE 1	NHE 1	NOT 1	NRU 1
PFL M1	RSR 2	SAC 1	SAC 2
SAL 1	SFI 1	SNA 1	SNA B1
SPH 1	XHO 1	XMA 3	

APPENDIX B

RESTRICTION SITES FOR TYPE III HEXOKINASE cDNA

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
AAT 1 (AGGCCT)						
	2					
		1210	1824	(49.4)	1868	3692
		1868	1210	(32.8)	1	1210
			658	(17.8)	1210	1868
AAT 2 (GACGTC)						
	1					
		2534	2534	(68.6)	1	2534
			1158	(31.4)	2534	3692
ACC 1 (GTVWAC)						
	2					
		807	1916	(51.9)	1776	3692
		1776	969	(26.2)	807	1776
			807	(21.9)	1	807
ACC 2 (CGCG)						
	4					
		873	1952	(52.9)	1740	3692
		1236	873	(23.6)	1	873
		1258	482	(13.1)	1258	1740
		1740	363	(9.8)	873	1236
			22	(0.6)	1236	1258
ACC 3 (TCCGGA)						
	2					
		295	1875	(50.8)	295	2170
		2170	1522	(41.2)	2170	3692
			295	(8.0)	1	295
ACY 1 (GPCGQC)						
	2					
		2534	2534	(68.6)	1	2534
		3453	919	(24.9)	2534	3453
			239	(6.5)	3453	3692
AFL 3 (ACPQGT)						
	4					
		389	1486	(40.2)	2206	3692
		1130	741	(20.1)	389	1130
		1739	609	(16.5)	1130	1739
		2206	467	(12.6)	1739	2206
			389	(10.5)	1	389
AHA 2 (GPCGQC)						
	2					
		2534	2534	(68.6)	1	2534
		3453	919	(24.9)	2534	3453
			239	(6.5)	3453	3692
ALU 1 (AGCT)						
	26					
		16	684	(18.5)	2743	3427
		129	615	(16.7)	2044	2659
		165	282	(7.6)	1270	1552
		217	207	(5.6)	565	772

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		364	195	(5.3)	1030	1225
		493	193	(5.2)	1626	1819
		565	173	(4.7)	857	1030
		772	147	(4.0)	217	364
		857	137	(3.7)	3555	3692
		1030	129	(3.5)	364	493
		1225	120	(3.3)	1924	2044
		1262	113	(3.1)	16	129
		1270	105	(2.8)	1819	1924
		1552	85	(2.3)	772	857
		1557	72	(2.0)	493	565
		1567	69	(1.9)	3427	3496
		1626	59	(1.6)	3496	3555
		1819	59	(1.6)	1567	1626
		1924	52	(1.4)	165	217
		2044	45	(1.2)	2698	2743
		2659	39	(1.1)	2659	2698
		2698	37	(1.0)	1225	1262
		2743	36	(1.0)	129	165
		3427	16	(0.4)	1	16
		3496	10	(0.3)	1557	1567
		3555	8	(0.2)	1262	1270
			5	(0.1)	1552	1557
APA 1 (GGGCCC)						
	1					
		2923	2923	(79.2)	1	2923
			769	(20.8)	2923	3692
APA L1 (GTGCAC)						
	2					
		1253	1503	(40.7)	1253	2756
		2756	1253	(33.9)	1	1253
			936	(25.4)	2756	3692
ASP718 1 (GGTACC)						
	3					
		753	1550	(42.0)	2142	3692
		816	1326	(35.9)	816	2142
		2142	753	(20.4)	1	753
			63	(1.7)	753	816
ASU 2 (TTCGAA)						
	1					
		1368	2324	(62.9)	1368	3692
			1368	(37.1)	1	1368
AVA 1 (CQCGPG)						
	3					
		110	2080	(56.3)	110	2190
		2190	1048	(28.4)	2190	3238
		3238	454	(12.3)	3238	3692
			110	(3.0)	1	110
AVA 2 (GGRCC)						
	18					
		319	555	(15.0)	1148	1703
		674	484	(13.1)	2911	3395
		709	356	(9.6)	2555	2911
		869	355	(9.6)	319	674
		962	319	(8.6)	1	319

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1049	231	(6.3)	1703	1934
		1070	224	(6.1)	3395	3619
		1148	210	(5.7)	2288	2498
		1703	208	(5.6)	2080	2288
		1934	160	(4.3)	709	869
		2080	146	(4.0)	1934	2080
		2288	93	(2.5)	869	962
		2498	87	(2.4)	962	1049
		2516	78	(2.1)	1070	1148
		2555	73	(2.0)	3619	3692
		2911	39	(1.1)	2516	2555
		3395	35	(0.9)	674	709
		3619	21	(0.6)	1049	1070
			18	(0.5)	2498	2516
AVA 3 (ATGCAT)						
	2					
		535	1720	(46.6)	1972	3692
		1972	1437	(38.9)	535	1972
			535	(14.5)	1	535
AVR 2 (CCTAGG)						
	3					
		476	1986	(53.8)	1706	3692
		926	780	(21.1)	926	1706
		1706	476	(12.9)	1	476
			450	(12.2)	476	926
BAL 1 (TGGCCA)						
	4					
		485	1759	(47.6)	1348	3107
		1135	650	(17.6)	485	1135
		1348	585	(15.8)	3107	3692
		3107	485	(13.1)	1	485
			213	(5.8)	1135	1348
BAM H1 (GGATCC)						
	1					
		3358	3358	(91.0)	1	3358
			334	(9.0)	3358	3692
BAN 1 (GGQPCC)						
	12					
		333	730	(19.8)	2962	3692
		414	697	(18.9)	2265	2962
		753	559	(15.1)	816	1375
		777	429	(11.6)	1713	2142
		816	339	(9.2)	414	753
		1375	333	(9.0)	1	333
		1541	172	(4.7)	1541	1713
		1713	166	(4.5)	1375	1541
		2142	87	(2.4)	2142	2229
		2229	81	(2.2)	333	414
		2265	39	(1.1)	777	816
		2962	36	(1.0)	2229	2265
			24	(0.7)	753	777
BAN 2 (GPGCQC)						
	7					
		902	1118	(30.3)	1805	2923
		1202	902	(24.4)	1	902

