NOVEL SPERM-SPECIFIC GLYCOLYTIC ISOZYMES GENERATED BY RETROTRANSPOSITION AND ALTERNATIVE SPLICING

Soumya Vemuganti

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Cell and Development Biology.

Chapel Hill
2009

Approved by:

Dr. Deborah A. O'Brien

Dr. Mohanish Deshmukh

Dr. Edward M. Eddy

Dr. Fernando Pardo Manuel de Villena

Dr. Dazhi Wang

ABSTRACT

Soumya Vemuganti: Novel Sperm-Specific Glycolytic Isozymes Generated by Retrotransposition and Alternative Splicing

(Under the direction of Dr. Deborah A. O'Brien)

Targeted gene disruption of glycolytic enzymes expressed only during spermatogenesis indicates that glycolysis is essential for sperm motility and male fertility. This pathway is compartmentalized in the principal piece of the sperm flagellum, where several spermatogenic cell-specific isozymes are bound to a cytoskeletal structure known as the fibrous sheath. Fructose-1,6-bisphosphate aldolase, or aldolase, is the fourth enzyme in the glycolytic pathway. Genomic and proteomic analyses identified three aldolase A (*Aldoa*) transcripts generated by retrotransposition and alternative splicing: *Aldoa_v2, Aldoart1,* and *Aldoart2*. Unique nucleotide sequence in *Aldoart1* provided evidence for an alternatively spliced exon in *Aldoa_v2*. Expression of all three isozymes was restricted to the male germline and was regulated at transcriptional and translational levels. Both ALDOART1 and ALDOA_V2 are tightly bound to the fibrous sheath and have unique N-terminal extensions that may mediate this binding. ALDOA_V2 is conserved across species and is expressed in rat and human sperm. Preliminary modeling data predicted unique amino acids near functional domains in each isozyme, suggesting distinctive binding and/or catalytic properties. Recombinant sperm ALDOA-related isozymes demonstrated reduced activity when expressed in *E. coli*, perhaps due to difficulties producing proteins with native conformations. However, analysis of aldolase kinetic parameters in mouse sperm identified a significant portion of the total activity in insoluble fractions, providing initial evidence that the novel ALDOA-related isozymes bound to the fibrous sheath are active.

Glycolytic enzymes with restricted expression during spermatogenesis arose via gene duplication (*Gapdhs*) and retrotransposition (*Pgk2, Aldoart1, Aldoart2*). A genomic approach identified all retroposed sequences matching glycolytic enzymes in the human and mouse genomes. Each glycolytic enzyme is encoded by a family of genes, and there is frequent retrotransposition of a single gene in each family. The same orthologous gene is independently retroposed in both species. Results from this study identified an alternative form of *Gpi1* transcribed during mouse spermatogenesis. Annotation and expression analysis of all glycolytic enzymes expressed during spermatogenesis will help to understand the regulation of energy metabolism. Since glycolysis is required for sperm motility and male fertility, spermatogenic-cell specific glycolytic enzymes are potential targets for contraceptives. Determining whether defects in the glycolytic pathway are a significant cause of male infertility is also an important clinical need.

ACKNOWLEDGMENTS

I would like to thank:

My thesis advisor, Debbie O'Brien

My parents Gangadhar and Sunita Vemuganti

My husband William for all of his love, support and advice

TABLE OF CONTENTS

LIST OF TABLES

Table

Supplemental Table

LIST OF FIGURES

Figure

Supplemental Figure

ABBREVIATIONS

α-GDH	α-glycerophosphate dehydrogenase
ACTB	Actin–β
AKAP	A kinase (PKA) anchor protein
ALDO	Aldolase
ALDOA	Aldolase A
ALDOA_V2	Aldolase A splice variant 2
ALDOART1	Aldolase A retrogene 1
ALDOART2	Aldolase A retrogene 2
ALDOB	Aldolase B
ALDOC	Aldolase C
ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosomes
BCA	Bicinchoninic acid
BLAST	Basic local alignment search tool
cAMP	Cyclic adenosine monophosphate
CCCP	Carbonyl cyanide m-chlorophenylhydrazone
cDNA	Complementary deoxyribonucleic acid
CS	Condensing spermatids
CYT Ct	Cytochrome Ct
dCTP	2'-deoxycytidine 5'-triphosphate
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylene diamine tetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
ENO	Enolase
EST	Expressed sequence tag
ExPASy	Expert Protein Analysis System
F1P	Fructose-1-phosphate
FBP	Fructose-1,6-bisphosphate
FS	Fibrous sheath
G3P	Glyceraldehyde-3-phosphate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GAPDHS	Glyceraldehyde-3-phosphate dehydrogenase-S
GCK	Glucokinase
GLUT	Glucose transporter

GPI1	Glucose phosphate isomerase
GST	Glutathion-s-transferase
HCl	Hydrochloric acid
HEM1	HEPES, EDTA, β-mercaptoethanol containing buffer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HK	Hexokinase
HK1-S	Hexokinase 1 spermatogenic cell-specific
HKDC	Hexokinase domain containing
HPLC	High performance liquid chromatography
Hs	*Homo sapiens*
IgG	Immunoglobulin G
IPTG	Isopropyl-beta-D-thiogalactopyranoside
KCl	Potassium chloride
KH_2PO_4	Potassium dihydrogen phosphate
LB	Luria-Bertani broth
LC	Liquid chromatography
LDH	Lactate dehydrogenase
LDHA	Lactate dehydrogenase A
LDHA_V2	Lactate dehydrogenase varient 2
LDHC	Lactate dehydrogenase C
LINE	Long interspersed nucleotide element
LTR	Long terminal repeat
MALDI-MS	Matrix-assisted laser desorption ionization-mass spectrometry
Mf	*Macaca mulatta*
Mm	*Mus musculus*
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
Na_2HPO_4	Sodium phosphate
NaCl	Sodium chloride
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (reduced form)
NCBI	National Center for Biotechnology Information
ODF	Outer dense fiber
ORF	Open reading frame
P^{32}	Phosphate-32
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFK	Phosphofructokinase
PFKL	Phosphofructokinase liver form
PFKM	Phosphofructokinase muscle form

PFKP	Phosphofructokinase platelet form
PGAM	Phosphoglycerate mutase
PGK	Phosphoglycerate kinase
PHYLIP	Phylogeny inference package
PI	Protease inhibitors
PK	Protein kinase A
PKA	Pyruvate kinase
PKLR	Pyruvate kinase liver and red blood cell form
PKM2	Pyruvate kinase muscle form
PMSF	Phenylmethylsulphonyl fluoride
Pn	*Pan troglodytes*
ppm	Parts per million
PRM1	Protamine 1
PS	Pachytene spermatocytes
PVDF	Polyvinylidene fluoride
Rn	*Rattus norvegicus*
RNA	Ribonucleic acid
RS	Round spermatids
RT-PCR	Reverse transcriptase-polymerase chain reaction
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEM	Standard error of the mean
SINE	Short interspersed nucleotide element
SSCP	Single strand conformation polymorphism
SSR	Sperm-specific region
TBS	Tris(hydroxymethyl)aminomethane buffered saline
TBS-T	Tris(hydroxymethyl)aminomethane buffered saline-tween
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoric acid
TPI	Triose phosphate isomerase
Tris	Tris(hydroxymethyl)aminomethane
TX-100	Triton X-100
UTR	Untranslated region

CHAPTER I

INTRODUCTION

Sperm motility is essential for fertilization.

Fertilization is a complex process dependent upon the proper completion of many events, including the delivery of the male genome in sperm to the egg. Before binding and penetrating the egg, sperm must travel through the female reproductive tract to the site of fertilization in the oviduct and penetrate the zona pellucida of the egg [1, 2]. Both motility and hyperactivated motility of sperm are essential for penetration of the zona pellucida and fertilization [3].

Rapid movement of sperm is propagated by the flagellum, a complex structure that is conserved across species and cell types [4]. The sperm flagellum consists of a connecting piece, middle piece, principal piece, and end piece (Figure 1.1). A “9+2” axoneme structure extends from the distal centriole reminant throughout the length of the flagellum [5]. The axoneme contains nine outer microtubule doublets and a central pair of microtubules. Each outer microtubule double is attached to the adjacent doublet and possesses dynein arms, which function as the motor of the sperm flagellum and facilitate movement. Dynein ATPases are localized along the entire sperm flagellum and are activated during sperm movement, resulting in sliding between adjacent outer microtubule doublets [4].

Axoneme structure is conserved in all cilia and flagella, but mammalian sperm flagella contain two accessory structures not found in other flagella: outer dense fibers and the fibrous sheath [4]. Nine outer dense fibers (ODFs) surround the outer microtubule doublets in the middle piece. The mitochondria are localized in the middle piece and are

arranged as a sheath structure encompassing the axoneme and ODFs. In the principal piece two ODFs (3 and 8) are replaced by the longitudinal columns of the fibrous sheath. The fibrous sheath, an insoluble cytoskeletal structure which defines the principal piece, consists of two longitudinal columns and connecting transverse ribs. Both the ODFs and fibrous sheath are thought to contribute to the rigid structure of the sperm flagellum, which helps to regulate movement along with activation of dynein ATPases associated with the microtubule doublets [5, 6].

Novel contraceptive approaches can target and inhibit sperm motility, thereby preventing zona pellucida penetration and fertilization. Most current contraceptive methods are hormonal and taken by women. Hormonal contraceptives are effective but carry with them undesirable side effects [7]. A non-hormonal male contraceptive that targets molecules found only in sperm would have fewer side effects. Therefore, proteins essential for sperm motility and expressed only during spermatogenesis are potential male contraceptive targets.

Conversely, impaired sperm motility can lead to male infertility [6, 8]. One study from 1085 patients with male factor infertility concluded that approximately 81% had defects in sperm motility, with 19% having no other defects in sperm morphology or count [9]. Proteomic profiling of sperm from 20 patients with defects in sperm motility compared to control donors showed differential expression of proteins localized to the flagellum, including cytoskeletal β-actin, ACTB [10]. In fact, multiple studies have identified mutations in genes encoding flagellar proteins in patients with low sperm motility. This includes mutations in dynein proteins, resulting in primary ciliary dyskinesia; tektin-t, an axonemal protein resulting in ultrastructural defects in the sperm tail; and AKAP3/AKAP4 deletions resulting in malformation (dysplasia) of the fibrous sheath and short tails [11-13]. Taken together, mutations in several genes encoding proteins localized to the sperm flagellum lead to decreased sperm motility and male infertility.

Glucose is required for hyperactivated motility and fertilization.

Spermatogenesis is the process by which germline stem cells differentiate into spermatozoa and can be divided into three phases: a proliferative mitotic phase that serves to increase cell number (spermatogonia), a meiotic phase when recombination occurs and number of chromosome sets are reduced (spermatocytes), and a haploid differentiation phase (spermiogenesis) producing highly specialized cells (round spermatids, condensing spermatids, and spermatozoa).

Cellular components are reorganized during spermiogenesis. Most of the cytoplasm is lost, the nucleus is condensed in the head, the acrosome forms and the flagellum forms. The flagellum contains the energy production system and machinery required for motility. The mitochondria are stacked into the middle piece of the flagellum and the glycolytic enzymes are localized to the principal piece (Figure 1.1). Therefore, ATP-generating metabolic pathways are compartmentalized in sperm, with oxidative phosphorylation occurring in the middle piece and glycolysis localized to the principal piece of the tail (Figure 1.1). Glycolysis converts glucose into two pyruvate molecules, with a net production of two ATP per glucose molecule (Figure 1.2). Pyruvate that is produced during glycolysis can be further metabolized in the mitochondria via the Krebs cycle and oxidative phosphorylation, or alternatively, converted to lactate in the cytoplasm.

Dynein ATPases located along the entire flagellum require ATP for sperm movement. It was originally thought that ATP produced via oxidative phosphorylation in the mitochondria was shuttled to the principal piece. However, blocking mitochondrial ATP production in mouse sperm by the use of the inhibitors oligomycin or carbonyl cyanide m-chlorophenylhydrazone (CCCP), or by targeted gene disruption of testis-specific cytochrome c (cyt c_t) demonstrate that oxidative phosphorylation is not required for motility and fertilization [3, 14, 15]. Together these results indicate that ATP generated via glycolysis is

required for sperm motility and fertilization in the mouse. The role of glycolysis in human sperm ATP production and motility is still unclear since inhibitors of both glycolysis and oxidative phosphorylation reduced ATP levels and motility [16-18]. Contradictory results suggest that human sperm motility is supported by substrates metabolized by oxidative phosphorylation, as well as glycolysable substrates [18-21]. It is clear that there are species differences in terms of the glycolytic and oxidative capacity of sperm. Two potential substrates of the glycolytic pathway include glucose and fructose.

Glucose is metabolized by hexokinase at the beginning of the glycolytic pathway (Figure 1.2). *In vitro* fertilization studies in mouse demonstrated that exogenous glucose, but not lactate or pyruvate, support hyperactivated sperm motility [3, 22, 23]. Similar studies using purified populations of human sperm without exogenous glucose resulted in a decrease in overall motility, a lack of hyperactivated motility, and reduced zona penetration [18, 24, 25].

Fructose can be metabolized either by hexokinase at the beginning of the glycolytic pathway (Figure 1.2), or by an aldolase-dependent three-step reaction (Figure 1.3). Hexokinase has a higher affinity for glucose than fructose, but will still bind and phosphorylate fructose [26]. Fructose-1,6-bisphosphate aldolase, or aldolase, is the fourth enzyme in the glycolytic pathway (Figure 1.2). When fructose is metabolized through an aldolase-dependent pathway, the cleavage product re-enters the glycolytic pathway as glyceraldehyde-3-phosphate, the substrate of glyceraldehyde-phosphate dehydrogenase (GAPDH). One of the enzymes required for aldolase-dependent metabolism of fructose (ketohexokinase) is expressed at low levels in mouse testis, while the there is no evidence for expression of the other enzyme (triose kinase) in testis or in sperm [27].

Fructose is metabolized by mouse sperm *in vitro*, but sperm exhibit sluggish motility after 6 hours of incubation [3, 28]. In human sperm, high levels of fructose in seminal plasma (0.016M in 10^8/ml sperm) are sufficient to support glycolytic ATP production [26].

Sperm from both normal and oligospermic patients metabolize fructose at the same rate, while sperm from necrospermic patients demonstrate lower levels of fructolysis [29]. Also, men with no sperm in their seminal plasma (nonobstructive azoospermia) have higher fructose levels in their semen, possibly due to a lack of sperm consuming the substrate [30].

Multiple hexose transporters, or glucose transporters (GLUTs), facilitate the entry of glucose and fructose into sperm for metabolism and are localized along the sperm membrane. GLUT3, the most abundant transporter, is found in both mouse and human sperm and has a high affinity for glucose [31, 32]. GLUT8, 9a, and 9b facilitate glucose transport in mouse sperm and testis, with each isoform demonstrating unique localization in sperm [33]. Only one transporter, GLUT5, has a high affinity for fructose and has been detected in human sperm [34].

Taken together, it is clear that glucose is metabolized by mouse sperm and supports motility required for fertilization [3]. Fructose is also metabolized by mouse sperm, resulting in slower motility which does not support all of the events required for fertilization [3, 22, 28]. In human sperm, both glucose and fructose support motility and high ATP levels [18]. There is a great deal of evidence supporting a critical role for glucose with *in vitro* fertilization using human sperm, but the role of fructose has not been analyzed as extensively [18, 24, 25].

Spermatogenic cell-specific glycolytic enzymes are required for sperm motility and fertilization.

The glycolytic pathway is modified substantially in mammalian sperm and includes multiple isozymes with novel structural or enzymatic properties. Glyceraldehyde 3-phosphate dehydrogenase-S (GAPDHS), phosphoglycerate kinase-2 (PGK2), and lactate dehydrogenase-C (LDHC) are three glycolytic isozymes encoded by genes expressed only during spermatogenesis [35-37]. The gene encoding GAPDHS was first identified through screening of a mouse spermatogenic cell expression library, and *Gapdhs* expression is first

detected in haploid cells [37, 38]. *Pgk2* is a well-characterized retrogene that is first expressed during meiotic prophase, although translation of the *Pgk2* transcript is delayed until spermiogenesis [35, 39-42]. *Ldhc* (formerly *Ldh-x*) is first transcribed in preleptotene spermatocytes, coinciding with the detection of LDHC activity [43-45].

Mechanisms of gene creation include retrotransposition and gene duplication. Spermatogenic-cell specific isozymes in the glycolytic pathway provide examples of both mechanisms. Gene duplication is responsible for *Gapdhs* and *Ldhc*, while *Pgk2* was retroposed from the X-linked, ubiquitously expressed *Pgk1* [35, 38, 39, 46]. Other examples of testis-specific retrogenes include, polyA binding protein 2, pyruvate dehydrogenase E1α2, TATA-binding protein and associated factor $TAF_{II}250$, and glucose-6-phosphate dehydrogenase [47-50].

Other glycolytic enzymes with restricted expression during spermatogenesis exhibit modified structures and/or functions, including hexokinase-1 (HK1), glucose phosphate isomerase (GPI), triose phosphate isomerase (TPI), phosphoglycerate mutase (PGAM), and enolase (ENO) [51-55]. Three splice variants of *Hk1* expressed during spermatogenesis possess unique 5' sequence encoding a sperm-specific region (SSR) [54, 56]. A unique species of GPI in mouse testis and sperm was detected by cellulose acetate electrophoresis coupled with glucose phosphate isomerase activity staining [51]. In rat, a larger *Tpi* transcript is expressed in haploid cells during spermatogenesis [55]. Sperm-specific forms of ENO were detected using cellulose acetate electrophoresis and activity staining [52, 57]. Modification of the glycolytic pathway through the expression of novel isozymes, combined with the requirement of glucose for sperm motility, highlights an essential role for proper regulation of glycolysis in sperm.

Sperm motility is dependent upon the production of high levels of ATP in the flagellum [15, 18, 26]. Glycolysis, rather than oxidative phosphorylation, is required for sperm motility and male fertility in mice [14, 58]. Early studies developing reversible

contraceptives identified chlorinated compounds as inhibitors of GAPDHS function, resulting in reduced glycolysis and motility [59]. Targeted disruption of *Gapdhs* resulted in marked reduction in ATP levels, the absence of progressive sperm motility and male infertility in mice [58]. Disruption of the genes for *Pgk2* and *Ldhc* produced sperm with lower ATP levels and reduced motility [60, 61]. Gene disruption and inhibitor studies collectively demonstrate a critical role for spermatogenic-specific glycolytic enzymes in male fertility.

Many sperm-specific glycolytic enzymes are tightly bound to the fibrous sheath.

Enzymes in the glycolytic pathway are localized to the principal piece of the sperm flagellum and several are tightly bound to the fibrous sheath, a cytoskeletal structure that defines the limits of the principal piece (Figure 1.1). Glycolytic enzymes resistant to urea extraction and tightly bound to the fibrous sheath include GAPDHS, aldolase A (ALDOA), and LDHA [62]. Localization of glycolytic enzymes provides ATP production along the length of the principal piece, which makes up approximately 2/3 of the entire tail [4]. A similar pattern of localization is seen in the green algae *Chlamydomonas reinhardtii,* where multiple glycolytic enzymes are targeted to the flagellar compartment [63, 64]. Here, as in sperm, ATP is supplied along the length of the flagellum through glycolysis.

Tight binding of glycolytic enzymes to the fibrous sheath suggests the formation of a multi-enzyme complex, which may result in higher catalytic activity. Such complexes exist in nature (DNA processing, RNA/protein synthesis, pyruvate dehydrogenase complex, and the electron transport chain), and direct binding between glycolytic enzymes has been suggested [65-67]. One example includes localized glycolytic ATP production in the *Drosophila* flight muscle [68, 69]. In flies lacking glycerol-3-phosphate dehydrogenase, localization of ALDOA and GAPDH to the Z-discs and M-lines is lost, along with the ability to fly, supporting a correlation between localization and function [68, 69]. It is possible that the

close nature of glycolytic enzymes in the sperm flagellum may help “shuttle” substrates through the pathway in a more efficient manner.

The most abundant proteins in the fibrous sheath are cAMP-dependent protein kinase anchoring proteins, or AKAPs [6]. AKAPs serve as a scaffold to anchor proteins to the fibrous sheath, including cAMP protein kinase (PKA) subunits and components of the Rho signaling pathway [6]. However, there has been no evidence to date supporting a direct interaction between glycolytic enzymes and AKAPs found in the fibrous sheath. Modification of proteins through the addition of novel amino acid residues, such as the N-terminal extension of GAPDHS, may help target and bind proteins to the fibrous sheath [37]. *Trypanosome Brucei* aldolase A contains a ten residue N-terminal extension that is hypothesized to direct localization of this glycolytic enzyme to a structure termed the glycosome [70]. A larger form of ALDOA is tightly bound to the mouse fibrous sheath, while a smaller form migrating at the expected molecular weight for ALDOA is lost during purification [62]. It is possible that spermatogenic-cell specific glycolytic enzymes, including ALDOA, are modified through the addition of N-terminal extensions to promote tight binding to the fibrous sheath where glycolytic ATP is produced.

Research presented in this dissertation

The overall goal of this dissertation is to identify and characterize the expression of novel spermatogenic-cell specific glycolytic enzymes in order to better understand metabolism and the regulation of glycolysis in sperm. The work presented in chapter two details the identification of all aldolase isoforms expressed during mouse spermatogenesis using genomic, molecular and proteomic methods. Three spermatogenic-cell specific aldolase transcripts are generated through retrotransposition and alternative splicing of the *Aldoa* gene. Unique nucleotide sequence in two spermatogenic-cell specific *Aldoa* transcripts encode N-terminal extensions hypothesized to be important for localization of

ALDOA to the fibrous sheath, the site of glycolytic ATP production. Chapter three details the kinetic properties of aldolase in mouse sperm and the quaternary structure of sperm-specific aldolase proteins expressed in *E. coli*. Preliminary three-dimensional modeling draws attention to unique structural differences in the outer face and near the substrate binding pocket of the isoforms, suggesting unique binding and catalytic properties. Results from the aldolase study supported the use of a similar genomic approach to identify all retroposed sequences matching glycolytic enzymes in the mouse and human genomes. Chapter four summarizes the genomic analysis of the glycolytic pathway, including characterization of gene structures in order to identify sequences with novel features, and detailed expression analysis of candidate retrogenes in mouse and human testes. These studies provide evidence for the frequent retrotransposition of orthologous genes encoding glycolytic enzymes in the mouse and human genome. Identification of all glycolytic enzymes expressed during spermatogenesis is important for understanding sperm metabolism, developing new contraceptive strategies, and discovering mutations in glycolytic enzymes that impair sperm motility and male fertility.

Figure 1.1

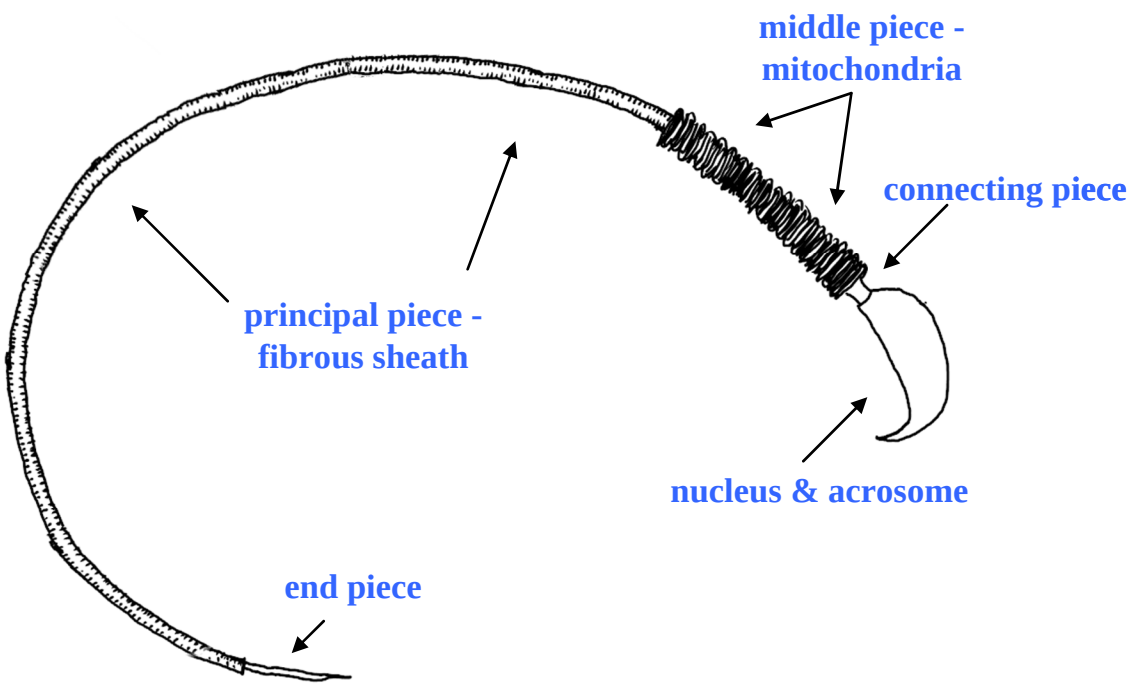


Figure 1.1. Mammalian sperm structure and compartmentalization of metabolic pathways. Sperm possess unique cellular structures, including a head and tail, or flagellum. The nucleus and acrosome are localized in the head. The tail is divided into three sections: the middle piece, the principal piece and the end piece. Metabolism is compartmentalized in the sperm tail. Mitochondria are compacted in the middle piece, and here oxidative phosphorylation occurs. The principal piece is defined by a highly insoluble cytoskeletal structure, the fibrous sheath. Glycolytic enzymes are localized to the principal piece. (Adapted from [71]).

Figure 1.2

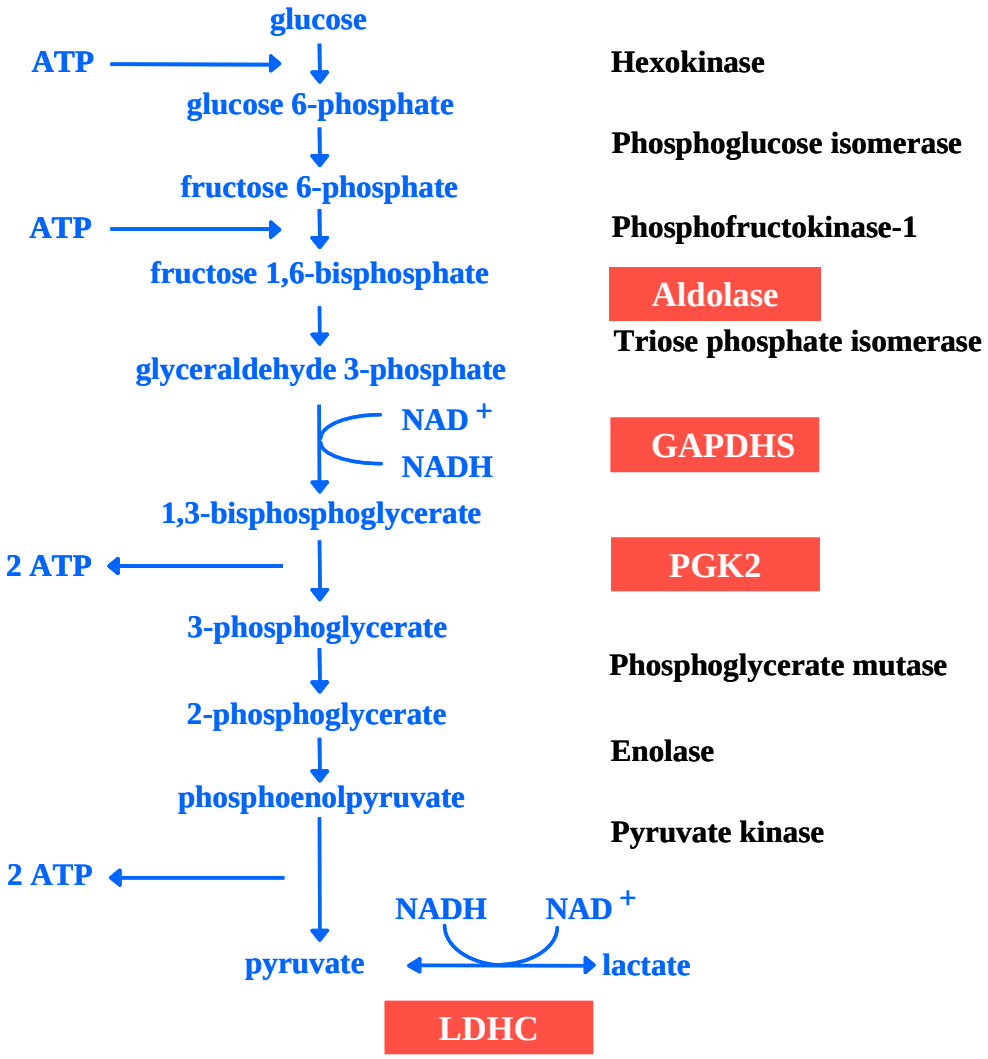


Figure 1.2. The glycolytic pathway is modified in mature sperm. Glucose is metabolized by ten sequential enzymatic reactions to yield 2 ATP and pyruvate, which can then enter the citric acid cycle. Enzymes shown in red are encoded by genes that are expressed only during spermatogenesis.

Figure 1.3

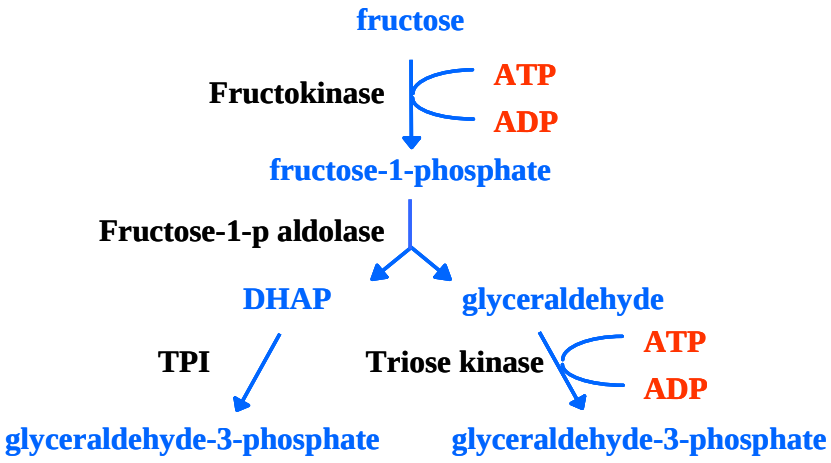


Figure 1.3. Fructose metabolism occurs through an aldolase-dependent pathway. Fructose is metabolized by fructokinase to produce fructose-1-phosphate, the substrate for fructose-1-phosphate aldolase. The final product, glyceraldehyde-3-phosphate, enters the glycolytic pathway and is metabolized by GAPDH. TPI is triose phosphate isomerase and DHAP is dihydroxyacetone phosphate.

References

1. Stauss CR, Votta TJ, Suarez SS: **Sperm motility hyperactivation facilitates penetration of the hamster zona pellucida**. *Biol Reprod* 1995, **53**(6):1280-1285.

2. Suarez SS: **Regulation of sperm storage and movement in the mammalian oviduct**. *Int J Dev Biol* 2008, **52**(5-6):455-462.

3. Fraser LR, Quinn PJ: **A glycolytic product is obligatory for initiation of the sperm acrosome reaction and whiplash motility required for fertilization in the mouse**. *J Reprod Fertil* 1981, **61**(1):25-35.

4. Turner RM: **Tales from the tail: what do we really know about sperm motility?** *J Androl* 2003, **24**(6):790-803.

5. Fawcett DW: **The mammalian spermatozoon**. *Dev Biol* 1975, **44**(2):394-436.

6. Eddy EM, Toshimori K, O'Brien DA: **Fibrous sheath of mammalian spermatozoa**. *Microsc Res Tech* 2003, **61**(1):103-115.

7. Rosenfeld JA, Zahorik PM, Saint W, Murphy G: **Women's satisfaction with birth control**. *J Fam Pract* 1993, **36**(2):169-173.

8. Boyle CA, Khoury MJ, Katz DF, Annest JL, Kresnow MJ, DeStefano F, Schrader SM: **The relation of computer-based measures of sperm morphology and motility to male infertility**. *Epidemiology* 1992, **3**(3):239-246.

9. Curi SM, Ariagno JI, Chenlo PH, Mendeluk GR, Pugliese MN, Sardi Segovia LM, Repetto HE, Blanco AM: **Asthenozoospermia: analysis of a large population**. *Arch Androl* 2003, **49**(5):343-349.

10. Martinez-Heredia J, de Mateo S, Vidal-Taboada JM, Ballesca JL, Oliva R: **Identification of proteomic differences in asthenozoospermic sperm samples**. *Hum Reprod* 2008, **23**(4):783-791.

11. Schwabe GC, Hoffmann K, Loges NT, Birker D, Rossier C, de Santi MM, Olbrich H, Fliegauf M, Failly M, Liebers U *et al*: **Primary ciliary dyskinesia associated with normal axoneme ultrastructure is caused by DNAH11 mutations**. *Hum Mutat* 2008, **29**(2):289-298.

12. Zuccarello D, Ferlin A, Cazzadore C, Pepe A, Garolla A, Moretti A, Cordeschi G, Francavilla S, Foresta C: **Mutations in dynein genes in patients affected by isolated non-syndromic asthenozoospermia**. *Hum Reprod* 2008, **23**(8):1957-1962.

13. Baccetti B, Collodel G, Estenoz M, Manca D, Moretti E, Piomboni P: **Gene deletions in an infertile man with sperm fibrous sheath dysplasia**. *Hum Reprod* 2005, **20**(10):2790-2794.

14. Narisawa S, Hecht NB, Goldberg E, Boatright KM, Reed JC, Millan JL: **Testis-specific cytochrome *c*-null mice produce functional sperm but undergo early testicular atrophy**. *Mol Cell Biol* 2002, **22**(15):5554-5562.

15. Mukai C, Okuno M: **Glycolysis plays a major role for adenosine triphosphate supplementation in mouse sperm flagellar movement**. *Biol Reprod* 2004, **71**(2):540-547.

16. Bone W, Jones AR, Morin C, Nieschlag E, Cooper TG: **Susceptibility of glycolytic enzyme activity and motility of spermatozoa from rat, mouse, and human to inhibition by proven and putative chlorinated antifertility compounds in vitro**. *J Androl* 2001, **22**(3):464-470.

17. Ruiz-Pesini E, Lapena AC, Diez-Sanchez C, Perez-Martos A, Montoya J, Alvarez E, Diaz M, Urries A, Montoro L, Lopez-Perez MJ *et al*: **Human mtDNA haplogroups associated with high or reduced spermatozoa motility**. *Am J Hum Genet* 2000, **67**(3):682-696.

18. Williams AC, Ford WC: **The role of glucose in supporting motility and capacitation in human spermatozoa**. *J Androl* 2001, **22**(4):680-695.

19. Ford WC, Harrison A: **The role of oxidative phosphorylation in the generation of ATP in human spermatozoa**. *J Reprod Fertil* 1981, **63**(1):271-278.

20. Peterson RN, Freund M: **ATP synthesis and oxidative metabolism in human spermatozoa**. *Biol Reprod* 1970, **3**(1):47-54.

21. Suter D, Chow PY, Martin IC: **Maintenance of motility in human spermatozoa by energy derived through oxidative phosphorylation and addition of albumin**. *Biol Reprod* 1979, **20**(3):505-510.

22. Urner F, Sakkas D: **Glucose is not essential for the occurrence of sperm binding and zona pellucida-induced acrosome reaction in the mouse**. *Int J Androl* 1996, **19**(2):91-96.

23. Cooper TG: **The onset and maintenance of hyperactivated motility of spermatozoa in the mouse.** *Gamete Research* 1984, **9**:55-74.

24. Hoshi K, Tsukikawa S, Sato A: **Importance of Ca2+, K+ and glucose in the medium for sperm penetration through the human zona pellucida**. *Tohoku J Exp Med* 1991, **165**(2):99-104.

25. Rogers BJ, Perreault SD: **Importance of glycolysable substrates for in vitro capacitation of human spermatozoa**. *Biol Reprod* 1990, **43**(6):1064-1069.

26. Peterson RN, Freund M: **Glycolysis by washed suspensions of human spermatozoa. Effect of substrate, substrate concentration, and changes in medium composition on the rate of glycolysis**. *Biol Reprod* 1969, **1**(3):238-246.

27. Funari VA, Crandall JE, Tolan DR: **Fructose metabolism in the cerebellum**. *Cerebellum* 2007, **6**(2):130-140.

28. Hoppe PC: **Glucose requirement for mouse sperm capacitation in vitro**. *Biol Reprod* 1976, **15**(1):39-45.

29. Mann T, R. J, Sherins R: **Fructolysis in human spermatozoa under normal and pathological conditions**. *Journal of Andrology* 1980, **1**(5):229-233.

30. Buckett WM, Lewis-Jones DI: **Fructose concentrations in seminal plasma from men with nonobstructive azoospermia**. *Arch Androl* 2002, **48**(1):23-27.

31. Haber RS, Weinstein SP, O'Boyle E, Morgello S: **Tissue distribution of the human GLUT3 glucose transporter**. *Endocrinology* 1993, **132**(6):2538-2543.

32. Urner F, Sakkas D: **A possible role for the pentose phosphate pathway of spermatozoa in gamete fusion in the mouse**. *Biol Reprod* 1999, **60**(3):733-739.

33. Kim ST, Moley KH: **The expression of GLUT8, GLUT9a, and GLUT9b in the mouse testis and sperm**. *Reprod Sci* 2007, **14**(5):445-455.

34. Burant CF, Takeda J, Brot-Laroche E, Bell GI, Davidson NO: **Fructose transporter in human spermatozoa and small intestine is GLUT5**. *J Biol Chem* 1992, **267**(21):14523-14526.

35. Boer PH, Adra CN, Lau YF, McBurney MW: **The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon**. *Mol Cell Biol* 1987, **7**(9):3107-3112.

36. Sakai I, Sharief FS, Li SS: **Molecular cloning and nucleotide sequence of the cDNA for sperm-specific lactate dehydrogenase-C from mouse**. *Biochem J* 1987, **242**(2):619-622.

37. Welch JE, Schatte EC, O'Brien DA, Eddy EM: **Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells**. *Biol Reprod* 1992, **46**(5):869-878.

38. Welch JE, Brown PR, O'Brien DA, Eddy EM: **Genomic organization of a mouse glyceraldehyde 3-phosphate dehydrogenase gene (Gapd-s) expressed in post-meiotic spermatogenic cells**. *Dev Genet* 1995, **16**(2):179-189.

39. McCarrey JR, Thomas K: **Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene**. *Nature* 1987, **326**(6112):501-505.

40. Kramer JM: **Immunofluorescent localization of PGK-1 and PGK-2 isozymes within specific cells of the mouse testis**. *Dev Biol* 1981, **87**(1):30-36.

41. Vandeberg JL, Lee CY, Goldberg E: **Immunohistochemical localization of phosphoglycerate kinase isozymes in mouse testes**. *J Exp Zool* 1981, **217**(3):435-441.

42. Bluthmann H, Cicurel L, Kuntz GW, Haedenkamp G, Illmensee K: **Immunohistochemical localization of mouse testis-specific phosphoglycerate kinase (PGK-2) by monoclonal antibodies**. *Embo J* 1982, **1**(4):479-484.

43. Hintz M, Goldberg E: **Immunohistochemical localization of LDH-x during spermatogenesis in mouse testes**. *Dev Biol* 1977, **57**(2):375-384.

44. Li SS, O'Brien DA, Hou EW, Versola J, Rockett DL, Eddy EM: **Differential activity and synthesis of lactate dehydrogenase isozymes A (muscle), B (heart), and C (testis) in mouse spermatogenic cells**. *Biol Reprod* 1989, **40**(1):173-180.

45. Alcivar AA, Trasler JM, Hake LE, Salehi-Ashtiani K, Goldberg E, Hecht NB: **DNA methylation and expression of the genes coding for lactate dehydrogenases A and C during rodent spermatogenesis**. *Biol Reprod* 1991, **44**(3):527-535.

46. Kao FT, Wu KC, Law ML, Hartz JA, Lau YF: **Assignment of human gene encoding testis-specific lactate dehydrogenase C to chromosome 11, region p14.3-p15.5**. *Somat Cell Mol Genet* 1988, **14**(5):515-518.

47. Kleene KC, Mulligan E, Steiger D, Donohue K, Mastrangelo MA: **The mouse gene encoding the testis-specific isoform of Poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes**. *J Mol Evol* 1998, **47**(3):275-281.

48. Wang PJ, Page DC: **Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis**. *Hum Mol Genet* 2002, **11**(19):2341-2346.

49. Hendriksen PJ, Hoogerbrugge JW, Baarends WM, de Boer P, Vreeburg JT, Vos EA, van der Lende T, Grootegoed JA: **Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse**. *Genomics* 1997, **41**(3):350-359.

50. Fitzgerald J, Hutchison WM, Dahl HH: **Isolation and characterisation of the mouse pyruvate dehydrogenase E1 alpha genes**. *Biochim Biophys Acta* 1992, **1131**(1):83-90.

