# EXPRESSION PATTERN AND GENOMIC ORGANIZATION OF HUMAN 3-PHOSPHOGLYCERATE DEHYDROGENASE

by

Jill Mauthe

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science

Department of Biochemistry and Medical Genetics
Faculty of Medicine

University of Manitoba

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

# EXPRESSION PATTERN AND GENOMIC ORGANIZATION OF HUMAN 3-PHOSPHOGLYCERATE DEHYDROGENASE

by

Jill Mauthe

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of Manitoba in partial fulfillment of the requirement of the degree

of

Masters of Science

## Acknowledgments

I would like to thank my supervisor, Steve Pind, for his valuable leadership, for challenging me to learn all I could and to grow as both a scientist and as a person, and for his neverending support in all my endeavours.

And to Elzbieta Slominski, many thanks for graciously sharing your knowledge of laboratory techniques with me. You're a talented technician and I'm glad I got the opportunity to learn from you.

To my committee members, Terry Zelinski and David Eisenstat, thank you for your advice on my labwork and thesis. Terry, thank you for always being a vocal advocate of my efforts (in the lab and out). David, thank you for welcoming me into your laboratory and sharing your knowledge of immunohistochemistry and brain physiology.

Thank you to the other collaborators who have shared their reagents, knowledge or advice, including but not limited to Dr. Marvin Natowicz and staff, Dr. Barbara Triggs-Raine, Dr. David Eisenstat, and Dr. Marc Del Bigio.

The students who have spent time in the Pind laboratory during my time there, Jackie Ching and Anjali Ghandi, thank you for your friendship and the opportunity to share my knowledge.

To the many smiling faces and inquiring minds in the department, thank you for giving me valuable advice when I asked, and for applauding my efforts when things were going slow. Your friendship is appreciated and thank you for making this a wonderful time in my life.

Many thanks go to my family and friends for their neverending support and love, especially my parents and grandmother.

As well, my thanks to Manitoba Health Research Council and Childrens Hospital Foundation for funding opportunities.

**Table of Contents**

**List of Figures**

## List of Tables

## List of Abbreviations

| | |
|---|---|
| BAC | bacterial artificial chromosome |
| BLAST | basic local alignment search tool |
| CIHR | Canadian Institutes for Health Research |
| $ddH_2O$ | distilled deionized water |
| DNase | deoxyribonuclease I |
| dNTP | deoxynucleotide triphosphate |
| EDTA | ethylenediaminetetraacetic acid |
| LB | Luria-Bertani media/plates |
| OMIM | Online Mendelian Inheritance in Man database |
| PBS | phosphate buffered saline |
| PCR | polymerase chain reaction |
| PHGDH | 3-phosphoglycerate dehydrogenase |
| RACE | rapid amplification of cDNA ends |
| RNaseA | ribonuclease A |
| rpm | rotations per minute |
| RT-PCR | reverse-transcriptase-polymerase chain reaction |
| SDS | sodium dodecyl sulfate |
| SSC | saline sodium citrate |

| | |
|---|---|
| TAE | Tris acetate EDTA |
| TBE | Tris borate EDTA buffer |
| TBEs | Tris borate EDTA for sequencing |
| TE | Tris EDTA buffer |
| TEMED | N,N,N',N'-tetramethylethylenediamine |
| Tris-HCl | tris(hydroxymethyl)aminomethane hydrochloride |
| X | times (e.g. 10X) |

# ABSTRACT

A deficiency of 3-phosphoglycerate dehydrogenase (PHGDH) results in congenital microcephaly, severe psychomotor retardation, and epilepsy. As PHGDH catalyzes the first step in the biosynthesis of L-serine, deficient patients are characterized by abnormally low levels of this amino acid, especially in their cerebrospinal fluid. Two point mutations have been described in the coding sequence of *PHGDH*, but no information on the structure and expression of the human gene was available at the onset of this thesis project.

Prior to my studies, two children of an Ashkenazi Jewish family were diagnosed with a deficiency of PHGDH and the causative mutation was determined to be a 1468G→A substitution. To analyze whether this mutant allele occurs more frequently in the Ashkenazi Jewish population, the genomic DNA of over 800 Jewish and non-Jewish individuals was screened for the 1468A allele using the polymerase chain reaction (PCR) and an endonuclease digestion assay. One heterozygous individual of Jewish descent was detected, suggesting that this allele is rare, but does exist outside of the affected family.

The expression pattern of PHGDH was investigated at the mRNA and protein level. High levels of the mRNA encoding PHGDH were detected in many tissues, including several regions of the brain. Preliminary studies using immunohistochemistry

support a cytoplasmic localization of the PHGDH protein within glial cell populations in the human brain.

The genomic organization of *PHGDH* was characterized using a polymerase-chain reaction-based cloning and sequencing method, which revealed that the gene spans approximately 33 kb with twelve exons and eleven introns. Sequences at these junctions were determined. The transcriptional intitiation sites of human liver *PHGDH* were also mapped to four sites between 82 and 88 bp upstream of the translational start site. Similarity between the rat and human *PHGDH* genes suggested that alternative splicing seen in exon 1 of the rat gene might also occur in human tissues. A PCR-based assay of cDNA pools was performed, and none of the 24 human tissues tested displayed this alternatively spliced product. These results establish the expression profile and genomic organization of human 3-phosphoglycerate dehydrogenase.

# 1. INTRODUCTION

A deficiency of 3-phosphoglycerate dehydrogenase (PHGDH) is a recently described inborn error of metabolism that results in congenital microcephaly, epilepsy, and mental retardation (1). As one of three enzymes responsible for the *de novo* biosynthesis of L-serine, a deficiency of PHGDH results in a lowered ability to synthesize this amino acid. The resulting serine insufficiency impacts upon protein, lipid and nucleotide metabolism, with dramatic effects upon the central nervous system. The role of PHGDH in the development and function of the central nervous system is not well understood, despite the recent characterization of two disease-causing point mutations in this gene (2;3). Further research of the human gene and protein should increase our understanding of this disease and our ability to diagnose and treat affected children.

This thesis introduction will begin with a discussion of microcephaly and the identification of a deficiency of PHGDH as one cause of this phenotype. This is followed by a review of the metabolism of serine in mammalian cells. As the experimental portion of this thesis will focus on PHGDH, the introduction will conclude with a review of PHGDH, including its function, expression pattern, regulation and structure/function relationships.

The relative infancy of PHGDH research at the outset of this thesis project meant that a wide variety of investigations could be undertaken. As a result, the three major objectives of this thesis cover distinct and separate areas of exploration. They are: (1) a

population study to screen for the allele encoding the V490M mutation in the Ashkenazi Jewish population, (2) the expression profile of PHGDH in human tissues, and (3) the structure and expression of the *PHGDH* gene. These objectives seemed the logical places to begin our investigation into PHGDH and serine metabolism. During the course of these experiments similar experiments were performed and reported by others in this field. In this introduction, I have described the current state of knowledge, rather than the state that existed in 1999 when I began these studies.

## 1.1 Microcephaly is one phenotype common to a multitude of etiologies

Microcephaly is a phenotype that is based on external signs and often correlates with errors in brain development. Physicians routinely examine a child's head size at birth and during the first few years of life. In practice, the child's head is measured at the circumference around the base of the head and the forehead (occipitofrontal circumference). A diagnosis of microcephaly is made when this measurement is three or more standard deviations below the mean for age and sex (4;5). A small brain is often the result of hidden abnormalities in brain development (6), such as internal calcifications (hardening of the brain tissues) (5;7), or small or absent brain structures (gyri, corpus callosum, or level of myelination) (4;5;7).

Although intelligence is not directly linked to brain size, the majority of individuals with microcephaly suffer from mental retardation and psychomotor problems

(8). The association between microcephaly and intelligence is believed to be due to the underlying pathology of the brain (5;6;8); those with normal intelligence often have small brains with no abnormalities, whereas those with lowered intelligence often have atrophy or dysmyelination (7). It is this abnormal pathology that influences the neurological outcome of microcephalic individuals more than the physical measurement of head size (6;8). In addition to the children affected with microcephaly, Qazi and Reed noted that a proportion of normocephalic parents and siblings of these children are also affected with lowered intelligence (9). They concluded that the gene mutation involved in these cases may act in a heterozygous state. The idea of a single gene causing microcephaly has been disproven in the 25 years since this paper was published, but the theory that certain genetic mutations may influence phenotype in the heterozygous state is still valid.

The severity of microcephaly varies widely based on timing, etiology, inheritance pattern, and underlying structural abnormalities. The insults responsible for congenital microcephaly occur early in gestation, and the microcephaly is evident at birth (5;6). The most vulnerable period appears to be the first half of the second trimester of pregnancy (4;10), when neurons proliferate and migrate (5;6;8). Incidents in the last trimester of gestation, during birth, or during early infancy, lead to secondary microcephaly, where the head size is normal at birth but then lags behind the growth of the body (4). Secondary microcephaly can be the result of a lack of oxygen during birth or serious infection in the neonatal period (4;5). The structural pathology of the brain in these cases

often includes cystic degeneration or scarring (5), and the anoxic period or infectious agent responsible can also affect other tissues (5;10).

The insults resulting in congenital microcephaly can be split into a variety of categories, including teratogenic, syndromic, and genetic. Teratogenic insults include intrauterine infections (rubella, cytomegalovirus, human immunodeficiency virus), maternal alcohol or drug use, radiation, or maternal metabolic disorders (phenylketonuria, uncontrolled diabetes mellitus) (4;10). Syndromic insults are chromosomal aberrations, such as trisomies 17, 18 and 21, and chromosome 18 deletions, that are responsible for syndromes with congenital microcephaly as one of many symptoms (4;5). When obvious teratogenic insults and chromosomal alterations can be ruled out, or when more than one affected individual is found in a family, a diagnosis of genetic microcephaly is considered (5).

Genetic microcephaly results from mutations in single genes, but most of the genes involved remain unknown. Several genes are probably responsible, as the genetic mutations occur with autosomal recessive, autosomal dominant, and X-linked inheritance patterns (4;10). Genetic microcephalies have mainly neurological symptoms, whereas teratogenic insults often affect other tissues (eye, heart, ear) as well as the brain (4;10).

Although autosomal dominant and X-linked forms of genetic microcephaly exist, they are not as prevalent as the many forms of autosomal recessive microcephaly (5). Autosomal recessive microcephaly usually presents a more severe phenotype, as the

affected individuals often have mild to severe mental retardation, and varying levels of psychomotor retardation, spasticity leading to paraplegia or quadriplegia, and seizures. Originally, the first reports on microcephaly focused on one phenotype, which was called "true microcephaly" or *microcephaly vera*. This phenotype involved patients with mild to moderate mental retardation but a preserved and happy personality, large ears, sloping forehead, without seizures or spasticity (11). If a patient failed to fit this description, it was believed that their microcephaly was not genetic in etiology. This classical phenotype dominated the early literature and only in the last two decades have papers appeared describing siblings with microcephaly that did not fit this description. These siblings often suffer from epileptic seizures, spasticity of all limbs, nystagmus (roving eyes), and severe psychomotor retardation (12-14). These autosomal recessive microcephalies are often present in families and cultures that condone consanguineous marriages. Although several loci associated with autosomal recessive microcephaly in the human genome have been identified through linkage studies, the genes and proteins they encode remain unknown. These loci have been given the designation *MCPH* and reside on separate chromosomes: *MCPH*1 on 8p22-8pter (15), *MCPH*2 on 19q (16), *MCPH*3 on 9q34 (17), *MCPH*4 on 15q15-q21 (18), and *MCPH*5 on 1q31 (19;20). As was described above, a deficiency of PHGDH has also been identified as a cause of congenital microcephaly (1). This association was made following biochemical rather than genetic studies and remains the only known cause of congenital microcephaly at this time.

## 1.2 A deficiency of PHGDH as a cause of congenital microcephaly

A deficiency of PHGDH as an inborn error of metabolism found in patients with congenital microcephaly and severe neurological abnormalities was first reported by Jaeken *et al.* in 1996 (1). This deficiency was detected in a sibling pair born to a consanguineous Turkish couple. In the children, low levels of serine in the serum (fasting state) and very low levels in the cerebrospinal fluid of the children suggested that there was a perturbation in the biosynthesis of L-serine. Fibroblasts of the patients were cultured and enzyme assays were performed on the cell lysates, revealing PHGDH enzyme activities of only approximately 20% of normal levels (1) (Online Mendelian Inheritance in Man database (21), OMIM: 601815). The clinical portrait included severe developmental delays, hypertonia, growth retardation, and frequent seizures unresponsive to anti-convulsant therapies (22). Magnetic resonance imaging revealed cortical atrophy and hypomyelination (22;23). Five additional children have been reported in the literature with similar phenotypes sometimes including involuntary eye movements, pale optic disks, feeding difficulties, and hyperexcitability (22). Oral L-serine replacement therapy increased the level of serine in the cerebrospinal fluid and reduced the frequency and severity of the epileptic seizures (24). The amount of serine in the replacement therapy requires careful monitoring, and the addition of glycine was reported to be necessary to reduce the seizures in one child with very low enzyme activity (25). In some children the severity of microcephaly was reduced with therapy, and an improvement in alertness and

eye movements was also seen (24). As well, increased myelination has been observed following treatment, as detected by magnetic resonance imaging studies on two of the affected children (23). Recently, adverse reactions to the oral therapy were noted in one sibling pair, when high doses of serine (more than twice the level previously described) caused an arrest of head growth and decreased other neutral amino acids in the cerebrospinal fluid (26). While the reports of reduced symptoms and positive responses to therapy remain encouraging, it should be remembered that oral L-serine replacement therapy is not a cure for the deficiency.

One family of Ashkenazi Jewish descent, with two affected children, was brought to our attention in 1997. The clinical portrait of these siblings was very similar to the previous reports, with profound psychomotor retardation, spasticity, seizures, pale optic disks, hyperexcitability, and congenital microcephaly (personal correspondence with Dr. M. Natowicz). The RNA from cultured fibroblasts from the family was reverse transcribed and the resulting cDNA was amplified by the polymerase chain reaction (PCR), cloned into a plasmid, and sequenced. A 1468G→A transition was identified, encoding a valine to methionine substitution at position 490 in the PHGDH protein (V490M) (3). This mutation was present in a heterozygous state in the parents and a homozygous state in the children, supporting an autosomal recessive inheritance pattern associated with this deficiency. Further experiments were performed in our laboratory, including transfection of the cDNA into HeLa cells and transformation into *E. coli* cells.

The HeLa cells were used as a mammalian expression system for testing the enzyme activity and the *E. coli* cells were used to overexpress the protein for purification purposes.

During the course of this project, a report appeared that confirmed our results concerning the V490M mutation and indicates that this mutation is common in other ethnic populations. Klomp *et al.* (2) identified two mutations in the seven previously reported children; the V490M mutation we had detected in the Ashkenazi Jewish family was also present in six of the children, and a V425M mutation was present in the seventh child. To date, the children reported are from diverse ethnic groups. Two sibling pairs are the children of consanguineous Turkish parents (1;24), although the two families are unrelated. One family with two affected children was classified as Western European, with no information on ethnicity or consanguinity (2). A seventh child was born to consanguineous Moroccan parents (25). A fourth sibling pair who experienced adverse effects of the high dose serine therapy has only recently been reported, with no information provided regarding ethnicity, consanguinity status, or the mutation involved (26). The low number of children diagnosed with this deficiency is believed to underrepresent the actual number of affected children, as the deficiency was only recently described and testing methods are not routine and include invasive procedures, relying upon cerebrospinal fluid sampling and skin biopsies. Without definitive symptoms, it is

only recently that children with congenital microcephaly and intractable seizures are being tested for low levels of serine in their cerebrospinal fluid.

Children of consanguineous matings are believed to have a higher incidence of several autosomal recessive diseases through identity by descent (27;28). Close relatives may carry mutations inherited from a common ancestor, and if the relatives are carriers, their children will have a much higher risk (25%) of inheriting two identical copies of a gene than children born to a non-consanguineous union. Since the mutation frequencies for most autosomal recessive diseases is low in the general population, inheriting two copies of a mutation from unrelated individuals occurs less frequently. Unlike the families described previously, the Ashkenazi Jewish family in our study is non-consanguineous, suggesting that the 1468G→A mutant PHGDH allele may be present in the Ashkenazi Jewish population and not isolated in one family. The Ashkenazi Jewish community has a higher incidence of other inborn errors of metabolism, such as Tay-Sachs disease (OMIM: 272800), Gaucher disease (OMIM: 230800), Canavan disease (OMIM: 271900), and Nieman Pick disease (OMIM: 257200). Tay-Sachs disease is a deficiency of the hexosaminidase A enzyme that leads to a buildup of the GM2 gangliosides in the cells (especially brain cells) of affected children, who rarely live to age 5 (OMIM: 272800). The estimated frequency of carriers in the Jewish population of New York City is 0.013-0.016, or about one hundred times higher than the incidence in non-Jewish New York citizens (OMIM: 272800). Another disease with over two-thirds of

American cases involving Ashkenazi Jewish individuals is Gaucher disease, a glucocerebrosidase deficiency also affecting the brain (OMIM: 230800). It has been estimated that one of every sixteen Ashkenazi Jewish individuals is a carrier for Gaucher disease (OMIM: 230800). It is one of the goals of this thesis to examine whether the allele 1468A, responsible for the PHGDH deficiency in the Ashkenazi Jewish family, is present at an increased frequency in a population of Ashkenazi Jewish individuals.

## 1.3 Pathways of serine biosynthesis in mammalian cells

The major route of serine synthesis involves three enzymes and begins with the 3-phosphoglycerate molecule from glycolysis (see Figure 1). The first reaction is the oxidation of 3-phosphoglycerate to 3-phosphohydroxypyruvate, with a concomitant reduction of nicotinamide adenine dinucleotide ($NAD^+$) to its reduced form, NADH. This reaction is reversible and is catalyzed by PHGDH (E.C. 1.1.1.95), an NAD-dependent oxidoreductase. The second step, also reversible, is catalyzed by phosphoserine aminotransferase (E.C. 2.6.1.52), and involves adding an amino group from glutamate to phosphohydroxypyruvate, resulting in 2-oxoglutarate and 3-phosphoserine. The final step is the dephosphorylation of 3-phosphoserine to yield L-serine, catalyzed by phosphoserine phosphatase (E.C. 3.1.3.3). It has been estimated that in rats, this pathway produces serine at a rate of 370 mg/100 g body weight per day (29). This pathway is the major source of *de novo* serine biosynthesis in mammalian cells.

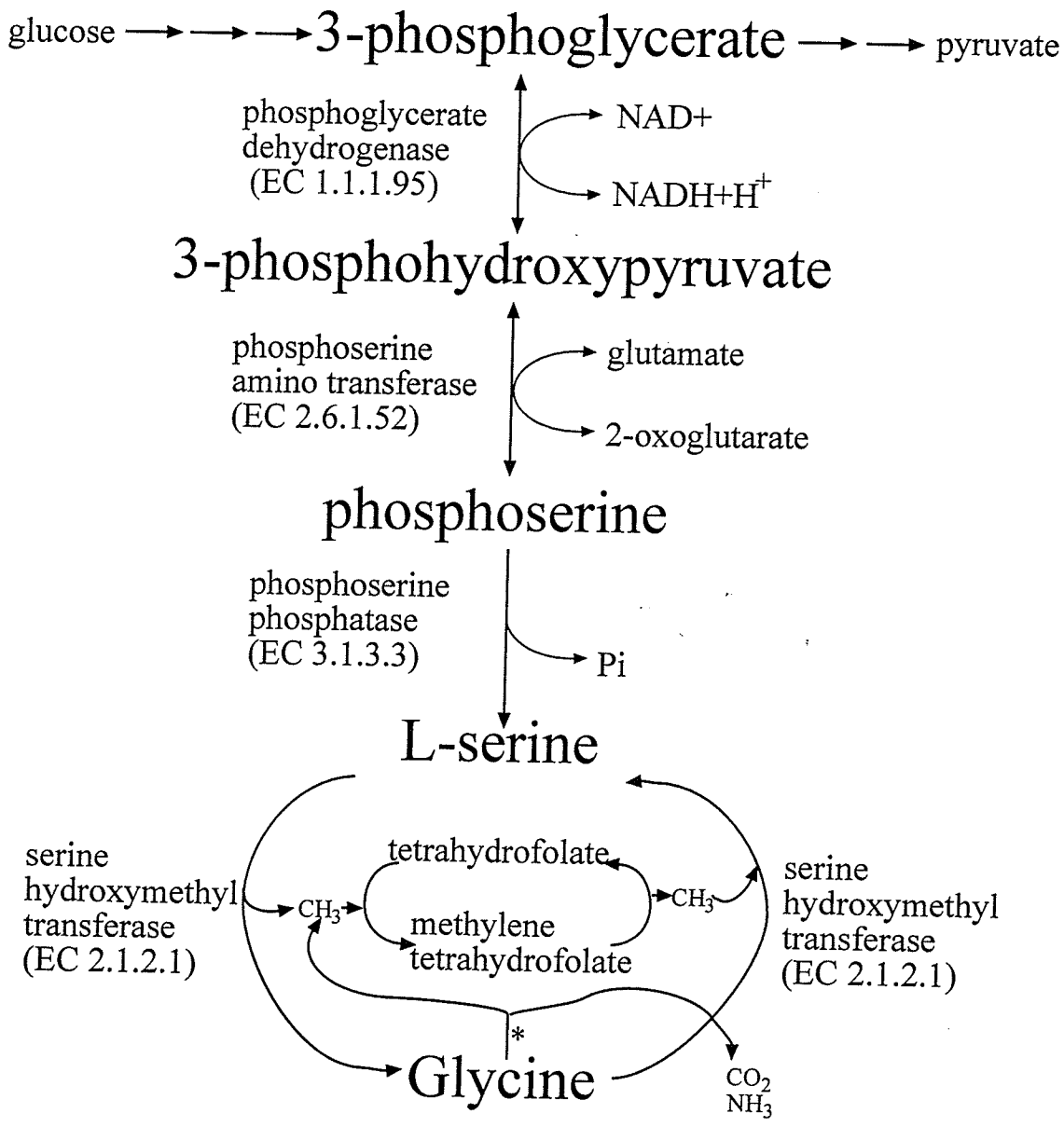



Figure 1. The biosynthetic pathway of L-serine.
The major source of serine in cells is the *de novo* biosynthesis from carbohydrate 3-phosphoglycerate. A possible alternative pathway involves the conversion of two molecules of glycine into one molecule of serine and byproducts (carbon dioxide and ammonia) by the glycine cleavage enzymes (shown by the *) and serine hydroxymethyl transferase enzymes.

A secondary pathway of serine synthesis involves the conversion of two molecules of glycine to form one molecule of serine (see the bottom half of Figure 1). The glycine cleavage enzymes of the mitochondria hydrolyze one molecule of glycine into carbon dioxide, an amino group, and a methyl group that is transferred to a tetrahydrofolate cofactor. The serine hydroxymethyl transferase enzyme (SHMT, E.C. 2.1.2.1) catalyzes the condensation of that methyl group and a second molecule of glycine, to form serine. The SHMT enzyme also catalyzes the reverse reaction, a conversion of one molecule of serine to one molecule of glycine, with the formation of a one-carbon unit that is transferred to a tetrahydrofolate cofactor. The SHMT activity is due to the action of two spatially distinct isozymes. One resides in the cytosol and the other is within the mitochondria (30). It is believed that the cytosolic isozyme is incapable of glycine synthesis, whereas the mitochondrial isozyme can catalyze either the formation or degradation of serine, depending upon the available cofactors and cellular demands (31;32). Although the availability of glycine, the presence of the necessary enzymes in various tissues, and the capacity of this alternative pathway of serine biosynthesis is not known, it appears that this pathway is not capable of adequately supplementing a deficiency of PHGDH. If it was capable, the symptoms of a deficiency of PHGDH would be much milder or completely absent.

## 1.4 Metabolism of serine in mammalian cells

L-serine is a key intermediate in metabolism. As shown in Figure 2, serine and its derivatives cysteine and glycine, are used for the synthesis of proteins. Serine is a major source of one-carbon units in the cell, carried by folate cofactors and used for synthesis of molecules such as thymidine and the purine nucleotides. The byproduct of the decarboxylation of serine is glycine, which also forms part of the carbon backbone of the purine nucleotides. In addition, glycine is utilized for the biosynthesis of glutathione, creatine, porphyrins, and in the metabolism of bile acids. Serine may be converted to pyruvate and used in the production of energy. In the liver and kidney, the pyruvate molecule can also re-enter the gluconeogenic pathway and form carbohydrates or lipids. L-serine is used to form the head groups of phosphatidylserine molecules, which are a component of cellular membranes. Serine is also a component of sphingolipids, including, but not restricted to, sphingomyelin, glucocerebrosides, cerebrosides, and glycolipids.

The variety of metabolic fates of serine is even wider in the brain, where serine is also used in the production of neurotransmitters. Glycine acts as an inhibitory neurotransmitter in several regions of the brain (33). Serine racemase catalyzes the conversion of L-serine to its enantiomer, D-serine (33). D-serine acts as an excitatory neurotransmitter involved in co-activation of the *N*-methyl-D-aspartate receptor (glutamate also activates this receptor) (34).

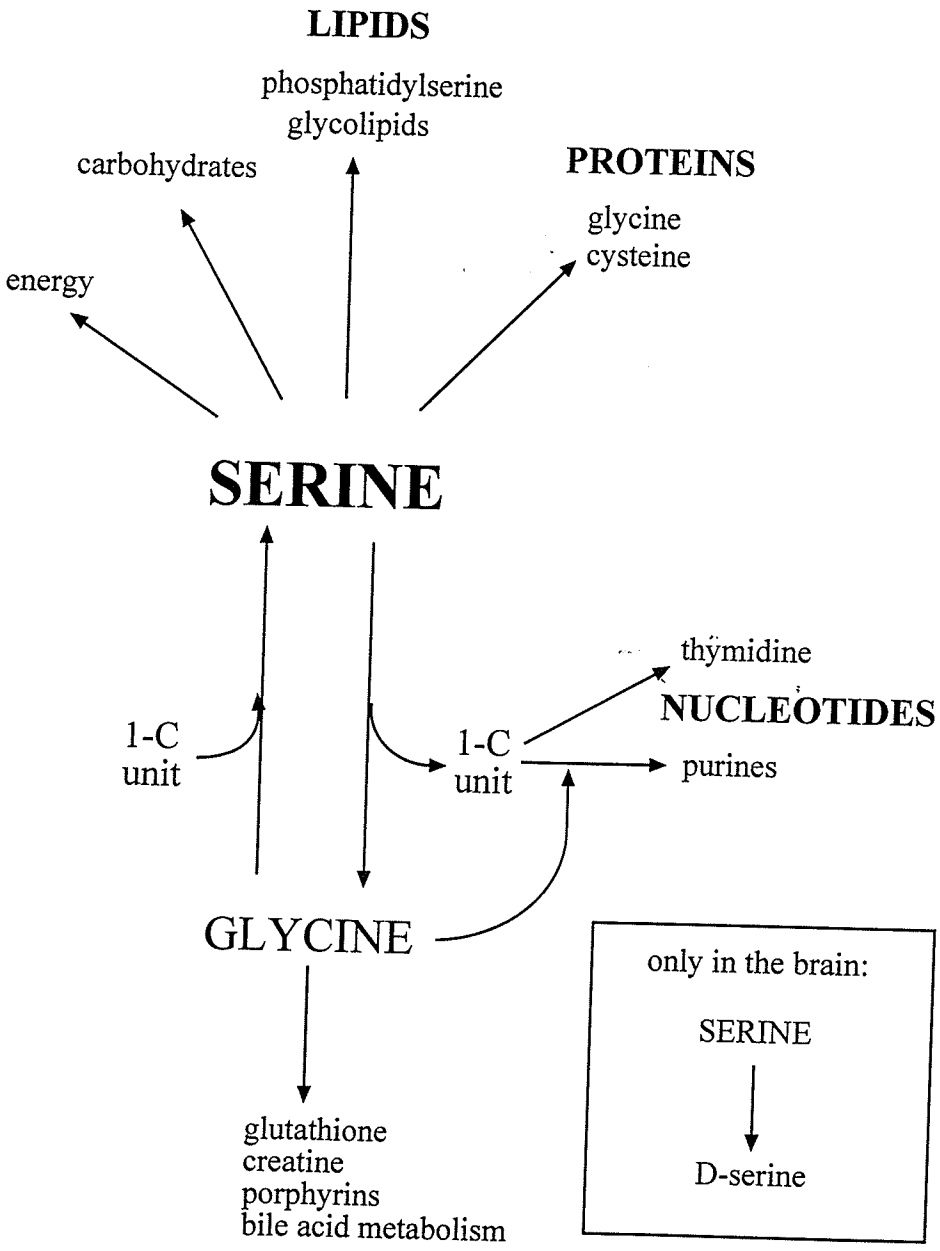



Figure 2. The metabolic fates of serine.

In summary, serine can be metabolized into proteins, nucleotides, energy, carbohydrates, lipids, and neurotransmitters. The complexity of serine metabolism limits our ability to pinpoint the exact biochemical disturbance responsible for each phenotypic change seen in patients with a deficiency of PHGDH. Attempts to solve each abnormality resulting from the serine insufficiency would be difficult, with so many reactions, endproducts, and metabolites to investigate, so the best method of treating this condition is increasing the amount of serine available to cells, especially those of the central nervous system.

The dietary intake of serine is likely insufficient to meet the body's daily requirements, otherwise a serine deficiency due to a deficiency of PHGDH would not have such serious consequences. Unlike metabolic disorders that usually affect enzymes in degradative pathways, resulting in a buildup of intermediates that lead to the disease phenotypes, a deficiency of PHGDH affects a biosynthetic pathway (35). As the first step in serine biosynthesis, a lower level of functional PHGDH would result in less flux through the pathway (36), but a buildup of toxic intermediates would not be expected to occur. Therefore, a decrease in the availability of serine leads to the observed phenotype.

The neurological profile of children with a deficiency of PHGDH suggests that serine production in the human central nervous system is heavily dependent upon synthesis through the PHGDH pathway. This dependency may be due, in part, to the blood brain barrier. This physical barrier is composed of tight junctions between the

endothelial cells of the blood vessels in the central nervous system (37). These junctions prevent compounds from entering the brain without passing through the endothelial cells and the astrocytic cells surrounding the blood vessels, in many cases requiring a carrier protein (37). Serine is transported across this barrier via a neutral amino acid carrier (38). This carrier also transports phenylalanine, leucine, tryptophan, methionine, isoleucine, tyrosine, histidine, valine, threonine, asparagine, and glutamine (38). Evidence suggests that this carrier has higher affinity for, and is probably saturated with, these larger, more hydrophobic amino acids (38;39). This saturation would result in lower levels of transport of the smaller neutral amino acids such as alanine, cysteine, and serine. The low levels of serine transport from the circulation to the brain, and a deficiency in PHGDH leading to a lowered production of endogenous serine, would be expected to result in very low levels of serine in the central nervous system. The dietary supplementation that benefits these children requires high levels of serine before the levels in the central nervous system begin to regain normal values (24), possibly due to the competition with the larger, more hydrophobic neutral amino acids for transport across the blood brain barrier.

## 1.5 Expression of PHGDH in mammalian cells

Given the many metabolic fates of serine, one might expect that all tissues and cells would possess this ability to synthesize serine from carbohydrates. This is not the case,

however, and the tissues that do produce serine have variable amounts of the biosynthetic enzymes. This was first illustrated in the 1960s, when the presence or absence of an enzyme in tissues was determined by enzyme assays. Early reports revealed that PHGDH is present in numerous animal tissues, including mouse brain (40), and the livers of chickens (41), rabbits (42), and those of pigeon, pig, dog, frog, and cow (43). In the rat, there is high activity of PHGDH in kidney, testis, epididymal fat pad, brain, spleen, and pancreas (44). Tissues with low activity include skeletal muscle, heart, and intestine (44). The expression of PHGDH in human tissues had not been investigated prior to the commencement of my thesis work.

During the course of this project, two articles were published with Northern blots showing the distribution of the mRNA encoding PHGDH in various tissues. Cho *et al.* (45) demonstrated a variable expression of PHGDH in 16 tissues, with high levels of expression in prostate, testis, pancreas, ovary, brain, liver, kidney, colon, and thymus, and low levels in spleen, small intestine, peripheral blood leukocytes, heart, placenta, lung, and skeletal muscle. Two transcripts were visible on the blot, a 2.1 kb band and a 0.7 kb band. The latter was visible in heart, skeletal muscle, pancreas, and less so in prostate, testis, and ovary. Cho *et al.* believed the former to be the full-length PHGDH transcript and that the latter was a transcript that had undergone alternative splicing (45). They made no attempt to investigate the origin of the smaller band, nor did they explain why the smaller band might result from alternative splicing.

The variable expression of the mRNA encoding PHGDH was confirmed by a 12-tissue Northern blot published by Klomp *et al.* (2), with a similar pattern of expression in most tissues studied. The one exception was thymus, as it displayed intermediate levels of expression in Cho *et al.*'s data and had minimal expression in Klomp *et al.*'s. In the central nervous system (CNS), all eight tissues studied express similar levels of mRNA encoding PHGDH (2). Fetal liver and kidney have high levels of expression whereas fetal brain and lung have medium to low expression. Klomp *et al.* also detected two transcripts, a 2.3 kb band considered to be the full-length transcript and a longer 4.5 kb transcript present in a subset of tissues (2). The heart and skeletal muscle had high levels of the 4.5 kb band, with low levels of the 2.3 kb band, whereas the other tissues had significant bands at both 2.3 kb and 4.5 kb. After longer exposure, Klomp *et al.* stated that both transcripts were visible in all tissues, but they did not speculate on the origin of the 4.5 kb band (2).

