Evolution of glucose utilization: Glucokinase and glucokinase regulator protein

David M. Irwin^{a,b,c,*} and Huanran Tan^c

^aDepartment of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

^bBanting and Best Diabetes Centre, University of Toronto, Toronto, Canada

^cDepartment of Pharmacology, Peking University, Health Sciences Centre, Beijing, China

Abstract

Glucose is an essential nutrient that must be distributed throughout the body to provide energy to sustain physiological functions. Glucose is delivered to distant tissues via the blood stream, and complex systems have evolved to maintain the levels of glucose within a narrow physiological range. Phosphorylation of glucose, by glucokinase, is an essential component of glucose homeostasis, both from the regulatory and metabolic point-of-view. Here we review the evolution of glucose utilization from the perspective of glucokinase. We discuss the origin of glucokinase, its evolution within the hexokinase gene family, and the evolution of its interacting regulatory partner, glucokinase regulatory protein (GCKR). Evolution of the structure and sequence of both glucokinase and GCKR have been necessary to optimize glucokinase in its role in glucose metabolism.

Keywords

Glucokinase (GCK); Hexokinase (HK); Glucokinase regulatory protein (GCKR); Liver; Glucose

1. Mammalian glucose metabolism

Glucose is an essential energy molecule for many species as it can readily be metabolized to generate ATP by either aerobic or anaerobic respiration. In mammals, several tissues, such as the brain, depend upon glucose for the generation of ATP for cellular functions (Wasserman, 2009; Thorens, 2011). To supply these tissues, glucose is distributed by the blood stream, and complex systems have evolved to maintain appropriate blood glucose levels in the face of changes due to the availability of food and the expenditures of energy (Suh et al., 2007; Wasserman, 2009; Polakof et al., 2011; Thorens, 2011). The diet is the ultimate source of glucose in mammals, where it is either directly absorbed from the digested food in the intestine or is generated by gluconeogenesis (generally from amino acids) from precursors obtained from the diet. Whether glucose comes directly from the diet,

*Address for correspondence: David M. Irwin, Department of Laboratory Medicine and Pathobiology, Room 6211 Medical Sciences Building, 1 King's College Circle, University of Toronto, Toronto, Ont., Canada, M5S 1A8, Tel: +1 416 978-0519, david.irwin@utoronto.ca.

or is synthesized via gluconeogenesis, depends upon the carbohydrate content of the diet. Counter-regulatory hormone systems have evolved to maintain blood glucose within a narrow range. Insulin promotes glucose uptake from the blood by diverse tissues when blood glucose levels are high; e.g., after feeding, while glucagon induces gluconeogenesis by the liver to release glucose and prevent hypoglycemia (Bansal and Wang, 2008; Wasserman, 2009). Additional physiological systems, including other hormones and the central nervous system, contribute to the regulation of blood glucose levels (Polakoff et al., 2011; Thorens, 2011; Grayson et al., 2013). Excess energy obtained from the diet is stored as glycogen in the liver and muscle and as lipid in adipose tissue (Klover and Mooney, 2004). Glycogen in muscle can be rapidly broken down to release glucose when it is needed by muscle tissue (Jensen and Richter, 2012); similarly, glycogen is a source for the production of glucose that can be released by the liver to maintain blood glucose levels (Klover and Mooney, 2004).

While glucose can be dispersed to distant tissues of the body through the blood, it still must enter cells, as it cannot cross the lipid membranes of cells by simple diffusion, as it is hydrophilic. In mammals, glucose is transported across the cell membrane by transporters, which belong to one of two families: the glucose transporters (SLC2 or GLUT gene family) (Augustin, 2010; Wilson-O'Brien et al., 2010; Mueckler and Thorens, 2013) and the sodium-coupled glucose transporters (SGLT or SLC5 gene family) (Wright et al., 2011; Wright, 2013). Members of the SGLT are expressed in the kidney and intestine where they actively transport glucose and allow reabsorption against a concentration gradient (Wright et al., 2011; Wright, 2013). The SLC2 family members (or GLUT proteins) are facilitative transports and only move glucose in the direction of a concentration gradient (Augustin, 2010; Mueckler and Thorens, 2013). The diverse members of the SLC2 family have varying substrate specificities, kinetics, and expression profiles allowing cells to have tissue-specific differences in their glucose uptake (Augustin, 2010; Mueckler and Thorens, 2013).

2. Glucokinase and the vertebrate hexokinase gene family

Once glucose (or other simple sugars, e.g., fructose) enter cells it is phosphorylated. Phosphorylation has two purposes: (1) the first step in metabolism, and (2) decreasing the intracellular concentration of the unphosphorylated form of the sugar, thus driving the uptake of the sugar from the external environment. Enzymes, called hexokinases, phosphorylate glucose, and other six-carbon sugars. In mammals, and other vertebrates, four hexokinase isozymes have been identified (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). The different isozymes of hexokinase were initially distinguished by letters (i.e., hexokinase (HK) A, B, C, and D) based on their elution time from DEAE cellulose columns (González et al., 1964), but subsequently given numbers (i.e., hexokinase I, II, III, and IV) based on their migration in electrophoretic gels (Katzen et al., 1965). Hexokinase IV (or D) is most often called glucokinase (GCK), although it is not specific for glucose (Cárdenas et al., 1998; Wilson, 2004). Genes encoding these hexokinases use Arabic numbers; e.g., HK1 encodes hexokinase I. Mammalian hexokinases have been extensively characterized, with possibly their most striking difference being their molecular weights (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). Hexokinases I, II, and III have a molecular weight of approximately 100kD, while glucokinase has a molecular mass of about 50kD. Hexokinases having a molecular weight of 50kD, but not 100kD, have

also been found, and characterized, in many other eukaryotes species, including non-vertebrate animals, plants, and yeast (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). In some of these non-mammalian species multiple isozymes have been characterized (Wilson, 1995, 2004; Cárdenas et al., 1998). While enzymes that phosphorylate sugars have also found in bacteria, these sequences show no significant sequence similarity to the eukaryotic hexokinases. However there is some similarity in the three-dimensional structures of hexokinases from bacteria and eukaryotes, which has been used to suggest that they share a common ancestor (Bork et al., 1993; Cárdenas et al., 1998; Kawai et al., 2005).