51. Buehr M, McLaren A: **An electrophoretically detectable modification of glucosephosphate isomerase in mouse spermatozoa**. *J Reprod Fertil* 1981, **63**(1):169-173.

52. Edwards YH, Grootegoed JA: **A sperm-specific enolase**. *J Reprod Fertil* 1983, **68**(2):305-310.

53. Fundele R, Winking H, Illmensee K, Jagerbauer EM: **Developmental activation of phosphoglycerate mutase-2 in the testis of the mouse**. *Dev Biol* 1987, **124**(2):562-566.

54. Mori C, Welch JE, Fulcher KD, O'Brien DA, Eddy EM: **Unique hexokinase messenger ribonucleic acids lacking the porin-binding domain are developmentally expressed in mouse spermatogenic cells**. *Biol Reprod* 1993, **49**(2):191-203.

55. Russell DL, Kim KH: **Expression of triosephosphate isomerase transcripts in rat testis: evidence for retinol regulation and a novel germ cell transcript**. *Biol Reprod* 1996, **55**(1):11-18.

56. Mori C, Nakamura N, Welch JE, Gotoh H, Goulding EH, Fujioka M, Eddy EM: **Mouse spermatogenic cell-specific type 1 hexokinase (*mHk1-s*) transcripts are expressed by alternative splicing from the *mHk1* gene and the HK1-S protein is localized mainly in the sperm tail**. *Mol Reprod Dev* 1998, **49**(4):374-385.

57. Force A, Viallard JL, Saez F, Grizard G, Boucher D: **Electrophoretic characterization of the human sperm-specific enolase at different stages of maturation**. *J Androl* 2004, **25**(5):824-829.

58. Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility**. *Proc Natl Acad Sci U S A* 2004, **101**(47):16501-16506.

59. Jones AR, Cooper TG: **A re-appraisal of the post-testicular action and toxicity of chlorinated antifertility compounds**. *Int J Androl* 1999, **22**(3):130-138.

60. Danshina PV: **Male fertility and sperm function are severely impaired in mice lacking phosphoglycerate kinase-2.** *Journal of Andrology* 2006, **27**(Suppl):35.

61. Odet F, Duan C, Willis WD, Goulding EH, Kung A, Eddy EM, Goldberg E: **Expression of the gene for mouse lactate dehydrogenase C (Ldhc) is required for male fertility**. *Biol Reprod* 2008, **79**(1):26-34.

62. Krisfalusi M, Miki K, Magyar PL, O'Brien D A: **Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa**. *Biol Reprod* 2006, **75**(2):270-278.

63. Pazour GJ, Agrin N, Leszyk J, Witman GB: **Proteomic analysis of a eukaryotic cilium**. *J Cell Biol* 2005, **170**(1):103-113.

64. Mitchell BF, Pedersen LB, Feely M, Rosenbaum JL, Mitchell DR: **ATP production in Chlamydomonas reinhardtii flagella by glycolytic enzymes**. *Mol Biol Cell* 2005, **16**(10):4509-4518.

65. Vertessy BG, Orosz F, Kovacs J, Ovadi J: **Alternative binding of two sequential glycolytic enzymes to microtubules. Molecular studies in the**

phosphofructokinase/aldolase/microtubule system. *J Biol Chem* 1997, **272**(41):25542-25546.

66. Ouporov IV, Knull HR, Huber A, Thomasson KA: **Brownian dynamics simulations of aldolase binding glyceraldehyde 3-phosphate dehydrogenase and the possibility of substrate channeling**. *Biophys J* 2001, **80**(6):2527-2535.

67. Westhoff D, Kamp G: **Glyceraldehyde 3-phosphate dehydrogenase is bound to the fibrous sheath of mammalian spermatozoa**. *J Cell Sci* 1997, **110**(15):1821-1829.

68. Sullivan DT, MacIntyre R, Fuda N, Fiori J, Barrilla J, Ramizel L: **Analysis of glycolytic enzyme co-localization in *Drosophila* flight muscle**. *J Exp Biol* 2003, **206**(Pt 12):2031-2038.

69. Wojtas K, Slepecky N, von Kalm L, Sullivan D: **Flight muscle function in *Drosophila* requires colocalization of glycolytic enzymes**. *Mol Biol Cell* 1997, **8**(9):1665-1675.

70. Clayton CE: **Structure and regulated expression of genes encoding fructose biphosphate aldolase in Trypanosoma brucei**. *Embo J* 1985, **4**(11):2997-3003.

71. Bloom W, Fawcett DW: **A Textbook of Histology**, Tenth edn. Philadelphia, London, Toronto: W.B. Saunders Company; 1975.

CHAPTER 2

THREE MALE GERMLINE-SPECIFIC ALDOLASE A ISOZYMES ARE GENERATED BY ALTERNATIVE SPLICING AND RETROTRANSPOSITION

Abstract

Enzymes in the glycolytic pathway of mammalian sperm are modified extensively and are localized in the flagellum, where several are tightly bound to the fibrous sheath. This study provides the first evidence for three novel aldolase isozymes in mouse sperm, two encoded by *Aldoart1* and *Aldoart2* retrogenes on different chromosomes and another by *Aldoa_v2*, a splice variant of *Aldoa*. Phylogenetic analyses and comparative genomics indicate that the retrogenes and splice variant have remained functional and have been under positive selection for millions of years. Their expression is restricted to the male germline and is tightly regulated at both transcriptional and translational levels. All three isozymes are present only in spermatids and sperm and have distinctive features that may be important for localization in the flagellum and/or altered metabolic regulation. Both ALDOART1 and ALDOA_V2 have unusual N-terminal extensions not found in other aldolases. The N-terminal extension of ALDOA_V2 is highly conserved in mammals, suggesting a conserved function in sperm. We hypothesize that the N-terminal extensions are responsible for localizing components of the glycolytic pathway to the fibrous sheath and that this localization is required to provide sufficient ATP along the length of the flagellum to support sperm motility.

Introduction

Mammalian spermatogenesis includes three phases: a mitotic proliferation period that expands the number of spermatogonia, a prolonged meiotic prophase allowing spermatocytes to undergo recombination followed by two meiotic divisions, and a post-meiotic period where haploid spermatids differentiate into highly polarized sperm that are specialized to achieve fertilization. The program of gene expression that directs this developmental process has several distinct features and produces a large number of transcripts that are restricted to spermatogenic cells [1-3]. Microarray analyses estimate that at least 4% of the mouse genome produces male germ cell-specific transcripts, predominantly during the meiotic and post-meiotic phases of spermatogenesis [4-6]. Essential processes that occur during these two phases include the generation and subsequent packaging of the haploid genome within an extremely condensed nucleus, formation of specialized sperm structures such as the acrosome and flagellum, organization of surface domains essential for fertilization, and the development of complex signaling and metabolic cascades that regulate sperm function and gamete interaction. Multiple mechanisms contribute to this extensive diversification of gene function including gene duplication, retrotransposition and alternative splicing. Many conserved pathways, as well as germ cell-specific structures and processes, have unique variants with restricted expression during spermatogenesis.

Glycolysis is an important conserved pathway that has been modified substantially during mammalian spermatogenesis. This central metabolic pathway is composed of ten enzymes that convert glucose to pyruvate, with a net production of 2 ATPs per glucose molecule. Pyruvate can be further metabolized in the mitochondria via the Krebs cycle and oxidative phosphorylation, or alternatively converted to lactate by lactate dehydrogenase (LDH). Eight glycolytic enzymes and LDH have multiple isoforms with distinct patterns of expression in different tissues [7]. At least three of these gene families include members with

restricted expression in germ cells. Two germ cell-specific isozymes with distinct enzymatic properties, glyceraldehyde 3-phosphate dehydrogenase-S (GAPDHS) and lactate dehydrogenase C (LDHC), are encoded by intron-containing genes that arose by gene duplication. In the mouse *Gapdhs* is expressed only in spermatids [8, 9], while *Ldhc* is expressed in spermatogenic cells and oocytes [10-12]. Phosphoglycerate kinase-2 (*Pgk2*), which encodes another germ cell-specific isozyme, is an intronless gene that evolved by retrotransposition (i.e., a retrogene) from the *Pgk1* gene [13, 14]. Although *Pgk2* is transcribed in primary spermatocytes, the PGK2 protein is translated only in post-meiotic spermatids [15, 16].

Other glycolytic enzymes, including hexokinase 1 (Mori et al., 1993, 1998), phosphoglucose isomerase [17], phosphofructokinase 1 [18], aldolase [19] and enolase [20, 21], have unique structural or functional properties in spermatogenic cells and sperm. The molecular basis for these distinctive properties has not been determined in most cases. One exception is HK1-S, the hexokinase 1 isozyme that is derived from alternative splicing and lacks the porin-binding domain responsible for binding to mitochondria [22, 23].

Energy production in sperm is compartmentalized in distinct regions of the flagellum, with mitochondria and oxidative phosphorylation restricted to the middle piece and glycolysis localized in the longest segment known as the principal piece [24]. Although there are differences between species, mammalian sperm typically exhibit a high rate of glycolysis that is correlated with motility [25, 26]. Multiple studies indicate that glycolysis provides a significant proportion of ATP in both mouse and human sperm [27-29]. Gene targeting studies provide compelling evidence that glycolysis in spermatozoa [30] rather than mitochondrial ATP production [31] is essential for maintaining sperm motility and male fertility in the mouse. Distinctive features of sperm glycolytic enzymes may be important for localization in the principal piece and/or altered regulation of this key metabolic pathway.

Several sperm glycolytic enzymes are difficult to solubilize because they are tightly

bound to the fibrous sheath, a cytoskeletal structure that defines the limits of the principal piece in the sperm flagellum. We found that GAPDHS, aldolase A (ALDOA), lactate dehydrogenase A (LDHA), and pyruvate kinase remain attached to the fibrous sheath throughout a rigorous isolation procedure [8, 32]. GAPDHS is larger than other GAPDH family members and contains a novel proline-rich extension at the N-terminus [9, 33] that may mediate binding to the fibrous sheath. Our proteomic and immunoblot analyses identified two ALDOA bands in mouse sperm, with the larger 50,000 molecular weight band always present in isolated fibrous sheaths [32]. ALDOA, along with several other glycolytic enzymes, was also identified in fibrous sheaths isolated from human sperm [34]. Consistent with this localization, an earlier study found that 90% of aldolase activity in bovine sperm could not be solubilized with detergents [19]. That study also determined that sperm and muscle aldolases had distinct kinetic properties.

The fructose-1,6-bisphosphate aldolase gene family in vertebrates contains three well-characterized members: *Aldoa*, which is ubiquitously expressed with particularly high levels in muscle, aldolase B (*Aldob*) which is expressed at high levels in liver and kidney, and aldolase C (*Aldoc*) which is predominantly expressed in the nervous system [35-37]. All three of these isozymes have molecular weights of ~39,000, significantly smaller than the isoform detected in fibrous sheath. To determine the identity and origin of this larger ALDOA isoform, we used genomic, molecular and proteomic methods to examine all the aldolase variants expressed during mouse spermatogenesis and present in mature sperm. Our analyses identified three spermatogenic cell-specific aldolase isoforms in mouse, two encoded by retrogenes and a third resulting from alternative splicing of the *Aldoa* gene.

Materials and methods

Genomic analyses to identify Aldoa-related sequences

Ensembl Mouse BlastView (http://www.ensembl.org/Mus_musculus/blastview) was

used to search the mouse genome (NCBI database, mouse build 34) for sequences highly related to the mRNA sequence of mouse *Aldoa* (accession number NM_007438). Chromosomal regions containing significant matches were aligned to the mRNA sequence of *Aldoa* using ClustalW (http://www.ebi.ac.uk/clustalw/). All chromosome numbers included refer to Assembly: NCBIm36, December 2005; Genebuild: Ensembl, April 2006; Database version: 42.36c. For each sequence, we identified all insertions and deletions resulting in a shift in open reading frame (ORF) and internal stop codons. We assume that sequences containing such insertions/deletions or stop codons are pseudogenes that do not encode functional aldolase enzymes. Intronless sequences that contain a full-length ORF with conserved start and stop codons were classified as putative retrogenes. These included sequences found on chromosome 4 and chromosome 12, now termed *Aldoart1* and *Aldoart2*. The Mouse Genome Informatics database (http://www.informatics.jax.org/) currently identifies both of these regions as *Aldoa* pseudogenes. *Aldoart1* contains an extended ORF capable of encoding a protein with an additional 55 amino acids at the N-terminus. Further genomic comparisons identified a homologous potential coding sequence within the first intron of the *Aldoa* gene on chromosome 7, indicating the existence of an additional *Aldoa* exon. Transcripts that include this newly identified exon 2 would result in a splice variant (*Aldoa_v2*) that also encodes a larger protein with a novel N-terminus.

Phylogenetic analyses

Phylogenetic analyses were performed with two sets of sequences. The first set comprises nucleotide sequences for the ORFs of mouse *Aldoa*, *Aldob*, *Aldoc*, *Aldoart1*, *Aldoart2* and their orthologs in rat, human and chimpanzee. All these sequences were obtained from Ensembl. To identify orthologous sequences of the predicted mouse retrogenes, we used the *Aldoart1* and *Aldoart2* sequence as queries in BLAST searches in each of these additional species. Putative matches were then tested for conservation of

synteny with sequences flanking the mouse retrogenes and for presence of a conserved ORF. The second set of sequences includes the 165 nucleotides of mouse *Aldoart1* encoding 55 additional amino acids at the N-terminus, the 162 nucleotides of *Aldoa_v2* that encode a similar 54 amino acid sequence at the N-terminus, and the homologous sequences in rat, human, chimpanzee and macaque identified through BLAST searches in the appropriate species. Each set of sequences were aligned separately using ClustalW. All phylogenetic analyses were performed using the PHYLIP phylogeny inference software package, version 3.6 (http://evolution.genetics.washington.edu/phylip.html). For each set of sequences, we generated 100 bootstrapped datasets using the Seqboot program. We then determined the phylogeny of the aldolase gene family with a distance matrix method (Neighbor) using distances obtained with the Dnadist program, a maximum likelihood method without a molecular clock (Dnaml), and a maximum parsimony method (Dnapars). The Consense program was used to construct a majority-rule consensus tree.

EST database searches

Sequences for the full-length *Aldoa* mRNA, individual exons of this gene including the proposed alternative exon 2, *Aldoart1* and *Aldoart2* were used as queries in BLAST searches against the mouse EST database, using the NCBI nucleotide-nucleotide BLAST (blastn) software. The EST database includes libraries from several cell types isolated from testis including type A spermatogonia, type B spermatogonia, pachytene spermatocytes, round spermatids and Sertoli cells [38]. EST libraries are also available from cell lines produced from Leydig cells. To query specific EST libraries, searches were completed using either tissue type or testicular cell type as a limiting field (i.e., muscle, brain, testis or testicle, spermatogonia, spermatocyte and spermatid). Positive hits from libraries derived from pooled tissues were excluded. ESTs with very high levels of identity for substantial contiguous sequences (>98% identity for >100 bp) were scored as positive. Positive ESTs

were then blasted against the mouse genomic database using Ensembl Mouse BlastView, and the best genomic sequence match (with at least 98% identity) was used to identify the gene from which the EST was transcribed.

Tissue and cell isolations

Outbred CD-1 mice and Sprague-Dawley rats were obtained from Charles River (Raleigh, NC) and inbred strains (C57BL/6J and A/J) were obtained from the Jackson Laboratory (Bar Harbor, ME). All procedures involving animals were approved by the University of North Carolina at Chapel Hill Animal Care and Use Committee and conducted in accordance with the Guide for the Care and Use of Laboratory Animals (Institute for Laboratory Animal Research, National Academy of Sciences). Testes and other tissues were processed immediately or quick-frozen in liquid nitrogen and stored at -70^{o}C. Mixed germ cells were isolated by sequential enzymatic dissociation of testes from 8-day-old, 17-day-old, or adult mice [39]. Spermatogonia (92% purity, including type A through type B spermatogonia) were isolated from mixed germ cells from 8-day-old mice by differential plating [40]. Sertoli cells were isolated from 17-day-old mice as described previously [41]. Pachytene spermatocytes, round spermatids, and condensing spermatids were purified from adult mixed germ cell suspensions by unit gravity sedimentation [39]. Both pachytene spermatocytes and round spermatids (steps 1-8) had purities >90%. Condensing spermatids contained 30-40% nucleated cells (steps 9-16 spermatids) and cytoplasts derived primarily from these same cells.

Mouse and rat sperm were collected from the cauda epididymis following a 15 min incubation of clipped tissue at 37ºC in phosphate buffer saline with protease inhibitors (PBS + PI) containing 140 mM NaCl, 10 mM phosphate buffer (pH 7.4) and Complete protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). Cryopreserved human sperm samples from healthy donors were provided by the Andrology Laboratory, Department of

Obstetrics and Gynecology, University of North Carolina School of Medicine. These samples were washed twice with PBS to remove seminal plasma.

RT-PCR to detect transcription of Aldoa-related genes

Total RNA was isolated using Trizol (Invitrogen, Carlsbad, CA) from tissues or cells pooled from at least three mice, including brain, heart, ovary, kidney, spleen, liver, juvenile (8 - 34 days of age) and adult testis, and purified testicular cells. Genomic DNA contamination was removed from RNA preparations using the Qiagen RNeasy Midi Kit (Qiagen Incorporation, Valencia, CA). RNA was quantified using the NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). Equal loading and quality was confirmed using the RNA 6000 Nano Assay (Agilent Technologies, Waldbronn, Germany).

Reverse transcription followed by gene-specific polymerase chain reaction (cMaster RT_{plus}PCR System, Eppendorf, Hamburg, Germany) was used to amplify transcripts from total RNA samples. Primers were designed to amplify and incorporate α-[^{32}P]-dCTP into the same size product (530 base pairs) from *Aldoa* (including both the ubiquitously expressed *Aldoa_v1* transcript and the predicted *Aldoa_v2* transcript), *Aldoart1* and *Aldoart2*. The forward primer sequence was 5'AGAAGGCAGATGTGGACGTCC3' and the reverse primer sequence was 5'GGCACTGTGCGACGAAGTGCTGTGAC3'. Both primers are 100% identical to matching sequences in *Aldoa, Aldoa_v2, Aldoart1,* and *Aldoart2.* Filled triangles in Fig. 2.2.1 denote the location of these primers in each gene. Following PCR amplification with both primer sets, the products were digested with the *Hae*III restriction enzyme (New England BioLabs, Ipswich, MA). Unique *Hae*III restriction sites in each amplified product produced fragments with distinct sizes that were resolved by electrophoresis on 5% acrylamide gels at 95 watts for 2 h. Appropriate controls (a cDNA clone for *Aldoa,* and BAC clones for *Aldoart1* [RP24-191C1] (Invitrogen, Carlsbad, CA) and *Aldoart2* [RP23-63H11])

were used as template in parallel PCR reactions to confirm the expected digestion pattern. Gels were exposed to Super RX x-ray film (Fujifilm, Tokyo, Japan) using intensifying screens to detect incorporation of α-[^{32}P]-dCTP into amplified products.

For semi-quantitative analysis of RT-PCR products, we first determined the linear phase for PCR amplification by varying the amount of input RNA and the number of amplification cycles. Once conditions for the linear phase of amplification were established, RT-PCR was repeated as described above with RNAs (75 ng/reaction) from isolated testicular cells (pachytene spermatocytes, round spermatids, condensing spermatids and Sertoli cells) using triplicate reactions from two preparations of each cell type. For each reaction, incorporation of α-[^{32}P]-dCTP into *Hae*III-digested products was measured by densitometry using the Fluor-S MultiImager (Bio-Rad, Hercules, CA). Corrections were made for varying amounts of α-[^{32}P]-dCTP incorporated into each gene product due to nucleotide differences and for *Hae*III-digested PCR fragments that were outside the resolution range of our gels.

A second set of primers (open triangles in Fig. 2.2.1) was used to amplify and incorporate α-[^{32}P]-dCTP into products from the N-terminal regions of *Aldoa* and *Aldoart1*. The forward primer sequence for this reaction was 5'CCGCGTTCGCTCCTTAGT3' and the reverse primer sequence was 5'ATCTGCAGCCAGGATGCCC3'. Both primers are 100% identical to the matching sequences in the *Aldoa_v1* and *Aldoa_v2* transcripts, and will amplify different-sized RT-PCR products from each transcript. However, the forward primer contained one mismatched residue when amplifying *Aldoart1* (5' CCG/ACGTTCGCTCCTTAGT3'). Products from *Aldoa_v2* and *Aldoart1* can be distinguished following *Hae*III-digestion, since only the *Aldoa_v2* product is cleaved by this enzyme. [^{32}P]-labeled RT-PCR products were resolved on 5% polyacrylamide gels and detected by exposure to film using intensifying screens.

Reverse transcription of RNA from CD1 testis using an oligo d(T) primer generated a

testis cDNA sample used for full length subcloning of *Aldoa*, *Aldoa_v2*, *Aldoart1*, and *Aldoart2.* All four aldolase A genes were PCR amplified by gene-specific primer sets. The following primer pairs were using for PCR amplification of each *Aldoa*-related transcript: *Aldoa* (forward primer: 5'ACGAGGTTCTGGTGACCCTA3', reverse primer: 5'GTGATGGGAAAGAGCCTGAA3'), *Aldoa_v2* (forward primer: 5'GAATTCATGGCAACGCACAG3', reverse primer: 5'CTCGAGTTCAATAGCAAGTGG3'), *Aldoart1* (forward primer: 5'CGGAATTCATGGCAACGCACAGGCA3', reverse primer: 5'AGCGTCGACACATGAGGGCA3'), and *Aldoart2* (forward primer: 5'GAATTCATGTCTTACCCCTACC3', reverse primer: 5'CTCGAGACCTCTGCTCAGTA3'). Amplified products were subcloned into the pCR®4-TOPO® vector using the TOPO TA Cloning® Kit for Sequencing (Invitrogen, Carlsbad, CA). Positive clones were confirmed by *Eco*RI restriction digest and were directly sequenced using M13 forward and reverse primers by the UNC-CH Genome Analysis Facility. All three newly identified mouse aldolase A genes were deposited into GenBank using the following accession numbers: *Aldoa_v2* (EF662059), *Aldoart1* (EF662061), and *Aldoart2* (EF662062).

Immunoblotting of ALDOA-related proteins

Proteins were extracted from tissues or isolated cells by homogenization in lysis buffer (2% SDS, 100 mM DTT, 125 mM Tris pH 6.8, 18% glycerol). The lysates were centrifuged at 16,000 x g for 10 min at 4ºC. The protein concentration in the supernatants was determined using the micro BCA assay according to the manufacturer's instructions (Pierce Biotechnology, Rockford, IL). Samples with equal protein amounts were separated by SDS polyacrylamide gel electrophoresis (SDS-PAGE) on 10% polyacrylamide gels and electrophoretically transferred to Immobilon-P PVDF (polyvinylidene fluoride) membranes (Millipore Corp, Bedford, MA). Protein transfer and equal loading were confirmed by staining the membranes with 0.1% Coomassie blue R250 in 45% methanol, 10% acetic

acid. Membranes were destained, rinsed with TBS-T (140 mM NaCl, 3 mM KCl, 0.05% Tween-20, 25 mM Tris-HCl, pH 7.4) and incubated in blocking buffer (5% nonfat dry milk in TBS-T) overnight at 4ºC. Membranes were then incubated for 1 h at room temperature with a polyclonal antibody raised against purified rabbit skeletal muscle aldolase (Polysciences, Warrington, PA) diluted 1:2000 in blocking buffer. After three 15-min washes with TBS-T, the membranes were incubated for 20 min at room temperature with secondary antibody (affinity-purified horseradish peroxidase-conjugated rabbit anti-goat IgG, KPL, Gaithersburg, MD) diluted 1:10,000 in blocking buffer. TBS-T washes were repeated and immunoreactive proteins were detected by enhanced chemiluminescence using the SuperSignal West Pico substrate (Pierce Biotechnology, Rockford, IL) and HyBlot CL autoradiography film (Denville Scientific, Metuchen, NJ).

Proteomic analysis of mouse sperm proteins

Mouse sperm were sonicated to separate sperm heads and tails, and were further fractionated to obtain purified fibrous sheath preparations as described previously [32]. Briefly, fragmented sperm tails were pelleted at 32,500 x g after removing the sperm heads by low-speed centrifugation. Soluble sperm proteins released by sonication were precipitated with 10% trichloroacetic acid. Fibrous sheaths were isolated from the tail pellets by sequential incubations in Triton X-100 extraction buffer (1% Triton X-100, 2 mM dithiothreitol [DTT], 1 mM EDTA, 1 mM EGTA, 2 mM PMSF, 50 mM Tris-HCl, pH 9.0 and PI), potassium thiocyanate extraction buffer (0.6 M potassium thiocyanate, 2 mM DTT, 50 mM Tris-HCl, pH 8.0), and urea extraction buffer (6 M urea, 20 mM DTT, 50 mM Tris-HCl, pH 8.0). The final urea-insoluble pellet contained purified fibrous sheaths.

Proteins from the soluble, sperm tail and fibrous sheath fractions were resuspended in 2X SDS sample buffer (4% SDS, 18% glycerol, 125 mM Tris-HCl, pH 6.8) containing 50 mM TCEP reducing agent (Tris(2-carboxyethyl)phosphine, Pierce Biotechnology, Rockford,

II), heated at 95ºC for 10 min, and separated by SDS-PAGE using pre-cast 10% Tris-HCl Ready Gels (Bio-Rad, Hercules, CA). The Benchmark™ Protein ladder (Invitrogen, Carlsbad, CA) was also resolved on these gels to determine approximate molecular weights of unknown proteins. SDS gels were fixed for 1 h at room temperature with gentle agitation in 45% methanol, 10% acetic acid in deionized water. Fixed gels were stained with SYPRO Ruby (Molecular Probes, Eugene, OR) overnight at room temperature with gentle agitation, followed by two 30-min washes in 10% methanol, 7% acetic acid in deionized water. Gels were stored in SYPRO Ruby stain until processed.

Gel bands corresponding to aldolase proteins identified by immunoblotting were cut from gels and processed using an in-gel digestion protocol as described previously [42]. Peptides from the digested proteins were resuspended in 5% acetonitrile/95% water/0.1% trifluoroacetic acid (Buffer A) to facilitate liquid chromatography. Following digestion, the samples were subjected to MALDI-MS (Matrix-Assisted Laser Desorption Ionization-Mass Spectrometry) and MS/MS analysis using the Applied Biosystems 4700 Proteomics Analyzer (Applied Biosystems Incorporated, Foster City, CA). The spectra were then searched against the NCBI and MSDB databases using the MASCOT search engine. After protein identification, samples were analyzed by liquid chromatography (LC)-MALDI-MS. Briefly, peptides in the sample (5 µL) was separated on a reverse-phase C18 nano-HPLC column (Dionex Corp., Sunnyvale, CA) using an Agilent 1100 Capillary HPLC system (Agilent Technologies Inc. Santa Clara, CA) with a post-pump split to deliver 250 nL/min through a manual injection valve. The column was coupled to a Probot Microfraction Collector (Dionex Corp., Sunnyvale, CA). A gradient from 5-65% Buffer B (95% acetonitrile/0.09% TFA) was run and the column eluate was spotted along with an equivalent volume of 60% acetonitrile supplied by the sheath flow from the Probot syringe pump. After drying, the fractions were overlaid with a saturated solution of α-cyano-4-hydroxycinamic acid in 50% acetonitrile/20 mM ammonium citrate. Spectra were collected

and examined for the presence of aldolase isoform-specific peptide masses. Peptides within 100 ppm of predicted masses were subjected to MS/MS to confirm their identity.

Results

Genomic analyses predict two Aldoa-related retrogenes and a novel Aldoa splice variant

Ensembl BLAST searches with the coding sequence of the *Aldoa* gene on chromosome 7 detected related sequences on ten additional mouse chromosomes (Supplemental Table 2.1). These sequences lack introns and are consistent with a previous report of several *Aldoa* pseudogenes in mouse derived from multiple retrotransposition events [43]. In contrast, we did not detect retroposed sequences related to *Aldob* or *Aldoc* in comparable BLAST searches. Two of the *Aldoa*-related sequences have conserved full-length ORFs (including start and stop codons), suggesting that they may be functional retrogenes. These predicted retrogenes are located on chromosome 4 (*Aldoart1*; chromosome 4, minus strand, 72,337,671-72,339,071, Ensembl gene ID: ENSMUSG00000059343) and chromosome 12 (*Aldoart2*; chromosome 12, plus strand, 56,483,303-56,484-873, Ensembl gene ID: ENSMUSG00000063129). The start and end positions of the retrogenes were determined based on sequence alignments to *Aldoa*, combined with the identification of ESTs spanning both the 5'-UTR and 3'-UTR of each gene (see below).

Aldoart2 encodes a protein of the same size as ALDOA (Fig. 2.1, white boxes corresponding to homologous coding exons of *Aldoa*). The 5'- and 3'-UTRs of *Aldoart2* contain regions with homology to the 5'- and 3'-UTRs of *Aldoa* (boxes with diagonal lines), along with unique sequences (boxes with horizontal lines) that are proximal and distal to the *Aldoa*-homologous regions. *Aldoart1* is predicted to encode a larger protein because it contains sequence (black box, chromosome 4, minus strand, 72,338,786-72,338,943) that extends the ORF by 165 nucleotides at the N-terminus. This additional sequence encodes

a 55-amino acid N-terminal extension that shows no homology with any of the previously described aldolase isoforms. The remaining *Aldoart1* coding sequence (white boxes), 5'-UTR and initial segment of the 3'-UTR (boxes with diagonal lines) are homologous to *Aldoa.* Only the terminal portion of the *Aldoart1* 3'-UTR has unique sequence that is not homologous to *Aldoa* or *Aldoart2*.

The *Aldoart1* sequence that encodes the novel N-terminal extension is 90% identical to a region within the first intron of *Aldoa*, providing evidence that *Aldoa* contains a previously unidentified exon. As shown at the top of Figure 2.1, alternative splicing of *Aldoa* can generate both the ubiquitously transcribed mRNA with eight coding exons (*Aldoa_v1*) and a longer transcript (*Aldoa_v2*) from nine exons, including the newly identified exon 2 (154 nucleotides, chromosome 7, minus strand, 126,590,193-126,590,347). The first 21 nucleotides of exon 3 are part of the 5'UTR in *Aldoa_v1*, but are part of the coding sequence in *Aldoa_v2*. The ORF of *Aldoa_v2* contains an additional 162 nucleotides encoding a 54-amino acid N-terminal extension. There is an in-frame stop codon immediately upstream of the predicted start codon in both *Aldoa_v2* and *Aldoart1*.

Aldoart1 and Aldoart2 are derived from Aldoa and arose in the rodent lineage

Phylogenetic analyses of coding sequences conclusively demonstrate that *Aldoart1* and *Aldoart2* belong to the *Aldoa* subfamily (Fig. 2.2). Both *Aldoart1* and *Aldoart2* arose by retrotransposition in the rodent lineage after the primate-rodent split. All phylogenetic methods robustly predict that *Aldoart2* arose prior to the divergence of rat and mouse. This prediction was confirmed by the identification of the orthologous sequence for *Aldoart2* on rat chromosome 6 (plus strand, position 75,812,016 – 75,813,706). The rat *Aldoart2* has a conserved ORF with 94% identity to mouse *Aldoart2* (not shown), supporting the hypothesis that it is a functional retrogene in both species. It is uncertain whether *Aldoart1* arose before or after the rat-mouse split. Although phylogenetic analyses support a retrotransposition

event just before the divergence of these two lineages, we were unable to find evidence for a homologous sequence anywhere in the rat genome, including the region of conserved synteny (90.5 Mb on rat chromosome 5). We conclude that the ORF and critical amino acids of both retrogenes have been conserved for a minimum of 15 million years in the mouse lineage for *Aldoart1* and for a much longer period in both the mouse and rat lineages in the case of *Aldoart2*. However, the retrogenes have evolved with higher substitution rates than other aldolase genes in rat and mouse, as indicated by the relative length of the retrogene branches in the phylogenetic tree (Fig. 2.2).

Aldoa_v2 is conserved in multiple mammalian species

We also performed phylogenetic analyses with the nucleotide sequences encoding the novel N-terminal extensions of *Aldoa_v2* from mouse (Fig. 2.3A, black segments of exon 2 and 3) and four other mammalian species, and mouse *Aldoart1*. The tree (Fig. 2.3B) replicates the known divergence order of these species and supports the sequence of events depicted in Figure 2. The most remarkable feature of this tree is the striking difference in substitution rate in *Aldoart1* compared to *Aldoa_v2*. When amino acid sequences are compared (Fig. 2.3C), the N-terminal extensions of ALDOART1 and mouse ALDOA_V2 have 12 amino acid differences (78% identity). The retention of an ORF with numerous substitutions confirms that the evolution of *Aldoart1* is not simply the result of purifying selection, but most likely involves positive selection for novel functions.

The ALDOA-related isozymes have distinct amino acid sequences

Excluding the N-terminal extensions, the predicted amino acid sequences of ALDOART1 and ALDOART2 are 95.3% and 95.1% identical to ALDOA, respectively, while these isoforms are 92.3% identical to each other (Fig. 2.4). Active site residues [44] are conserved among all aldolase isoforms (Fig. 2.4, boxed residues). Both ALDOART1 and

ALDOART2 contain unique amino acids that are not found in any other ALDO-related proteins (indicated by white letters on a black background). When the predicted amino acid sequences of mouse and rat *Aldoart2* are aligned, eight of the unique amino acids are conserved, including Y3, D130, G133, Q140, K156, S246, P266, and E348. ALDOART1 and ALDOART2 also contain residues that are identical to ALDOB and/or ALDOC (residues highlighted in grey). In ALDOART1 there are three adjacent amino acids (MGN, amino acids 95-97) between the first two active site residues that are identical to the homologous sequence in ALDOB (amino acids 39-41), but are different from the homologous IAK sequence in ALDOA. This region contains two amino acids (I39 in ALDOA and G40 in ALDOB, arrows in Fig. 2.4) that have been identified as isozyme-specific residues based on sequence conservation [45]. ALDOART1 also contains F112 in its sequence (arrow), which aligns with another ALDOB-specific residue (F58 in ALDOB).

ESTs support the restricted expression of Aldoa-related retrogenes

We searched mouse EST databases in our initial assessment of whether the predicted *Aldoa*-related retrogenes and splice variant are transcribed (Table 2.1). As expected, *in silico* analyses detected abundant transcripts for *Aldoa* in muscle, *Aldob* in liver, and *Aldoc* in brain. Transcripts for *Aldoa*, but not *Aldob* or *Aldoc*, were detected in testis. Multiple *Aldoart1* and *Aldoart2* ESTs were identified, mainly in testis libraries. Only 1 of 10 *Aldoart1* ESTs was detected in a library of unknown origin and 3 of 33 *Aldoart2* ESTs were detected in a spleen library. However, this spleen library also contains a similar number of ESTs from *Prm1*, a male germ cell-specific transcript (accession numbers AA290514, AI509226, and AI661727). ESTs specific for the alternatively spliced exon 2 of *Aldoa_v2* were found in libraries from testis, retina, a mammary tumor cell line (RCB0527) and other carcinomas. Each EST was blasted against the mouse genome to confirm the gene from which it was transcribed (>98% identity for >100 bp). EST analyses indicate that

the *Aldoart2* ortholog in rat is also expressed in the testis.

EST libraries are available for germ cells isolated at distinct stages of spermatogenesis and for testicular somatic cells, including Sertoli cells from the seminiferous epithelium and cell lines derived from the interstitial Leydig cells that are responsible for testosterone production. *Aldoa* transcripts were detected in Leydig and Sertoli cells and at earlier stages of spermatogenesis, including type B spermatogonia and pachytene spermatocytes (Table 2.1). *Aldoart1* and *Aldoart2* transcripts were detected only in round spermatids, which undergo haploid differentiation following meiosis. Frequency analysis supports the abundant expression of *Aldoart1* and *Aldoart2* in round spermatids (Table 2.2) compared to other testicular cell types (Table 2.2) and other tissues (Table 2.3). These EST analyses provide support for the expression of *Aldoa*-related retrogenes and the *Aldoa_v2* splice variant, with restricted expression of the retrogenes primarily during the later stages of spermatogenesis.

ESTs preferentially span the 3' and 5' ends of transcripts. The 3' ends of the *Aldoart1* and *Aldoart2* transcripts and the 5' end of *Aldoart2* were determined directly from the alignment of multiple ESTs to the corresponding genomic sequences. Since the 5'-UTRs of *Aldoart1* and *Aldoa* are homologous, we predict that *Aldoart1* has a transcription start site that is comparable to *Aldoa*, although available 5'-derived ESTs do not contain the first 88 bp of the predicted 5'-UTR sequence.

RT-PCR confirms restricted transcription of Aldoa_v1, Aldoa_v2, Aldoart1 and Aldoart2 in the testis

Although there is high sequence similarity between *Aldoa, Aldoart1* and *Aldoart2,* we developed an RT-PCR strategy using a single primer pair followed by *Hae*III digestion to distinguish among transcripts from these genes (Fig. 2.5A, also see Materials and Methods). Following RT-PCR and *Hae*III digestion, we observed restriction fragments characteristic of

Aldoa transcripts (*Aldoa_v1*or *Aldoa_v2*) in all tissues examined (Fig. 2.5B). Two distinct cleavage products each from *Aldoart1* and *Aldoart2* mRNAs were seen only in testis. Products were not amplified when the RT enzyme was omitted, confirming that the observed products were not derived from contaminating genomic DNA (data not shown). Analysis of RNAs isolated from the testes of juvenile mice (middle panel, Fig. 2.5B) indicated that *Aldoart1* and *Aldoart2* transcripts are both expressed beginning at 20 days of age, coincident with the appearance of round spermatids in the developing testis. PCR products specific for each of these retrogenes were present in testis RNAs isolated throughout subsequent pubertal development between 22 and 34 days of age (not shown). The appearance of *Aldoart1* and *Aldoart2* transcripts at day 20 was confirmed by three independent PCR replicates.

Semi-quantitative RT-PCR using the same primer set (right panel, Fig. 2.5B) followed by densitometry of the resulting products (Fig. 2.5C) showed that *Aldoart2* mRNA levels were elevated compared to other *Aldoa*-related transcripts, both in testis and in isolated round spermatids (R) and condensing spermatids (C). *Aldoart1* mRNA was detected in both round and condensing spermatids, at levels comparable to the expression of *Aldoa* and/or *Aldoa_v2* transcripts (Fig. 2.5B, C). When analyzed in the same linear range of PCR amplification, the levels of all *Aldoa*-related transcripts were reduced or absent in pachytene spermatocytes (P) and Sertoli cells. Minor contamination of isolated pachytene spermatocytes with round spermatids may account for the *Aldoart2* signal detected in this cell population.

Expression of the novel N-terminal coding sequences in *Aldoa_v2* and *Aldoart1* was confirmed using a similar RT-PCR strategy to amplify regions near the 5' end of the *Aldoa* and *Aldoart1* transcripts (Fig. 2.6A). Primers used in these assays do not amplify a product from the *Aldoart2* transcript. *Hae*III restriction digestion cleaves only the *Aldoa_v2* RT-PCR product, generating distinct bands from *Aldoa_v1*, *Aldoa_v2* and *Aldoart1* transcripts that

are resolved by gel electrophoresis. This assay detected expression of both *Aldoa_v2* and *Aldoart1* transcripts only in testis (Fig. 2.6B). *Aldoart1* transcripts first appeared at 20 days of age (Fig. 2.6C), confirming our previous analyses (Fig. 2.5B). In contrast, *Aldoa_v2* transcripts were detected throughout postnatal development of the testis between 8 and 34 days of age (Fig. 2.6C), indicating that this splice variant is expressed during earlier stages of spermatogenesis.

Proteins encoded by Aldoa_v1, Aldoart1 and Aldoart2 are present in mouse sperm

Our initial proteomic and Western analyses identified two ALDOA bands, with apparent molecular weights of ~39,000 and 50,000, in fibrous sheaths isolated from mouse sperm [32]. Western blotting detected a ~39,000 molecular weight aldolase isoform in the soluble fraction following sonication of sperm (soluble, Fig. 2.7B) and in isolated sperm tails (tail P, Fig. 2.7B). Small amounts of this protein were sometimes detected in the fibrous sheath fraction [32]. In contrast, the 50,000 molecular weight aldolase band was consistently enriched in isolated fibrous sheaths (FS, Fig. 2.7A and B). Detailed proteomic analyses of these bands identified unique peptide sequences that match the three ALDOA-related isoforms expressed during spermatogenesis (Fig. 2.7A, Supplemental Figs. 2.1-2.3). Peptide sequences identified in the 39,000 molecular weight band from the soluble sperm fraction (Fig. 2.7A) matched the predicted amino acid sequence of ALDOART2 (Supplemental Fig. 2.1). We did not detect peptides unique to the somatic ALDOA isoform in this band, even though *in silico* tryptic digestion predicts that there are at least six peptides with sufficient difference to be assigned only to ALDOA.