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1224	433	(11.7)	3259	3692
		1625	401	(10.9)	1224	1625
		1805	336	(9.1)	2923	3259
		2923	300	(8.1)	902	1202
		3259	180	(4.9)	1625	1805
			22	(0.6)	1202	1224
BBV 1 (GCTGC)						
	24					
		130	635	(17.2)	1981	2616
		218	497	(13.5)	2802	3299
		221	341	(9.2)	514	855
		491	270	(7.3)	221	491
		507	247	(6.7)	1675	1922
		514	229	(6.2)	855	1084
		855	186	(5.0)	2616	2802
		1084	172	(4.7)	1325	1497
		1232	151	(4.1)	3402	3553
		1263	148	(4.0)	1084	1232
		1290	139	(3.8)	3553	3692
		1325	130	(3.5)	1	130
		1497	88	(2.4)	130	218
		1565	78	(2.1)	1565	1643
		1643	75	(2.0)	3327	3402
		1675	68	(1.8)	1497	1565
		1922	59	(1.6)	1922	1981
		1981	35	(0.9)	1290	1325
		2616	32	(0.9)	1643	1675
		2802	31	(0.8)	1232	1263
		3299	28	(0.8)	3299	3327
		3327	27	(0.7)	1263	1290
		3402	16	(0.4)	491	507
		3553	7	(0.2)	507	514
			3	(0.1)	218	221
BCL 1 (TGATCA)						
	3					
		1603	1603	(43.4)	1	1603
		1826	1360	(36.8)	2332	3692
		2332	506	(13.7)	1826	2332
			223	(6.0)	1603	1826
BGL 1 (GCCNNNNNGG						
	1					
		1500	2192	(59.4)	1500	3692
			1500	(40.6)	1	1500
BIN 1 (GGATC)						
	8					
		227	1633	(44.2)	473	2106
		458	773	(20.9)	2106	2879
		473	479	(13.0)	2879	3358
		2106	231	(6.3)	227	458
		2879	227	(6.1)	1	227
		3358	173	(4.7)	3359	3532
		3359	160	(4.3)	3532	3692
		3532	15	(0.4)	458	473
			1	(0.0)	3358	3359

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
BSM 1 (GAATGC)						
	1					
		145	3547	(96.1)	145	3692
			145	(3.9)	1	145
BSP 1286 (G2GC3C)						
	16					
		413	569	(15.4)	2187	2756
		614	433	(11.7)	3259	3692
		776	413	(11.2)	1	413
		902	382	(10.3)	1805	2187
		985	372	(10.1)	1253	1625
		1202	296	(8.0)	2963	3259
		1224	217	(5.9)	985	1202
		1253	201	(5.4)	413	614
		1625	167	(4.5)	2756	2923
		1712	162	(4.4)	614	776
		1805	126	(3.4)	776	902
		2187	93	(2.5)	1712	1805
		2756	87	(2.4)	1625	1712
		2923	83	(2.2)	902	985
		2963	40	(1.1)	2923	2963
		3259	29	(0.8)	1224	1253
			22	(0.6)	1202	1224
BSPH 1 (TCATGA)						
	1					
		3463	3463	(93.8)	1	3463
			229	(6.2)	3463	3692
BSP M1 (ACCTGC)						
	4					
		2268	2268	(61.4)	1	2268
		2607	578	(15.7)	3114	3692
		3053	446	(12.1)	2607	3053
		3114	339	(9.2)	2268	2607
			61	(1.7)	3053	3114
BSP M2 (TCCGGA)						
	2					
		295	1875	(50.8)	295	2170
		2170	1522	(41.2)	2170	3692
			295	(8.0)	1	295
BST E2 (GGTNACC)						
	1					
		203	3489	(94.5)	203	3692
			203	(5.5)	1	203
BST N1 (CCRGG)						
	39					
		54	354	(9.6)	1937	2291
		137	228	(6.2)	3095	3323
		353	216	(5.9)	137	353
		439	209	(5.7)	2886	3095
		446	208	(5.6)	3484	3692
		530	178	(4.8)	781	959
		628	170	(4.6)	2716	2886

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		665	162	(4.4)	1124	1286
		781	156	(4.2)	1538	1694
		959	146	(4.0)	2570	2716
		974	143	(3.9)	981	1124
		981	124	(3.4)	1381	1505
		1124	116	(3.1)	3332	3448
		1286	116	(3.1)	665	781
		1381	98	(2.7)	530	628
		1505	95	(2.6)	1286	1381
		1526	90	(2.4)	1802	1892
		1538	86	(2.3)	2348	2434
		1694	86	(2.3)	353	439
		1772	84	(2.3)	446	530
		1802	83	(2.2)	54	137
		1892	78	(2.1)	1694	1772
		1931	54	(1.5)	1	54
		1937	51	(1.4)	2519	2570
		2291	49	(1.3)	2434	2483
		2307	41	(1.1)	2307	2348
		2348	39	(1.1)	1892	1931
		2434	37	(1.0)	628	665
		2483	36	(1.0)	3448	3484
		2510	30	(0.8)	1772	1802
		2519	27	(0.7)	2483	2510
		2570	21	(0.6)	1505	1526
		2716	16	(0.4)	2291	2307
		2886	15	(0.4)	959	974
		3095	12	(0.3)	1526	1538
		3323	9	(0.2)	3323	3332
		3332	9	(0.2)	2510	2519
		3448	7	(0.2)	974	981
		3484	7	(0.2)	439	446
			6	(0.2)	1931	1937
BST X1 (CCANNNNNN						
	1					
		1802	1890	(51.2)	1802	3692
			1802	(48.8)	1	1802
CFR 1 (QGGCCP)						
	11					
		82	1327	(35.9)	1744	3071
		485	585	(15.8)	3107	3692
		875	403	(10.9)	82	485
		1135	390	(10.6)	485	875
		1342	278	(7.5)	1466	1744
		1348	260	(7.0)	875	1135
		1466	207	(5.6)	1135	1342
		1744	118	(3.2)	1348	1466
		3071	82	(2.2)	1	82
		3097	26	(0.7)	3071	3097
		3107	10	(0.3)	3097	3107
			6	(0.2)	1342	1348
DDE 1 (CTNAG)						
	20					
		18	566	(15.3)	3000	3566
		158	517	(14.0)	167	684
		167	499	(13.5)	2253	2752