Sequences of hexokinases were initially deduced and predicted based on cDNA clones (Andreone et al., 1989; Schwab and Wilson, 1989, 1991; Griffin et al., 1991; Thelen and Wilson, 1991). Analysis of genome sequence data identified, in contrast to the 4 expected known hexokinase enzymes (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998), a total of 6 hexokinase-like genes in the human genome, and 5 in the genomes of most other vertebrates (Irwin and Tan, 2008; González-Alvarez et al., 2009). In addition to the expected genes encoding the known hexokinases (HK1, HK2, HK3, and GCK), a fifth hexokinase-like gene was found in the genome searches that is conserved throughout vertebrates. The fifth hexokinase-like gene was named hexokinase domain containing 1 (HKDC1), however, the biological function of this gene is currently unknown. SNPs near the HKDC1 gene have been associated with attention-deficit/hyperactivity disorder (Neale et al., 2010) and the protein was found to potentially interact with iNOS2 (Foster et al., 2013). The human genome, as well as those of a few other primates, contains a sixth hexokinase-like sequence, which is a reverse-transcribed pseudogene copy of the HK2 gene (Ardehali et al., 1995; Irwin and Tan, 2008). The HK2 pseudogene was generated relatively recently during primate evolution (Ardehali et al., 1995; Irwin and Tan, 2008).

Characterization of the protein sequences of the hexokinases provided an explanation for the difference in the molecular weights of the 100kD hexokinases (hexokinases I, II and III) and the 50kD glucokinase – the larger hexokinases contain two kinases domains, while the smaller glucokinase has only a single kinase domain (Andreone et al., 1989; Schwab and Wilson, 1989, 1991; Griffin et al., 1991; Thelen and Wilson, 1991; Cárdenas et al., 1998). This conclusion is illustrated in Fig. 1, where similarity between the different kinase domains is shown as long diagonal lines in dotplots. Glucokinase contains only a single kinase domain while hexokinase I has two. When hexokinase I is compared to glucokinase (Fig. 1A) two diagonal lines are generated since the N-terminal half of hexokinase I is similar to all of glucokinase (lower diagonal, Fig. 1A), and the C-terminal half of hexokinase I also is similar to all of glucokinase (upper diagonal in Fig. 1A). If glucokinase is compared to itself (Fig. 1B), no repetitive structure is identified, only similarity to itself along the entire sequence (diagonal from lower left to upper right), although there are short sequences that show limited similarity (short lines that are off the central diagonal, Fig. 1, also seen in the other comparisons). If hexokinase I is compared to itself (Fig. 1C) or to hexokinase II (Fig. 1D) a total of three long diagonal lines are seen due to the repetitive structures in both proteins. The central diagonal line indicates that the two proteins are similar to each other over their entire lengths, while the diagonal lines in the upper left and lower right indicate that similarity between the N-terminal and C-terminal halves of the is

observed. Similar results are observed with the hexokinase III and HKDC1 (which also predicts an approximately 100kD protein) sequences (results not shown).

Isolation and partial characterization of the genes encoding hexokinase II and glucokinase (Magnuson et al., 1989; Thelen and Wilson, 1991; Kogure et al., 1993) strengthened the conclusion that the larger hexokinases were generated by duplication of the kinase domain. The intron-exon structures of the two halves of the HK2 gene are similar to each other and to the glucokinase gene, where the sizes of the exons are similar and the phase of the codons interrupted by the introns are identical (Fig. 2). These observations suggesting that the two halves of HK2 were generated by a duplication of glucokinase-like sequence (Magnuson et al., 1989; Thelen and Wilson, 1991; Kogure et al., 1993). The similarity in intron-exon structure is seen not only for HK2, but also for all other members of the mammalian hexokinase gene family (Fig. 2). Hexokinase genes from non-vertebrate species also share similarities in gene structure with glucokinase (Fig. 2).

3. Origin of the vertebrate hexokinase gene family

Multiple hexokinases have been found in vertebrates, plants and yeast (Wilson, 1995; Cárdenas et al., 1998). A hexokinase in both vertebrates and yeast has, unfortunately, been named glucokinase. The sharing of the name glucokinase has suggested to some that the vertebrate and yeast glucokinases are more closely related to each other than to other hexokinases (see Cárdenas et al., (1998) for examples), however, analyses of hexokinase and glucokinase sequences from mammals, yeast, and plants show that mammalian glucokinase is more closely related to other mammalian hexokinases and yeast glucokinase is more closely related to other yeast hexokinases than the two glucokinases are to each other (Griffin et al., 1991; Bork et al., 1993; Wilson, 1995; Cárdenas et al., 1998). Hexokinases from non-vertebrate species have a molecular weight of about 50kD, indicating that the 100kD mammalian hexokinases originated on the vertebrate lineage (Cárdenas et al., 1998). It was suggested that a duplication of the hexokinase domain occurred within an ancestral hexokinase gene in a pre-vertebrate ancestor (Ureta, 1982). The intragenic duplication of the hexokinase domain also occurred at about the same time as the gene duplications that generated the diverse hexokinase genes (Ureta, 1982). How glucokinase was related to the 100kD hexokinase could not be resolved, thus Ureta (1982) suggested two possible evolutionary relationships: (1) glucokinase diverged prior to the duplication of the hexokinase domain or (2) glucokinase lost a domain after duplication (see Fig. 3A).

With the cloning and sequencing of mammalian hexokinase genes (Andreone et al., 1989; Magnuson et al., 1989; Schwab and Wilson, 1989, 1991; Printz et al., 1993a) it became possible to better investigate the origin and evolution of these genes. Griffin et al. (1991) compared the sequences of the two halves of mammalian hexokinase I to mammalian glucokinase and yeast hexokinases, and showed that mammalian glucokinase is more closely related to mammalian hexokinase I than to any of the yeast hexokinases (Griffin et al., 1992). This conclusion is in agreement with the multiple mammalian and yeast hexokinases have independent origins (Griffin et al., 1991; Bork et al., 1993; Wilson, 1995; Cárdenas et al., 1998). The phylogeny presented by Griffin et al. (1992) was consistent with a relatively ancient origin of the 100kD hexokinase, and with the glucokinase sequence being more

closely related to the C-terminal, rather than the N-terminal, portion of hexokinase I. The study by Griffin et al. (1991) implied that the duplication of the hexokinase domain occurred before the divergence of glucokinase from hexokinase I (Griffin et al., 1991), consistent with first possibility suggested by Ureta (1982) (see Fig. 3A left panel). Bork et al. (1993) followed this with a study that used the sequences of all four mammalian hexokinases, and while again they were able to show independent origins for the mammalian and yeast hexokinase gene families, they were unable to resolve the relationships among the mammalian sequences. Cárdenas et al. (1998) unfortunately used the UPGMA (unweighted pair-group method with arithmetic mean (Sokal and Michener, 1958)) method to calculate a tree for mammalian hexokinases, using protein sequences distances calculated by Fothergill-Gilmore and Michels (1993). As acknowledged by the authors (Cárdenas et al., 1998), UPGMA assumes that the sequences have evolved at the same rate, but it was already clear from the distance data (Fothergill-Gilmore and Michels, 1993) that the N-terminal domains of hexokinases evolve more rapidly than the C-terminal domains. The UPGMA distance tree generated by Cárdenas et al. (1998), unsurprisingly, yielded an unexpected tree where the two halves of hexokinase III have different origins. The authors attempted to rationalize their finding by placing glucokinase as the earliest diverging lineage, a phylogeny that is in accord with the second model presented by Ureta (1982) (see Fig. 3A, right panel).