Both *Aldoa_v2* and the *Aldoart1* retrogene encode larger proteins with 54 and 55 additional amino acids at the N-terminus, respectively. Our proteomic analyses identified peptide sequences matching each of these novel N-terminal sequences in the 50,000 molecular weight aldolase band that is enriched in isolated fibrous sheaths (Fig. 2.7A,

Supplemental Figs. 2.2B and 2.3B). Additional peptide sequences corresponding to other regions of ALDOA_V2 or ALDOART1 were also identified (Supplemental Figs. 2.2 and 2.3, respectively).

The larger ALDOA-related isoforms are synthesized late during spermatogenesis

Western analysis with a polyclonal antibody raised against rabbit muscle aldolase detected a 50,000 molecular weight band in mouse sperm (Fig. 2.8A) and testis (Fig. 2.8B), but not in several other tissues (Fig. 2.8A). We did not detect this band in mammary gland (virgin or lactating) or retina (data not shown), suggesting that *Aldoa_v2* transcripts detected by EST analysis in these tissues may represent artifacts or may not be translated. Despite the appearance of *Aldoart1* transcripts in round spermatids and *Aldoa_v2* transcripts during earlier stages of spermatogenesis (Figs. 2.5 and 2.6), the larger ALDOA-related protein band was first detected late in spermatogenesis in condensing spermatids (C in Fig. 2.8B). This protein was not detected in germ cells at earlier stages of spermatogenesis or in Sertoli cells. Therefore, expression of *Aldoa_v2* and *Aldoart1* during spermatogenesis appears to be regulated at both transcriptional and translational levels.

A ~50,000 molecular weight aldolase is also present in rat and human sperm (Fig. 2.8C). Alternatively spliced *Aldoa_v2* transcripts are likely to encode this isoform, since we did not identify *Aldoart1* retrogenes in either of these species. This result, along with the high level of sequence conservation across species (Fig. 2.3C), strongly supports a functional role for the extra amino acid residues located at the N-terminus of ALDOA_V2 and ALDOART1.

Discussion

This study identifies three new aldolase isozymes in mouse with restricted expression during male germ cell development. All three isozymes first appear late during

the haploid phase of spermatogenesis and are retained in mature sperm. ALDOART1 and ALDOART2 are encoded by *Aldoa*-related retrogenes which arose in the rodent lineage. ALDOA_V2 is a splice variant of *Aldoa* that is conserved in several mammalian species, including human. Both ALDOART1 and ALDOA_V2 have novel N-terminal extensions that may be responsible for tight binding to the fibrous sheath. We hypothesize that these N-terminal extensions are important for the proper localization of glycolytic ATP production, required for sperm motility. Our data exemplify the diverse mechanisms that generate genes with restricted expression patterns and novel functions in the germline, and illustrate the role of positive selection in maintaining expression of newly evolved genes with unique attributes.

Retrotransposition is a common feature of mammalian evolution, producing >4,800 retrotroposed sequences in the mouse genome [46, 47] and ~8,000 in the human genome [47-49]. LINE-1 transposable elements facilitate this process, initiating reverse transcription of an mRNA and insertion of the resulting intronless sequence into the genome at a new location [50]. Only those retrotransposition events that occur in the germline are heritable and, therefore, can be exploited by natural selection to generate new gene family members. Although most retroposed sequences are pseudogenes incapable of coding functional proteins, EST analyses indicate that >1,000 human retroposed sequences may be transcribed [51]. Further studies have confirmed that at least 100 of these sequences are functional retrogenes, with many exhibiting specific roles during spermatogenesis [48, 51-53]. We identified 17 retroposed sequences derived from the ancestral *Aldoa* gene located on mouse chromosome 7 (Supplemental Table 2.1), but none from *Aldob* or *Aldoc*, consistent with the observation that genes with multiple retroposed copies are frequently housekeeping genes that are highly expressed in the germline [47, 49]. Only *Aldoart1* and *Aldoart2*, which arose from independent retrotransposition events in the rodent lineage,

have full-length conserved ORFs. These retrogenes, like *Pgk2* [13, 14], contribute to the diversity of the glycolytic pathway in the male germline.

It has been proposed that expression of retrogenes during spermatogenesis is facilitated by a transcriptionally permissive environment [54, 55], particularly during the haploid stages when multiple components of the RNA polymerase II transcription machinery are substantially upregulated [56, 57]. Both *Aldoart1* and *Aldoart2* mRNAs appear in the testis at 20 days of age and are selectively expressed in spermatids, suggesting that they may share common regulatory mechanisms. Although the promoters of these retrogenes have not been defined, comparison of upstream flanking sequences indicates that they are not derived from *Aldoa,* their ubiquitiously expressed progenitor gene.

Unique constituents with restricted expression in spermatogenic cells have been identified in the multi-protein complexes required for transcriptional initiation [1, 3, 58, 59], alternative splicing [3, 59, 60], and 3'-processing of mRNAs [60]. Alterations in these processes contribute to the complexity of gene expression during spermatogenesis. Analysis of human ESTs indicates that the testis, along with brain and liver, have the highest frequencies of alternative splicing [61]. In this study, identification of the *Aldoart1* retrogene led to the discovery of an alternatively spliced exon that encodes an N-terminal extension of ALDOA. The *Aldoa_v2* splice variant, which served as the template for the *Aldoart1* retrogene, is expressed predominantly in testis. The restricted expression of this alternatively spliced exon and its conservation across distantly related species provide evidence that the encoded N-terminal sequence has an important functional role in male germ cells.

Western analysis demonstrated a delay in translation of the larger aldolases encoded by *Aldoart1* and the *Aldoa_v2* splice variant. Storage of mRNAs and delayed translation are common features of mammalian spermatogenesis, particularly during the haploid period of differentiation where extensive morphogenesis continues for a week or

longer after transcription ceases [2, 62]. Although *Aldoa_v2* is transcribed throughout spermatogenesis and *Aldoart1* is initially transcribed in round spermatids, the 50,000 molecular weight aldolase band was first detected in condensing spermatids, which appear in the testis during postnatal development approximately one week later than the round spermatids. GAPDHS [8, 63] and PGK2 [64, 65], two additional spermatogenesis-specific isoforms in the glycolytic pathway, are also synthesized in condensing spermatids following significant periods of translation repression. Both global and mRNA-specific mechanisms may regulate the translation of *Aldoart1* and *Aldoa_v2*, along with many other mRNAs transcribed during meiosis or the initial haploid phase of spermatogenesis [66].

Novel features of the newly identified ALDOA-related isozymes, including their restricted expression and sequence diversity, suggest neofunctionalization and positive selection within the *Aldo* gene family. ALDOA, ALDOB and ALDOC, the three known aldolase isozymes in vertebrates, catalyze the reversible cleavage of fructose-1,6-bisphosphate to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate [67, 68]. ALDOA and ALDOC preferentially catalyze the cleavage reaction essential for glycolysis, while the catalytic properties of ALDOB favor the reverse reaction required for gluconeogenesis [35, 67, 69, 70]. ALDOB also cleaves fructose-1-phosphate, a reaction that is essential for fructose metabolism [37, 71]. Amino acid differences in the ALDOA-related isozymes may contribute to distinctive enzymatic or structural properties, as reported for bovine sperm aldolases [19]. Three of the amino acids in *Aldoart1* (M95, G96, F112) that match ALDOB rather than ALDOA have been identified as isozyme-specific residues [45]. The *Aldoart1* MGN sequence between the first two active site residues that is identical to ALDOB is particularly interesting since it is in a region that contributes to the distinct properties of ALDOA and ALDOB [72-74]. Further studies will be necessary to assess the catalytic properties of the ALDOA-related isozymes. Although the potential for metabolizing fructose via an ALDOB-like isozyme would appear to be beneficial for sperm, fructose is

apparently unable to support hyperactivated sperm motility and *in vitro* fertilization in the mouse [75-77]. It is interesting to note that most of the isozyme-specific residues identified for the aldolases are not within hydrogen bonding distance of the active site and are localized on the surface where they may exert long-distance conformational effects or mediate tissue-specific protein interactions [45]. Several of these residues are localized in the flexible C-terminal region that has been shown to contribute to differences between the aldolase isozymes [74, 78]. Multiple unique residues are present within the C-terminal regions of both *Aldoart1* and *Aldoart2* where they could contribute to novel enzymatic properties.

Both ALDOART1 and ALDOA_V2, the larger aldolase isozymes, are present in purified fibrous sheath preparations. Like GAPDHS [8], these isozymes have distinctive N-terminal extensions that could mediate targeting to the fibrous sheath and/or tight binding to this cytoskeletal structure. Aldolase functions as a tetramer, and heterotetramers of ALDOA with ALDOB or ALDOC have been identified in various tissues [37, 69]. Our proteomic analyses did not detect ALDOA in mouse sperm. The restricted expression of the newly identified ALDOA-related isoforms in spermatids suggests that functional tetramers containing one or more of these isoforms are formed late during spermatogenesis and are tethered to the forming fibrous sheath by the N-terminal extensions of ALDOART1 and ALDOA_V2. Binding of aldolase tetramers to an insoluble structure such as the fibrous sheath could affect their conformation and modulate their kinetic properties.

Compartmentalization of the glycolytic enzymes in the principal piece may facilitate energy production along the entire length of the sperm flagellum, providing a localized supply of ATP to the dynein ATPases in the axoneme to maintain motility. Enzymatic complexes are common in nature and direct binding between glycolytic enzymes has been suggested [79-81]. Recent proteomic studies identified multiple glycolytic enzymes in the flagella of *Chlamydomonas reinhardtii,* including some that are associated with the

axoneme [82, 83]. These studies suggest that there may be a more general requirement for glycolytic enzyme localization and ATP production in flagella. Some glycolytic enzymes, including ALDOA and GAPDH, are co-localized at the Z-discs and M-lines in *Drosophila* flight muscles [84, 85]. Flies that lack glycerol-3-phosphate dehydrogenase lose the ability to fly and the localization of both ALDOA and GAPDH in the flight muscle, providing another link between the co-localization and function of glycolytic enzymes [84, 85]. Similarly, the co-localization of glycolytic enzymes in the principal piece of mammalian sperm may result in higher catalytic activity to support sperm motility.

There is ample evidence that the glycolytic pathway is extensively modified during mammalian spermatogenesis and is required for sperm function and male fertility. This study indicates that both retrotransposition and alternative splicing are responsible for the restricted expression of three novel ALDOA-related isoforms during the haploid phase of mouse spermatogenesis. All three are present in mature sperm and two isoforms with N-terminal extensions are tightly bound to the fibrous sheath. Further studies to assess the functional properties of the ALDOA-related isozymes and other sperm-specific glycolytic enzymes may provide new insights regarding potential genetic causes of infertility and the rational design of contraceptives. In addition, detailed genomic and molecular analyses of novel glycolytic enzymes in the male germline offer a promising approach to identify elements that regulate gene expression during spermatogenesis and to increase our understanding of the causes underlying the differential rates and non-random locations of retrotransposition.

The fate of most retrotransposition events is to fade into evolutionary history as pseudogenes. It is, therefore, remarkable that *Aldoart1* and *Aldoart2* show a striking combination of negative selection at all active site residues (Fig. 2.4), convergent evolution at isozyme-specific residues (Fig. 2.4), and positive selection at many other amino acids (Fig. 2.2 and Fig. 2.3B). Since these two retrogenes are only expressed during the final

stages of spermatogenesis and their products are present in sperm, we hypothesize that reproductive performance is the most likely force that has shaped their evolution.

Figure 2.1

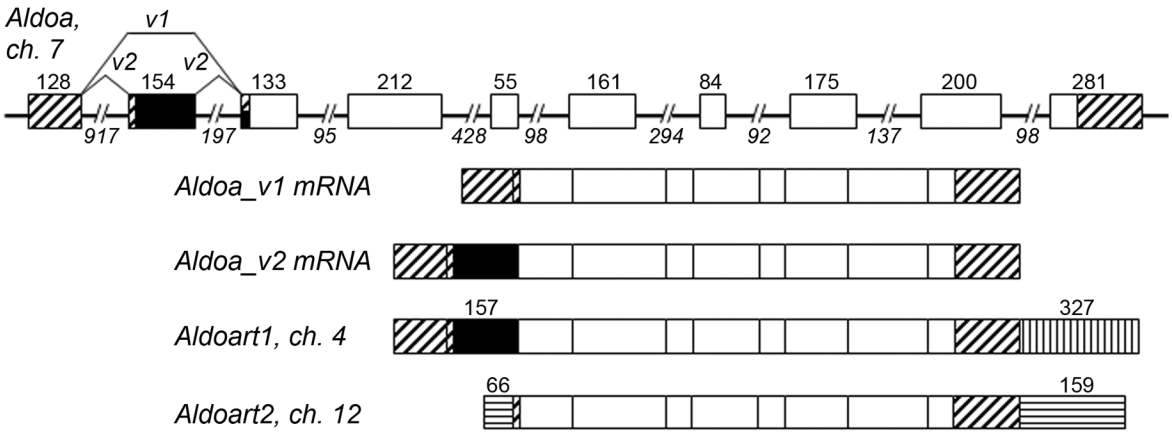


Figure 2.1. Genomic organization and mRNA structure of the mouse *Aldoa* gene and related retrogenes. The *Aldoa* gene on chromosome 7 is shown at the top, with numbers indicating the lengths of exons (boxes) and introns (lines). *Aldoart1* and *Aldoart2* are intronless retrogenes on chromosomes 4 and 12, respectively. Coding sequences are indicated by white and black boxes, with black boxes representing novel N-terminal sequences. Boxes with diagonal lines denote homologous UTRs in *Aldoa*, *Aldoart1* and *Aldoart2.* Boxes with vertical or horizontal lines denote UTR sequences found only in *Aldoart1* or *Aldoart2,* respectively. The lengths of novel regions are noted.

Figure 2.2

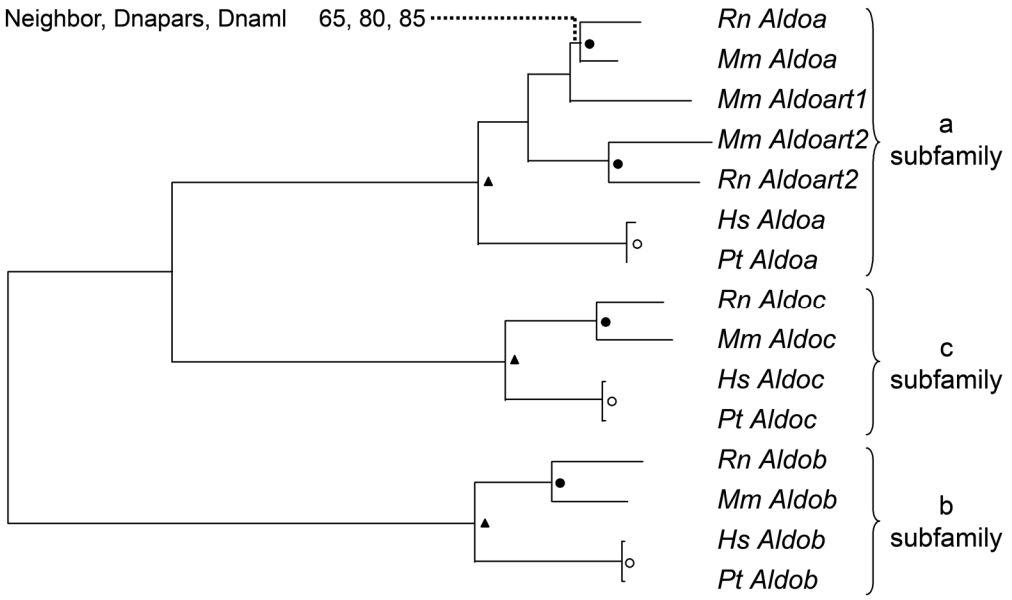


Figure 2.2. Phylogenetic relationships between aldolase genes and retrogenes. The tree depicts the relationships between orthologs in mouse (*Mm*), rat (*Rn*), human (*Hs*) and chimpanzee (*Pt*), and is based on a consensus cladogram from three consensus trees obtained by three different phylogenetic methods (Neighbor, Dnapars and Dnam). The figure shows the number of times out of 100 that a branch is observed in each method for the less well-supported branch. All remaining branches are supported by more than 90% of trees in all three methods. Triangles denote the primate/rodent split. Open circles represent the split between the human and chimpanzee lineages. Filled circles represent the split between the mouse and rat lineages. The length of the branches is proportional to the ones obtained in a representative tree using the Neighbor method.

Figure 2.3

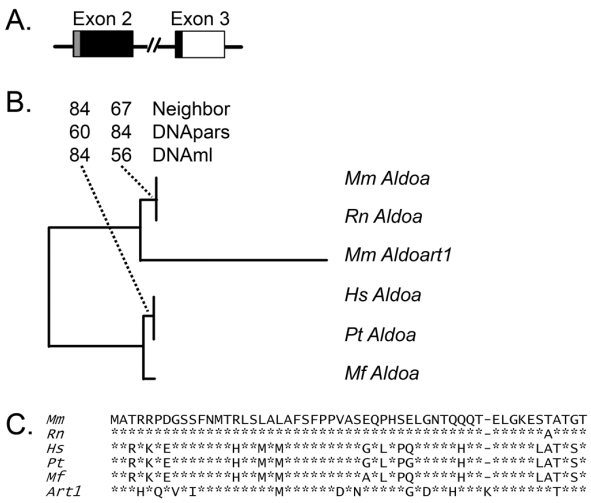


Figure 2.3. Phylogenetic relationships between novel N-terminal sequences in the aldolase A family. (*A*) This diagram shows the two exons that generate the novel *Aldoa* splice variant (*Aldoa_v2*). Exon 2 includes a fragment of the 5'-UTR (shown as a grey box) and 144 nucleotides that encode the first 48 amino acids of the N-terminal extension of ALDO_V2 (black box). Exon 3 contains 21 nucleotides that encode the remaining 7 amino acids of the novel N-terminal sequence (black box). (*B*) The tree depicts the relationships between homologous sequences encoding the N-terminal extensions of *Aldoa_v2* and *Aldoart1*. The tree is based on a consensus cladogram from three consensus trees obtained by three different phylogenetic methods (Neighbor, Dnapars and Dnaml). The figure also shows the number of times out of 100 that a branch is observed in each method for the two less well-supported branches. All remaining branches are supported by more than 90% of trees in all three methods. (*C*) Conservation of the amino acid sequence of the N-terminal extensions across species. Amino acid alignment of the N-terminal extension of ALDOA_V2 in mouse (*Mm*), rat (*Rn*), macaque (*Mf*), chimpanzee (*Pt*), and human (*Hs*), and mouse ALDOART1 (*Art1*). An asterisk (*) denotes residues that are identical while a dash (-) denotes the absence of that residue.

Figure 2.4

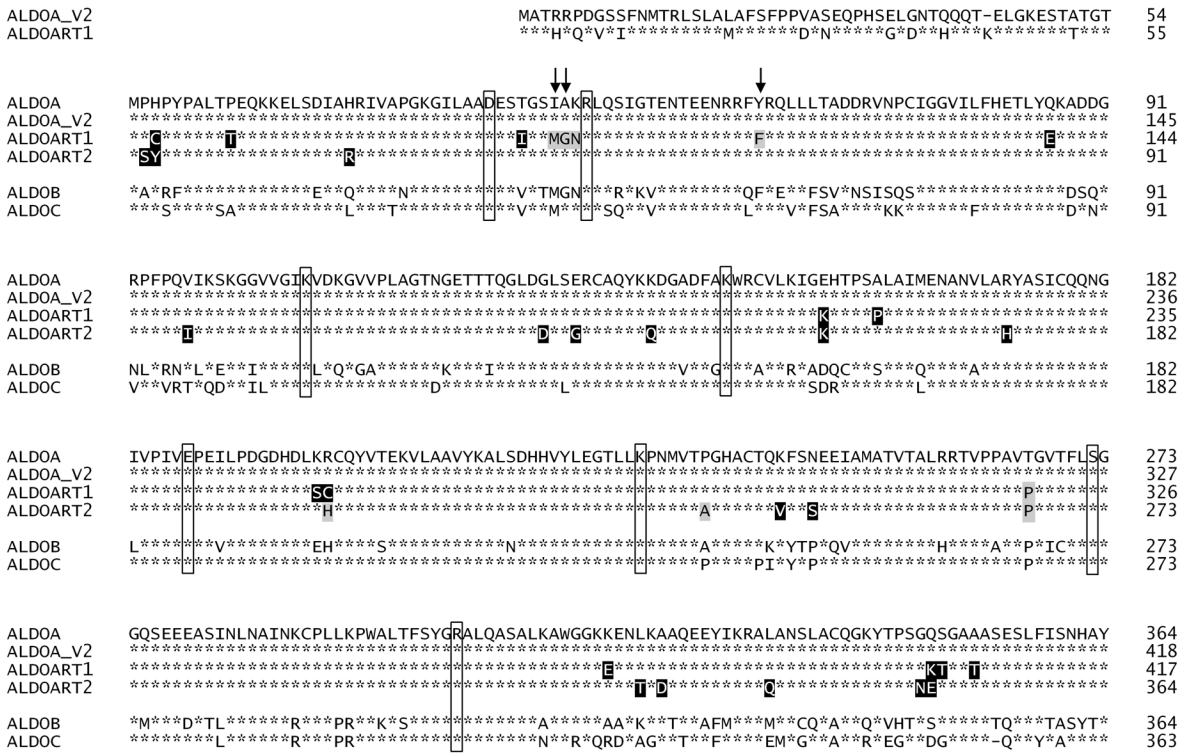


Figure 2.4. **Amino acid alignment for the mouse aldolase genes.** Asterisks (*) denotes identical amino acid residues. ALDOA_V2 and ALDOART1 have novel N-terminal extensions. Boxed amino acids denote active site residues, which are conserved in all members of the gene family. Amino acids that are unique to ALDOART1 or ALDOART2 are highlighted in black with white lettering. Additional residues in each retrogene that match ALDOB or ALDOC rather than ALDOA are highlighted in grey. Three isozymes-specific residues in ALDOART1 that match ALDOB are noted by black arrows.

Figure 2.5

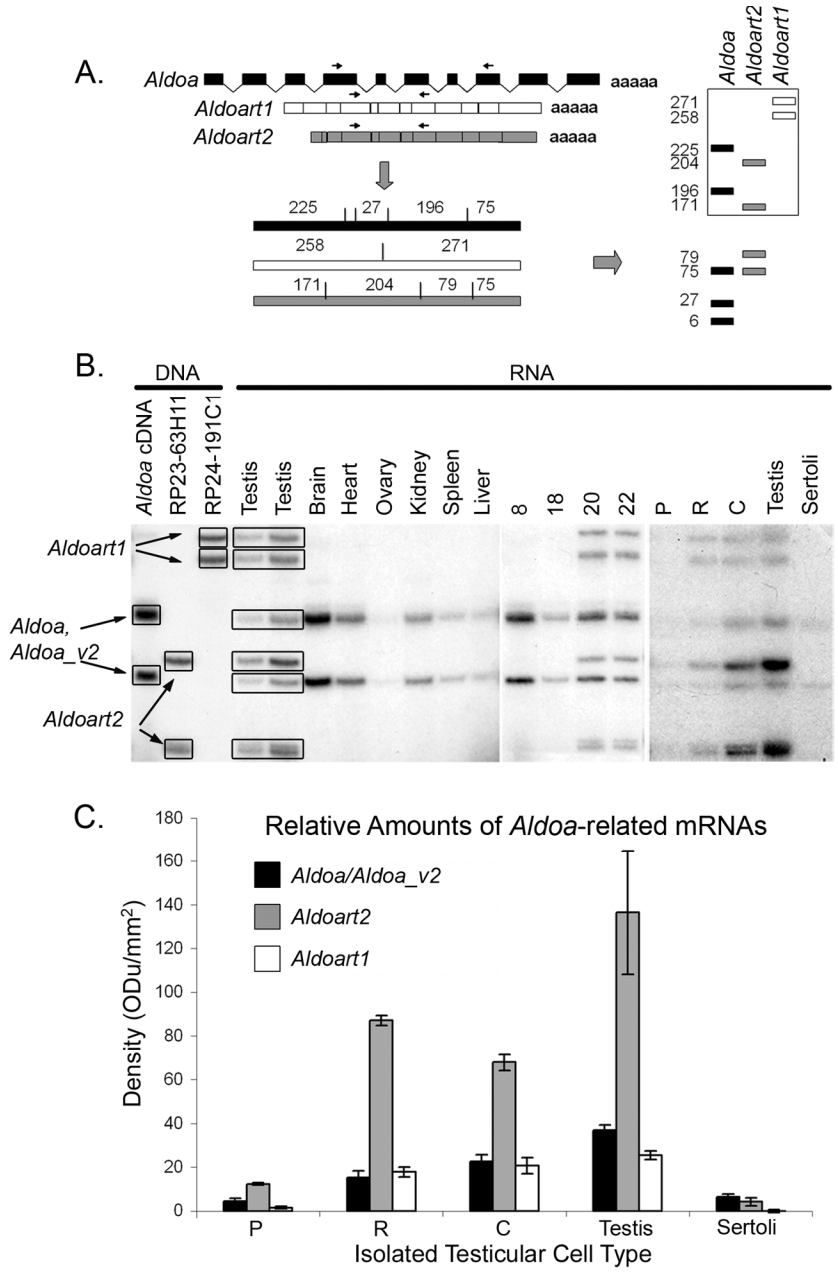

A.
Aldoa
Aldoart1
Aldoart2
aaaaa
225
27
196
75
258
271
171
204
79
75
Aldoa
Aldoart2
Aldoart1
271
258
225
204
196
171
79
75
27
6
B.
DNA
RNA
Aldoa cDNA
RP23-63H11
RP24-191C1
Testis
Testis
Brain
Heart
Ovary
Kidney
Spleen
Liver
8
18
20
22
P
R
C
Testis
Sertoli
Aldoart1
Aldoa,
Aldoa_v2
Aldoart2
C.
Relative Amounts of Aldoa-related mRNAs
Aldoa/Aldoa_v2
Aldoart2
Aldoart1
Density (ODu/mm2)
180
160
140
120
100
80
60
40
20
0
P
R
C
Testis
Sertoli
Isolated Testicular Cell Type

Figure 2.5. Restricted expression of *Aldoart1* and *Aldoart2* during the late stages of spermatogenesis. (*A*) Diagram of the RT-PCR approach used to distinguish *Aldoa*-related transcripts. Arrows denote a single primer set which amplified a 530 bp product from *Aldoa/Aldoa_v2*, *Aldoart1* and *Aldoart2* cDNAs. *Hae*III digestion of these products produced distinct fragments that were resolved by electrophoresis on 5% acrylamide gels. (*B*) The expected fragments were amplified from DNA positive controls, including an *Aldoa* cDNA subclone and BAC (bacterial artificial chromosome) clones for the intronless retrogenes (clone RP24-191C1 for *Aldoart1,* clone RP23-63H11 for *Aldoart2*). RT-PCR of total RNA samples detected *Aldoart1* and *Aldoart2* transcripts in testis, but not in other tissues. Transcripts from the *Aldoa* gene (*Aldoa_v1* or *Aldoa_v2*) were present in all tissues examined. RNAs for the second testis lane and remaining tissues were isolated from (C57BL/6J x A/J)F1 mice, while RNAs for the first testis lane, testes from juvenile animals between 8 and 22 day of age (8, 18, 20, 22) and isolated testicular cells were from CD-1 mice. During postnatal development, *Aldoart1* and *Aldoart2* are first expressed in the testis at 20 days of age. Semi-quantitative RT-PCR conditions were used to compare transcript levels in isolated pachytene spermatocytes (P), round spermatids (R), condensing spermatids (C) and Sertoli cells. (*C)* Analysis of the semi-quantitative RT-PCR products by densitometry. Bars represent mean densities ± SEM calculated from 3-6 replicated PCR reactions.

Figure 2.6

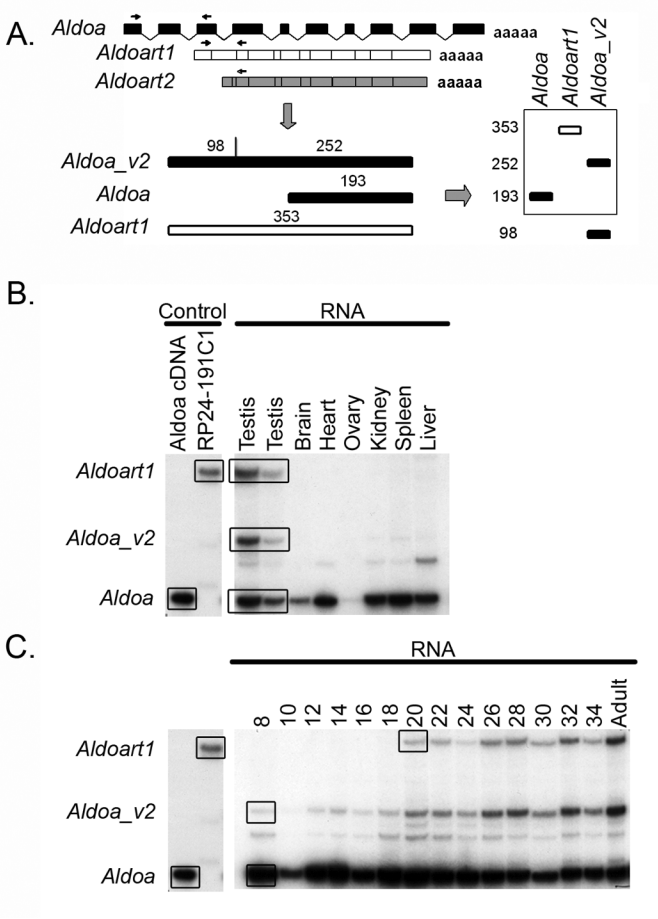

A.
Aldoa
Aldoart1
Aldoart2
aaaaa
aaaaa
aaaaa
Aldoa
Aldoart1
Aldoa_v2
353
252
193
98
98
252
193
353
Aldoa_v2
Aldoa
Aldoart1
B.
Control
RNA
Aldoa cDNA
RP24-191C1
Testis
Testis
Brain
Heart
Ovary
Kidney
Spleen
Liver
Aldoart1
Aldoa_v2
Aldoa
C.
RNA
8
10
12
14
16
18
20
22
24
26
28
30
32
34
Adult
Aldoart1
Aldoa_v2
Aldoa

Figure 2.6. Restricted expression of *Aldoa_v2* in testis. (*A*) Diagram of the RT-PCR strategy used to distinguish between *Aldoa, Aldoa_v2, Aldoart1.* Arrows denote a single primer set which amplified 350, 193 and 353 bp products from *Aldoa_v2*, *Aldoa*, and *Aldoart1,* respectively. *Hae*III digestion cleaved only the *Aldoa_v2* product, resulting in distinct bands for each transcript on 5% acrylamide gels. (*B*) This strategy detected *Aldoa_v2* transcripts only in testis (CD-1 testis in the first RNA lane, other tissue RNAs from (C57BL/6J x A/J)F_1 mice). (*C*) Testis RNAs from juvenile CD-1 mice (8-34 days of age) were also analyzed to compare the appearance of *Aldoa_v2* and *Aldoart1* transcripts during postnatal development. RNAs were isolated from pooled testes and RT-PCR was repeated three times to confirm reproducibility. Control DNAs for the assays in (*B*) and (*C*) included an *Aldoa* cDNA subclone and a BAC clone for *Aldoart1* (clone RP24-191C1).

Figure 2.7

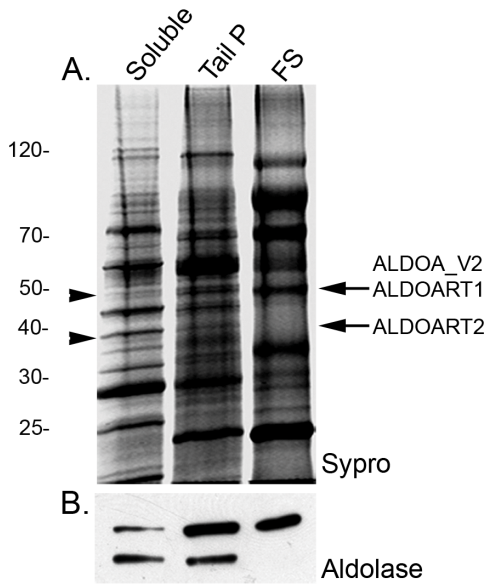


Figure 2.7. Proteomic and Western analyses identify novel ALDOA-related proteins in mouse sperm. Proteins in the soluble fraction following sonication (soluble), the tail pellet (tail P), and isolated fibrous sheaths (FS) were separated by SDS-PAGE. Protein from 10^6 sperm was loaded in the soluble and tail P lanes, while the FS lane contains protein from 10^7 sperm. (*A*) Bands indicated by arrows on this representative Sypro Ruby-stained gel were isolated and analysed by mass spectrometry. MALDI identified the bands as aldolase-A related proteins. LC-MALDI identified peptides specific for ALDOART2 in the soluble ~39,000 molecular weight protein and peptides specific for ALDOA_V2 and ALDOART1 in the ~50,000 molecular weight band enriched in the tail P and FS fractions (see Supplemental Figs S1-S3). (*B*) Western analysis of the three fractions using a polyclonal antibody raised against rabbit skeletal muscle aldolase.

Figure 2.8

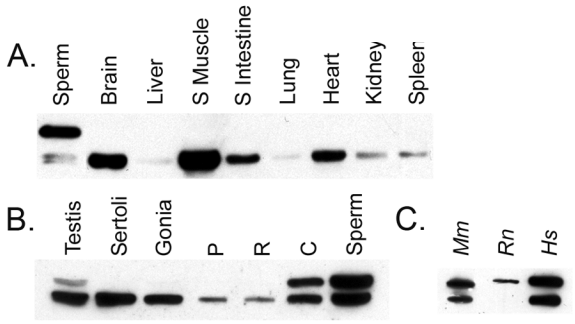


Figure 2.8. Larger ALDOA-related isoforms are first translated in condensing spermatids and are conserved in mouse, rat and human sperm. Following SDS-PAGE, Western blots were probed with a polyclonal antibody raised against rabbit skeletal muscle aldolase. Equal protein loads were confirmed by Coomassie blue staining of the blots (not shown). (*A*) A ~50,000 molecular weight ALDOA-related band was detected in mouse sperm, but not in eight somatic tissues including skeletal muscle (S Muscle) and small intestine (S Intestine). (*B*) The larger ALDOA-related band was also detected in testis and condensing spermatids (C), but not in Sertoli cells, spermatogonia (Gonia), pachytene spermatocytes (P) or round spermatids (R). (*C*) A larger ALDOA-related band is present in mouse (*Mm*), rat (*Rn*) and human (*Hs*) sperm.

Table 2.1. *Total number of aldolase ESTs in selected tissues and isolated testicular cells*

	Selected Tissues					Isolated Testicular Cells					
	All	Brain	Liver	Musc	Testis	L	S	A	B	P	R
Aldoc	>100	>100	1[b]	3[b]	0	0	0	0	0	0	0
Aldob	>100	0	>100	21	0	0	0	0	0	0	0
Aldoa	>100	>100	12	>100	8	1[b]	8[c]	0	3	1	0
Aldoa_v2[a]	17	0	0	0	1	0	0	0	0	0	0
Aldoart1	10	0	0	0	7	0	0	0	0	0	2
Aldoart2	33	0	0	0	27	0	0	0	0	0	3

ESTs were analyzed from selected mouse tissues and isolated testicular cells, including muscle (Musc), Leydig cells (L), Sertoli cells (S), type A spermatogonial (A), type B spermatogonial (B), pachytene spermatocytes (P) and round spermatids (R). ESTs from pooled libraries were excluded in this analysis.

[a] refers to exon 2

[b] BLAST hits from cell lines

[c] BLAST hits from primary Sertoli cell cultures

Table 2.2. *Expression frequency of Aldoa-related genes*

	Testis	Leydig	Sertoli	B Gonia	Pachytene	R Spermatid
Aldoa	0.8 (0.3)	11.4 (11.4)	7.5 (2.6)	3.7 (2.1)	1.1 (1.1)	0.0
Aldoart1	0.6 (0.2)	0.0	0.0	0.0	0.0	5.0 (3.5)
Aldoart2	2.3 (0.4)	0.0	0.0	0.0	0.0	7.5 (4.3)

Data are presented as the number of ESTs for each gene per 10,000 sequenced ESTs in libraries derived from testis and isolated cells including type B spermatogonia (B Gonia), pachytene spermatocytes (Pachytene) and round spermatids (R Spermatid). Standard errors are shown in parentheses.

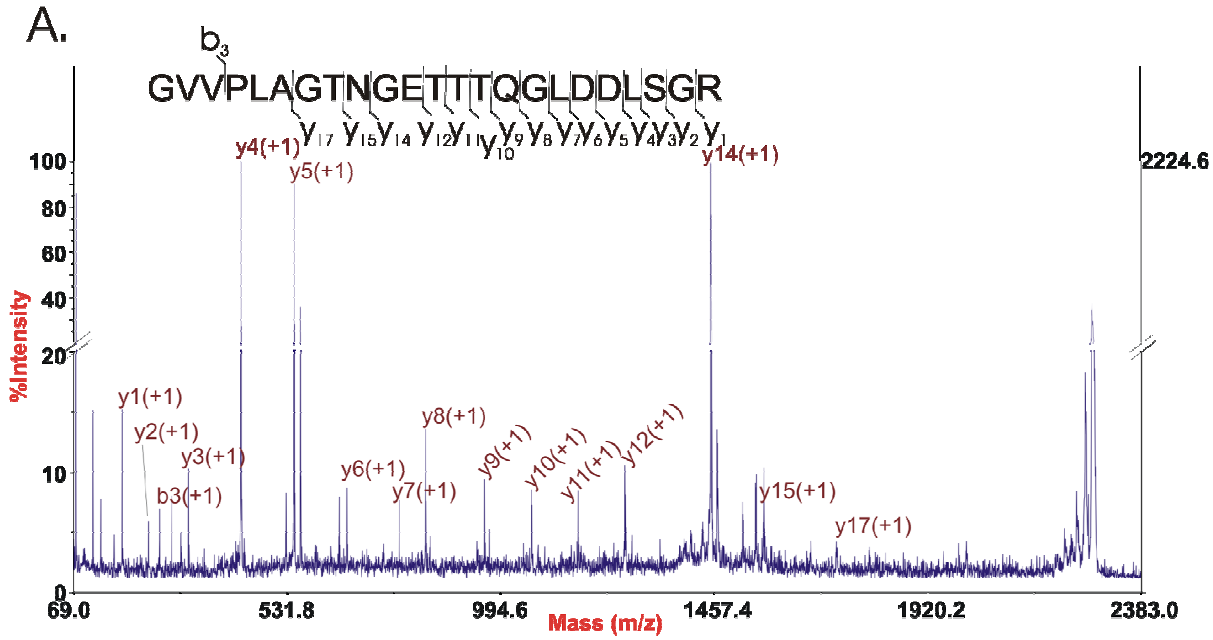
Table 2.3. *Aldoa-related gene expression in other tissues*

	Tissue	Number of ESTs	Frequency[a]
Aldoart1	Unknown	1	0.5 (0.5)
Aldoart2	Spleen	3	0.4 (0.2)

[a] Number of ESTs for each gene per 10,000 sequenced ESTs in each library, with standard errors in parentheses

Supplemental Figure 2.1

A.



B.