Despite the relatively high expression of PHGDH in the human brain (2;45), it appears that certain cells within the brain do not produce PHGDH and rely on other cells for their serine needs. Astrocytes have been shown to actively take up glycine from the extracellular space and metabolize the glycine into serine, creatine, glutathione (46), and lactate (47). These compounds are exported to the extracellular space (46-48), where they are believed to be taken up by the neighboring neurons (48). Hippocampal neurons do not appear to produce PHGDH and the serine biosynthetic enzymes (49). When they were

grown in serine-free conditions, the hippocampal neurons produced phosphatidylthreonine instead of the normal phosphatidylserine (50). The base exchange enzyme responsible for this alteration appears to substitute threonine for serine when serine concentrations are limited, despite a much lower affinity for threonine. Other neuronal cell types (chicken dorsal root ganglion nerve cells (51), and rat Purkinje neurons (52)) have similar requirements for exogenous serine when grown in culture. This implies that these cells do not possess the serine biosynthetic enzymes in sufficient quantities to supply their cellular demands for serine. When grown in medium supplemented with serine (49;51;52), or when co-cultured with glial cells (i.e. astrocytes), the neurons survive, grow, and differentiate at a much higher rate than growth without serine allows (49;51).

Recent studies using immunohistochemistry have extended these results to the cellular level. In the rat cerebellum, PHGDH was shown to be expressed in the Bergmann glia but was absent from the Purkinje neurons, suggesting that the Bergmann glia were responsible for serine synthesis in this region of the brain (52). In the mouse, a developmental study showed that PHGDH was expressed throughout the brain in early development, but became localized to radial glial cells and the glia that surround olfactory nerves at later stages (53). The authors stated that serine acted as a neurotrophic factor, since it is known that many neurons migrate alongside the radial glial processes during neuronal migration, and serine might be a chemoattractant secreted by the glia to

attract the neurons (53). If the neurons cannot produce serine, but require serine for protein and lipid synthesis, they will actively take up exogenous serine and may grow towards the source of serine. In this situation, a neuron expressing its own PHGDH and producing serine would be less attracted to the radial glia and might not migrate into proper position within the brain (54). In this manner, tightly regulated production of L-serine or its derivatives may aid in neuronal migration, differentiation, and survival.

The knowledge of the expression pattern of PHGDH in human central nervous system tissues was not known at the beginning of this project, and we believed it would be a first step towards understanding the role of this enzyme in the brain. One of my objectives was to characterize the expression profile of PHGDH in human tissues, with an emphasis on the central nervous system. This was accomplished primarily by Northern blotting, and these results were confirmed by a second method utilizing reverse-transcriptase-PCR amplification of mRNAs from various tissues. As monoclonal antibodies specific for human PHGDH became available during the course of my studies, I examined whether these could be used to establish the cellular expression profile of PHGDH using immunohistochemistry.

As reviewed above, serine can also be synthesized from glycine using the glycine cleavage system and SHMT. Although the mitochondrial form of SHMT was reported to be present at low levels in total brain RNA (55), its expression in these tissues was not examined in detail. A secondary goal of my expression studies was to determine the

expression profile of the SHMT isozymes in the central nervous system as a possible alternative source of serine.

## 1.6 Regulated expression of PHGDH

The neurological abnormalities evident in children with a deficiency of PHGDH indicate that there is a critical minimum level of PHGDH required for normal development and function of the central nervous system. It is not known what regulates PHGDH levels in human cells, although a number of different studies have shown that PHGDH from rat liver undergoes regulated expression (56-58). Study of the regulation of rat hepatic PHGDH began with controversy when several researchers had difficulty detecting any PHGDH enzyme activity in adult rat liver (57). This was resolved by studies in the developing rat that revealed very high PHGDH activity in fetal liver and decreasing enzyme activity throughout the neonatal period, culminating with very low levels in the adult (59). The reason for the low levels of PHGDH detected in rat liver, when this pathway is the major source of *de novo* serine in this tissue (29), was eventually resolved as due to inhibition by dietary protein. Regular rat chow represses the expression of PHGDH to very low levels (57), explaining the difficulty earlier researchers had in measuring PHGDH activity in rat liver preparations. It was unknown why only the hepatic PHGDH enzyme in rat would respond to regulation cues from the diet, nor was it discussed what was the benefit obtained from such a design. Hayashi

*et al.* (57) noted that the output of serine from the liver was roughly equal to the total output of serine from all other tissues, when the hepatic PHGDH enzyme was not being repressed by dietary protein. This suggests that the regulation of hepatic PHGDH plays a significant role in the amount of serine produced in an animal. This repression by dietary protein was also seen in rabbit liver (60) and rat hepatocytes in culture (56). A half-life for the rat liver PHGDH mRNA and protein of 8 hours and 2.5 days, respectively, was demonstrated by reintroducing protein to animals on low-protein diets and observing the declines in the mRNA and protein levels (56). The induction by low protein diet raised PHGDH mRNA levels rapidly within 12 to 72 hours (short-term effect) and continued to rise for up to five days (long-term effect) (56). The total increase in PHGDH mRNA varied between studies, up to 100-fold in experimental animals compared to controls (56). Mauron *et al.* (61) reported that only five amino acids were responsible for the decrease in PHGDH expression, and these were cysteine, methionine, tryptophan, threonine, and valine. It appeared that cysteine may be primarily responsible for regulating the expression of mRNA encoding PHGDH, since injection of cysteine into rats on a low protein diet had similar effects to refeeding with protein. Achouri *et al.* (56) speculated that cysteine levels drop during protein starvation, providing a good signal of proper nutrition to genes influenced by nutritional status. Though low-protein diets can induce PHGDH activity, complete starvation did not. It is believed that this lack of induction with starvation was due to lower levels of insulin (62). Support for this theory

was the knowledge that alloxan-induced diabetes in rats prevents the induction of PHGDH by a low-protein diet (62). This suggests that insulin, although it cannot overcome the repression by dietary protein, was required for the induction of PHGDH by a low-protein diet (56;57). Glucocorticoids repressed the induction of PHGDH by a low-protein diet (57). This was expected since glucocorticoids induce enzymes of serine metabolism towards gluconeogenesis (57). Glucagon repressed PHGDH expression in a method additive to the effect of cysteine. Achouri *et al.* believed that glucagon affects transcription of *PHGDH* and cysteine acts to destabilize the mRNA (56). All three of the serine synthetic enzymes responded similarly to dietary regulation in the rat liver (57).

In addition to the adaptive changes to diet by the rat liver PHGDH, it appears that certain forms of cancer reprogram tumour cells to up-regulate PHGDH expression, as well as other enzymes involved in serine metabolism. Comparisons of enzyme activity in transplanted cancerous tumours and host organs show that the serine biosynthetic enzymes are greatly increased (63;64), but that serine dehydratase and serine aminotransferase, two enzymes that are responsible for the metabolism of serine, are completely absent in the tumours (63;65). This reprogramming was shown for rat hepatomas (44;63;64), rat sarcomas (66), and human colon carcinoma (66). Unlike other enzymes of serine metabolism, SHMT activity was maintained in the tumours, likely because it provides the one-carbon units and glycine required for the synthesis of nucleotide precursors used in the biosynthesis of DNA. This reprogramming may result

in more serine being fed toward nucleotide synthesis, as was demonstrated by the incorporation of radiolabelled serine into the DNA of liver carcinoma 3924A cultured cells (44;65). Snell *et al.* found that after six hours of culture with radiolabelled serine, the level of labelled nucleotides is 5-6 fold higher in actively dividing cells compared to resting cells (64). These results confirmed the label incorporation rates seen with non-malignant, mitogenically transformed proliferating lymphocytes (65). Similar to neoplastic cells, there was also a correlation between high PHGDH enzyme activity and tissues with a high proliferative ability (29), since these tissues would also require large amounts of serine for nucleotide, protein, and lipid synthesis.

## 1.7 The crystal structure of the *E. coli* PHGDH enzyme has allowed structure-function analysis

The residues important to structure and function of PHGDH have been investigated using the *E. coli* protein. The nucleotide sequence encoding this enzyme was determined in 1989 (67). The gene is 1233 nucleotides in length and encodes a 410-amino acid protein. The enzyme shows approximately 50% sequence similarity to other D-2-hydroxyacid dehydrogenases (67), mainly due to conservation of the NAD-binding domain (68). The X-ray crystal structure of the serine-bound (inactive) form of the enzyme was reported in 1996 (69). The *E. coli* PHGDH has three domains: the nucleotide binding domain (residues 108-294), the substrate binding domain (residues 7-107 and

295-336), and the regulatory domain (residues 337-410) (70). Whereas the other members of this family are homodimers, PHGDH is a homotetramer (67). In addition to the association of monomers, the PHGDH enzyme is unique in the existence of its carboxyl-terminal regulatory domain (70), defined as an ACT motif (71), which binds serine to inhibit the catalytic activity. The monomers associate at two interfaces; the nucleotide binding domains associate to form dimers, and the regulatory domains of two dimers associate to form a donut-shaped tetramer (see Figure 3). The crystal structure led to a model of catalysis known as the "rigid domains about flexible hinges" hypothesis (69). The $NAD^+$ cofactor (nucleotide) and the 3-phosphoglycerate (substrate) bind to their respective domains, near the junction of the domains where an active site cleft is formed (shown by arrows in Figure 3) (70). A conformational shift allows closure of the cleft that brings the substrates close enough to facilitate the hydride transfer. The protein returns to its original conformation and the products are released. This conformational shift is regulated by serine binding in the regulatory domain (70), so a serine-bound protein is unable to undergo the conformational changes required for completion of catalysis (69).

The *E. coli* protein has been manipulated with site-directed mutagenesis, revealing several amino acid residues critical to various functions. In the last decade, Grant *et al.* have elucidated the critical residues for domain interaction (72;73), substrate

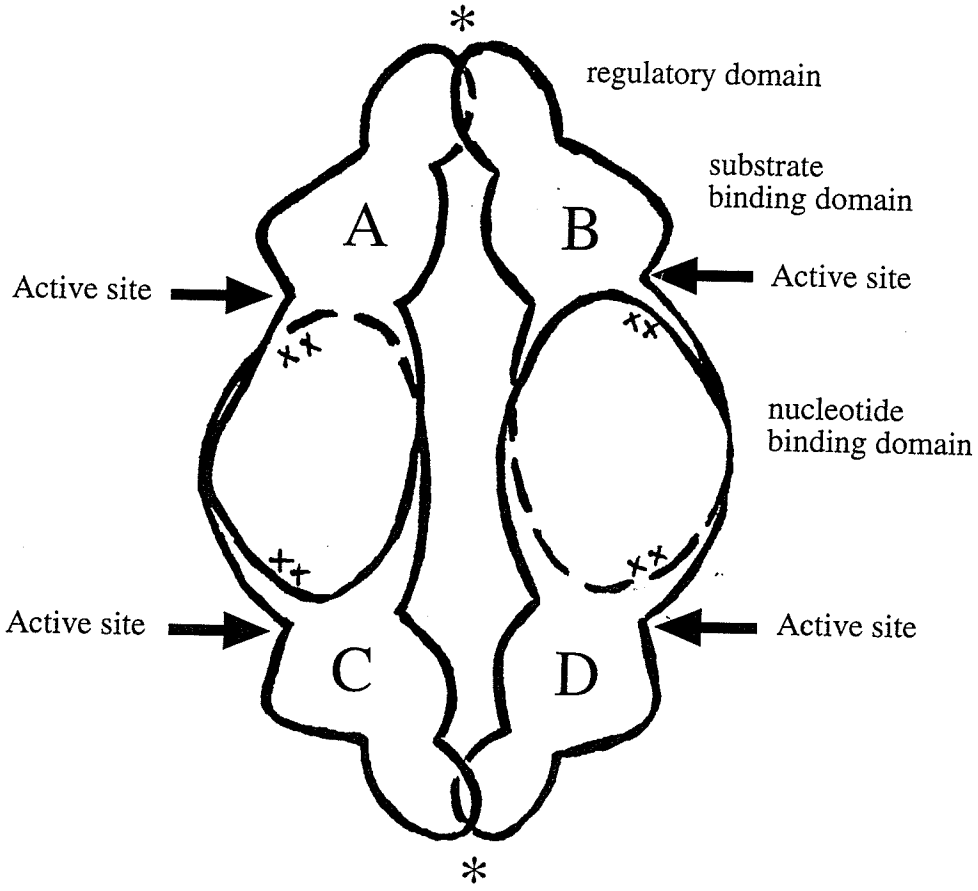



Figure 3. The orientation of the *E. coli* PHGDH subunits (A-D) in a functional homotetramer. The arrows indicate the location of the active site clefts and the asterisks (*) indicate the interface at which serine binds (two binding sites per interface). The crosses (X) near the active sites represent Trp-139 and Lys-141 on each monomer. These residues interact with active site residues of the adjacent monomer. The domains of monomer B are noted.
[adapted from Grant *et al.* JBC **274**:5357-5361 (1999)]

binding (74), serine binding (75), cooperation between subunits (72;73), and flexibility required for catalytic activity (72).

The association of monomers to form dimers occurs at the nucleotide domains of two monomers (70). The interface between adjacent nucleotide domains was believed to be extensive, with over 1600 $Å^2$ of contact (70), but this interaction can be abolished by one mutation (73). When the side chain of tryptophan-139 was mutated, the enzyme did not associate at the nucleotide domains and instead formed dimers about the regulatory domain interface. The side chain of the tryptophan-139 from one monomer fits into a slot formed by the active site of the adjacent monomer, and this interaction is critical to the association at the nucleotide-binding domain interface (73;74).

Substrate binding occurs in the substrate binding domain and requires several positively charged residues that form a pocket which attracts and holds the negatively charged phosphoglycerate (70). Residues important to substrate binding include lysine-39, arginines at positions 60, 62, and 240 and lysine-141 from the adjacent monomer (74). Specifically, arginine-240 binds the C1 carbon of phosphoglycerate while the other two arginines and two lysines act to bind the C3 carbon. The PHGDH of *E. coli* is also capable of acting on α-ketoglutarate (76), and in this instance, the positively charged pocket would attract the carboxylate group of this substrate. The catalytic activity involves a hydride transfer from the substrate to cofactor $NAD^+$, and this is

believed to occur by the histidine at position 292 and the glutamate at position 269 (74). Similar histidine-aspartate pairs function in this capacity in other oxidoreductases (68).

It was also shown that serine interacts with the regulatory domain by creating a network of hydrogen bonds that anchor the domain interface (77), creating a conformational change that is transmitted to the other domains by flexible glycine pairs and thus inhibits catalytic activity (72). This was illustrated by mutating the residues involved in serine binding and in hydrogen bond formation along the regulatory domain interface, resulting in a loss of the inhibitory effect of serine (78). A disulfide bridging experiment was a second method used to study the flexibility and conformational changes required for catalysis (79). By engineering cysteine residues along the regulatory domain interfaces and exposing the altered enzyme to reducing or oxidizing agents and measuring catalytic activity, the act of serine binding and release could be mimicked (79). When the cysteines were oxidized, disulfide bonds formed between the regulatory domains were similar to the hydrogen bonding that occurs when serine binds, and the catalytic activity was completely inhibited without serine. This inhibition was reversible by the addition of a reduction agent (79).

Glycine-glycine pairs provide the flexibility required at the hinge regions between the rigid domains, as well as allowing cooperation to occur when serine is binding and for the subsequent inhibition of catalytic activity (72). The first glycine pair to be studied resides between the regulatory domain and the substrate binding domain at positions 336

and 337. When steric hindrance was introduced by mutating the glycines to residues with larger side chains, serine binding still occurred but had less of an inhibitory effect on the catalytic activity of the enzyme (72). The second glycine-glycine pair at positions 294 and 295 allows the movement required for catalysis, as mutants showed little change to serine binding but a significantly decreased catalytic activity (80).

These mutagenesis studies have provided a good starting point for similar experiments with the human PHGDH protein, once its structure has been characterized. However, the *E. coli* protein is not an ideal model for extrapolating protein structure/function relationships to the human PHGDH protein since there are several major differences between these two proteins. The *E. coli* PHGDH protein is regulated by serine feedback inhibition, but in mammals serine levels have no effect on PHGDH. Instead, serine inhibits the third enzyme of the pathway (3-phosphoserine phosphatase) (29). The length of the *E. coli* protein is considerably shorter than the PHGDH proteins from other organisms, with the difference primarily due to the size of the carboxyl-terminal regulatory domain. In addition, the *E. coli* protein acts on an alternative substrate (α-ketoglutarate) not used by the mammalian proteins (81) (E. Slominski and S. Pind, unpublished observations).

## 1.8 Characterization of the mammalian PHGDH genes and proteins

The genomic organization of the human *PHGDH* had not been reported, although the sequence of the cDNA had been known for two years at the start of this thesis. The cDNA encoding the human PHGDH protein was first deposited in Genbank (http://www.ncbi.nlm.nih.gov) in 1997. Information on the gene and protein has been slower to appear. The human cDNA that first appeared (Genbank accession number AF006043) was identified from a scan of transcripts in a T-cell cDNA library (45). The cDNA was 2478 nucleotides, with 692 bp of 5' untranslated region, 1602 bp open reading frame encoding a predicted 533 amino acid protein, and 184 bp of 3' untranslated region including the polyadenosine tail. The coding sequence was supported by the cDNA sequence obtained in the Pind laboratory and others, with the exception of a base pair change (2;3;3). Alternative cDNA sequences have since appeared (Genbank accession numbers AF171236, AF171237), which do not include the 590 bp of unrelated sequence in the 5' untranslated region (82). Using the cDNA as a probe, the *PHGDH* gene has been localized to the pericentromeric region of chromosome 1. It is found between 1p12 and 1q12, with Baek *et al.* (83) and Genbank supporting 1p12 while Klomp *et al.* (2) and Celera support a 1q12 localization. The repetitive nature of pericentromeric DNA is probably the reason for the confusion, as clones from the region may have significant rearrangements and repetitive elements and may be misassigned (84;85). The identification of the splice sites and sequence of the introns near these sites

would be beneficial to the study of PHGDH deficiency, by allowing researchers to amplify individual exons for mutation analysis in patients that have neither the V490M nor the V425M mutation. We determined the genomic organization of the human *PHGDH* gene using a PCR-based sequencing strategy. We also investigated the transcriptional start site of *PHGDH* transcripts in human liver using a 5' RACE kit and sequencing.

The sequence of the *PHGDH* gene from rat (*R. norvegicus*) may provide useful information towards the determination of the human *PHGDH* gene organization. The rat cDNA sequence was reported in 1997 (81) (Genbank accession number X97772), with the rat gene appearing in 2000 (82) (Genbank accession number AJ271975). The rat gene consists of twelve exons and eleven introns. The rat gene was also reported to have an alternative exon 1, 275 base pairs upstream from the translational start site (ATG). A study of eight tissues revealed that this alternative exon 1 (termed exon 1') is expressed only in rat testis, and only one in four transcripts from testis had the alternative exon present. The alternative exon 1' is 40 bp long and is spliced onto exon 1 at position -30, resulting in the same translational start site as the non-testis transcripts (see Figure 4). Hence the only change is the substitution of 40 bp in the 5' untranslated region. This raised the possibility that the human PHGDH also undergoes alternative splicing upstream of exon 1. Pools of cDNA from various human tissues were amplified with two

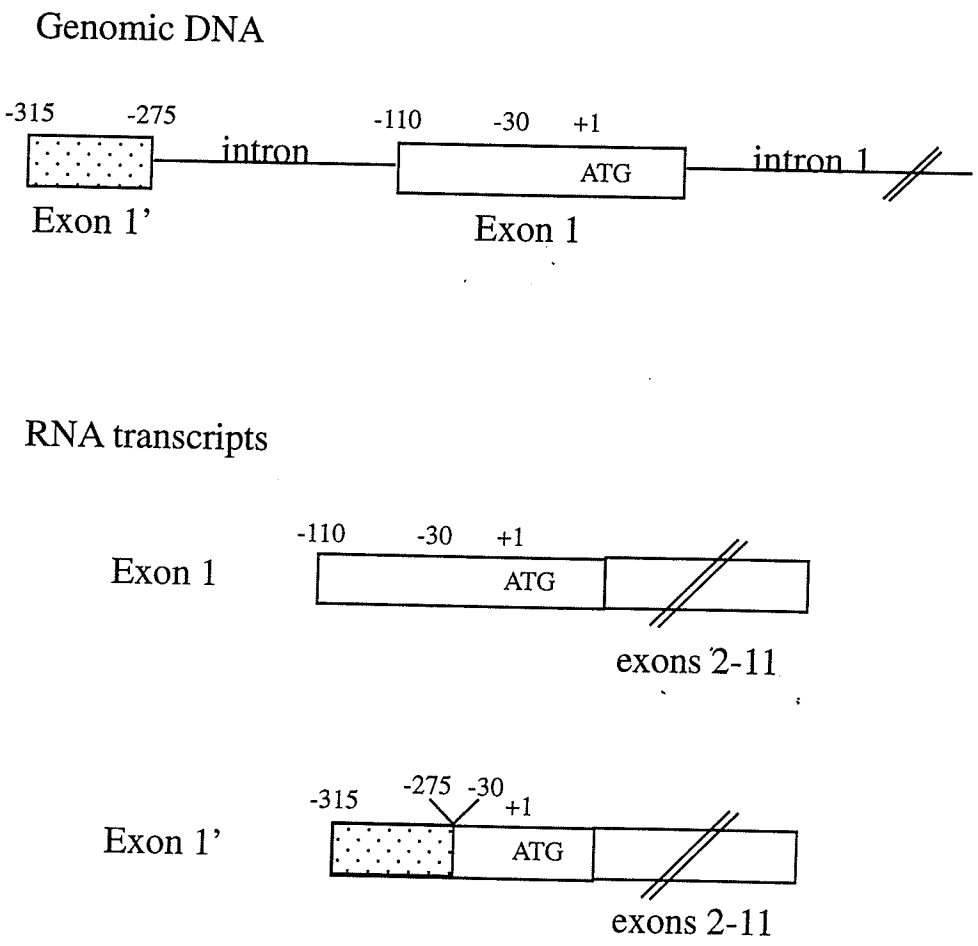



Figure 4. Alternative splicing of the rat PHGDH gene in the 5' untranslated region. The exons are represented by boxed regions, intronic or non-transcribed DNA is represented by a line. Exon 1' splices on to exon 1 at position -30, as described by Robbi *et al.* Mamm. Genome **11**:1034-1036 (2000). This alternative splicing was seen for one transcript from rat testis.

sets of primers to determine whether the transcripts had exon 1 or exon 1' sequences at their 5' ends.

## 1.9 Alignment of PHGDH amino acid sequences suggests conserved residues

Comparisons between the amino acid sequences of PHGDH homologues will yield information regarding the importance of critical residues in the PHGDH enzyme. Essential residues will be maintained throughout evolution, whereas residues with less of an impact on critical function or structure may undergo mutation without selective pressure. With structure/function relationships known only for the *E. coli* PHGDH protein, comparisons would be with the residues already mentioned above.

Alignment of the human and *E.coli* amino acid sequences reveal that the human protein is 123 amino acids longer, and these residues reside in the C-terminal regulatory domain. There appear to be two forms of PHGDH. A shorter protein (~410 amino acids) is common to a subset of bacteria (*E. coli, H. influenzae*, etc.) and yeast (*S. pombe, S. cerevisiae*). A longer protein (>525 amino acids) is common to some bacteria (*B. subtilis*, *H. pylori*, etc.), mammals, and the plant *A. thaliana*. The longest PHGDH homologue is that of *A. thaliana* at 622 amino acids, due to a 60 amino acid localization signal at the amino-terminal region that targets the protein to the plastids within plant cells and additional length at the carboxyl-terminal regulatory domain (86).

With the explosion of submissions in the Genbank repository, alignment of several PHGDH sequences with bioinformatic software such as ClustalW (87) can be performed. A ClustalW alignment of the *E. coli* amino acid sequence with sequences from bacteria (*H. influenzae, B. subtilis*), Archaebacteria (*A. fulgidus*), yeast (*S. pombe, S. cerevisiae*), mammals (*H. sapiens, R. norvegicus, M. musculus*), and plant (*A. thaliana*) is provided in Figure 5, with the Genbank accession numbers for these sequences reported in the figure legend. To provide a clear comparison of the residues that will be discussed, (shown in bold), the regions of homology detected in the ClustalW alignment are not indicated.

The alignment of PHGDH homologues is not consistent or reproducible in their entirety. The nucleotide and substrate binding domains are highly conserved and alignment of these regions is reproducible. The differences in the size and composition of the carboxyl-terminal regulatory domain cause the alignment programs to align the carboxyl-terminal region in two main ways. The first alignment is one in which the shorter sequences align and the carboxyl-terminal region of the longer proteins extends past this point. The second alignment occurs with a gap introduced between *E. coli* residues 317-318, with the carboxyl-termini of both the shorter and longer proteins aligning at the distal end of the sequence. Figure 5 shows this latter alignment.


Figure 5. ClustalW alignment of ten PHGDH amino acid sequences. Abbreviations are: Ec=*E. coli*, Hi=*H. influenzae*, Sp=*S. pombe*, Sc=*S. cerevisiae*, Rn=*R. norvegicus*, Mm=*M. musculus*, Hs=*H. sapiens*, At=*A. thaliana*, Af=*A. fulgidus*, Bs=*B. subtilis*. The alignment continues on the next page. Genbank accession numbers are: Ec=P08328, Hi=C64070, Sp=T41375, Sc=S50584, Rn=CAA66374, Mm=(translated cDNA from overlapping BG146773 and cDNA L21027), Hs=AAD51415, At=BAA20405, Af=NP_069647, Bs=NP_390188.

```
                    H292
Ec   278   TSPLCEFDNVLLTPHIGGSTQEAQENIGLEVAGKLIKYSD----------
Hi   279   ISPLREFDNVILTPHIGGSTAEAQENIGFEVAGKFVKYSD----------
Sp   330   TSELTHCKNIILTPHIGGSTEEAQYNIGIEVSEALTRYIN----------
Sc   333   TSELVSLPNIILTPHIGGSTEEAQSSIGIEVATALSKYIN----------
Rn   272   ---LVDHENVISCPHLGASTKEAQSRCGEEIAVQFVDMVG-KSLTGVVNAQALTSAFSPHTKPWIGLAEALGTLMHAWAGSPKGTIQVVTQGTSLKNAGTCLSPAVIVGLLR---EASKQ
Mm   272   ---LVDHENVISCPHLGASTKEAQDRCGEEIAVQFVDMVKGKSLTGVVNAQALTSAFSPHTKPWIGLAEAMGTLMHAWAGSPKGTIQVVTQGTSLKNAGTWLSPAVIVALLR---EASKQ
Hs   272   ---LVDHENVISCPHLGASTKEAQSRCGEEIAVQFVDMVKGKSLTGVVNAQALTSAFSPHTKPWIGLAEAMGTLMHAWAGSPKGTIQVVTQGTSLKNAGTWLSPAVIVALLR---EASKQ
At   348   --RLIQHENVTVTPHLGASTKEAQEGVAIEIAEAVAGALKGELSATAVNAPMVAPEVLSELTPYIVLAEKLGRLAVQLASGGKGVQSIRVVYRSARDRDDLDTRLLRAMITKGIIEPISD
Af   266   --PLLKLDNVVTTPHIAASTREAQLNVGMIIAEDIVNMAKGLPVRNAVNLPSIEPSDFEFMMPFLTLAEKMGKIASVRLGGAIRKVKVTCSGKLATKNTEFVTRALLKG----------
Bs   265   ---LVDHPLVIATPHLGASTKEAQLNVAAQVSEEVLQFAKGLPVMSAINLPAMTKDEFAKIKPYHQIAGKIGSLVSQCMKEPVQDVAIQYEGTIAKLETSFITKALLSG----------

Ec   318   --------------NGSTLSAVNFPEVSLPLHGG-----RRLMHIHENRPGVLTALNKIFAEQGVNIAAQYLQTSAQM----------GYVVIDIEADEDV-AEKALQAMKAIPGTIRAR
Hi   319   --------------NGSTLSSVNFPEVSLPEHEGT----KRLLHIHENRPGILNKLNQIFVEANLNIAAQYLQTDPKI----------GYVVVDVETND---ASPLLTKLKEIDGTIRAR
Sp   370   --------------EGNSIGAVNFPEVSLRSLTEADRNAARVLFVHRNVPGVLRQVNELFIDH--NIKSQFSDSRGDI----------AYLVADISDCTPGSLEALHQKLESLPCKINTR
Sc   373   --------------EGNSVGSVNFPEVALKSLSYDQENTVRVLYIHQNVPGVLKTVNDILSNH--NIEKQFSDSNGEI----------AYLMADISSVDQSDIKDIYEQLNQTSAKISIR
Rn   385   ADVNLVNAKLLVKEAGLNVTTSHSPGVPGEQGIGECLLTVALAGAPYQAVGLVQGTTPMLQMLNGAVFRPEVPLRRGQ----------PLLLFRAQPSDPVMLPTMIGLLAEAGVQLLSY
Mm   386   ADVNLVNAKLLVKEAGLNVTTSHNPGVPGEQGSGECLLTVALAGAPYQAVGLVQGTTPMLQMLNGAVFRPEVPLRRGQ----------PLLLFRAQPSDPVMLPTMIGLLAEAGVQLLSY
Hs   386   ADVNLVNAKLLVKEAGLNVTTSHSPAAPGEQGFGECLLAVALAGAPYQAVGLVQGTTPVLQGLNGAVFRPEVPLRRDL----------PLLVFRAQPSDPGMLPTMIGLLAEAGVQLLSY
At   466   SYVNLVNADFIAKQKGLRISEERMVVDSSPEYPVDS-IQVQILNVESNFAGAVSDAGDISIEGKVKYGVPHLTCVGSFG---VDVSLEGNLILCRQVDQPGMIGQVGNILGEQNVNVNFM
Af   373   --------LFEPILSNEINLVSAKPVAVERGITIEESKVESVEHYESLLEVWVESNGKEMYLAGTCFGNEYRILKIDVY--NVNFVPKGHYIISLHEDKPGVIGRVGTLFGRNNINIAGM
Bs   371   ---------FLKPRVDSTVNEVNAGGVAKERGISFSEKISSSESGYDNCISVKVTGDRSTFTVTATYIPHFGERIVEINGFNIDFYPTGHLVYIQHQDTTGVIGRVGRILGDNDINIATM

Ec   408   LLY-------------------------------------------
Hi   408   VLY-------------------------------------------
Sp   464   LLY-------------------------------------------
Sc   467   LLY-------------------------------------------
Rn   495   QTSKVSD--GDTWHVMGLSSLLPSLDAWKQHVSEAFQFCF------
Mm   496   QTSMVSD--GEPWHVMGLSSLLPSLETWKQHVLEAFQFCF------
Hs   496   QTSLVSD--GETWHVMGISSLLPSLEAWKQHVTEAFQFHF------
At   582   SVGRTVL--RKQAIMAGVDEEPDNKTLERIGGVSAIEEFVFLKL--
Af   483   IVGRSGDKPGGIQLMLLLVDDPPTPEVLEEMTKLDGIIDATYVEL-
Bs   482   QVGRKEK--GGEAIMMLSFDRHLEDKIVKELTNVPDIVSVKLIDLP
```

Figure 5 (continued). ClustalW alignment of ten PHGDH amino acid sequences. Abbreviations are: Ec=*E. coli*, Hi=*H. influenzae*, Sp=*S.pombe*, Sc=*S. cerevisiae*, Rn=*R. norvegicus*, Mm=*M. musculus*, Hs=*H. sapiens*, At=*A. thaliana*, Af=*A. fulgidus*, Bs=*B. subtilis*. The alignment begins on the previous page. Genbank accession numbers are: Ec=P08328, Hi=C64070, Sp=T41375, Sc=S50584, Rn=CAA66374, Mm=(translated cDNA from overlapping BG146773 and cDNA L21027), Hs=AAD51415, At=BAA20405, Af=NP_069647, Bs=NP_390188.

The residue tryptophan-139 in the *E. coli* protein is conserved in all ten sequences of the ClustalW alignment. It would be expected that this residue, critical to the interactions at the nucleotide binding domain interfaces, would be highly conserved. Arginine-240 of the *E. coli* sequence, responsible for binding the C1 carbon of phosphoglycerate, is conserved in all ten sequences of the ClustalW alignment. The other basic residues that attract and hold the rest of the substrate, retain the positively charged nature of the *E. coli* sequence, especially the arginine-60 and lysine-141 residues. In the ClustalW alignment, arginine-60 is conserved in all ten sequences. Lysine-141 is present in the *E. coli*, *H. influenzae*, and the two yeast species (the shorter homologues), with arginine-141 in the remaining sequences. The arginine-141 retains the positive side chain necessary to attract the substrate, but may influence the success of using α-ketoglutarate as an alternative substrate (mammalian enzymes cannot use this substrate, as previously mentioned).