With the advent of genome sequencing, hexokinase gene sequences are now available from a far greater number of vertebrate species, as well as from diverse non-vertebrate animal species. As indicated above, searches of vertebrate genome sequences revealed the existence of a novel fifth member of the hexokinase gene family, the hexokinase domain containing 1 (HKDC1) gene (Irwin and Tan, 2008). The HKDC1 gene had not been considered in earlier phylogenetic analyses (Griffin et al., 1991; Bork et al., 1993; Cárdenas et al., 1998). Phylogenetic analysis of this larger diverse set of vertebrate hexokinase genes, combined with the use of outgroups that were closer than the yeast and plant sequences previously used, yielded well supported phylogenies similar to that shown in Fig. 3B (Irwin and Tan, 2008). A similar phylogenetic conclusion was reached in an analysis of hexokinase genes from zebrafish and mammals (González-Alvarez et al., 2009). Both Irwin and Tan (2008) and González-Alvarez et al. (2009) found phylogenetic trees where the divergence of the amino- and carboxy-terminal hexokinase domains of the 100kD hexokinases is the most ancient, and with glucokinase being most closely related to the C-terminal hexokinase domains (Fig. 3B). The HKDC1 gene, in both studies, is most closely related to the hexokinase 1 gene (Irwin and Tan, 2008; González-Alvarez et al., 2009). Among the remaining hexokinases, hexokinase 2 is most closely related to hexokinase 1, with hexokinase 3 being most divergent (Fig. 3B), in agreement with earlier analyses (Bork et al., 1993; Cárdenas et al., 1998). The use of gene sequences from multiple vertebrate species and the use of closer outgroup species yielded a more confident resolution of the phylogeny and placement of the glucokinase sequence as being most closely related to the C-terminal domains. This model of the evolution of glucokinase is in agreement with that shown in the left of Fig. 3A, as suggested by Ureta (1982), and agrees that glucokinase, along with the other three hexokinase isozymes diverged early in vertebrate evolution (Irwin and Tan, 2008).

Duplication of the genes for glucokinase and the three other hexokinase isozymes prior to the earliest divergence of vertebrates would be consistent with these genes originating via the pair of genome duplications (2R hypothesis) that occurred on the early vertebrate lineage (Furlong and Holland, 2002; Hokamp et al., 2003; Panopoulou and Poustka, 2005). While the phylogeny of glucokinase and hexokinase 1, 2, and 3 is not in complete agreement with the 2R hypothesis, this pattern is commonly seen with genes that originated at this time (Friedman and Hughes, 2001; Furlong and Holland, 2002). Further support for the origin of these four genes via genome duplication is the observation that the glucokinase and hexokinase genes are dispersed on four different chromosomes (Irwin and Tan, 2008). HKDC1 appears to be a more recent product of the duplication of the hexokinase I gene and was via a tandem gene duplication, as these two genes are arranged head-to-tail in all vertebrate genomes examined (Irwin and Tan, 2008). Despite being a more recent gene duplication event, this must still have occurred very early in vertebrate evolution as hexokinase 1 and HKDC1 genes are found in all vertebrate classes examined (Irwin and Tan, 2008; González-Alvarez et al., 2009).

4. Evolution of the vertebrate hexokinase gene family

Combining gene structure and phylogenetic data allows an inference of the events involved in the evolution of the vertebrate 100kD hexokinases and glucokinase. The ancestor of vertebrate hexokinases had a single hexokinase domain and a molecular weight of about 50kD (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998; Irwin and Tan, 2008). The phylogenetic analysis indicates that a tandem duplication of the hexokinase domain within the hexokinase gene yielded a 100kD hexokinase the common ancestor of vertebrates. This 100kD hexokinase was the ancestor of all vertebrate hexokinases, including glucokinase, and contained two hexokinase domains as shown in the left panel of Fig. 3A as proposed by Ureta (1982). How did the domain duplication occur? Comparison of the intron-exon structure of hexokinase genes (see Fig. 2) showed that exons 11 through 17 are similar in size to exons 3 through 9 (Magnuson et al., 1989; Thelen and Wilson, 1991; Kogure et al., 1993). Sequences encoded by exon 1 are unique and present in a single copy. While the 5' end of exon 10 encodes protein sequence similar to that encoded by exon 18, the 3' end encodes sequence similar to that of exon 2. This observation suggested, as shown in Fig. 4A, that a recombination event between exon 10 of a single-domain hexokinase gene and exon 2 of a second copy of this gene would result in a gene with a structure similar to the genes for HK1, HK2, HK3, and HKDC1. The different genes (HK1, HK2, HK3, and HKDC1) can then be generated by gene duplication events and subsequent sequence divergence (Fig. 4A). A gene with duplicated hexokinase domains is also the ancestor of the vertebrate glucokinase gene (Ureta, 1982), with the glucokinase sequence being most closely related to the C-terminal hexokinase domain (see section 3 and Fig. 3).

The glucokinase protein sequence is most closely related to the C-terminal domain of 100kD hexokinases, thus for glucokinase to evolve from a larger hexokinase ancestor it must have lost its N-terminal kinase domain. The genetic mechanism by which this occurred is currently unknown. Here we propose a possible model for the generation of a glucokinase gene (Fig. 4B). The C-terminal kinase domain of the 100kD hexokinases is encoded by exons 10 through 18. Glucokinase, however, is unlikely to have been generated simply for

the 3' half of a 100kD hexokinase genes. Exon 10, the 5' end of the second half of the two kinase-domain containing hexokinases, encodes sequences from both exon 2 and 10 of the ancestral single kinase domain-containing gene (Fig. 4A). The hybrid exon 10 was generated due to recombination generating the 100kD hexokinases (see Fig. 4A, exon 10 has sequences from exon 2a and 10b) and thus encodes sequences from the C-terminus of the hexokinase domain followed by sequences from near the N-terminus of the domain. The N-terminal sequence of glucokinase shows no similarity to the C-terminal sequences of kinases domains (see Fig. 1A), thus does not appear to have evolved from an exon 10-like sequence. Instead, exon 2 (and exon 1) of glucokinase appears to be more similar, in sequence and structure, to exon 2 (and exon 1) of the 100kD hexokinase genes. This suggests that the 5' end of the glucokinase gene is derived from the 5' end (exons 1 and 2) of a 100kD hexokinase gene (and would also provide a promoter for expression), while the 3' end of glucokinase is derived from the 3' end (exons 11–18, named 3b–10b in Fig. 4B) of the larger gene (since the glucokinase protein sequence is most closely related to the C-terminal hexokinase domain) (Fig. 4B). This model would imply an internal deletion of exons 3 to 10 (named 3a to 10a/2b in Fig. 4B). Whether this loss was initially due to the deletion of these exons, or a change in splicing pattern to skip these exons is unknown, however none of the extant glucokinase genes examined possesses genomic sequences bearing similarity to the lost exons (results not shown).