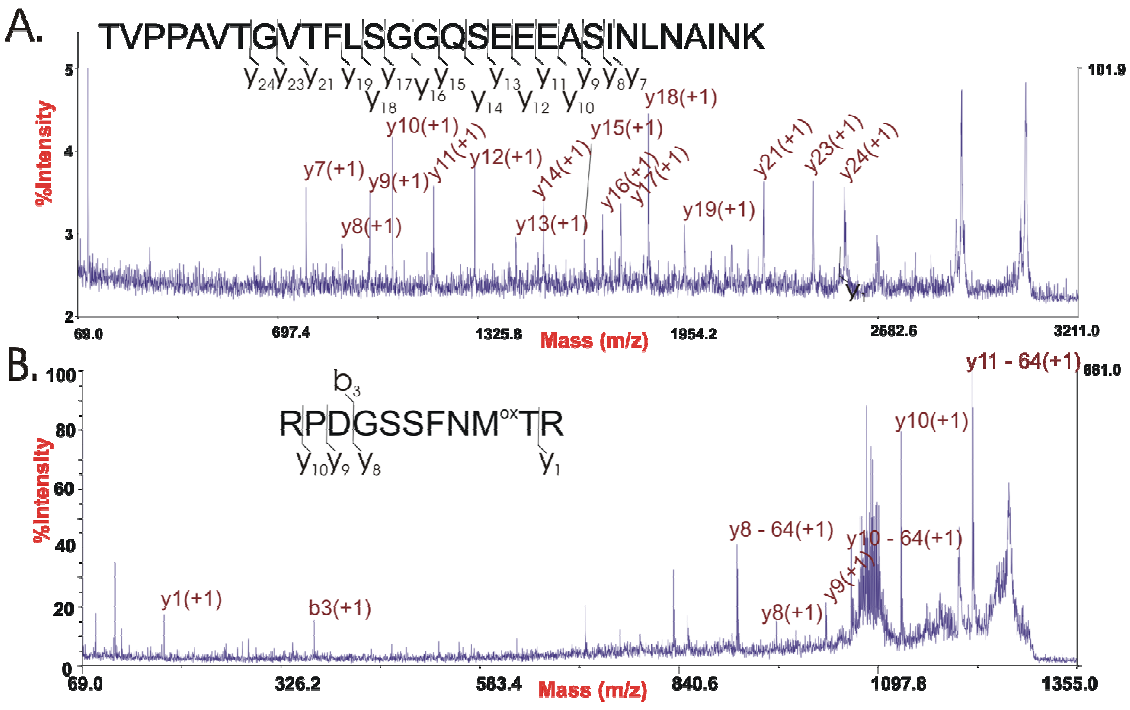
Observed Mass	Predicted Mass	Mass Accuracy	Peptide Sequence	MS/MS Ion Matches
931.4625	931.5207	62 ppm	KELSDIAR	Y1 y2 y3 y4 y5 y6 b3 b5 b6
995.5646	995.4680	97 ppm	DAQEEYIK	y1 y2 y3 y4 y5 y6 y7 b3 b4 b5 b6 b7
1087.5790	1087.6224	40 ppm	KELSDIARR	y1 y8 y9
1090.6110	1090.5315	73 ppm	AQANSLACQGK	b3 b6 y10 y11
1151.6023	1151.5691	29 ppm	DAQEEYIKR	y1 y2 y4 y6 y7 y8 b2 b3 b4 b6
1356.7306	1356.7270	3 ppm	ADDGRPFPQIIK	y1 y3 y4 y5 y6 y7 y9 y10 b3 b4 b5 b7 b9 b11
1467.5859	1467.6962	75 ppm	ENTKDAQEEYIK	y5 y6 y7 y8 y9 y10 b5 b6 b7 b8 b9 b10 b11
1623.6835	1623.7973	70 ppm	ENTKDAQEEYIKR	y1 y4 y5 y8 y10
2258.2263	2258.1259	44 ppm	GVVPLAGTNGETT-TQGLDDLSGR	y1 y2 y3 y4 y5 y6 y7 y8 y9 y10 y11 y12 y13 - y14 y15 y17 y18 y20 b3 b4 b7 b8
2316.1160	2316.0420	32 ppm	YTPSNESGAAASE-SLFISNHAY	y3 y5 y6 y7 y8 y9 y10 y11 y13 y16 y18 y20

C.

Observed Mass	Predicted Mass	Mass Accuracy	Isoforms Matched
959.5602	959.5269	34 ppm	ART2
1090.5249	1090.5310	6 ppm	ART2
1246.6190	1246.6321	11 ppm	ART2
1571.7847	1571.8540	44 ppm	ART2
2316.0293	2316.0415	5 ppm	ART2
1332.6239	1332.7005	57 ppm	ART2, A
1510.8085	1510.8013	5 ppm	ART2, A
1802.8899	1802.9104	11 ppm	ART2, A
851.3716	851.3894	21 ppm	ART2, A, ART1
1044.5543	1044.5684	13 ppm	ART2, A, ART1
1186.6620	1186.7154	45 ppm	ART2, A, ART1
1193.5338	1193.5698	30 ppm	ART2, A, ART1
1444.6044	1444.6525	33 ppm	ART2, A, ART1
1490.6588	1490.7087	33 ppm	ART2, A, ART1
1646.7917	1646.8098	11 ppm	ART2, A, ART1
1652.6940	1652.8353	85 ppm	ART2, A, ART1
1668.7765	1668.8302	32 ppm	ART2, A, ART1

Supplemental Figure 2.1. Mass spectrometry (MS) identifies ALDOART2 as a soluble protein in mouse sperm. Proteins in the soluble fraction following sonication were concentrated and separated on SDS-polyacrylamide gels. The band corresponding to the ~39,000 molecular weight aldolase isoform was subjected to in-gel digestion followed by LC-MALDI analysis. (*A*) A representative MS/MS spectrum showing fragmentation products from the m/z 2258.22 precursor ion. The m/z 2258 ion corresponds to the peptide GVVPLAGTNGETTTQGLDDLSGR, which is unique to the ALDOART2 isoform. Amino-terminal (b) and carboxyl-terminal (y) ions are indicated. (*B*) Summary of MS/MS data for all peaks unique to ALDOART2 that were analyzed in the LC-MALDI experiments. (*C*) Summary of additional peptide masses (within 100 ppm) found in the LC-MALDI experiments that could be derived from the ALDOART2 (ART2) isoform. Note that some peptides could be derived from more than one isoform due to the degree of homology between ALDOART2, ALDOA (A) and ALDOART1 (ART1).

Supplemental Figure 2.2



C.

Observed mass	Predicted mass	Mass accuracy	Peptide sequence	MS/MS Ion matches
1283.6101	1283.5797	24 ppm	RPDGSSFNM^{ox}TR	y1 y8 y9 y10 (y8-64) (y10-64) (y11-64) b3
3043.4092	3043.5582	49 ppm	TVPPAVTGVTFLSGG-QSEEEASINLNAINK	y7 y8 y9 y10 y11 y12 y13 y14 y15 y16y17 y18 y19 y21 y23 y24

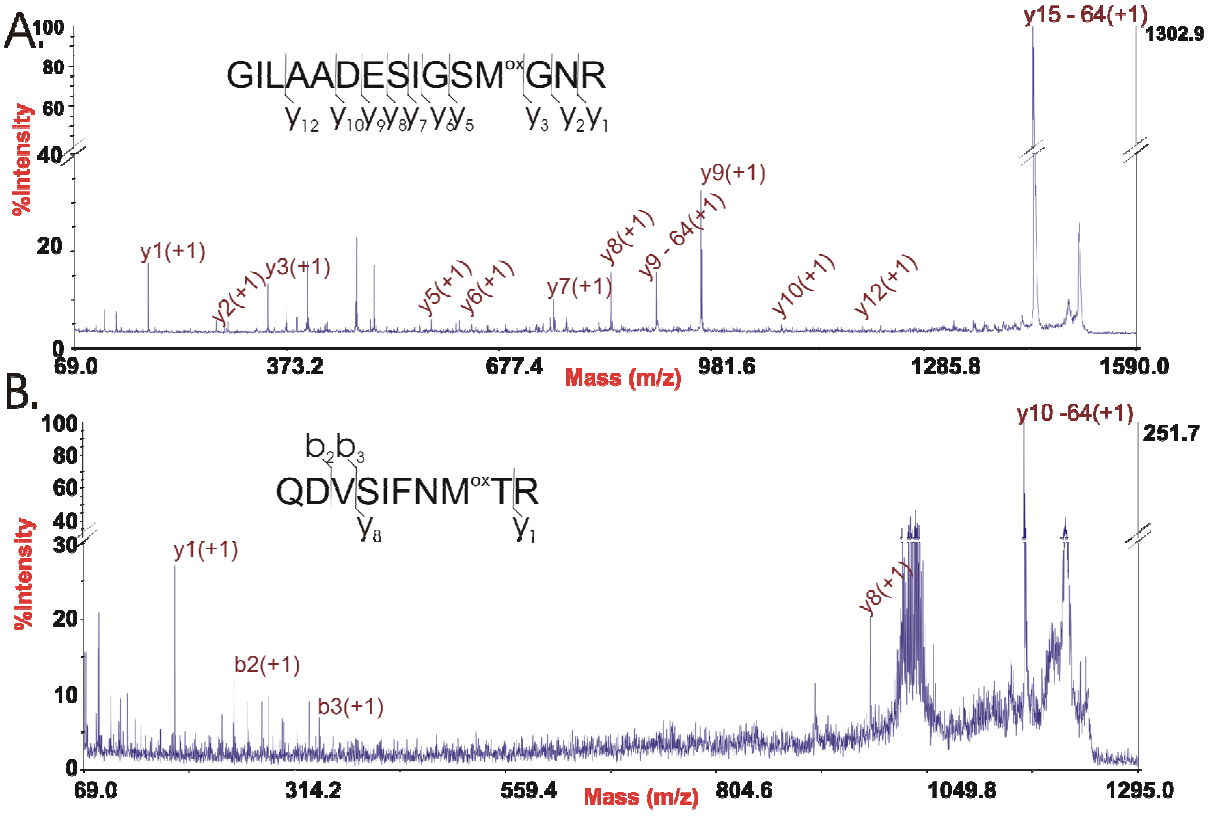
Supplemental Figure 2.2

D.

Observed Mass	Predicted Mass	Mass Accuracy	Isoforms Matched
870.3393	870.4026	73 ppm	A_V2
1267.5366	**1267.5848**	**38 ppm**	**A_V2**
1614.7423	1614.8560	70 ppm	A_V2
1726.8081	**1726.8112**	**2 ppm**	**A_V2**
1758.8347	**1758.8010**	**19 ppm**	**A_V2**
2107.0193	2107.0964	37 ppm	A_V2
2123.0288	2123.0914	29 ppm	A_V2
3199.4814	3199.6593	56 ppm	A_V2
3424.5813	**3424.7383**	**46 ppm**	**A_V2**
940.4711	940.4847	14 ppm	A_V2, ART1
951.4604	951.4782	19 ppm	A_V2, ART1
1065.5206	1065.5112	9 ppm	A_V2, ART1
1068.5457	1068.5796	32 ppm	A_V2, ART1
1075.5336	1075.5565	22 ppm	A_V2, ART1
1107.5684	1107.5793	10 ppm	A_V2, ART1
1332.6593	1332.7005	31 ppm	A_V2, ART2
1342.6946	1342.7114	13 ppm	A_V2, ART1
1557.7552	1557.8384	53 ppm	A_V2, ART1
1824.8812	1824.9273	26 ppm	A_V2, ART1
2112.9778	2113.0421	31 ppm	A_V2, ART2
2171.9810	2172.0278	22 ppm	A_V2, ART1
2257.9700	2258.0360	29 ppm	A_V2, ART2
2272.0220	2272.1415	53 ppm	A_V2, ART1
2614.2097	2614.3319	47 ppm	A_V2, ART1
851.4117	851.4258	17 ppm	A_V2, ART1, ART2
1044.5406	1044.5684	27 ppm	A_V2, ART1, ART2
1186.6427	1186.7154	61 ppm	A_V2, ART1, ART2
1193.6019	1193.6062	4 ppm	A_V2, ART1, ART2
1444.6323	1444.6889	39 ppm	A_V2, ART1, ART2
1646.7648	1646.8093	27 ppm	A_V2, ART1, ART2
1652.7739	1652.8312	35 ppm	A_V2, ART1, ART2
1668.7750	1668.8261	31 ppm	A_V2, ART1, ART2

Supplemental Figure 2.2. Mass spectrometry identifies ALDOA_V2 in mouse sperm. Following the separation of sperm tail or fibrous sheath proteins by SDS-PAGE, LC-MALDI was used to analyze the 50,000 molecular weight band. (*A*) A representative MS/MS spectrum showing fragmentation of the peak at m/z 3043.40. As in Supplemental Fig. 1, b and y ions from this ALDOA_V2 peptide are labeled. (*B*) A representative MS/MS spectrum from the m/z 1283.61 precursor ion, which is derived from the unique region of the ALDOA_V2 N-terminal extension. B and y ions are labeled. Y ions labeled -64 result from the loss of $HSOCH_3$ from the oxidized methionine. (*C*) Summary of the MS/MS spectra obtained in our LC-MALDI experiments that could be derived from ALDOA_V2. (*D*) Summary of additional peptide masses (within 100 ppm) observed in LC-MALDI experiments that could be derived from ALDOA_V2 (A_V2). Peptides in **bold** are derived specifically from the N-terminal extension of ALDOA_V2. Peptides in normal typeface are found in both ALDOA and ALDOA_V2, but are attributed to ALDOA_V2 because they were isolated from the ~50,000 molecular weight band. Again, due to the high level of sequence identity between ALDOA_V2, ALDOART1 (ART1) and ALDOART2 (ART2), certain peptides could be derived from more than one isoform.

Supplemental Figure 2.3



C.

Observed mass	Predicted mass	Mass accuracy	Peptide sequence	MS/MS Ion matches
1226.6870	1226.5834	84 ppm	QDVSIFNM^{ox}TR	y1 y8 b2 b3
1506.6857	1506.7217	24 ppm	GILAADESIGSM^{ox}GNR	y1 y2 y3 y5 y6 y7 y8 y9 y10 y12
1849.8660	1849.9589	50 ppm	HTPSPLAIMENANVLAR	y1 y2 (y5-17) y6 y7 y8 b2

D.

Observed Mass	Predicted Mass	Mass Accuracy	Isoforms Matched
1131.5425	1131.5793	33 ppm	ART1
1494.6714	1494.8063	90 ppm	ART1
1564.6633	1564.7853	78 ppm	ART1
1708.7996	1708.8687	40 ppm	ART1
1806.7949	1806.8738	43 ppm	ART1
1833.8595	1833.9640	57 ppm	ART1
2064.0107	2064.0396	14 ppm	ART1
2097.0918	2097.0472	21ppm	ART1
3197.3381	3197.4957	49 ppm	ART1

Supplemental Figure 2.3. Mass spectrometry identifies ALDOART1 in mouse sperm. Following the separation of sperm tail or fibrous sheath proteins by SDS-PAGE, LC-MALDI also identified ALDOART1 in the 50,000 molecular weight band. (*A*) A representative MS/MS spectrum from the m/z 1506.68 precursor ion. Fragment ions in this spectrum can correspond to b and y ions from the GILAADESIGMGNR peptide, which is unique to the ALDOART1 isoform. (*B*) MS/MS spectrum from the m/z 1226.68 precursor ion. Fragment ions in this spectrum correspond to b and y ions from the QDVSIFNMTR peptide unique to the N-terminal extension of ALDOART1. (*C*) Summary of MS/MS results for peptides unique to the ALDOART1 isoform. D. Summary of additional peptide masses found in our LC-MALDI experiments that correspond (within 100 ppm) to ALDOART1 (ART1) peptides. For brevity, masses that could be derived from multiple aldolase isoforms are not repeated from Figure 2D.

Supplemental Table 2.1. *Mouse aldolase A gene, retrogenes and pseudogenes*

Chromosome	Start	End	Status
7	126586384	126590331	Gene
4	72337488	72339071	Retrogene
12	56483411	56484712	Retrogene
X	152721872	152723250	Pseudogene
6	39772630	39774080	Pseudogene
3	57450178	57451543	Pseudogene
X	153493899	153495466	Pseudogene
10	29773608	29774775	Pseudogene
X	128234278	128235510	Pseudogene
X	114814571	114815824	Pseudogene
X	54006718	54007872	Pseudogene
9	79132459	79133400	Pseudogene
X	127515370	127516260	Pseudogene
14	55135918	55136753	Pseudogene
9	79131982	79132447	Pseudogene
2	83195325	83195668	Pseudogene
4	112228575	112228962	Pseudogene
1	92957529	92957664	Pseudogene

Assembly: NCBI m36, December 2005; Genebuild: Ensembl, April 2006; Database version: 42.36c

References

1. DeJong J: **Basic mechanisms for the control of germ cell gene expression**. *Gene* 2006, **366**(1):39-50.

2. Eddy EM: **Male germ cell gene expression**. *Recent Prog Horm Res* 2002, **57**:103-128.

3. Elliott DJ, Grellscheid SN: **Alternative RNA splicing regulation in the testis**. *Reproduction* 2006, **132**(6):811-819.

4. Pang AL, Johnson W, Ravindranath N, Dym M, Rennert OM, Chan WY: **Expression profiling of purified male germ cells: stage-specific expression patterns related to meiosis and postmeiotic development**. *Physiol Genomics* 2006, **24**(2):75-85.

5. Schultz N, Hamra FK, Garbers DL: **A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets**. *Proc Natl Acad Sci U S A* 2003, **100**(21):12201-12206.

6. Shima JE, McLean DJ, McCarrey JR, Griswold MD: **The murine testicular transcriptome: characterizing gene expression in the testis during the progression of spermatogenesis**. *Biol Reprod* 2004, **71**(1):319-330.

7. Steinke D, Hoegg S, Brinkmann H, Meyer A: **Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates**. *BMC Biol* 2006, **4**:16.

8. Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis**. *Biol Reprod* 1998, **58**(3):834-841.

9. Welch JE, Schatte EC, O'Brien DA, Eddy EM: **Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells**. *Biol Reprod* 1992, **46**(5):869-878.

10. Coonrod S, Vitale A, Duan C, Bristol-Gould S, Herr J, Goldberg E: **Testis-specific lactate dehydrogenase (LDH-C_4; Ldh3) in murine oocytes and preimplantation embryos**. *J Androl* 2006, **27**(4):502-509.

11. Goldberg E: **Isozymes in testes and spermatozoa**. *Isozymes: Curr Topics Biol Med Res* 1977, **1**:79-124.

12. Li SS, O'Brien DA, Hou EW, Versola J, Rockett DL, Eddy EM: **Differential activity and synthesis of lactate dehydrogenase isozymes A (muscle), B (heart), and C (testis) in mouse spermatogenic cells**. *Biol Reprod* 1989, **40**(1):173-180.

13. Boer PH, Adra CN, Lau YF, McBurney MW: **The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon**. *Mol Cell Biol* 1987, **7**(9):3107-3112.

14. McCarrey JR, Thomas K: **Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene**. *Nature* 1987, **326**(6112):501-505.

15. Bluthmann H, Cicurel L, Kuntz GW, Haedenkamp G, Illmensee K: **Immunohistochemical localization of mouse testis-specific phosphoglycerate kinase (PGK-2) by monoclonal antibodies**. *EMBO J* 1982, **1**(4):479-484.

16. Vandeberg JL, Lee CY, Goldberg E: **Immunohistochemical localization of phosphoglycerate kinase isozymes in mouse testes**. *J Exp Zool* 1981, **217**(3):435-441.

17. Buehr M, McLaren A: **An electrophoretically detectable modification of glucosephosphate isomerase in mouse spermatozoa**. *J Reprod Fertil* 1981, **63**(1):169-173.

18. Yamada S, Nakajima H, Kuehn MR: **Novel testis- and embryo-specific isoforms of the phosphofructokinase-1 muscle type gene**. *Biochem Biophys Res Commun* 2004, **316**(2):580-587.

19. Gillis BA, Tamblyn TM: **Association of bovine sperm aldolase with sperm subcellular components**. *Biol Reprod* 1984, **31**(1):25-35.

20. Edwards YH, Grootegoed JA: **A sperm-specific enolase**. *J Reprod Fertil* 1983, **68**(2):305-310.

21. Gitlits VM, Toh BH, Loveland KL, Sentry JW: **The glycolytic enzyme enolase is present in sperm tail and displays nucleotide-dependent association with microtubules**. *Eur J Cell Biol* 2000, **79**(2):104-111.

22. Mori C, Nakamura N, Welch JE, Gotoh H, Goulding EH, Fujioka M, Eddy EM: **Mouse spermatogenic cell-specific type 1 hexokinase (*mHk1-s*) transcripts are expressed by alternative splicing from the *mHk1* gene and the HK1-S protein is localized mainly in the sperm tail**. *Mol Reprod Dev* 1998, **49**(4):374-385.

23. Mori C, Welch JE, Fulcher KD, O'Brien DA, Eddy EM: **Unique hexokinase messenger ribonucleic acids lacking the porin-binding domain are developmentally expressed in mouse spermatogenic cells**. *Biol Reprod* 1993, **49**(2):191-203.

24. Eddy EM, Toshimori K, O'Brien DA: **Fibrous sheath of mammalian spermatozoa**. *Microsc Res Tech* 2003, **61**(1):103-115.

25. Hoskins DD: **Adenine nucleotide mediation of fructolysis and motility in bovine epididymal spermatozoa**. *J Biol Chem* 1973, **248**(4):1135-1140.

26. Mann T, Lutwak-Mann C: **Male reproductive function and semen**. New York, NY: Springer-Verlag; 1981.

27. Peterson RN, Freund M: **Glycolysis by washed suspensions of human spermatozoa. Effect of substrate, substrate concentration, and changes in medium composition on the rate of glycolysis**. *Biol Reprod* 1969, **1**(3):238-246.

28. Mukai C, Okuno M: **Glycolysis plays a major role for adenosine triphosphate supplementation in mouse sperm flagellar movement**. *Biol Reprod* 2004, **71**(2):540-547.

29. Williams AC, Ford WC: **The role of glucose in supporting motility and capacitation in human spermatozoa**. *J Androl* 2001, **22**(4):680-695.

30. Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility**. *Proc Natl Acad Sci U S A* 2004, **101**(47):16501-16506.

31. Narisawa S, Hecht NB, Goldberg E, Boatright KM, Reed JC, Millan JL: **Testis-specific cytochrome *c*-null mice produce functional sperm but undergo early testicular atrophy**. *Mol Cell Biol* 2002, **22**(15):5554-5562.

32. Krisfalusi M, Miki K, Magyar PL, O'Brien D A: **Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa**. *Biol Reprod* 2006, **75**(2):270-278.

33. Welch JE, Brown PL, O'Brien DA, Magyar PL, Bunch DO, Mori C, Eddy EM: **Human glyceraldehyde 3-phosphate dehydrogenase-2 gene is expressed specifically in spermatogenic cells**. *J Androl* 2000, **21**(2):328-338.

34. Kim YH, Haidl G, Schaefer M, Egner U, Herr JC: **Compartmentalization of a unique ADP/ATP carrier protein SFEC (Sperm Flagellar Energy Carrier, AAC4) with glycolytic enzymes in the fibrous sheath of the human sperm flagellar principal piece**. *Dev Biol* 2006, **302**:463-476.

35. Penhoet EE, Kochman M, Rutter WJ: **Molecular and catalytic properties of aldolase C**. *Biochemistry* 1969, **8**(11):4396-4402.

36. Rutter WJ, Rajkumar T, Penhoet E, Kochman M, Valentine R: **Aldolase variants: structure and physiological significance**. *Ann N Y Acad Sci* 1968, **151**(1):102-117.

37. Penhoet E, Rajkumar T, Rutter WJ: **Multiple forms of fructose diphosphate aldolase in mammalian tissues**. *Proc Natl Acad Sci U S A* 1966, **56**(4):1275-1282.

38. McCarrey JR, O'Brien DA, Skinner MK: **Construction and preliminary characterization of a series of mouse and rat testis cDNA libraries**. *J Androl* 1999, **20**(5):635-639.

39. A. OBD: **Isolation, separation, and short-term culture of spermatogenic cells**. In: *Methods in Toxicology* Edited by Chapin RE, Heindel JJ, vol. 3A; 1993: 246-264.

40. Tsuruta JK, Eddy EM, O'Brien DA: **Insulin-like growth factor-II/cation-independent mannose 6-phosphate receptor mediates paracrine interactions during spermatogonial development**. *Biol Reprod* 2000, **63**(4):1006-1013.

41. O'Brien DA, Gabel CA, Rockett DL, Eddy EM: **Receptor-mediated endocytosis and differential synthesis of mannose 6-phosphate receptors in isolated spermatogenic and Sertoli cells**. *Endocrinology* 1989, **125**(6):2973-2984.

42. Parker C.E. MV, Warren M.R., Greer S.F., Scarlett C.O., and Borchers C.H.: **Mass spectrometric determination of protein ubiqitination**. In: *Ubiquitin-proteasome protocols, methods in molecular biology series.* Edited by Patterson C, Cyr, D.M. Patterson, N.J. : Humana Press; 2005: 117-152.

43. Cortinas MN, Lessa EP: **Molecular evolution of aldolase A pseudogenes in mice: multiple origins, subsequent duplications, and heterogeneity of evolutionary rates**. *Mol Biol Evol* 2001, **18**(9):1643-1653.

44. Arakaki TL, Pezza JA, Cronin MA, Hopkins CE, Zimmer DB, Tolan DR, Allen KN: **Structure of human brain fructose 1,6-(bis)phosphate aldolase: linking isozyme structure with function**. *Protein Sci* 2004, **13**(12):3077-3084.

45. Pezza JA, Choi KH, Berardini TZ, Beernink PT, Allen KN, Tolan DR: **Spatial clustering of isozyme-specific residues reveals unlikely determinants of isozyme specificity in fructose-1,6-bisphosphate aldolase**. *J Biol Chem* 2003, **278**(19):17307-17313.

46. Sakai H, Koyanagi KO, Imanishi T, Itoh T, Gojobori T: **Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes**. *Gene* 2007, **389**(2):196-203.

47. Zhang Z, Carriero N, Gerstein M: **Comparative analysis of processed pseudogenes in the mouse and human genomes**. *Trends Genet* 2004, **20**(2):62-67.

48. Nishimune Y, Tanaka H: **Infertility caused by polymorphisms or mutations in spermatogenesis-specific genes**. *J Androl* 2006, **27**(3):326-334.

49. Zhang Z, Gerstein M: **Large-scale analysis of pseudogenes in the human genome**. *Curr Opin Genet Dev* 2004, **14**(4):328-335.

50. Kazazian HH, Jr.: **Mobile elements: drivers of genome evolution**. *Science* 2004, **303**(5664):1626-1632.

51. Vinckenbosch N, Dupanloup I, Kaessmann H: **Evolutionary fate of retroposed gene copies in the human genome**. *Proc Natl Acad Sci U S A* 2006, **103**(9):3220-3225.

52. Emerson JJ, Kaessmann H, Betran E, Long M: **Extensive gene traffic on the mammalian X chromosome**. *Science* 2004, **303**(5657):537-540.

53. Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H: **Emergence of young human genes after a burst of retroposition in primates**. *PLoS Biol* 2005, **3**(11):e357.

54. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH, Jr.: **L1 integration in a transgenic mouse model**. *Genome Res* 2006, **16**(2):240-250.

55. Kleene KC, Mulligan E, Steiger D, Donohue K, Mastrangelo MA: **The mouse gene encoding the testis-specific isoform of Poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes**. *J Mol Evol* 1998, **47**(3):275-281.

56. Schmidt EE, Ohbayashi T, Makino Y, Tamura T, Schibler U: **Spermatid-specific overexpression of the TATA-binding protein gene involves recruitment of two potent testis-specific promoters**. *J Biol Chem* 1997, **272**(8):5326-5334.

57. Schmidt EE, Schibler U: **High accumulation of components of the RNA polymerase II transcription machinery in rodent spermatids**. *Development* 1995, **121**(8):2373-2383.

58. Hochheimer A, Tjian R: **Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression**. *Genes Dev* 2003, **17**(11):1309-1320.

59. Sergeant KA, Bourgeois CF, Dalgliesh C, Venables JP, Stevenin J, Elliott DJ: **Alternative RNA splicing complexes containing the scaffold attachment factor SAFB2**. *J Cell Sci* 2007, **120**(Pt 2):309-319.

60. Liu D, Brockman JM, Dass B, Hutchins LN, Singh P, McCarrey JR, MacDonald CC, Graber JH: **Systematic variation in mRNA 3'-processing signals during mouse spermatogenesis**. *Nucleic Acids Res* 2007, **35**(1):234-246.

61. Yeo G, Holste D, Kreiman G, Burge CB: **Variation in alternative splicing across human tissues**. *Genome Biol* 2004, **5**(10):R74.

62. Kleene KC: **Patterns of translational regulation in the mammalian testis**. *Mol Reprod Dev* 1996, **43**(2):268-281.

63. Mori C, Welch JE, Sakai Y, Eddy EM: **In situ localization of spermatogenic cell-specific glyceraldehyde 3-phosphate dehydrogenase (*gapd-s*) messenger ribonucleic acid in mice**. *Biol Reprod* 1992, **46**(5):859-868.

64. McCarrey JR, Berg WM, Paragioudakis SJ, Zhang PL, Dilworth DD, Arnold BL, Rossi JJ: **Differential transcription of *pgk* genes during spermatogenesis in the mouse**. *Dev Biol* 1992, **154**(1):160-168.

65. McCarrey JR, Kumari M, Aivaliotis MJ, Wang Z, Zhang P, Marshall F, Vandeberg JL: **Analysis of the cDNA and encoded protein of the human testis-specific *pgk-2* gene**. *Dev Genet* 1996, **19**(4):321-332.

66. Kleene KC: **Patterns, mechanisms, and functions of translation regulation in mammalian spermatogenic cells**. *Cytogenet Genome Res* 2003, **103**(3-4):217-224.

67. Rutter WJ, Blostein RE, Woodfin BM, Weber CS: **Enzyme variants and metabolic diversification**. *Adv Enzyme Regul* 1963, **17**:39-56.

68. Penhoet EE, Rutter WJ: **Detection and isolation of mammalian fructose-diphosphate aldolases**. *Methods Enzymol* 1975, **42**:240-249.

69. Penhoet EE, Rutter WJ: **Catalytic and immunochemical properties of homomeric and heteromeric combinations of aldolase subunits**. *J Biol Chem* 1971, **246**(2):318-323.

70. Penhoet EE: **Isolation of fructose diphosphate aldolases A, B, and C.** *Biochemistry* 1969, **8**(11):4391-4395.

71. Cori GT, Ochoa S, Slein MW, Cori CF: **The metabolism of fructose in liver. Isolation of fructose-2-phosphate and inorganic pyrophosphate**. *Biochim Biophys Acta* 1951, **7**(2):304-317.

72. Kitajima Y, Takasaki Y, Takahashi I, Hori K: **Construction and properties of active chimeric enzymes between human aldolases A and B. Analysis of molecular regions which determine isozyme-specific functions**. *J Biol Chem* 1990, **265**(29):17493-17498.

73. Kusakabe T, Motoki K, Sugimoto Y, Takasaki Y, Hori K: **Human aldolase B: liver-specific properties of the isozyme depend on type B isozyme group-specific sequences**. *Protein Eng* 1994, **7**(11):1387-1393.

74. Motoki K, Kitajima Y, Hori K: **Isozyme-specific modules on human aldolase A molecule. Isozyme group-specific sequences 1 and 4 are required for showing characteristics as aldolase A**. *J Biol Chem* 1993, **268**(3):1677-1683.

75. Cooper TG: **The onset and maintenance of hyperactivated motility of spermatozoa in the mouse.** *Gamete Research* 1984, **9**:55-74.

76. Fraser LR, Quinn PJ: **A glycolytic product is obligatory for initiation of the sperm acrosome reaction and whiplash motility required for fertilization in the mouse**. *J Reprod Fertil* 1981, **61**(1):25-35.

77. Hoppe PC: **Glucose requirement for mouse sperm capacitation in vitro**. *Biol Reprod* 1976, **15**(1):39-45.

78. Berthiaume L, Tolan DR, Sygusch J: **Differential usage of the carboxyl-terminal region among aldolase isozymes**. *J Biol Chem* 1993, **268**(15):10826-10835.

79. Vertessy BG, Orosz F, Kovacs J, Ovadi J: **Alternative binding of two sequential glycolytic enzymes to microtubules. Molecular studies in the phosphofructokinase/aldolase/microtubule system**. *J Biol Chem* 1997, **272**(41):25542-25546.

80. Ouporov IV, Knull HR, Huber A, Thomasson KA: **Brownian dynamics simulations of aldolase binding glyceraldehyde 3-phosphate dehydrogenase and the possibility of substrate channeling**. *Biophys J* 2001, **80**(6):2527-2535.

81. Westhoff D, Kamp G: **Glyceraldehyde 3-phosphate dehydrogenase is bound to the fibrous sheath of mammalian spermatozoa**. *J Cell Sci* 1997, **110**(15):1821-1829.

82. Pazour GJ, Agrin N, Leszyk J, Witman GB: **Proteomic analysis of a eukaryotic cilium**. *J Cell Biol* 2005, **170**(1):103-113.

83. Mitchell BF, Pedersen LB, Feely M, Rosenbaum JL, Mitchell DR: **ATP production in Chlamydomonas reinhardtii flagella by glycolytic enzymes**. *Mol Biol Cell* 2005, **16**(10):4509-4518.

84. Sullivan DT, MacIntyre R, Fuda N, Fiori J, Barrilla J, Ramizel L: **Analysis of glycolytic enzyme co-localization in *Drosophila* flight muscle**. *J Exp Biol* 2003, **206**(Pt 12):2031-2038.

85. Wojtas K, Slepecky N, von Kalm L, Sullivan D: **Flight muscle function in *Drosophila* requires colocalization of glycolytic enzymes**. *Mol Biol Cell* 1997, **8**(9):1665-1675.

CHAPTER 3

KINETIC PROPERTIES OF SPERMATOGENIC-CELL SPECIFIC ALDOLASE A ISOZYMES

Abstract

Fructose-1,6-bisphosphate (FBP) aldolase, or aldolase, is a metabolic enzyme which cleaves FBP during glycolysis. Aldolase also functions during gluconeogenesis and fructose metabolism. Glycolysis is modified during mouse spermatogenesis through the expression of spermatogenic-cell specific enzymes, including three novel aldolase A (ALDOA)-related proteins: ALDOA_V2, ALDOART1 and ALDOART2. Both ALDOA_V2 and ALDOART1 possess N-terminal extensions that are not present in ALDOA or other aldolases. Conserved spermatogenic-cell specific expression of ALDOA_V2 across species supports the functional importance of this isozyme. Here we report high levels of ALDOA FBP activity in insoluble fractions of mouse sperm. Amino acid alignment and preliminary three-dimensional modeling suggests that ALDOART1 and ALDOART2 may have unique catalytic properties. Recombinant ALDOA_V2, ALDOART1, and ALDOART2 do not form tetramers when expressed in *E. coli*, which may contribute to dramatically reduced FBP activity as compared to ALDOA. Understanding energy metabolism in sperm is important in the context of male fertility. Previous studies have identified glycolysis as the primary source of ATP production required for sperm motility and male fertility. Therefore, defining the kinetic properties of spermatogenic-cell specific aldolase A isozymes may provide insights on the regulation of sperm metabolism.

Introduction

Fructose-1,6-bisphosphate aldolase, or aldolase, is an enzyme which catalyzes the cleavage of fructose-1,6-bisphosphate (FBP) in the glycolytic pathway, forming two triose phosphates. Since aldolase is a reversible enzyme, it is known to participate in the reverse reaction of glycolysis, gluconeogenesis. Aldolase also cleaves fructose-1phosphate (F1P) during fructose metabolism, or fructolysis.

A spectrophotometric assay measures aldolase activity through the addition of exogenous enzyme to a reaction mixture containing the FBP substrate, two coupling enzymes (triose phosphate isomerase and α-glycerophosphate dehydrogenase), and β-NADH (Figure 3.1A). When aldolase cleaves FBP, β-NADH is oxidized to β-NAD in the coupled reaction catalyzed by α-glycerophosphate dehydrogenase. The loss of β-NADH is monitored at 340 nm, reflecting increases in aldolase activity [1, 2].

A hyperbolic saturation curve is created by measuring the activity of a particular enzyme over varying substrate concentrations (Figure 3.1B). The K_m and V_{max} of the enzyme is calculated from this curve. V_{max} is the enzyme's maximum specific activity, while K_m is the concentration of substrate where the V_{max} is ½. A small K_m value means an enzyme requires a small amount of substrate to reach saturation, while a large K_m indicates that an enzyme reaches maximum velocity at high substrate concentrations.

Aldolase is encoded by a family of genes in vertebrates, *Aldoa*, *Aldob*, and *Aldoc*. Each aldolase enzyme exhibits unique expression patterns and enzymatic properties. *Aldoa* is expressed ubiquitously and is highly expressed in skeletal muscle, *Aldob* is expressed in the liver and kidney, and *Aldoc* is expressed primarily in the nervous system [3, 4]. The aldolases function as tetramers [5], although studies using recombinant proteins have confirmed active dimers and monomers [6]. In tissues where more than one aldolase isozyme is expressed, tetramers of multiple aldolase isozymes, or heterotetramers, form [4].

Studies of aldolases purified from different tissues demonstrated unique kinetic properties of ALDOA, ALDOB, and ALDOC [3, 7-11]. ALDOA purified from rabbit skeletal muscle or brain possesses the highest V_{max} for FBP, a very high K_m for F1P (only active at high concentrations), and a high FBP/F1P ratio, indicating that ALDOA preferentially catalyzes glycolysis over gluconeogenesis and fructolysis (Table 3.1, adapted from [8]). ALDOB has greater catalytic efficiency for gluconeogenesis and fructolysis, as shown by its lower V_{max} for FBP, higher V_{max} for F1P and lower FBP/F1P ratio compared to ALDOA. It is expressed in the liver and kidney, tissues which promote both reactions. Finally, ALDOC possesses intermediate FBP and F1P activity when compared to ALDOA and ALDOB and has a primary glycolytic role.

We previously identified two additional *Aldoa*-like genes in the mouse genome, *Aldoart1* and *Aldoart2* [12]. Both *Aldoart1* and *Aldoart2* encode proteins that are expressed only in haploid cells during mouse spermatogenesis. Analysis of the *Aldoart1* sequence provided evidence for an alternative splice variant (*Aldoa_v2*) of aldolase A. *Aldoa_v2* is expressed only during spermatogenesis and encodes a protein present in mouse, rat and human sperm. Both *Aldoa_v2* and *Aldoart1* encode N-terminal extensions that appear to be important for tight binding to the fibrous sheath and compartmentalization with other glycolytic enzymes in the principal piece of the sperm flagellum.

In this study we measured the catalytic activity of spermatogenic cell-specific aldolase A isozymes *in vivo* and *in vitro*. Preliminary three-dimensional modeling draws attention to unique structural differences in the outer face and near the substrate binding pocket of these isozymes, suggesting unique binding and catalytic properties. To assess the kinetic properties of each isozyme, we expressed recombinant forms of each protein, monitored tetramer formation and measured FBP catalytic activity. We hypothesize that these novel aldolase A isozymes may have distinctive catalytic properties that promote enhanced glycolytic activity in sperm.

Methods

Isolation of mouse sperm from cauda epididymis

Outbred CD-1 mice were obtained from Charles River (Raleigh, NC). All procedures involving animals were approved by the University of North Carolina at Chapel Hill Animal Care and Use Committee and conducted in accordance with the Guide for the care and Use of Laboratory Animals (Institute for Laboratory Animal Research, National Academy of Sciences).

Mouse sperm was collected as previously described [12]. Briefly, clipped cauda epididymis were incubated for 15 minutes at 37 ℃ in p hosphate-buffered saline with protease inhibitors (PBS + PI) containing 140 mM NaCl, 10 mM phosphate buffer (pH 7.4) and Complete protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). Sperm were washed in PBS+PI to remove debris and counted to determine the final concentration.

SDS-PAGE analysis of mouse sperm protein lysates

In order to measure activity of endogenous ALDOA-related proteins, sperm lysates were prepared using a published protocol [13]. Proteins were extracted in HEM1 buffer (50mM HEPES, 5mM EDTA, 4mM β-mercaptoethanol, 0.05M sodium phosphate, 0.1% TX-100, pH 7.4) containing Complete protease inhibitor cocktail. Cauda epididymal sperm were resuspended in HEM1 buffer at a concentration of 1 x 10^5/μL and sonicated on ice with a pre-chilled probe for six 5-second intervals at 40% output using a Branson Sonifier Cell Disruptor 185 (Danbury, CT). An aliquot was taken for the “lysate” sample. The sample was then centrifuged at 16,000 x g for 10 minutes at 4℃. The supernatant was removed, and the pellet was resuspended in HEM1 extraction buffer. The catalytic activity of samples was measured.

Protein samples were heated at 95℃ in 2X sample load ing buffer (2% SDS, 100mM DTT, 125 mM Tris-HCl pH 6.8, 18% glycerol, 0.002% Bromophenol blue) containing 50mM TCEP reducing agent (*tris*-(2-carboxyethyl) phosphine, Pierce Biotechnology, Rockford, IL), and resolved by 10% SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). The Benchmark™ protein ladder (Invitrogen, Carlsbad, CA) was included on these gels. Protein loads were adjusted to 5 x 10^5 cell equivalents for the lysate and pellet samples and 7.5 x 10^5 cell equivalents for the supernatant sample. Following electrophoresis, samples were electrophoretically transferred to Immobilon-P PVDF (polyvinylidene fluoride) membranes (Millipore Corp, Bedford, MA). Approximate equal loading of samples was confirmed by 0.1% Coomassie blue R250 staining in 45% methanol and 10% acetic acid. Membranes were destained and rinsed with TBS-T (140mM NaCL, 3mM KCl, 0.05% Tween-20, 25mM Tris-HCl, pH 7,4), following by incubation in blocking buffer (5% nonfat dry milk in TBS-T) overnight at 4°C. Immunodetection of aldolase was performed at room temperature. Membranes were incubated with a polyclonal anti-aldolase A primary antibody (Polysciences, Warrington, PA) at a dilution of 1:2,000 in blocking buffer for 1 hour at room temperature. Following three 15-minute washes in TBS-T, membranes were incubated with secondary antibody (affinity-purified horseradish peroxidase-conjugated rabbit anti-goat IgG, KPL, Gaithersburg, MD) at 1:10,000 for 30 minutes at room temperature in blocking buffer. After three 15-minute TBS-T washes, immunoreactive proteins were detected by enhanced chemiluminescence using SuperSignal West Pico substrate (Pierce Biotechnology, Rockford, IL) with HyBlot CL autoradiography film (Denville Scientific, Metuchen, NJ).