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		684	332	(9.0)	1222	1554
		988	304	(8.2)	684	988
		1104	238	(6.4)	1960	2198
		1200	235	(6.4)	1554	1789
		1207	140	(3.8)	2752	2892
		1214	140	(3.8)	18	158
		1222	126	(3.4)	3566	3692
		1554	116	(3.1)	988	1104
		1789	108	(2.9)	2892	3000
		1872	96	(2.6)	1104	1200
		1960	88	(2.4)	1872	1960
		2198	83	(2.2)	1789	1872
		2253	55	(1.5)	2198	2253
		2752	18	(0.5)	1	18
		2892	9	(0.2)	158	167
		3000	8	(0.2)	1214	1222
		3566	7	(0.2)	1207	1214
			7	(0.2)	1200	1207
DRA 3 (CACNNNGTG)						
	2					
		385	3266	(88.5)	426	3692
		426	385	(10.4)	1	385
			41	(1.1)	385	426
EAE 1 (QGGCCP)						
	11					
		82	1327	(35.9)	1744	3071
		485	585	(15.8)	3107	3692
		875	403	(10.9)	82	485
		1135	390	(10.6)	485	875
		1342	278	(7.5)	1466	1744
		1348	260	(7.0)	875	1135
		1466	207	(5.6)	1135	1342
		1744	118	(3.2)	1348	1466
		3071	82	(2.2)	1	82
		3097	26	(0.7)	3071	3097
		3107	10	(0.3)	3097	3107
			6	(0.2)	1342	1348
ECO O109 (PGGNCCQ						
	9					
		153	1368	(37.1)	1147	2515
		708	595	(16.1)	2647	3242
		922	555	(15.0)	153	708
		1069	376	(10.2)	3242	3618
		1147	214	(5.8)	708	922
		2515	153	(4.1)	1	153
		2647	147	(4.0)	922	1069
		3242	132	(3.6)	2515	2647
		3618	78	(2.1)	1069	1147
			74	(2.0)	3618	3692
ECO R1 (GAATTC)						
	2					
		525	2504	(67.8)	525	3029
		3029	663	(18.0)	3029	3692
			525	(14.2)	1	525

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
FNU 4H1 (GCNGC)						
	32					
		84	608	(16.5)	1981	2589
		130	341	(9.2)	514	855
		218	270	(7.3)	221	491
		221	264	(7.2)	2827	3091
		491	247	(6.7)	1675	1922
		507	210	(5.7)	874	1084
		514	186	(5.0)	2616	2802
		855	177	(4.8)	3091	3268
		874	172	(4.7)	1325	1497
		1084	148	(4.0)	1084	1232
		1232	139	(3.8)	3553	3692
		1263	134	(3.6)	3419	3553
		1290	88	(2.4)	130	218
		1325	84	(2.3)	1	84
		1497	78	(2.1)	1565	1643
		1565	75	(2.0)	3327	3402
		1643	68	(1.8)	1497	1565
		1675	59	(1.6)	1922	1981
		1922	46	(1.2)	84	130
		1981	35	(0.9)	1290	1325
		2589	32	(0.9)	1643	1675
		2616	31	(0.8)	3268	3299
		2802	31	(0.8)	1232	1263
		2827	27	(0.7)	2589	2616
		3091	27	(0.7)	1263	1290
		3268	25	(0.7)	3302	3327
		3299	25	(0.7)	2802	2827
		3302	19	(0.5)	855	874
		3327	17	(0.5)	3402	3419
		3402	16	(0.4)	491	507
		3419	7	(0.2)	507	514
		3553	3	(0.1)	3299	3302
			3	(0.1)	218	221
FNU D2 (CGCG)						
	4					
		873	1952	(52.9)	1740	3692
		1236	873	(23.6)	1	873
		1258	482	(13.1)	1258	1740
		1740	363	(9.8)	873	1236
			22	(0.6)	1236	1258
FOK 1 (GGATG)						
	21					
		298	420	(11.4)	1523	1943
		533	304	(8.2)	3388	3692
		668	298	(8.1)	1	298
		689	297	(8.0)	1943	2240
		863	285	(7.7)	2783	3068
		1118	276	(7.5)	1118	1394
		1394	255	(6.9)	863	1118
		1445	235	(6.4)	298	533
		1457	211	(5.7)	2296	2507
		1523	174	(4.7)	2507	2681
		1943	174	(4.7)	689	863
		2240	152	(4.1)	3068	3220
		2296	135	(3.7)	533	668
		2507	133	(3.6)	3220	3353

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2681	80	(2.2)	2703	2783
		2703	66	(1.8)	1457	1523
		2783	56	(1.5)	2240	2296
		3068	51	(1.4)	1394	1445
		3220	35	(0.9)	3353	3388
		3353	22	(0.6)	2681	2703
		3388	21	(0.6)	668	689
			12	(0.3)	1445	1457
GDI 2 (QGGCCG)						
	7					
		82	1327	(35.9)	1744	3071
		875	793	(21.5)	82	875
		1342	595	(16.1)	3097	3692
		1466	467	(12.6)	875	1342
		1744	278	(7.5)	1466	1744
		3071	124	(3.4)	1342	1466
		3097	82	(2.2)	1	82
			26	(0.7)	3071	3097
HAE 1 (RGGCCR)						
	14					
		485	641	(17.4)	2201	2842
		662	520	(14.1)	1348	1868
		783	485	(13.1)	1	485
		1013	372	(10.1)	3320	3692
		1135	273	(7.4)	1928	2201
		1210	265	(7.2)	2842	3107
		1348	230	(6.2)	783	1013
		1868	177	(4.8)	485	662
		1928	138	(3.7)	1210	1348
		2201	122	(3.3)	1013	1135
		2842	121	(3.3)	662	783
		3107	117	(3.2)	3203	3320
		3203	96	(2.6)	3107	3203
		3320	75	(2.0)	1135	1210
			60	(1.6)	1868	1928
HAE 2 (PGCGCQ)						
	1					
		260	3432	(93.0)	260	3692
			260	(7.0)	1	260
HAE 3 (GGCC)						
	37					
		83	371	(10.0)	3321	3692
		154	278	(7.5)	1467	1745
		371	217	(5.9)	154	371
		486	207	(5.6)	2202	2409
		663	194	(5.3)	2649	2843
		784	177	(4.8)	486	663
		876	148	(4.0)	2924	3072
		923	135	(3.7)	1989	2124
		931	132	(3.6)	1211	1343
		1014	124	(3.4)	1745	1869
		1136	122	(3.3)	1014	1136
		1211	121	(3.3)	663	784
		1343	115	(3.1)	371	486