Enzymes encoded by members of the hexokinase gene family differ in their regulatory and enzymatic properties (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998), presumably due to changes in their amino acid sequences. Although hexokinases I, II, and III each have a pair of hexokinase domains, only hexokinase II retains hexokinase activity for both of these domains (Tsai and Wilson, 1996). Only the C-terminal kinase domains of hexokinase I and II have been demonstrated to have kinase activity (White and Wilson, 1989; Tsai and Wilson, 1997). Loss of the hexokinase activity may account for the more rapid evolution of the N-terminal domains of the hexokinase I and III protein sequences, and thus explain, at least in part, the observed greater divergence in sequence of the N-terminal, compared to C-terminal, domains of the 100kD hexokinases (Fothergill-Gilmore and Michels, 1993; Cárdenas et al., 1998). The more rapid evolution of the N-terminal regions of hexokinase I and III suggests that a smaller portion of these sequences are constrained by evolution, however, the observation that these sequence are retained in hexokinases from diverse vertebrates suggest that these sequences likely still have functional roles. Experimental studies have shown that while the N-terminal kinase domains of hexokinase I and III have lost kinase activity, they retain regulatory activity by being able to bind glucose-6-phosphate (Tsai and Wilson, 1997).

It is unknown whether HKDC1 has any hexokinase activity, as its amino acid sequence contains all of the known functional active residues (Irwin and Tan, 2008). Extensive searches of diverse tissues in many vertebrate species have not suggested the presence of any additional hexokinase activities that could not be accounted for by the known hexokinase isozymes (Ureta, 1982; Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). These observations suggest that HKDC1 does not phosphorylate glucose, and likely has a different enzymatic or regulatory function.

An important difference between glucokinase and the other hexokinases is that it has a much lower affinity for glucose, such that it is not saturated at physiological blood glucose concentrations (Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). In addition, glucokinase is not inhibited by its product, glucose-6-phosphate (Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). These unique properties of glucokinase, compared to other vertebrate hexokinases, means that the activity of glucokinase varies with glucose concentrations, thus can act as a sensor of glucose levels (Wilson, 1995, 1997, 2003, 2004; Cárdenas et al., 1998). Hexokinases I, II, and III likely function to immediately phosphorylate any glucose (or hexose sugar) that enters a cell, thus driving glucose metabolism and the survival of the cells. Glucokinase, in contrast, has been adapted for a regulatory role, where it can control physiological processes at the whole organism level. The changes in the sequences of glucokinase compared to hexokinases that account for these functional differences are currently unknown.

5. Glucokinase regulatory protein and the tissue-specific roles of glucokinase

While all hexokinases can phosphorylate glucose, glucokinase is the hexokinase with the most important role in glucose homeostasis in mammals. Glucokinase is expressed in tissues with essential roles in the regulation of blood glucose levels, and mutations in this gene, MODY2 mutations, lead to diabetes (Printz et al., 1993b; Iynedjian, 1993, 2009). Glucokinase is expressed in the liver, pancreatic islets, and select cells in the gut and the brain. In the liver glucokinase functions to regulate glucose metabolism, while in the pancreatic islets, intestinal cells and neurons it functions as a glucose sensor (Iynedjian, 1993, 2009; Postic et al., 2001; Agius, 2008). As a key component of the glucose sensor in pancreatic islet beta-cells, glucokinase regulates the release of insulin in response to blood glucose levels (Postic et al., 2001; Agius, 2008). In intestinal and neuronal cells, glucokinase functions similarly to regulate responses of these tissues to changes in blood glucose levels (Iynedjian, 2009). To function as a glucose sensor, glucokinase must always be present in a cell, thus it can instantly respond to change in blood glucose levels. In the liver, in contrast, glucokinase must only function when blood glucose levels are elevated; therefore its activity in these cells is regulated, where this is achieved at both the transcriptional and post-transcriptional level (Iynedjian, 1993, 2009).

Tissue-specific expression of glucokinase is mediated through a pair of unique promoters, where an upstream promoter drives expression in pancreatic beta-cells as well as select gut and neuronal cells, and a downstream promoter drives expression in the liver (Magnuson et al., 1989; Tanizawa et al., 1992; Postic et al., 1995). Little, if any, transcriptional regulation occurs at the upstream promoter that functions in pancreatic islet, intestinal and neuronal cells, likely due to these cells requiring a constant presence of glucokinase for glucose sensing (Postic et al., 2001; Iynedjian, 2009). In contrast, the liver-specific promoter is transcriptionally regulated, yielding glucokinase activity when blood glucose levels are elevated (Postic et al., 2001; Iynedjian, 1993, 2009). However, since changes in blood glucose levels can occur rapidly, at rates more rapid than can be addressed by transcriptional

regulation, changes in glucokinase activity levels is also regulated at the post-translational level (van Schaftingen et al., 1997; Agius, 2008; Iynedjian, 2009).

Post-translationally, glucokinase is regulated by an interaction with glucokinase regulatory protein (GCKR), where binding to GCKR results in the loss of enzymatic activity and localization to the nucleus, while active glucokinase is found in the cytoplasm and is not bound to GCKR (Agius and Peak, 1997; van Schaftingen et al., 1997). Levels of fructose-6-phosphate, which change in parallel with blood glucose levels, regulate the binding of glucokinase with GCKR, allowing glucokinase to be active only when blood glucose levels are elevated (Agius and Peak, 1997; van Schaftingen et al., 1997). GCKR is only abundantly expressed in the liver, and not appreciably in the other sites of glucokinase expression (Detheux et al., 1993; Vandercammen and Van Schaftingen, 1993). The tissue-specific expression of GCKR in the liver therefore permits glucokinase to be post-transcriptionally regulated only in the liver, the one tissue where activity must be regulated to prevent function when blood glucose levels are low.