Preliminary three-dimensional protein structure modeling

Modeling data was obtained using previously described structures for rabbit aldolase A, since the structure for mouse aldolase A has not been solved. Mouse aldolase A has six

residues that are different when compared to rabbit aldolase A. Of these six residues, two are unique to ALDOART1 and ALDOART2 when compared to mouse ALDOA. Therefore, we marked these two residues as unique and the remaining four residues as identical since they are found in all mouse aldolase A isozymes. In addition to these two unique residues, we also marked 15 unique residues in ALDOART1 and 17 unique residues in ALDOART2. The aldolase monomer model used was 3B8D with a resolution of 2.0 Å [14]. For the aldolase tetramer, we used the 1ZAL structure, with a resolution of 1.8 Å [15]. Modifications were made using PyMOL (DeLano Scientific, Palo Alto, CA).

Expression of recombinant ALDOA-like proteins

We used testis cDNA and cDNA from condensing spermatids [12] to subclone *Aldoa*, *Aldoa_v2*, *Aldoart1*, and *Aldoart2* using primer-based PCR amplification with restriction enzyme sites (*EcoRI*, *XhoI*, and *SalI*) for insertion into an expression vector. For *Aldoa*, the primers were: forward 5'GAATTCATGCCCCACCCATA 3', reverse 5'CTCGAGTTCAATAGCAAGTGG3'. For *Aldoa_v2*, the primers were: forward 5'GAATTCATGGCAACGCGCAG3', reverse 5'CTCGAGTTCAATAGCAAGTGG3'. For *Aldoart1*, the primers were: forward 5'CGGAATTCATGGCAACGCACAGGCA3', reverse 5'AGCGTCGACACATGAGGGCA3'. For *Aldoart2*, the primers were: forward 5'GAATTCATGTCTTACCCCTACC3', reverse 5'CTCGACACCTCTGCTCAGTA3'.

Full-length amplified products and open reading frames were confirmed by DNA sequencing (UNC-Genome Analysis Facility, Chapel Hill, NC). These PCR products were subcloned into a bacterial expression vector (pGEX4T-1, GE Healthcare Life Sciences, Piscataway, NJ) containing an N-terminal GST sequence that can be cleaved using thrombin protease. Constructs were transformed and expressed in *E. coli*, using BL21 competent cells (Novagen, Gibbstown, NJ). For protein expression, single colonies were selected and grown in a 25 ml LB-ampicillin (10g Tryptone, 5g Yeast Extract, 10g NaCl, 50

μg/mL ampicillin) overnight at 37℃, shaking at 210 RP M. This culture was transferred to 225 ml of LB-ampicillin and grown for 1 hour. Once cells reached a density measured as 0.6 at 600 nm, they were induced with 0.1mM IPTG (isopropyl-beta-D-thiogalactopyranoside) for 4 more hours. Cells were pelleted by centrifugation at 6000 x g, 4℃.

To isolate protein, cells were resuspended in PBS (137mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.47 mM KH_2PO_4, pH 7.4) + Complete protease inhibitor cocktail, followed by sonication on ice with a pre-chilled probe for six 5-second intervals at 40% output using Branson Sonifier Cell Disruptor 185 (Danbury, CT). Lysates were treated with 0.1% Triton X-100 on ice for 1 hour, and then cleared by centrifugation at 16,000 x g at 4℃. Washed and pre-swelled Glutathione Sepharose 4B beads (GE Healthcare Life Sciences, Piscataway, NJ) at 50% in PBS +PI were added to cleared lysates and incubated overnight at 4℃ with rocking. Beads were pelleted and washed three times using PBS alone at 4℃, 1,000 x g, for 5 minutes. Thrombin protease (from bovine plasma, Sigma-Aldrich, St. Louis, MO) diluted in PBS was added (at a concentration of 1unit/25μL) directly to the beads to cleave the GST tag and incubated with rocking for three days at 4℃. Cleaved recombinant aldolase A proteins were separated from beads by centrifugation at 1,000 x g for 5 minutes. Beads were washed once with PBS. Both fractions were kept for analysis. Protein concentrations were calculated as previously described [6, 16-20] by measuring the absorbance at 280 nm and calculating the extinction coefficient for each ALDOA-related protein using the ProtParam tool on ExPASy (Swiss Institute of Bioinformatics, Switzerland, http://ca.expasy.org/tools/protparam.html). The following extinction coefficients were calculated: 0.874 (ALDOA), 0.762 (ALDOA_V2), 0.725 (ALDOART1) and 0.911 (ALDOART2). Purity of recombinant proteins and equal loading was confirmed by 10% SDS-PAGE followed by Coomassie R250 staining (described above).

Native PAGE analysis of tetramer formation

Recombinant proteins were diluted in 2X native sample loading buffer (125 mM Tris-HCl pH 6.8, 18% glycerol, 0.002% bromophenol blue) and resolved by native PAGE. Equal protein amounts (0.5 μg) were resolved on precast 4-15% gradient gels (Bio-Rad laboratories, Hercules, CA) at 100V for 4 hours on ice. Gels were fixed for 1 hour in 45% methanol, 10% acetic acid, followed by staining in Sypro Ruby (Molecular Probes, Eugene, OR) overnight. Gels were destained in 10% methanol, 7% acetic acid (2 x 30 minutes). All staining procedures were done at room temperature with gentle agitation. Gels were visualized using UV illumination.

Measuring aldolase FBP activity

Catalytic activity of endogenous and recombinant proteins was measured using an enzymatic assay for aldolase (Sigma-Aldrich, St. Louis, MO). The presence of either 10^6 sperm equivalent of protein or 100ng of active aldolase in combination with 2.6mM fructose-1,6-bisphosphate (Sigma F-6803), α-glycerophosphate dehydrogenase (α-GDH), triose phosphate isomerase (TPI), and 29 mM β-NADH (Sigma N-8129) will oxidize β-NADH to β-NAD (measured at A_{340nm}, Figure 3.1). α-GDH and TPI were added as a mixture at 25 units/mL (Sigma G-1881). All reagents were purchased from Sigma-Aldrich (St. Louis, MO) and all solutions were prepared in 100mM Tris-HCl, pH 7.4.

To correct for background activity, we refrained from adding the enzyme to each reaction and allowed for the reaction to reach zero activity (approximately 10 minutes). We then added the enzyme and measure the change (slope) in β-NADH absorption over 10 minutes. Kinetic parameters were calculated by measuring the velocity (slope) of the enzyme during the linear phase of the curve at various substrate concentrations. To

calculate the K_m and V_{max} we used the Michaelis-Menten equation ($Y=V_{max}X/(K_m+X)$, where X is the substrate concentration and Y is the velocity) in GraphPad Prism4 (La Jolla, CA).

Results

Fructose-1,6-bisphosphate activity is found in insoluble fraction of sperm lysate

Previous studies of purified bovine sperm aldolase demonstrated maximum specific activity in a Triton X-100 and β-mercaptoethanol-containing buffer [13]. Based upon this data, we extracted mouse sperm using the same buffer and conditions. We detected aldolase isozymes with two molecular weights in both the supernatant and pellet fraction (Figure 3.2A). Our proteomic analyses (Chapter 2) indicate that the smaller molecular weight immunoreactive band contains ALDOART2, while the larger molecular weight band contains ALDOA_V2 and ALDOART1. The pellet fraction contains more of the larger molecular weight ALDOA-related proteins.

By measuring activity in the total lysate and the fractionated samples, we see that the kinetic values for FBP aldolase activity in the supernatant fraction are similar to the previously described values for purified bovine sperm aldolase (Table 3.2 and [13]). However, most of the endogenous aldolase FBP activity in the total lysate is found in the insoluble pellet fraction (Figure 3.2B), where ALDOA_V2 and ALDOART1 are found. These data also match studies using bovine sperm aldolase A [13]. Since the sperm lysates and fractions contain multiple ALDOA isozymes, it would be advantageous to measure activity of the individual isozymes.

Individual spermatogenic-cell specific aldolase A isozymes possess amino acid residues which may confer unique activity

Proteomic analysis indicates that mouse sperm lysates contain at least three, and possibly four, ALDOA-related isozymes [12]. ALDOA_V2 is also expressed in rat and

human sperm [12]. Amino acid alignment of all ALDOA-related proteins expressed in testis demonstrates conservation of all known active site residues [12]. The only difference between ALDOA and ALDOA_V2 is the presence of a 54 amino acid N-terminal extension in ALDOA_V2 [12]. Although ALDOART1 and ALDOART2 are 95% identical to ALDOA, each has several residues that are more like ALDOB than ALDOA (green in Fig. 3.3) and three of these have been identified as isozyme-specific residues (arrows, Figure 3.3 and [21]). ALDOB catalyzes gluconeogenesis (the reverse reaction of glycolysis) and fructolysis more efficiently than ALDOA or ALDOC (Table 3.1, [8]). Preliminary modeling data of aldolase monomers shows the position of unique residues (red) and ALDOB-like residues (light blue) in ALDOART1 and ALDOART2, located to the outside of the protein structure or surrounding the substrate binding pocket (dark blue, Figure 3.4, modified from [22]). Unique peripheral amino acid residues may participate in proper localization in the sperm flagellum. Residues that are located closer to the active site, including the three ALDOB isozyme-specific residues in ALDOART1, may contribute to distinct catalytic activity. Differences in amino acid residues may direct unique localization and/or catalytic function of isozymes.

Recombinant ALDOA proteins may not form tetramers and have reduced activity

To determine the catalytic activity of ALDOA-related proteins in mouse sperm, we expressed recombinant forms of each isozyme in *E. coli*. *In vivo*, the aldolases function as tetramers, and heterotetramers have been observed in tissues that express multiple isozymes [4]. We measured activity using GST-tagged ALDOA and ALDOA_V2 (Figure 3.5). In general, ALDOA expresses at much higher levels in *E. coli* than the sperm ALDOA isozymes. Despite conservation of all active site residues and 100% amino acid sequence identity with exception to the N-terminal extension, ALDOA_V2 FBP activity was lower than ALDOA activity. The V_{max} was lower and the K_m was dramatically higher for GST-ALDOA_V2 (Table 3.3). Since aldolase functions as a tetramer, we hypothesized that the

presence of the GST-tag might interfere with tetramer formation of any GST-ALDOA-related proteins. In fact, the N-terminal of ALDOA extends into the center of the tetramer structure (highlighted in red in Figure 3.6). Based upon these results, we measured activity of ALDOA-related proteins following cleavage of the GST tag.

We used native gel electrophoresis to detect tetramer formation of all recombinant ALDOA-related proteins following cleavage of the GST tag. While we loaded equal amounts of protein as calculated by extinction coefficients, the amount detected by Sypro stain appears to show greater amounts of ALDOA protein (Figure 3.7A). Out attempt to use native gel electrophoresis standards (NativeMark, Invitrogen, Carlsbad, CA) to identify tetramers by molecular weight was unsuccessful. A very large molecular weight protein species is present only in the ALDOA lane. Therefore, only ALDOA formed what appears to be a tetramer, while ALDO_V2, ALDOART1, and ALDOART2 formed dimers and monomers (Figure 3.7A). This lack of tetramer formation corresponded with a dramatic reduction in FBP aldolase activity for ALDOA_V2, ALDOART1, and ALDOART2 (Figure 3.7B).

Discussion

Glycolysis is essential for mammalian sperm motility and male fertility [23]. The glycolytic pathway is modified during mammalian spermatogenesis through the expression of novel spermatogenic-cell specific isozymes [12, 24, 25]. Mammalian sperm express multiple isozymes of aldolase A, the glycolytic enzyme which cleaves fructose-1,6-bisphosphate. In mouse sperm at least three ALDOA-related proteins are present, ALDOA_V2, ALDOART1, and ALDOART2 [12]. While *Aldoart1* and *Aldoart2* are only present in the mouse genome, the alternatively spliced exon present in *Aldoa_v2* is conserved across species, and ALDOA_V2 protein is detected in human and rat sperm [12]. Both ALDOA_V2 and ALDOART1 possess N-terminal extensions which may facilitate tight

binding to the fibrous sheath, a highly insoluble structure where multiple glycolytic enzymes are localized [26].

We detected high levels of aldolase FBP activity in the insoluble fraction from mouse sperm lysates. The K_m value is close to previously published report for mouse sperm aldolase A activity [27]. The kinetic values of soluble aldolase FBP activity from mouse sperm were comparable to levels from bovine sperm [13]. Previous measurements of bovine sperm aldolase excluded activity from insoluble fractions (which may contain novel isozymes) and only measured 10% of the total activity in sperm [13]. Unique solubility properties of bovine sperm aldolase in various detergents and buffers suggest that there are tight stable interactions with other proteins and subcellular structures [13]. The high level of activity found in the insoluble fraction correlates with a greater amount of the larger ALDOA-related proteins in this fraction. Since we know the fibrous sheath is highly insoluble and does not solubilize in Triton X-100-containing buffer, the high levels of aldolase FBP activity may be tightly bound to the fibrous sheath where many other glycolytic enzymes are present [26]. In fact, proteomic data has previously identified both ALDOA_V2 and ALDOART1 tightly bound to the fibrous sheath [12]. Components of the fibrous sheath may act as a scaffold [28] and may be required for tetramer formation and/or sperm aldolase A activity.

In order to measure activity of the individual ALDOA-like proteins expressed in sperm, we expressed these protein using an *E. coli* system. Modeling data suggested that ALDOART1 and ALDOART2 might have unique kinetic properties due to modifications near the active site. However, sperm ALDOA-related proteins do not appear to form tetramers when expressed in *E. coli*, and this lack of tetramer formation may contribute to reduced FBP activity. Previous data using recombinant ALDOA demonstrated tetramer formation even at very low concentrations (0.2ug/ml) and reported that only very high temperatures disrupt tetramer formation [6, 29]. However, aldolase A monomers and dimers are still active, suggesting that tetramer formation plays a role in protein stability rather than activity

[6]. It is possible that the N-terminal extensions present in ALDOA_V2 and ALDOART1 may interfere with tetramer formation. However, ALDOART2, which does not contain an N-terminal extension, also does not form tetramers when expressed in *E. coli*.

Another possibility is that the sperm ALDOA-related proteins are not properly folded, which is required for tetramer formation and activity. The sperm ALDOA-related proteins were soluble using an *E. coli* expression system. We assessed quaternary structure formation through native gel electrophoresis, but we did not confirm accurate secondary and tertiary structure formation to confirm proper folding. Circular dichroism may prove useful for verifying proper folding of ALDOA-related enzymes expression in *E. coli.*

The sperm ALDOA-related proteins may require post-translational modifications for proper folding and tetramer formation, which cannot be completed in an *E. coli* expression system. In fact, many proteins expressed in sperm are known to form complex disulfide bonds not found in any other cell type. One example is the spermatogenic cell-specific glycolytic enzyme, hexokinase (HK1S). This protein contains a novel 5' exon which encodes an N-terminal domain, the SSR domain. A cysteine residue located in the SSR domain may be responsible for disulfide bond-dependent dimer formation of HK1 [30]. In the case of aldolase, ALDOART1 has two unique cysteine residues not found in ALDOA (C58 and C253) [12]. Whether disulfide bond formation and other post-translational modifications are required for proper folding, tetramer formation, and subsequent activity of the sperm ALDOA-related isozymes has not been addressed.

Expression of sperm ALDOA-related proteins in mammalian cell systems may promote proper folding and tetramer formation. Once tetramers are formed, both FBP and F1P activity can be measured to determine the role of sperm aldolase in glycolysis and fructolysis. There is some evidence of fructolytic activity in mouse and human sperm [31-33], and fructose is found at very high levels in human seminal plasma [34]. Previously, endogenous aldolase activity was measured using enzyme purified from tissue. It would be

difficult to separate the four aldolase isozymes in mouse sperm since they are highly insoluble. The sperm ALDOA-related isozymes may be soluble when they are first expressed during spermatogenesis in the testis [12]. With that in mind, one possibility is to measure activity using testicular lysates combined with gel filtration chromatography to distinguish between activity from larger sperm ALDOA-like proteins (ALDOA_V2 and ALDOART1) and activity from ALDOART2.

Sperm ALDOA-related proteins may not exhibit unique catalytic activity and instead may play a primary role in facilitating proper localization of aldolase activity to the site of glycolytic ATP production in sperm. Tight binding to the fibrous sheath may be accomplished through the presence of N-terminal extension. Only ALDOA_V2 is conserved across species, again supporting a functional role for the N-terminal extension. In fact, we see a distinct trend when we look at multiple glycolytic enzymes localized to the fibrous sheath. GAPDHS, ALDOA_V2, ALDOART1, and a newly identified splice variant of lactate dehydrogenase, LDHA_V2, are all tightly bound to the fibrous sheath and contain N-terminal extensions when compared to the somatic protein ([12, 26] and unpublished data). While there is no obvious sequence similarity between these N-terminal extensions, the enzymes in the glycolytic pathway are being modified through the addition of N-terminal extensions which may facilitate tight binding to the fibrous sheath, the site of glycolytic ATP production.

We detected high levels of aldolase FBP activity in insoluble fractions of mouse sperm lysate. Although we were unable to determine the kinetic properties of individual aldolase isozymes, difficulties experienced when expressing these proteins in an *E. coli* system may highlight distinctive attributes of sperm ALDOA-related proteins. Defining the kinetic properties of aldolase A isozymes will help to understand the regulation of sperm metabolism in the context of proper male fertility.

Figure 3.1

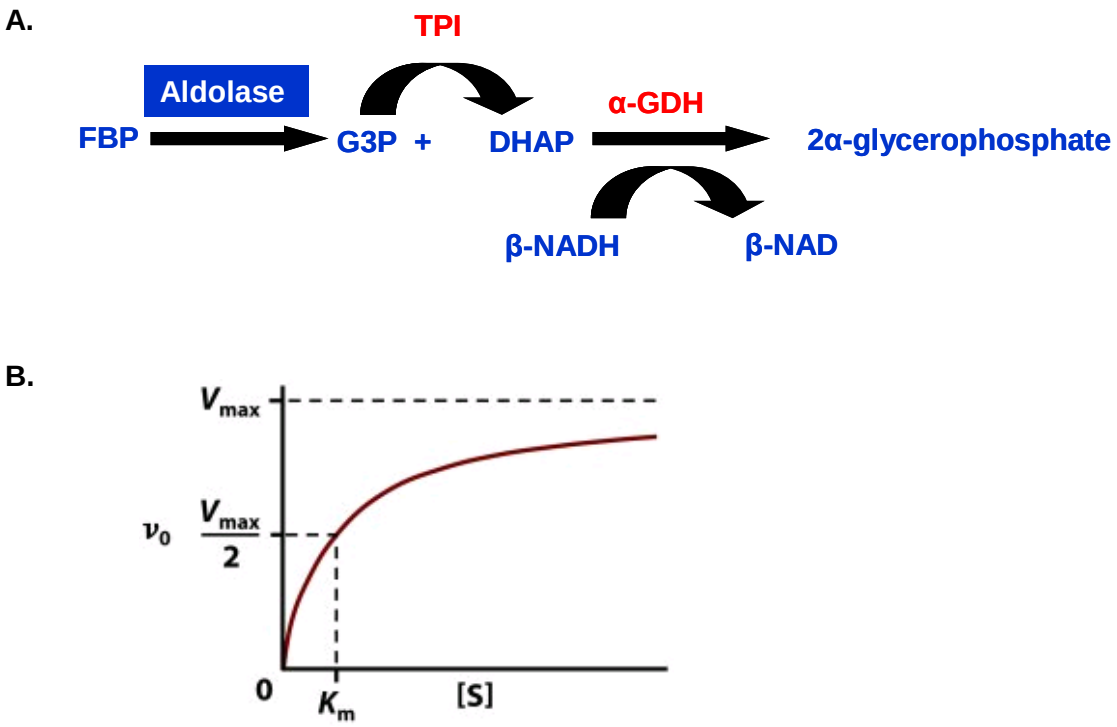


Figure 3.1. Enzymatic assay for FBP aldolase activity and definition of kinetic parameters. (A) Diagram depicting the enzymatic reaction used to measure FBP aldolase activity. The substrate, fructose-1,6-bisphosphate (FBP) is cleaved by Aldolase and produces glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP). G3P is converted to DHAP by triose phosphate isomerase (TPI). DHAP is then metabolized by α-glycerophosphate dehydrogenase (α-GDH) to produce 2α-glycerophosphate. This final step oxidizes β-NADH to β-NAD, which can be read by the decrease in absorbance at 340nm, therefore measuring FBP aldolase activity. (B) Graph depicting the calculation of kinetic parameters K_m and V_{max}. By measuring the velocity of an enzymatic reaction (v_0) over varying concentrations of substrate [S], one can plot an exponential curve using the Michaelis-Menten equation, $Y=V_{max}X/(K_m+X)$, where X is the substrate concentration [S] and Y is the velocity (v_0). This equation is used to calculate the V_{max}, or the maximum velocity of the reaction, and the K_m, the concentration of substrate at which the Km is one-half.

Figure 3.2

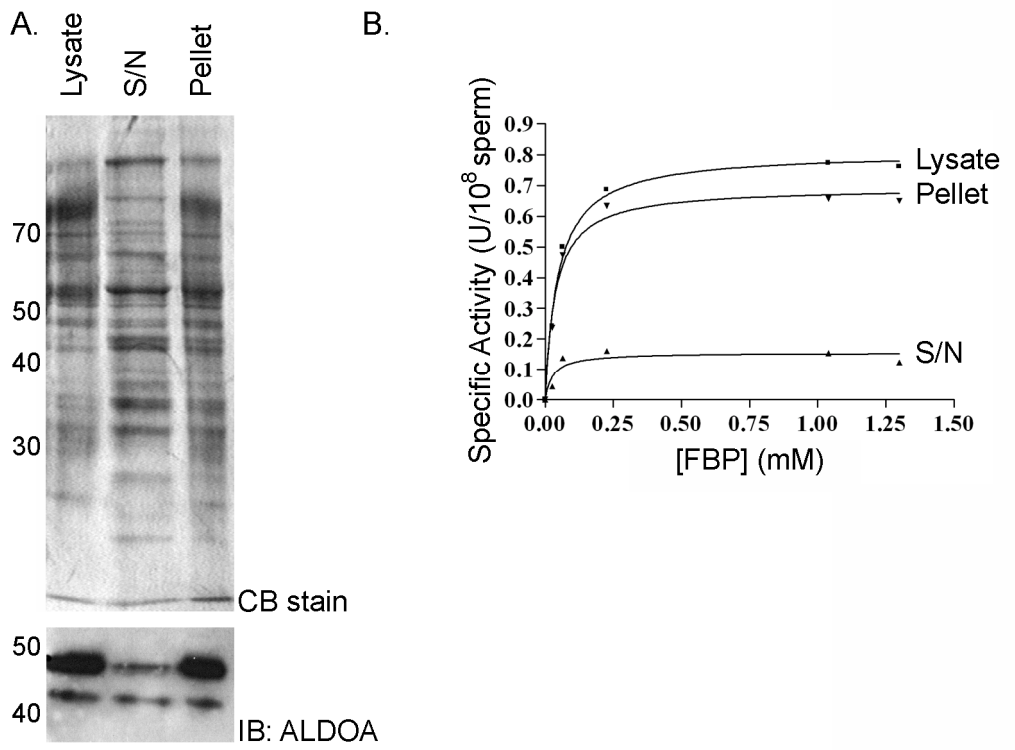


Figure 3.2. Kinetic analysis of endogenous mouse sperm FBP aldolase. (A) Protein from sperm isolated in 1% Triton-X-100-containing buffer and ß-mercaptoethanol were separated by centrifugation. Protein loads were adjusted to 5 x 10^5 cell equivalents for the lysate and pellet samples and 7.5 x 10^5 cell equivalents for the supernatant sample. Total protein lysate, supernatant (S/N) and pellet were resolved by 10% SDS-PAGE, transferred to PVDF and stained with coomassie blue to demonstrate equal loading. The same samples were processed for western blotting using an anti-aldolase antibody. Larger amounts of the 50kDa aldolase species are found in the pellet fraction. (B) Michaelis-Menten curves showing FBP aldolase activity of the three sperm samples. Specific activity is measure as units per 10^8 sperm. For values of kinetic parameters see Table 3.2.

Figure 3.3

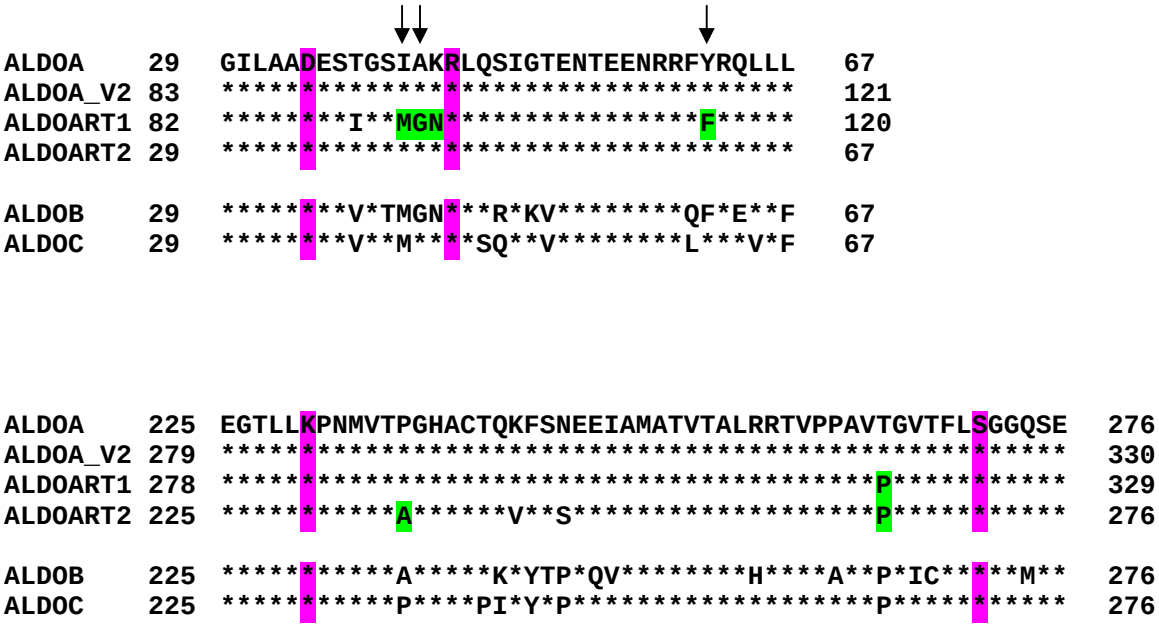


Figure 3.3. ALDOB-like residues found in ALDOART1 and ALDOART2 sit in between active site residues. Amino acid alignment of all ALDO-related proteins highlighting four of the eight known active site residues [14]. Asterisks (*) denotes identical amino acid residues. Amino acids highlighted in pink are active site residues, conserved in all ALDO-related proteins. Residues highlighted in green are unique to ALDOART1 and ALDOART2 and more similar to ALDOB. Three isozyme-specific residues in ALDOB are found in ALDOART1 (denoted by black arrows).

Figure 3.4

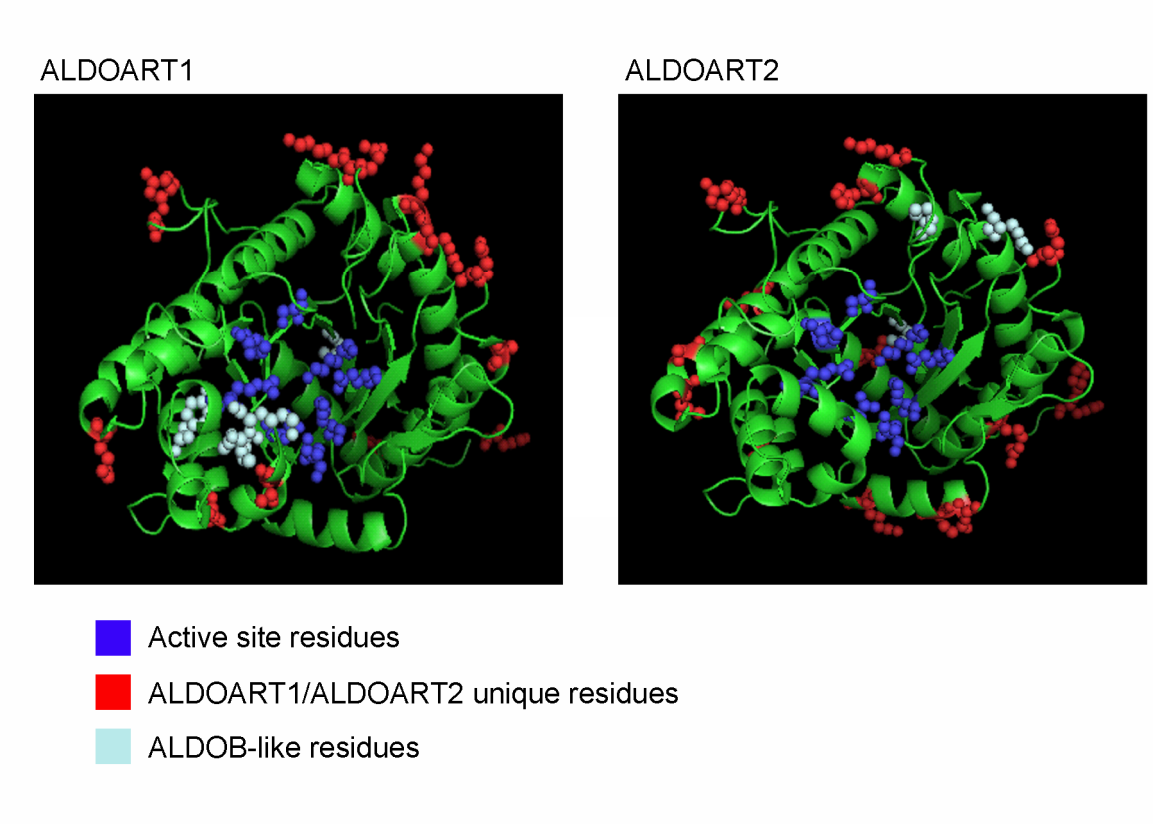


Figure 3.4. Unique amino acid residues in ALDOART1 and ALDOART2 are found on the outside of the protein or near the substrate binding pocket. Three-dimensional protein structure models highlighting the differences between ALDOA isozymes expressed in sperm. Using the ALDOA monomer structure, we highlighted active site residues (blue, [14]) and unique residues in ALDOART1 and ALDOART2 (red). We also highlighted residues in ALDOART1 and ALDOART2 which are identical to ALDOB (light blue).

Figure 3.5

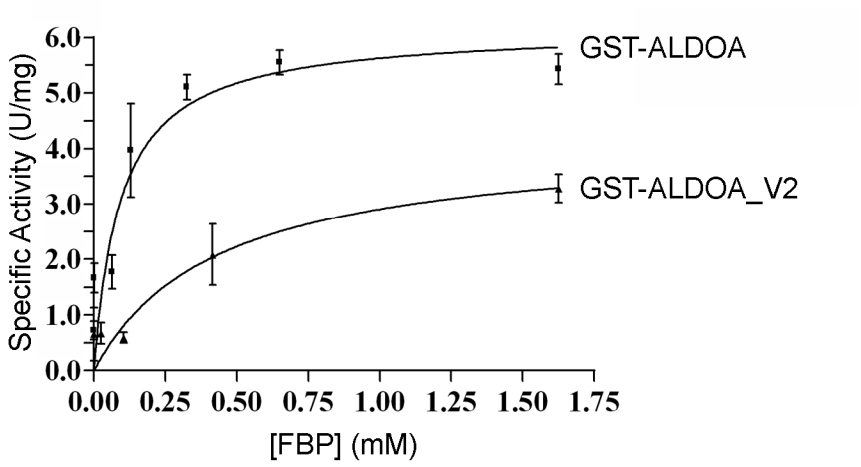


Figure 3.5. GST-ALDOA_V2 has reduced FBP aldolase activity when compared to GST-ALDOA. Michaelis-Menten curves showing FBP aldolase activity of recombinant GST-ALDOA and GST-ALDOA_V2. Specific activity is shown as units per mg of recombinant protein. For values of kinetic parameters see Table 3.3.

Figure 3.6

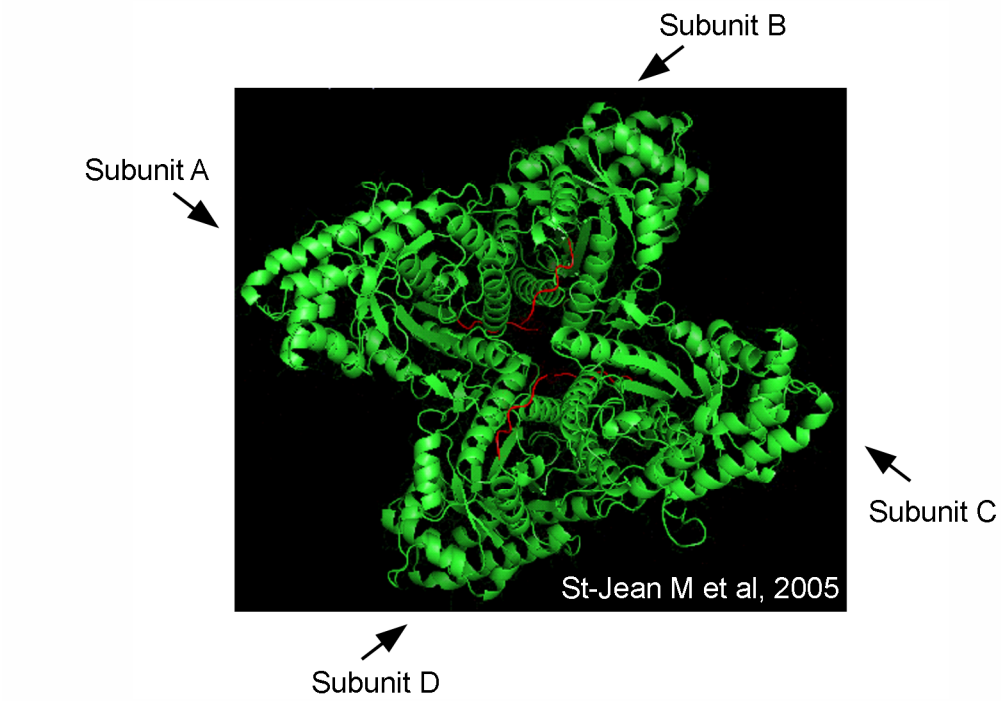


Figure 3.6. The N-terminus of ALDOA sits in the center of the tetramer structure. Three-dimensional protein structure of aldolase A tetramer. Image modified from [15]. Each subunit of the tetramer is labeled A-D. Highlighted in red are the first seven residues (excluding methionine not included in the structure) of the N-terminus of ALDOA.

Figure 3.7

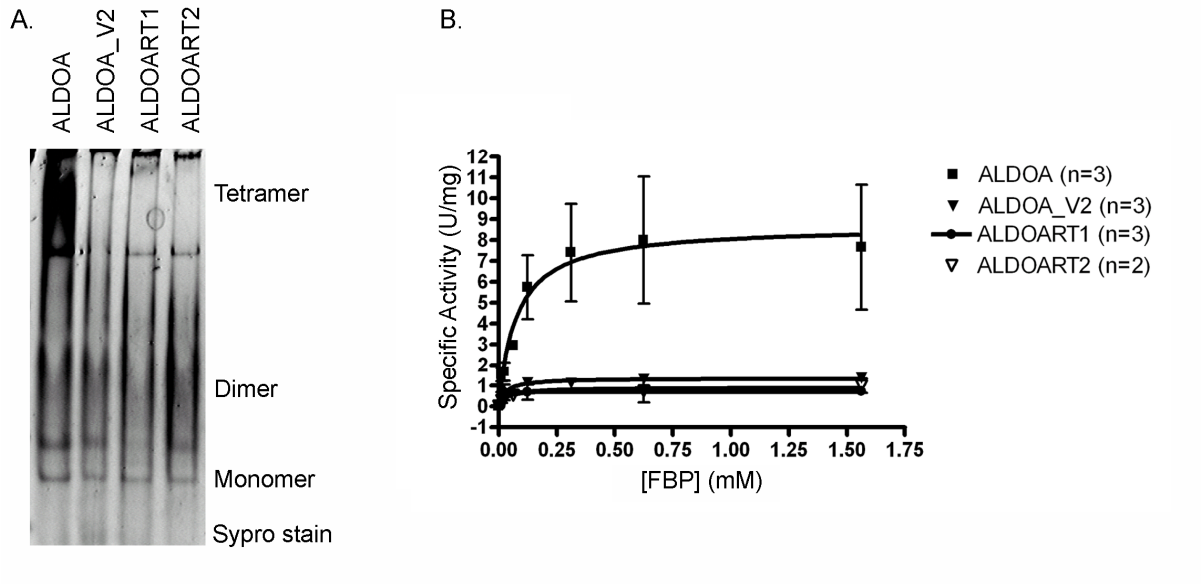


Figure 3.7. Spermatogenic-cell specific ALDOA-related recombinant proteins do not form tetramers and have reduced activity when expressed in *E. coli*. (A) Native gel electrophoresis of ALDOA-related recombinant proteins followed by Sypro Ruby staining. Equal protein load was measured by absorbance, although Sypro stain appears to show varying protein load. Large molecular weight protein species may represent tetramers, only present in ALDOA lane. (B) Michaelis-Menten curve showing FBP aldolase activity of recombinant spermatogenic-cell specific ALDOA-related proteins. Specific activity is measured as units per mg of recombinant protein. ALDOA_V2, ALDOART1, and ALDOART2 have dramatically reduced activity when compared to ALDOA. Number in parenthesis next to protein name in key denotes the number of protein preparations used.

Table 3.1. *Kinetic properties of purified rabbit aldolase isozymes [8]*

Aldolase	A	B	C
Tissue source	Muscle (Brain)	Liver	Brain
K_m FBP (μM)	4	1	2
V_{max} FBP (units/mg)	16.5	1.0	6.5
K_m F1P (μM)	5000	300	4000
V_{max} F1P (units/mg)	0.33	1.0	0.65
FBP/F1P ratio (V_{max} FBP/V_{max} F1P)	50	1	10

V_{max} is the enzyme's maximum rate, while K_m is the concentration of substrate where the V_{max} is ½.

Table 3.2. *Kinetic properties of endogenous mouse sperm aldolase isozymes*

Sperm sample	Lysate	S/N	Pellet	Purified [13]
K_m FBP (μM)	47	28	37	7
V_{max} FBP (units/10^8 sperm)	0.81	0.16	0.69	0.104

V_{max} is the enzyme's maximum rate, while K_m is the concentration of substrate where the V_{max} is ½.

Table 3.3. *Kinetic properties of recombinant sperm aldolase isozymes*

GST protein	ALDOA	ALDOA_V2
K_m FBP (μM)	94	430
V_{max} FBP (units/mg)	6.2	4.2

V_{max} is the enzyme's maximum rate, while K_m is the concentration of substrate where the V_{max} is ½.

References

1. Richards OC, Rutter WJ: **Preparation and properties of yeast aldolase**. *J Biol Chem* 1961, **236**:3177-3184.

2. Blostein R, Rutter WJ: **Comparative Studies of Liver and Muscle Aldolase. Ii. Immunochemical and Chromatographic Differentiation**. *J Biol Chem* 1963, **238**:3280-3285.

3. Rutter WJ, Rajkumar T, Penhoet E, Kochman M, Valentine R: **Aldolase variants: structure and physiological significance**. *Ann N Y Acad Sci* 1968, **151**(1):102-117.

4. Penhoet E, Rajkumar T, Rutter WJ: **Multiple forms of fructose diphosphate aldolase in mammalian tissues**. *Proc Natl Acad Sci U S A* 1966, **56**(4):1275-1282.

5. Penhoet E, Kochman M, Valentine R, Rutter WJ: **The subunit structure of mammalian fructose diphosphate aldolase**. *Biochemistry* 1967, **6**(9):2940-2949.

6. Beernink PT, Tolan DR: **Disruption of the aldolase A tetramer into catalytically active monomers**. *Proc Natl Acad Sci U S A* 1996, **93**(11):5374-5379.

7. Lebherz HG, Rutter WJ: **Distribution of fructose diphosphate aldolase variants in biological systems**. *Biochemistry* 1969, **8**(1):109-121.

8. Penhoet EE, Rutter WJ: **Catalytic and immunochemical properties of homomeric and heteromeric combinations of aldolase subunits**. *J Biol Chem* 1971, **246**(2):318-323.

9. Penhoet EE: **Isolation of fructose diphosphate aldolases A, B, and C.** *Biochemistry* 1969, **8**(11):4391-4395.

10. Penhoet EE, Rutter WJ: **Detection and isolation of mammalian fructose-diphosphate aldolases**. *Methods Enzymol* 1975, **42**:240-249.