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		1349	102	(2.8)	2486	2588
		1421	92	(2.5)	784	876
		1467	83	(2.2)	931	1014
		1745	83	(2.2)	1	83
		1869	81	(2.2)	2843	2924
		1929	78	(2.1)	3243	3321
		1989	78	(2.1)	2124	2202
		2124	77	(2.1)	3127	3204
		2202	77	(2.1)	2409	2486
		2409	75	(2.0)	1136	1211
		2486	72	(2.0)	1349	1421
		2588	71	(1.9)	83	154
		2649	61	(1.7)	2588	2649
		2843	60	(1.6)	1929	1989
		2924	60	(1.6)	1869	1929
		3072	47	(1.3)	876	923
		3078	46	(1.2)	1421	1467
		3090	39	(1.1)	3204	3243
		3098	19	(0.5)	3108	3127
		3108	12	(0.3)	3078	3090
		3127	10	(0.3)	3098	3108
		3204	8	(0.2)	3090	3098
		3243	8	(0.2)	923	931
		3321	6	(0.2)	3072	3078
			6	(0.2)	1343	1349
HGA 1 (GACGC)						
	4					
		2417	2417	(65.5)	1	2417
		2492	767	(20.8)	2687	3454
		2687	238	(6.4)	3454	3692
		3454	195	(5.3)	2492	2687
			75	(2.0)	2417	2492
HGI A1 (GRGCRC)						
	6					
		614	1131	(30.6)	1625	2756
		985	936	(25.4)	2756	3692
		1224	614	(16.6)	1	614
		1253	372	(10.1)	1253	1625
		1625	371	(10.0)	614	985
		2756	239	(6.5)	985	1224
			29	(0.8)	1224	1253
HGI C1 (GGQPCC)						
	12					
		333	730	(19.8)	2962	3692
		414	697	(18.9)	2265	2962
		753	559	(15.1)	816	1375
		777	429	(11.6)	1713	2142
		816	339	(9.2)	414	753
		1375	333	(9.0)	1	333
		1541	172	(4.7)	1541	1713
		1713	166	(4.5)	1375	1541
		2142	87	(2.4)	2142	2229
		2229	81	(2.2)	333	414
		2265	39	(1.1)	777	816
		2962	36	(1.0)	2229	2265
			24	(0.7)	753	777

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
HGI J2 (GPGCQC)						
	7					
		902	1118	(30.3)	1805	2923
		1202	902	(24.4)	1	902
		1224	433	(11.7)	3259	3692
		1625	401	(10.9)	1224	1625
		1805	336	(9.1)	2923	3259
		2923	300	(8.1)	902	1202
		3259	180	(4.9)	1625	1805
			22	(0.6)	1202	1224
HHA 1 (GCGC)						
	6					
		261	1254	(34.0)	2027	3281
		1087	826	(22.4)	261	1087
		1235	507	(13.7)	1520	2027
		1520	411	(11.1)	3281	3692
		2027	285	(7.7)	1235	1520
		3281	261	(7.1)	1	261
			148	(4.0)	1087	1235
HIND 3 (AAGCTT)						
	1					
		564	3128	(84.7)	564	3692
			564	(15.3)	1	564
HINF 1 (GANTC)						
	9					
		46	1090	(29.5)	1106	2196
		123	739	(20.0)	2897	3636
		551	428	(11.6)	123	551
		969	418	(11.3)	551	969
		1106	369	(10.0)	2528	2897
		2196	332	(9.0)	2196	2528
		2528	137	(3.7)	969	1106
		2897	77	(2.1)	46	123
		3636	56	(1.5)	3636	3692
			46	(1.2)	1	46
HPA 2 (CCGG)						
	10					
		111	1173	(31.8)	296	1469
		296	671	(18.2)	1469	2140
		1469	467	(12.6)	3074	3541
		2140	429	(11.6)	2645	3074
		2171	240	(6.5)	2405	2645
		2191	214	(5.8)	2191	2405
		2405	185	(5.0)	111	296
		2645	151	(4.1)	3541	3692
		3074	111	(3.0)	1	111
		3541	31	(0.8)	2140	2171
			20	(0.5)	2171	2191
HPH 1 (GGTGA)						
	11					
		33	928	(25.1)	2764	3692
		203	864	(23.4)	1900	2764
		470	458	(12.4)	1026	1484
		593	289	(7.8)	737	1026

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		737	267	(7.2)	203	470
		1026	202	(5.5)	1484	1686
		1484	170	(4.6)	33	203
		1686	144	(3.9)	593	737
		1765	135	(3.7)	1765	1900
		1900	123	(3.3)	470	593
		2764	79	(2.1)	1686	1765
			33	(0.9)	1	33
KPN 1 (GGTACC)						
	3					
		753	1550	(42.0)	2142	3692
		816	1326	(35.9)	816	2142
		2142	753	(20.4)	1	753
			63	(1.7)	753	816
MBO 2 (GAAGA)						
	13					
		239	679	(18.4)	239	918
		918	624	(16.9)	2414	3038
		1001	529	(14.3)	1001	1530
		1530	461	(12.5)	3083	3544
		1859	329	(8.9)	1530	1859
		2163	304	(8.2)	1859	2163
		2401	239	(6.5)	1	239
		2414	238	(6.4)	2163	2401
		3038	97	(2.6)	3595	3692
		3083	83	(2.2)	918	1001
		3544	48	(1.3)	3547	3595
		3547	45	(1.2)	3038	3083
		3595	13	(0.4)	2401	2414
			3	(0.1)	3544	3547
MLU 1 (ACGCGT)						
	1					
		1739	1953	(52.9)	1739	3692
			1739	(47.1)	1	1739
MNL 1 (CCTC)						
	58					
		27	192	(5.2)	2126	2318
		141	189	(5.1)	2562	2751
		157	180	(4.9)	1450	1630
		317	160	(4.3)	157	317
		382	142	(3.8)	2372	2514
		462	140	(3.8)	1986	2126
		468	136	(3.7)	701	837
		518	114	(3.1)	27	141
		620	113	(3.1)	1749	1862
		701	106	(2.9)	3048	3154
		837	102	(2.8)	518	620
		845	87	(2.4)	3154	3241
		867	84	(2.3)	1862	1946
		921	81	(2.2)	2967	3048
		992	81	(2.2)	992	1073
		1073	81	(2.2)	620	701
		1090	80	(2.2)	382	462
		1121	73	(2.0)	2781	2854

