Molecular cloning and sequence analysis of *sglt1* and tertiary structure prediction of deduced protein in *Cyprinus carpio* L.

Guoxing Nie*, Caixia Hou¶, Junli Wang, Jianxin Zhang, Dongying Song, Bei Wang, Xuejun Li, Xianghui Kong

College of Life Science, Henan Normal University, Xinxiang 453007, Henan, P R .China
niegx@htu.cn

Abstract: Na^+/glucose cotransporter (Sglt1) plays an important role in transporting Na^+ and glucose and maintaining the adjustment of metabolism. The aim to study *sglt1* is to further understand the regulation mechanism of *sglt1* gene in fish. In this study, the full-length cDNA of Na^+/glucose cotransporter gene was cloned in intestine of *Cyprinus carpio* L. using RT-PCR and RACE methods, which included 2856 bp involved in 113 bp 5'-untranslated region, 766 bp 3'-untranslated region, and 1977 bp open reading frame (ORF) which encoded 658 amino acids. The predicted amino acid sequence was the highest similar with that of *Danio rerio* (90.70%), and the lowest similar with that of rabbit (71.40%). Fourteen transmembrane domains were predicted in the 3-D protein model using comparative protein modeling program SWISS-MODEL. The structural core was comparative of 5 TM helices (TM2-TM6 and TM7-TM11) with the inverted repeat. It was demonstrated that Glucose might be bounded in the center of the structural core, and a possible Na^+-binding site was located at the intersection of TM2 and TM9. Thereby, the functional roles and regulation mechanism of Sglt would provide unique opportunities to investigate the biochemical processes in intestine of *Cyprinus carpio* L., and lay the foundation for artificial culture of the species involved.

[Guoxing Nie, Caixia Hou, Junli Wang, Jianxin Zhang, Dongying Song, Bei Wang, Xuejun Li, Xianghui Kong. **Molecular cloning and sequence analysis of *sglt1* and tertiary structure prediction of deduced protein in *Cyprinus carpio* L.** Life Science Journal. 2011;8(2):204-212] (ISSN:1097-8135). http://www.lifesciencesite.com.

Keywords: *sglt1*; cDNA sequence analysis; Protein tertiary structure; *Cyprinus carpio* L.

1. Introduction

Na^+/substrate cotransport (or symport) is a widespread mechanism of solute transportation across cytoplasmic membranes of prokaryotic and eukaryotic cells, which is mainly performed by cotransport proteins. Based on sequences similarities, the members of these proteins are classified into different families (Reizer *et al.*, 1994). Proteins of these families generally utilize electrochemical sodium gradient to drive transportation of substance in reverse concentration like sugars, amino acids, vitamins, ions, myoinositol, phenyl acetate, urea and water (Reizer *et al.*, 1994; Frank *et al.*, 1998). Among them, human sodium-glucose transporter (hSglt1) is one of the best characterized members of the Solute Sodium Symporters Families (SSSF). In addition, the sodium-iodide transporter (NIS) and the sodium-proline transporter (PutP) in *E. coli* and sodium-galactose symporter (*vsglt*) in *Vibrio parahaemolyticus* also were described in publishments (Frank *et al.*, 1998; Schwan *et al.*, 1998; Zeuthen *et al.*, 2001; Hirayama *et al.*, 1997). Sglt1 and NIS catalyze the uptake of substance with a 2:1 sodium-substrate stoichiometry while the value of sodium-proline transported by PutP is 1:1 (Zeuthen *et al.*, 2001; Hirayama *et al.*, 1997; Eskandari *et al.*, 1997). In addition, it was also reported that Sglt1 could couple the uptake of two sodium and one sugar with the transportation of 264 water molecules (Wegener *et al.*, 2000).