11. Penhoet EE, Kochman M, Rutter WJ: **Molecular and catalytic properties of aldolase C**. *Biochemistry* 1969, **8**(11):4396-4402.

12. Vemuganti SA, Bell TA, Scarlett CO, Parker CE, de Villena FP, O'Brien DA: **Three male germline-specific aldolase A isozymes are generated by alternative splicing and retrotransposition**. *Dev Biol* 2007, **309**(1):18-31.

13. Gillis BA, Tamblyn TM: **Association of bovine sperm aldolase with sperm subcellular components**. *Biol Reprod* 1984, **31**(1):25-35.

14. Arakaki TL, Pezza JA, Cronin MA, Hopkins CE, Zimmer DB, Tolan DR, Allen KN: **Structure of human brain fructose 1,6-(bis)phosphate aldolase: linking isozyme structure with function**. *Protein Sci* 2004, **13**(12):3077-3084.

15. St-Jean M, Lafrance-Vanasse J, Liotard B, Sygusch J: **High resolution reaction intermediates of rabbit muscle fructose-1,6-bisphosphate aldolase: substrate cleavage and induced fit**. *J Biol Chem* 2005, **280**(29):27262-27270.

16. Baranowski T, Niederland TR: **Aldolase activity of myogen A**. *J Biol Chem* 1949, **180**(2):543-551.

17. Morris AJ, Tolan DR: **Site-directed mutagenesis identifies aspartate 33 as a previously unidentified critical residue in the catalytic mechanism of rabbit aldolase A**. *J Biol Chem* 1993, **268**(2):1095-1100.

18. Tolan DR, Schuler B, Beernink PT, Jaenicke R: **Thermodynamic analysis of the dissociation of the aldolase tetramer substituted at one or both of the subunit interfaces**. *Biol Chem* 2003, **384**(10-11):1463-1471.

19. Choi KH, Shi J, Hopkins CE, Tolan DR, Allen KN: **Snapshots of catalysis: the structure of fructose-1,6-(bis)phosphate aldolase covalently bound to the substrate dihydroxyacetone phosphate**. *Biochemistry* 2001, **40**(46):13868-13875.

20. Lubini DG, Christen P: **Paracatalytic modification of aldolase: a side reaction of the catalytic cycle resulting in irreversible blocking of two active-site lysyl residues**. *Proc Natl Acad Sci U S A* 1979, **76**(6):2527-2531.

21. Pezza JA, Choi KH, Berardini TZ, Beernink PT, Allen KN, Tolan DR: **Spatial clustering of isozyme-specific residues reveals unlikely determinants of isozyme specificity in fructose-1,6-bisphosphate aldolase**. *J Biol Chem* 2003, **278**(19):17307-17313.

22. Maurady A, Zdanov A, de Moissac D, Beaudry D, Sygusch J: **A conserved glutamate residue exhibits multifunctional catalytic roles in D-fructose-1,6-bisphosphate aldolases**. *J Biol Chem* 2002, **277**(11):9474-9483.

23. Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility**. *Proc Natl Acad Sci U S A* 2004, **101**(47):16501-16506.

24. Boer PH, Adra CN, Lau YF, McBurney MW: **The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon**. *Mol Cell Biol* 1987, **7**(9):3107-3112.

25. Welch JE, Schatte EC, O'Brien DA, Eddy EM: **Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells**. *Biol Reprod* 1992, **46**(5):869-878.

26. Krisfalusi M, Miki K, Magyar PL, O'Brien D A: **Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa**. *Biol Reprod* 2006, **75**(2):270-278.

27. Arcelay E, Salicioni AM, Wertheimer E, Visconti PE: **Identification of proteins undergoing tyrosine phosphorylation during mouse sperm capacitation**. *Int J Dev Biol* 2008, **52**(5-6):463-472.

28. Eddy EM, Toshimori K, O'Brien DA: **Fibrous sheath of mammalian spermatozoa**. *Microsc Res Tech* 2003, **61**(1):103-115.

29. Lebherz HG: **Stability of quaternary structure of mammalian and avian fructose diphosphate aldolases**. *Biochemistry* 1972, **11**(12):2243-2250.

30. Nakamura N, Miranda-Vizuete A, Miki K, Mori C, Eddy EM: **Cleavage of disulfide bonds in mouse spermatogenic cell-specific type 1 hexokinase isozyme is associated with increased hexokinase activity and initiation of sperm motility**. *Biol Reprod* 2008, **79**(3):537-545.

31. Fraser LR, Quinn PJ: **A glycolytic product is obligatory for initiation of the sperm acrosome reaction and whiplash motility required for fertilization in the mouse**. *J Reprod Fertil* 1981, **61**(1):25-35.

32. Hoppe PC: **Glucose requirement for mouse sperm capacitation in vitro**. *Biol Reprod* 1976, **15**(1):39-45.

33. Mann T, R. J, Sherins R: **Fructolysis in human spermatozoa under normal and pathological conditions**. *Journal of Andrology* 1980, **1**(5):229-233.

34. Peterson RN, Freund M: **Glycolysis by washed suspensions of human spermatozoa. Effect of substrate, substrate concentration, and changes in medium composition on the rate of glycolysis**. *Biol Reprod* 1969, **1**(3):238-246.

CHAPTER 4

FREQUENT RETROTRANSPOSITION OF ORTHOLOGOUS GENES IN THE GLYCOLYTIC PATHWAY

Abstract

In the central metabolic pathway of glycolysis, each reaction is catalyzed by multiple isozymes encoded by a multigene family. Gene targeting studies of multiple spermatogenic cell-specific glycolytic isozymes indicate that glycolysis is essential for sperm function and male fertility in mouse. In addition to *Pgk2*, a well-characterized retrogene expressed in all eutherian mammals, at least two other spermatogenic cell-specific glycolytic enzymes in the mouse (*Aldoart1*, and *Aldoart2*) are encoded by retrogenes. The restricted expression of these retrogenes during spermatogenesis suggests that retrotransposition may play a significant role in the evolution of sperm glycolytic enzymes. We conducted a comprehensive genomic analysis of glycolytic enzymes in the human and mouse genomes and identified several intronless copies for all enzymes in the pathway, except *Pfk*. Within each gene family, a single orthologous gene was typically retrotransposed frequently and independently in both species. Sequence alignment of each retroposed sequence against the parent mRNA transcript identified those sequences with alternatively spliced exons and/or open reading frames. Retroposed sequences that maintained open reading frames and <99% sequence identity in the coding region were tested for expression in the testis and in isolated germ cells. These studies provided evidence for the expression of an alternative *Gpi1* transcript in mouse spermatogenic cells. A striking number of related sequences were

immediately adjacent to autonomous repetitive elements (LINEs and LTRs) known to be actively retrotransposed in the germline. Our analysis detected frequent, recent, and lineage-specific retrotransposition of orthologous glycolytic enzymes in the human and mouse genomes. Retrotransposition events are LINE/LTR-dependent and genomic integration is random. Many retroposed sequences have maintained open reading frames, suggesting a functional role for these genes.

Introduction

Although glycolysis is highly conserved, this central metabolic pathway is modified extensively during spermatogenesis. There are several glycolytic isozymes with restricted expression in the male germline including glyceraldehyde 3-phosphate dehydrogenase-S (GAPDHS) [1, 2], phosphoglycerate kinase 2 (PGK2) [3], lactate dehydrogenase-C (LDHC) [4, 5] and two aldolase A-related isozymes, ALDOART1/ALDOART2 in mouse [6]. Other unique sperm isozymes in this pathway are generated by alternative splicing, including hexokinase-1 (HK1S*)*; [7-9], ALDOA_V2 [6], and a pyruvate kinase muscle form isozyme (PK-S) [10, 11]. There is also evidence that other glycolytic enzymes have unique functional or structural properties in mammalian sperm, including glucose phosphate isomerase (GPI1) [12, 13], triose phosphate isomerase (TPI) [14], enolase (ENO) [15-17], and phosphofructokinase (PFK) [18]. Targeted gene disruption of genes encoding three spermatogenic cell-specific glycolytic enzymes (*Gapdhs*, *Pgk2*, and *Ldhc*) demonstrates an essential role of these enzymes in sperm motility and male fertility [19-21].

The glycolytic pathway is comprised of ten enzymes, each encoded by a family of genes [10]. Seven of these gene families have two to five intron-containing genes, while the *Gpi1*, *Tpi1*, and *Pgk* families each have only one. Within a family, each gene encodes a different isoform with a unique expression pattern [10]. Many of these gene families arose by multiple rounds of gene duplication in the last 100 million years [10]. However, the *Pgk*

gene family expanded through retrotransposition [3, 22]. In addition, we recently identified two aldolase A (*Aldoa*) retrogenes in the mouse genome [6].

Spermatogenic cell-specific glycolytic isozymes are expressed from genes generated by either gene duplication (*Gapdhs, Ldhc*) or retrotransposition (*Pgk2, Aldoart1, Aldoart2*) [3, 6, 22-24]. *Pgk2* represent an ancient retrotransposition event shared by all eutherian mammals, while *Aldoart1* and *Aldoart2* are only found in the rodent lineage and are much more recent [6, 25]. In addition, frequent retrotransposition of the *Gapdh* and *Aldoa* genes has been reported in both human and mouse, based on an abundance of pseudogenes [26-29].

Theoretically, retrotransposition can occur in any cell type, but the retrotransposition event is only transmitted to future generations when it takes place in the germline [30-33]. Retrotransposition is facilitated by repetitive elements (including LINE and LTR elements) resulting in the creation of pseudogenes or retrogenes [34]. In the human lineage most LTR elements have been extinct for over 40 million years. However, LINE elements are still active and are therefore thought to be responsible for most retroposed mRNA sequences [34]. The proteins encoded by LINE elements, ORF1 and ORF2, are expressed in testicular germ cells undergoing meiosis (spermatocytes), a period when retrotransposition is thought to occur [30, 33, 35]. In fact, retrotransposition is responsible for the creation of many retrogenes translated only during the haploid phase of spermatogenesis, including but not limited to *Pgk2*, *G6pd*, and *Pabp2* [3, 22, 36, 37]. At least 10% of retroposed sequences with open reading frames are transcribed during spermatogenesis [35, 38-40]. Positive selection of sperm proteins, combined with frequent retrotransposition to create genes encoding sperm-specific proteins, results in the faster evolution of genes involved in sperm function [41].

Sperm motility is dependent upon the production of high levels of ATP in the flagellum [42-44]. It has been shown that glycolysis, rather than mitochondrial oxidative

phosphorylation, is required for sperm motility and male fertility in mice [19, 45]. Fertilization is an evolutionarily fine-tuned process that requires the successful completion of a cascade of events, including changes in sperm motility required for reaching the site of fertilization in the female reproductive tract and penetrating the zona pellucida surrounding the oocyte. A recent study of 1085 patients with male factor infertility found that approximately 81% exhibit defects in sperm motility, with 19% having no other defects in sperm count or morphology [46]. The expression of genes that promote high sperm motility can increase reproductive fitness, while disruptive mutations in genes essential for sperm motility can hinder proper fertilization, leading to infertility. In humans, genes involved in spermatogenesis and sperm motility demonstrate the strongest evidence for positive selection, and proteins involved in reproduction are among the most rapidly evolving genes across multiple species [47, 48].

Based on the existence of *Pgk2* and *Aldoa*-related retrogenes and their restricted expression during spermatogenesis, we hypothesized that there may be additional retrogenes that encode novel sperm glycolytic enzymes. Therefore, we conducted a comprehensive genomic analysis to identify all human and mouse retroposed sequences that are derived from genes encoding glycolytic enzymes. We analyzed the gene structure of these sequences and determined which copies maintain open reading frames, are transcribed, and may encode sperm-specific isoforms of glycolytic enzymes. Unique features of sperm glycolytic isozymes may be important for localization of this pathway in the prinicipal piece of the sperm flagellum or for altered regulation or kinetic properties that may be required to sustain sperm metabolism and motility in this highly polarized cell. Taken together, identification of all sperm-specific glycolytic enzymes will improve our understanding of sperm metabolism at a molecular level and may provide insights regarding the rapid evolution of genes required for reproduction.

Methods

Identification of gene families

The Ensembl (http://www.ensembl.org) Interpro Domain was used to identify the intron-containing genes for each glycolytic enzyme [49]. We used Ensembl release 48 (Dec 2007) to identify all genes, the mRNA sequence, and their chromosome location. Accession numbers used for BLAST are as follows:

Human: *HK1* [ENSG00000156515] *HK2* [ENSG00000159399] *HK3* [ENSG00000160883], *GCK* [ENSG00000106633], *HKDC1* [ENSG00000156510], *GPI1* [ENSG00000105220], *PFKL* [ENSG00000141959], *PFKM* [ENSG00000152556], *PFKP* [ENSG00000067057], *ALDOA* [ENSG00000149925], *ALDOB* [ENSG00000136872], *ALDOC* [ENSG00000109107], *TPI1* [ENSG00000111669], *GAPDH* [ENSG00000111640], *GAPDHS* [ENSG00000105679], PGK1 [ENSG00000102144], PGAM1 [ENSG00000171314], *PGAM2* [ENSG00000164708], *PGAM5* [ENSG00000176894], *ENO1* [ENSG00000074800], *ENO2* [ENSG00000111674], *ENO3* [ENSG00000108515], *DKFZp781N1041* [ENSG00000188316], *PKLR* [ENSG00000143627], *PKM2* [ENSG00000067225].

Mouse: *Hk1* [ENSMUSG00000037012], *Hk2* [ENSMUSG00000000628], *Hk3* [ENSMUSG00000025877], *Gck* [ENSMUSG00000041798], *Hkdc1* [ENSMUSG00000020080], *Gpi1* [ENSMUSG00000036427], *Pfkl* [ENSMUSG00000020277], *Pfkm* [ENSMUSG00000033065], *Pfkp* [ENSMUSG00000021196], *Aldoa* [ENSMUSG00000030695], *Aldob* [ENSMUSG00000028307], *Aldoc* [ENSMUSG00000017390], *Tpi1* [ENSMUSG00000023456], *Gapdh* [ENSMUSG00000057666], *Gapdhs* [ENSMUSG00000061099], *Pgam1* [ENSMUSG00000011752], *Pgam2* [ENSMUSG00000020475], *Pgam5* [ENSMUSG00000029500], *Eno1*

[ENSMUSG00000063524], *Eno2* [ENSMUSG00000004267], *Eno3* [ENSMUSG00000060600], *6430537H07Rik* [ENSMUSG00000048029], *Pklr* [ENSMUSG00000041237], *Pkm2* [ENSMUSG00000032294].

BLAST search for retroposed sequences

We blasted the mRNA sequence for each gene encoding a glycolytic enzyme using Ensembl BlastView in order to search both the mouse and human genome for retroposed sequences. We grouped BLAST hits based upon chromosome location and excluded hits that were less than 50 base pairs. By comparing BLAST results between gene family members, we identified the parent gene for each retroposed sequence. For each retroposed sequence, we identified the parent gene by choosing matches with the longest hit and the highest percentage match. Using the BLAST results, we calculated the weighted average of the nucleotide identity of all retroposed sequences matching glycolytic enzymes. Ensembl was used to retrieve the FASTA sequence for each retroposed sequence on the appropriate strand.

Sequence alignment

Sequencher 4.8 (Gene Codes Corporation, Ann Arbor, MI) was used to align all retroposed sequences to their parent gene. We used large gap parameters and 60% identity threshold to align all sequences to the reference sequence (the parent gene). At the nucleotide level we looked for insertions and/or deletions, base pair substitutions, and exon structure. We then calculated the percent identity of the coding sequence and looked for an open reading frame. Amino acid sequence alignments were performed using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html) [50].

Tissue and cell isolations

Outbred CD-1 mice were obtained from Charles River (Raleigh, NC). All procedures involving animals were approved by the University of North Carolina at Chapel Hill Animal Care and Use Committee and conducted in accordance with the Guide for the care and Use of Laboratory Animals (Institute for Laboratory Animal Research, National Academy of Sciences).

All tissues were quick frozen in liquid nitrogen and kept at -80℃ until use. Testicular germ cells were isolated using an established protocol [51]. Briefly, we purified pachytene spermatocytes, round spermatids, and condensing spermatids by unit gravity sedimentation from adult mixed germ cell suspensions [51]. Pachytene spermatocytes and round spermatids have purities of > 90%, while fractions containing condensing spermatids have 30-40% nucleated cells and cytoplasts. Testes from 17-day-old mice were used to isolated Sertoli cells, as previously described [52].

Mouse sperm was collected as previously described [6]. Briefly, clipped cauda epididymis were incubated for 15 minutes at 37 ℃ in p hosphate-buffered saline with protease inhibitors (PBS + PI) containing 140 mM NaCl, 10 mM phosphate buffer (pH 7.4) and Complete protease inhibitor cocktail (Roche Diagnostics, Mannheim, Germany). Human sperm samples were obtained from healthy donors provided by the Andrology Laboratory, Department of Obstetrics and Gynecology, University of North Carolina School of Medicine. These samples were washed twice with PBS to remove seminal plasma and cryopreserved.

RT-PCR expression analysis of newly identified retroposed sequences in mouse and human tissues and cells

Total RNA was isolated using Trizol (Invitrogen, Carlsbad, CA) from either cells or tissues pooled from at least three mice. Adult tissues included brain, heart, ovary and testis,

and cells included isolated testicular cells. The Qiagen RNeasy Midi Kit (Qiagen Incorporation, Valencia, CA) was used to remove genomic DNA contamination from RNA preparations. RNA was quantified using the NanoDrop spectrophotometer (NanoDrop Technologies, Wilmington, DE). Human RNA was purchased from Clontech (Mountainview, CA) and was prepared from pooled tissues from 39 individuals. Genomic DNA isolated from CD1 mice and two human subjects were used as positive controls for detected of PCR amplified retrogenes. Reverse transcription followed by gene-specific polymerase chain reaction (Superscript RT II, Invitrogen, Carlsbad, CA; Taq DNA polymerase, New England BioLabs, Ipswich, MA) was used to amplify transcripts from total RNA samples.

Two primer pairs were designed to detect expression of *Gpi1*-related transcripts in mouse RNA samples. The first primer pair distinguishes between transcripts that contain alternatively spliced exons 5 and 6 (Figure 4.4A): *Gpi1*F in exon 4 (5'GAGGTGAACAGGGTTCTGGA3'), *Gpi1*R in exon 11 (5'GCTCGAAGTGGTCAAAACC3'). The expected product sizes are as follows: *Gpi1*, 520 base pairs; *Gpi1_v2/Gpi1-rs1*, 288 base pairs. The second primer pair is specific for *Gpi1-rs1*F in exon 4 (5'ATCAAGGTGGTCGGG3'), *Gpi1-rs1* R in exon 10 (5'CAATGGAAGGTCCAG3'). We also included a negative control with no reverse transcriptase to control for genomic DNA contamination. All PCR products were resolved by 2% agarose gel electrophoresis and visualized by ethidium bromide staining using UV detection.

To detect expression of human retroposed sequences primers were designed to amplify and incorporate α-[^{32}P]-dCTP into the same size product from the parent gene and the retroposed sequences of interest: *Tpi1/Tpi1-rs1, Pgam1/Pgam1-rs1,* and *Eno1/Eno1-rs1*. The forward primer sequence for *Tpi1/Tpi1-rs1* was 5′CCGACACCGAGGTGGTTT3′ and the reverse primer sequence was 5′GTTCTGCGCAGCCACAGCAA3′. The forward primer sequence for *Pgam1/Pgam1-rs1* was 5′GCAGACCTCACAGAAGATCAG3′ and the reverse primer sequence was 5′ACAGATGTGGTCAGTGTGACAT3′. The forward primer

sequence for *Eno1/Eno1-rs1* was 5′TTGGGAAAGCTGGCTACACT3′ and the reverse primer sequence was 5′ CCAGTCATCCTGGTCAAAGG 3′. Arrows in Figure 4.3A denote the location of these primers in each gene.

As a positive control to confirm proper spermatogenesis in human testis samples, we detected expression of protamine 1 (*Prm1*) in RNA samples. We also included a negative control with no reverse transcriptase to control for genomic DNA contamination. The forward primer sequence for *Prm1* was 5'TCACAGGTTGGCTGGCTC3'and the reverse primer sequence was 5'CATTGTTCCTTAGCAGGCTCC3' [53]. Following PCR amplification with both primer sets, the products were resolved by Single Strand Conformation Polymorphism (SSCP) electrophoresis using MDE gel solution (Cambrex, East Rutherford, NJ) at 0.5W for 19 hours. Genomic DNA was used as a control template in parallel PCR reactions to confirm the expected electrophoretic pattern of the retroposed sequences. Gels were exposed to Super RX X-ray film (Fujifilm, Tokyo, Japan) using intensifying screens to detect incorporation of α-[^{32}P]-dCTP into amplified products.

Western analysis of GPI1-related proteins

Lysis buffer (2% SDS, 100 mM DTT, 125 mM Tris pH 6.8, 18% glycerol) was used to extract proteins from tissues or isolated cells. Samples were centrifuged at 16,000×*g* for 10 min at 4 ℃ following homogenization. Concentrat ion of protein samples was determined using the micro-BCA assay (Pierce Biotechnology, Rockford, IL). SDS polyacrylamide gel electrophoresis (SDS-PAGE) on 7.5% polyacrylamide gels were used to separate samples of equal protein amounts, followed by electrophoretically transfer to Immobilon-P PVDF (polyvinylidene fluoride) membranes (Millipore Corp, Bedford, MA). Coomassie blue R250 staining (0.1% Coomassie blue R250 in 45% methanol, 10% acetic acid) of staining detected equal loading following protein transfer. Following staining, membranes were destained, rinsed with TBS-T (140 mM NaCl, 3 mM KCl, 0.05% Tween-20, 25 mM Tris–HCl,

pH 7.4) and incubated in blocking buffer (5% nonfat dry milk in TBS-T) overnight at 4 ℃. Antibody incubations were performed at room temperature in blocking buffer. Membranes were then incubated with a 1:500 dilution of a polyclonal antibody raised against a recombinant human glucose phosphate isomerase protein fragment (Strategic Diagnostic Incorporation, Newark, DE) for 2 hours. Membranes were incubated for 45 min at room temperature with secondary antibody (affinity-purified horseradish peroxidase-conjugated rabbit anti-goat IgG, KPL, Gaithersburg, MD) diluted 1:10,000. Following antibody incubations membranes were rinsed for 5 minutes with TBS-T. Immunoreactive proteins were detected by enhanced chemiluminescence using the SuperSignal West Pico substrate (Pierce Biotechnology, Rockford, IL) and HyBlot CL autoradiography film (Denville Scientific, Metuchen, NJ).

Repetitive Element Analysis

Galaxy (http://galaxy.psu.edu) was used to obtain both 1kb and 10kb FASTA format sequence flanking retroposed sequences and genes encoding glycolytic enzymes [54]. We analyzed FASTA format sequence for all retroposed sequences, 1kb flanking retroposed sequences and 1kb flanking genes encoding glycolytic enzymes for the presence of repetitive elements using Repeatmasker (http://www.repeatmasker.org). We calculated the percent frequency of repetitive elements (LINE, LTR, and SINE) in each base pair within 1 kb upstream or downstream of retroposed sequences or genes encoding glycolytic enzymes. Chi-square values were calculated using a contingency table comparing mouse and human sequences versus glycolytic enzymes and retroposed sequences for each repetitive element. (G+C) content was calculated using the eMBOSS geecee program (http://cat.toulouse.inra.fr/apps/emboss/emboss.cgi/geecee) [55].

BLAST search for extensions

Repeatmasker (http://www.repeatmasker.org) was used to generate sequence with the repetitive elements masked (represented by "n") [56]. We repeatmasked the 1kb sequence flanking all retroposed sequences and used Ensembl BLAST to compare this sequence to the mouse or human genome. We looked for matches with genomic locations close to either the parent gene or other retroposed sequences, indicative of a sequence extension at the end of the gene.

Results

Frequent retrotransposition of orthologous genes encoding glycolytic enzymes occurred independently in the mouse and human genomes

Ten gene families that compose the glycolytic pathway consist of 25 intron-containing genes in the human and mouse genomes (Table 4.1). We identified retroposed sequences in the human and mouse genomes for each family of glycolytic enzymes, except phosphofructokinase (*Pfk)*. Major conclusions from this analysis are:

- Retrotransposition of genes encoding glycolytic enzymes is frequent. We identified 94 matching retroposed sequences in the human genome and 291 in the mouse genome. Our analysis confirms that the mouse genome contains significantly more retroposed sequences than the human genome [57].
- As a rule, only one gene within each family is retroposed.
- The same orthologous gene is retroposed in the human and mouse genomes (bolded font in Table 4.1). This is always true in cases where there is more than one matching retroposed sequence. The two exceptions to this rule, hexokinase (*Hk*) and phosphoglycerate mutase (*Pgam*) have a single retroposed sequence in one or both species. In the human genome *HK2* is retroposed, while *Hk1* is retroposed in

the mouse genome. There is a single *Pgam5* retroposed sequence in mouse in addition to multiple retroposed sequences for *Pgam1* in both species.

- Human retroposed sequences are more divergent from their parent genes when compared to mouse retroposed sequences (Figure 4.1).
- Retrotransposition events occurred independently in each lineage following the divergence of primates and rodents. Phylogenetic analysis was inconclusive in determining the strict order of retrotransposition events due to the high levels of sequence identity. However, analysis of flanking sequence confirmed a lack of conservation of synteny between sequences in the human and mouse genome. Taken together, our analysis found frequent, independent retrotransposition of orthologous genes encoding glycolytic enzymes.
- The location of retroposed sequences in the human (Supplemental Table 4.1) and mouse (Supplemental Table 4.2) genomes appears to be random. There is no region or chromosome with an overrepresentation of retroposed sequences. There is also evidence for gene duplication of retrotransposed sequences (see legend for Supplemental Tables 4.1 and 4.2).

3-6% of retroposed sequences have open reading frames

6.3% of human retroposed sequences derived from genes encoding glycolytic enzymes and 3.4% of mouse sequences contain open reading frames (ORFs) (Figure 4.2). This value includes the previously identified *Pgk2*, *Aldoart1*, and *Aldoart2* retrogenes. In this study we identified five new retroposed sequences with ORFs in the human genome and seven new retroposed sequences with ORFs in the mouse genome (gene structure with red and yellow boxes representing segments of the ORF that match exons in the parent gene, Figure 4.2). Numbers next to each gene structure indicate the percent identity of each retroposed sequence compared to the parent gene. Most retroposed sequences with ORFs

are more similar to the parent gene (>99%) and represent recent retrotransposition events (eg., *PGAM1-rs4*). However, there are examples of retroposed sequences that are much older (eg., *ENO1-rs1*).

Three of the five human retroposed sequences with ORFs were derived from *PGAM1* (Figure 4.2). Both *PGAM1-rs1* and *PGAM1-rs4* have greater than 99% sequence identity at the nucleotide level compared to *PGAM1*, indicative of recent retrotransposition, while *PGAM1-rs7* is only 98.3% identical and has 11 unique amino acid residues (Supplemental Figure 4.1). Three of the mouse retroposed sequences have less than 99% sequence identity at the nucleotide level, including *Pgk1-rs1* and *Pgk1-rs2* (Figure 4.2). Figure 3 shows the percent identity of the coding sequence in retroposed sequences, which is 100% identical for both *Pgk1-rs1* and *Pgk1-rs2*. Both *Pgk1*-related sequences have less than 99% sequence identity in the 3'UTR due to a 5 base pair insertion. We did not detect transcription of either *Pgk1*-related sequence in mouse testis cDNA (data not shown). These results indicate that several retroposed sequences matching glycolytic enzymes in both the human and mouse genomes have open reading frames, supporting possible expression of these sequences.

Detection of splice variants in the glycolytic enzyme parent genes

Analysis of retroposed sequences in both the human and mouse genomes support the expression of alternative transcripts from the parent genes. Two retroposed sequences in humans (*ALDOA-rs1* and *TPI1-rs3*) suggest alternative splicing of internal exons (represented as boxes with diagonal lines in Figure 4.2). For example, *TPI1-rs3* is missing the last 2.5 exons, but still contains part of the last exon when compared to *TPI1* (box with lines, Figure 4.2). Alternative splicing is also supported by two mouse retroposed sequences (*Gpi1-rs1* and *Eno1-rs5*) (boxes with diagonal lines, Figure 4.2). For example,

Eno1-rs5 matches full length *Eno1*, except for a deletion of the last 1.5 exons (box with lines, Figure 4.2). The remaining 3'UTR is maintained, without the coding sequence.

Detection of N-terminal extensions in the glycolytic enzyme parent genes

Multiple mouse and human retroposed sequences have upstream start codons, supporting the expression of transcripts that encode glycolytic enzymes with N-terminal extensions. Three human sequences (*TPI1-rs1*, *PGK1-rs1* and *PGAM1-rs6*) and eight mouse sequences (*Gpi1-rs1*, *Tpi1-rs5, Eno1-rs5,9,* and *Pkm2-rs1,2,3,8*) contain upstream start codons (black exons, Figure 4.2). In most cases, comparison of the amino acid sequence in these N-terminal extensions reveal a unique origin for these extensions that is independent from the parent genes (Supplemental Figure 4.2). Alignment of the amino acid sequence does not show a high level of identity. Five retroposed sequences matching *Pkm2* in mice contain N-terminal extensions. Previous data support the expression of a larger *Pkm2* transcript in bovine and mouse sperm [11, 58]. Proteomic evidence from bovine sperm suggests extension of the N-terminus by at least four amino acids [11]. Our sequence analysis of *Pkm2* pseudogenes with upstream start codons in the mouse genome agrees with the previously identified four amino acid extension, but does not clearly elucidate the start codon responsible for the larger protein product detected in sperm (Supplemental Figure 4.2B).

Novel open reading frames with divergent sequences are not expressed in human testis

Expression of retroposed sequences with greater than 99% identity at the nucleotide level would be difficult to distinguish from parent gene expression. However, we did analyze expression of three human sequences with ORFs and less than 99% nucleotide identity in the coding sequence (*TPI1-rs1, PGAM1-rs7, ENO1-rs1,* and Supplemental Figure 1). Due to high sequence similarity at the nucleotide level, we used Single Strand Conformation

Polymorphism (SSCP) gel electrophoresis to examine potential expression of these retroposed sequences in human testis (Figure 4.3A). Here, RT-PCR products are denatured into two strands and separated based upon individual nucleotide differences. This method allowed us to distinguish between sequences with very high levels of identity. We used genomic DNA to identify the migration pattern of the PCR products amplified from the retroposed sequences (G1 and G2, Figure 4.3B). Expression of protamine 1 (*PRM1*), a spermatid-specific transcript, was detected in human cDNA preparations, confirming complete spermatogenesis in the pooled testes tissues used for RNA isolation. With primers specific for *TPI-rs1, PGAM1-rs7,* and *ENO1-rs1*, RT-PCR did not amplify products from human testis RNA that match the retroposed sequence fragments amplified from genomic DNA (Figure 4.3B). Therefore, we did not detect expression of any of the human retroposed sequences with ORFs in testis.

Expression of an alternative Gpi1 transcript in mouse spermatogenic cells

Gpi1-rs1 maintains an open reading frame, despite missing two internal exons (Figure 4.4A). The open reading frame of *Gpi1-rs1* suggests the expression of this retroposed sequence and/or a *Gpi1* splice variant missing exons 5 and 6 (*Gpi1_v2*, Supplemental Figure 4.3). RT-PCR expression analysis in mouse tissues revealed a testis-specific transcript of glucose phosphate isomerase, representative of *Gpi1_v2* and/or *Gpi1-rs1* (Figure 4.4B). This transcript was also detected in both pachytene spermatocytes and round spermatids isolated from mouse testis, but not in later germ cells (condensing spermatids) or Sertoli cells. The same band was detected in human testis, but due to the absence of *Gpi1-rs1* in the human genome, must represent the expression of *GPI1_V2* (data not shown).

To distinguish between *Gpi1_v2* and *Gpi1-rs1* expression, we designed PCR primers to specifically detect expression of *Gpi1-rs1* (Figure 4.4A). Using this approach, we did not

detect a *Gpi1-rs1*-specific product (Figure 4.4B, bottom panel), indicating that PCR products initially detected in pachytene spermatocytes (Figure 4.4A, top panel) are most likely derived from *Gpi1_v2* transcripts.

We detected expression of the GPI1 protein in various tissues and germ cells isolated from mouse testis (Figure 4.4C). GPI1 has 553 amino acids, while GPI1_V2 has 476 amino acids since it is missing sequence encoded by exons 5 and 6 (Supplemental Figure 4.3). GPI1-rs1 is also missing sequence encoding exons 5 and 6 but contains an N-terminal extension and is, therefore, 502 amino acids (Supplemental Figure 4.3). The predicted molecular weights of GPI1, GPI1_V2, and GPI1-rs1 are, 62,800, 54,500, and 55,100, respectively. We detected a protein band that migrates with an apparent molecular weight of ~55,000 in all tissues analyzed. This band is assumed to be GPI1 due to its ubiquitous expression pattern. We also identified a larger immunoreactive band that was seen only in isolated spermatogenic cells (Figure 4.4C). This protein is not present in human or mouse sperm and is larger than the predicted molecular weights of GPI1_V2 and GPI1-rs1. We also found that glucose phosphate isomerase is soluble in the supernatant fraction following sonication of sperm (s/n, Figure 4.4C). Since GPI1 is not found in insoluble fractions of mouse sperm, it is not tightly bound to the fibrous sheath, the cytoskeletal structure in the sperm flagellum that binds multiple glycolytic enzymes with unique N-terminal extensions [6]. Although we were unable to distinguish GPI1_V2 in our Western analysis, we identified an alternative splice variant of *Gpi1* that is transcribed in spermatogenic cells of the mouse testis.

Repetitive elements are overrepresented in sequence flanking retroposed sequences

The *Aldoart1* sequence provided evidence for an alternative splice variant (*Aldoa_v2*) of aldolase A that is also expressed during spermatogenesis [6]. In our analyses of other retroposed sequences, we examined flanking sequences for evidence of alternative

splicing or additional coding sequence particularly at the N-terminus. Analysis of 1kb sequence both upstream and downstream of all human and mouse retroposed sequences did not identify additional coding sequences. Instead, we found a significant ($P < 0.01$) increase in the number of repetitive elements, particularly LINE and LTR elements, in regions that flank retroposed sequences compared to comparable regions that flank the parent genes (intron-containing genes encoding the glycolytic enzymes) (Supplemental Figure 4.4).

We calculated the percent frequency of both LINE and LTR elements at each base pair within 1 kb upstream and downstream of each retroposed sequences, as compared to intron-containing parent genes encoding glycolytic enzymes. We observed an increase in LINE and LTR elements along the entire 1kb immediately upstream or downstream of retroposed sequences (Figure 4.5A). Because LINE elements are found preferentially in (A+T)-rich regions of the genome [57], we expected a low (G+C) content in the the flanking regions (10kb) of retroposed sequences. Surprisingly, we found that the (G+C) content matched the (G+C) content of the entire genome for both species (Figure 4.5B). Therefore, these retroposed sequences and flanking repetitive elements are not preferentially located in (A+T) rich regions.

Discussion

We conducted a comprehensive analysis of the glycolytic pathway in the human and mouse genomes. The mouse genome contains a larger number of retroposed sequences originating from the glycolytic enzymes when compared to the human genome. There is frequent retrotransposition of orthologous genes encoding glycolytic enzymes in humans and mice. The total number of retroposed sequences matching glycolytic enzymes in the human (94) and mouse (291) genomes is striking. If the approximate 30,000 protein coding genes in the mouse and human genome contained the same number of matching

retroposed sequences as the 25 intron-containing genes in the glycolytic pathway, we would expect upwards of 270,000 total retroposed sequences [57]. In contrast, current estimates of the total retroposed sequence population is much lower, roughly 5000 in humans and 8000 in mice [59, 60]. The striking number of retroposed sequences matching glycolytic enzymes is most likely due to gene expression in the male germline.

Only retrotransposition events that occur in the germline are inherited and contribute to gene family expansion. In order for retrotransposition events to be maintained in the genome they must occur in the germline. Along with cellular localization, there are other characteristics that preferentially promote retrotransposition of certain mRNA transcripts, including stability of RNA in the cytoplasm where LINE ORF1 protein is shown to localize [61, 62]. In our study, we identified preferential retrotransposition of gene family members of the glycolytic enzymes expressed in the testis. We found retrotransposition of *Pgam1*, not *Pgam2*, in the human and mouse genome. Both *Pgam1* and *Pgam2* are expressed in the mouse testis, but *Pgam1* is expressed in early spermatogenesis, while *Pgam2* is expressed in haploid spermatogenic cells following meiosis [63, 64]. *Eno1* is expressed in the mouse testis at 16-20 days post-natal, when spermatocytes are first expressed [65]. In mouse *Pgk1* is expressed in spermatogonia and early spermatocytes, while *Pgk2* expression is first detected in preleptotene spermatocytes [66]. In all three cases, the gene that is preferentially retroposed is the one expressed at early stages of spermatogenesis, when retrotransposition is thought to occur. A recent study detected retrotransposition events during spermatogenesis, although the majority occurred during early stages of embryogenesis immediately following fertilization [67].

Most retroposed sequences are pseudogenes. On average, pseudogenes in the human genome are 86% identical at the nucleotide level to their parent gene [33], while pseudogenes in the mouse genome are 77% identical at the nucleotide level to their parent gene [59]. Pseudogenes matching the glycolytic enzymes are similar to the human average

in terms of nucleotide identity, ranging from 82-100% with an average of 89.2%. However, retroposed sequences in mice have a greater nucleotide identity to their parent gene, also ranging from 82-100% with an average of 93.4%. These values are much greater than the average values, again supporting recent divergence of retroposed sequences matching glycolytic enzymes. The average level of sequence identity between the human and mouse genomes is 69.1% [57]. The related sequences matching glycolytic enzymes in the mouse and human genomes are greater than 80% identical to their parent gene, indicating that these retrotransposition events occurred independently following the divergence of primates and rodents.

Greater than 45% of the human genome and 35% of the mouse genome is composed of repetitive elements [57, 68]. The amount of LINE and LTR repetitive elements was significantly higher flanking retroposed sequences when compared to intron-containing glycolytic enzymes. SINE elements were equally abundant (Supplemental Figure 4.4) flanking both types of sequences and match the total percentage of SINE elements in the human (13.6%) and mouse (8.2%) genomes [57]. There are examples of LINE and LTR elements that are either primate or rodent specific [69]. Our study identified retroposed sequences that diverged following the primate/rodent split, which we would expect since retrotransposition is facilitated by repetitive elements. The number of retroposed sequences combined with the nucleotide divergence of these sequences supports lineage-specific divergence of retroposed sequences.

Recent evidence supports both repression of repetitive elements in male germ cells through methylation and small RNA pathways [70-72]. In most cases the insertion and deletion of genomic sequence that occurs in the germline is deleterious, but it some cases it can be advantageous. Our study supports a role of retrotransposition in the germline for creating gene diversity. A strict regulation of repetitive element expression is achieved

through multiple mechanisms, in order to promote the generation of new genes with functional difference while also preserving existing genomic sequence.

Genes with the most processed pseudogenes include housekeeping genes and metabolic enzymes, and both *Ldha* and *Gapdh* are among the most retroposed transcripts in the human genome [60, 73]. We excluded the glyceraldehyde-phosphate dehydrogenase (*GAPDH*) genes from this analysis due to the previous literature supporting the presence of at least 30 retroposed sequences matching *GAPDH* in human and over 400 in mouse and rat [26, 57, 74]. It has been shown that the *GAPDH* retroposed sequences are located in heavily methylated and DNAse I-insensitive regions of the rat genome [75]. Based upon sequence and phylogenetic analysis, it has been speculated that the *GAPDH* family arose through retrotransposition of retroposed sequences, suggesting transcription of retroposed sequences [74].

The functional importance of our analysis was to identify retrogenes and/or novel sequence in parent genes that are expressed in the testis. We found multiple retroposed genes with either evidence for maintained ORFs and/or alternative splice variants in the human and mouse genomes. We identified five and seven retroposed sequences with ORFs in humans and mice, respectively. We found two retroposed sequences in both the mouse and human genome supporting alternative splicing of the parent gene. Retroposed sequences with ORFs and greater than 99% sequence identity to their parent gene resulted from recent retrotransposition events and were not analyzed for expression. While we did not detect testis expression for any of the putative human retrogenes identified in this study, we cannot exclude the possibility that these genes are expressed in other tissues. In fact, transcription of *PGAM1-rs7* on the X chromosome has been confirmed in human leukocytes [39, 76].

A *Gpi1* retroposed sequence in the mouse genome gave evidence of an alternative splice variant of this gene, expressed in pachytene spermatocytes and round spermatids in

the mouse testis. Our data shows expression of *Gpi1_v2* in spermatocytes, supporting the retrotransposition of this transcript during meiosis to create *Gpi1-rs1*. Meiosis occurs in spermatocyte cells, and methylation of genomic DNA decreases specifically in leptotene and pachytene spermatocytes [77, 78]. DNA double strand breaks required for meiotic recombination are detected in leptotene spermatocytes [79], and LINE elements can take advantage of these naturally occurring double strand breaks. There is evidence for the expression of the proteins encoded by LINE elements (ORF1 and ORF2) in leptotene and zygotene spermatocytes, which includes cells where prophase I of meiosis takes place [30]. Our data shows expression of *Gpi1_v2* in spermatocytes, supporting the retrotransposition of this transcript during meiosis to create *Gpi1-rs1*. Taken together, it is clear that retrotransposition during meiosis can exploit both hypomethylation and double strand breaks.