	# SITES	FRAGMENTS		FRAGMENT ENDS	
	1146	71	(1.9)	921	992
	1206	66	(1.8)	1213	1279
	1213	65	(1.8)	1684	1749
	1279	65	(1.8)	317	382
	1307	62	(1.7)	1388	1450
	1357	61	(1.7)	2854	2915
	1388	60	(1.6)	3407	3467
	1450	60	(1.6)	1146	1206
	1630	56	(1.5)	3506	3562
	1634	55	(1.5)	3562	3617
	1684	54	(1.5)	2318	2372
	1749	54	(1.5)	867	921
	1862	52	(1.4)	2915	2967
	1946	51	(1.4)	3241	3292
	1949	50	(1.4)	1634	1684
	1986	50	(1.4)	1307	1357
	2126	50	(1.4)	468	518
	2318	48	(1.3)	3622	3670
	2372	48	(1.3)	2514	2562
	2514	44	(1.2)	3363	3407
	2562	39	(1.1)	3467	3506
	2751	37	(1.0)	1949	1986
	2781	31	(0.8)	1357	1388
	2854	31	(0.8)	1090	1121
	2915	30	(0.8)	2751	2781
	2967	28	(0.8)	1279	1307
	3048	27	(0.7)	3292	3319
	3154	27	(0.7)	1	27
	3241	25	(0.7)	1121	1146
	3292	22	(0.6)	3670	3692
	3319	22	(0.6)	3341	3363
	3341	22	(0.6)	3319	3341
	3363	22	(0.6)	845	867
	3407	17	(0.5)	1073	1090
	3467	16	(0.4)	141	157
	3506	8	(0.2)	837	845
	3562	7	(0.2)	1206	1213
	3617	6	(0.2)	462	468
	3622	5	(0.1)	3617	3622
	3670	4	(0.1)	1630	1634
		3	(0.1)	1946	1949
MST 2 (CCTNAGG)					
	2				
	157	2486	(67.3)	1206	3692
	1206	1049	(28.4)	157	1206
		157	(4.3)	1	157
NCI 1 (CCSGG)					
	8				
	110	1358	(36.8)	111	1469
	111	721	(19.5)	1469	2190
	1469	618	(16.7)	3074	3692
	2190	429	(11.6)	2645	3074
	2191	240	(6.5)	2405	2645
	2405	214	(5.8)	2191	2405
	2645	110	(3.0)	1	110
	3074	1	(0.0)	2190	2191
		1	(0.0)	110	111

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
NCO 1 (CCATGG)						
	2					
		250	2959	(80.1)	733	3692
		733	483	(13.1)	250	733
			250	(6.8)	1	250
NHE 1 (GCTAGC)						
	2					
		2280	2280	(61.8)	1	2280
		2744	948	(25.7)	2744	3692
			464	(12.6)	2280	2744
NLA 3 (CATG)						
	18					
		80	1038	(28.1)	2258	3296
		209	444	(12.0)	1640	2084
		251	404	(10.9)	330	734
		330	357	(9.7)	1244	1601
		734	278	(7.5)	787	1065
		758	228	(6.2)	3464	3692
		787	168	(4.6)	3296	3464
		1065	129	(3.5)	80	209
		1131	123	(3.3)	2084	2207
		1244	113	(3.1)	1131	1244
		1601	80	(2.2)	1	80
		1640	79	(2.1)	251	330
		2084	66	(1.8)	1065	1131
		2207	42	(1.1)	209	251
		2219	39	(1.1)	2219	2258
		2258	39	(1.1)	1601	1640
		3296	29	(0.8)	758	787
		3464	24	(0.7)	734	758
			12	(0.3)	2207	2219
NLA 4 (GGNNCC)						
	32					
		333	366	(9.9)	1713	2079
		369	333	(9.0)	1	333
		378	297	(8.0)	3395	3692
		414	266	(7.2)	442	708
		442	264	(7.2)	2647	2911
		708	240	(6.5)	2407	2647
		753	227	(6.1)	1148	1375
		777	172	(4.7)	1541	1713
		816	167	(4.5)	3076	3243
		903	142	(3.8)	2265	2407
		977	120	(3.3)	1421	1541
		1069	115	(3.1)	3243	3358
		1148	114	(3.1)	2962	3076
		1375	92	(2.5)	977	1069
		1384	87	(2.4)	2142	2229
		1421	87	(2.4)	816	903
		1541	79	(2.1)	1069	1148
		1713	74	(2.0)	903	977
		2079	57	(1.5)	2079	2136
		2136	45	(1.2)	708	753
		2142	39	(1.1)	2923	2962
		2229	39	(1.1)	777	816
		2265	37	(1.0)	3358	3395
		2407	37	(1.0)	1384	1421

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		2647	36	(1.0)	2229	2265
		2911	36	(1.0)	378	414
		2923	36	(1.0)	333	369
		2962	28	(0.8)	414	442
		3076	24	(0.7)	753	777
		3243	12	(0.3)	2911	2923
		3358	9	(0.2)	1375	1384
		3395	9	(0.2)	369	378
			6	(0.2)	2136	2142
NSI 1 (ATGCAT)						
	2					
		535	1720	(46.6)	1972	3692
		1972	1437	(38.9)	535	1972
			535	(14.5)	1	535
NSP B2 (CVGCWG)						
	7					
		216	1356	(36.7)	2336	3692
		512	763	(20.7)	1573	2336
		872	679	(18.4)	872	1551
		1551	360	(9.8)	512	872
		1566	296	(8.0)	216	512
		1573	216	(5.9)	1	216
		2336	15	(0.4)	1551	1566
			7	(0.2)	1566	1573
NSP C1 (PCATGQ)						
	5					
		1130	1130	(30.6)	1	1130
		1639	1038	(28.1)	2257	3295
		2206	567	(15.4)	1639	2206
		2257	509	(13.8)	1130	1639
		3295	397	(10.8)	3295	3692
			51	(1.4)	2206	2257
PFL M1 (CCANNNNNT						
	2					
		665	2415	(65.4)	665	3080
		3080	665	(18.0)	1	665
			612	(16.6)	3080	3692
PPU M1 (PGGRCCQ)						
	5					
		708	1368	(37.1)	1147	2515
		1069	1103	(29.9)	2515	3618
		1147	708	(19.2)	1	708
		2515	361	(9.8)	708	1069
		3618	78	(2.1)	1069	1147
			74	(2.0)	3618	3692
PST 1 (CTGCAG)						
	5					
		219	1326	(35.9)	1291	2617
		1185	1040	(28.2)	2652	3692
		1291	966	(26.2)	219	1185
		2617	219	(5.9)	1	219
		2652	106	(2.9)	1185	1291
			35	(0.9)	2617	2652