The high-affinity Na^+/glucose contransporter (Sglt1) (Martin *et al.*, 1996) is an important member of the sodium: solute symporter family (SSSF) (Ernest *et al.*, 2004) with more than 700 different sequences (Wright *et al.*, 2004; Turk *et al.*, 1997). Sglt1, belonging to the homologous family 5 (SLC5), is abundantly expressed in small intestine(Balen *et al.*, 2008; Hirsh *et al.*, 1998) , while at a lower level in kidney (Ikeda *et al.*, 1989). The major function of Sglt1 is to accumulate sugar in intestinal or kidney epithelial cells adverse concentration gradient. Since the driving force of Na^+-coupled transporters is provided by the Na^+ electrochemical potential gradient across the plasma membrane. It serves as the principal uptake pathway for glucose derived from diet. Mutations of *sglt1* could result in the dysfunctional of this pathway, which would affect intestinal glucose/galactose absorption (Martin *et al.*, 1996). Recently, Sglt1 has been studied as a target protein for diabetes treatment (Ikumi *et al.*, 2008; Sabino-Silva *et al.*, 2010). Ryuichi *et al.* findings indicate that Sglt1 serves as the intestinal glucose sensor for glucose-induced incretin secretion and that a noncalorigenic Sglt1 substrate ameliorates hyperglycemia by stimulating incretin secretion (Ryuichi *et al.*, 2009).

The crystal structure of the *Vibrio parahaemolyticus* sodium/galactose symporter (vSglt) had been reported (Faham *et al.*, 2008). So far, few studies on Na^+/glucose cotransporter were reported in the freshwater fishes. In this study, *sglt1* gene with the high affinity was first cloned in intestine of *Cyprinus carpio* L. and subsequently used to obtain new insights into the molecular mechanism of Na^+/glucose cotransporter. The secondary structure and tertiary structure of Sglt1 protein in *C. carpio* have been predicted with several computational algorithms.

2. Material and Methods

Fish acclimation

In this study, *C. carpios* with body weight of (12.6±0.38) g were used as experimental animals, which were acclimated in a 200-L tank filled with dechlorinated water with constant aeration (DO: 6.2±0.2 mg /L) and a 12/12 h light/dark photoperiod. Water temperature was controlled at (26.8±0.68)℃. During the period of acclimation, the fishes were fed for four times each day (8:30am, 11:30am, 14:30pm and 17:30pm) with commercial pellet feed. After the acclimation, ten fishes were randomly arrested to be used as experimental fishes. Subsequently, the intestines of which were obtained by fish dissection respectively after general anesthesia, and scissored immediately into intestine, and the contents in guts were cleared rapidly. All the operations were conducted under aseptic condition on the ice.

Total RNA extraction and 5' and 3' RACE

The prepared guts were quickly frozen in liquid nitrogen and used to extract the total RNA with TRIzol reagent (purchased from Invitrogen) respectively. The total RNA (5mg) was used to synthesize the first-stand cDNA using AMV reverse transcriptase (from Shanghai Sangon) and oligo-p $(dT)_{18}$ Primer (from Shanghai Sangon) in a 20μL reaction, according to the manufacturer's instruction. The *sglt1* cDNAs were then amplified by PCR in a total volume of 50μL, containing 10mM Tris-HCl(pH9.0), 50mM KCl, 1.25mM $MgCl_2$, 0.2mM dNTPs, 1 units of Taq polymerase (from Takara, Japan), 40pmol primer for each one and 5μL template cDNA (Table 1). A initial reaction of 3 min at 94℃ was followed by 35 cycles (denaturation at 95℃ for 35 s, annealing at specific temperature for 35 s and extension at 72℃ for 2 min), and a final extension for 10 min at 72℃. The annealing temperatures of PCR reaction were depended on the *sglt1* to be amplified (Table I). The PCR product was resolved on 1% agarose gels via electrophoresis. Photographs of the gels stained with ethidium bromide are show in an inverted black/white format.