6. Glucose metabolism and diet

Glucokinase is essential for sensing blood glucose levels, thus controls metabolism (Iynedjian, 1993, 2009; Postic et al., 2001; Agius, 2008). Loss of glucokinase activity in humans and mice leads to diabetes (Printz et al., 1993b; Postic et al., 2001; Agius, 2008; Iynedjian, 2009); however, this does not appear to be the case for all vertebrate species. Several mammals, and other vertebrate species, have been reported to be deficient in liver glucokinase activity (reviewed in Ureta, 1982). The initial reports of liver glucokinase deficiency were based largely on the electrophoretic separation of hexokinase isozymes (see Ureta, 1982). More recent investigations that used more sensitive approaches have often detected low levels of glucokinase in the livers of many of these species (e.g., Panserat et al., 2000a; Berradi et al., 2005; Hiskett et al., 2009), but confirm, at least in some species, that the levels of glucokinase in their liver are low. It has been argued that low levels of glucokinase in the liver of these species is unlikely to be due to inactivating mutations in the glucokinase gene, as these mutations would likely disrupt essential glucokinase function (i.e., regulating insulin secretion) in the pancreas (Cárdenas et al., 1998; Wang et al., 2013). These observations raise several questions, including what is the genetic basis for the low levels of glucokinase in the liver and what are functional consequences of the low level of activity (Cárdenas et al., 1998).

Reduced liver glucokinase levels, without changes in pancreatic levels, were observed in GCKR knockout mice (Farrelly et al., 1999; Grimsby et al., 2000). This result suggested that GCKR not only has a role in regulating glucokinase activity, via subcellular localization, but also enhances the stability of the enzyme. GCKR is also absent from the livers of the cat, a species who are deficient in liver glucokinase activity (Hiskett et al., 2009). These observations prompted searches for glucokinase and GCKR genes from the genomes of diverse vertebrate species, including those that do and do not have reported liver glucokinase activity (Wang et al., 2012, 2013). Potentially intact glucokinase genes could be found in every genome searched, however, mutated GCKR genes were found in the genomes of multiple species (Wang et al., 2012, 2013). GCKR genes that contain mutations that should

prevent function (i.e., frame-shift or splicing mutations) were found in species that were previously reported to lack (or have low) glucokinase activity in the liver; while potentially intact (i.e., have complete open reading frames and splice consensus sequences) GCKR genes were found in species that were reported to have activity (Wang et al., 2013). The correlation between the possessing an intact GCKR gene with glucokinase activity (and mutated gene with loss of activity) suggests that the loss of the GCKR may be the main molecular mechanism for deficiency of liver glucokinase activity. Intriguingly, loss of liver-specific glucokinase activity must have occurred multiple times, as neither loss of liver glucokinase activity nor disrupted GCKR genes are monophyletic, where both intact and mutated GCKR genes as well as presence and absence of liver glucokinase activity are interspersed in the vertebrate phylogeny (Wang et al., 2013).

Why would glucokinase activity, and thus GCKR genes, be dispensable? The liver only needs glucokinase activity if blood glucose levels can exceed the preferred level, and glucose levels would only exceed this level if too much is made by gluconeogenesis or if too much is absorbed from the diet. Since gluconeogenesis is a regulated process (Postic et al., 2001; Agius, 2008), this should not cause excess blood glucose levels. Acquiring glucose from the diet is not a regulated process, therefore it can lead to elevated blood glucose levels, but only if the diet is high in carbohydrate as they are broken down to sugars in the digestive tract. Species that lack liver glucokinase activity have a ruminant lifestyle that ferments plant material to generate short-chain fatty acids (e.g., cow and sheep), have elevated blood glucose levels (e.g., birds), or are carnivores (e.g., cat) (Hiskett et al., 2009; Aschenbach et al., 2010; Polakof et al., 2011; Verbrugghe et al., 2012). These species, though, will still require glucose sensing in pancreatic islets, as well as gut and neuronal cells, to prevent hypoglycemia, therefore must retain an intact glucokinase gene to function in those cells. The loss of a requirement for glucokinase in ruminants can be explained by the fact that these animals derive almost all of their energy needs from short chain fatty acids (Aschenbach et al., 2010). Very little glucose is acquired from the ruminant diet, thus they do not need to remove excess glucose from the blood. Birds have the highest blood glucose levels of any vertebrate group (Braun and Swezaea, 2008; Polakof et al., 2011). The lower levels of glucocokinase activity in the liver of birds may contribute to this, with the increased blood glucose levels contributing to the high metabolic rates required for flight. Further studies are needed to understand how birds cope with high blood glucose levels. Carnivores derive most of their energy from protein, thus one might expect to be similar to ruminants and not require glucose uptake by the liver. The cat provides an example of a carnivore that has lost GCKR and thus glucokinase activity (Hiskett et al., 2009; Verbrugghe et al., 2012). However, there are many other carnivores, such as the dog that have glucokinase activity (Tanaka et al., 2005) and have retained both GCKR and glucokinase genes (Wang et al., 2012, 2013). The retention of glucokinase activity in the liver may allow dogs (Batchelor et al., 2011) or carnivorous fish (Panserat et al., 2000b; Polakof et al., 2011) to eat more varied diets or to regulate blood glucose levels when glucose is synthesized from other substrates. Species that have lost GCKR function may be more restricted in their diets.

A number of evolutionary events have occurred that have allowed the development of complex physiological process to efficiently and precisely regulate vertebrate blood glucose levels. This complexity is seen even at the level of the phosphorylation of glucose as it enters

a liver cell. First there was a diversification of the hexokinase gene family that allowed specialization of tissue-specific expression of genes. Of the duplicated hexokinase genes, glucokinase became specialized for expression and function in glucose sensing (e.g., pancreatic islets) and energy storing (i.e., liver) tissues. In the role as a glucose sensor, glucokinase had to evolve such that it had a maximum sensitivity in the normal physiological range of blood glucose levels, thus allowing its activity to change in response to changes in the blood glucose levels. Given the differing roles of glucokinase in glucose sensing, and energy storing tissues, the glucokinase also had to evolve tissue-specific mechanisms to regulate expression, which was achieved for gene expression using tissue-specific alternative promoters and for enzymatic activity through liver-specific interactions with a regulatory protein GCKR. Finally, as the diet of animals change, the necessity of glucokinase activity in the liver became reduced in some species. This reduction in the need for expression and function of glucokinase in the liver, while retaining expression and function in other glucose-sensing tissues, was achieved by the loss of the GCKR interacting partner. Loss of GCKR in the liver results in decreased protein stability, which leads to degradation of glucokinase and loss of enzymatic function without effecting the glucokinase sequence and function in other tissues. Glucose uptake is only a small part of the regulation of glucose metabolism, and clearly equally, or more, complicated evolutionary pathways were followed by other physiological systems, such as hormonal regulation and digestion.

Acknowledgments

This work was supported by a grant CIHR-NSFC China-Canada Joint Health Research from the National Science Foundation of China (Grant number 81061120525) and the Canadian Institutes for Health Research (Grant number CCI-109605). The authors thank two anonymous reviewers for helpful suggestions.