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
PVU 2 (CAGCTG)						
	3					
		216	2126	(57.6)	1566	3692
		1551	1335	(36.2)	216	1551
		1566	216	(5.9)	1	216
			15	(0.4)	1551	1566
RRU 1 (AGTACT)						
	1					
		935	2757	(74.7)	935	3692
			935	(25.3)	1	935
RSA 1 (GTAC)						
	10					
		754	801	(21.7)	936	1737
		817	773	(20.9)	2345	3118
		880	754	(20.4)	1	754
		936	406	(11.0)	1737	2143
		1737	336	(9.1)	3145	3481
		2143	211	(5.7)	3481	3692
		2345	202	(5.5)	2143	2345
		3118	63	(1.7)	817	880
		3145	63	(1.7)	754	817
		3481	56	(1.5)	880	936
			27	(0.7)	3118	3145
SAC 1 (GAGCTC)						
	2					
		1224	2067	(56.0)	1625	3692
		1625	1224	(33.2)	1	1224
			401	(10.9)	1224	1625
SAC 2 (CCGCGG)						
	1					
		872	2820	(76.4)	872	3692
			872	(23.6)	1	872
SAU 1 (CCTNAGG)						
	2					
		157	2486	(67.3)	1206	3692
		1206	1049	(28.4)	157	1206
			157	(4.3)	1	157
SAU 3A (GATC)						
	12					
		227	1131	(30.6)	473	1604
		458	523	(14.2)	2357	2880
		473	479	(13.0)	2880	3359
		1604	279	(7.6)	1827	2106
		1763	231	(6.3)	227	458
		1827	227	(6.1)	2106	2333
		2106	227	(6.1)	1	227
		2333	173	(4.7)	3359	3532
		2357	160	(4.3)	3532	3692
		2880	159	(4.3)	1604	1763
		3359	64	(1.7)	1763	1827
		3532	24	(0.7)	2333	2357

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
			15	(0.4)	458	473
SAU 96 (GGNCC)						
	35					
		154	304	(8.2)	370	674
		319	282	(7.6)	1421	1703
		370	273	(7.4)	1148	1421
		674	263	(7.1)	2648	2911
		709	231	(6.3)	1703	1934
		869	224	(6.1)	3395	3619
		923	165	(4.5)	2123	2288
		930	165	(4.5)	154	319
		962	160	(4.3)	709	869
		1049	154	(4.2)	1	154
		1070	153	(4.1)	2924	3077
		1148	152	(4.1)	3243	3395
		1421	120	(3.3)	2288	2408
		1703	117	(3.2)	3126	3243
		1934	92	(2.5)	1988	2080
		1988	87	(2.4)	962	1049
		2080	78	(2.1)	2408	2486
		2123	78	(2.1)	1070	1148
		2288	73	(2.0)	3619	3692
		2408	61	(1.7)	2587	2648
		2486	54	(1.5)	1934	1988
		2498	54	(1.5)	869	923
		2516	51	(1.4)	319	370
		2555	43	(1.2)	2080	2123
		2587	39	(1.1)	2516	2555
		2648	37	(1.0)	3089	3126
		2911	35	(0.9)	674	709
		2923	32	(0.9)	2555	2587
		2924	32	(0.9)	930	962
		3077	21	(0.6)	1049	1070
		3089	18	(0.5)	2498	2516
		3126	12	(0.3)	3077	3089
		3243	12	(0.3)	2911	2923
		3395	12	(0.3)	2486	2498
		3619	7	(0.2)	923	930
			1	(0.0)	2923	2924
SCA 1 (AGTACT)						
	1					
		935	2757	(74.7)	935	3692
			935	(25.3)	1	935
SCR F1 (CCNGG)						
	47					
		54	253	(6.9)	1937	2190
		110	228	(6.2)	3095	3323
		111	216	(5.9)	137	353
		137	208	(5.6)	3484	3692
		353	188	(5.1)	2886	3074
		439	178	(4.8)	781	959
		446	170	(4.6)	2716	2886
		530	162	(4.4)	1124	1286
		628	156	(4.2)	1538	1694
		665	143	(3.9)	981	1124
		781	116	(3.1)	3332	3448
		959	116	(3.1)	665	781
		974	100	(2.7)	2191	2291

	# SITES	FRAGMENTS		FRAGMENT ENDS	
	981	98	(2.7)	530	628
	1124	95	(2.6)	1286	1381
	1286	90	(2.4)	1802	1892
	1381	88	(2.4)	1381	1469
	1469	86	(2.3)	353	439
	1505	84	(2.3)	446	530
	1526	78	(2.1)	1694	1772
	1538	75	(2.0)	2570	2645
	1694	71	(1.9)	2645	2716
	1772	57	(1.5)	2348	2405
	1802	56	(1.5)	54	110
	1892	54	(1.5)	1	54
	1931	51	(1.4)	2519	2570
	1937	49	(1.3)	2434	2483
	2190	41	(1.1)	2307	2348
	2191	39	(1.1)	1892	1931
	2291	37	(1.0)	628	665
	2307	36	(1.0)	3448	3484
	2348	36	(1.0)	1469	1505
	2405	30	(0.8)	1772	1802
	2434	29	(0.8)	2405	2434
	2483	27	(0.7)	2483	2510
	2510	26	(0.7)	111	137
	2519	21	(0.6)	3074	3095
	2570	21	(0.6)	1505	1526
	2645	16	(0.4)	2291	2307
	2716	15	(0.4)	959	974
	2886	12	(0.3)	1526	1538
	3074	9	(0.2)	3323	3332
	3095	9	(0.2)	2510	2519
	3323	7	(0.2)	974	981
	3332	7	(0.2)	439	446
	3448	6	(0.2)	1931	1937
	3484	1	(0.0)	2190	2191
		1	(0.0)	110	111
SDU 1 (G2GC3C)					
16					
	413	569	(15.4)	2187	2756
	614	433	(11.7)	3259	3692
	776	413	(11.2)	1	413
	902	382	(10.3)	1805	2187
	985	372	(10.1)	1253	1625
	1202	296	(8.0)	2963	3259
	1224	217	(5.9)	985	1202
	1253	201	(5.4)	413	614
	1625	167	(4.5)	2756	2923
	1712	162	(4.4)	614	776
	1805	126	(3.4)	776	902
	2187	93	(2.5)	1712	1805
	2756	87	(2.4)	1625	1712
	2923	83	(2.2)	902	985
	2963	40	(1.1)	2923	2963
	3259	29	(0.8)	1224	1253
		22	(0.6)	1202	1224
SFA N1 (GATGC)					
15					
	534	614	(16.6)	2544	3158
	690	534	(14.5)	3158	3692
	892	534	(14.5)	1	534