The histidine-glutamate pair responsible for the hydride transfer in the *E. coli* protein is highly conserved, as expected. Glutamate at position 269 is conserved in all ten sequences as is histidine 292. Since this hydride transfer is essential to the catalytic function of PHGDH (or related dehydrogenase enzymes), it is expected that these two residues would be retained throughout evolution.

The variability seen in alignments of the carboxyl-terminal regions of the PHGDH homologues creates difficulty in the accurate assessment of the possible

conservative nature of not only the glycine pair at *E. coli* positions 336, 337, but also the mutations seen in human PHGDH (V425 and V490).

The structure of the human PHGDH protein remains unsolved, though the wild-type protein has been purified and both the wild-type and the V490M PHGDH proteins have been studied enzymatically. During the course of my studies, the human PHGDH protein was overexpressed in *E. coli* cells and purified by Pind *et al.* (3). Purified human PHGDH protein was sent for X-ray crystallization studies and was used to induce monoclonal antibodies in mice. The monoclonal antibodies were used to recognize the PHGDH protein on Western blots and for immunoprecipitation. These studies show that the V490M mutation reduces the total amount of correctly folded PHGDH protein in cells, which in turn decreases the enzyme activity of cells expressing the mutant PHGDH protein (3). The mutation has no apparent effect on the activity or stability of the portion of the mutant protein that achieves a correctly folded active form, as shown with pulse-chase and thermal inactivation experiments. It was concluded that the V490M mutation likely effects protein folding (3). Experiments on the V425M mutation are required to confirm whether this mutation has a similar effect, or if the V425M mutation acts in some other method to disrupt enzyme activity or regulation.

## 1.10 The research aims of this thesis

A deficiency in PHGDH is known to result from two point mutations: a rare 1326G→A (V425M substitution) and a more common 1468G→A (V490M substitution) (2;3). The latter was characterized in four families of different ethnic groups (two Turkish, one Ashkenazi Jewish, one unknown). No evidence of an increased frequency of the mutant allele was detected in sample populations of Turkish and Western European individuals (2). The Ashkenazi Jewish family from North America was not consanguineous and therefore raises a question concerning the frequency of heterozygotes for the 1468A allele in this ethnic population. The introduction of a novel *Hsp*92 II recognition site by the 1468G→A substitution is the basis of a simple PCR-based assay to identify individuals carrying this allele (3). In this manner, DNA samples from individuals with four Jewish grandparents will be tested for the allele encoding the V490M mutation, and will be compared to a similar group of samples from non-Jewish individuals.

The predominant neurological symptoms of PHGDH deficiency are unusual considering the published expression pattern in rat, animal, and human tissues. These studies show that despite expression in most tissues, there are tissue-specific variations in the level of expression. All tissues within the human brain appear to express the mRNA encoding PHGDH, but the number of tissues studied remains small. The possible role of SHMT acting to supplement the production of serine in the central nervous system via

glycine to serine conversion has not been considered in previous studies. The expression of PHGDH and the two isozymes of SHMT will be characterized by hybridizing radiolabelled RNA probes to commercially obtained Northern blots. A secondary method of studying PHGDH expression will be employed, by PCR amplification of a section of *PHGDH* from pooled cDNAs isolated from an array of human tissues. As mouse monoclonal PHGDH antibodies became available during the course of this thesis, these antibodies were used to investigate the localization of expression of the PHGDH protein in paraffin-embedded sections of human brain tissue. These experiments should localize the expression of PHGDH to individual cell types within regions of the human brain.

The genomic organization of PHGDH from *R. norvegicus* has been characterized (82) but the human gene structure remained unsolved as late as November 2000. The localization of introns in the human gene was determined using a PCR-based cloning strategy, and intronic regions near the intron-exon boundaries were sequenced. The transcriptional start site of human liver *PHGDH* was also determined using a 5' RACE kit and sequencing of cloned RACE-PCR products. The possibility of alternative splicing upstream of the translational start site of human PHGDH was also investigated by a PCR-based strategy using cDNA from multiple tissues.

# 2. METHODS

A variety of experiments were performed during the course of these studies. However, many of these experiments had similar components (PCR amplification, electrophoresis, or cloning, for example). In this chapter I will first describe some of the general methods that were used repeatedly, and then present a more detailed description of the methods used for specific objectives. Specifics related to basic molecular biology procedures can be found in a general molecular biology text, such as *Short Protocols in Molecular Biology* (88).

## 2.1 Isolation of DNA from various sources

### 2.1.1 Isolation of human genomic DNA

In order to screen large numbers of individuals for the 1468G→A mutation, it was necessary to isolate genomic DNA from frozen white cell pellets. The pellets were prepared from whole blood samples and frozen by Dr. M. Natowicz and colleagues (E.K. Shriver Center, Division of Medical Genetics, Waltham, MA). All identifying markers were removed prior to shipment; samples were simply grouped as 'Jewish' or 'non-Jewish'. The samples marked Jewish were known to have four Ashkenazi Jewish grandparents, as documented by the records of Dr. M. Natowicz's laboratory. The frozen cell pellets were kept at –80°C until used for genomic DNA preparation.

The tubes containing frozen white cell pellets were warmed to room temperature and labelled for future reference. The pellets were lysed and the cellular protein was digested in a 0.5 ml aliquot of Proteinase K lysis buffer (1% SDS, 50 mM Tris-HCl, 10 mM EDTA, pH 8.0, 200 µg/ml Proteinase K) by incubation overnight at 50-55°C. 100 µl aliquots were removed for isolation of genomic DNA, with the remainder of the digested samples stored at –80°C. The aliquots were well mixed with 200 µl of Low TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0), then with 300 µl of buffer-saturated phenol, and the phases were separated by centrifugation for 5 minutes at 7000 rpm. The aqueous phase underwent a second phase separation with 300 µl of 25:24:1 buffer-saturated phenol:chloroform:isoamyl alcohol. The aqueous phase was subjected to a third separation with 250 µl of 24:1 chloroform:isoamyl alcohol, before precipitation of the DNA with 10 µl of 5 M NaCl and 500 µl of ice-cold ethanol. The DNA was pelleted by centrifugation at 13,200 rpm for 10 minutes at 4°C. The supernatant was aspirated and the DNA pellets were allowed to air dry, before being resuspended in 200 µl of Low TE buffer and stored at 4°C. For long-term storage, the DNA was kept at –20°C or –80°C.

### 2.1.2 Isolation of plasmid DNA from *E. coli* by a small scale method

This method, also known as a rapid alkaline lysis method or miniprep, rapidly isolates a small amount of plasmid DNA from bacterial cultures (89).

Overnight cultures were made by selecting isolated bacterial colonies and incubation in 3 ml aliquots of Luria-Bertani (LB) media (1% tryptone, 0.5% yeast extract,

1% NaCl, pH 7.2) and appropriate antibiotic at 37°C overnight with shaking (250-300 rpm). There were three types of plasmids used during the course of this thesis and each required a different antibiotic. The pcDNA3.1(-) plasmid (Invitrogen) was grown in 20 μg/ml carbenicillin and was used in the cloning of the DNA templates for RNA transcription. The Topo®XL plasmids (Invitrogen) were grown in 50 μg/ml kanamycin and were used for the cloning of the PCR products used to identify the introns within the *PHGDH* gene, as well as the cloning of the 5' rapid amplification of cDNA ends (5' RACE) products. The bacterial artificial chromosomes (BACs) were grown in 10 μg/ml chloramphenicol and were used solely as a source of human genomic DNA for the PCR amplification of *PHGDH* introns.

About 1.5 ml of overnight culture was pelleted by centrifugation at 13,000 rpm for 1 minute. The pellet was resuspended in 100 μl of a resuspension buffer (50 mM glucose, 10 mM EDTA, 25 mM Tris-HCl, pH 8.0) and incubated for 5 minutes at room temperature. A 200 μl aliquot of the alkaline lysis buffer (1% SDS, 0.2 M NaOH) was added, and the tube was then inverted 3-4 times to mix and placed on ice. This lysed the cells and denatured the bacterial proteins. 150 μl of 3 M potassium acetate was added and mixed by inverting the tube for 10-15 seconds. This was incubated on ice for 5 minutes before the insoluble material was pelleted by centrifugation at 13,000 rpm for 5 min at 4°C. The supernatant (400 μl) was transferred to a fresh tube and 400 μl of 25:24:1 buffer-saturated phenol:chloroform:isoamyl alcohol was added, mixed briefly and

partitioned by centrifugation at 13,000 rpm for 5 minutes. The aqueous phase (300 µl) was removed to a fresh tube, mixed well with 600 µl of ice-cold ethanol and incubated at room temperature for 2 minutes, prior to centrifugation at 13,000 rpm for 5 minutes. The supernatant was aspirated and the DNA pellet was washed with 70% ethanol. The pellet was air-dried and resuspended in 30 µl of Low TE buffer with 50 µg/ml RNase A (Fisher Scientific), to degrade any RNA present in the sample.

### 2.1.3 Isolation of plasmid DNA from *E. coli* by a large scale method

Large-scale isolations of plasmid DNA were performed with the QIAfilter Plasmid Maxiprep purification kit (Qiagen). Starter cultures from isolated colonies were incubated in 3 ml LB media and antibiotic in the 37°C shaker-incubator at 250 rpm for 8 hours. 500 µl of this culture was diluted into 250 ml of LB media with antibiotic in a 1 L flask and incubated overnight with shaking. The bacterial cells were harvested by centrifugation at 8,000 rpm for 15 minutes at 4°C. The pellets were resuspended in 10 ml of cold resuspension buffer (containing RNase A) and 10 ml of alkaline lysis buffer was added. This was inverted 4-6 times to mix and incubated at room temperature for 5 minutes. Chilled neutralization buffer was added and inverted 4-6 times to mix. This mixture was poured into the barrel of a Qiagen filter and incubated at room temperature for 10 minutes. The lysate was forced through the filter, which removed cellular debris, and allowed to flow through the pre-equilibrated Qiagen column where the plasmid DNA bound to the resin. The column was washed with a total of 60 ml of wash buffer, before

the DNA was eluted with 15 ml of elution buffer into a 30-ml glass tube. Room-temperature isopropanol (10.5 ml) was added to precipitate the DNA and was mixed before centrifugation for 30 minutes at 11,000 rpm at 4°C. After the supernatant was removed by aspiration, the DNA pellet was washed with 5 ml of 70% ethanol and centrifuged at 11,000 rpm for 10 minutes at room temperature. The pellet was air-dried and resuspended in 500 µl of 10 mM Tris-HCl, pH 8.0. This was allowed to sit at room temperature several hours, or at 4°C overnight, to fully resuspend the DNA. Quantification of the DNA was done on an agarose gel or by spectrophotometry (see below).

### 2.1.4 Isolation of plasmid DNA from TOP10 *E. coli* cells by a modified large scale method

The TOP10 *E. coli* cells provided with the Topo®XL cloning kit did not produce high yields of DNA using the maxiprep method described above. The handbook provided by Qiagen stated that TOP10 cells grow to a higher density, resulting in a high biomass that impedes the release of the plasmid DNA from the cells, and obstructs the syringe filter. They recommended doubling the volumes of lysis buffers and starting with less bacterial culture. We used smaller overnight cultures (100 ml of LB with 50 µg/ml of kanamycin), and doubled the amount of lysis buffers used. In addition, we also incorporated a 20 minute incubation on ice and an extra 20 minute centrifugation step (9,000 rpm at 4°C) after the addition of the neutralization buffer, to pellet the majority of insoluble material prior to the application of the lysate to the Qiagen filter. These changes

were successful in providing larger amounts of DNA, and these DNA pellets were resuspended in 300 µl to 1 ml of 10 mM Tris-HCl, pH 8.0, depending on the size of the pellet. Quantification was done by spectrophotometry and the DNA was stored at -20°C until needed.

### 2.1.5 Isolation of Bacterial artificial chromosome (BAC) DNA

The identification of BAC clones containing the *PHGDH* gene required a cDNA probe of the gene of interest to screen the RPCI-11 human BAC library (CIHR Genome Resource Facility, Toronto, ON). The full-length cDNA of *PHGDH* contained within a pcDNA3.1(-) plasmid was already available within the Pind laboratory. The probe was removed from the plasmid by a double restriction enzyme digest using *Bam*H I and *Xba* I. The digested DNA was precipitated with linear polyacrylamide, separated via electrophoresis (0.8% agarose and guanosine) and the 1.6 kb band was gel purified using the Qiagen gel extraction kit. The purified DNA was quantified by agarose gel and ethanol precipitated to obtain a concentration of 30 ng/µl (see below for methods) and sent to Toronto for the identification of BAC clones.

The bacterial artificial chromosome clones: H_NH0334J05, H_NH0401N19, and H_NH0425C16 were obtained from the CIHR Genome Resource Facility, Department of Genetics, The Hospital for Sick Children, Toronto, ON. A small-scale DNA isolation protocol was provided with the BAC clones. The large size of the genomic DNA inserts

(average size 140 kb) required gentle handling during isolation to prevent shearing of the DNA fragments.

The isolation protocol began with pelleting the 3.0 ml of overnight culture (LB media with 10 μg/ml chloramphenicol) at 3,000 rpm for 10 minutes. The pellet was resuspended in 0.3 ml of solution 1 (15 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 μg/ml RNase A), and 0.3 ml of solution 2 (0.2 N NaOH, 1% SDS) was added and inverted to mix. This was incubated at room temperature for 5 minutes prior to the addition of 0.3 ml of solution 3 (3M potassium acetate, pH 5.5) and the incubation was continued on ice for 5 minutes. The cellular debris was pelleted by centrifugation at 3,000 rpm for 10 minutes and the supernatant was transferred to a clean tube and clarified by an additional centrifugation for 10 minutes at 13,000 rpm. The supernatant was removed to a tube containing 0.8 ml of cold isopropanol, inverted to mix, incubated on ice 5 minutes, and then centrifuged at 13,000 rpm for 15 minutes at 4°C. The pellet was washed with 70% ethanol, air-dried and resuspended in 30 μl of $ddH_2O$. The DNA was stored at -20°C.

### 2.1.6 Precipitation and quantification of DNA

Two general methods of concentrating a DNA sample were used. Ethanol precipitation involved mixing one volume of DNA with 0.1 volumes of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ice-cold ethanol. This solution was mixed and stored at –80°C overnight or longer, before pelleting the DNA by centrifugation for 10 minutes

at 4°C at 13,000 rpm. The supernatant was aspirated, the DNA allowed to air dry, and then resuspended in an appropriate volume of 10 mM Tris-HCl, pH 8.0.

The second method of precipitation involved using linear polyacrylamide as a carrier for the DNA molecules. DNA samples were made up to 100 µl in Low TE buffer and mixed with 3 µl of 3 M sodium acetate (pH 5.2) and 3 µl of 5 mg/ml linear polyacrylamide (GenElute, Sigma). Samples were mixed with 250 µl of ice-cold ethanol and precipitated at 13,000 rpm for 5 minutes. The supernatant was removed by aspiration, and the DNA was resuspended in an appropriate volume of Low TE buffer.

A description of the quantification of small amounts of DNA using agarose gel electrophoresis by comparison to a DNA mass ladder (Invitrogen) is provided below. Larger amounts of DNA were quantified by absorption, as measured using a spectrophotometer (Amersham Biosciences). In this manner, a small aliquot (1-10 µl) of DNA was mixed with 10 mM Tris-HCl, pH 8.0, to a volume of 1.0 ml, and it's absorbance at 260 and 280 nm was determined. A sample without DNA (buffer only) was used to standardize the spectrophotometer at each wavelength. It has been shown that for the typical 1 cm path (cuvette), 50 µg/ml of double-stranded DNA gives an absorbance value of 1.0 at 260 nm (90). The concentration of double stranded DNA (in µg/ml) was then determined by multiplying the absorbance at 260 nm by 50. The ratio of DNA absorbance at 260 nm to its absorbance at 280 nm was also used to estimate the purity of the sample, with the highest purity obtaining a ratio of 1.8 for double stranded DNA (90).

## 2.2 Resolution of DNA by gel electrophoresis

Pools of DNA from PCR, restriction digests, or DNA sequencing, were separated by size using gel electrophoresis. The sizes of the DNA bands were estimated visually, by comparing the distance migrated from the well with the distances that standards of known sizes migrated (DNA ladders, Invitrogen). The degree of resolution depended upon the concentration and type of gel used, with agarose providing a less refined separation than that of polyacrylamide, which is capable of resolving single base pair differences.

### 2.2.1 Polyacrylamide gel electrophoresis (PAGE)

PAGE was used to separate the products of the *Hsp*92 II digestions in our population studies of the 1468G→A mutation (see below). A 10% polyacrylamide gel solution [10% polyacrylamide, in 1X TBE buffer (130 mM Tris-HCl, 45 mM boric acid, 2 mM EDTA)] was deaerated and 10% ammonium persulfate and TEMED were added to catalyze polymerization. This solution was poured between two glass plates separated by 1.5 mm spacers and allowed to polymerize. DNA samples were separated by electrophoresis in a 1X TBE running buffer at 200 volts for approximately 1.5 hours, or until the dye front had migrated approximately two-thirds of the gel length. The gel was post-stained in a solution of 1X TBE and ethidium bromide for approximately 20 minutes with shaking, before destaining in 1X TBE buffer for about 30 minutes with shaking. The bands were visualized by ultraviolet illumination and photographed.

### 2.2.2 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis provided a quick method to estimate the size of a DNA band from PCR amplification, restriction digests, or DNA prepared after cloning. DNA bands were cut from agarose gels and purified, when required (see below). Gels with low percentages of agarose were used to resolve the PCR products containing introns, which were generally large, and the high concentration gels were used to study smaller products.

Agarose gels of 0.8, 1.0, 1.5, or 2.0 % concentration were prepared with 1X TAE buffer (40 mM Tris-HCl, 20 mM acetic acid, 1 mM EDTA). Prestained gels were made by adding ethidium bromide to a final concentration of approximately 0.5 μg/ml. A running buffer of cold 1X TAE buffer was used and the electrophoresis conditions were generally 80 volts for 1 hour. Visualization of the DNA bands was done under ultraviolet light and Polaroid photographs were taken, if required.

### 2.2.3 Agarose gel electrophoresis for purification of DNA fragments

Agarose gels were also used to purify DNA fragments that were to be cloned. In these cases, guanosine was added to the stock agarose solution at a final concentration of 0.28 mg/ml (91). After electrophoresis, a small band of the agarose containing the DNA to be purified was cut from the gel under ultraviolet light. The Qiagen QIAquick gel extraction kit was used to extract DNA from agarose gel slice and the manufacturer's protocol was followed. Briefly, this involved incubation of the gel slice at 50°C for

10 minutes in a volume of buffer (dependent on the size of the gel slice) and 10 µl of a resin that binds DNA. This mixture was pelleted by centrifugation and the supernatant was discarded. The pellet was washed with 500 µl of buffer, resuspended by mixing, and pelleted again. This was repeated twice with 500 µl of a second buffer before air-drying the pellet, and eluting the DNA in 20 µl of 10 mM Tris-HCl (pH 8.0). The eluate was separated from the beads by centrifugation and removed to a fresh tube.

### 2.2.4 Quantification of DNA using an agarose gel and a DNA mass ladder

In addition to sizing a DNA sample by comparison with a DNA ladder, a DNA sample can be quantified by comparison with a DNA mass ladder that contains bands of known quantities. An aliquot of the DNA sample to be quantified was separated by electrophoresis in an agarose gel next to 2 µl of a DNA mass ladder (Invitrogen). The intensity of the DNA samples was compared to the mass ladder standards to estimate the concentration of DNA in the sample.

### 2.2.5 Electrophoresis of sequencing products

The sequencing gel solution [6% acrylamide, 7 M urea in 1X TBEs buffer (89 mM Tris-HCl, 89 mM boric acid, 2 mM EDTA)] was deaerated and 25% ammonium persulfate and TEMED were added to catalyze the polymerization, just prior to injection via syringe between the 0.4 mm spaced plates. Sharkstooth combs were used, and one of the plates was siliconized regularly. The gel was allowed to polymerize for a minimum of 1 hour and was then preheated at 60 watts for about 1 hour until the temperature of the

gel was approximately 50°C. The running buffer was 1X TBEs buffer and the sequencing samples were preheated for 2 minutes at 75-80°C. Samples were separated by electrophoresis at 60 watts for 2-6 hours. The gel was fixed by washing in 1 L of a 10% methanol:10% acetic acid (v/v) for 15 minutes, then transferred to blotting paper and dried at 80°C for 1 hour under vacuum. The dried gel was exposed to autoradiography film in a cassette for 15-50 hours. The level of radioactivity was estimated using a Geiger counter and this level was used to predict exposure times.

## 2.3 Restriction endonuclease digests

Restriction enzyme digests were used to analyze small-scale plasmid preparations, to detect the mutant 1468A allele in PCR products, and for cloning insert DNA into the pcDNA3.1 (-) vector. The reactions were performed as per manufacturers recommendations (enzymes were purchased from New England Biolabs, unless otherwise indicated). Generally, a reaction mix consisted of a 10X buffer and 10X bovine serum albumin (if required) with a final concentration of 1X, 10-20 units of restriction enzyme and DNA (2-5 µl), in a total volume of 30 or 50 µl. This mixture was incubated at 37°C for 2-3 hours. The digests were separated by electrophoresis immediately following the digestion, or stored at –20°C.

Due to the large number of samples tested for the 1468G→A mutation, the typical reaction mix was modified slightly to simplify the process. The 10X buffer K (Promega),

supplied with the enzyme, was pre-diluted to a 1X concentration. A standard master mix (for 17 reactions) included: 102 µl of 1X buffer K, 51 µl of 10X bovine serum albumin, and 17 µl of the *Hsp*92 II enzyme. A 20 µl aliquot of the PCR reaction was mixed with the 10 µl of the digest mix and incubated at 37°C for 2-3 hours. Orange G loading buffer [6 µl of 5X concentration, (5X: 15% Ficoll 400, 0.25-0.3% Orange G)] was added to each tube and the samples were analyzed immediately or stored overnight at 4°C. An uncut control was created by mixing 10 µl of the PCR-amplified sample with 20 µl $ddH_2O$ and 6 µl 5X Orange G loading buffer.

## 2.4 Amplification of DNA templates by the polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) method of amplifying DNA uses gene-specific primers that anneal to template DNA, and a heat-stable DNA polymerase enzyme that extends the primers to create the DNA product. The reactions must be optimized for each primer set, DNA template, and PCR enzyme, but a general reaction mixture and thermal cycler program is recommended by the manufacturer of each PCR enzyme. All PCR mixtures were assembled on ice and were performed in thin-walled 500 µl tubes in a MJ-Research thermal cycler. Instead of oil overlays, a heated lid on the thermal cycler was used to prevent evaporation of the samples during cycling. The sequences of all primers described in this thesis are listed in Appendix I.

### 2.4.1 Amplification of the region of *PHGDH* around nucleotide 1468

The amplification of genomic DNA was performed to produce DNA samples that could be digested to reveal the presence of an adenine nucleotide at position 1468. The primers 1 and 2 were used to amplify a 238-base pair (bp) region near the C-terminus of the gene. Primer 1 was within an intron, 115-93 bp upstream of cDNA nucleotide 1448. The reverse primer (#2) was localized to cDNA nucleotides 1572 to 1553.

The genomic DNA samples to be amplified (2 µl each) were mixed with 8 µl of autoclaved $ddH_2O$ in 500-µl tubes. The master mix was assembled on ice as follows (for 16 samples): 489.8 µl of $ddH_2O$, 80 µl of 10X PCR buffer, 16 µl of 10 mM dNTP mixture, 16 µl of 100 pmol/µl forward primer (#1), 16 µl of 100 pmol/µl reverse primer (#2), 24 µl of 50 mM $MgCl_2$, and 4.8 µl of *Taq* polymerase (Invitrogen). After thorough mixing, 40 µl of this solution was dispensed to each tube containing an aliquot of genomic DNA. The PCR conditions were: 94°C for 2 minutes, followed by 31 cycles of 94°C for 30 seconds, 65°C for 30 seconds and 72°C for 30 seconds. One final elongation step of 10 minutes at 72°C occurred before the cycler stored the samples at 4°C. The samples were used directly or stored at 4°C for 1-3 days.

Some samples were amplified using PCR Supermix (Invitrogen). Supermix is a pre-mixed solution of buffer, enzyme, dNTPs and water, and only requires the addition of gene-specific primers and DNA template. 720 µl of Supermix (enough for 16 reactions) was mixed with 16 µl of forward primer, 16 µl of reverse primer, and 16 µl of $ddH_2O$,

before 48 µl aliquots of this master mix were dispensed into 500 µl PCR tubes. Genomic DNA (2 µl) was added and amplified as described above.

### 2.4.2 Amplification of DNA templates for the SHMT probes

The production of the two SHMT RNA probes first required the synthesis of DNA templates. Since we wanted an accurate DNA template that would produce RNA transcripts with minimal errors, the genomic DNA was amplified with *Pfu* polymerase (Stratagene) that had a higher fidelity rate than *Taq* polymerase. Careful study of the known cDNA and genomic sequences of the *SHMT* genes ensured that the PCR primers amplified regions with minimal similarity to the other isozyme and lay within one exon (preventing the inadvertent amplification of intronic DNA in the PCR reaction).

Primers 3 and 4 were used to amplify the cytosolic *SHMT* fragment, and primers 5 and 6 amplified the mitochondrial *SHMT* fragment. The upstream primers for each pair, primers 3 and 5, had a *Hind* III restriction enzyme recognition site engineered on the 5' end, indicated in bold (Appendix I). The downstream, or reverse, primers (4 and 6) had *Xba* I restriction sites added to their 5' ends, shown underlined. These restriction sites allowed for unidirectional ligation into the pcDNA3.1(-) vector. When the DNA template was used for RNA transcription from the T7 promoter, the gene fragments were transcribed in the opposite orientation, so that the RNA transcripts were antisense strands, and would bind to the sense strands present on the Northern blots.

Primer 3 was located in exon 13 of cSHMT from nucleotides 1319-1342 and primer 4 was located in exon 13 at nucleotides 1633-1610. These primers amplified a 315 bp piece of the 3' untranslated region of the cytosolic *SHMT1* gene (Genbank accession number L23928). The forward primer (#5) used to amplify the mitochondrial *SHMT2* fragment was located in exon 11, from nucleotides 1358-1379 (Genbank accession number L11932). The reverse primer (#6) was located at nucleotides 1671-1647. The PCR product from this primer set was a 290 bp fragment of *mSHMT*.

The PCR master mixes contained (for 4 reactions): 156 µl of $ddH_2O$, 20 µl of *Pfu* 10X PCR buffer (containing 20 mM $MgS0_4$), 4 µl of 100 pmol/µl forward primer, 4 µl of 100 pmol/µl reverse primer, 4 µl of 10 mM dNTP mix, and 4 µl of *Pfu* polymerase. Aliquots of 48 µl were mixed with either 2 µl of $ddH_2O$ (water blank) or 2 µl of human genomic DNA. The PCR conditions were an initial denaturation of 2 minutes at 94°C, 31 cycles of 30 seconds at 94°C, annealing for 30 seconds at 55°C for cSHMT or 60°C for mSHMT, and elongation for 1 minute at 72°C. A final elongation time of 10 minutes at 72°C occurred before incubation at 4°C for an indefinite amount of time.

### 2.4.3 Amplification of the introns within the *PHGDH* gene

To amplify the introns within the *PHGDH* gene, an initial set of primers, listed in Appendix I, was designed to divide the coding region into segments of approximately 200-300 bp. A high fidelity product was desired to provide accurate sequences at the intron-exon boundaries, but the enzyme had to be capable of producing long PCR

products since the size of the introns were not known. The Expand High Fidelity PCR enzyme (Roche) was chosen for these experiments.

The PCR mixture was (for one reaction): 40.4 µl of $ddH_2O$, 5 µl of 10X PCR buffer (containing 15 mM $MgCl_2$), 1 µl of 10 mM dNTP mixture, 1 µl of 100 pmol/µl of forward primer, 1µl of 100 pmol/µl of reverse primer, 1 µl of BAC DNA and 0.8 µl of enzyme. This was assembled on ice, and mixed prior to thermal cycling. The same PCR mixture was used for the amplification of introns 2 through 11, but the PCR conditions varied with each fragment. The PCR mixture for intron 1 had 5% dimethylsulfoxide (2.5 µl) and only 37.9 µl of $ddH_2O$; all other components were the same. The general PCR conditions began with an initial denaturation of 2 minutes at 94°C, then 10 cycles of denaturation at 94°C for 30 seconds, annealing at X°C for 30 seconds and elongation at 68°C for Y minutes (see the Results section for exact annealing temperatures and elongation times for each primer pair). This was followed by 20 cycles of denaturing and annealing as in the first 10 cycles, but with an additional 5 seconds per cycle added to the elongation time from the first 10 cycles. A final elongation time of 10-30 minutes at 72°C assisted with the correct synthesis of the PCR products (the time increased with length of product).

### 2.4.4 Amplification of the 5' end of the *PHGDH* cDNA from human liver

To determine the transcriptional start site of the *PHGDH* gene in human liver, a FirstChoice™ RACE-Ready cDNA kit (Ambion) was purchased. The manufacturer

removes the cap structure from full-length mRNA, replacing this with an RNA linker and this new sequence with linker is reverse transcribed to create cDNA products enriched for full-length sequences. The kit includes two primers specific for the linker DNA at the 5' end of the cDNA sequences. By combining these primers with gene-specific primers in a nested PCR approach, the 5' ends of the cDNA were amplified.

Nested PCR is a two-step process. The first PCR reaction used a forward primer located on the 5' end of the linker region and a gene-specific reverse primer. A small aliquot from this reaction was used for a second PCR amplification with a second forward primer located near the 3' region of the linker DNA and a second gene-specific reverse primer that was upstream from the first reverse primer. See Figure 6 for a schematic diagram that illustrates the position of the primers and the resulting PCR products.

The first PCR reaction was performed in a mixture of (for one reaction): 40 µl of $ddH_2O$, 5 µl of 10X buffer (containing 1.5 mM $MgCl_2$), 1 µl 10 mM dNTP mix, 1 µl of outer linker primer (Appendix I), 1 µl of gene-specific reverse primer (primer 16, nucleotides 603 to 581), 1 µl of Expand High Fidelity enzyme (Roche), and 1 µl of liver cDNA (provided in the Ambion kit). The PCR conditions were: 1 cycle of 94°C for 3 minutes, followed by 10 cycles of 30 seconds at 94°C, 30 seconds at 60°C and

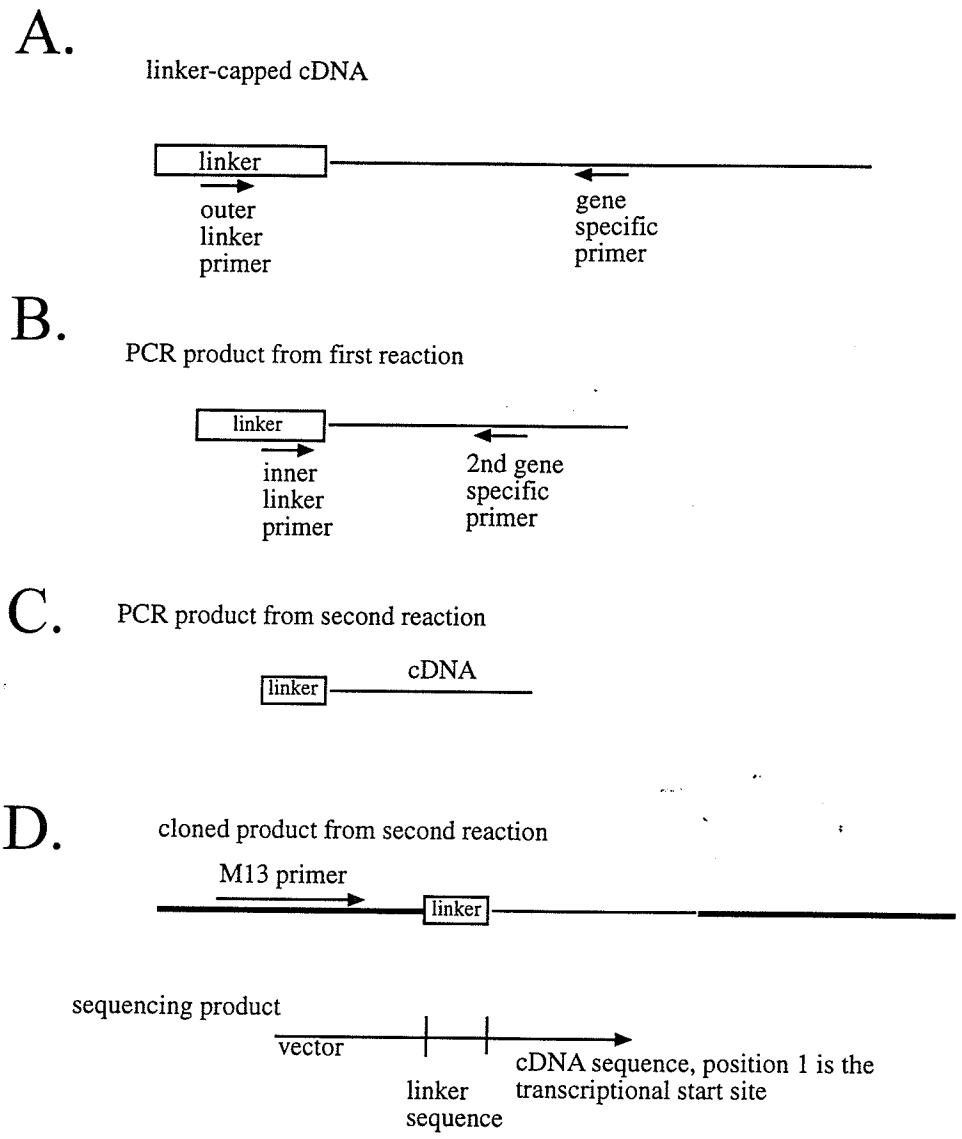



Figure 6. Approach to determine the transcriptional start site of human liver using the Ambion 5' RACE kit. The horizontal line represents the *PHGDH* cDNA, the box represents the linker DNA that has been added to the cDNA's 5' end. A. The outer linker primer and the first gene specific primer are used to PCR amplify a product with a section of linker and a segment of PHGDH. B. The product of the initial PCR reaction is amplified with the inner linker primer and a second gene specific primer. C. The product of the second PCR reaction contains a small piece of linker and a segment of PHGDH. D. The product from C, cloned into a TopoXL vector (shown in bold) prior to sequencing. The sequencing product is read from the vector-based sequence, through the linker sequence, and the first non-linker based sequence is the first nucleotide in the cDNA, which is the transcriptional start site.