References

- Agius L. Glucokinase and molecular aspects of liver glycogen metabolism. *Biochem J.* 2008; 414:1–18. [PubMed: 18651836]
- Agius L, Peak M. Binding and translocation of glucokinase in hepatocytes. *Biochem Soc Trans.* 1997; 25:145–150. [PubMed: 9056861]
- Andreone TL, Printz RL, Pilkis SJ, Magnuson MA, Granner DK. The amino acid sequence of rat liver glucokinase deduced from cloned cDNA. *J Biol Chem.* 1989; 264:363–369. [PubMed: 2909525]
- Ardehali H, Printz RL, Koch S, Phillips JA 3rd, Granner DK. Isolation, characterization and chromosomal localization of a human pseudogene for hexokinase II. *Gene.* 1995; 164:357–361. [PubMed: 7590357]
- Aschenbach JR, Kristensen NB, Donkin SS, Hammon HM, Penner GB. Gluconeogenesis in dairy cows: the secret of making sweet milk from sour dough. *IUBMB Life.* 2010; 62:869–877. [PubMed: 21171012]
- Augustin R. The protein family of glucose transport facilitators: It's not only about glucose after all. *IUBMB Life.* 2010; 62:315–333. [PubMed: 20209635]
- Bansal P, Wang Q. Insulin as a physiological modulator of glucagon secretion. *Am J Physiol Endocrinol Metab.* 2008; 295:E751–761. [PubMed: 18647881]
- Batchelor DJ, Al-Rammahi M, Moran AW, Brand JG, Li X, Haskins M, German AJ, Shirazi-Beechey SP. Sodium/glucose cotransporter-1, sweet receptor, and disaccharidase expression in the intestine of the domestic dog and cat: two species of different dietary habit. *Am J Physiol Regul Integr Comp Physiol.* 2011; 300:R67–75. [PubMed: 20980625]

- Berradi H, Taouis M, Cassy S, Rideau N. Glucokinase in chicken (*Gallus gallus*). Partial cDNA cloning, immunodetection and activity determination. *Comp Biochem Physiol*. 2005; 141B:129–139.
- Bork P, Sander C, Valencia A. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci*. 1993; 2:31–40. [PubMed: 8382990]
- Braun EJ, Sweazea KL. Glucose regulation in birds. *Comp Biochem Physiol*. 2008; 151B:1–9.
- Cárdenas ML, Cornish-Bowden A, Ureta T. Evolution and regulatory role of the hexokinases. *Biochim Biophys Acta*. 1998; 1401:242–264. [PubMed: 9540816]
- Detheux M, Vandekerckhove J, Van Schaftingen E. Cloning and sequencing of rat liver cDNAs encoding the regulatory protein of glucokinase. *FEBS Lett*. 1993; 321:111–115. [PubMed: 7682971]
- Farrelly D, Brown KS, Tieman A, Ren J, Lira SA, Hagan D, Gregg R, Mookhtiar KA, Hariharan N. Mice mutant for glucokinase regulatory protein exhibit decreased liver glucokinase: a sequestration mechanism in metabolic regulation. *Proc Natl Acad Sci USA*. 1999; 96:14511–14516. [PubMed: 10588736]
- Foster MW, Thompson JW, Forrester MT, Sha Y, McMahon TJ, Bowles DE, Moseley MA, Marshall HE. Proteomic analysis of the NOS2 interactome in human airway epithelial cells. *Nitric Oxide*. 2013 pii: S1089-8603(13)00093-1.
- Fothergill-Gilmore LA, Michels PA. Evolution of glycolysis. *Prog Biophys Mol Biol*. 1993; 59:105–235. [PubMed: 8426905]
- Friedman R, Hughes AL. Pattern and timing of gene duplication in animal genomes. *Genome Res*. 2001; 11:1842–1847. [PubMed: 11691848]
- Furlong RF, Holland PW. Were vertebrates octoploid? *Philos Trans R Soc Lond B*. 2002; 357:531–544. [PubMed: 12028790]
- González C, Ureta T, Sánchez R, Niemeyer H. Multiple molecular forms of ATP:hexose 6-phosphotransferase from rat liver. *Biochem Biophys Res Commun*. 1964; 16:347–352. [PubMed: 5871820]
- González-Alvarez R, Ortega-Cuellar D, Hernández-Mendoza A, Moreno-Arriola E, Villaseñor-Mendoza K, Gálvez-Mariscal A, Pérez-Cruz ME, Morales-Salas I, Velázquez-Arellano A. The hexokinase gene family in the zebrafish: structure, expression, functional and phylogenetic analysis. *Comp Biochem Physiol*. 2009; 152B:189–195.
- Grayson BE, Seeley RJ, Sandoval DA. Wired on sugar: the role of the CNS in the regulation of glucose homeostasis. *Nat Rev Neurosci*. 2013; 14:24–37. [PubMed: 23232606]
- Griffin LD, Gelb BD, Wheeler DA, Davison D, Adams V, McCabe ER. Mammalian hexokinase 1: evolutionary conservation and structure to function analysis. *Genomics*. 1991; 11:1014–1024. [PubMed: 1783373]
- Grimsby J, Coffey JW, Dvorozniak MT, Magram J, Li G, Matschinsky FM, Shiota C, Kaur S, Magnuson MA, Grippo JF. Characterization of glucokinase regulatory protein-deficient mice. *J Biol Chem*. 2000; 275:7826–7831. [PubMed: 10713097]
- Hiskett EK, Suwitheechon OU, Lindbloom-Hawley S, Boyle DL, Schermerhorn T. Lack of glucokinase regulatory protein expression may contribute to low glucokinase activity in feline liver. *Vet Res Commun*. 2009; 33:227–240. [PubMed: 18780155]
- Hokamp K, McLysaght A, Wolfe KH. The 2R hypothesis and the human genome sequence. *J Struct Funct Genomics*. 2003; 3:95–110. [PubMed: 12836689]
- Iynedjian PB. Mammalian glucokinase and its gene. *Biochem J*. 1993; 293:1–13. [PubMed: 8392329]
- Iynedjian PB. Molecular physiology of mammalian glucokinase. *Cell Mol Life Sci*. 2009; 66:27–42. [PubMed: 18726182]
- Irwin DM, Tan H. Molecular evolution of the vertebrate hexokinase gene family: Identification of a conserved fifth vertebrate hexokinase gene. *Comp Biochem Physiol*. 2008; 3D:96–107.
- Jensen TE, Richter EA. Regulation of glucose and glycogen metabolism during and after exercise. *J Physiol*. 2012; 590:1069–1076. [PubMed: 22199166]