Our analysis confirmed that retroposed sequences are recent and lineage-specific, which supports the completion of similar genomic analyses in other organisms to confirm all genes encoding glycolytic enzymes. We found frequent, lineage-specific retrotransposition of certain glycolytic enzymes in the human and mouse genomes. Integration of retrotransposition events is random and LINE-1 and LTR dependent. Some of these sequences support the expression of alternative splice variants. Many of these retroposed sequences have maintained open reading frames. While these sequences were not found to be expressed in the testis, the fact that they have open reading frames suggests functional importance of the proteins encoded by these genes. In order for the coding sequence of these genes to be preserved, they should undergo positive selection, possibly in the context of reproductive fitness [47, 48]. In fact, multiple studies have confirmed misexpression of glycolytic enzymes (including *Gpi1*) in sperm from infertile patients [80, 81]. Glycolysis is essential for sperm motility and fertilization, supporting the creation of new genes encoding glycolytic enzymes that promote reproductive success.

Figure 4.1

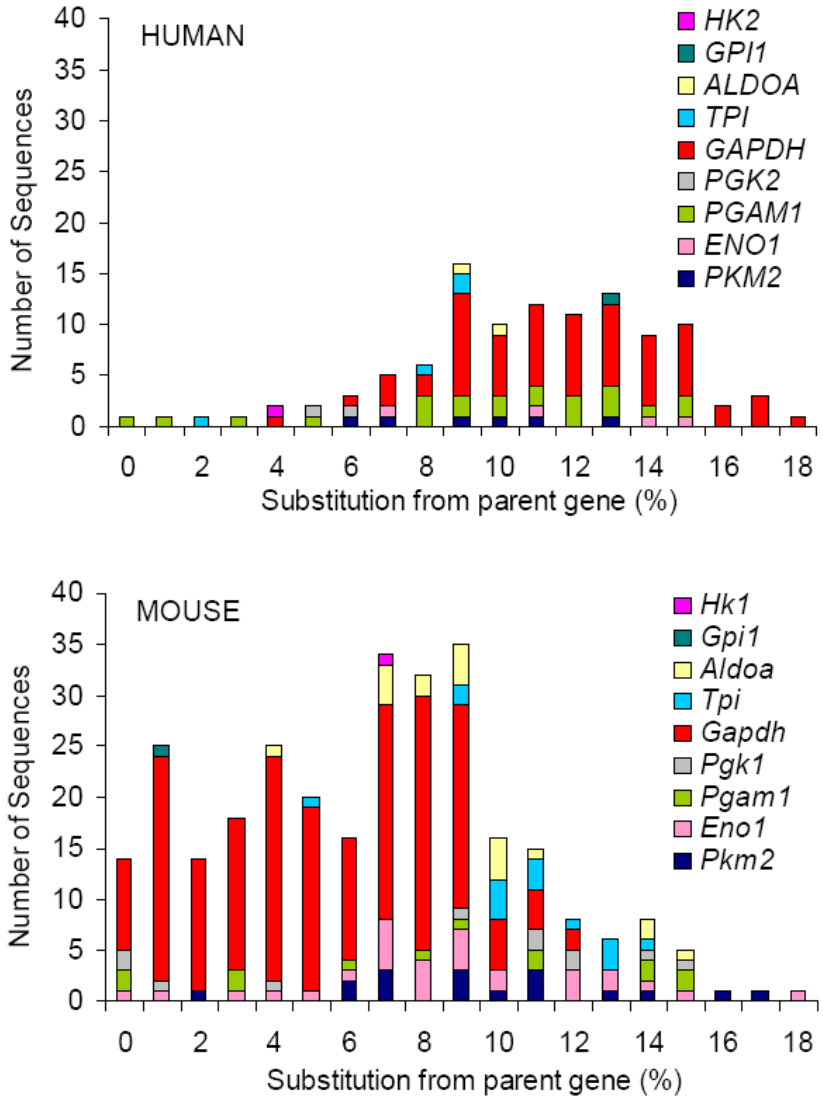


Figure 4.1. Gene and species-specific divergence of human and mouse retroposed sequences. Retroposed sequences matching each enzyme are represented by a different color, as shown in the figure legend. Genes are stacked in each bar as listed in figure legend.

Figure 4.2

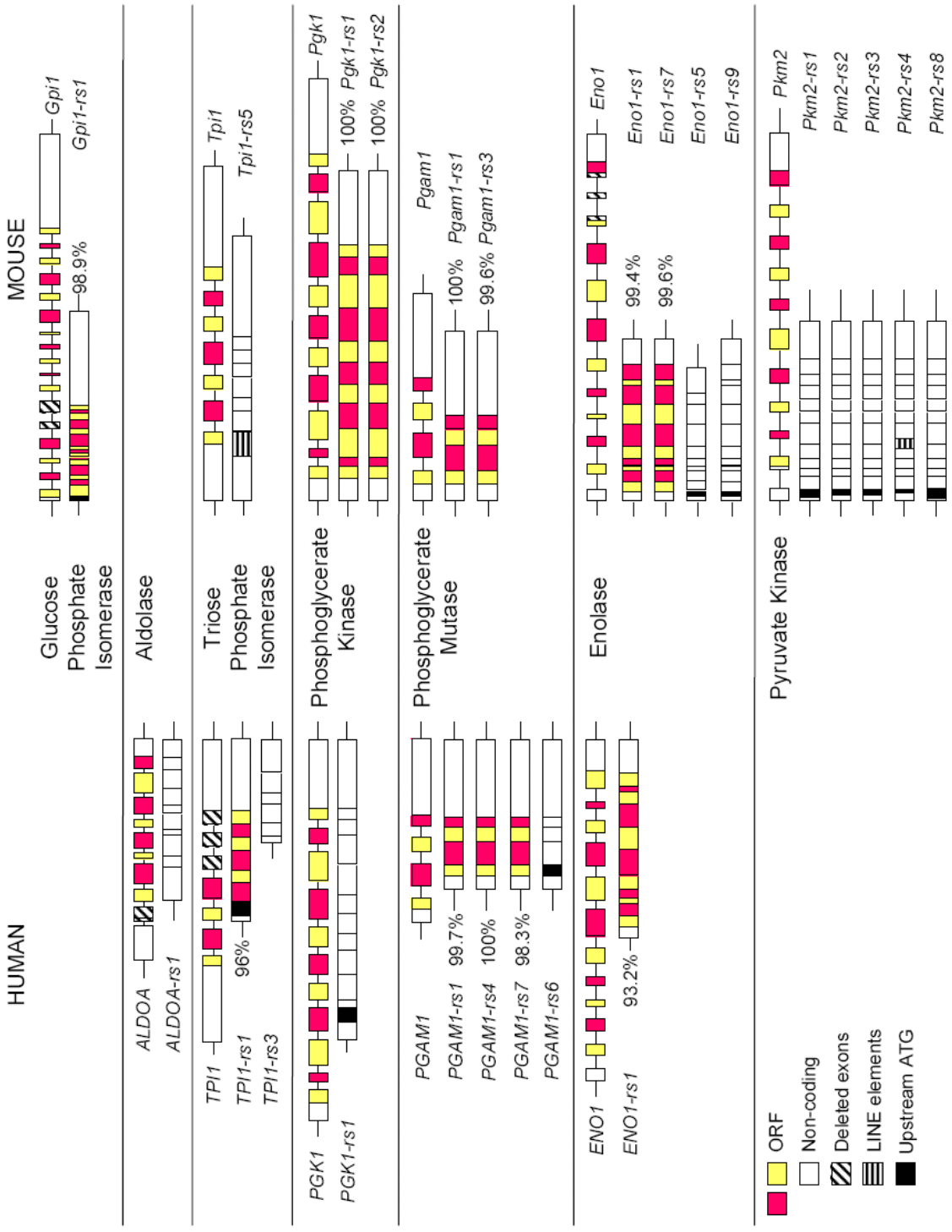

HUMAN
MOUSE
Glucose Phosphate Isomerase
Gpi1
98.9%
Gpi1-rs1
Aldolase
ALDOA
ALDOA-rs1
Triose Phosphate Isomerase
TPI1
96%
TPI1-rs1
TPI1-rs3
Tpi1
Tpi1-rs5
Phosphoglycerate Kinase
PGK1
PGK1-rs1
100%
Pgk1-rs1
100%
Pgk1-rs2
Pgk1
Phosphoglycerate Mutase
PGAM1
PGAM1-rs1
99.7%
PGAM1-rs4
100%
PGAM1-rs7
98.3%
PGAM1-rs6
100%
Pgam1-rs1
99.6%
Pgam1-rs3
Pgam1
Enolase
ENO1
ENO1-rs1
93.2%
99.4%
Eno1-rs1
99.6%
Eno1-rs7
Eno1-rs5
Eno1-rs9
Eno1
Pyruvate Kinase
Pkm2
Pkm2-rs1
Pkm2-rs2
Pkm2-rs3
Pkm2-rs4
Pkm2-rs8
ORF
Non-coding
Deleted exons
LINE elements
Upstream ATG

Figure 4.2. Retroposed sequences support the expression of novel transcripts. Gene structure of parents genes and newly identified retroposed sequences in the human and mouse genome with ORFs (red and yellow exons), upstream start codons (black exons), and/or alternatively spliced exons (diagonal lined boxes). Within these gene structures, sequences containing LINE elements are denoted by horizontal lines. Coding regions for retroposed sequences with open reading frames were compared to their parent gene, and the percent identity at the nucleotide level is shown next to the corresponding gene structure.

Figure 4.3

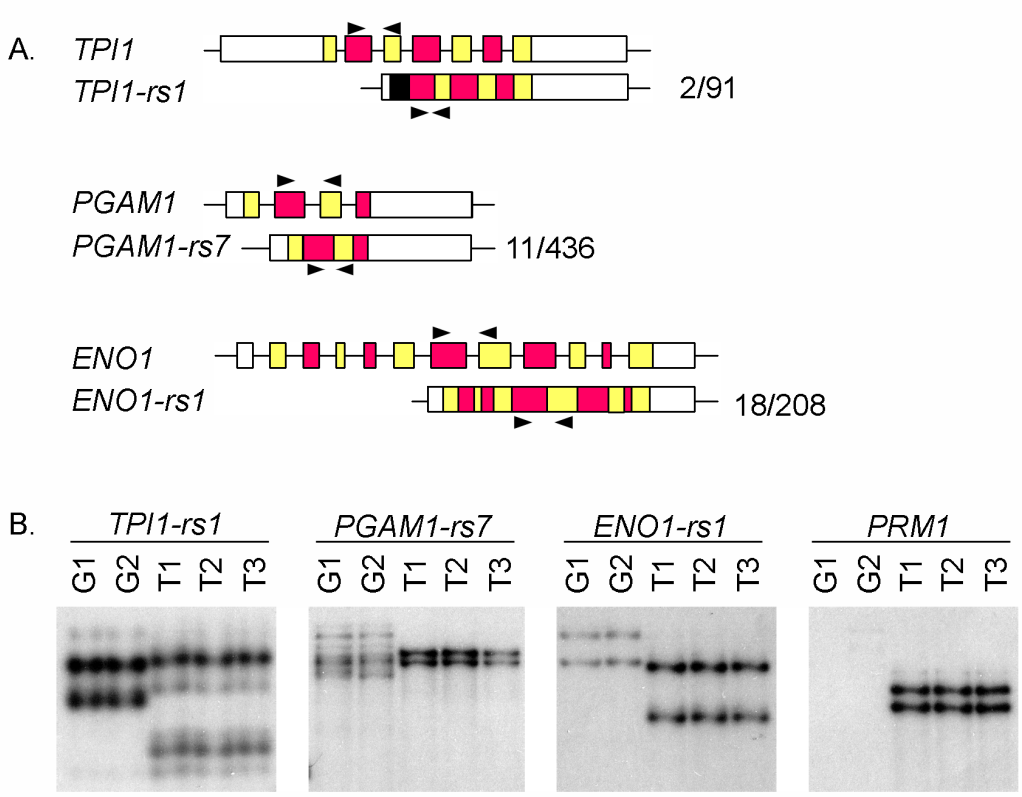


Figure 4.3. Human open reading frames with divergent sequences are not expressed in testis. (A) Diagram of RT-PCR approach used to distinguish expression of transcripts. Black arrows denote primer sets used to amplify both parent gene and retroposed sequence. Fraction next to retroposed sequence gene structure shows the number of unique nucleotide residues in the amplified product. (B) RT-PCR using primers that would amplify both the retroposed sequence and the parent glycolytic enzyme, followed by single-strand conformation polymorphism (SSCP) gel electrophoresis confirmed the absence of *TPI1-rs1*, *PGAM1- rs7*, and *ENO1- rs1* transcripts in pooled human testis RNA samples. PCR products amplified from human genomic DNA (G1 and G2; two individuals) show the expected position of transcripts from retroposed sequences. PCR products amplified from pooled testis total RNA are located in lanes T1, T2, and T3 (triplicate cDNA preparations).

Figure 4.4

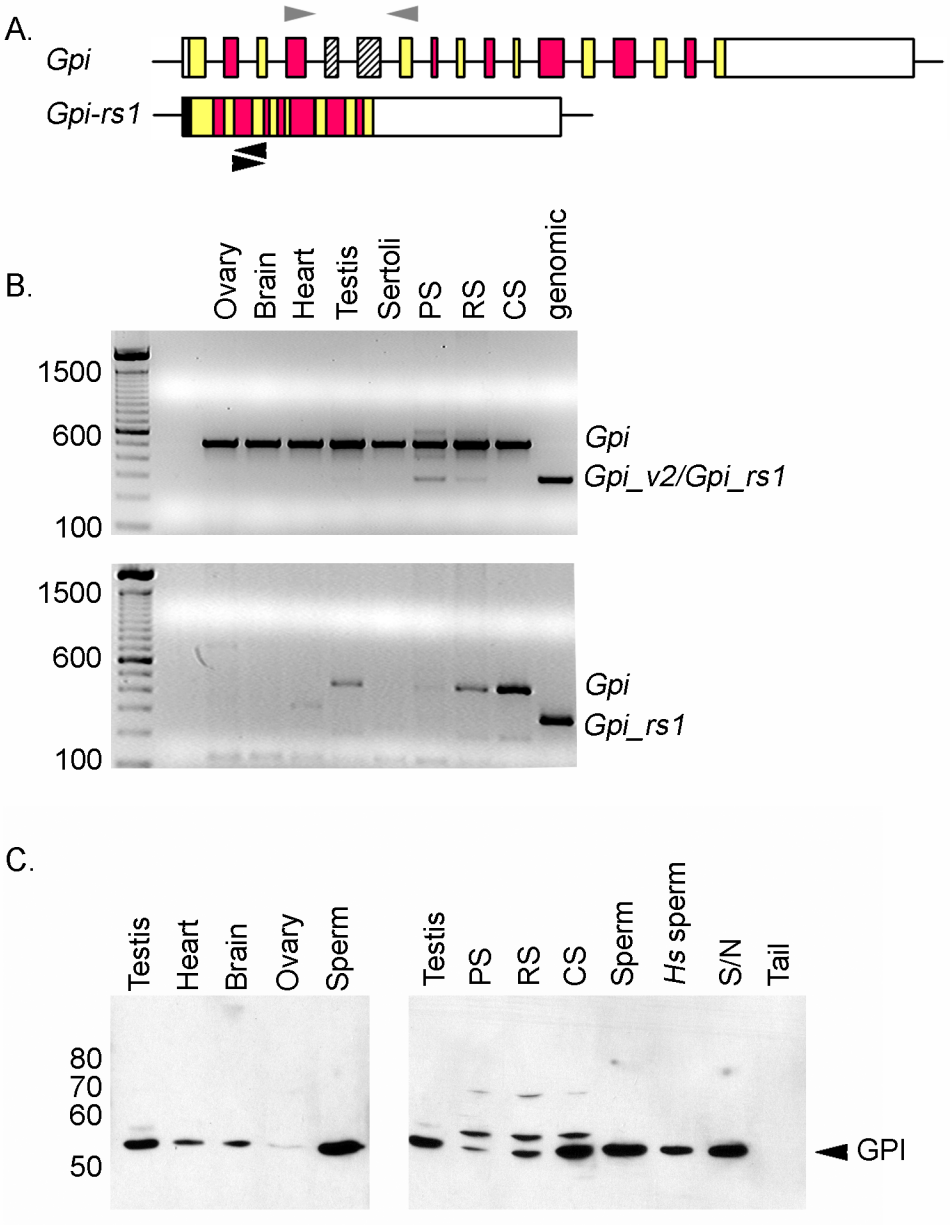

A.
Gpi
Gpi-rs1
B.
Ovary
Brain
Heart
Testis
Sertoli
PS
RS
CS
genomic
1500
600
100
Gpi
Gpi_v2/Gpi_rs1
1500
600
100
Gpi
Gpi_rs1
C.
Testis
Heart
Brain
Ovary
Sperm
Testis
PS
RS
CS
Sperm
Hs sperm
S/N
Tail
80
70
60
50
GPI

Figure 4.4. Expression of an alternative *Gpi1* transcript in mouse spermatogenic cells. (A) Diagram of RT-PCR approach used to distinguish expression of *Gpi1*-related transcripts. Gray arrows denote primer set used to differentiate transcripts containing alternatively spliced exons 5 and 6 (boxes with diagonal lines). Black arrows denote *Gpi1-rs1*-specific primer set. (B) Transcripts from *Gpi1* were detected in all mouse tissues and isolated testicular cells. *Gpi1-rs1* was amplified from genomic DNA to identify the expected size of PCR products from *Gpi1* transcripts not containing exons 5 and 6. A product of the same size was detected in isolated pachytene spermatocytes (PS) and round spermatids (RS), but not condensing spermatids (CS). This PCR fragment appears to be derived from *Gpi1_v2*, since *Gpi1_rs1*-specific primers did not amplify a product. (C) A smaller GPI1_V2 protein was not detected by western analysis using a polyclonal antibody raised against human GPI1. A larger protein product was seen in isolated testicular cell, but not in mouse or human sperm.

Figure 4.5

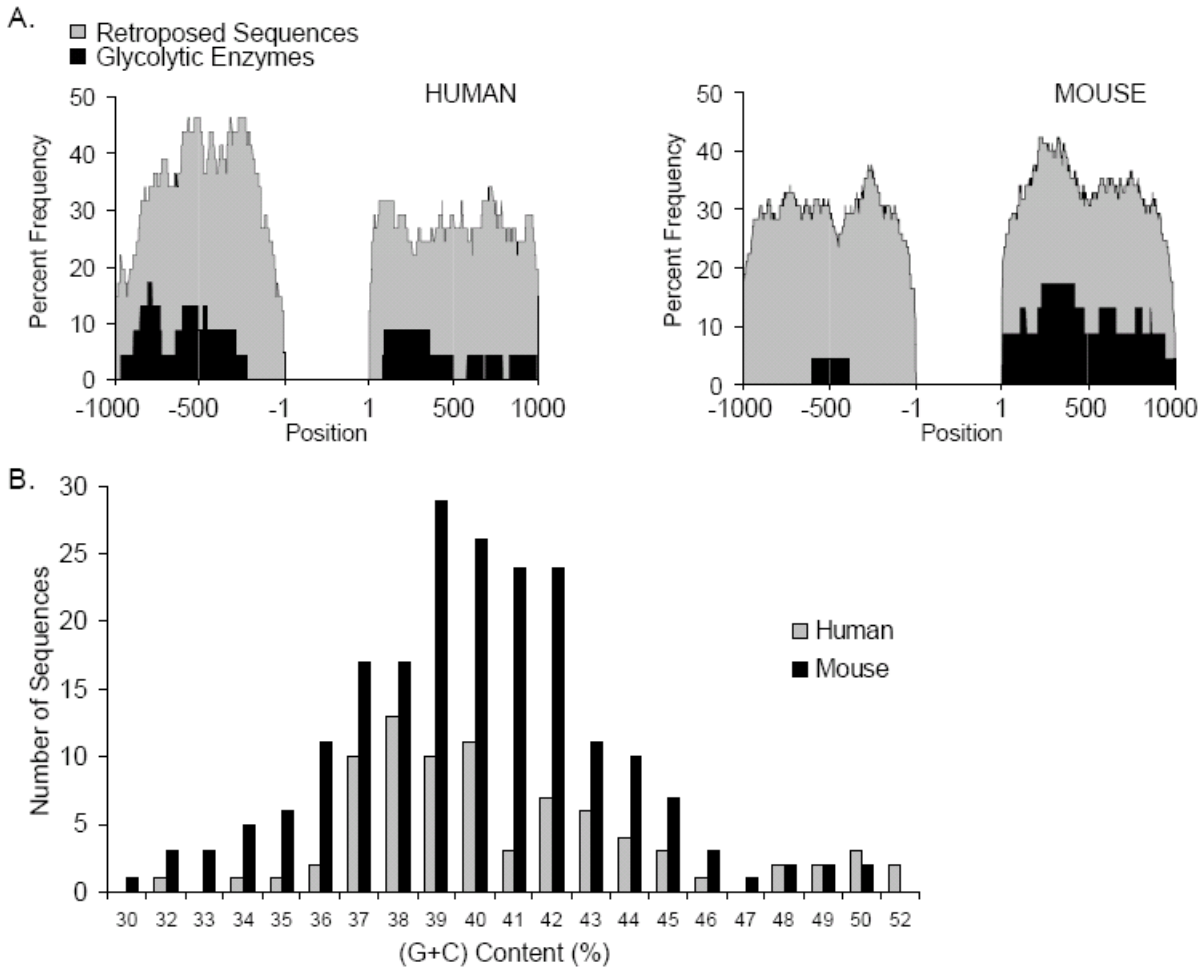


Figure 4.5. Abundance of repetitive elements flanking retroposed sequences and (G+C) content. (A) Analysis confirmed a significant different in LINE and LTR, but not SINE elements within 1 kb flanking either retroposed sequences or genes encoding the glycolytic enzymes. Position of LINE and LTR elements flanking retroposed sequences (grey) or genes encoding all glycolytic enzymes (black). (B) (G+C) content (%) of 10kb sequence flanking human (grey) and mouse (black) retroposed sequences. Pooled upstream and downstream flanking sequences.

Table 4.1. *Gene families encoding glycolytic enzymes and the parent genes that are retroposed.*

	Human			Mouse		
Gene Family	Gene Name	Position	Ret.seq.	Gene Name	Position	Ret.seq.
Hexokinase	*HK1*	10:71	No	***Hk1***	10:62	Yes (1)
	HK2	02:75	Yes (1)	*Hk2*	6:83	No
	HK3	05:176	No	*Hk3*	13:55	No
	GCK	07:44	No	*Gck*	11:6	No
	HKDC1	10:71	No	*Hkdc1*	10:62	No
Glucose Phosphate Isomerase	***GPI1***	19:40	Yes (1)	***Gpi1***	7:34	Yes (1)
Phosphofructokinase	*PFKL*	21:45	No	*Pfkl*	10:77	No
	PFKM	12:47	No	*Pfkm*	15:98	No
	PFKP	10:3	No	*Pfkp*	13:7	No
Aldolase	***ALDOA***	16:30	Yes (2)	***Aldoa***	7:127	Yes (18)
	ALDOB	09:103	No	*Aldob*	4:50	No
	ALDOC	17:24	No	*Aldoc*	11:78	No
Triosephosphate isomerase	***TPI1***	12:7	Yes (4)	***Tpi1***	6:125	Yes (15)
Glyceraldehyde 3-phosphate dehydrogenase	***GAPDH***	12:7	Yes (52)	***Gapdh***	6:125	Yes (188)
	GAPDHS	19:41	No	*Gapdhs*	7:30	No
Phosphoglycerate kinase	***PGK1***	X:77	Yes (3)	***Pgk1***	X:102	Yes (11)
Phosphoglycerate mutase	***PGAM1***	10:99	Yes (21)	***Pgam1***	19:42	Yes (12)
	PGAM2	07:44	No	*Pgam2*	11:6	No
	PGAM5	12:132	No	***Pgam5***	5:111	Yes (1)
Enolase	***ENO1***	01:9	Yes (4)	***Eno1***	4:149	Yes (27)
	ENO2	12:7	No	*Eno2*	6:125	No
	ENO3	17:5	No	*Eno3*	11:70	No
	DKFZp781N1041	10:119	No	*6430537H07Rik*	19:59	No
Pyruvate kinase	*PKLR*	01:154	No	*Pklr*	3:89	No
	PKM2	15:70	Yes (6)	***Pkm2***	9:59	Yes (17)

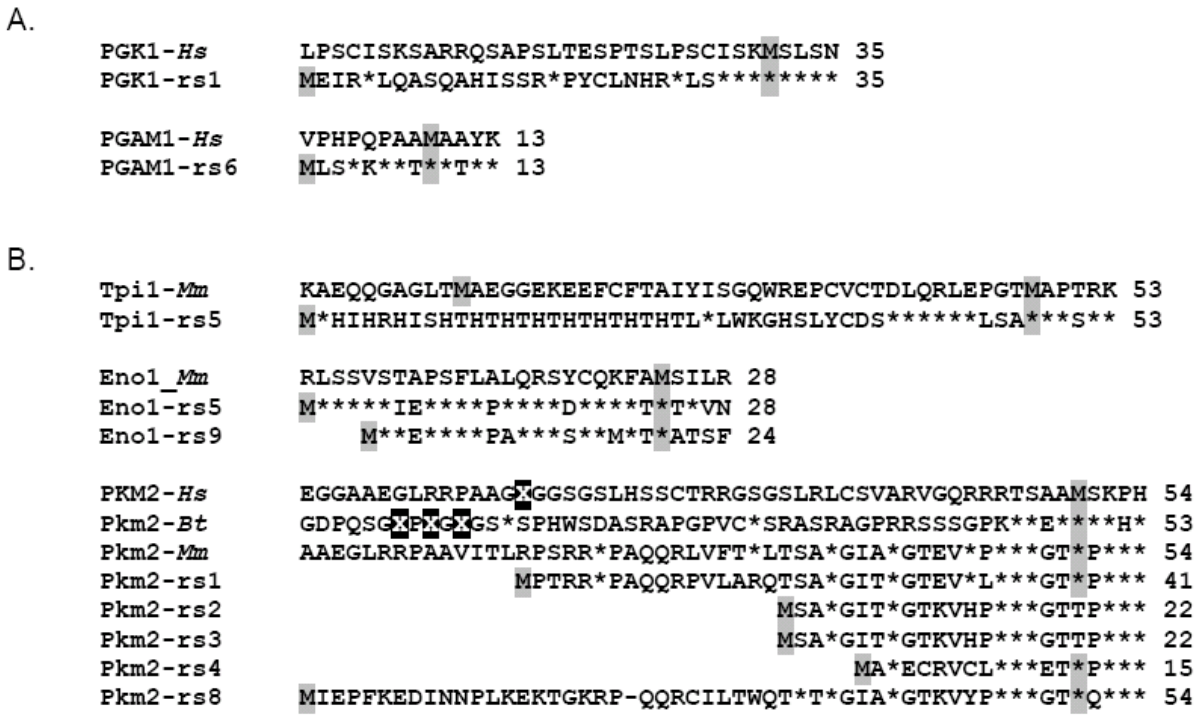
Ensembl release 48 Dec 2007

Table 4.1. Gene families encoding glycolytic enzymes and the parent genes that are retroposed. Number next to “Yes” in retroposed sequence column represents the total number of retroposed sequences matching the parent gene. “Ret. seq.” refers to retroposed sequence. Total number of retroposed sequences matching glycolytic enzymes is 94 in the human genome and 291 in the mouse genome.

Supplemental Figure 4.1

Supplemental Figure 4.1. Amino acid alignment of retroposed sequences in the human genome with maintained open reading frames. Asterisks (*) denote identical residues. Methionine residues are highlighted in grey boxes, residues marked as “X” in a black box denote stop codons, and dashes denote deleted codons.

Supplemental Figure 4.2



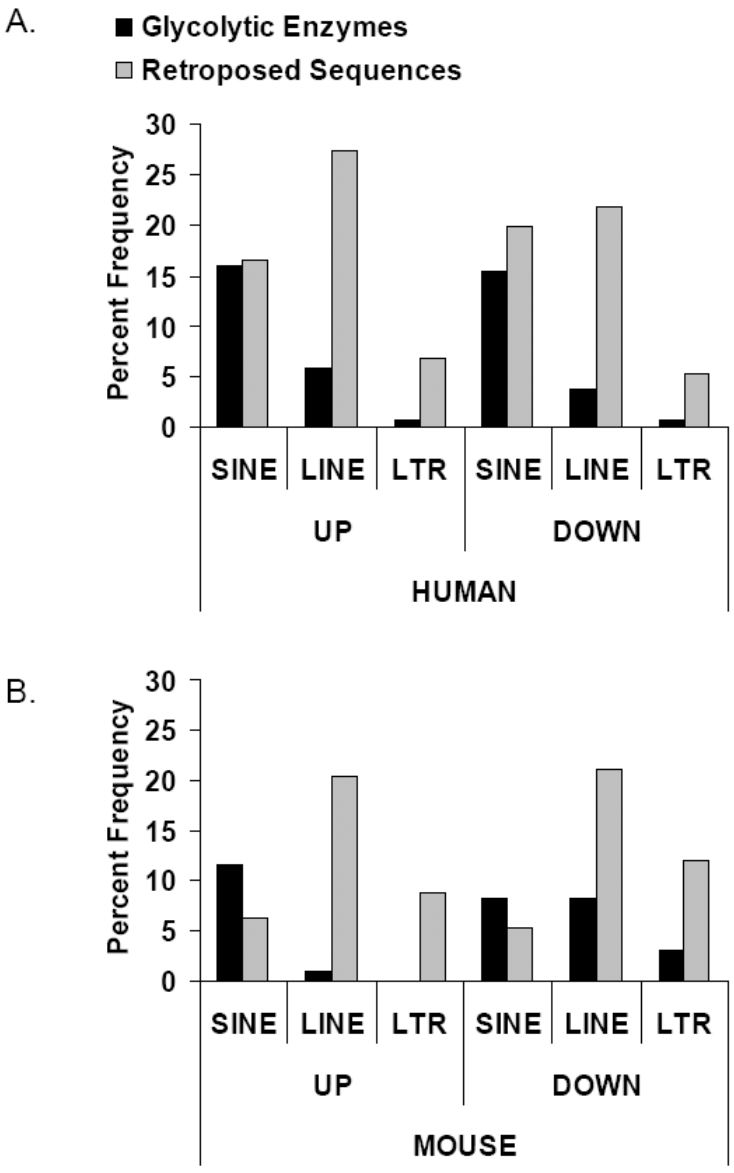
Supplemental Figure 4.2. Amino acid alignments are shown for retroposed sequences containing upstream start codons. (A) Amino acid sequence alignment comparing upstream extensions of human retroposed sequences to the parent glycolytic enzymes. (B) Amino acid sequence alignment of mouse retroposed sequences with upstream start codons and the parent glycolytic enzymes. Asterisks (*) denote identical residues. Methionine residues are highlighted in grey boxes, residues marked as “X” in a black box denote stop codons, and dashes indicate deleted codon

Supplemental Figure 4.3

```
Gpi1       LRVPLGSLAMAALTRNPQFQKLLEWHRANSANLKLRELFEADPERPNNFSLNLNTNHGHILVDYSKNLVNKEVMQMLVELAKSRGVEAARDNMFSGSKINY 100
Gpi1-rs1   MSCSVYLSGPSPWL*S*GTRSSRSCWSG--T*********************************************************************** 98

Gpi1       TENRAVLHVALRNRSNTPIKVDGKDVMPEVNRVLDKMKSFCQRVRSGDWKGYTGKSITDIINIGIGGSDLGPLMVTEALKPYSKGGPRVWFVSNIDGTHIA 200
Gpi1-rs1   **D**********************V*********************---------------------------------------------------- 140

Gpi1       KTLASLSPETSLFIIASKTFTTQETITNAETAKEWFLEAAKDPSAVAKHFVALSTNTAKVKEFGIDPQNMFEFWDWVGGRYSLWSAIGLSIALHVGFDHFE 300
Gpi1-rs1   ----------------------*L*********************************************************T*P********************* 223

Gpi1       QLLSGAHWMDQHFLKTPLEKNAPVLLALLGIWYINCYGCETHALLPYDQYMHRFAAYFQQGDMESNGKYITKSGARVDHQTGPIVWGEPGTNGQHAFYQLI 400
Gpi1-rs1   ************************************************************Y*********************************************** 323

Gpi1       HQGTKMIPCDFLIPVQTQHPIRKGLHHKILLANFLAQTEALMKGKLPEEARKELQAAGKSPEDLEKLLPHKVFEGNRPTNSIVFTKLTPFILGALIAMYEH 500
Gpi1-rs1   ************************************M*********************************************************************** 423

Gpi1       KIFVQGIMWDINSFDQWGVELGKQLAKKIEPELEGSSAVTSHDSSTNGLISFIKQQRDTKLF 562
Gpi1-rs1   ************************************************************W****** 485
```

Supplemental Figure 4.3. Amino acid alignment of GPI1-related sequences in the mouse genome with maintained open reading frames (GPI1-rs1). Asterisks (*) denote identical residues. Methionine residues are highlighted in grey boxes and dashes denote deleted codons.

Supplemental Figure 4.4



Supplemental Figure 4.4. Percent frequency of repetitive elements flanking retroposed sequences and genes encoding glycolytic enzymes in the (A) human and (B) mouse genome. Gray bars denote percent frequency of SINE, LINE and LTR elements both upstream and downstream of retroposed sequences. Black bars represent percent frequency of SINE, LINE and LTR elements flanking genes encoding all glycolytic enzymes.

Supplemental Table 4.1. *Human retroposed sequences matching genes encoding glycolytic enzymes*

Parent Gene	Gene name	Chr	Start	End	Strand	FL CDS
HK2	*HK2-rs1*	X	79711597	79717193	-	Y
GPI1	*GPI1-rs1*	3	53014814	53015169	-	
ALDOA	*ALDOA-rs1*	10	127345239	127346612	-	Y
	ALDOA-rs2	3	52202205	52204044	+	
TPI1	*TPI1-rs1*	1	76938028	76939260	+	Y
	TPI1-rs2	7	128483271	128484502	+	Y
	TPI1-rs3	4	48713051	48713961	-	Y
	TPI1-rs4	6	116466583	116467789	+	Y
PGK1	*PGK1-rs1*	X	67206648	67208343	-	Y
	PGK1-rs2	19	12531320	12533092	-	Y
PGAM1	*PGAM1-rs1*	12	102948637	102950343	+	Y
	PGAM1-rs2	15	94784888	94786599	-	Y
	PGAM1-rs3	3	9365078	9366726	-	Y
	PGAM1-rs4	12	92557837	92559523	-	Y
	PGAM1-rs5	12	94590196	94591866	+	Y
	PGAM1-rs6	6	73821525	73823177	+	Y
	PGAM1-rs7	X	77110148	77111803	-	Y
	PGAM1-rs8	11	91738423	91740066	-	Y
	PGAM1-rs9	X	46390044	46391709	-	
	PGAM1-rs10	9	35932321	35933832	-	
	PGAM1-rs11	20	11549802	11551012	+	
	PGAM1-rs12	X	54715275	54717274	+	
	PGAM1-rs13	11	64698204	64699538	-	
	PGAM1-rs14	4	116903998	116905323	-	
	PGAM1-rs15	13	59794725	59795890	+	
	PGAM1-rs16	12	46570073	46570979	-	UTR
	PGAM1-rs17	2	23948642	23949404	-	
	PGAM1-rs18	1	10041994	10042676	+	
	PGAM1-rs19	5	57492909	57493336	+	
	PGAM1-rs20	8	108728816	108729137	+	
	PGAM1-rs21	7	7410252	7410423	-	UTR
ENO1	*ENO1-rs1*	1	234713027	234714826	+	Y
	ENO1-rs2	15	90765047	90765198	-	
	ENO1-rs3	15	90766554	90766652	-	
	ENO1-rs4	9	22053949	22054012	-	UTR
PKM2	*PKM2-rs1*	1	114880051	114882605	+	Y
	PKM2-rs2	6	5917597	5919858	+	Y
	PKM2-rs3	X	65634319	65636611	+	Y
	PKM2-rs4	8	76451507	76452423	-	
	PKM2-rs5	5	8779377	8780055	+	
	PKM2-rs6	6	86426308	86426934	-	

Supplemental Table 1. Human retroposed sequences matching genes encoding glycolytic enzymes. Gene name given to each retroposed sequence in the human genome, along with the chromosome position and strand. FL CDS refers to those sequences containing full length coding sequence (Y) or only untranslated sequence (UTR), regardless of if sequences are in frame. In some cases, more than one parent gene and/or retroposed sequence is located at the same chromosome position. This includes: 7:44 is *GCK* and *PGAM2*; 10:71 is *HK1* and *HKDC1;* 10:99 is *PGAM1, PGAM1-rs16, PGAM1-rs20,* and *PGAM1-rs23*; 12:7 is TPI1, *GAPDH*, and *ENO2*; 15:91 is *ENO1-rs2* and *ENO1-rs3*; X:66 is *PKM2-rs3* and *PKM2-rs7*

Supplemental Table 4.2. *Mouse retroposed sequences matching genes encoding glycolytic enzymes*

Parent Gene	Gene name	Chr	Start	End	Strand	FL CDS
Hk1	*Hk1-rs1*	18	29504342	29504735	+	UTR
Gpi1	*Gpi1-rs1*	8	52202205	52204044	+	Y
Tpi1	*Tpi1-rs1*	14	17779166	17780521	+	Y
	Tpi1-rs2	10	44309731	44312928	+	
	Tpi1-rs3	9	31755930	31757215	+	Y
	Tpi1-rs4	18	20706529	20707859	-	Y
	Tpi1-rs5	10	18821352	18822679	+	Y
	Tpi1-rs6	3	65932328	65933674	-	Y
	Tpi1-rs7	13	25547615	25548396	+	
	Tpi1-rs8	5	83621602	83622074	-	UTR
	Tpi1-rs9	2	9191857	9192103	-	
	Tpi1-rs10	7	54502936	54503214	+	UTR
	Tpi1-rs11	7	54777639	54777893	+	UTR
	Tpi1-rs12	7	55039236	55039446	+	UTR
	Tpi1-rs13	7	54608077	54608280	+	UTR
	Tpi1-rs14	7	55246555	55246649	+	UTR
	Tpi1-rs15	7	54503234	54503328	+	UTR
Pgk1	*Pgk1-rs1*	12	10905372	10907119	-	Y
	Pgk1-rs2	3	57906493	57908218	-	Y
	Pgk1-rs3	6	29911075	29912501	+	
	Pgk1-rs4	3	18464493	18465979	+	Y
	Pgk1-rs5	5	85295224	85295820	-	
	Pgk1-rs6	18	6760085	6760954	-	
	Pgk1-rs7	12	94954868	94955589	-	
	Pgk1-rs8	16	81909252	81909959	+	
	Pgk1-rs9	5	85290883	85291246	-	UTR
	Pgk1-rs10	14	86629605	86629757	+	UTR
	Pgk1-rs11	14	86629502	86629607	-	
Pgam1	*Pgam1-rs1*	7	28253905	28255638	-	Y
	Pgam1-rs2	11	21582397	21584105	-	Y
	Pgam1-rs3	3	11013514	11015192	-	Y
	Pgam1-rs4	11	79710680	79712360	+	Y
	Pgam1-rs5	18	80359524	80361179	+	
	Pgam1-rs6	15	59760341	59761759	-	
	Pgam1-rs7	9	50023246	50024730	+	
	Pgam1-rs8	1	9429117	9430492	+	
	Pgam1-rs9	X	136493319	136494149	+	UTR
	Pgam1-rs10	15	24059372	24059829	-	
	Pgam1-rs11	1	23984587	23984906	-	
	Pgam1-rs12	9	123442121	123442509	+	UTR
Pgam5	*Pgam5-rs1*	12	92720490	92723471	+	Y
Eno1	*Eno1-rs1*	18	48206301	48208032	+	Y
	Eno1-rs2	14	15279830	15281567	-	Y
	Eno1-rs3	3	130631969	130633696	-	Y
	Eno1-rs4	17	16918472	16920172	+	Y
	Eno1-rs5	13	88686724	88688161	+	Y
	Eno1-rs6	18	48227671	48229782	-	Y

	Eno1-rs7	2	80677252	80678919	+	Y
	Eno1-rs8	X	126935186	126936782	-	Y
	Eno1-rs9	1	43842756	43844341	+	
	Eno1-rs10	2	100247310	100248846	-	Y
	Eno1-rs11	17	91191800	91193200	-	
	Eno1-rs12	4	64517715	64519081	-	
	Eno1-rs13	X	104474124	104475051	-	
	Eno1-rs14	10	18981708	18982788	+	
	Eno1-rs15	3	60573575	60574378	-	
	Eno1-rs16	10	68153941	68154564	+	
	Eno1-rs17	7	4902248	4902901	-	
	Eno1-rs18	2	83378374	83378827	-	
	Eno1-rs19	12	64765950	64766281	+	
	Eno1-rs20	1	165944207	165944509	-	UTR
	Eno1-rs21	1	193075309	193075608	+	
	Eno1-rs22	8	51205678	51205966	-	UTR
	Eno1-rs23	3	60574381	60574576	+	
	Eno1-rs24	14	53128289	53128461	-	
	Eno1-rs25	10	10895171	10895287	-	UTR
	Eno1-rs26	3	13489047	13489153	+	UTR
	Eno1-rs27	16	82107093	82107155	-	
Pkm2	*Pkm2-rs1*	12	31702976	31705204	+	Y
	Pkm2-rs2	2	100299806	100301960	-	Y
	Pkm2-rs3	2	100466315	100468469	-	Y
	Pkm2-rs4	9	97108961	97111315	-	Y
	Pkm2-rs5	X	111135597	111137719	+	Y
	Pkm2-rs6	X	8499059	8501032	-	
	Pkm2-rs7	5	22230028	22232077	-	Y
	Pkm2-rs8	15	69893125	69895146	-	Y
	Pkm2-rs9	18	21650759	21652605	-	Y
	Pkm2-rs10	13	13987730	13988873	-	
	Pkm2-rs11	X	159282193	159283455	-	
	Pkm2-rs12	12	88056199	88056992	+	
	Pkm2-rs13	17	93117277	93118318	-	
	Pkm2-rs14	3	43567795	43568688	-	
	Pkm2-rs15	16	54827438	54827861	-	
	Pkm2-rs16	13	14002561	14002728	-	
	Pkm2-rs17	19	54333380	54333485	-	UTR

Supplemental Table 4.2. Mouse retroposed sequences matching genes encoding glycolytic enzymes. Gene name given to each retroposed sequence in the mouse genome, along with the chromosome position and strand. FL CDS refers to those sequences containing full length coding sequence (Y) or only untranslated sequence (UTR), regardless of if sequences are in frame. In some cases, more than one parent gene and/or retroposed sequence is located at the same chromosome position. This includes: 2:100 is *Eno1-rs10, Pkm2-rs2, Pkm2-rs3*; 3:61 is *Eno1-rs15, Eno1-rs23*; 5:85 is *Pgk1-rs5* and *Pgk1-rs9*; 6:125 *Tpi1, Gapdh, Eno2*; 7:55 is *Tpi1-rs11, Tpi1-rs12, Tpi1-rs13, Tpi1-rs14, Tpi1-rs15, Tpi1-rs16*; 10:14 is *Tpi1-rs4* and *Eno1-rs14*; 10:44 is *Tpi1-rs7* and *Tpi1-rs8*; 11:6 is *Gck* and *Pgam2*; 13:14 is *Pkm2-rs9* and *Pkm2-rs17*; 14:87 is *Pgk1-rs10* and *Pgk1-rs11*; 16:82 is *Pgk1-rs8* and *Eno1-rs27*; 18:48 is *Eno1-rs1* and *Eno1-rs6*.