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
		954	321	(8.7)	1522	1843
		1117	303	(8.2)	1974	2277
		1218	249	(6.7)	2295	2544
		1393	202	(5.5)	690	892
		1522	175	(4.7)	1218	1393
		1843	163	(4.4)	954	1117
		1942	156	(4.2)	534	690
		1974	129	(3.5)	1393	1522
		2277	101	(2.7)	1117	1218
		2295	99	(2.7)	1843	1942
		2544	62	(1.7)	892	954
		3158	32	(0.9)	1942	1974
			18	(0.5)	2277	2295
SMA 1 (CCCGGG)						
	2					
		110	2080	(56.3)	110	2190
		2190	1502	(40.7)	2190	3692
			110	(3.0)	1	110
SPH 1 (GCATGC)						
	3					
		1639	1639	(44.4)	1	1639
		2257	1038	(28.1)	2257	3295
		3295	618	(16.7)	1639	2257
			397	(10.8)	3295	3692
SSP 1 (AATATT)						
	1					
		717	2975	(80.6)	717	3692
			717	(19.4)	1	717
STU 1 (AGGCCT)						
	2					
		1210	1824	(49.4)	1868	3692
		1868	1210	(32.8)	1	1210
			658	(17.8)	1210	1868
STY 1 (CCRRGG)						
	7					
		150	1986	(53.8)	1706	3692
		250	780	(21.1)	926	1706
		418	257	(7.0)	476	733
		476	193	(5.2)	733	926
		733	168	(4.6)	250	418
		926	150	(4.1)	1	150
		1706	100	(2.7)	150	250
			58	(1.6)	418	476
TAQ 1 (TCGA)						
	4					
		895	1870	(50.7)	1369	3239
		949	895	(24.2)	1	895
		1369	453	(12.3)	3239	3692
		3239	420	(11.4)	949	1369
			54	(1.5)	895	949

	#	SITES	FRAGMENTS		FRAGMENT ENDS	
TTH111 1 (GACNNNG						
	1					
		2663	2663	(72.1)	1	2663
			1029	(27.9)	2663	3692
TTH111 2 (CCAPCA)						
	9					
		358	1390	(37.6)	1732	3122
		482	557	(15.1)	488	1045
		488	422	(11.4)	1310	1732
		1045	358	(9.7)	1	358
		1061	343	(9.3)	3349	3692
		1310	249	(6.7)	1061	1310
		1732	227	(6.1)	3122	3349
		3122	124	(3.4)	358	482
		3349	16	(0.4)	1045	1061
			6	(0.2)	482	488
XHO 1 (CTCGAG)						
	1					
		3238	3238	(87.7)	1	3238
			454	(12.3)	3238	3692
XHO 2 (PGATCQ)						
	4					
		226	2653	(71.9)	226	2879
		2879	479	(13.0)	2879	3358
		3358	226	(6.1)	1	226
		3531	173	(4.7)	3358	3531
			161	(4.4)	3531	3692
XMA 3 (CGGCCG)						
	1					
		875	2817	(76.3)	875	3692
			875	(23.7)	1	875

The following do not appear:

AFL 2	AHA 3	AOS 1	BGL 2
BSS H2	CLA 1	ECO R5	HINC 2
HPA 1	MST 1	NAE 1	NAR 1
NDE 1	NOT 1	NRU 1	PVU 1
RSR 2	SAL 1	SFI 1	SNA 1
SNA B1	SPE 1	XBA 1	XMN 1

LIST OF REFERENCES

LIST OF REFERENCES

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

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

3. Griffin, L.D., Gelb, B.D., Wheeler, D.A., Davison, D., Adams, V., and McCabe, E.R.B. (1991) *Genomics* **11**, 1014-1024.

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

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

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

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

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

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

10. Tanizawa, Y., Koranyi, L.I., Welling, C.M., and Permutt, M.A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7294-7297.

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

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

13. Frohlich, K., Entian, K., Mecke, D. (1985) *Gene* **36**, 105-111.

14. Albig, W., and Entian, K.D. (1988) *Gene* **73**, 141-152.

15. Personal communication from Dr. Charles Shoemaker, School of Tropical Medicine, Harvard University.

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

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

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

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

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

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

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

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

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

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

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

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

28. Sols, A. and Crane, R.K., (1954) *J. Biol. Chem.* **210**, 581-595.

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

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

31. Ureta, T. (1975) *in* Isozymes (Markert, C.L., Ed.), Vol. III, pp. 575-602, Academic Press, New York, NY.

32. Magnani, M., Stocchi, V., Serafini, N., Piatti, E., Dacha, M., and Fornaini, G. (1983) *Arch. Biochem. Biophys.* **226**, 377-387.

33. Siano, D.B., Zyskind, J.W., and Fromm, H.J. (1975) *Arch. Biochem. Biophys.* **170**, 587-600.

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

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

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

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

38. Preller, A., and Wilson, J.E. (1992) *Arch. Biochem. Biophys.* **294**, 482-492.

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

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

41. Ashcroft, F.M., Harrison, D.E., and Ashcroft, S.J.H. (1984) *Nature* **312**, 446-448.

42. Nelson, T.Y., Gaines, K.L., Rajan, A.S., Berg, M., and Boyd III, A.E. (1987) *J. Biol. Chem.* **262**, 2608-2612.

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

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

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

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

47. Xie, G. and Wilson, J.E. (1988) *Arch. Biochem. Biophys.* **276**, 285-293.

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

49. Felgner, P.L., Messer, J.L., and Wilson, J.E. (1979) *J. Biol. Chem.* **260**, 4946-4949.

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

51. Ureta, T., Bravo, R., and Babul,J. (1976) *Enzyme* **20**, 334-348.

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

53. Purich, D.L., Fromm, H.J., and Rudolph, F.B. (1973) *Adv. Enzymol.* **39**,249-326.

54. Anderson, C., McDonald, R., and Steitz, T. (1978) *J. Mol. Biol.* **123**, 1-13.

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

56. Anderson, C., McDonald, R., and Steitz, T. (1978) *J. Mol. Biol.* **123**, 15-33.

57. Bennett, W., and Steitz, T. (1980) *J. Mol. Biol.* **140**, 183-209.

58. Bennett, W., and Steitz, T. (1980) *J. Mol. Biol.* **140**, 211-230.

59. Steitz, T., Anderson, W., Fletterick, R., and Anderson, C. (1977) *J. Biol. Chem.* **252**, 4494-4500.

60. Bennett, W., and Steitz, T. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4848-4852.

61. Anderson, C.M., Zucker, F.H., and Steitz, T.A. (1979) *Science* **204**, 375-380.

62. Lobo, Z., and Maitra, P. (1977) *Arch. Biochem. Biophys.* **182**, 639-645.

63. Maitra, P. (1975) *Methods Enzymol.* **42**, 25-30.

64. Rossmann, M., and Argos, P. (1977) *J. Mol. Biol.* **109**, 99-129.

65. McLachlan, A. (1979) *Eur. J. Biochem.* **100**, 181-187.

66. Harrison, R. (1985) *Crystallographic Refinement of Two Isozymes of Yeast Hexokinase and Relationship of Structure to Function*. Ph.D. thesis, Yale University, New Haven, CT.

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

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

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

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

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

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

73. Schirch, D.M. and Wilson, J.E. (1987) *Arch. Biochem. Biophys.* **259**, 402-411.

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

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

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

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

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

79. Hutny, J., and Wilson, J.E. (1990) *Arch. Biochem. Biophys.* **283**, 173-183.

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

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

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

83. Pittler, S.J., Kozak, L.P., and Wilson, J.E. (1985) *Biochim. Biophys. Acta* **843**, 186-192.

84. Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).