- Katzen HM, Soderman DD, Nitowsky HM. Kinetic and electrophoretic evidence for multiple forms of glucose-ATP phosphotransferase activity from human cell cultures and rat liver. *Biochem Biophys Res Commun.* 1965; 19:377–382. [PubMed: 14317406]
- Kawai S, Mukai T, Mori S, Mikami B, Murata K. Hypothesis: structures, evolution, and ancestor of glucose kinases in the hexokinase family. *J Biosci Bioeng.* 2005; 99:320–330. [PubMed: 16233797]
- Klover PJ, Mooney RA. Hepatocytes: critical for glucose homeostasis. *Int J Biochem Cell Biol.* 2004; 36:753–758. [PubMed: 15061128]
- Kogure K, Shinohara Y, Terada H. Evolution of the type II hexokinase gene by duplication and fusion of the glucokinase gene with conservation of its organization. *J Biol Chem.* 1993; 268:8422–8424. [PubMed: 8473284]
- Magnuson MA, Andreone TL, Printz RL, Koch S, Granner DK. Rat glucokinase gene: structure and regulation by insulin. *Proc Natl Acad Sci USA.* 1989; 86:4838–4842. [PubMed: 2662183]
- Mueckler M, Thorens B. The SLC2 (GLUT) family of membrane transporters. *Mol Aspects Med.* 2013; 34:121–138. [PubMed: 23506862]
- Neale BM, Medland S, Ripke S, Anney RJ, Asherson P, Buitelaar J, Franke B, Gill M, Kent L, Holmans P, Middleton F, Thapar A, Lesch KP, Faraone SV, Daly M, Nguyen TT, Schäfer H, Steinhausen HC, Reif A, Renner TJ, Romanos M, Romanos J, Warnke A, Walitza S, Freitag C, Meyer J, Palmason H, Rothenberger A, Hawi Z, Sergeant J, Roeyers H, Mick E, Biederman J. IMAGE II Consortium Group. Case-control genome-wide association study of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry.* 2010; 49:906–920. [PubMed: 20732627]
- Panserat S, Blin C, Médale F, Plagnes-Juan E, Brèque J, Krishnamoorthy J, Kaushik S. Molecular cloning, tissue distribution and sequence analysis of complete glucokinase cDNAs from gilthead seabream (*Sparus aurata*), rainbow trout (*Oncorhynchus mykiss*) and common carp (*Cyprinus carpio*). *Biochim Biophys Acta.* 2000a; 1474:61–69. [PubMed: 10699491]
- Panserat S, Médale F, Blin C, Brèque J, Vachot C, Plagnes-Juan E, Gomes E, Krishnamoorthy R, Kaushik S. Hepatic glucokinase is induced by dietary carbohydrates in rainbow trout, gilthead seabream, and common carp. *Am J Physiol Regul Integr Comp Physiol.* 2000b; 278:R1164–1170. [PubMed: 10801283]
- Panopoulou G, Poustka AJ. Timing and mechanism of ancient vertebrate genome duplications -- the adventure of a hypothesis. *Trends Genet.* 2005; 21:559–567. [PubMed: 16099069]
- Polakof S, Mommsen TP, Soengas JL. Glucosensing and glucose homeostasis: from fish to mammals. *Comp Biochem Physiol.* 2011; 160B:123–149.
- Postic C, Niswender KD, Decaux JF, Parsa R, Shelton KD, Gouhot B, Pettepher CC, Granner DK, Girard J, Magnuson MA. Cloning and characterization of the mouse glucokinase gene locus and identification of distal liver-specific DNase I hypersensitive sites. *Genomics.* 1995; 29:740–750. [PubMed: 8575768]
- Postic C, Shiota M, Magnuson MA. Cell-specific roles of glucokinase in glucose homeostasis. *Recent Prog Horm Res.* 2001; 56:195–217. [PubMed: 11237213]
- Printz RL, Koch S, Potter LR, O'Doherty RM, Tiesinga JJ, Moritz S, Granner DK. Hexokinase II mRNA and gene structure, regulation by insulin, and evolution. *J Biol Chem.* 1993a; 268:5209–5219. [PubMed: 8444897]
- Printz RL, Magnuson MA, Granner DK. Mammalian glucokinase. *Annu Rev Nutr.* 1993b; 13:463–496. [PubMed: 8369155]
- Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000; 16:276–277. [PubMed: 10827456]
- Schwab DA, Wilson JE. Complete amino acid sequence of rat brain hexokinase, deduced from the cloned cDNA, and proposed structure of a mammalian hexokinase. *Proc Natl Acad Sci USA.* 1989; 86:2563–2567. [PubMed: 2704734]
- Schwab DA, Wilson JE. Complete amino acid sequence of the type III isozyme of rat hexokinase, deduced from the cloned cDNA. *Arch Biochem Biophys.* 1991; 285:365–370. [PubMed: 1897938]
- Sharp PA. Speculations on RNA splicing. *Cell.* 1981; 23:643–646. [PubMed: 7226224]