1.5 minutes at 68°C. The next 25 cycles were had an additional 5 seconds per cycle added to the elongation time. A final extension time of five minutes at 72°C was performed before storage at 4°C.

The second PCR reaction used (for one reaction): 39 µl of $ddH_2O$, 5 µl of 10X buffer (containing 1.5 mM $MgCl_2$), 1 µl of 10 mM dNTP mix, 1 µl of 10 µM of inner linker primer (see Appendix I), 1 µl of reverse primer (primer 10, nucleotides 321 to 345), 1 µl of Expand High Fidelity enzyme, and 2 µl of PCR product from the first reaction. The amplification conditions were identical to the first reaction.

### 2.4.5 Amplification strategy to survey for possible alternate transcription of exon 1 in human tissues

With the publication of a possible alternative exon 1 of *PHGDH* in rat testis (82), it was decided that the possibility of alternative transcription of human *PHGDH* should be investigated. A cDNA array of 24 human tissues (Origene) would allow PCR amplification of the 5' region of human *PHGDH* with gene-specific primers located in the two possible exon 1 regions and a downstream primer in exon 5. Primer 29 is a forward primer that corresponds to exon 1 (nucleotides -40 to -61) and should amplify a 485 bp fragment with the reverse primer 30 (located at nucleotides +403 to +424) when exon 1 of *PHGDH* is present in the cDNA population. Primer 31 is a forward primer corresponding to the possible exon 1' at nucleotides -315 to -290 and should amplify

(with reverse primer 30) a 486 bp fragment if exon 1' is spliced onto exon 1 during transcription.

The PCR mixture involved (for 28 reactions): 357 µl of $ddH_2O$, 140 µl of 10X PCR buffer, 70 µl of dimethylsulfoxide, 42 µl of 50 mM $MgCl_2$, 28 µl of 10 mM dNTP mix, 28 µl of 100 pmol/µl forward primer, 28 µl of 100 pmol/µl reverse primer, and 7 µl of *Taq* DNA polymerase (Invitrogen). 25 µl of this mixture was added to 25 µl of cDNA (approximately 100 or 200 pg) for the *PHGDH* products. For amplification of the *β-actin* control, the primers were supplied with the kit (see Appendix I for sequences), the dimethylsulfoxide was omitted, and 25 µl of cDNA (approximately 10 pg) was used.

The PCR conditions were: 94°C for 3 minutes, 35 cycles of 94°C for 30 seconds, 57°C (primers 30 and 31) or 60°C (primers 29 and 30) for 30 seconds and 72°C for 1 minute, followed by a final extension of 10 minutes at 72°C. The β-actin control reaction involved 94°C for three minutes, 31 cycles of 94°C for 30 seconds, 55°C for 30 seconds and 72°C for two minutes, followed by a final extension of five minutes at 72°C.

## 2.5 Cloning

### 2.5.1 Ligation of insert DNA into the pcDNA3.1(-) vector

DNA fragments were cloned into the pcDNA3.1(-) vector using an approximate 3:1 molar ratio of insert to vector and approximately 50 ng of vector. The ligation

reaction was assembled on ice and consisted of 1 µl of 10X ligation buffer, 50 ng of vector DNA, a volume of insert DNA, and 1µl of T4 ligase (Roche) in a total volume of 10 µl. Ligations were performed in a thermal cycler at 14°C for eight hours, followed by a 20 minute incubation at 65°C to inactivate the enzyme, and a final storage step at 4°C.

### 2.5.2 Ligation of insert DNA into the Topo® XL vector

The fragments of *PHGDH* that were amplified to study the location of the introns within the *PHGDH* gene were cloned using Invitrogen's Topo® XL PCR cloning kit. This kit was also used to clone the PCR products from the 5' RACE reaction. This kit operates on a T/A overlap system, and does not require PCR primers engineered with unique restriction sites. Purified DNA from a PCR reaction (4 µl) was mixed with the Topo®XL vector (1 µl) and incubated at room temperature for five minutes, and then stopped with 1 µl of 6X stop solution.

### 2.5.3 Transformation of ligated plasmids into *E. coli* cells

The bacteria transformed with the pcDNA3.1(-) vector were *E. coli* XL-1 Blue cells (Stratagene) that had previously been treated with calcium chloride to increase their ability to incorporate plasmid DNA. These were stored at –80°C in 150 µl aliquots. The cells were thawed and 30 µl of cells removed to a new tube where 1 µl of ligation mixture was added.

The Topo® XL kit provided chemically competent TOP10 bacterial cells, and these were used to clone the large-sized *PHGDH* intron fragments for sequencing. 2 µl of ligation mixture was added to one vial (50 µl) of the *E. coli* TOP10 cells.

The general transformation protocol, used for both pcDNA3.1(-) and Topo® XL vectors, was 30 minutes on ice, 30-60 seconds of heat shock at 42°C and then 2-5 minutes on ice. To each tube, 250 µl of SOC media (2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $Mg_2SO_4$, 20 mM glucose, pH 7.0) was added before incubation for one hour at 37°C with shaking (250-300 rpm). The transformed bacteria was spread plated onto prewarmed LB plates (with antibiotic), and incubated overnight at 37°C. The plates were wrapped with Parafilm and stored inverted at 4°C.

### 2.5.4 Preparation of frozen stocks of bacteria

Plasmids of interest were stored as purified DNA and also in small amounts of bacterial cells. A 0.85 ml aliquot of overnight bacterial culture was added to 0.15 ml of glycerol, mixed well, and stored at -80°C in a 2 ml cryovial.

## 2.6 Manual dideoxy sequencing of plasmid DNA

Small sections of the *PHGDH* gene from BAC DNA, as well as the 5' rapid amplification of the cDNA ends, were PCR amplified and cloned into Topo® XL vectors, as described. A T7 sequencing kit (Amersham Pharmacia Biotech) was used to sequence the Topo® XL-DNA templates purified from the bacterial colonies (in either small or

large scale preparations). The general scheme was to denature the double-stranded DNA with heat and high pH, then neutralize the solution and precipitate the DNA. The sequencing primers were annealed to the single-stranded template and elongated in the presence of both radioactively labelled nucleotide ($\alpha^{35}S$) and a mixture of dideoxynucleotides (92). This created a pool of shortened products that when separated by size in a urea:polyacrylamide gel revealed the sequence of the DNA template.

### 2.6.1 Sequencing the cloned *PHGDH* fragments

The DNA template (0.2-0.3 µg of DNA in Low TE to a total of 10 µl) was denatured by the addition of 10 µl of denaturing solution (0.4 M NaOH, 0.4 mM EDTA) and incubation at 65°C for 15 minutes. This was neutralized with 2 µl of 2 M ammonium acetate, pH 4.5, and precipitated using 55 µl of ice-cold ethanol. This solution was stored at –80°C for at least eight hours. The template DNA was precipitated by centrifugation at 13,200 rpm at 4°C for five minutes, and the supernatant was gently removed by pipette. When miniprep DNA was used, the pellet was washed with 70% ethanol and the centrifugation step repeated. The pellet was air dried and dissolved in 9 µl of $ddH_2O$ (except intron 1 fragments, with 6 µl of $ddH_2O$). The M13 primers (forward and reverse) that anneal to the Topo® XL vector were supplied with the Topo®XL cloning kit and were diluted to 20 pmol/µl in 10 mM Tris-HCl, pH 8.0, prior to use (see Appendix I).

For each sample of denatured template DNA (9 µl), 2 µl of primer and 2 µl of annealing buffer were incubated at 65°C for five minutes, then at room temperature for

10 minutes. The annealed primer/template was labelled with 2 µl of $\alpha^{35}$S-dATP (NEN, 1250 Ci/mmol), 3 µl of labelling buffer, and 2 µl of T7 DNA polymerase (1.6 U/µl) during a room-temperature incubation of 5 minutes. During this time, the 2.5 µl aliquots of "short" mixes A, C, G, T were pre-warmed at 37°C. The label mixture was dispensed into 4.5 µl aliquots, mixed with the short mixes, and then incubated at 37°C for five minutes before the addition of 5 µl of stop buffer (with formamide). Samples were analyzed immediately or stored at -20°C for up to two days.

When the sequencing reaction was performed with the DNA fragment containing intron 1, the sequencing gel displayed a large smear. Troubleshooting information provided by the manufacturer suggested the addition of dimethylsulfoxide in the annealing procedure may alleviate secondary structure or aid in the denaturation of difficult templates. The denatured, precipitated template DNA of intron 1 was resuspended in 6 µl of $ddH_2O$, mixed with 2 µl of annealing buffer, 2 µl of 20 pmol/µl M13 primer, and 3 µl of dimethylsulfoxide. This was incubated at 37°C for 20 minutes, then 10 minutes at room temperature. The labelling and extension portions of the sequencing protocol were identical to the protocols used for introns 2 through 11.

## 2.7 Northern blots

The DNA templates of the *SHMT* and *PHGDH* fragments, contained in the pcDNA3.1(-) vectors, were transcribed into RNA probes before hybridization with the

Northern blots. As mentioned previously, the cloning was done in a specific direction, to provide anti-sense RNA transcripts that would anneal to the sense-strand RNA on the Northern blots. The RNA probes were synthesized with a Strip-EZ™ kit (Ambion) and radiolabelled UTP ($\alpha^{33}$P). The DNA template was then degraded by DNase treatment, and the probe separated from unused nucleotides and degraded DNA template via spin column purification. The Strip-EZ™ kit created a probe with modified cytidine nucleotides that can be degraded with a buffer under mild conditions after the blot had been analyzed. The mild removal of the labelled probe was supposed to extend the life of the Northern blots by avoiding the harsh temperatures usually required to strip and reuse blots.

### 2.7.1 Synthesis of RNA probes by transcription of DNA templates

All reagents were thawed and kept on ice, except for the transcription buffer that was kept at room temperature. The reaction mixture was assembled at room temperature and consisted of: 7 µl of 0.1 µg/µl DNA template, 2 µl of 10X transcription buffer, 1 µl each of ATP, GTP, modified CTP and diluted, unlabelled UTP, 5 µl of $\alpha^{33}$P-UTP (NEN, 3000 Ci/mmol), and 2 µl of RNA polymerase/RNase inhibitor (Ambion). This was mixed and incubated at 37°C for one hour. The RNA transcription reaction was treated with DNase (1 µl, Ambion) at 37°C for 15 minutes, and brought to 100 µl with 10 mM Tris-HCl buffer before purification via spin columns (ProbeQuant™ G50 Spin columns, Amersham Biosciences). The purified probe was mixed with UltraHyb™ solution

(Ambion) to a final concentration of approximately 1 million cpm per ml of hybridization solution.

### 2.7.2 Hybridization of RNA probes with Northern blots and analysis

All incubations were performed at 68°C with a gentle, continuous rotation of the hybridization bottles within a hybridization incubator (Robbins Scientific Corp., Sunnyvale, CA). The blots were prehybridized in UltraHyb™ solution for one hour at 68°C, before addition of the labelled probe and incubation overnight. The blots were washed twice with preheated 2X SSC (0.3 M NaCl, 30 mM sodium citrate), 0.1% SDS solution for 5 minutes, then twice with 0.1X SSC (15 mM NaCl, 1.5 mM sodium citrate), 0.1% SDS solution for 15 minutes at 68°C. The blots were removed to pre-soaked (in 0.1% SDS) blot paper and wrapped with plastic wrap before exposure to a phosphor screen for 20-70 hours. The phosphor screen was read by a Personal Molecular Imager® FX (BioRad) and the computer program QuantityOne (BioRad) was used to quantify the data.

### 2.7.3 Stripping and storage of Northern blots

The protocol for stripping the blots was provided with the StripEZ™ kit. The probe degradation buffer (200X) and blot reconstitution buffer (100X) were diluted in 0.1% SDS before added to the blots. The degradation buffer was incubated for 10 minutes at 68°C, and discarded before the blot reconstitution buffer was applied to the blots for a 10 minute incubation. The blots were washed with 0.1% SDS and removed to

damp blotting paper (pre-soaked in 0.1% SDS), wrapped in plastic wrap and exposed to a phosphor screen to ensure the complete removal of the previous probe. Blots were stored at -20°C for possible further use.

## 2.8 Detection of the PHGDH protein in human brain tissue using immunohistochemistry

During the course of these studies, mouse monoclonal antibodies to human PHGDH were developed. They were produced by the Canadian Genetic Disease Network Immunoprobes Facility (Winnipeg, MB) using recombinant human PHGDH protein purified from *E. coli* in the Pind laboratory (3). Thirty-six positive clones were identified and tested for their ability to recognize PHGDH in paraffin-embedded human brain tissue. Two were selected based upon high levels of immunoreactivity and were grown in large-scale cultures. The antibodies from these two clones (1C2 and 8B6) were purified from hybridoma lysates by protein-G affinity purification.

The basic peroxidase-antiperoxidase immunohistochemistry protocol involved deparaffinization, rehydrating the tissue, incubating with primary antibodies, biotinylated secondary goat anti-mouse antibodies, avidin-biotin complex (ABC) reagent (Vector) and diaminobenzidine substrate, counterstaining, dehydrating the tissue and mounting. The analysis was done with light microscopy (Olympus BX51). Photographs were obtained using image acquisition software and a digital camera (SPOT 32) mounted on the microscope.

Immunohistochemistry reagents and protocols were provided by Dr. D. Eisenstat (of the Manitoba Institute of Cell Biology). Paraffin-embedded sections of human brain tissues on slides were provided by Dr. D. Eisenstat and Dr. M. Del Bigio (Department of Pathology, University of Manitoba). These were baked in a hybridization oven at 60°C for 20 minutes to melt the paraffin. The slides were incubated 10 minutes in xylene, followed by two minutes each in xylene (twice), 100% ethanol (twice), 95% ethanol, 85% ethanol, 75% ethanol, and 50% ethanol, before a five minute incubation in distilled water. The slides were then incubated for 15 minutes in a solution of 1X PBS (0.9% NaCl, 10 mM sodium phosphate, pH 7.2), 0.05% TritonX-100, and 0.3% hydrogen peroxide. The slides were washed three times for five minutes each in a solution of 1X PBS, 0.05% TritonX-100.

The slides were placed onto trays in a humidity chamber (Fisher Scientific) and incubated for two hours at room temperature with 300 µl of blocking solution (1X PBS, 5% goat serum, 0.2% TritonX-100, 0.02% sodium azide, 0.1% bovine serum albumin) per slide. The excess blocking solution was poured off and 300 µl of the primary antibody (diluted in blocking solution) was added to each slide. The primary antibody was incubated with the slides overnight at 4°C in the humidity chamber. The primary antibodies were incubated at full-strength to first test the immunoreactivity of each clone, and immunoreactive clones were tested at lower dilutions (1:5, 1:100, etc.) to determine a good working titre. The purified 1C2 antibody (with an approximate protein

concentration of 0.5 mg/ml) was used at a dilution of 1:2500. Following the overnight incubation, the excess solution was poured off and the slides were washed in 1X PBS, 0.05% TritonX-100, three times for five minutes each. The slides were again placed in the humidity chamber and the secondary biotinylated goat anti-mouse antibody was added. Each slide was incubated in 300 µl of the secondary antibody, diluted 1:400 in blocking solution, for two hours at room temperature. The excess liquid was poured off and the slides washed again, three times for five minutes each. The tertiary antibody solution (Vectastain ELITE PK6100, Vector Laboratories, Burlingame, CA) was prepared by adding one drop of buffer A and one drop of buffer B to 2.5 ml of 1X PBS, mixing well by vortexing between each addition. This solution was covered with foil and incubated at room temperature for 30 minutes. The slides were placed in the humidity chamber and 300 µl of the tertiary antibody solution was added to each slide and incubated at room temperature for 30 minutes. The slides were washed again (3 x 5 min), and the diaminobenzidine (DAB, Vector Laboratories) substrate was prepared. To 2.5 ml of distilled water, one drop of buffer, two drops of substrate, and one drop of hydrogen peroxide were added separately and mixed by vortexing after each addition. The slides were placed in clear dishes and 125 µl of DAB was added to each slide, a few slides at a time. The reaction was allowed to progress until significant brown staining appeared (usually 2-10 minutes). The reaction was stopped by submerging the slides in distilled water. After all the slides had been treated with DAB, the slides were incubated in water

for five minutes. The slides were counterstained in hematoxylin, which stains the nuclei of cells, for one minute and washed in approximately seven changes of water (about one minute each), until no more hematoxylin was released into the water.

The stained slides were dehydrated by reversing the order of the rehydration solutions, that is, five minutes in distilled water, then two minutes each in 50% ethanol, 75% ethanol, 85% ethanol, 95% ethanol, 100% ethanol (twice), and xylene (twice). The slides were mounted with coverslips and mounting media (Permount, Fisher Scientific), and allowed to dry at room temperature for several days.

The slides were observed using light microscopes. Areas with distinct brown staining were analyzed in consultation with Dr. Eisenstat. Computer-assisted digital photography was performed.

## 2.9 General Approaches

### 2.9.1 Detection of the 1468A allele in the genomic DNA of Jewish and non-Jewish individuals

With the occurrence of the 1468G→A substitution in two unrelated Ashkenazi Jewish individuals, the prevalence of this allele in the Ashkenazi Jewish population was unknown and we wanted to determine its frequency. To analyze the alleles at position 1468, where G encodes the wild-type and A is a mutation leading to a deficiency of PHGDH, genomic DNA was isolated from white cell pellets, a 238 base pair fragment

was amplified by PCR and digested with the restriction enzyme *Hsp*92 II. This enzyme recognizes and cuts at CATG sequences, the sequence around position 1468 is usually CGTG, with the 1468G shown in bold. When the 1468A allele is present, the sequence becomes CATG, which will be recognized and cut by the enzyme. The digested PCR products were resolved by electrophoresis in polyacrylamide gels and visualized by ultraviolet light. In this manner, the DNA from over 400 individuals of Ashkenazi Jewish ancestry and 400 non-Jewish control individuals were analyzed.

### 2.9.2 Determination of the PHGDH and SHMT mRNA expression profiles in human tissues

The expression profile of PHGDH in the human central nervous system (and other tissues) was investigated due to the neurological problems evident in children with a deficiency of PHGDH. The two isozymes of SHMT were also investigated as possible alternative sources of serine biosynthesis in the brain. In order to characterize the expression profile of the mRNAs encoding human PHGDH, cytosolic and mitochondrial SHMT isozymes, DNA templates were created and transcribed into RNA probes. The PHGDH cDNA was cut by *Apa* I and *EcoR* I, and the 298 bp fragment, from nucleotides 446 to 744, was ligated into a pcDNA3.1(-) vector also cut with these two enzymes. Approximately 300 base pairs of the two 3' untranslated regions of the *SHMT* genes were PCR amplified from a random sample of genomic DNA, and the *Hin*d III and *Xba* I restriction sites engineered onto the 5' ends of the PCR primers were used to clone the

inserts into pcDNA3.1(-). The DNA template was linearized with *BamH* I, purified and quantified by electrophoresis in agarose gels, and ethanol precipitated, if necessary. A concentration of 0.1-0.5 μg/μl was recommended for the T7 RNA transcription kit that was used to synthesize the radioactively labelled RNA probes. The radioactive probes were hybridized (separately) to the Clontech Northern blots overnight, washed, analyzed via phosphorimager, and the probe was then stripped from the blots.

The ability to strip and reuse the blots allowed us to study all three probes on one set of blots. The use of RNA probes required one consideration, and that was that the template DNA must be in reverse orientation to provide an anti-sense probe upon transcription. The RNA on the blots is sense strand, and an anti-sense probe is required to hybridize successfully. This reversal of DNA orientation was performed by selective use of restriction endonucleases for the ligation of digested cDNA (PHGDH) or PCR products (*SHMT*). The scheme for synthesis of the SHMT probes is shown in Figure 7. The scheme for the PHGDH probe is shown in Figure 8.

Since monoclonal mouse antibodies became available during these studies, these antibodies were used to localize PHGDH in specific cell types. Thirty-six clones were tested, and two were subjected to further cloning and purification. Slides with paraffin-embedded human brain tissues were rehydrated, incubated with antibodies, treated with DAB and dehydrated before mounting with coverslips. Detection of immunostaining was captured with a digital camera mounted on an Olympus BX51 microscope.

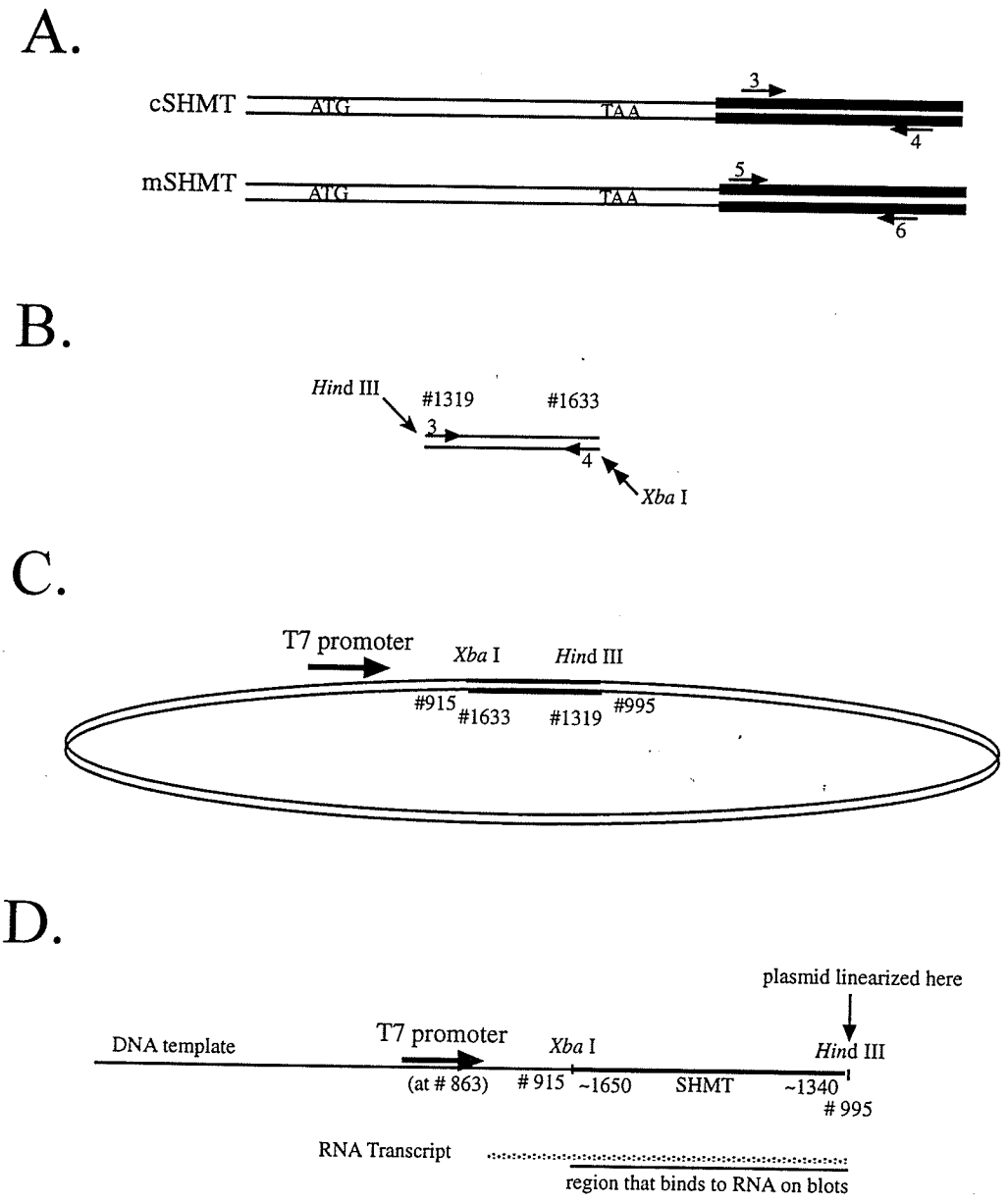
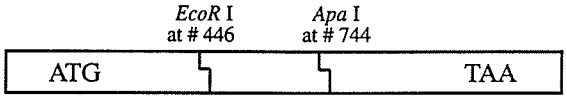
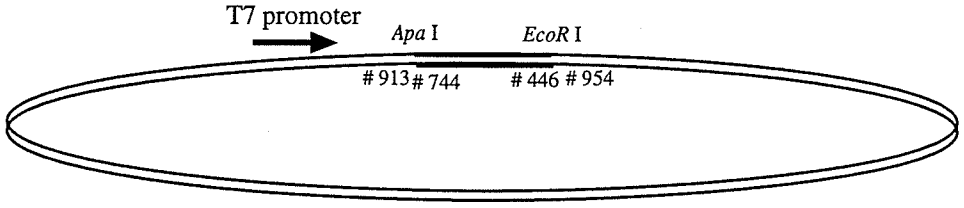



Figure 7. Construction of radiolabelled RNA probes of the *SHMT* isozymes. A. Primers were designed to amplify the 3-untranslated region of each isozyme. B. The PCR products (only one shown) were digested with *Xba* I and *Hind* III, whose recognition sites were engineered into the PCR primers. C. The digested products were ligated into a prepared pcDNA3.1(-) vector, in reverse orientation to the T7 promoter. D. The plasmid was purified and linearized with *Hind* III. A T7 transcription kit (Ambion) was used to produce a radiolabelled RNA transcript, which was hybridized to the Northern blots. See Figures 13 and 14 for the results.

A.




B.




C.

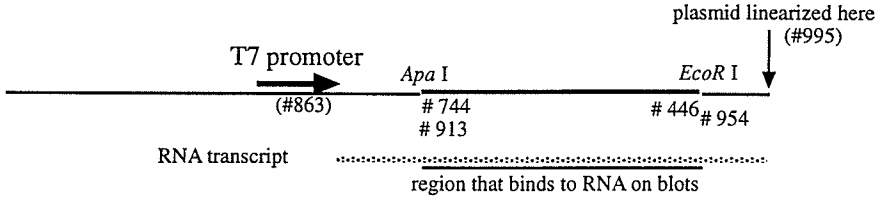



Figure 8. Construction of a radiolabelled RNA probe of *PHGDH*.
A. The cDNA encoding PHGDH was digested with *Apa* I and *EcoR* I, and the 298 bp fragment was gel purified. B. The purified fragment was ligated into a prepared pcDNA3.1(-) vector, in reverse orientation to the T7 promoter. C. The plasmid was purified and linearized with *Hind* III. A T7 transcription kit (Ambion) was used to produce a radiolabelled RNA transcript, which was hybridized to the Northern blots. See Figures 11-B and 12 for the results.

### 2.9.3 Identification of the introns of human genomic *PHGDH*

The goal of this study was to locate the intron-exon boundaries within the cDNA sequence and deduce the approximate length for each intron. Primers located approximately 200-300 base pairs apart along the length of the coding region were used to amplify human genomic DNA carried within BAC clones (Figure 9-A). When the PCR products were greater than expected (i.e. >200bp) the products were purified, and cloned using the Topo® XL T/A cloning kit (Figure 9-B). The DNA was isolated by small scale preparations and the insert size was confirmed by digestion with *EcoR* I and agarose gel electrophoresis. A large-scale plasmid preparation was done before the DNA was manually sequenced [Sanger (dideoxynucleotide) method] (92) (Figure 9-C). The sequence was then compared to the known cDNA sequence and where it diverged was the location of the intron-exon boundary (Figure 9-D).

### 2.9.4 Determination of the possible alternative splicing of the 5' UTR of human *PHGDH*

Analysis of the published rat *PHGDH* gene indicated that one of four transcripts from rat testis had a transcription start site located approximately 250 bp upstream of exon 1 (82). The authors concluded this resulted from alternative splicing of an exon 1' onto the 5' untranslated region, approximately 30 bp upstream from the translational start site (see Figure 4). The human gene has a similar sequence to the rat exon 1' approximately 275 base pairs upstream from human exon 1. To determine if this possible

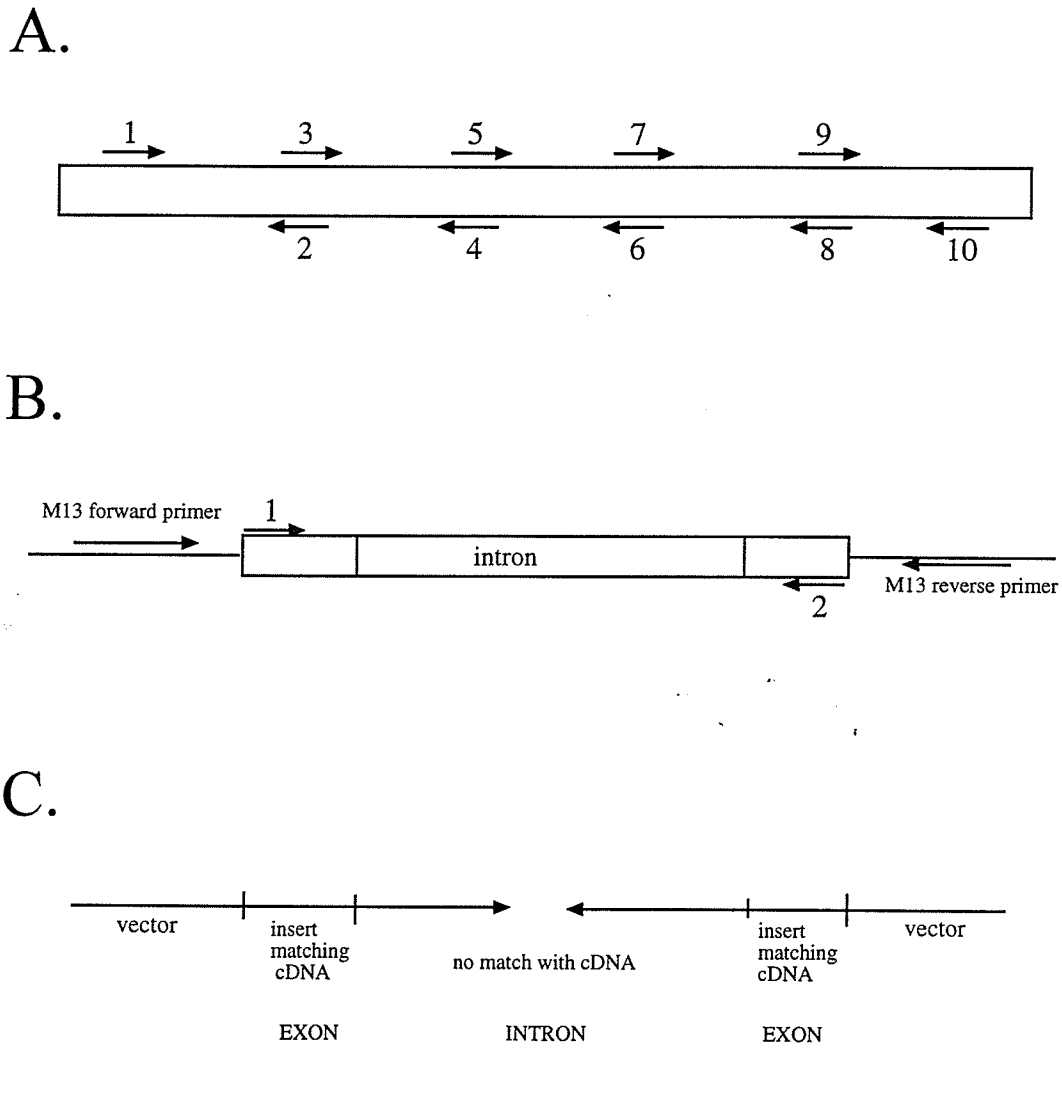



Figure 9. Schematic showing the approach taken to localize and sequence the intron/exon boundaries of human *PHGDH*. A. Primers were designed to amplify ~200 bp fragments of the gene. B. PCR products greater than the expected size were cloned into TopoXL T/A vectors. C. Cloned fragments were subjected to manual dideoxy sequencing. The sequences were compared to the known cDNA sequence, with sites of divergence identified as intron/exon boundaries.

alternative exon 1' was expressed in human tissues, an array of pooled cDNA from 24 tissues was used to amplify two fragments: exons 1-5 or exons 1'-5. The PCR products were resolved on 1.5% agarose gels.