Table 1. Sequences of oligonucleotide primers used for PCR and rapid amplification of cDNA ends (RACE)

Names	Oligonucleotide sequence (5'→3')	Length (bp)
3' GSP1 *sglt1*	GGTGGATTTGAATGGAATGCTCT	23
3' GSP2 *sglt1*	ACCTCTCCGTGCTCTCCCTGTTT	23
3' RACE outer	TACCGTCGTTCCACTAGTGATTT	23
3' RACE inner	CGCGGATCCTCCACTAGTGATTTCACTATAGG	32
5' GSP1 *sglt1*	ATCCCACGACCATGATGATTGTC	23
5' GSP2 *sglt1*	CTGTGTGGCAGCGTTTGGAGGAG	23
5' RACE outer	CATGGCTACATGCTGACAGCCTA	23
5' RACE inner	CGCGGATCCACAGCCTACTGATGATCAGTCGATG	34

Note: The primers were based on the *sglt1* sequences of Zebrafish and other animals deposited in GenBank.

Cloning and sequencing of *C. Carpios* intestion cDNAs

The amplified bands corresponding to *sglt1* cDNAs were accurately excised from the 1% agarose gel and purified using the Gel extraction kit (from Takara, Japan) respectively. The purified *sglt1* cDNAs were ligated into the pGEM-T Easy vectors (from Promega, USA), and the resultant recombinant plasmids were transferred into competent *Escherichia coli* strain JM109. For each cDNA, 4-6 plasmid clones containing *sglt1* cDNAs were sequenced by ABI3730 using M13+/-universal primers (from Takara, Japan).

3. Results

Isolation of the *C. Carpios sglt1* cDNA by RACE

The primer was originally designed from highly conserved regions of *sglt1* based on the sequence alignment of zebrafish, spiny dogfish, mouse, rat, human, bat, horse, bovine and rabbit *sglt1* cDNA from GenBank. Employing the RACE strategy,

the full-length *sglt1* of *C. Carpio* were cloned. The 5'-RACE and 3'-RACE results were sequenced and spliced to obtain the full-length cDNA (Figure 1). The complete coding sequence of the *C. carpio sglt1* cDNA with 2856 nucleotides comprised coding sequence region with 1974-bp open reading frame (ORF), a 113-bp 5'-untranslated region and a 766-bp 3'-untranslated region including poly (A), encodes a putative protein of 658 amino acids (Figure 2).