References

1. Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis**. *Biol Reprod* 1998, **58**(3):834-841.

2. Welch JE, Schatte EC, O'Brien DA, Eddy EM: **Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells**. *Biol Reprod* 1992, **46**(5):869-878.

3. Boer PH, Adra CN, Lau YF, McBurney MW: **The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon**. *Mol Cell Biol* 1987, **7**(9):3107-3112.

4. Goldberg E: **Isozymes in testes and spermatozoa**. *Isozymes: Curr Topics Biol Med Res* 1977, **1**:79-124.

5. Li SS, O'Brien DA, Hou EW, Versola J, Rockett DL, Eddy EM: **Differential activity and synthesis of lactate dehydrogenase isozymes A (muscle), B (heart), and C (testis) in mouse spermatogenic cells**. *Biol Reprod* 1989, **40**(1):173-180.

6. Vemuganti SA, Bell TA, Scarlett CO, Parker CE, de Villena FP, O'Brien DA: **Three male germline-specific aldolase A isozymes are generated by alternative splicing and retrotransposition**. *Dev Biol* 2007, **309**(1):18-31.

7. Mori C, Nakamura N, Welch JE, Gotoh H, Goulding EH, Fujioka M, Eddy EM: **Mouse spermatogenic cell-specific type 1 hexokinase (*mHk1-s*) transcripts are expressed by alternative splicing from the *mHk1* gene and the HK1-S protein is localized mainly in the sperm tail**. *Mol Reprod Dev* 1998, **49**(4):374-385.

8. Mori C, Welch JE, Fulcher KD, O'Brien DA, Eddy EM: **Unique hexokinase messenger ribonucleic acids lacking the porin-binding domain are developmentally expressed in mouse spermatogenic cells**. *Biol Reprod* 1993, **49**(2):191-203.

9. Nakamura N, Shibata H, O'Brien DA, Mori C, Eddy EM: **Spermatogenic cell-specific type 1 hexokinase is the predominant hexokinase in sperm**. *Mol Reprod Dev* 2008, **75**(4):632-640.

10. Steinke D, Hoegg S, Brinkmann H, Meyer A: **Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates**. *BMC Biol* 2006, **4**:16.

11. Feiden S, Stypa H, Wolfrum U, Wegener G, Kamp G: **A novel pyruvate kinase (PK-S) from boar spermatozoa is localized at the fibrous sheath and the acrosome**. *Reproduction* 2007, **134**(1):81-95.

12. Buehr M, McLaren A: **An electrophoretically detectable modification of glucosephosphate isomerase in mouse spermatozoa**. *J Reprod Fertil* 1981, **63**(1):169-173.

13. Yakirevich E, Naot Y: **Cloning of a glucose phosphate isomerase/neuroleukin-like sperm antigen involved in sperm agglutination**. *Biol Reprod* 2000, **62**(4):1016-1023.

14. Russell DL, Kim KH: **Expression of triosephosphate isomerase transcripts in rat testis: evidence for retinol regulation and a novel germ cell transcript**. *Biol Reprod* 1996, **55**(1):11-18.

15. Edwards YH, Grootegoed JA: **A sperm-specific enolase**. *J Reprod Fertil* 1983, **68**(2):305-310.

16. Gitlits VM, Toh BH, Loveland KL, Sentry JW: **The glycolytic enzyme enolase is present in sperm tail and displays nucleotide-dependent association with microtubules**. *Eur J Cell Biol* 2000, **79**(2):104-111.

17. Force A, Viallard JL, Saez F, Grizard G, Boucher D: **Electrophoretic characterization of the human sperm-specific enolase at different stages of maturation**. *J Androl* 2004, **25**(5):824-829.

18. Yamada S, Nakajima H, Kuehn MR: **Novel testis- and embryo-specific isoforms of the phosphofructokinase-1 muscle type gene**. *Biochem Biophys Res Commun* 2004, **316**(2):580-587.

19. Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility**. *Proc Natl Acad Sci U S A* 2004, **101**(47):16501-16506.

20. Danshina PV: **Male fertility and sperm function are severely impaired in mice lacking phosphoglycerate kinase-2.** *Journal of Andrology* 2006, **27**(Suppl):35.

21. Odet F, Duan C, Willis WD, Goulding EH, Kung A, Eddy EM, Goldberg E: **Expression of the gene for mouse lactate dehydrogenase C (Ldhc) is required for male fertility**. *Biol Reprod* 2008, **79**(1):26-34.

22. McCarrey JR, Thomas K: **Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene**. *Nature* 1987, **326**(6112):501-505.

23. Kao FT, Wu KC, Law ML, Hartz JA, Lau YF: **Assignment of human gene encoding testis-specific lactate dehydrogenase C to chromosome 11, region p14.3-p15.5**. *Somat Cell Mol Genet* 1988, **14**(5):515-518.

24. Welch JE, Brown PR, O'Brien DA, Eddy EM: **Genomic organization of a mouse glyceraldehyde 3-phosphate dehydrogenase gene (Gapd-s) expressed in post-meiotic spermatogenic cells**. *Dev Genet* 1995, **16**(2):179-189.

25. Vandeberg JL, Cooper DW, Sharman GB, Poole WE: **Somatic expression and autosomal inheritance of phosphoglycerate kinase B in kangaroos**. *Genetics* 1980, **95**(2):413-424.

26. Piechaczyk M, Blanchard JM, Riaad-El Sabouty S, Dani C, Marty L, Jeanteur P: **Unusual abundance of vertebrate 3-phosphate dehydrogenase pseudogenes**. *Nature* 1984, **312**(5993):469-471.

27. Tolan DR, Niclas J, Bruce BD, Lebo RV: **Evolutionary implications of the human aldolase-A, -B, -C, and -pseudogene chromosome locations**. *Am J Hum Genet* 1987, **41**(5):907-924.

28. Vemuganti SA: **Genomic and proteomic analysis of aldolase A retrogenes expressed during spermatogenesis.** *Journal of Andrology* 2006, **27**(Suppl):39.

29. Cortinas MN, Lessa EP: **Molecular evolution of aldolase A pseudogenes in mice: multiple origins, subsequent duplications, and heterogeneity of evolutionary rates**. *Mol Biol Evol* 2001, **18**(9):1643-1653.

30. Branciforte D, Martin SL: **Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition**. *Mol Cell Biol* 1994, **14**(4):2584-2592.

31. Dupressoir A, Heidmann T: **Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice**. *Mol Cell Biol* 1996, **16**(8):4495-4503.

32. Ergun S, Buschmann C, Heukeshoven J, Dammann K, Schnieders F, Lauke H, Chalajour F, Kilic N, Stratling WH, Schumann GG: **Cell type-specific expression of LINE-1 open reading frames 1 and 2 in fetal and adult human tissues**. *J Biol Chem* 2004, **279**(26):27753-27763.

33. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, Kazazian HH, Jr.: **A mouse model of human L1 retrotransposition**. *Nat Genet* 2002, **32**(4):655-660.

34. Babushok DV, Kazazian HH, Jr.: **Progress in understanding the biology of the human mutagen LINE-1**. *Hum Mutat* 2007, **28**(6):527-539.

35. Vinckenbosch N, Dupanloup I, Kaessmann H: **Evolutionary fate of retroposed gene copies in the human genome**. *Proc Natl Acad Sci U S A* 2006, **103**(9):3220-3225.

36. Hendriksen PJ, Hoogerbrugge JW, Baarends WM, de Boer P, Vreeburg JT, Vos EA, van der Lende T, Grootegoed JA: **Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse**. *Genomics* 1997, **41**(3):350-359.

37. Kleene KC, Mulligan E, Steiger D, Donohue K, Mastrangelo MA: **The mouse gene encoding the testis-specific isoform of Poly(A) binding protein (Pabp2) is an expressed retroposon: intimations that gene expression in spermatogenic cells facilitates the creation of new genes**. *J Mol Evol* 1998, **47**(3):275-281.

38. Emerson JJ, Kaessmann H, Betran E, Long M: **Extensive gene traffic on the mammalian X chromosome**. *Science* 2004, **303**(5657):537-540.

39. Marques AC, Dupanloup I, Vinckenbosch N, Reymond A, Kaessmann H: **Emergence of young human genes after a burst of retroposition in primates**. *PLoS Biol* 2005, **3**(11):e357.

40. Nishimune Y, Tanaka H: **Infertility caused by polymorphisms or mutations in spermatogenesis-specific genes**. *J Androl* 2006, **27**(3):326-334.

41. Torgerson DG, Kulathinal RJ, Singh RS: **Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes**. *Mol Biol Evol* 2002, **19**(11):1973-1980.

42. Mukai C, Okuno M: **Glycolysis plays a major role for adenosine triphosphate supplementation in mouse sperm flagellar movement**. *Biol Reprod* 2004, **71**(2):540-547.

43. Peterson RN, Freund M: **Glycolysis by washed suspensions of human spermatozoa. Effect of substrate, substrate concentration, and changes in medium composition on the rate of glycolysis**. *Biol Reprod* 1969, **1**(3):238-246.

44. Williams AC, W.C. F: **The role of glucose in supporting motility and capacitation in human spermatozoa**. *J Androl* 2001, **22**(4):680-695.

45. Narisawa S, Hecht NB, Goldberg E, Boatright KM, Reed JC, Millan JL: **Testis-specific cytochrome *c*-null mice produce functional sperm but undergo early testicular atrophy**. *Mol Cell Biol* 2002, **22**(15):5554-5562.

46. Curi SM, Ariagno JI, Chenlo PH, Mendeluk GR, Pugliese MN, Sardi Segovia LM, Repetto HE, Blanco AM: **Asthenozoospermia: analysis of a large population**. *Arch Androl* 2003, **49**(5):343-349.

47. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ *et al*: **A scan for positively selected genes in the genomes of humans and chimpanzees**. *PLoS Biol* 2005, **3**(6):e170.

48. Swanson WJ, Vacquier VD: **The rapid evolution of reproductive proteins**. *Nat Rev Genet* 2002, **3**(2):137-144.

49. Hubbard TJ, Aken BL, Beal K, Ballester B, Caccamo M, Chen Y, Clarke L, Coates G, Cunningham F, Cutts T *et al*: **Ensembl 2007**. *Nucleic Acids Res* 2007, **35**(Database issue):D610-617.

50. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**(22):4673-4680.

51. O'Brien DA: **Isolation, separation, and short-term culture of spermatogenic cells**. In: *Methods in Toxicology* Edited by Chapin RE, Heindel JJ; 1993: 246-264.

52. O'Brien DA, Gabel CA, Rockett DL, Eddy EM: **Receptor-mediated endocytosis and differential synthesis of mannose 6-phosphate receptors in isolated spermatogenic and Sertoli cells**. *Endocrinology* 1989, **125**(6):2973-2984.

53. Steger K, Pauls K, Klonisch T, Franke FE, Bergmann M: **Expression of protamine-1 and -2 mRNA during human spermiogenesis**. *Mol Hum Reprod* 2000, **6**(3):219-225.

54. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J *et al*: **Galaxy: a platform for interactive large-scale genome analysis**. *Genome Res* 2005, **15**(10):1451-1455.

55. Rice P, Longden I, Bleasby A: **EMBOSS: the European Molecular Biology Open Software Suite**. *Trends Genet* 2000, **16**(6):276-277.

56. Smit A, Hubley R, Green P: **RepeatMasker Open-3.0. <http://www.repeatmasker.org>**. 1996-2004.

57. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P *et al*: **Initial sequencing and comparative analysis of the mouse genome**. *Nature* 2002, **420**(6915):520-562.

58. Krisfalusi M, Miki K, Magyar PL, O'Brien D A: **Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa**. *Biol Reprod* 2006, **75**(2):270-278.

59. Zhang Z, Carriero N, Gerstein M: **Comparative analysis of processed pseudogenes in the mouse and human genomes**. *Trends Genet* 2004, **20**(2):62-67.

60. Zhang Z, Harrison PM, Liu Y, Gerstein M: **Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome**. *Genome Res* 2003, **13**(12):2541-2558.

61. Goodier JL, Zhang L, Vetter MR, Kazazian HH, Jr.: **LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex**. *Mol Cell Biol* 2007, **27**(18):6469-6483.

62. Pavlicek A, Gentles AJ, Paces J, Paces V, Jurka J: **Retroposition of processed pseudogenes: the impact of RNA stability and translational control**. *Trends Genet* 2006, **22**(2):69-73.

63. Fundele R, Winking H, Illmensee K, Jagerbauer EM: **Developmental activation of phosphoglycerate mutase-2 in the testis of the mouse**. *Dev Biol* 1987, **124**(2):562-566.

64. Broceno C, Walsh K, Pons G: **A 1.3-kb upstream 5' region of the rat phosphoglycerate mutase m gene confers testis and skeletal muscle-specific expression in transgenic mice**. *Biochem Biophys Res Commun* 1999, **263**(1):244-250.

65. Couldrey C, Carlton MB, Ferrier J, Colledge WH, Evans MJ: **Disruption of murine alpha-enolase by a retroviral gene trap results in early embryonic lethality**. *Dev Dyn* 1998, **212**(2):284-292.

66. McCarrey JR, Berg WM, Paragioudakis SJ, Zhang PL, Dilworth DD, Arnold BL, Rossi JJ: **Differential transcription of *pgk* genes during spermatogenesis in the mouse**. *Dev Biol* 1992, **154**(1):160-168.

67. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH, Jr.: **L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism**. *Genes Dev* 2009, **23**(11):1303-1312.

68. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**(6822):860-921.

69. Yang S, Smit AF, Schwartz S, Chiaromonte F, Roskin KM, Haussler D, Miller W, Hardison RC: **Patterns of insertions and their covariation with substitutions in the rat, mouse, and human genomes**. *Genome Res* 2004, **14**(4):517-527.

70. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ: **Developmentally regulated piRNA clusters implicate MILI in transposon control**. *Science* 2007, **316**(5825):744-747.

71. Bourc'his D, Bestor TH: **Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L**. *Nature* 2004, **431**(7004):96-99.

72. Carmell MA, Girard A, van de Kant HJ, Bourc'his D, Bestor TH, de Rooij DG, Hannon GJ: **MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline**. *Dev Cell* 2007, **12**(4):503-514.

73. Ohshima K, Hattori M, Yada T, Gojobori T, Sakaki Y, Okada N: **Whole-genome screening indicates a possible burst of formation of processed pseudogenes and Alu repeats by particular L1 subfamilies in ancestral primates**. *Genome Biol* 2003, **4**(11):R74.

74. Garcia-Meunier P, Etienne-Julan M, Fort P, Piechaczyk M, Bonhomme F: **Concerted evolution in the GAPDH family of retrotransposed pseudogenes**. *Mamm Genome* 1993, **4**(12):695-703.

75. Riad-el Sabrouty S, Blanchard JM, Marty L, Jeanteur P, Piechaczyk M: **The muridae glyceraldehyde-3-phosphate dehydrogenase family**. *J Mol Evol* 1989, **29**(3):212-222.

76. Betran E, Emerson JJ, Kaessmann H, Long M: **Sex chromosomes and male functions: where do new genes go?** *Cell Cycle* 2004, **3**(7):873-875.

77. Mertineit C, Yoder JA, Taketo T, Laird DW, Trasler JM, Bestor TH: **Sex-specific exons control DNA methyltransferase in mammalian germ cells**. *Development* 1998, **125**(5):889-897.

78. La Salle S, Mertineit C, Taketo T, Moens PB, Bestor TH, Trasler JM: **Windows for sex-specific methylation marked by DNA methyltransferase expression profiles in mouse germ cells**. *Dev Biol* 2004, **268**(2):403-415.

79. Mahadevaiah SK, Turner JM, Baudat F, Rogakou EP, de Boer P, Blanco-Rodriguez J, Jasin M, Keeney S, Bonner WM, Burgoyne PS: **Recombinational DNA double-strand breaks in mice precede synapsis**. *Nat Genet* 2001, **27**(3):271-276.

80. Rockett JC, Patrizio P, Schmid JE, Hecht NB, Dix DJ: **Gene expression patterns associated with infertility in humans and rodent models**. *Mutat Res* 2004, **549**(1-2):225-240.

81. Song GJ, Lee H, Park Y, Lee HJ, Lee YS, Seo JT, Kang IS: **Expression pattern of germ cell-specific genes in the testis of patients with nonobstructive azoospermia: usefulness as a molecular marker to predict the presence of testicular sperm**. *Fertil Steril* 2000, **73**(6):1104-1108.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

Overview

Gene targeting studies in the mouse demonstrate that glycolysis is required for sperm motility and male fertility [1]. While the glycolytic pathway is highly conserved, it is modified during spermatogenesis through the expression of novel isozymes derived from distinct genes or alternative splicing [2-7]. Genes with restricted expression in the male germline were generated by either gene duplication, as in *Gapdhs*, or retrotransposition, as in *Pgk2*, *Aldoart1*, and *Aldoart2* [2, 3, 6, 8].

Retrotransposition events which occur in the germline are carried on to future generations, accumulate in the genome, and can create new genes. Retrogenes expressed in the germ cells are acted upon by selective pressure in the context of reproductive fitness. Retrotransposition has created multiple testis-specific retrogenes, including at least three glycolytic enzymes: *Pgk2, Aldoart1, Aldoart2* [2, 3, 6]. The identification of *Aldoart1* also provided evidence for an N-terminal extension encoded by a splice variant of *Aldoa*, *Aldoa_v2*. Both ALDOA_V2 and GAPDHS contain N-terminal extensions when compared to their somatic counterpart and are expressed across mammalian species [6, 7]. There is evidence that these N-terminal extensions facilitate binding to the fibrous sheath, a cytoskeletal structure which defines the limits of the principal piece. Many glycolytic enzymes are tightly bound to the fibrous sheath, including ALDOA, GAPDHS, LDHA, and PKM2 [9]. Other glycolytic enzymes, including HK1, ENO1, and LDHC, are also localized in the principal piece [4, 10-12]. Thus, glycolysis is compartmentalized in a distinct segment of

the sperm flagellum away from the middle piece where oxidative phosphorylation occurs in mitochondria.

The studies presented in this dissertation detail the expression of three spermatogenic-cell specific aldolase A isozymes generated by retrotransposition and alternative splicing. Two of the ALDOA-related proteins contain N-terminal extensions and are tightly bound to the fibrous sheath. Analysis of tertiary structure and kinetic properties of the individual sperm ALDOA-related enzymes demonstrated a lack of tetramer formation and reduced activity when expressed in *E. coli*. Since retrotransposition is responsible for the creation of at least three spermatogenic-cell specific glycolytic enzymes, we analyzed the remainder of the pathway in the mouse and human genomes for retroposed copies of genes encoding glycolytic enzymes. We found frequent retrotransposition of orthologous genes in both species. Analysis of these retroposed sequences demonstrated independent divergence in the human and mouse genomes and an abundance of flanking LINE and LTR elements, supporting the role of autonomous repetitive elements in the retrotransposition of processed mRNA transcripts.

Kinetic analysis and targeted localization of novel spermatogenic-cell specific ALDOA isozymes

Two aldolase isozymes, ALDOA_V2 and ALDOART1, contain N-terminal extensions, and ALDOA_V2 is conserved in primates and rodents. Proteomic analysis demonstrated tight binding of both ALDOA_V2 and ALDOART1 to the fibrous sheath in mouse sperm, suggesting a role for the N-terminal extension in localization [6]. Most of the aldolase activity in mouse sperm is found in insoluble fractions following Triton-X-100 and ß-mercaptoethanol extraction. Since mouse sperm contain all three ALDOA-related proteins (ALDOART1, ALDOART2, and ALDOA_V2), measuring activity in whole lysates does not

distinguish the activities of individual isozymes. It is particularly interesting to analyze the kinetic properties of ALDOA_V2 since this isozyme is conserved across species and is expressed in human sperm [6]. Recombinant forms of the sperm ALDOA isozymes expressed in *E. coli* do not form tetramers and activity is significantly reduced when compared to ALDOA.

Most of the endogenous aldolase activity was found in the insoluble fraction. Many proteins expressed in sperm, including endogenous ALDOA_V2 and ALDOART1, are highly insoluble. The fibrous sheath itself is an insoluble structure resistant to urea extraction [9]. However, recombinant forms of sperm ALDOA-related proteins are soluble when expressed in *E. coli*, providing evidence that insolubility of sperm ALDOA isozymes is due to fibrous sheath binding.

Recombinant forms of the sperm ALDOA-related proteins did not form tetramers and had reduced activity compared to ALDOA, which may highlight distinctive properties of sperm enzymes which cannot be recapitulated in a bacterial expression system. Class I ALDOA is expressed in all eukaryotes and some prokaryotes, but the sperm ALDOA-related proteins are expressed only in some mammals, including rodents and primates [6, 13]. While we have utilized multiple *E. coli* host strains to attempt to promote expression of rare codons, it may be advantageous to design expression constructs using codon optimization to promote higher levels of expression, which has been successful with recombinant GAPDHS expression ([14] and P. Danshina, unpublished data). Also, a mammalian expression system may help to facilitate tetramer formation of sperm ALDOA-related proteins.

It is possible that the sperm aldolase isozymes require post-translational modifications and/or disulfide bond formation for proper folding, which can only be recapitulated in a mammalian expression system. The fibrous sheath is stabilized by disulfide bonds, and disulfide bonds also form during sperm maturation in the epididymis

[15, 16]. While ALDOA does not require disulfide bonds for proper folding and tetramer formation, the sperm ALDOA-related proteins may need additional modifications to promote proper folding, tetramer formation and subsequent activity.

Expression of sperm ALDOA-related proteins in *E. coli* produced slightly lower yields than ALDOA, and we have yet to confirm by circular dichroism if these proteins are properly folded. Therefore, it is also possible that folding difficulties contributed to reduced activity of sperm ALDOA-related proteins. If we can overcome experimental difficulties with protein expression, measuring the kinetic properties of recombinant sperm ALDOA-related proteins would allow us to determine whether each isozyme possesses distinct catalytic functions. Amino acid alignment [6] and modeling data suggests that ALDOART1 and ALDOART2 may have ALDOB-like function, so it would be advantageous to measure both glycolytic (with fructose-1,6-bisphosphate as a substrate) and fructolytic (with fructose-1-phosphate as a substrate) activity for these isozymes.

Compartmentalization of glycolytic enzymes to the fibrous sheath may help to facilitate effective glycolytic ATP production in the flagellum. It is possible that these enzymes form a complex which may help to “shuttle” substrates through the pathway in an efficient manner[17]. Several glycolytic enzymes are tethered to the fibrous sheath [9]. Similar targeted localization of glycolytic enzymes is seen in human erythrocytes, another glycolysis-dependent cell type. In the erythrocyte, GAPDH, ALDOA, PFK, and LDHA all bind directly to another protein, Band 3, which acts as a scaffold to form a glycolytic enzyme complex [18, 19]. In *Drosophila* both GAPDH and ALDOA are localized to the Z-discs and M-lines in flight muscles, again supporting localized ATP production in glycolysis-dependent fast moving muscles [20, 21]. Finally, ALDOA expressed in *Trypanosome brucei* possesses a ten residue N-terminal extension that is hypothesized to direct localization to the glycosome, an organelle containing the glycolytic enzymes [22]. In a similar fashion, the expression of novel ALDOA-related proteins with unique N-terminal extensions may direct

proper complex formation or localization of glycolytic enzymes to the principal piece of mammalian sperm, facilitating high levels of ATP production.

Multiple glycolytic enzymes are tightly bound to the sperm fibrous sheath, including GAPDHS, ALDOA, LDHA, and PKM2 [9]. Immunoblot analysis indicates that the larger form of each variant is present in the fibrous sheath fraction following purification [9]. GAPDHS, ALDOA_V2 and ALDOART1 contain N-terminal extensions when compared to their somatic counterparts [6, 7]. We have also recently identified a splice variant of LDHA (LDHA_V2) which contains an N-terminal extension when compared to somatic LDHA. Proteomic analysis confirmed tight binding of LDHA_V2 to the fibrous sheath in mouse sperm. Evidence for an N-terminal extension of PKM2 in boar sperm also supports modification of this enzyme to promote tight binding to the fibrous sheath [23, 24]. The genomic location of sequence encoding the entire N-terminal extension has not been identified. However, a larger PKM2 immunoreactive band is found in the final fibrous sheath fraction, suggesting tight binding of this enzyme via an N-terminal extension [9]. These studies suggest that N-terminal extensions may promote tight binding of several glycolytic enzymes to the fibrous sheath for localized ATP production along the sperm flagellum.

Our lab is currently using transgenic and knockout mice to address the role of N-terminal extensions in tight binding to the fibrous sheath. GAPDHS demonstrates restricted localization to the principal piece [25]. Transgenic mice containing FLAG-tagged GAPDHS with and without the N-terminal extension have been generated and crossed to mice lacking the *Gapdhs* gene (Z. Huang, unpublished data). Only transgenic GAPDHS with the N-terminal extension was localized exclusively in the principal piece, supporting an essential role of the N-terminal extension in correct targeting of this isozyme.

If the N-terminal extensions in ALDOA_V2 and ALDOART1 also promote restricted expression to the principal piece and tight binding to the fibrous sheath, this conserved function would be an example of convergent evolution. Using a similar transgenic/knockout

approach (as with GAPDHS) to address the role of the N-terminal extension in ALDOA_V2 and ALDOART1 would be extremely difficult, since at least three endogenous ALDOA isoforms are expressed in sperm [6]. Targeted gene disruption of the individual sperm ALDOA isozymes would likely have no phenotype due to functional redundancy, and would require a triple knockout of all three variants. While it would be possible to disrupt both the ALDOART1 and ALDOART2 genes, targeted disruption of the exon encoding the alternatively spliced exon of ALDOA_V2 could be more difficult.

To look for protein-protein interactions and complex formation, we can purify endogenous sperm ALDOA-related proteins in soluble fractions from testis lysates combined with gel filtration chromatography. Preliminary data from this approach demonstrated co-fractionation of the larger immunoreactive ALDOA-related band, containing ALDOA_V2 and ALDOART1, with GAPDHS (K. Miki, unpublished data). This data suggests that glycolytic enzymes tightly bound to the fibrous sheath directly interact or form a complex, possibly via N-terminal extensions. GST pull down assays using both testicular and sperm lysates did not confirm a direct interaction between any of the sperm ALDOA-related proteins and GAPDHS or AKAP4, both abundant proteins of the fibrous sheath (data not shown).

Retrotransposition creates genes with novel functions implicated in reproductive fitness

Comparative genomics and expression analysis of genes encoding the glycolytic enzymes supports retrotransposition as a mechanism to create new genes with novel functions. Retrotransposition events are only transmitted when they occur in the germline. If new genes generated by retrotransposition are expressed in the germline, they undergo selective pressure in the germline. Expression of genes in spermatogenic cells can have a direct effect reproductive fitness, in the context of proper spermatogenesis and male

fertility. In fact, genes that are involved in reproduction and spermatogenesis are under strong positive selection and include some the most rapidly evolving genes across multiple species [26-28].

When analyzing a particular gene of interest at the molecular level it is extremely important to identify all genes expressed, keeping in mind the possible redundant function of multiple expressed isoforms. Previously identified testis-specific retrogenes include *Pgk2*, *Aldoart1*, and *Aldoart2* [2, 3, 6]. In our study we detected expression of an alternative *Gpi1* transcript in mouse spermatogenic cells. It is possible that the remaining newly identified retroposed sequences may be expressed in other tissues, as with *PGAM1-rs7* which is transcribed in human leukocytes. In particular, cells and tissues that preferentially depend upon glycolysis for ATP may benefit from modification of the glycolytic pathway, such as mitochondria-deficient red blood cells and fast skeletal muscle fibers [29, 30]. Performing a similar genomic and expression analysis for genes in other pathways modified during spermatogenesis, such as oxidative phosphorylation or tyrosine phosphorylation, may detect expression of additional novel genes. Our analysis confirmed recent, independent divergence of retroposed sequences in the human and mouse genomes. Lineage-specific retrotransposition events suggest that similar genomic analysis in other species would be useful to identify all expressed genes and conserved, functionally important glycolytic enzyme variants.

Only certain genes in the glycolytic pathway are retroposed, and there is preliminary data supporting expression of these genes in testicular germ cells. Identification of the cell types where the parent genes are expressed during spermatogenesis may help to pinpoint the spermatogenic cell type where retrotransposition occurs. Retrotransposition that occurs during meiosis can take advantage of double strand breaks, chromatin remodeling, and reduced DNA methylation [31]. In fact, the open reading frames expressed from LINE elements are expressed during meiosis in spermatocytes [32]. A recent study demonstrated

active retrotransposition of human and mouse LINE elements during spermatogenesis, however the majority of these events occurred following fertilization in early stages of embryogenesis [33].

A more detailed understanding of sperm metabolism and its regulation has potential clinical implications, particularly for assessing potential causes of male infertility. It is likely that disruptive mutations in genes that encode sperm glycolytic enzymes would impair energy production and sperm motility. Each step of the glycolytic pathway is catalyzed by multiple isozymes which complicates the genetic analysis for genes contributing to male infertility. Annotation of all genes, retrogenes and pseudogenes matching glycolytic enzymes will facilitate targeted sequencing of human genes encoding glycolytic enzymes in patients with male factor infertility. Identification of candidate mutations in these genes would provide insights into the genetic basis of some cases of male infertility.

Enzymes are excellent pharmaceutical targets, since binding of an inhibitor molecule into the active can has the ability to selectively block enzyme activity. Sperm-specific glycolytic enzymes are novel contraceptive targets that could possibly be selectively inhibited to impair sperm ATP production and motility. GAPDHS is an example of one candidate contraceptive target. Early studies developing reversible contraceptives identified chlorinated compounds as inhibitors of GAPDHS, resulting in reduced glycolytic activity and motility [34]. In addition, targeted gene disruption of the *Gapdhs* gene resulted in dramatically reduced sperm motility associate with male infertility [1]. GAPDHS has unique kinetic properties when compared to GAPDH and can therefore be selectively inhibited (P. Danshina, unpublished data).

ALDOA_V2 is the only sperm ALDOA-related protein expressed in human sperm. The only difference between ALDOA_V2 and ALDOA is the N-terminal extension, so it is unlikely that ALDOA_V2 could be specifically inhibited without blocking the ubiquitously expressed ALDOA. However, completion of the kinetic analysis of ALDOA_V2 will help to

determine if this enzyme has distinct catalytic properties that can be selectively inhibited. Taken together, understanding the role of glycolytic isozymes in sperm energy metabolism will allow for the rational design of targeted contraceptive strategies and may open new avenues for the treatment of infertility.

References

1. Miki K, Qu W, Goulding EH, Willis WD, Bunch DO, Strader LF, Perreault SD, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S, a sperm-specific glycolytic enzyme, is required for sperm motility and male fertility**. *Proc Natl Acad Sci U S A* 2004, **101**(47):16501-16506.

2. Boer PH, Adra CN, Lau YF, McBurney MW: **The testis-specific phosphoglycerate kinase gene *pgk-2* is a recruited retroposon**. *Mol Cell Biol* 1987, **7**(9):3107-3112.

3. McCarrey JR, Thomas K: **Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene**. *Nature* 1987, **326**(6112):501-505.

4. Mori C, Nakamura N, Welch JE, Gotoh H, Goulding EH, Fujioka M, Eddy EM: **Mouse spermatogenic cell-specific type 1 hexokinase (*mHk1-s*) transcripts are expressed by alternative splicing from the *mHk1* gene and the HK1-S protein is localized mainly in the sperm tail**. *Mol Reprod Dev* 1998, **49**(4):374-385.

5. Mori C, Welch JE, Fulcher KD, O'Brien DA, Eddy EM: **Unique hexokinase messenger ribonucleic acids lacking the porin-binding domain are developmentally expressed in mouse spermatogenic cells**. *Biol Reprod* 1993, **49**(2):191-203.

6. Vemuganti SA, Bell TA, Scarlett CO, Parker CE, de Villena FP, O'Brien DA: **Three male germline-specific aldolase A isozymes are generated by alternative splicing and retrotransposition**. *Dev Biol* 2007, **309**(1):18-31.

7. Welch JE, Schatte EC, O'Brien DA, Eddy EM: **Expression of a glyceraldehyde 3-phosphate dehydrogenase gene specific to mouse spermatogenic cells**. *Biol Reprod* 1992, **46**(5):869-878.

8. Welch JE, Brown PR, O'Brien DA, Eddy EM: **Genomic organization of a mouse glyceraldehyde 3-phosphate dehydrogenase gene (Gapd-s) expressed in post-meiotic spermatogenic cells**. *Dev Genet* 1995, **16**(2):179-189.

9. Krisfalusi M, Miki K, Magyar PL, O'Brien D A: **Multiple glycolytic enzymes are tightly bound to the fibrous sheath of mouse spermatozoa**. *Biol Reprod* 2006, **75**(2):270-278.

10. Beyler SA, Wheat TE, Goldberg E: **Binding of antibodies against antigenic domains of murine lactate dehydrogenase-C4 to human and mouse spermatozoa**. *Biol Reprod* 1985, **32**(5):1201-1210.

11. Gitlits VM, Toh BH, Loveland KL, Sentry JW: **The glycolytic enzyme enolase is present in sperm tail and displays nucleotide-dependent association with microtubules**. *Eur J Cell Biol* 2000, **79**(2):104-111.

12. Travis AJ, Foster JA, Rosenbaum NA, Visconti PE, Gerton GL, Kopf GS, Moss SB: **Targeting of a germ cell-specific type 1 hexokinase lacking a porin-binding domain to the mitochondria as well as to the head and fibrous sheath of murine spermatozoa**. *Mol Biol Cell* 1998, **9**(2):263-276.

13. Marsh JJ, Lebherz HG: **Fructose-bisphosphate aldolases: an evolutionary history**. *Trends Biochem Sci* 1992, **17**(3):110-113.

14. Wu ZL, Qiao J, Zhang ZG, Guengerich FP, Liu Y, Pei XQ: **Enhanced bacterial expression of several mammalian cytochrome P450s by codon optimization and chaperone coexpression**. *Biotechnol Lett* 2009.

15. Bedford JM, Calvin HI: **Changes in -S-S- linked structures of the sperm tail during epididymal maturation, with comparative observations in sub-mammalian species**. *J Exp Zool* 1974, **187**(2):181-204.

16. Shalgi R, Seligman J, Kosower NS: **Dynamics of the thiol status of rat spermatozoa during maturation: analysis with the fluorescent labeling agent monobromobimane**. *Biol Reprod* 1989, **40**(5):1037-1045.

17. Ouporov IV, Knull HR, Huber A, Thomasson KA: **Brownian dynamics simulations of aldolase binding glyceraldehyde 3-phosphate dehydrogenase and the possibility of substrate channeling**. *Biophys J* 2001, **80**(6):2527-2535.

18. Campanella ME, Chu H, Low PS: **Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane**. *Proc Natl Acad Sci U S A* 2005, **102**(7):2402-2407.

19. Chu H, Low PS: **Mapping of glycolytic enzyme-binding sites on human erythrocyte band 3**. *Biochem J* 2006, **400**(1):143-151.

20. Sullivan DT, MacIntyre R, Fuda N, Fiori J, Barrilla J, Ramizel L: **Analysis of glycolytic enzyme co-localization in *Drosophila* flight muscle**. *J Exp Biol* 2003, **206**(Pt 12):2031-2038.

21. Wojtas K, Slepecky N, von Kalm L, Sullivan D: **Flight muscle function in *Drosophila* requires colocalization of glycolytic enzymes**. *Mol Biol Cell* 1997, **8**(9):1665-1675.

22. Clayton CE: **Structure and regulated expression of genes encoding fructose biphosphate aldolase in Trypanosoma brucei**. *Embo J* 1985, **4**(11):2997-3003.

23. Feiden S, Stypa H, Wolfrum U, Wegener G, Kamp G: **A novel pyruvate kinase (PK-S) from boar spermatozoa is localized at the fibrous sheath and the acrosome**. *Reproduction* 2007, **134**(1):81-95.

24. Feiden S, Wolfrum U, Wegener G, Kamp G: **Expression and compartmentalisation of the glycolytic enzymes GAPDH and pyruvate kinase in boar spermatogenesis**. *Reprod Fertil Dev* 2008, **20**(6):713-723.

25. Bunch DO, Welch JE, Magyar PL, Eddy EM, O'Brien DA: **Glyceraldehyde 3-phosphate dehydrogenase-S protein distribution during mouse spermatogenesis**. *Biol Reprod* 1998, **58**(3):834-841.

26. Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ *et al*: **A scan for positively selected genes in the genomes of humans and chimpanzees**. *PLoS Biol* 2005, **3**(6):e170.

27. Swanson WJ, Vacquier VD: **The rapid evolution of reproductive proteins**. *Nat Rev Genet* 2002, **3**(2):137-144.

28. Torgerson DG, Kulathinal RJ, Singh RS: **Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes**. *Mol Biol Evol* 2002, **19**(11):1973-1980.

29. Gronowicz G, Swift H, Steck TL: **Maturation of the reticulocyte in vitro**. *J Cell Sci* 1984, **71**:177-197.

30. Peter JB, Barnard RJ, Edgerton VR, Gillespie CA, Stempel KE: **Metabolic profiles of three fiber types of skeletal muscle in guinea pigs and rabbits**. *Biochemistry* 1972, **11**(14):2627-2633.

31. Bestor TH, Tycko B: **Creation of genomic methylation patterns**. *Nat Genet* 1996, **12**(4):363-367.

32. Branciforte D, Martin SL: **Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition**. *Mol Cell Biol* 1994, **14**(4):2584-2592.

33. Kano H, Godoy I, Courtney C, Vetter MR, Gerton GL, Ostertag EM, Kazazian HH, Jr.: **L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism**. *Genes Dev* 2009, **23**(11):1303-1312.

34. Jones AR, Cooper TG: **A re-appraisal of the post-testicular action and toxicity of chlorinated antifertility compounds**. *Int J Androl* 1999, **22**(3):130-138.