### 2.9.5 Characterization of the transcriptional start site of *PHGDH* in human liver

The determination of the transcriptional start site in human liver was performed by using a PCR-based RACE kit (Ambion). The cDNA from human liver contained modified 5' ends, where a DNA linker has been attached. Gene specific primers and primers that anneal to the linker region were used in a nested PCR protocol and the PCR products were purified by electrophoresis in agarose gels and cloned into a Topo® XL vector. The clones were digested with *EcoR* I to determine the insert size, and with *BamH* I to determine the direction of the insert. This allowed sequencing from only one direction (the end nearest the linker) to decrease the amount of sequencing reactions required. A total of 22 separate colonies were harvested, and their DNA was purified and sequenced, as described.

# 3. RESULTS

## 3.1 An estimation of the frequency of the 1468A allele in the Ashkenazi Jewish population

Consanguinity increases the incidence of several autosomal recessive disorders by raising the likelihood of identity by descent (28). A deficiency of phosphoglycerate dehydrogenase (PHGDH) has been identified in five families, three of which were reported to be due to consanguineous unions (2). It is not known if the 1468A allele is present outside of these families. The point mutation 1468G→A was not detected in sample populations of individuals of Turkish (47) or Western European descent (50) (2), suggesting that this is not a common polymorphism. The discovery of a deficiency of PHGDH in a non-consanguineous Ashkenazi Jewish family (3) raised the possibly that the V490M mutation might exist in the Ashkenazi Jewish population. I performed a PCR-based survey of over 400 Jewish and 400 control individuals. For this purpose, a small region of *PHGDH* was amplified and digestion with *Hsp*92 II revealed whether an individual carried the 1468A allele. This method is illustrated in Figure 10-A, and a sample polyacrylamide gel displaying the possible results is shown in Figure 10-B.

The 238 bp PCR product (Figure 10-B, lane 1) from a control individual without the 1468A allele, yields one 193 bp band (upper band, lane 2) and one 45 bp band (lower band, lane 2) when digested. When an affected individual is homozygous for the 1468A

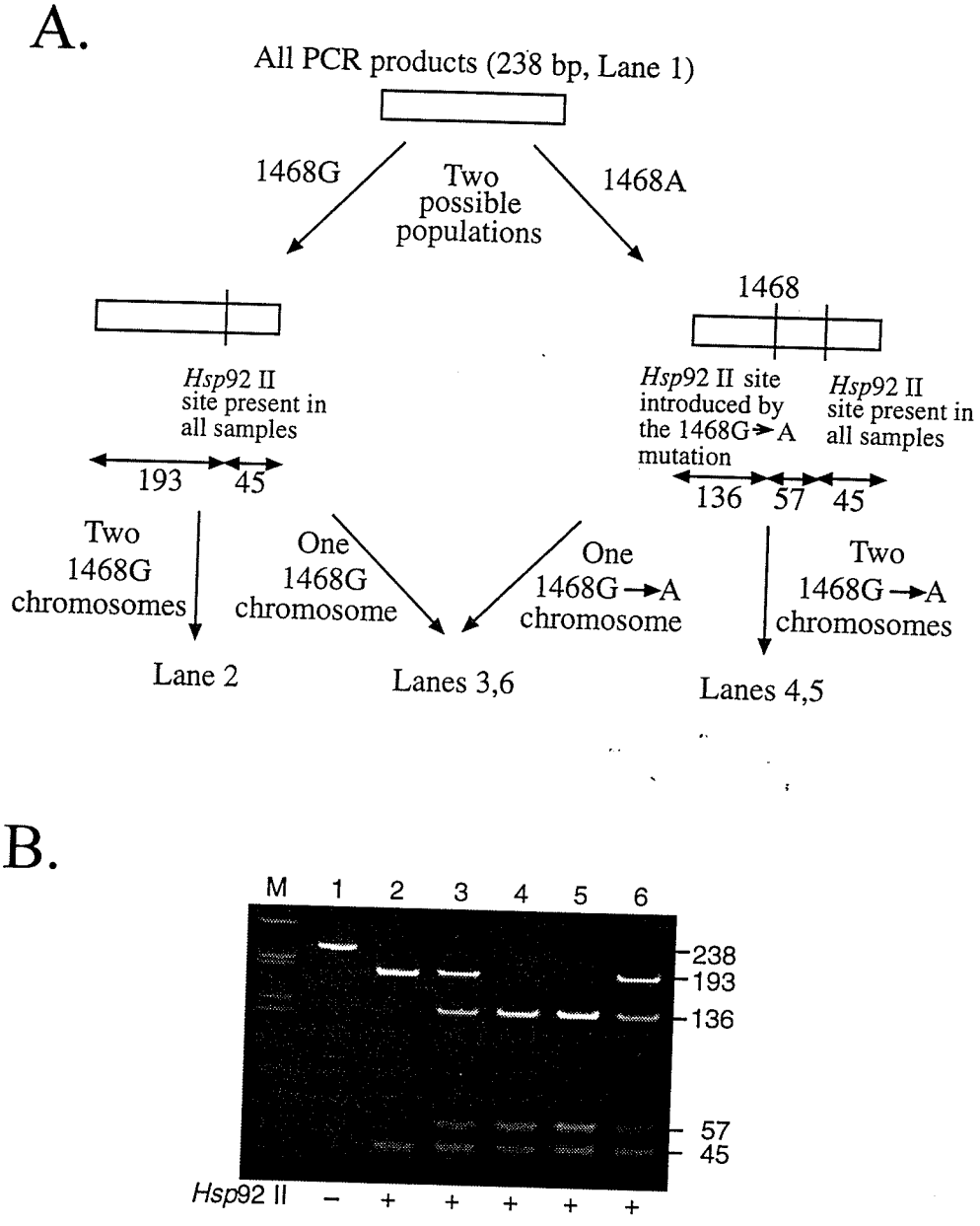



Figure 10. Detection of the alleles present at position 1468 in human *PHGDH*. A 238 bp region of human genomic DNA was PCR amplified, digested with *Hsp*92 II and separated by gel electrophoresis. A. Schematic diagram illustrating the detection of two alleles per sample of genomic DNA. B. Photograph of a polyacrylamide gel with an undigested (lane 1) and digested (lane 2) sample from a control individual homozygous for the 1468G allele, and digested samples from individuals heterozygous (lanes, 3,6) and homozygous for the 1468A allele (lanes 4,5).

allele, the PCR products will be cut twice, once at position 193 and once at nucleotide 1468 (position 136). This digestion pattern produces three bands: a 45 bp band common to the individuals without the mutation, and 136 and 57 bp bands originating from the former 193 bp fragment (lanes 4, 5). An individual heterozygous for the allele will have PCR products from the wild-type (1468G) chromosome and from the chromosome with a 1468A allele, giving four bands: 193, 136, 57, and 45 bp (lanes 3,6).

It was in this manner that the alleles of over 800 individuals of Jewish and non-Jewish descent were monitored for the V490M mutation. The results are shown in Table 1. One heterozygous individual of Jewish descent was identified. These results suggest that although the allele is present outside of the proband family, it is not present at a much higher frequency in the Ashkenazi Jewish community. This confirms previous reports indicating that this mutation is not a common polymorphism. Determining the frequency with any degree of accuracy is difficult with only one positive individual, but we can estimate the 95% confidence limits for the frequency of heterozygosity in this population to be between 1/70 and 1/16,000 (3). The large range represents the fact that there was only one positive sample out of 400 and would be expected to narrow with the determination of additional positive samples.

Table 1.
Results of investigation into the frequency of the 1468A allele in Ashkenazi Jewish and non-Jewish individuals.

| Ethnic Group | Individuals | Chromosomes | Mutant Alleles |
|---|---|---|---|
| Ashkenazi Jewish | 401 * | 802 | 1 |
| Non-Jewish | 409 *+ | 818 | 0 |

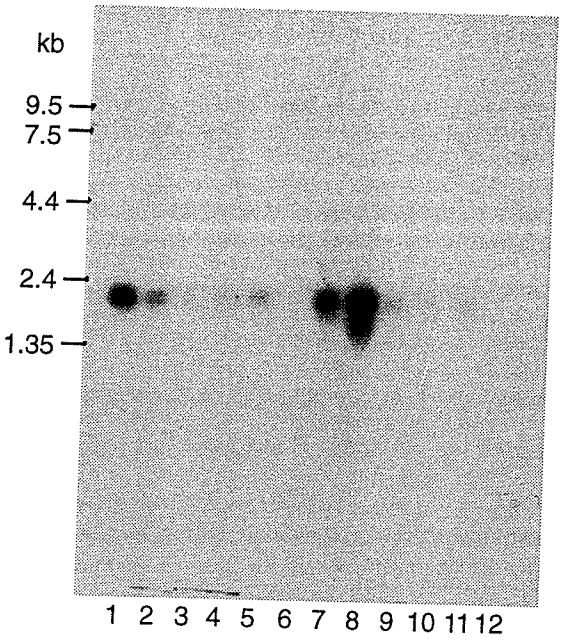
*Assays on 101 Jewish samples and 109 non-Jewish samples were performed by A. Gandhi; no mutant alleles were observed.

+Assays on 50 non-Jewish samples were performed by E. Slominski and S. Pind; no mutant alleles were observed.

## 3.2 Characterization of the presence of PHGDH and SHMT isozymes in human tissues

It was previously determined that PHGDH enzyme activity exists in many animal tissues (29). Reports were published during the course of these studies that revealed the expression pattern of the mRNA encoding PHGDH in human tissues (2;45). High enzyme activity and expression of mRNA encoding PHGDH was consistently observed in liver, kidney, and brain (2;29;45). Prior to my joining the Pind laboratory, a probe containing 595 bp of *PHGDH* cDNA sequence was hybridized to a Northern blot containing mRNA from twelve different tissues (Clontech). A band of approximately 2.2 kb was visible, with the highest intensities seen in liver, kidney, and brain (see Figure 11-A). A second, slightly smaller band was also observed in human liver. To extend these studies and to focus on the expression of the mRNA encoding PHGDH in the human brain, an eight-tissue brain blot (Multiple Tissue Northern Brain Blot II, Clontech) and a 75-tissue dot blot (Multiple Tissue Expression array II, Clontech) were hybridized with a 0.3 kb radiolabelled RNA probe of PHGDH. The same blots were stripped and reused with probes of the cytosolic and mitochondrial *SHMT* genes to analyze their expression and explore whether they might be alternative sources of serine synthesis in human tissues.

A.




B.

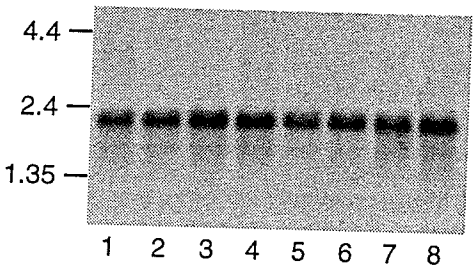


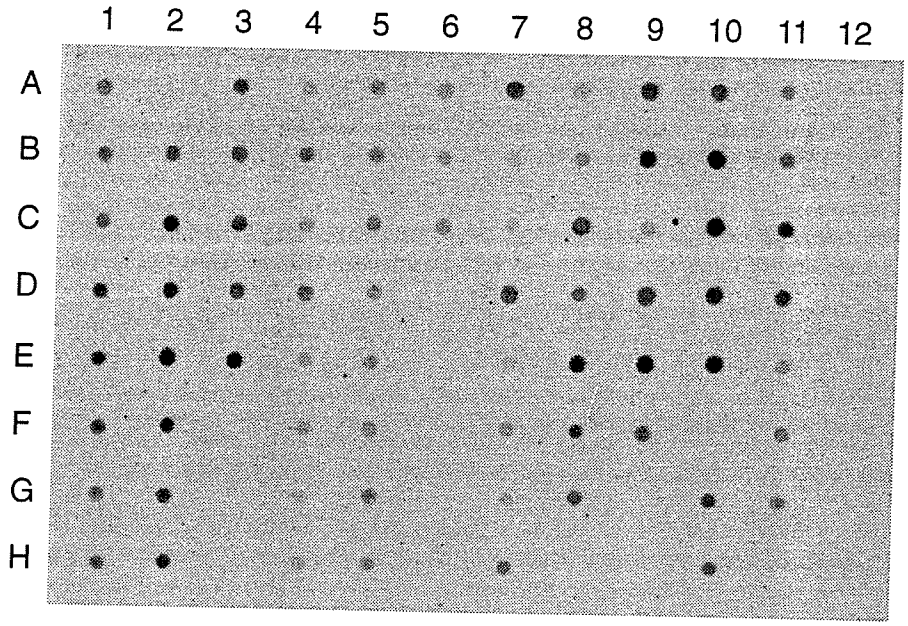

Figure 11. The expression of the mRNA encoding PHGDH.
A. A twelve tissue blot hybridized with a 595 bp radiolabelled cDNA probe. Lanes 1-12: brain, heart, skeletal muscle, colon, thymus, spleen, kidney, liver, small intestine, placenta, lung, peripheral blood leukocytes.
B. A radiolabelled RNA probe was hybridized to a Multiple Tissue brain blot and analyzed using a phosphorimager. Lanes 1-8: cerebellum, cerebral cortex, medulla, spinal cord, occipital pole, frontal lobe, temporal lobe, putamen.

### 3.2.1 PHGDH mRNA expression in human tissues

The distribution of the mRNA encoding PHGDH in central nervous system tissues was explored using a 298-bp RNA probe and the eight tissue brain blot. An ~2.2 kb band was visible in all lanes of the blot, as shown in Figure 11-B. This band, which corresponds to the full-length PHGDH transcript, was observed in all regions of the brain, unlike the tissue-specificity seen in the 12-tissue Northern blot (Figure 11-A). The bands were not equal in intensity, as detected by eye and phosphorimager analysis, indicating that some tissue-specific differences may exist within the different regions of the brain. However, the principal conclusion taken from this experiment was that all regions of the human brain examined on this blot expressed significant and quite similar levels of the mRNA encoding PHGDH.

The same RNA probe and hybridization conditions were used with the Clontech Human MTE Array II. An image of the blot is shown in Figure 12-A and a key to identify the samples present as well as a quantitative analysis of their intensity is shown in Figure 12-B. The value in parentheses below each tissue name in Figure 12-B was the relative intensity of that spot (as measured using the phosphorimager program QuantityOne) normalized to the intensity of the human kidney spot (Figure 12, row C, column 7), which was assigned a value of 1.0. The RNA from seventy-five human tissues was represented on the blot, along with the RNA and DNA from eight controls (column 12 in Figure 12). Seven of the human tissues represented were fetal tissues (16-36 weeks

A.




B.

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | whole brain (0.35) | | substantia nigra (0.60) | heart (0.09) | esophagus (0.17) | colon, transverse (0.18) | kidney (1.00) | lung (0.10) | liver (1.02) | HL-60 cells (0.81) | fetal brain (0.20) | yeast total RNA (0) |
| B | cerebral cortex (0.44) | cerebellum (right) (0.48) | nucleus accumbens (0.49) | aorta (0.35) | stomach (0.26) | colon, descending (0.13) | skeletal muscle (0.06) | placenta (0.17) | pancreas (2.07) | HeLa S3 cells (3.64) | fetal heart (0.35) | yeast tRNA (0) |
| C | frontal lobe (0.47) | corpus callosum (3.18) | thalamus (0.81) | atrium (left) (0.13) | duodenum (0.33) | rectum (0.24) | spleen (0.05) | bladder (1.10) | adrenal gland (0.11) | K-562 cells (3.94) | fetal kidney (1.52) | *E. coli* tRNA (0) |
| D | parietal lobe (1.23) | amygdala (1.64) | pituitary gland (0.60) | atrium (right) (0.46) | jejunum (0.21) | | thymus (0.99) | uterus (0.41) | thyroid gland (1.10) | MOLT-4 cells (2.02) | fetal liver (1.16) | *E.coli* DNA (0) |
| E | occipital lobe (1.10) | caudate nucleus (2.42) | spinal cord (2.61) | ventricle (left) (0.09) | ileum (0.19) | | leukocyte (c) (0.07) | prostate (1.80) | salivary gland (2.76) | Raji(d) cells (2.58) | fetal spleen (0.15) | poly r(A) (0) |
| F | temporal lobe (0.87) | hippo-campus (1.18) | | ventricle (right) (0.13) | ileocecum (0.18) | | lymph node (0.13) | testis (0.74) | mammary gland (0.49) | Daudi(d) cells (0.01) | fetal thymus (0.26) | human Cot-1 DNA (0) |
| G | p.g.(a) (0.42) | medulla oblongata (1.07) | | I.V.S.(b) (0.07) | appendix (0.38) | | bone marrow (0.10) | ovary (0.51) | | SW480(e) cells (1.03) | fetal lung (0.31) | human DNA 100 ng (0) |
| H | pons (0.55) | putamen (1.14) | | apex of the heart (0.09) | colon, ascending (0.19) | | trachea (0.42) | | | A549(f) cells (0.59) | | human DNA 500 ng (0) |

Figure 12. The expression profile of human *PHGDH*. A radiolabelled RNA probe (nucleotides 446-744) was hybridized to Clontech's Multiple Tissue Array II.
A.The blot's image as captured by the phosphorimager.
B. Identification of the samples present on the blot, along with the values derived from the QuantityOne analysis of the image in A, with the value for human adult kidney arbitrarily set to 1.0 for comparison. Abbreviations used are: (a)=paracentral gyrus of the cerebral cortex, (b)=interventricular septum, (c)=peripheral blood leukocyte, (d)=Burkitt's lymphoma, (e)=colorectal adenocarcinoma, (f)=lung carcinoma.

gestation) and eight were derived from cancer cell lines. The Multiple Tissue Array blot II was loaded with polyA+ RNA, and normalized to eight housekeeping genes: β-actin, glyceraldehyde-3-phosphate dehydrogenase, ubiquitin, phopholipase A2, α-tubulin, 23-kilodalton highly basic protein, ribosomal protein S9, and hypoxanthine guanine phosphoribosyl transferase. The manufacturer used eight controls that belong to different functional classes and showed minimal variation between tissues. They claim that this method results in a consistent signal from any gene that displays minimal tissue-to-tissue variation in its expression.

PHGDH expression was highest in the cancer cell lines, with the exception of Daudi cells (Figure 12, row F, column 10) where no signal was detected. The levels of PHGDH mRNA in cancer cell lines varied: K-562 cells (chronic myelogenous leukemia), HeLa S3 cells (epitheloid cervix carcinoma), Raji cells (Burkitt's lymphoma), and MOLT-4 cells (lymphoblastic leukemia) had relatively high levels; SW480 cells (colorectal adenocarcinoma), HL-60 cells (acute promyelocytic leukemia), and A549 cells (lung carcinoma) had medium levels, and Daudi cells (Burkitt's lymphoma) had an undetectable level.

Central nervous system tissues are displayed in columns 1-3. The highest signals were observed in corpus callosum, spinal cord, and caudate nucleus. The areas with lower values included: cerebral cortex, frontal lobe, paracentral gyrus, pons, and right cerebellum.

The cardiac tissues (column 4) and gastrointestinal tract (columns 5, 6) displayed relatively low levels of PHGDH expression. The organs and glands varied in their relative levels of intensity detected (columns 7-9). Those with high intensities included salivary gland, pancreas, prostate, bladder, thyroid gland, liver, kidney, and thymus. Trachea, uterus, mammary gland, ovary, and testis had medium levels. The tissues with low values were spleen, skeletal muscle, peripheral blood leukocytes, lung, bone marrow, adrenal gland, and lymph node.

Relatively high levels of the mRNA encoding PHGDH were observed in fetal kidney and liver (column 11). Lower levels were apparent in the other fetal tissues, including brain, heart, spleen, thymus, and lung. These levels were relatively higher than those of their respective adult tissues, with the exception of fetal thymus and brain.

### 3.2.2 Characterization of the expression of the mRNA encoding cSHMT and mSHMT enzymes

The expression profile of the cytosolic SHMT isoenzyme (cSHMT), is shown in Figure 13. High levels of cSHMT mRNA were present in kidney and liver. The rest of the tissues ranged between undetectable (pons, putamen, interventricular septum, apex of the heart, appendix, ascending colon, bone marrow, trachea, testis, ovary, mammary gland, fetal thymus, and fetal lung) to low levels of expression (thyroid gland and fetal liver).

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | whole brain (0.03) | | substantia nigra (0.03) | heart (0.03) | esophagus (0.06) | colon, transverse (0.04) | kidney (1.00) | lung (0.03) | liver (1.02) | HL-60 cells (0.04) | fetal brain (0.02) | yeast total RNA (0) |
| B | cerebral cortex (0.03) | cerebellum (right) (0.06) | nucleus accumbens (0.03) | aorta (0.03) | stomach (0.03) | colon, descending (0.04) | skeletal muscle (0.06) | placenta (0.04) | pancreas (0.04) | HeLa S3 cells (0.06) | fetal heart (0.03) | yeast tRNA (0) |
| C | frontal lobe (0.03) | corpus callosum (0.04) | thalamus (0.03) | atrium (left) (0.03) | duodenum (0.06) | rectum (0.04) | spleen (0.06) | bladder (0.05) | adrenal gland (0.05) | K-562 cells (0.05) | fetal kidney (0.11) | *E. coli* tRNA (0) |
| D | parietal lobe (0.04) | amygdala (0.04) | pituitary gland (0.04) | atrium (right) (0.04) | jejunum (0.06) | | thymus (0.12) | uterus (0.04) | thyroid gland (0.24) | MOLT-4 cells (0.05) | fetal liver (0.26) | *E.coli* DNA (0) |
| E | occipital lobe (0.04) | caudate nucleus (0.04) | spinal cord (0.05) | ventricle (left) (0.03) | ileum (0.05) | | leukocyte (c) (0.03) | prostate (0.04) | salivary gland (0.05) | Raji(d) cells (0.01) | fetal spleen (0.01) | poly r(A) (0) |
| F | temporal lobe (0.03) | hippo-campus (0.03) | | ventricle (right) (0.02) | ileocecum (0.03) | | lymph node (0.01) | testis (0.003) | mammary gland (0.00) | Daudi(d) cells (0.00) | fetal thymus (0.00) | human Cot-1 DNA (0) |
| G | p.g.(a) (0.01) | medulla oblongata (0.004) | | I.V.S.(b) (0.00) | appendix (0.00) | | bone marrow (0.00) | ovary (0.00) | | SW480(e) cells (0.00) | fetal lung (0.00) | human DNA 100 ng (0) |
| H | pons (0.00) | putamen (0.00) | | apex of the heart (0.00) | colon, ascending (0.00) | | trachea (0.00) | | | A549(f) cells (0.00) | | human DNA 500 ng (0) |

Figure 13. The expression profile of human cytoplasmic SHMT (*SHMT1*). A radiolabelled RNA probe of the 3' untranslated region was hybridized to Clontech's Multiple Tissue Array II. The results shown are the values derived from the QuantityOne analysis of the image, with the value for human adult kidney arbitrarily set to 1.0 for quantitative comparisons. Abbreviations used are: (a)=paracentral gyrus of the cerebral cortex, (b)=interventricular septum, (c)=peripheral blood leukocyte, (d)=Burkitt's lymphoma, (e)=colorectal adenocarcinoma, (f)=lung carcinoma.

The levels of mRNA seen with the mitochondrial SHMT isoenzyme (mSHMT), (see Figure 14) revealed a broader expression in all tissues, but with significant variation between tissues. When normalized to kidney, liver and lymph node had very high levels of expression. In the central nervous system tissues (columns 1-3) the relative level of mRNA detected varied from low (substantia nigra, and nucleus accumbens) to medium (corpus callosum, putamen, medulla oblongata, spinal cord, caudate nucleus, and right cerebellum). The highest value in the central nervous system was localized to the pituitary gland.

Cardiac tissues (column 4) had a medium level of intensity of mSHMT, higher than many of the tissues in the central nervous system. In the gastrointestinal tract (columns 5-6), the values were increased slightly from those of the cardiac tissues. The tissues of the reproductive organs (ovary, testis, prostate, uterus, and placenta, in column 8, rows B, D-G) had medium levels of mRNA expression.

Immune system tissues (column 7, rows C-G: spleen, thymus, peripheral blood leukocytes, lymph node, and bone marrow) had fairly high levels of intensity, as did the digestive organs/glands, such as liver and pancreas.

The cancer cell lines (column 10) expressed higher levels of the mRNA encoding mSHMT than did normal kidney, except for MOLT-4 leukemia cells and lung carcinoma A549 cells. Compared to the tissues they originated from, the MOLT-4 cells have approximately one-third the level of mRNA expression of peripheral blood leukocytes,

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | whole brain (0.13) | | substantia nigra (0.13) | heart (0.34) | esophagus (0.27) | colon, transverse (0.49) | kidney (1.00) | lung (0.42) | liver (5.82) | HL-60 cells (1.18) | fetal brain (0.11) | yeast total RNA (0) |
| B | cerebral cortex (0.14) | cerebellum (right) (0.53) | nucleus accumbens (0.13) | aorta (0.27) | stomach (0.52) | colon, descending (1.02) | skeletal muscle (0.50) | placenta (0.69) | pancreas (1.21) | HeLa S3 cells (3.21) | fetal heart (0.72) | yeast tRNA (0) |
| C | frontal lobe (0.14) | corpus callosum (0.31) | thalamus (0.19) | atrium (left) (0.32) | duodenum (0.76) | rectum (0.43) | spleen (0.53) | bladder (0.39) | adrenal gland (0.56) | K-562 cells (2.44) | fetal kidney (0.66) | *E. coli* tRNA (0) |
| D | parietal lobe (0.23) | amygdala (0.26) | pituitary gland (0.80) | atrium (right) (0.46) | jejunum (0.58) | | thymus (0.98) | uterus (0.61) | thyroid gland (0.58) | MOLT-4 cells (1.25) | fetal liver (1.41) | *E.coli* DNA (0) |
| E | occipital lobe (0.24) | caudate nucleus (0.46) | spinal cord (0.31) | ventricle (left) (0.51) | ileum (0.37) | | leukocyte (c) (0.39) | prostate (0.41) | salivary gland (0.57) | Raji(d) cells (1.65) | fetal spleen (0.88) | poly r(A) (0) |
| F | temporal lobe (0.25) | hippo-campus (0.23) | | ventricle (right) (0.42) | ileocecum (0.48) | | lymph node (3.46) | testis (0.13) | mammary gland (0.82) | Daudi(d) cells (1.72) | fetal thymus (0.43) | human Cot-1 DNA (0) |
| G | p.g.(a) (0.15) | medulla oblongata (0.24) | | I.V.S.(b) (0.31) | appendix (0.46) | | bone marrow (0.73) | ovary (0.59) | | SW480(e) cells (1.84) | fetal lung (0.74) | human DNA 100 ng (0) |
| H | pons (0.20) | putamen (0.18) | | apex of the heart (0.32) | colon, ascending (0.54) | | trachea (0.42) | | | A549(f) cells (0.51) | | human DNA 500 ng (0) |

Figure 14. The expression profile of human mitochondrial SHMT (*SHMT2*). A radiolabelled RNA probe of the 3' untranslated region was hybridized to Clontech's Multiple Tissue Array II. The results shown are the values derived from the QuantityOne analysis of the image, with the value for human adult kidney arbitrarily set to 1.0 for quantitative comparisons. Abbreviations used are: (a)=paracentral gyrus of the cerebral cortex, (b)=interventricular septum, (c)=peripheral blood leukocyte, (d)=Burkitt's lymphoma, (e)=colorectal adenocarcinoma, (f)=lung carcinoma.

and the lung cancer cells (A549) expressed slightly more mSHMT mRNA than regular adult lung.

The fetal tissues, with the exception of fetal brain (low expression), had medium levels of mRNA detected in fetal thymus and fetal spleen, and high levels in fetal liver. The values for fetal kidney, liver, brain, and thymus were not higher than those of the corresponding adult tissues, but the samples of lung, spleen, and heart had higher levels of mSHMT mRNA expression in the fetus than in the adult.

### 3.2.3 Localization of human PHGDH in brain tissue using immunohistochemistry

The mRNA encoding PHGDH was detected at significant levels in all tissues of the human central nervous system (Figures 11 and 12). This supports the idea that the neurological system requires endogenous serine for its growth, differentiation, and functional capabilities. While a region of the brain may express PHGDH and produce serine, this does not mean that all cell types within that region possess this ability. This hypothesis arises from reports that certain neurons do not express PHGDH and require exogenous serine (49;51), and the recent immunohistochemical analysis of PHGDH expression in rodent brains (52;53). Cellular localization of PHGDH in human brain was pursued using immunohistochemistry with mouse monoclonal anti-human PHGDH antibodies and sections of human brain tissue. Thirty-six immunoreactive clones were screened for their ability to recognize the PHGDH protein in fixed and paraffin-

embedded brain tissues. One clone, 1C2, was chosen for large-scale isolation and antibody purification.

Immunoreactivity of the 1C2 primary antibody (at a dilution of 1:2500) was detected by a standard biotin-avidin-horseradish peroxidase detection system with diaminobenzidine (DAB) as the chromogenic substrate. This results in a brown coloured precipitate at the site of antibody binding. The sections were counterstained with hematoxylin (blue colour) which stains negatively charged compounds such as the nuclei of cells. A negative control was performed by incubating one slide with blocking solution (without primary antibody) and then the addition of the secondary antibody and avidin-biotin complex. This demonstrated the level of non-specific binding of the secondary antibody, the avidin-biotin complex, or the non-specific precipitation of the DAB substrate.

Digital photomicrographs of human cerebellum are shown in Figure 15 at a magnification of 40X. Bergmann glia were immunoreactive for the PHGDH protein (Figure 15-A). The cellular extensions were not clearly stained, though one is indicated with an arrow. The cell bodies of the Bergmann glia are located at the interface between the molecular layer (right side of photograph) and the granular layer (left side) of the cerebellum (52), and distinct DAB staining was detected at this interface. This suggests that the cytoplasm of the Bergmann glia expressed the PHGDH protein. Further studies are required to confirm this result, either with a lower dilution of the 1C2 antibody or

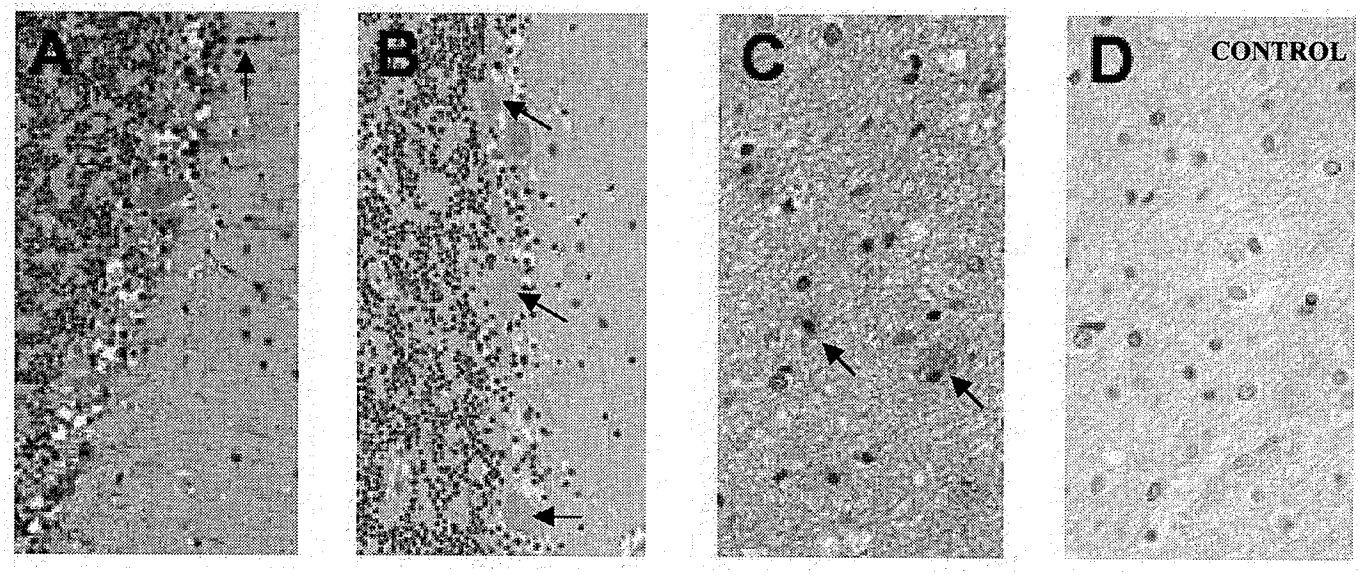



Figure 15. Immunolocalization of PHGDH in paraffin-embedded sections of human cerebellum. Sections were incubated with anti-PHGDH 1C2 antibodies (A-C), or in blocking solution (negative control, D). Immunoreactivity was detected using a biotin-avidin-horseradish peroxidase system with the chromogenic substrate diaminobenzidine, and counterstained with hematoxylin. A. Bergmann glia (arrow) are detected in this region. B. Granule cells (small blue cells, left half of photo) and Purkinje neurons (larger cells, arrows) are not detected by 1C2 antibodies. C. Astrocytes (arrows) in the white matter of the cerebellum are immunoreactive with the 1C2 antibody. D. A negative control of cerebellar white matter.

co-immunofluorescence of this antibody with one specific to glial cells, such as glial fibrillary acidic protein (GFAP). The 1C2 antibody did not detect granule cells (left half of Figure 15-B) or Purkinje neurons (as shown by arrows in Figure 15-B). The astrocytes of the cerebellar white matter exhibited PHGDH immunoreactivity (arrows, Figure 15-C). A negative control (Figure 15-D) of cerebellar white matter (same region as Figure 15-C) showed no diaminobenzidine staining.