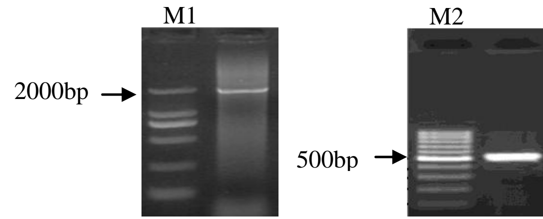

Figure 1. The results of RACE-PCR on *sglt1*. A, the result of 5′-RACE, the acquired gene was about 2000bp fragment; B, the result of 3′-RACE, the acquired was about 500bp fragment; M1, 2000bp DNA ladder; M2, 100bp DNA ladder.

```
-113                                                                                         gagaacagtcagtc
      acacacctgcaggtgagcgtctgatcc aggtatatagaccgtcgtgtcagagagggaaacagtgtgctgcccaccggcccctttgaaaccatcaaa
   1  ATG GGT GAA GAA TAT TTT GGA TTC TCC TGG GTT CGC AAT GAA AAC AGA AAA AAC GTC ACC ATA TAT GTT AAC AAT
   1   M   G   E   E   Y   F   G   F   S   W   V   R   N   E   N   R   K   N   V   T   I   Y   V   N   N
  76  CCA GCG GAC ATC TCA GTG ATA GTT ATA TAT TTT CTG GTG GTT CTG GCT GTG GGA ATA TGG GCA ATG GTA CGG ACC
  26   P   A   D   I   S   V   I   V   I   Y   F   L   V   V   L   A   V   G   I   W   A   M   V   R   T
 151  AAT CGA GCC ACA GTA GGA GGC TTT TTC CTG GCT GGC AGA AGT ATG GTG TGG TGG CCG ATT GGA GCC TCT CTT TTT
  51   N   R   A   T   V   G   G   F   F   L   A   G   R   S   M   V   W   W   P   I   G   A   S   L   F
 181  GCT AGT AAC ATC GGA AGT GGG CAT TTT GTG GGA ATA GCA GGA ACT GGA GCA GCT GCT GGT ATT GCT ATC GGA GGA
  76   A   S   N   I   G   S   G   H   F   V   G   I   A   G   T   G   A   A   A   G   I   A   I   G   G
 301  TTT GAG TGG AAT GCT CTT GTT GTA GTG ATT ATT CTG GGA TGG CTC TTC GTG CCC ATC TAC ATT AAG GCT GGG ATC
 101   F   E   W   N   A   L   V   V   V   I   I   L   G   W   L   F   V   P   I   Y   I   K   A   G   I
 376  GTG ACC ATG CCA GAG TAC CTG AAG AAA CGA TTT GGT GGG CAG CGG ATC CGT ATC TAC CTC TCC GTG CTC TCC CTG
 126   V   T   M   P   E   Y   L   K   K   R   F   G   G   Q   R   I   R   I   Y   L   S   V   L   S   L
 451  TTT CTC TAT GTG CTC ACA AAA ATC TCT GCG GAC ATG TTT GCA GGA GCC ATT TTC ATC AAC CAG GCT CTG GGA CTA
 151   F   L   Y   V   L   T   K   I   S   A   D   M   F   A   G   A   I   F   I   N   Q   A   L   G   L
 526  AAC ATT TAC CTG GCC GTA ATT ATT CTG CTG CTC ATT ACA GCT CTT TAT ACA GTA ACA GGT GGT CTA GCT GCA GTC
 176   N   I   Y   L   A   V   I   I   L   L   L   I   T   A   L   Y   T   V   T   G   G   L   A   A   V
 601  ATC TAC ACC GAC ACA CTC CAG ACC ATC ATC ATG GTT GTG GGT TCA TTC ATT CTT ATG GGC TTT GCG TTC AAT GAG
 201   I   Y   T   D   T   L   Q   T   I   I   M   V   V   G   S   F   I   L   M   G   F   A   F   N   E
 676  GTG GGA GGC TAT GAG AAC TTC AAA GAC CGA TAC ATG ACT GCG ATC CCG TCA GTG GTG GGT GTG AAC ATC AGT GAA
 226   V   G   G   Y   E   N   F   K   D   R   Y   M   T   A   I   P   S   V   V   G   V   N   I   S   E
 751  GAG TGC TAC ACT CCT CGT GCA GAC TCC TTC CAC ATC TTC AGA GAC CCC CTC AGT GGC GAT CTG CCC TGG CCT GGA
 251   E   C   Y   T   P   R   A   D   S   F   H   I   F   R   D   P   L   S   G   D   L   P   W   P   G
 826  CTG ATC TTT GGT CTT ACT ATC CAG GCT GGC TGG TAC TGG TGC ACT GAC CAG GTG ATT GTG CAG CGC TGT CTG TCT
 276   L   I   F   G   L   T   I   Q   A   G   W   Y   W   C   T   D   Q   V   I   V   Q   R   C   L   S
 901  GCC AAG AGC CTG TCT CAT GTG AAA GCA GGC TGC ATC CTG TGT GGT TAC CTC AAA CTT CTG CCC ATG TTC CTC ATG