- Sokal RR, Michener CD. A statistical method for evaluating systematic relationships. *Univ Kansas Sci Bull.* 1958; 28:1409–1438.
- Suh SH, Paik IY, Jacobs K. Regulation of blood glucose homeostasis during prolonged exercise. *Mol Cells.* 2007; 23:272–279. [PubMed: 17646701]
- Tanizawa Y, Matsutani A, Chiu KC, Permutt MA. Human glucokinase gene: isolation, structural characterization, and identification of a microsatellite repeat polymorphism. *Mol Endocrinol.* 1992; 6:1070–1081. [PubMed: 1354840]
- Thelen AP, Wilson JE. Complete amino acid sequence of the type II isozyme of rat hexokinase, deduced from the cloned cDNA: comparison with a hexokinase from novikoff ascites tumor. *Arch Biochem Biophys.* 1991; 286:645–651. [PubMed: 1897984]
- Thorens B. Brain glucose sensing and neural regulation of insulin and glucagon secretion. *Diabetes Obes Metab.* 2011; 13(Suppl 1):82–88. [PubMed: 21824260]
- Tsai HJ. Functional organization and evolution of mammalian hexokinases: mutations that caused the loss of catalytic activity in N-terminal halves of type I and type III isozymes. *Arch Biochem Biophys.* 1999; 369:149–156. [PubMed: 10462451]
- Tsai HJ, Wilson JE. Functional organization of mammalian hexokinases: both N- and C-terminal halves of the rat type II isozyme possess catalytic sites. *Arch Biochem Biophys.* 1996; 329:17–23. [PubMed: 8619630]
- Tsai HJ, Wilson JE. Functional organization of mammalian hexokinases: characterization of the rat type III isozyme and its chimeric forms, constructed with the N- and C-terminal halves of the type I and type II isozymes. *Arch Biochem Biophys.* 1997; 338:183–192. [PubMed: 9028870]
- Ureta T. The comparative isozymology of vertebrate hexokinases. *Comp Biochem Physiol.* 1982; 71B: 549–555.
- van Schaftingen E, Veiga-da-Cunha M, Niculescu L. The regulatory protein of glucokinase. *Biochem Soc Trans.* 1997; 25:136–140. [PubMed: 9056859]
- Vandercammen A, Van Schaftingen E. Species and tissue distribution of the regulatory protein of glucokinase. *Biochem J.* 1993; 294:551–556. [PubMed: 8373368]
- Verbrugghe A, Hesta M, Daminet S, Janssens GP. Nutritional modulation of insulin resistance in the true carnivorous cat: a review. *Crit Rev Food Sci Nutr.* 2012; 52:172–182. [PubMed: 22059962]
- Wang Y, Guo T, Zhao S, Li Z, Mao Y, Li H, Wang X, Wang R, Xu W, Song R, Jin L, Li X, Irwin DM, Niu G, Tan H. Expression of the human glucokinase gene: important roles of the 5' flanking and intron 1 sequences. *PLoS One.* 2012; 7:e45824. [PubMed: 23029263]
- Wang ZY, Jin L, Tan H, Irwin DM. Evolution of hepatic glucose Metabolism: Liver-specific glucokinase deficiency explained by parallel loss of the gene for Glucokinase regulatory protein (GCKR). *PLoS One.* 2013; 8:e60896. [PubMed: 23573289]
- Wasserman DH. Four grams of glucose. *Am J Physiol Endocrinol Metab.* 2009; 296:E11–E21. [PubMed: 18840763]
- White TK, Wilson JE. Isolation and characterization of the discrete N- and C-terminal halves of rat brain hexokinase: retention of full catalytic activity in the isolated C-terminal half. *Arch Biochem Biophys.* 1989; 274:375–393. [PubMed: 2802617]
- Wilson JE. Hexokinases. *Rev Physiol Biochem Pharmacol.* 1995; 126:65–198. [PubMed: 7886381]
- Wilson JE. An introduction to the isoenzymes of mammalian hexokinase types I-III. *Biochem Soc Trans.* 1997; 25:103–107. [PubMed: 9056852]
- Wilson JE. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *J Exp Biol.* 2003; 206:2049–2057. [PubMed: 12756287]
- Wilson, JE. The hexokinase gene family. In: Matschinsky, FM., Magnuson, MA., editors. *Glucokinase and Glycemic Disease: From Basic to Novel Therapeutics.* Vol. 16. Basel: Karger AG; 2004. p. 18-30. *Frontiers in Diabetes*
- Wilson-O'Brien AL, Patron N, Rogers S. Evolutionary ancestry and novel functions of the mammalian glucose transporter (GLUT) family. *BMC Evol Biol.* 2010; 10:152. [PubMed: 20487568]
- Wright EM. Glucose transport families SLC5 and SLC50. *Mol Aspects Med.* 2013; 34:183–196. [PubMed: 23506865]

Wright EM, Loo DD, Hirayama BA. Biology of human sodium glucose transporters. *Physiol Rev.* 2011; 91:733–794. [PubMed: 21527736]

Fig. 1.

Similarities between hexokinase protein sequences. Hexokinase proteins sequences were compared in a pairwise manner by dotplots (dotmatcher, Rice et al., 2000), where a dot indicates similarity (window size = 10, threshold = 23). In these dotplots a 10 amino acid long sequence of one sequence is compared to 10 amino acid segments of a second sequence and dot is generated if the similarity is greater than the threshold. This process is repeated so that every 10 amino acid long sequence in one protein is compared to every 10 amino acid long sequence in the second sequence. If the two sequences are homologous and the similarity is greater than the threshold then a diagonals line is generated. If two (or more) diagonal lines are generated, this would indicates that one sequence in one protein is similar to two (or more) sequences in the second sequence. Comparisons are: **A.** Hexokinase I and Glucokinase; **B.** Glucokinase with itself; **C.** Hexokinase I with itself; **D.** Hexokinase I with Hexokinase II. Similar plots to **A** were found if Hexokinase I was replaced with Hexokinase II, Hexokinase III, or HKDC1. Comparisons of Hexokinase I, Hexokinase II, Hexokinase III, or HKDC1 yielded plots similar to **D**.

Fig. 2. Structure of vertebrate hexokinase genes. The exon structure of coding regions of hexokinase genes from vertebrates (HK1, HK2, HK3, GCK, and HKDC1) as well as a gene from *Ciona savignyi* (ENSEMBL ID: ENSCSAVG00000011413) are illustrated. Coding sequence exons are shown as solid boxes that are proportional to their coding lengths (untranslated exon are not shown). Introns are thin lines and are not proportional to length. The phase of the intron (position of the codon interrupted by introns (Sharp, 1981)) is shown above each intron.

Fig. 3.

Evolution of the kinase domains in vertebrate hexokinase genes. **A.** Illustration of two models proposed for the evolution of vertebrate hexokinases. Open boxes represent kinase domains, with solid diamonds representing duplication (domain or gene) events. The + symbol represents a domain deletion event. The ancestral sequence is shown at the bottom as a single hexokinase domain gene. The scheme on the left indicates that the duplication of the hexokinase domain occurred before radiation of the hexokinase genes, with a subsequent loss of a hexokinase domain in GCK. The scheme on the right indicates that GCK diverged first, before the duplication of the hexokinase domain (Adapted from Ureta, 1982). **B.** Phylogeny of Hexokinase domains in hexokinase genes. The phylogeny has been rooted with several outgroups that yield the same phylogeny. The solid diamond indicated the hexokinase domain duplication event. (Adapted from Irwin and Tan, 2008).

Fig. 4.

Evolution of hexokinase gene structures. **A.** The ancestral hexokinase gene had a single hexokinase domain and an intron-exon structure similar to glucokinase with 10 exons (see Fig. 2). Duplication and recombination between exons 2 and 10 of the duplicated genes yields a gene with 18 exons, with two copies of each of exon 3 through 9, single copies of exons 1 and 2 and a hybrid exon containing sequences from a duplicated copies of exons 2 and 10, and single copy of exon 1. The duplicated exons diverge in sequence, and subsequently the gene duplicates to become the ancestor of the 100kD HK1, HK2, KH3, and HKDC genes. **B.** The glucokinase gene is also derived form the duplicated ancestor. After duplication of the exons, mutational events lead to the loss of exons 3a though to the hybrid exon 10a/2b that code for most of the N-terminal half of the protein. No trace of these exons can be found in extant genes.