Figure 16 shows that both astrocytes and oligodendrocytes were specifically recognized by the 1C2 antibody at a 1:2500 dilution (40X magnification). Oligodendrocytes can be discerned in these sections by a distinctive clearing around the nucleus (perinuclear 'halo'), which is an artifact of fixation (personal communication, Dr. D. Eisenstat). Oligodendrocytes from human striatum (white matter) are indicated in Figure 16-A by arrows, while asterisks highlight several astrocytes. Some diaminobenzidine precipitates were visible in the clearings around the oligodendrocytes in the negative control (Figure 16-B), however, it was the region surrounding the nucleus and not the clearing itself which was positive for the antigen recognized by the 1C2 antibody in Figure 16-A.

The human hippocampus also had PHGDH-positive astrocytes (Figure 17-A). The cells of the CA1 region of the hippocampus were negative (cells surrounding asterisk, 40X magnification). No non-specific staining was evident in the negative control (Figure 17-B).

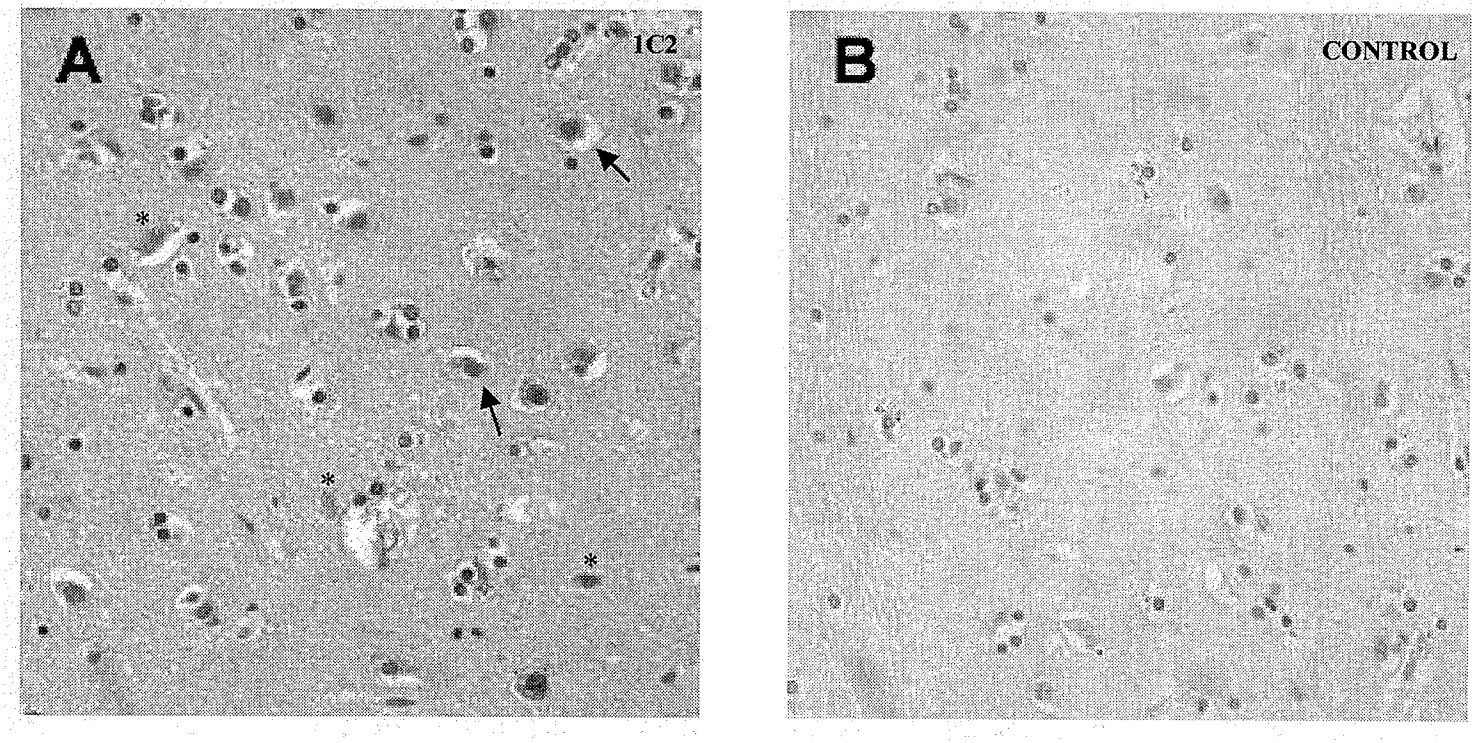



Figure 16. Immunostaining of oligodendrocytes (arrows) and astrocytes (asterisks) in a section of human striatum, with the 1C2 antibody (A), or without primary antibody (B).

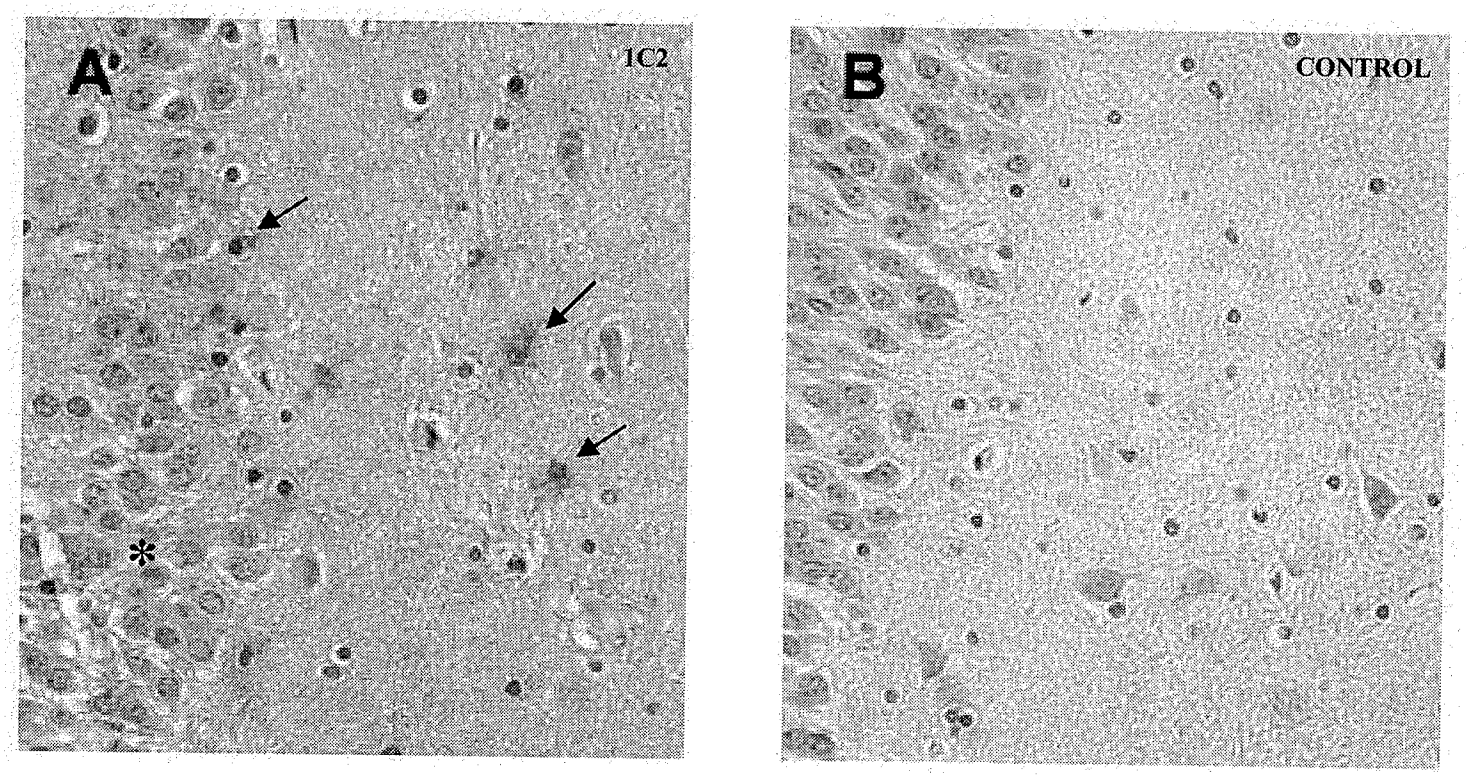



Figure 17. Localization of 1C2 antibody-immunoreactivity to the astrocytes (A, shown with arrows) but not neurons of the CA1 region (asterisk) in human hippocampus. No diaminobenzidine staining was present when primary antibody was omitted (B).

Similar to the striatum, the corpus callosum had astrocytes (shown by arrows in Figure 18-A at 40X magnification) and oligodendrocytes (arrowheads) that were immunoreactive with the PHGDH antibody. The negative control (Figure 18-B) had no diaminobenzidine precipitates.

Sections of frontal cortex revealed that oligodendrocytes were recognized by the PHGDH 1C2 antibody, but pyramidal neurons were not (preliminary data, not shown). Small astrocytes in the white matter of the midbrain (substantia nigra) were immunoreactive and it appeared that the oligodendrocytes might also be immunostained with the 1C2 antibody. Certain neurons in the substantia nigra region might be considered positive, as the neurons were dark brown or black in colour due to the natural production of melanin as a byproduct of the synthesis of dopamine (37). This colouring was easily distinguished because the melanin was darker and appeared more granular in nature than that of the diaminobenzidine staining (not shown). Furthermore, this pigment was detected in negative controls in which no primary antibody was used.

A second antibody, 8B6, has been partially characterized and initial experiments with different titres revealed a primarily astrocytic staining pattern. Further experiments are needed to determine the correct dilution and characterize the cell types recognized by this antibody.

These results indicate that the 1C2 antibody appears to successfully bind antigen present in certain cells types of human brain tissue, especially within the cytoplasm of

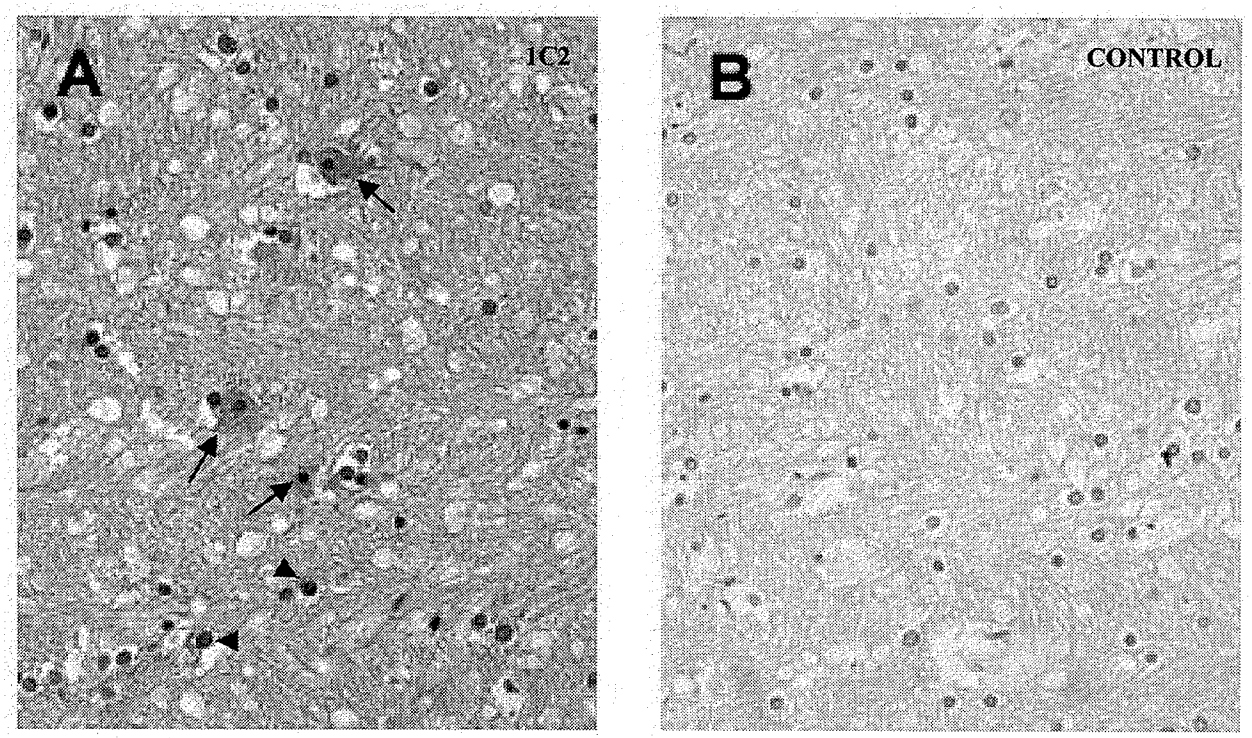



Figure 18. Immunostaining of cells within the human corpus callosum with PHGDH 1C2 antibody.
A. Astrocytes (arrows) and oligodendrocytes (arrowheads) were detected with the 1C2 antibody.
B. Negative control.

astrocytes and oligodendrocytes. This will be a useful reagent for characterizing the localization of PHGDH protein.

## 3.3 Elucidation of the structure of the human *PHGDH* gene

The coding DNA (cDNA) and amino acid sequences of human PHGDH had been deduced prior to my start in the laboratory, but the complete genomic sequence, including the number and size of introns within the *PHGDH* gene had not been reported. A PCR-based cloning strategy to obtain *PHGDH* fragments in plasmids for dideoxy-sequencing was employed. This method provided an estimate of intron size by agarose gel electrophoresis of the PCR products and comparison with a DNA ladder. The size of the intron(s) contained in those fragments is the difference between the total size and the number of nucleotides known to be located between the primers. A single sequencing reaction from both ends of each cloned PCR product provided the sequence of the intron-exon boundaries and the localization of the splice sites within that section of *PHGDH*. All intron-exon boundaries were sequenced at least three times to ensure accuracy, and the identification of the introns was not performed until every section of the *PHGDH* gene had been characterized. It was not our intention to sequence the introns in their entirety, but to provide the sequences surrounding the intron-exon boundaries.

Three types of results were obtained following the amplification and sequencing of the intron products. The first, and simplest, result occurred for introns 1 and 11 (see

Figure 19). With these introns, the initial primers amplified PCR products (Figure 19-A) in which all of the expected cDNA sequence was obtained by sequencing from both ends of the cloned PCR products, indicating that only one intron existed in those fragments (Figure 19-B).

The next category of result occurred for introns 5 through 10 (see Figure 20). In these instances, the initial primers yielded products (Figure 20-A) that upon sequencing did not yield the entire cDNA sequence known to reside between the primers (Figure 20-B). Single sequencing reactions from both ends revealed intronic sequence before all of the expected cDNA sequence was accounted for. This indicated that two or more introns resided within those fragments, and further PCR amplification and cloning was required. A new primer set was designed within the region of cDNA that was 'missing' and the new sets of primers were used to amplify two smaller fragments (Figure 20-C). Sequencing of these new products revealed only one intron within each fragment.

The third type of result occurred during amplification of introns 2, 3, and 4 (see Figure 21). The initial primer pair amplified a product containing all three of these introns (see Figure 21-A). Sequencing from one end of the clones revealed the beginning of intron 2 (Figure 21-B), but the other end revealed the cDNA sequence (exon), 99 bases of an intron (intron 4), 66 bp of cDNA sequence (exon 4), and more intronic sequence. Since the intron was so small, and we had sequenced its entire length, it was not

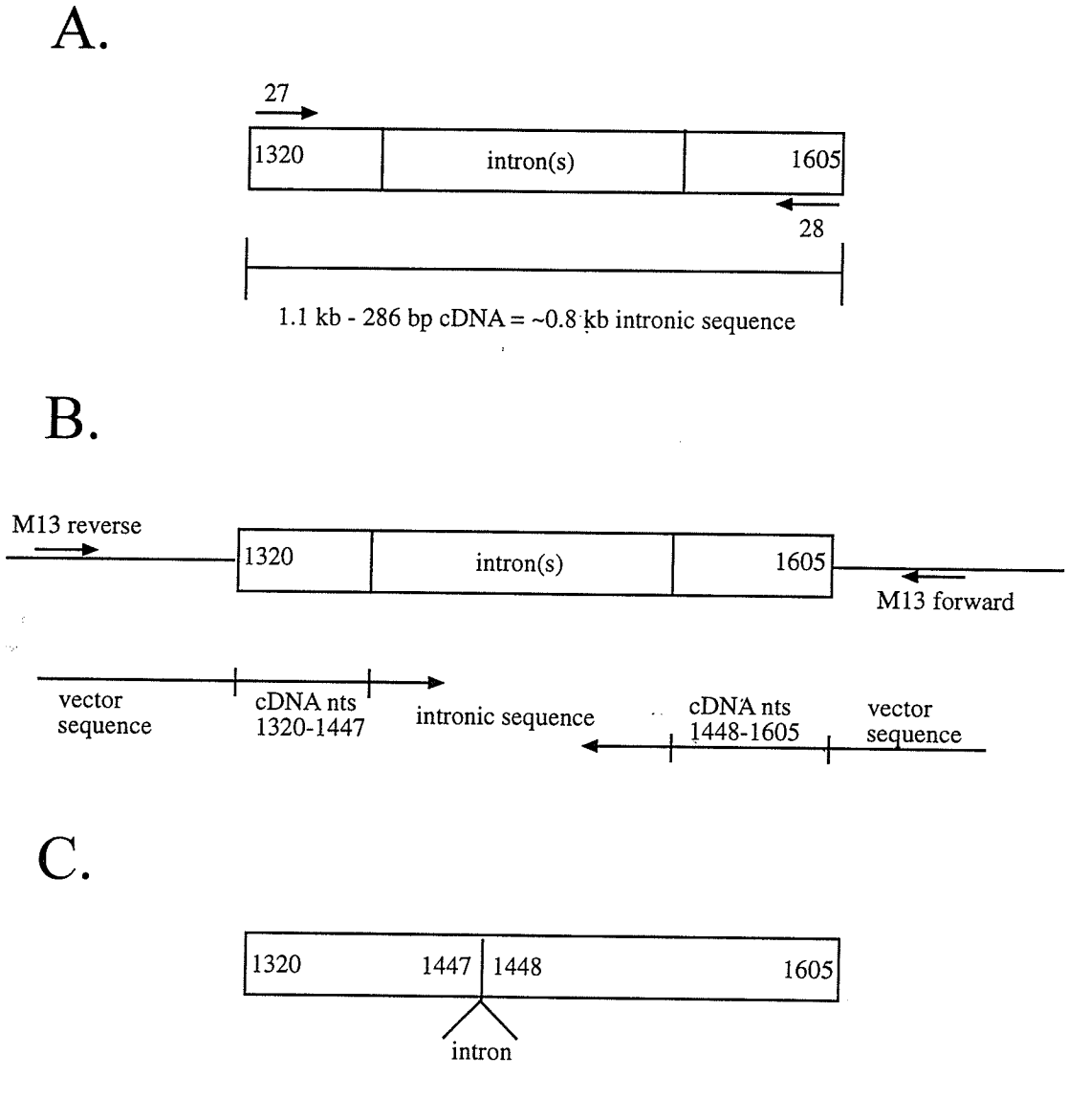



Figure 19. Strategy employed to determine the placement of one intron within a PCR product (intron 11 is illustrated). A. Initial primers were designed to amplify cDNA nucleotides 1320 to 1605, which produced a 1.1 kb fragment. B. Cloning into TopoXL vector and sequencing revealed products with vector, cDNA and intronic sequences. The intronic sequences were located between nucleotides 1447 and 1448. No cDNA sequence was missing, indicating there was only one intron between these primers. C. The placement of intron 11 within the fragment.

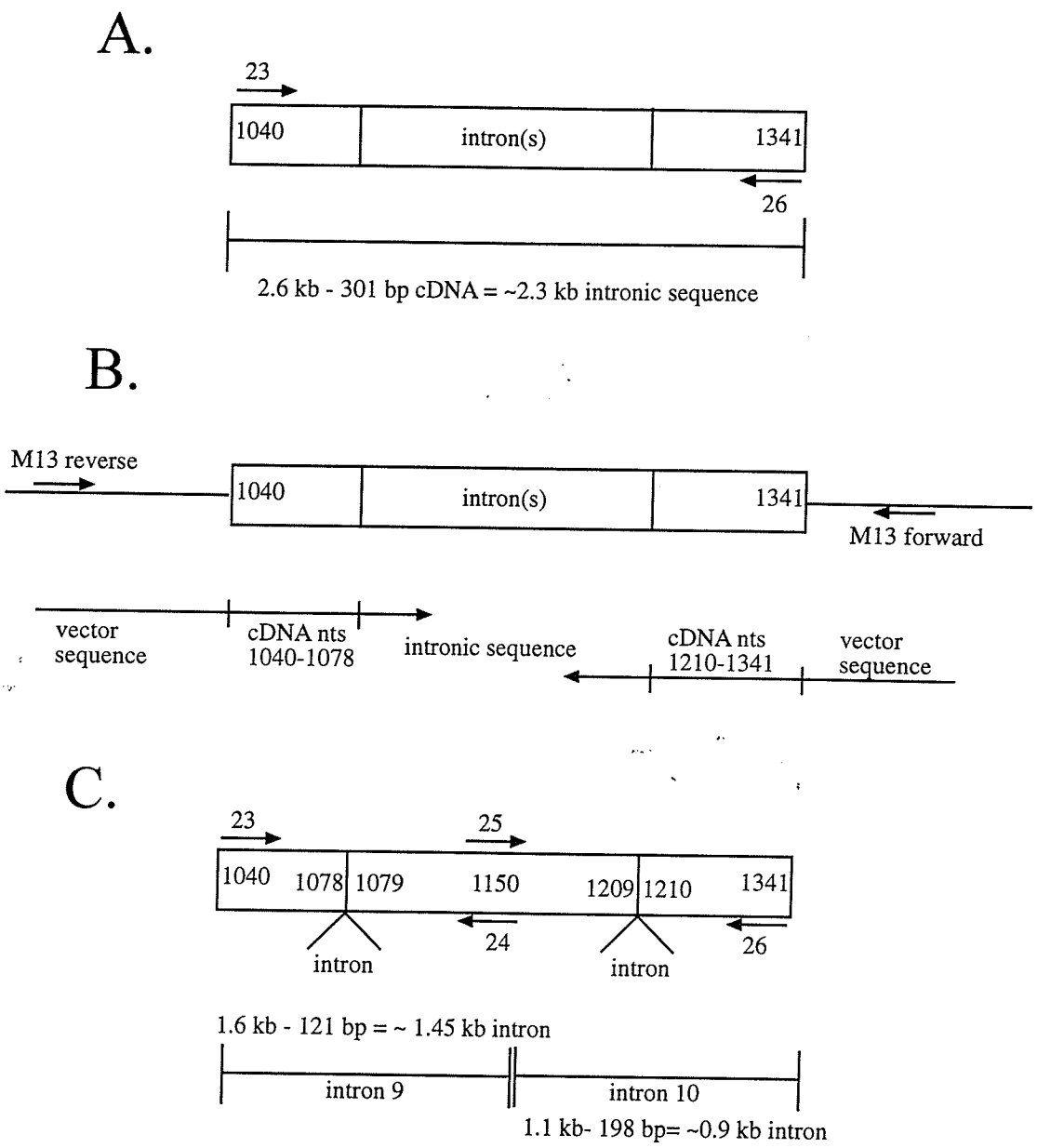



Figure 20. Strategy employed to determine the placement of two introns within a PCR product (Results from introns 9 and 10 illustrated). A. Initial primers were designed to amplify cDNA nucleotides 1040 to 1341, which resulted in a 2.6 kb fragment. B. The product was cloned into a TopoXL vector and sequenced, revealing vector, cDNA and intronic sequences. The introns were found to interrupt the cDNA sequence at occurred at nucleotides 1078 and 1210. The cDNA sequence between 1079-1209 was 'missing' from these initial sequencing reactions.
C. New primers (24 and 25) were designed with homology to the cDNA sequence between the introns. Amplification with 23 and 24 produced a fragment that when sequenced revealed a splice site between nucleotides 1078 and 1079 (intron 9). Amplification with primers 25 and 26 produced a fragment of 1.1 kb with a splice site between 1209-1210 (intron 10).

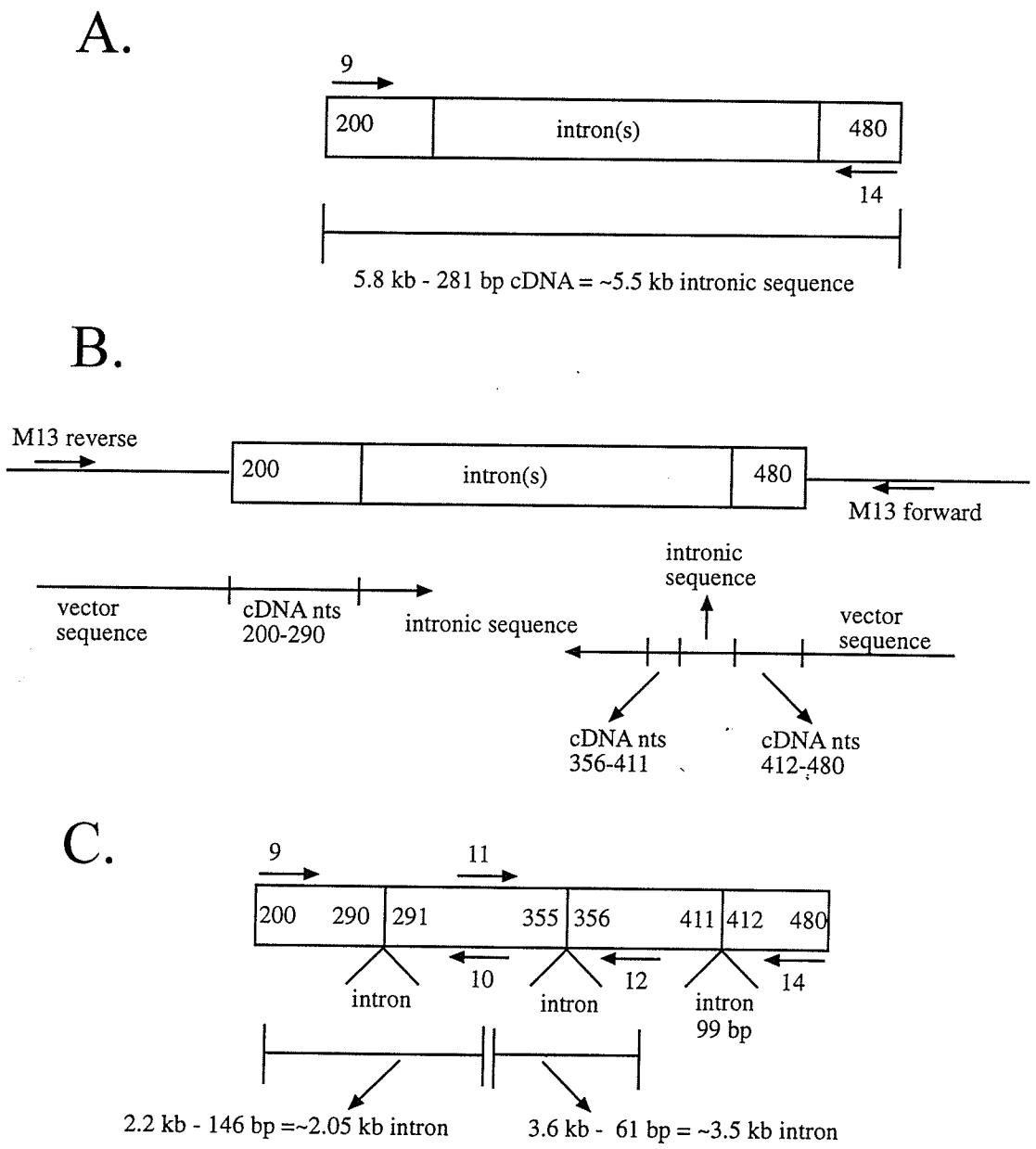



Figure 21. Strategy employed to determine the placement of three introns within a PCR product. A. Initial primers were designed to amplify cDNA nucleotides 200 to 480, which resulted in a 5.6 kb fragment. B. The product was cloned into a TopoXL vector, sequenced, revealing products with vector, cDNA and intronic sequences. One intron was located after nucleotide 290. The sequencing products from the opposite end revealed cDNA sequence 412-480, a 99 bp intron between nucleotides 411 and 412, an exon from 356-411, and another intron beginning upstream of nucleotide 356. C. New primers (10, 11, and 12) were designed with homology to the cDNA sequence between the introns. Amplification with primers 9 and 10 produced a fragment that revealed an intron between nucleotides 290 and 291. Similar amplification with primers 11 and 12 produced a fragment with an intron between nucleotides 355-356.

necessary to clone it separately to determine its size. The remaining region of 'missing' cDNA was characterized in a manner similar to introns 5 to 10, with new primers designed and PCR products amplified and sequenced (see Figure 21-C).

Starting from the 3' end of the gene, intron 11 was previously amplified and the junction sites were sequenced by E. Slominski and S. Pind (unpublished results). They determined the splice site to be between cDNA nucleotides 1447 and 1448 and the intron was ~800 base pairs in length. Confirmation of this finding was obtained by amplifying BAC DNA with primers 27 and 28 under the conditions of a 60°C annealing temperature and five minutes elongation per cycle. The primers were designed to amplify from nucleotides 1320 to 1605, or approximately 286 base pairs. The PCR product was approximately 1.1 kb in length, as shown in Figure 22, lane 7, and sequencing confirmed that the splice site was located between cDNA nucleotides 1447 and 1448. By subtracting the length of cDNA nucleotides known to reside in this PCR product, an intron length of approximately 800 base pairs was deduced.

The first attempt to clone and sequence intron 10 began with primers 23 and 26 and conditions of a 60°C annealing temperature and five minutes extension per cycle (Figure 20-A). This amplified the genomic DNA located between cDNA nucleotides 1040 and 1341 (302 bp of cDNA sequence). The 2.6 kb PCR product (Figure 22, lane 6) was cloned, and sequencing from one end with an M13 primer revealed the cDNA

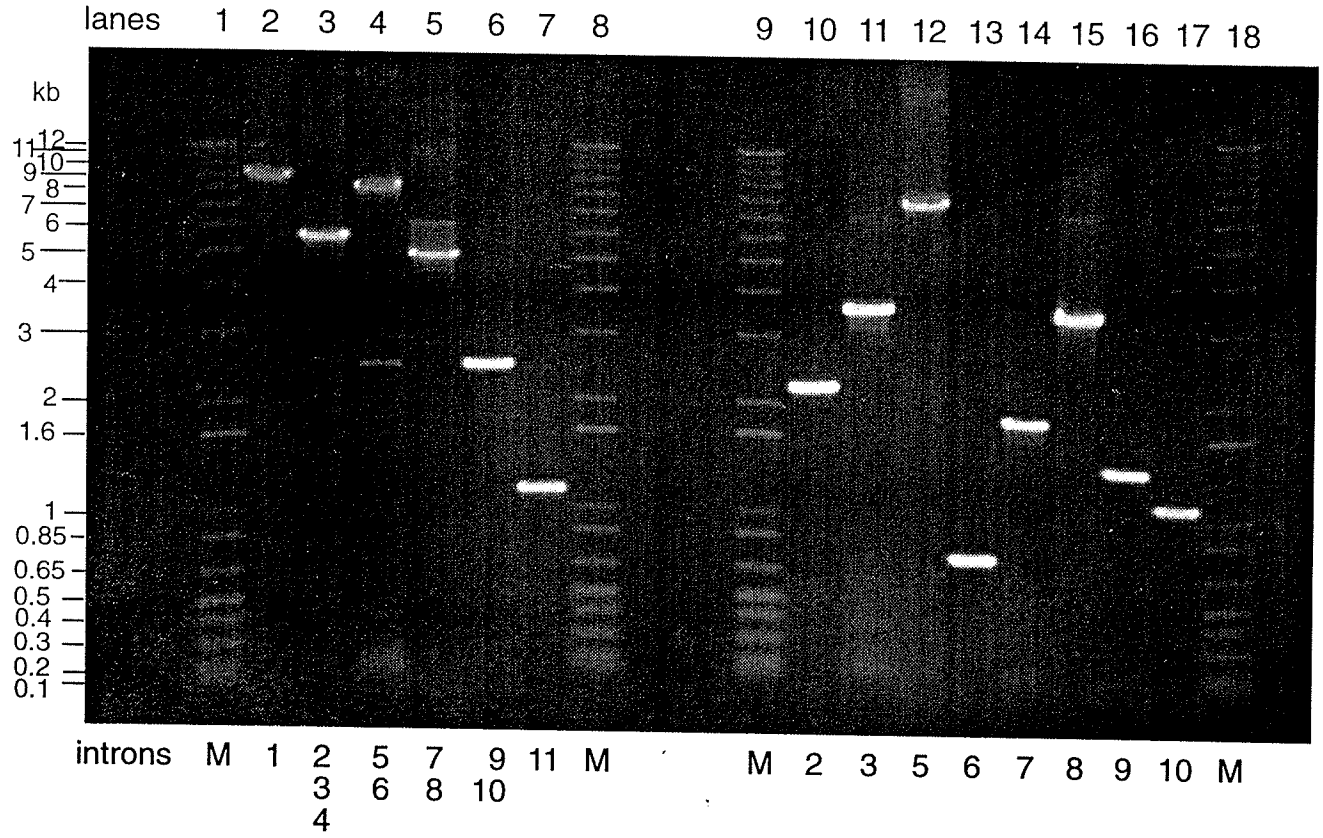



Figure 22. Summary agarose gel showing the sizes of the introns of the human *PHGDH* gene. Primers spaced along the length of the PHGDH cDNA were PCR amplified and the size of the resulting products was determined on an agarose gel. The localization of the intron-exon boundaries in each fragment was determined by sequencing. The first sets of primers amplified fragments containing 1-3 introns, see lanes 2-7. The introns represented within each fragment are listed below the lanes. New primers were designed to amplify single introns, and the new products are shown in lanes 10-17, with the introns listed below the lanes. Only one intron was not localized in this manner. Intron 4 was 99 bp and was sequenced in its entirety from the PCR fragment in lane 3. The 1kb Plus DNA ladder is shown in lanes 1, 8, 9, 18 (labelled M). The PCR fragments ranged in size from 0.7kb (lane 13) to approximately 9.5 kb (lane 2).

sequence from nucleotides 1040 to 1078. At this point, the sequence diverged from the cDNA sequence, indicating an intron between nucleotides 1078 and 1079. Sequencing from the opposite end of the cloned insert, with the other M13 primer, provided the cDNA sequence from position 1210 to1341, revealing an intron between 1209 and 1210 (Figure 18-B). The cDNA sequence from 1079 to 1209 lay between these two introns of unknown size. New primers (24 and 25) were designed to amplify the cDNA sequence located between nucleotides 1079 and 1209 (Figure 20-C).