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

86. Feinberg, A.P., and Vogelstein, B. (1984) *Anal. Biochem.* **137**, 266-267.

87. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

88. Henikoff, S. (1984) *Gene* **28**, 351-359.

89. Henikoff, S. (1987) in *Promega Notes*, No. 8, pp. 1-3, Promega Corp., Madison, WI

90. Takahara, M., Hibler, D., Barr, p., Gerlt, J., and Inouye, M. (1985) *J. Biol. Chem.* **260**, 2670-2674.

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

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

93. Wickens, M. and Stephenson, P. (1984) *Science* **226**, 1045-1051.

94. Creighton, T.E. (1983) *Proteins*, pp. 252-262, Freeman Publications, New York.

95. Lesk, A. M., and Chothia, C. (1986) *Philos. Trans. R. Soc. London B* **317**, 345-356.

96. Chothia, C., and Lesk, A.M. (1986) *EMBO J.* **5**, 823-826.

97. Craik, C.S., Rutter, W.J., and Fletterick, R. (1983) *Science* **220**, 1125-1129.

98. Steitz, T.A., Fletterick, R.J., Anderson, W.F., and Anderson, C.M. (1976) *J. Mol. Biol.* **104**, 197-222.

99. Smith, A.D. and Wilson, J.E., (1991) *Arch. Biochem. Biophys.* **287**, 359-366.

100. Smith, A.D. and Wilson, J.E. (1991) *Arch. Biochem. Biophys.*

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

102. Barnell, W., Yi, K.C., and Conway, T. (1990) *J. Bacteriology* **172**, 7227-7240.

103. Keim, P., Heinrickson, R.L., and Fitch, W.M. (1981) *J. Mol. Biol.* **151**, 179-197.

104. Rossman, M.G., Liljas, A., Branden, C.I., and Banaszak, L.J. (1975) *in* The Enzymes (Boyer,P.D., Ed.), 3rd ed., Vol. 11, pp. 61-102, Academic Press, New York.

105. Weber, I.T., Takio, K., Titasni, K., and Steitz, T.A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7979-7983.

106. Rudolph, F.B., and Fromm, H.J. (1971) *J. Biol. Chem.* **246**, 6611-6619.

107. Noat, G., Richard, J., Borel, M., and Got, C. (1970) *Eur. J. Biochem.* **13**, 347-360.

108. Branden, C., and Tooze, J. (1991) *Introduction to Protein Structure*, pp. 141-159, Garland Publishing, Inc., New York.

109. Hanks, S.K., Quinn, A.M., and Hunter, T. (1988) *Science* **241**, 42-52.

110. Hol, W.G.J. (1985) *Prog. Biophys. Mol. Biol.* **45**, 149-195.

111. Arora, K.K., Shenbagamurthi, P., Fanciulli, M., and Pedersen, P.L. (1990) *J. Biol. Chem.* **265**, 5324-5328.

112. Fry, D.C., Kuby, S.A., and Mildvan, A.S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 907-911.

113. Tamura, J.K., LaDine, J.R., and Cross, R.L. (1988) *J. Biol. Chem.* **263**, 7907-7912.

114. Bork, P., Sander, C., and Valencia, A. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7290-7294.

115. Pollard, T.D., and Cooper, J.A. (1986) *Ann. Rev. Biochem.* **55**, 987-1035.

116. Carlier, M. (1991) *J. Biol. Chem.* **266**, 1-4.

117. Hennessey, E.S., Drummond, D.R., and Sparrow, J.C. (1993) *Biochem. J.* **282**, 657-671.

118. Gething, M.J., and Sambrook, J. (1992) *Nature* **355**, 33-42.

119. Ingolia, T.D., and Craig, E.A. (1982) *Proc. Natl. Acad. Sci USA* **79**, 525-529.

120. Schlossman, D.M., Schmid, S.L., Braell, W.A., and Rothman, J.E. (1984) *J.Cell Biol.* **99**, 723-733.

121. Chirico, W.J., Waters, M.G., and Blobel, G. (1988) *Nature* **332**, 805-810.

122. Flaherty, K.M., DeLuca-Flaherty, C, and McKay, D.B. (1990) *Nature* **346**, 623-628.

123. Flaherty, K.M., McKay, D.B., Kabsch, W., and Holmes, K.C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 5041-5045.

124. Kabsch, W., Mannherz, H.G., Suck, D., Pai, E.F., and Holmes, K.C. (1990) *Nature* **347**, 37-44.

125. Hurley, J.H., Faber, H.R., Worthylake, D., Meadow, N.D., Roseman, S., Pettigrew, D.W., and Remington, S.J. (1993) *Science* **259**, 673-677.

126. Levitt, M., and Chothia, C. (1976) *Nature* **261**, 552-557.

127. Branden, C.I. (1980) *Q. Rev. Biophys.* **13**, 317-338.

128. Orr, G.A., Simon, J., Jones, S.R., Chin, G.J., Knowles, J.R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2230-2233.

129. Blattler, W.A., Knowles, J.R. (1979) *J. Am. Chem. Soc.* **101**, 510-511.

130. Ghrayeb, J., Kimura, H., Takahara, M., Hsiung, H., Masui, Y., and Inouye, M. (1984) *EMBO J.* **3**, 2437-2442.

131. Takahara, M., Sagai, H., Inouye, S., and Inouye, M. (1988) *Biotechnology* **6**,195-198.

132. Swamy, K.H.S., and Goldberg, A.L. (1982) *J. Bacteriol.* **149**, 1027.

133. Better, M., Chang, C.P., Robinson, R.R., Horwitz, A.H. (1988) *Science* **240**, 1041.

134. Pollitt, S., and Zalkin, H. (1983) *J. Bacteriol.* **153**, 27.

135. Duffaud, G.D., March, P.E., and Inouye, M. (1987) *Methods Enzymol.* **153**, 492-507.

136. Neu, H.C., and Heppel, L.A. (1965) *J. Biol. Chem.* **240**, 3685-3692.

137. Schein, C.H., and Noteborn, M.H.M. (1988) *Biotechnology* **6**, 291-294.

138. Takagi, H., Morinaga, Y., Tsuchiya, M., Ikemura, H., and Inouye, M. (1988) *Biotechnology* **6**, 948-950.

139. Fromm, H.J. (1981) *in* The Regulation of Carbohydrate Formation and Utilization in Mammals (C.M. Veneziale, Ed.), University Park Press, Baltimore.