Primers 25 and 26 amplified a 1.1 kb PCR product (Figure 22, lane 17) with annealing at 60°C and four minutes elongation per cycle. Sequencing confirmed the preliminary data and located intron 10 between cDNA nucleotides 1209 and 1210.

New PCR conditions with 58°C as the annealing temperature and three minutes elongation time per cycle were used with primers 23 and 24 to amplify a product of approximately 1.6 kb (Figure 22, lane 16), that when sequenced located intron 9 between cDNA nucleotides 1078 and 1079.

Introns located between cDNA nucleotides 760 and 1040 were amplified using primers 19 and 22. PCR with an annealing temperature of 60°C and an extension of eight minutes per cycle were required to amplify the 5.5 kb fragment (Figure 22, lane 5). Cloning and sequencing revealed one intron began after nucleotide 792 and another ended prior to nucleotide 946. This meant that cDNA nucleotides 793 to 945 were

contained within at least two introns. New primers 20 and 21 were designed and PCR was attempted with primer sets 19 & 20 and 21 & 22.

Intron 8 was found to lie between cDNA nucleotides 945 and 946 after sequencing the 2.4 kb PCR product (Figure 22, lane 15) obtained with primers 21 and 22. Conditions of a 60°C annealing temperature and a four minute elongation time per cycle were used to amplify this product.

Primers 19 and 20 amplified a 1.7 kb fragment (Figure 22, lane 14) of *PHGDH* when conditions of 60°C and four minutes per cycle were used. This PCR product, upon sequencing, localized intron 7 between cDNA nucleotides 792 and 793.

Primers 15 and 18 were used to amplify genomic DNA between cDNA nucleotides 480 to 760. PCR conditions of 57°C annealing and ten minute extensions were required to amplify a product of 8.5 kb (Figure 22, lane 4), that upon sequencing using the M13 forward primer revealed cDNA nucleotides 480 to 510, then intronic sequence. Sequencing from M13 reverse primer provided cDNA nucleotides 644 to 760, followed by intronic sequence. This meant that cDNA nucleotides 511 to 643 were located between two introns, so new primers 16 and 17 were created within this region.

Using primers 17 and 18, a PCR product of 0.6 kb (Figure 22, lane 13) was amplified with an annealing temperature of 57°C and four minutes extension per cycle. Cloning and sequencing revealed only one intron, intron 6, located between cDNA nucleotides 643 and 644.

Primers 15 and 16 were used to amplify the sequence between nucleotides 480 and 603 using a ten minute extension per cycle and an annealing temperature of 57°C. A PCR product of 8 kb (Figure 22, lane 12) was amplified and several attempts were made to clone this into the Topo® XL vector. The clones produced did not contain the proper insert, so the purified vector containing both introns 5 and 6 (cloned above using primers 15 and 18) was used to sequence the boundaries of intron 5. This was performed using the M13 forward primer to sequence from nucleotides 480 to 510, and primer 16 was used to sequence from nucleotide 511 to 603. Intron 5 was determined to be approximately 8 kb long and located between cDNA nucleotides 510 and 511.

The first attempt to determine the introns located between cDNA nucleotides 200 and 480 used primers 9 and 14, with an annealing temperature of 61°C, and an extension time of ten minutes per cycle. This amplified a product of 5.8 kb (Figure 22, lane 3) that was cloned into a Topo® XL vector. Sequencing with the M13 forward primer revealed cDNA sequence from nucleotides 200 to 289 and then intronic sequence. Sequencing from the M13 reverse primer revealed cDNA nucleotides 357 to 400, at which point the sequence diverged for 99 base pairs. Then cDNA nucleotides 290 to 356 were recovered before a second divergence. We could infer that the entire sequences of intron 4, then exon 3 were read prior to sequencing some of intron 3 (see Figure 21).

New primers were designed for cDNA nucleotides 318-340 (primer 11) and 379-359 (primer 12) to amplify intron 3. With PCR conditions of 57°C annealing temperature

and four minutes per cycle extension time, a PCR product of size 3.6 kb (Figure 22, lane 11) was obtained. Cloning and sequencing revealed that intron 3 lay between nucleotides 356 and 357.

Intron 2 was amplified using primers 9 and 10 (encompassing nucleotides 200-345), producing a 2.2 kb fragment of *PHGDH* (Figure 22, lane 10). The PCR conditions included an annealing temperature of 57°C and an elongation time of four minutes. Intron 2 was localized between the nucleotides 289 and 290.

The PCR product containing intron 1 was amplified with primers 7 and 8, using an annealing temperature of 57°C and an elongation time of 15 minutes. In addition, 5% dimethylsulfoxide was required in the PCR reaction in order to ensure the proper product. This PCR product was the largest, with a size of ~9.5 kb (Figure 22, lane 2) and intron 1 was located between cDNA nucleotides 138 and 139.

In summary, PCR amplification, cloning, and sequencing was used to determine that the human *PHGDH* gene has twelve exons and eleven introns. The sequences at the intron-exon boundaries and a summary of the location and sizes of the introns are provided in Table 2. The approximate size of the entire gene, from exon 1 to exon 12, is 33 kilobases. A scale diagram with the estimated sizes of the introns, as determined by sizing of the PCR fragments (see Figure 22) is shown in Figure 23-A. An illustration of the location of the splice sites within the length of the cDNA is shown in 23-B.

Table 2.
Intron-exon boundary sequences and intron sizes of the human *PHGDH* gene.
Exon nucleotides are in capitals, intronic sequences are in lower-case text.

| Intron number | Exon ends at cDNA # | Donor site | Intron size | Acceptor site |
|---|---|---|---|---|
| 1 | 138 | GTGCAG gtaaggcgagagagagaaaa | 9.5 kb | aaatgtttttctgcttcag GACTGT |
| 2 | 290 | TATGAA gtaagtcatggaggctgcgg | 2.1 kb | ccctttctttgatctttag CACCCC |
| 3 | 355 | GGCCAG gtaagtccctgacttctcag | 3.5 kb | acctgtgtctatccttgcag GCAGAT |
| 4 | 411 | AAGAAG gtagcagcggccttgactcg | 0.1 kb | ttttgcctgtttggttgcag TTCATG |
| 5 | 510 | ATGAAG gtaagatgttgctggaaccc | 8.0 kb | tctctcttgcttccaaccag ACTATA |
| 6 | 643 | CGACAG gtaggtgtgtcttacattgt | 0.5 kb | gatcgttcgctttcttccag GCTTGC |
| 7 | 792 | ACGGAA gtaagtgcctggcagcctca | 1.6 kb | aggtatctctttctgggcag GAGCCG |
| 8 | 945 | GGGGTT gtaagtatcaccacctgggg | 3.1 kb | ccgctctccatcctctgcag GTGAAT |
| 9 | 1078 | CACAGG gtgagctggggaccttgcag | 1.2 kb | tagcttctctctgtccccag GAACAT |
| 10 | 1209 | TGGCCT gtgcgcccctctcccccacg | 0.9 kb | accttgccttctccacacag GTCACC |
| 11 | 1447 | ACCATG gtgaggagggcctgtagggc | 0.8 kb | ggagtttcttctatttccag GCCTCC |

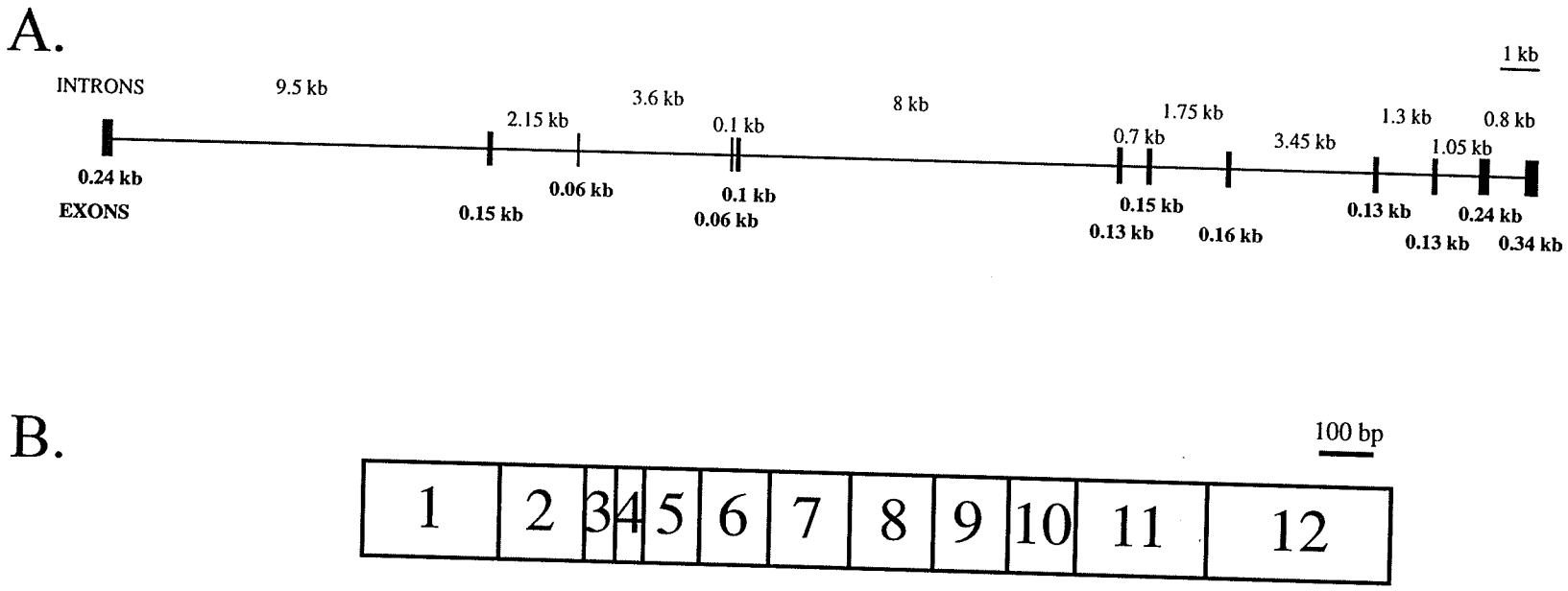



Figure 23. Scale diagram of the gene (A), and cDNA (B) encoding human PHGDH. A. Exons are represented by vertical bars, the introns are horizontal lines. The sizes of introns are given above the line, whereas the sizes of exons are below the line, all in kb. B. The *PHGDH* cDNA with vertical lines representing splice sites. Exons are numbered 1 through 12.

## 3.4 Identification of the transcriptional start sites for human liver *PHGDH*

The transcriptional start sites of *PHGDH* were characterized by sequencing cloned PCR products from a 5' RACE reaction (Ambion). A total of 22 different clones were sequenced in order to obtain an accurate estimate of the start sites utilized. These ranged from 9 to 15 bp upstream of the published 5' end of the gene (Genbank accession number AF171237), or 82 to 88 bases upstream of the translational start site (ATG). A schematic diagram illustrating the location of the transcriptional start sites is shown in Figure 24. The transcriptional initiation site found with the highest frequency was 88 bp upstream of the translational start site, with a total of 9 clones beginning at this position. Five clones began 87 bp upstream of the ATG, seven at position -83, and one clone began at nucleotide -82.

## 3.5 Examination of the possible alternative splicing of exon 1 of *PHGDH*

The similarity between rat and human genomic sequences upstream of exon 1, and the report that *PHGDH* from rat testis was alternatively spliced (82), led to the investigation of the splicing patterns of human *PHGDH* exon 1 in various human tissues. A panel of cDNAs from 24 tissues was subjected to PCR amplification with forward primers specific to either exon 1 or exon 1' sequences, and a common reverse primer (specific to exon 5). Amplification of the exon 1 forward primer with the reverse primer

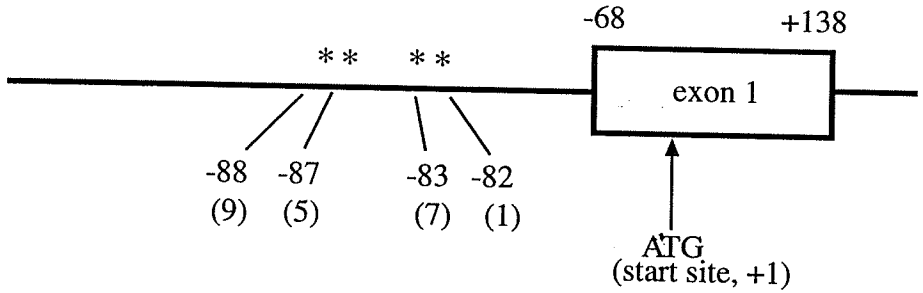



Figure 24. Transcriptional initiation sites of PHGDH in human liver. The location of the four transcriptional start sites found in human liver, measured in nucleotides from the translational start site. The number of clones with that transcriptional initiation site are given in parentheses.

produced bands of the expected size (485 bp) in nearly all tissues. There appears to be tissue-specific levels of expression that confirm the results of the Northern blots (refer to Figure 12). Minimal or no PHGDH cDNA was detected in bone marrow and peripheral blood leukocytes, and high levels of cDNA encoding PHGDH were present in liver, kidney, brain, testis, salivary gland, thyroid gland, placenta, ovary, and prostate (see Figure 25). In addition, relatively high levels of PHGDH cDNA were detected in the skin, a tissue not examined in the previous Northern blots. However, no PCR products were obtained using the exon 1' forward and exon 5 reverse primers (data not shown). The viability of the exon 1' primer was established using genomic DNA, so the absence of PCR products using the cDNA array suggests that the sequence designated exon 1' is not actively transcribed at detectable levels in the 24 human tissues tested.

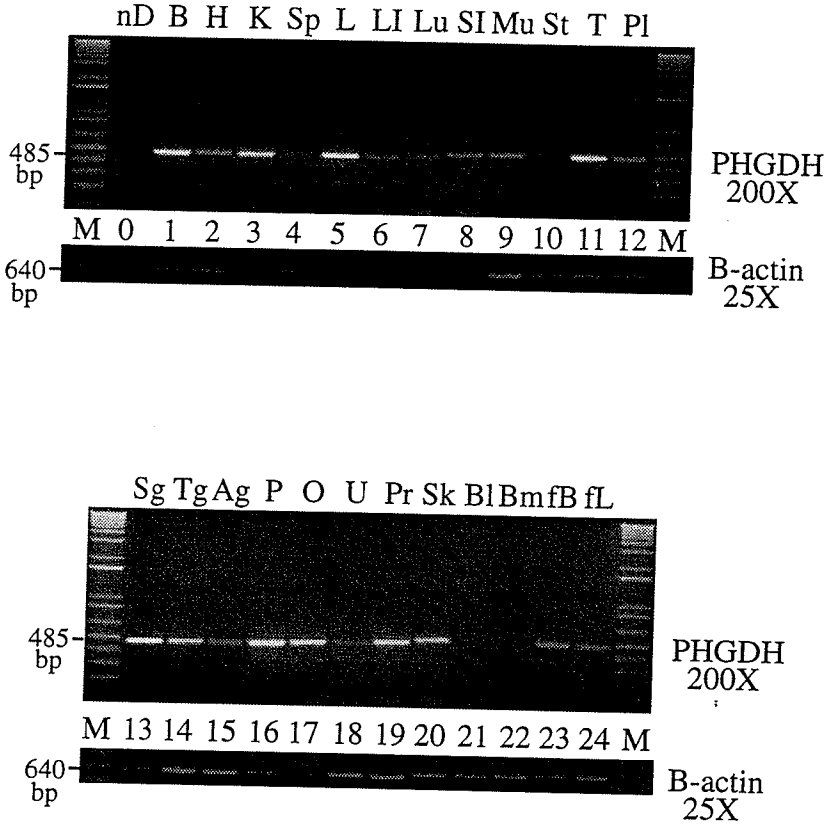



Figure 25. The expression profile of human PHGDH in multiple tissues. A fragment of *PHGDH* representing exons 1-5 was amplified from a cDNA array (Origene). Tissues represented included: 1 (B) brain, 2 (H) heart, 3 (K) kidney, 4 (Sp) spleen, 5 (L) liver, 6 (LI) large intestine, 7 (Lu) lung, 8 (SI) small intestine, 9 (Mu) muscle, 10 (St) stomach, 11 (T) testis, 12 (Pl) placenta, 13 (Sg) salivary gland, 14 (Tg) thyroid gland, 15 (Ag) adrenal gland, 16 (P) pancreas, 17 (O) ovary, 18 (U) uterus, 19 (Pr) prostate, 20 (Sk) skin, 21 (Bl) blood leukocyte, 22 (Bm) bone marrow, 23 (fB) fetal brain, 24 (fL) fetal liver. Marker lanes are labelled M, a negative control is labelled nD (no DNA, lane 0). The cDNA used to amplify *PHGDH* was at a 200X (~200 pg, upper panels) concentration. *B-actin* was amplified from the same cDNA samples at a 25X concentration (~25 pg, lower panels).

# 4. DISCUSSION

## 4.1 Presence of the 1468A allele of *PHGDH* in different populations

The purpose of our population study was to determine if the mutation (1468G→A) responsible for a deficiency of PHGDH in an Ashkenazi Jewish family (3) was present at an increased frequency in this ethnic group. A survey of 400 individuals of Jewish descent led to the discovery of one heterozygous individual (in addition to the affected family). Of over 400 persons tested, no non-Jewish individuals carried the 1468A allele. These results support a frequency of heterozygosity of 1/70 to 1/16,000 in this population, within 95% confidence limits (3). Given the relatively small sample size and the rarity of the allele, the range in the estimated frequency remains large. Nevertheless, one heterozygote in a relatively small sample of 400 individuals is a significant finding, suggesting that this allele is present at a measurable level in this ethnic group. The occurrence of the same mutation in both parents in our family may therefore be due to an unfortunate twist of fate, rather than being a case of unidentified consanguinity.

A similar study was performed on groups of Turkish and western European individuals, although only a few dozen individuals from these populations were screened for the 1468A allele (2). No heterozygous individuals were found. From these studies, we conclude that the 1468A allele is rare, in both Ashkenazi Jewish and non-Jewish

populations. However, its pan-ethnic distribution raises several interesting questions regarding the origin of this mutation and its distribution around the world.

With our present knowledge, there does not appear to be any particular group at high risk for this mutation. Therefore population testing would not be beneficial. Testing with biochemical methods (sampling the cerebrospinal fluid for abnormal serine levels) should remain the first step in diagnosing children for a possible deficiency of PHGDH. Once the serine biosynthetic pathway is implicated, genetic testing for the 1468A allele should be undertaken, with the understanding that it is but one mutation possible in the *PHGDH* gene. Should that result be negative, a skin biopsy to culture fibroblasts and test for enzyme activities should be performed to determine whether the deficiency is indeed due to a deficiency of PHGDH, or a deficiency of either of the other two enzymes of the serine biosynthetic pathway.

## 4.2 The expression of serine biosynthetic enzymes varies in human tissues

A series of experiments were performed to characterize the tissue distribution of PHGDH and the two SHMT isozymes in human tissues. The expression of PHGDH had been previously characterized in animal tissues (29), and during the course of these studies, in human tissues (2;45). Analyzing the expression pattern of the SHMT isozymes was pursued to provide further information regarding the distribution of these enzymes, as they have the potential to provide an alternate pathway for the biosynthesis of serine.

Previous expression studies indicated that *SHMT2* coding for the mitochondrial isozyme was expressed in total brain, but that *SHMT1* coding for the cytosolic isozyme was expressed at very low or negligible levels in this organ (55).

### 4.2.1 The mRNA encoding PHGDH is detected in many human tissues

Analysis of PHGDH expression in the human brain revealed a similar level of PHGDH mRNA in all eight regions tested, although slight variations in the signal intensity were observed between the different regions (Figure 11-B). These results were extended using the Multiple Tissue Expression array (Figure 12) which included samples from more regions of the brain, as well as many other tissue types. Comparison of the signal intensities on this blot, as well as the Multiple Tissue Northern that was completed before I joined the lab (Figure 11-A), indicates that the human brain expresses relatively high levels of the mRNA encoding PHGDH. These high levels of expression support the hypothesis that *de novo* synthesis of serine is required in the brain.

Unique to the brain is the blood-brain barrier, where tight junctions between the endothelial cells of the blood vessels restrict the amount of serine that can be transported from the circulation to the brain. It is assumed that the preference of the neutral amino acid carrier for amino acids, other than the small polar amino acids such as serine (39) is beneficial under normal conditions. Since the serine metabolites D-serine and glycine act as neurotransmitters, unregulated import across the blood-brain barrier might disrupt the

homeostatic regulation that one would deem necessary for such compounds. Indeed, elevated levels of glycine in the brain are associated with glycine encephalopathy (or non-ketotic hyperglycinemia, OMIM: 605899), propionic acidemia (ketotic hyperglycinemia, OMIM: 606054), and D-glyceric acidemia (OMIM: 220120) (21), all of which include some form of psychomotor retardation and seizures. This tight control over the production and distribution of serine may result in the brain's inability to counteract a serine biosynthetic insufficiency by increasing transport of serine from the circulation. This would negatively influence the development and integrity of the brain, as seen in individuals with this deficiency. It is assumed that the blood-brain barrier also affects the implementation of serine replacement therapy by requiring high daily doses before the serine levels within the central nervous system begin to rise (24).

Preliminary immunohistochemistry results show that the PHGDH antibody (1C2) and a second antibody (8B6, results not shown) recognized PHGDH in the cytoplasm of astrocytes (Figures 15-18). Astrocytes, once thought to be a supportive matrix for neurons, are now believed to play an important role in neuronal function and development. Some researchers have suggested that the existing concept of the synapse, with one axon sending a signal to the dendrite of the next neuron, should be replaced with a model of a tripartite synapse, with an upstream neuron, a downstream neuron, and an astrocyte encircling the synapse and modulating the neurochemical signals between the two neurons (93;94). Astrocytes also produce numerous substrates that direct neuronal

migration, survival, differentiation and maintenance. Previous reports on rat and mice brains revealed that astrocytes and glial cells produced PHGDH but the neuronal populations studied did not (52).

The highest level of expression of PHGDH in central nervous system tissues was found in the corpus callosum and spinal cord. These regions are composed of significant amounts of white matter, as myelinated tracts of axons connecting either two hemispheres of the brain (corpus callosum) or the brain to the peripheral nervous system (spinal cord). When examined at a cellular level, the 1C2 antibodies immunostained both astrocytes and oligodendrocytes in the white matter of several tissues (Figures 16, 17, 18), including corpus callosum. The localization of PHGDH to oligodendrocytes is a novel observation, as previous reports with rodent brain tissues noted only astrocytes or glial cells (52;53). Oligodendrocytes are the CNS cells responsible for the production of myelin used to surround the neuronal axons in white matter (37). A deficiency of PHGDH may reduce myelination by either impairing production of myelin or by reducing oligodendrocyte survival. The content of white matter in the brains of children with a deficiency of PHGDH is reduced, and is increased following serine replacement therapy, as shown by magnetic resonance imaging studies (23). This suggests that the serine insufficiency results in abnormal myelination patterns observed in these patients.

Although PHGDH was detected in astrocytes and oligodendrocytes, several types of neurons were not detected by our PHGDH antibodies. Previous reports have indicated

that Purkinje neurons in the cerebellum of rats do not express PHGDH (52), suggesting that these neurons rely on the neighboring Bergmann glia for their serine needs. Our results with the monoclonal 1C2 PHGDH antibody revealed that human Purkinje neurons and cerebellar granule cells were negative for PHGDH immunoreactivity (Figure 15-B). The Bergmann glia appear to express PHGDH (Figure 15-A) but this should be confirmed with further experiments. The molecular layer of the cerebellum (right half of Figure 15-A) had higher levels of DAB staining than the other sections of the cerebellum, and this likely represents background. The cell bodies of the Bergmann glia that lie between the molecular layer and the granular layer appear to express PHGDH. Co-immunofluorescence studies with both glial fibrillary acidic protein and the 1C2 PHGDH antibodies should be performed to confirm the presence of PHGDH in Bergmann glia and the astrocytes localized to the white matter layer in the human cerebellum. Similar experiments reported colocalization of glial fibrillary acidic protein and PHGDH in rat cerebellum (52). Other neuronal populations that were negative for PHGDH included the CA1 region in hippocampus (Figure 17-A), neurons of the dentate nucleus of the hippocampus (not shown), and pyramidal neurons of the frontal cortex (not shown).

The frontal cortex of rat brain was used in one series of experiments by Sugishita *et al.* to analyze the levels of PHGDH in astrocytes, neurons, and microglia (95). They noted that Western blots and RT-PCR supported a high level of PHGDH expression in

astrocytes, a medium level in neurons (approximately 1/3 the level in astrocytes), and a low level in microglia. Although cultured neurons were positive for PHGDH expression, the authors could not detect neuronal PHGDH expression in sections of rat brain using immunohistochemistry. Our immunohistochemistry results appear to mimic this pattern of expression, since the use of higher titres of the 1C2 antibody results in positive immunostaining of neuronal populations staining positive as well (personal communication with Dr. Del Bigio, Department of Pathology, University of Manitoba). If the endogenous levels of PHGDH protein are inversely proportional to the concentration of antibody required to recognize specific cell types, it would appear that the oligodendrocytes have less PHGDH protein than astrocytes but more than neuronal cells and microglia. Further characterization of the PHGDH antibodies (both 1C2 and 8B6) is needed at a variety of titres to confirm this preliminary result.

Tissues with white matter are not the only tissues with high expression of the mRNA encoding *PHGDH*, as the caudate nucleus consists of cell bodies of neurons and glia and displays high levels of intensity on the Multiple Tissue Expression array. Glial cells that were immunostained with the PHGDH antibody were found in several regions of the human brain, in both white and gray matter (data not shown). These preliminary studies should be extended to include colocalization of the 1C2 antibody with glial fibrillary acidic protein (GFAP) antibodies to confirm the presence of PHGDH in glial cells. PHGDH appears to be primarily cytosolic in its distribution, though it may be

present at lower levels in astrocytic processes (note the high background staining in the white matter tissues of Figures 15A and 15-C, and Figures 16-18). Since GFAP is a cytoskeletal protein present in the astrocytic processes and PHGDH appears to be primarily cytosolic in its distribution, colocalization studies might determine the subcellular localization of PHGDH. Further studies should help in the localization and identification of cell types expressing PHGDH.

High levels of the mRNA encoding PHGDH were not limited to tissues of the central nervous system. Other tissues with high levels of PHGDH included the salivary gland, pancreas, prostate, thyroid gland, bladder, liver, kidney, and thymus. Skeletal muscle has low expression of PHGDH, as do the cardiac tissues, and the gastrointestinal system. These tissues might require low amounts of serine, or may be able to obtain sufficient serine from the circulation to meet their cellular demands.

The results from the Northern blots were supported by our analysis of the putative alternative splicing of PHGDH. Amplification of cDNA from 24 human tissues was performed with primers specific for transcripts with exon 1 or exon 1' at their 5' end (Figure 25). None of the tissues tested showed alternative splicing, as determined by the lack of detection using the primer specific to exon 1' (data not shown). But the relative intensity of the products obtained with the exon 1 primer provided a semi-quantitative analysis of PHGDH mRNA in these tissues that correlated closely with the results

obtained from the Northern blots. Skin was not included in the tissues surveyed by Northern analysis, and had relatively high levels of expression on the cDNA array.

Relatively high levels of PHGDH mRNA were noted in fetal liver and kidney (Figure 12), indicating that the developing fetus can synthesize serine. The fetus may also have access to serine from the maternal circulation. Studies in sheep have indicated that there is a glycine-serine shuttle mechanism operating in the placenta (96). Labelled serine is removed from the maternal circulation by the placenta, but no corresponding increase in fetal serine is seen; instead, fetal glycine becomes labelled (97). It is believed that the fetal liver converts the glycine back to serine, allowing for regulation of serine levels in the fetus. Similar shuttles exist for the excitatory amino acids glutamate and aspartate (97). A fetus with a deficiency of PHGDH would have a lowered ability to produce endogenous serine, but the deficiency should not have a dramatic effect upon the levels of exogenous serine obtained from the maternal circulation. However, since growth retardation and congenital microcephaly exist at birth, it suggests that serine levels during gestation probably do not lie within acceptable levels. This may be compounded after birth when the infant no longer has access to maternal serine. In addition, the development of the blood-brain barrier would limit access to serine from the circulation, placing additional stress on the central nervous system. After birth, with increased function of this barrier, and loss of serine from the maternal circulation, it would be expected that serine deficiency, especially in the brain, increases in severity. There is

some clinical evidence to support this scenario, as the epileptic seizures in PHGDH deficient patients have been reported to develop only after the first few months of life (1;24;25).

The expression of the mRNA encoding PHGDH was highest in the transformed cell lines (Figure 12). Neoplastic cells are believed to reprogram their metabolism (66), increasing the content of enzymes that provide the greatest benefit to their reproductive status. Increasing the content of PHGDH and the SHMT isozyme responsible for the serine to glycine conversion would lead to an increased flux towards production of nucleotides for DNA and RNA synthesis as noted in liver 3924A carcinoma cells (65) and mitogenically-stimulated proliferating lymphocytes (64). This reprogramming appears to occur in several of the cancer cell lines studied on the blot, including K-562 cells (TYPE), HeLa S3 cells (TYPE), Raji cells (TYPE), and MOLT-4 cells (TYPE). One exception to this were Daudi cells that had undetectable levels of PHGDH (Figure 12, column 10, row F). Daudi cells are derived from lymphoblasts of a patient with Burkitt's lymphoma. As normal bone marrow and peripheral blood leukocytes have very low to negligible expression of the mRNA encoding PHGDH, the Daudi cell line must have increased its rate of proliferation over that of normal leukocytes without inducing the expression of PHGDH. It is expected that the Daudi cells have some other mechanism (uptake from the circulation, etc.) by which they obtain the serine necessary for increased proliferation.

The twelve-tissue Northern blot and the brain Northern blot (Figures 11-A and 11-B) provided information regarding the sizes of the transcripts detected by the various probes. We observed only one band of approximately 2.2 kb in all tissues, except liver which had a second band of ~2.0 kb. The ~2.2 kb band represents the full-length transcript that was also described by the other investigators. Blots by Cho *et al.* (45) and Klomp *et al.* (2) had multiple bands, as described in the thesis introduction. The smaller band seen by Cho *et al.* was likely not an alternate functional transcript, since 0.7 kb would code for less than half of the full-length protein. Klomp *et al.* (2) also reported two bands, the common 2.2 kb band and a second band at 4.5 kb. They did not venture an opinion on what this larger band could be, although one possibility would be a partially spliced *PHGDH* that still contains an intron. Further experiments to characterize the 4.5 kb band would have been interesting. The different results obtained by the other groups should not be attributed to different blots, since all three groups used blots produced by Clontech. It can be assumed that the manufacturer's production standards provide consistent quantities of mRNA on each tissue blot.

The probes and hybridization conditions used by the various groups could be likely sources of variation. Cho *et al.* (45) used a 2.4 kb cDNA probe (subsequently shown to be chimeric (82)), Klomp *et al.* used a 1.6 kb coding sequence cDNA probe (2), and we used a 0.3 kb RNA probe. While our 0.3 kb probe might not have hybridized to alternatively spliced transcripts lacking the region of our probe (nucleotides 446-744), it

should not have hybridized to non-*PHGDH* transcripts either. RNA probes form stable duplexes that can be hybridized and washed at higher temperatures, allowing higher stringency and specificity. This may be why two bands were not visualized. The twelve-tissue Northern blot performed in our laboratory prior to my project showed only one 2.2 kb band in all tissues studied, with the exception of liver where a second band of approximately 2.0 kb was visible (Figure 11-A). This blot was hybridized with an approximately 600 bp (nucleotides 150-744) cDNA probe. While both our probes were shorter than those used by the other groups, it remains possible that the stringency of the hybridization conditions varied sufficiently between the three groups to account for the observed differences.

### 4.2.2 The mRNAs encoding the *SHMT* isozymes display tissue-specific expression

During the course of my studies, it was shown that astrocytes can metabolize glycine into serine and other metabolites, so at least a subset of cells in the brain are capable of synthesizing serine from glycine (47;48). In order to further characterize this potentially important alternate pathway of serine biosynthesis, I determined the expression pattern of the cytoplasmic and mitochondrial isozymes of human SHMT using the Northern blots available from my previous studies. Other investigators had determined that low levels of the mitochondrial isoform were present in samples of mRNA from total brain (55). However, the distribution of mSHMT in the different

regions of the brain was not reported. These investigators also showed that the cytoplasmic isoform was confined primarily to liver and kidney, with low levels in skeletal muscle, the only other signal seen in the eight tissues they examined (55).

My analysis of the distribution of the mRNA encoding the cytoplasmic SHMT isoform confirmed the high levels reported in liver and kidney by Girgis *et al.* (55), and extended this study to over 60 different tissues (Figure 13). With the exception of thyroid gland and thymus (at 24% and 12%, respectively), all of the other tissues, including all of the brain tissues on this blot had 6% or less of the signal seen in liver and kidney. As the cytoplasmic isozyme of SHMT appears to catalyze only the conversion of glycine to serine and not the reverse reaction (32), it suggests that only liver and kidney could synthesize appreciable amounts of serine using the cytosolic SHMT isozyme. This conversion probably accounts for the ability of these organs to synthesize glucose from glycine (98;99).

The distribution of the mitochondrial SHMT isozyme was more ubiquitous than that of the cytosolic SHMT isozyme, with significant levels in nearly all tissues (Figure 13). The organs with the relatively highest levels of this mRNA were liver and lymph node, followed by pancreas and kidney. Combined with the view that this isoform of SHMT can reversibly interconvert glycine and serine, depending on intracellular conditions (100), its presence in all of the brain tissues (at 10-80% of the level seen in kidney) suggests that this pathway could produce serine in the brain. Indeed, the presence

of the glycine-cleavage complex in astrocytes (101) and the demonstrated synthesis of serine from glycine in astrocyte-rich cultures (47) supports this suggestion.

The ability to produce serine through the mitochondrial SHMT pathway could theoretically decrease the neurological serine insufficiency observed in children with a deficiency of PHGDH. The viability of this rescue would depend upon several factors, especially the presence of the glycine-cleavage complex and mitochondrial SHMT, as well as sufficient levels of glycine as a substrate for this pathway. However, children with a deficiency of PHGDH also have a glycine insufficiency (1;24;25), since glycine is produced from serine (1). It is unlikely that the SHMT pathway can counteract the deficiency of PHGDH using already lowered levels of glycine accounting for the dramatic clinical impairment we see in PHGDH deficient patients.

## 4.3 The genomic organization of human *PHGDH*

The structure and organization of the gene encoding PHGDH was characterized using a PCR-based cloning strategy. The gene is divided into 12 exons by 11 introns ranging in size from 99 bp to over 9000 bp. These were localized and their intron-exon boundary sequences were obtained. The splice sites all contained a gt/ag consensus sequence (102). My analysis represents the first experimental confirmation of the sequencing data generated by the Human Genome Project, which was released after I had cloned 10 out of the 11 introns. The availability of this information makes it now possible

to characterize individual *PHGDH* exons. This capability will be important for the identification of other mutations in this gene in the patient population.

The rat *PHGDH* gene also contains 12 exons and 11 introns (82). Comparison between the rat and human sequences reveals that the splice sites are conserved, though the sizes of the introns vary. Introns within the human gene are generally longer, with the exception being intron five: it is approximately 8 kb in human and 10.5 kb in rat.

A search of the Ensembl mouse genome (103) (http://www.ensembl.org/Mus_musculus/, version 4.1.1) using the cDNA of human PHGDH (Genbank accession number AF171237) revealed 22 homologous sequences. One of these on mouse chromosome 3 contains thirteen smaller alignments, approximately the same size as the exons of the human gene. The other 21 homologues appear to be pseudogenes, containing cDNA sequences without introns. Experimental confirmation of this data is required.

The major problem I encountered with the PCR-based method was difficulty in amplifying intron 1 of the human *PHGDH* gene. The region was subsequently found to be relatively GC-rich, and initial PCR experiments revealed multiple bands that could not be resolved by altering magnesium concentrations, the elongation times, or the annealing temperatures during the PCR reactions. Sequencing of the largest band from the PCR reactions revealed sequences with the reverse PCR primer (*i.e.* exon 2) at either end of the insert DNA. When the Celera sequence was published, the sequence between exons 1 and 2 was analyzed. A web-based program [(104), http://125.itba.mi.cnr.it/webgene] was

used to identify repetitive elements, and several were localized within the intron. Since repetitive elements interfere with the annealing of primers in PCR reactions and can self-anneal, a variety of different additives were tested for their ability to improve the reactions. A concentration of 5% dimethylsulfoxide allowed the successful amplification of an approximately 9.5 kb fragment with minimal other products (Figure 22, lane two). This fragment was cloned and revealed exon one and exon two sequences at either end, separated by a large intron. The method by which dimethylsulfoxide aids in the PCR reaction is not known, but it is believed to help in the denaturing of the DNA template (105). This additive was also required in the sequencing reaction before the fragment could be successfully sequenced.

A detailed study of the genomic organization of a gene may soon become obsolete with the publication of the human genome by the International Human Genome Consortium (84) and the genome sequence published by Celera (85). There are several differences between the data reported by the public Human Genome project and the privately funded Celera project. The former used a clone-by-clone or hierarchical shotgun sequencing method. This involved creating a library of BAC clones whose large inserts were mapped and sequenced (84). This should have minimized the amount of computational analysis needed to resolve the genome, as the sequences were already locally aligned and mapped. The Celera sequence was determined using a whole-genome shotgun sequencing approach (85). BAC libraries of different sizes (2, 10, or 50 kb) were

produced and the ends from the inserts (~550 bp of each end) were sequenced (85). The distance between the sequenced ends could be estimated by determining which library the clone was from (2, 10, or 50 kb insert sizes), and the ends were mapped using computational methods. Celera admits that clone-by-clone sequencing provides a more detailed map, especially near the pericentromeric regions (85). The reasons behind their decision to pursue the whole-genome shotgun sequencing strategy were the increased cost and lower efficiency of the clone-by-clone strategy. The major problem for either methodology is the highly repetitive nature of the human genome.

Repetitive elements are believed to occupy 40-50 % of the human genome (84;85). There are four main classes of repetitive elements: long interspersed nuclear elements (6-7 kb in length), short interspersed nuclear elements (~0.3 kb), long terminal repeats, and DNA transposons (102). Repetitive elements are present in the genome in many thousands of copies, in some cases up to millions of copies. Several short interspersed nuclear repeats were present in intron one of the human PHGDH gene, and their presence caused difficulty in successfully amplifying the PCR product containing this intron.

Sequencing fragments of the BAC clones and their subsequent alignment could be problematic if large or frequent repetitive elements exist in a region. Initial drafts of the human chromosome one placed exon one of PHGDH more than 30 kb upstream from the rest of the gene, due to misaligned contigs. Clones containing repetitive elements can

also suffer from rearrangements, supporting the idea that all information from public or private databases should be experimentally verified.

Pericentromeric regions consist of constitutive heterochromatin and were not sequenced in either of the genome projects, because they contain large amounts of repetitive sequences (84;85). The public human genome project has assigned *PHGDH* very close to the edge of the pericentromeric heterochromatin region of chromosome 1p (84). This is in contrast to Celera's placement of PHGDH on the long arm of chromosome 1 (85). Further analysis will be required to resolve this discrepancy.

## 4.4 Identification of the transcriptional initiation sites of human liver *PHGDH*

The transcriptional start sites in human liver *PHGDH* were characterized through a 5' RACE and sequencing strategy. Sequencing of 22 clones revealed four transcriptional start sites clustered between 82 and 88 base pairs upstream of the translational start site. Although there were multiple start sites, they were tightly clustered, suggesting the nearby presence of a promoter. These transcription start sites may also be used in other tissues besides liver, but this will need verification since tissue-specific transcription initiation is known to exist (102). As PHGDH is highly expressed in liver, and the liver 5' RACE kit was available, we began our studies in that tissue.

The transcriptional start sites in rat *PHGDH* were analyzed from five tissues: liver, lung, brain, testis, and kidney (82). One site was as far downstream as position –61,

seven were at -83, ten at -89, and one each at -92, -103, -108, and –110 (82). Similar to our results, they show clustering at the –83 and -89 positions, but several transcripts extended farther upstream. There was no preference for transcription initiation sites within tissues, and despite the dietary regulation reported only for rat liver PHGDH, the transcripts from that organ did not differ from the four other tissues studied (82).

Robbi *et al.* (82) concluded that the observed heterogeneity in the transcription initiation sites supported the idea that PHGDH was a 'housekeeping gene', based on the lack of a 'TATA' box. Although computer analysis of a 1 kb section of sequence upstream of the human PHGDH translation initiation site reveals 8 conserved TATA boxes [Webgene (104)], the closest lies 300 bp upstream and not the conserved -20 to -30 bp (102). However, a TATA box is not present upstream of the glucocorticoid receptor, and it is not considered a housekeeping gene (102). Housekeeping genes, generally required by many tissues, are believed to have a basal level of transcription and therefore do not require a specific promoter (102). With the highly variable levels of PHGDH expression observed in our Northern blots (Figures 11, 12) and our PCR analysis (Figure 25), the concept that all tissues possess a basal level of expression seems unlikely. Bone marrow and peripheral blood leukocytes have undetectable levels of PHGDH mRNA and even within tissues which have relatively levels of high expression (corpus callosum, hippocampus), not all cells express the PHGDH protein (Figures 17, 18). These variable

levels of expression suggest that some method of regulation of PHGDH mRNA or protein levels is present.

## 4.5 Human *PHGDH* does not undergo alternative splicing at the 5' end of the gene

The possibility that alternative splicing occurs in the 5' untranslated region of human *PHGDH* was investigated using a PCR-based strategy. Exon 1-containing products were amplified in 21 of 24 tissues, and the amount of product in each tissue was very similar to the level of expression observed on the Northern blots when hybridized with a PHGDH probe. This provided an alternate method of determining the expression pattern of PHGDH in human tissues.

Robbi *et al.* (82) found one clone that appeared to have alternative splicing during their transcription start site analysis of rat *PHGDH*. One of four clones from rat testis began at position –315, the sequence spanning –275 to –30 was missing, and the transcript sequence then continued with exon 1. Four clones is not a robust sample, and further testing should help determine the frequency of this splicing variant. This change in splicing affects the 5' untranslated region of the transcript, but the same coding sequence remained (82). It is not known what, if any, functional significance this might have for PHGDH. Exon 1'-containing cDNAs were not amplified in the 24 human tissues I tested. Although the splice acceptor site of rat *PHGDH* in exon 1 is not conserved in

humans, there is a potential splice site a few bases away. This region, although fairly conserved between rats and humans, does not appear to undergo alternative splicing in humans.

## 5. FUTURE STUDIES

The study of 3-phosphoglycerate dehydrogenase and serine biosynthesis and their effects on neurological development and function is still in a state of relative infancy. A wide variety of experiments can be performed that will add to our knowledge and eventually benefit children with a deficiency of PHGDH. For this portion of my thesis, I will focus on possible future experiments.

The 1C2 PHGDH antibody appears to work well with paraffin-embedded human tissue. Further characterization of the antibody and the cell types it recognizes would be required to complete the studies I began. Once the antibody and cell types are characterized, analysis of tissues from the central nervous systems of individuals of different ages would help determine the expression pattern of PHGDH at a cellular level during development. The experiments with the 1C2 antibody can also be expanded to other tissues that express PHGDH. There is a commercial paraffin-embedded tissue array with small micro-sections of different tissues of the human body that could be used to investigate the cellular profile of PHGDH expression in the different tissues that express high levels of this enzyme.

The experiments on the transcriptional start site and genomic organization of human PHGDH provide a basis for the characterization of the regulation of expression of *PHGDH* in humans. Knowledge of the promoter elements and the regulation of this gene could lead to experimental (and possibly even therapeutic) manipulation of PHGDH

expression. Determining methods to increase the expression of partial activity mutants present in children with a deficiency of PHGDH might help decrease the severity of symptoms, especially when used in conjunction with serine replacement therapy.

## 6. REFERENCES


1. Jaeken, J., Detheux, M., van Maldergem, L., Foulon, M., Carchon, H., and Van Schaftingen, E. (1996) *Arch.Dis.Child* **74,** 542-545

2. Klomp, L. W., de Koning, T. J., Malingre, H. E., van Beurden, E. A., Brink, M., Opdam, F. L., Duran, M., Jaeken, J., Pineda, M., van Maldergem, L., Poll-The, B. T., van, d. B., I, and Berger, R. (2000) *Am.J.Hum.Genet.* **67,** 1389-1399

3. Pind, S., Slominski, E., Mauthe, J., Pearlman, K., Swoboda, K. J., Wilkins, J. A., Sauder, P., and Natowicz, M. R. (2002) *J.Biol.Chem.* **277,** 7136-7143

4. Etheridge, Jr. J. E. Birth Defects and developmental disorders (Chapter 3) in Pediatric Neurology. Farmer, T. W. (editor). (1983) Philadelphia, PA, Harper and Row Publishers.

5. Menkes, J. H. and Till K. Malformations of the central nervous system (Chapter 4) in Textbook of Child Neurology. Menkes, J. H. (editor) 5th edition. (1995) Philadelphia, PA, Williams & Wilkins.

6. Cowie, V. A. (1987) *J.Ment.Defic.Res.* **31 ( Pt 3),** 229-233

7. Tolmie, J. L., McNay, M., Stephenson, J. B., Doyle, D., and Connor, J. M. (1987) *Am.J.Med.Genet.* **27,** 583-594

8. Dorman, C. (1991) *Dev.Med.Child Neurol.* **33,** 267-269

9. Qazi, Q. H. and Reed, T. E. (1975) *Clin.Genet.* **7,** 85-90

10. Volpe, J. J. Neurology of the Newborn. 3rd edition. (1995) Toronto, ON, W.B. Saunders Company.

11. Book, J. S., Schutt, J. W., Reed, S. C. (1953) *Am.J.Ment.Defic.* **57,** 637-660

12. Gross-Tsur, V., Joseph, A., Blinder, G., and Amir, N. (1995) *Clin.Genet.* **47,** 33-37

13. Silengo, M., Lerone, M., Martinelli, M., Martucciello, G., Caffarena, P. E., Jasonni, V., and Romeo, G. (1992) *Clin.Genet.* **42,** 152-155

14. Scheffer, I. E., Baraitser, M., Wilson, J., Godfrey, C., and Brett, E. M. (1992) *Neuropediatrics* **23,** 53-56

15. Jackson, A. P., McHale, D. P., Campbell, D. A., Jafri, H., Rashid, Y., Mannan, J., Karbani, G., Corry, P., Levene, M. I., Mueller, R. F., Markham, A. F., Lench, N. J., and Woods, C. G. (1998) *Am.J.Hum.Genet.* **63,** 541-546

16. Roberts, E., Jackson, A. P., Carradice, A. C., Deeble, V. J., Mannan, J., Rashid, Y., McHale, D. P., Markham, A. F., Lench, N. J., Woods, C. G. (1999) *Eur.J.Hum.Genet.* **7,** 815-820

17. Moynihan, L., Jackson, A. P., Roberts, E., Karbani, G., Lewis, I., Corry, P., Turner, G., Mueller, R. F., Lench, N. J., and Woods, C. G. (2000) *Am.J.Hum.Genet.* **66,** 724-727

18. Jamieson, C. R., Govaerts, C., Abramowicz, M. J. (1999) *Am.J.Hum.Genet.* **65,** 1465-1469

19. Pattison, L., Crow, Y. J., Deeble, V. J., Jackson, A. P., Jafri, H., Rashid, Y., Roberts, E., and Woods, C. G. (2000) *Am.J.Hum.Genet.* **67,** 1578-1580

20. Jamieson, C. R., Fryns, J. P., Jacobs, J., Matthijs, G., and Abramowicz, M. J. (2000) *Am.J.Hum.Genet.* **67,** 1575-1577

21. Online Mendelian Inheritance in Man, OMIM. McKusick-Nathans Institute for Genetic Medicine, Johns Hopkins University (Baltimore, MD) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, MD), 2000. World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/

    OMIM: 601815 (PHGDH deficiency, last updated 1/29/2001), OMIM:272800 (TaySachs disease, last updated 3/2/2001), OMIM:230800 (Gaucher disease, last updated 11/29/2001), OMIM:271900 (Canavan disease, last updated 3/27/2001), OMIM:257200 (Neiman Pick disease, last updated 11/1/2000), OMIM:605899 (nonketotic hyperglycinemia, last updated 3/12/2001), OMIM:606054 (propionic acidemia, last updated 12/21/2001), OMIM:220120 (D-glyceric acidemia, last updated 6/22/2001).

22. de Koning, T. J., Poll-The, B. T., and Jaeken, J. (1999) *Neuropediatrics* **30,** 1-4

23. de Koning, T. J., Jaeken, J., Pineda, M., van Maldergem, L., Poll-The, B. T., and van der Knaap, M. S. (2000) *Neuropediatrics* **31,** 287-292

24. de Koning, T. J., Duran, M., Dorland, L., Gooskens, R., Van Schaftingen, E., Jaeken, J., Blau, N., Berger, R., and Poll-The, B. T. (1998) *Ann.Neurol.* **44,** 261-265

25. Pineda, M., Vilaseca, M. A., Artuch, R., Santos, S., Garcia Gonzalez, M. M., Aracil, A., Van Schaftingen, E., and Jaeken, J. (2000) *Dev.Med.Child Neurol.* **42,** 629-633

26. Hausler, M. G., Jaeken, J., Monch, E., and Ramaekers, V. T. (2001) *Neuropediatrics* **32,** 191-195

27. Hamamy, H. and Alwan, A. (1997) *Eastern Mediterranean Health Journal* _, 123-132

28. Mange, A. P. and Mange E. J. Genetics: Human Aspects. 2nd edition. (1990) Sunderland, MA, Sinauer Associates, Inc.

29. Snell, K. (1984) *Adv.Enzyme Regul.* **22,** 325-400

30. Garrow, T. A., Brenner, A. A., Whitehead, V. M., Chen, X. N., Duncan, R. G., Korenberg, J. R., and Shane, B. (1993) *J.Biol.Chem.* **268,** 11910-11916

31. Pfendner, W. and Pizer, L. I. (1980) *Arch.Biochem.Biophys.* **200,** 503-512

32. Narkewicz, M. R., Sauls, S. D., Tjoa, S. S., Teng, C., and Fennessey, P. V. (1996) *Biochem.J.* **313 ( Pt 3),** 991-996

33. Wolosker, H., Blackshaw, S., and Snyder, S. H. (1999) *Proc.Natl.Acad.Sci.U.S.A* **96,** 13409-13414

34. Mothet, J. P., Parent, A. T., Wolosker, H., Brady, R. O., Jr., Linden, D. J., Ferris, C. D., Rogawski, M. A., and Snyder, S. H. (2000) *Proc.Natl.Acad.Sci.U.S.A* **97,** 4926-4931

35. Walter, J. H. (1996) *Lancet* **348,** 558-559

36. Fell, D. A. and Snell, K. (1988) *Biochem.J.* **256,** 97-101

37. Kingsley, R. E. Concise Text of Neuroscience. 2nd edition. (2000) Philadelphia, PA, Lippincott Williams & Wilkins.

38. Oldendorf, W. H. and Szabo, J. (1976) *Am.J.Physiol* **230,** 94-98

39. Smith, Q. R., Momma, S., Aoyagi, M., and Rapoport, S. I. (1987) *J.Neurochem.* **49,** 1651-1658

40. Bridgers, W. F. (1965) *J.Biol.Chem.* **240,** 4591-4597

41. Grant, G. A. and Bradshaw, R. A. (1978) *J.Biol.Chem.* **253,** 2727-2731

42. Lund, K., Merrill, D. K., and Guynn, R. W. (1986) *Biochem.J.* **238,** 919-922

43. Willis, J. E. S. H. J. (1964) *Biochim.Biophys.Acta* **81,** 39-54

44. Snell, K. and Weber, G. (1986) *Biochem.J.* **233,** 617-620

45. Cho, H. M., Jun, D. Y., Bae, M. A., Ahn, J. D., and Kim, Y. H. (2000) *Gene* **245,** 193-201

46. Dringen, R., Verleysdonk, S., Hamprecht, B., Willker, W., Leibfritz, D., and Brand, A. (1998) *J.Neurochem.* **70,** 835-840

47. Verleysdonk, S., Martin, H., Willker, W., Leibfritz, D., and Hamprecht, B. (1999) *Glia* **27,** 239-248

48. Verleysdonk, S. and Hamprecht, B. (2000) *Glia* **30,** 19-26

49. Mitoma, J., Furuya, S., and Hirabayashi, Y. (1998) *Neurosci.Res.* **30,** 195-199

50. Mitoma, J., Kasama, T., Furuya, S., and Hirabayashi, Y. (1998) *J.Biol.Chem.* **273,** 19363-19366

51. Savoca, R., Ziegler, U., and Sonderegger, P. (1995) *J.Neurosci.Methods* **61,** 159-167

52. Furuya, S., Tabata, T., Mitoma, J., Yamada, K., Yamasaki, M., Makino, A., Yamamoto, T., Watanabe, M., Kano, M., and Hirabayashi, Y. (2000) *Proc.Natl.Acad.Sci.U.S.A* **97,** 11528-11533

53. Yamasaki, M., Yamada, K., Furuya, S., Mitoma, J., Hirabayashi, Y., and Watanabe, M. (2001) *J.Neurosci.* **21,** 7691-7704

54. Gressens, P. (2000) *Pediatr.Res.* **48,** 725-730

55. Girgis, S., Nasrallah, I. M., Suh, J. R., Oppenheim, E., Zanetti, K. A., Mastri, M. G., and Stover, P. J. (1998) *Gene* **210,** 315-324

56. Achouri, Y., Robbi, M., and Van Schaftingen, E. (1999) *Biochem.J.* **344 Pt 1,** 15-21

57. Hayashi, S., Tanaka, T., Naito, J., and Suda, M. (1975) *J.Biochem.(Tokyo)* **77,** 207-219

58. Fallon, H. J., Davis, J. L., and Goyer, R. A. (1968) *J.Nutr.* **96,** 220-226

59. Johnson, B. E. W. D. A. S. H. J. (1964) *Biochim.Biophys.Acta* **85,** 202-205

60. Cheung, G. P., Cotropia, J. P., and Sallach, H. J. (1969) *Arch.Biochem.Biophys.* **129,** 672-682

61. Mauron, J., Mottu, F., and Spohr, G. (1973) *Eur.J.Biochem.* **32,** 331-342

62. Sallach, H. J., Sanborn, T. A., and Bruin, W. J. (1972) *Endocrinology* **91,** 1054-1063

63. Davis, J. L., Fallon, H. J., and Morris, H. P. (1970) *Cancer Res.* **30,** 2917-2920

64. Snell, K., Natsumeda, Y., and Weber, G. (1987) *Biochem.J.* **245,** 609-612

65. Snell, K. (1985) *Biochim.Biophys.Acta* **843,** 276-281

66. Snell, K., Natsumeda, Y., Eble, J. N., Glover, J. L., and Weber, G. (1988) *Br.J.Cancer* **57,** 87-90

67. Tobey, K. L. and Grant, G. A. (1986) *J.Biol.Chem.* **261,** 12179-12183

68. Schuller, D. J., Fetter, C. H., Banaszak, L. J., and Grant, G. A. (1989) *J.Biol.Chem.* **264,** 2645-2648

69. Grant, G. A., Schuller, D. J., and Banaszak, L. J. (1996) *Protein Sci.* **5,** 34-41

70. Schuller, D. J., Grant, G. A., and Banaszak, L. J. (1995) *Nat.Struct.Biol.* **2,** 69-76

71. Chipman, D. M. and Shaanan, B. (2001) *Curr.Opin.Struct.Biol.* **11,** 694-700

72. Grant, G. A., Xu, X. L., and Hu, Z. (2000) *Biochemistry* **39,** 7316-7319

73. Grant, G. A., Xu, X. L., and Hu, Z. (2000) *Arch.Biochem.Biophys.* **375,** 171-174

74. Grant, G. A., Kim, S. J., Xu, X. L., and Hu, Z. (1999) *J.Biol.Chem.* **274,** 5357-5361

75. Grant, G. A., Xu, X. L., and Hu, Z. (1999) *Protein Sci.* **8,** 2501-2505

76. Zhao, G. and Winkler, M. E. (1996) *J.Bacteriol.* **178,** 232-239

77. Grant, G. A., Hu, Z., and Xu, X. L. (2001) *J.Biol.Chem.* **276,** 1078-1083

78. Al Rabiee, R., Zhang, Y., and Grant, G. A. (1996) *J.Biol.Chem.* **271,** 23235-23238

79. Al Rabiee, R., Lee, E. J., and Grant, G. A. (1996) *J.Biol.Chem.* **271,** 13013-13017

80. Grant, G. A., Hu, Z., and Xu, X. L. (2001) *J.Biol.Chem.* **276,** 17844-17850

81. Achouri, Y., Rider, M. H., Schaftingen, E. V., and Robbi, M. (1997) *Biochem.J.* **323 ( Pt 2),** 365-370

82. Robbi, M., Achouri, Y., Szpirer, C., and Van Schaftingen, E. (2000) *Mamm.Genome* **11,** 1034-1036

83. Baek, J. Y., Jun, D. Y., Taub, D., and Kim, Y. H. (2000) *Cytogenet.Cell Genet.* **89,** 6-7

84. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., *et al.* (2001) *Nature* **409,** 860-921

85. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., *et al.* (2001) *Science* **291,** 1304-1351

86. Ho, C. L., Noji, M., Saito, M., and Saito, K. (1999) *J.Biol.Chem.* **274,** 397-402

87. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673-4680

88. Short Protocols in Molecular Biology. Ausubel, F. M. Brent R. Kingston R. E. Moore D. D. Seidman J. G. Smith J. A. Struhl K. 3rd ed. (1995) John Wiley & Sons, Inc.

89. Birnboim, H. C. and Doly, J. (1979) *Nucleic Acids Res.* **7,** 1513-1523

90. Molecular Cloning: A Laboratory Manual. Sambrook, J. Fritsch E. F. and Maniatis T. 2nd ed. 1989. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

91. Grundemann, D. and Schomig, E. (1996) *Biotechniques* **21,** 898-903

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

93. Vesce, S., Bezzi, P., Volterra, A. (1999) *Cell Mol Life Sci* **56,** 991-1000

94. Araque, A., Parpura, V., Sanzgiri, R. P., Haydon, P. G. (1999) *Trends Neurosci* **22,** 208-215

95. Sugishita, H., Kuwabara, Y., Toku, K., Doi, L., Yang, L., Mitoma, J., Furuya, S., Hirabayashi, Y., Maeda, N., Sakanaka, M., and Tanaka, J. (2001) *J.Neurosci.Res.* **64,** 392-401

96. Cetin, I. (2001) *Pediatr.Res.* **49,** 148-154

97. Moores, R. R., Jr., Rietberg, C. C., Battaglia, F. C., Fennessey, P. V., and Meschia, G. (1993) *Pediatr.Res.* **33,** 590-594

98. Cowin, G. J., Willgoss, D. A., and Endre, Z. H. (1996) *Biochim.Biophys.Acta* **1310,** 41-47

99. Cowin, G. J., Willgoss, D. A., Bartley, J., and Endre, Z. H. (1996) *Biochim.Biophys.Acta* **1310,** 32-40

100. Daly, E. C. and Aprison, M. H. (1974) *J.Neurochem.* **22,** 877-885

101. Sato, K., Yoshida, S., Fujiwara, K., Tada, K., and Tohyama, M. (1991) *Brain Res.* **567,** 64-70

102. Strachan, T., and Read A. P. Human Molecular Genetics. 2nd edition. (1999) Toronto, ON, John Wiley & Sons, Inc.

103. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., Mongin, E., Pettett, R., Pocock, M., Potter, S., Rust, A., Schmidt, E., Searle, S., Slater, G., Smith, J., Spooner, W., Stabenau, A., Stalker, J., Stupka, E., Ureta-Vidal, A., Vastrik, I., and Clamp, M. (2002) *Nucleic Acids Res.* **30,** 38-41 World Wide Web URL: http://www.ensembl.org/Mus_musculus/

104. Milanesi, L., D'Angelo, D., and Rogozin, I. B. (1999) *Bioinformatics.* **15,** 612-621 World Wide Web URL: http://125.itba.mi.cnr.it/webgene/

105. Frackman, S. Kobs G. Simpson D. Storts D. Betaine and DMSO : Enhancing agents for PCR. ProMega Notes 65, 27. 1998. World Wide Web URL: http://www.promega.com/pnotes/65/6921_27/default.html

## 7. APPENDIX I

PCR primers used to amplify the region of *PHGDH* surrounding nucleotide 1468.

| | Primer sequence (5' to 3') | location in cDNA |
|---|---|---|
| 1 | CTCCGCTCATTGCACCTTTGA | 115-93 bp upstream of cDNA +1448 |
| 2 | ATGCTGCTTCCACGCTTCCA | +1572 to +1553 |

PCR primers used to amplify the 3' UTR of the SHMT genes

Nucleotides shown in bold represent *Hind* III restriction enzyme sites, those underlined indicate *Xba* I restriction enzyme sites.

cytosolic SHMT (*SHMT1*) gene

| | Primer sequence (5' to 3') | location |
|---|---|---|
| 3 | TATACA**AAGCTT**GTGTCAGAGCCACCCTGAAAGAG | 1319-1342 |
| 4 | ATAGTCTAGAATGAGACTTAAACAAATTTTGAT | 1633-1610 |

mitochondrial SHMT (*SHMT2*) gene

| | Primer sequence (5' to 3') | location |
|---|---|---|
| 5 | TATAC**AAGCTT**AGGACTCAGAAACAAGTCAGC | 1358-1379 |
| 6 | ATAGTCTAGAGTGCCTGGGCCCTCCTTCTAAGTC | 1671-1647 |

PCR primers utilized for amplifying fragments of human *PHGDH* gene.

| | Primer sequence (5' to 3') | location in cDNA |
|---|---|---|
| 7 | CTCCAGCAATGGCTTTTG | -8 to +10 |
| 8 | TTGCGGCCTCCAGATCCACAT | +265 to +245 |
| 9 | CTGAGAAACTCCAGGTGG | +200 to +217 |
| 10 | CATGATCATTCCACAAGTGAGTTCT | +345 to +321 |
| 11 | CGCAGAACTCACTTGTGGAATGA | +318 to +340 |
| 12 | AAGCCGTCGCCTGGGGAATCT | +379 to +359 |
| 14 | GGACTGCATCCGGGTAGCT | +498 to +480 |
| 15 | AGCTACCCGGATGCAGTCC | +480 to +498 |
| 16 | ATCACAGAGAGGCCAGATCTCCT | +603 to +581 |
| 17 | CTGGCCTCTCTGTGATTTCATCA | +587 to +610 |
| 18 | CCAGTGCAGCCCCGGCACA | +778 to +760 |
| 19 | TGTGCCGGGGCTGCACTGG | +760 to +778 |
| 20 | CTCCCCACAGCGGCTCTGAGC | +891 to +871 |
| 21 | CAGAGCCGCTGTGGGGAG | +874 to +891 |
| 22 | CCCTTGGGGGACCCAGC | +1056 to +1039 |
| 23 | CTGGGTCCCCCAAGGG | +1040 to +1056 |
| 24 | TCCGCCTGCTTGGAAGCC | +1160 to +1143 |
| 25 | GCTTCCAAGCAGGCGGATGTG | +1144 to +1164 |
| 26 | CCCCTGCAGTACAGGTGTAGTG | +1341 to +1320 |
| 27 | CACTACACCTGTACTGCAGGGG | +1320 to +1341 |
| 28 | AGGTTAGAAGTGGAACTGGAGGC | +1605 to +1581 |

## 7. Appendix I (continued)

M13 primers used to sequence DNA within TopoXL vectors (used at 20 pmol/µl)

| | |
|---|---|
| M13 forward | CTGGCCGTCGTTTTAC |
| M13 reverse | CAGGAAACAGCTATGAC |

PCR primers used to amplify exon 1 vs exon 1' products

| | Primer sequence (5' to 3') | location in DNA |
|---|---|---|
| 29 | GCAGCTTCTTGGCTTAGGTAC | -60 to -40 |
| 30 | GCCTTCACAGTCCTGCATCTC | +424 to +403 |
| 31 | CGTAAATGTGCTTTTTATATGCTCAC | -315 to -290 |
| B-actin forward | GCATGGGTCAGAAGGAT | |
| B-actin reverse | CCAATGGTGATGACCTG | |

PCR primers supplied with 5' RACE kit

| | |
|---|---|
| outer linker primer | GCTGATGGCGATGAATGAACACTG |
| inner linker primer | CGCGGATCCGAACACTGCGTTTGCTGGCTTTGATG |