 301   A   K   S   L   S   H   V   K   A   G   C   I   L   C   G   Y   L   K   L   L   P   M   F   L   M
 976  GTT TTC CCT GGC ATG ATC AGC AGA GTT CTT TAT CCA GAT GAG ATT GCA TGT GTG GAT CCA AAA GAG TGT GAC TAC
 326   V   F   P   G   M   I   S   R   V   L   Y   P   D   E   I   A   C   V   D   P   K   E   C   D   Y
1051  TAC TGT GGA GCC AGT GTG GGC TGC AGT AAT ATT GCT TAT CCG AAA CTA GTG GTG GAT CTG ATG CCA AAC GGT CTC
 351   Y   C   G   A   S   V   G   C   S   N   I   A   Y   P   K   L   V   V   D   L   M   P   N   G   L
1126  AGA GGG TTG ATG TTG TCC GTC ATG CTG GCC TCT CTG ATG AGT TCA CTC ACT TCC ATC TTT AAC AGC GCT AGC ACA
 376   R   G   L   M   L   S   V   M   L   A   S   L   M   S   S   L   T   S   I   F   N   S   A   S   T
1201  CTC TTC ACT ATG GAC ATT TAC ACC AAG ATT CGC TCC AAT GCC AAG GAA AAG GAA CTC ATG ATT GCC GGG AGG GTG
 401   L   F   T   M   D   I   Y   T   K   I   R   S   N   A   K   E   K   E   L   M   I   A   G   R   V
1276  TTC ATG CTG TTT CTG ATC GGC GTG AGT ATC GCG TGG ATC CCC ATC GTT CAG TCG GCT CAG AGC GGG CAG CTC TTC
 426   F   M   L   F   L   I   G   V   S   I   A   W   I   P   I   V   Q   S   A   Q   S   G   Q   L   F
1351  GAC TAC ATT CAG TCC ATC ACC AGT TAT CTG GCT CCG CCC ATC GCT GCT GTC TTC ACC CTC GCC ATT TTC TGC AAG
 451   D   Y   I   Q   S   I   T   S   Y   L   A   P   P   I   A   A   V   F   T   L   A   I   F   C   K
1426  CGA GTC AAT GAA CCA GGT GCT TTC TAC GGG ATG TGT ATT GGT CTG TTG GTG GGT CTG GCA CGT ATG ATA ACT GAG
 476   R   V   N   E   P   G   A   F   Y   G   M   C   I   G   L   L   V   G   L   A   R   M   I   T   E
1501  TTT GCC TAC GGC ACG GGC AGC TGC GTG AGT CCC AGT AAC TGT CCT ACG ATC ATC TGT GGA GTG CAC TAC CTC TAT
 501   F   A   Y   G   T   G   S   C   V   S   P   S   N   C   P   T   I   I   C   G   V   H   Y   L   Y
1576  TTC GCC CTC ATC CTC TAC AGC CTG TCC TGT ATA TTG ATA CCG TGC ATC TCC CTC ATG ACT AAA CCC ATT GAC GAC
 526   F   A   L   I   L   Y   S   L   S   C   I   L   I   P   C   I   S   L   M   T   K   P   I   D   D
1651  AAA CAT CTG TAC AGG CTC TGC TGG AGT TTG AGG AAC AGC ACT GAG GAG AGG ATT GAT CTG GAA AAA GAT GAC TGG
 551   K   H   L   Y   R   L   C   W   S   L   R   N   S   T   E   E   R   I   D   L   E   K   D   D   W
1726  ACT GAA AAT CAG GAT TCA GAT TCT GTG CAA ACA GAA GAG GTG CGC AAG GAT CCA AGC TGT TGG AAG AAG GCC TAT
 576   T   E   N   Q   D   S   D   S   V   Q   T   E   E   V   R   K   D   P   S   C   W   K   K   A   Y
1801  AAC TGG TTC TGC GGT TTT GAT CAA GGC AAT GCA CCT AAA CTG ACT AAA GAA GAG GAG GCA GAG TTG AAT TTG AAG
 601   N   W   F   C   G   F   D   Q   G   N   A   P   K   L   T   K   E   E   E   A   E   L   N   L   K
1876  CTC ACA GAC ACC ACT GAG AAA CCG CTA TGG AGA AAC GTG GTC AAC GCT AAT GCT ATT ATC CTC CTC ACC GTC TGT
 626   L   T   D   T   T   E   K   P   L   W   R   N   V   V   N   A   N   A   I   I   L   L   T   V   C
1951  GTC TTC TTC CAT GGT TTC TTT GGC TAA
 651   V   F   F   H   G   F   F   G   *
      Aaatcatcagtatcaactaactctgttaaggatattagagctgtattctgctgacagaaatagtatcttatcagtaacatgctaaacataattaatgtgattgtaag
      aagttacgaaagttcctcttgaaacacttcacacttccttataatatcctccacaaaaacgtccttactgctgcctactgttttgatttgatggcaggaagtagga
      aaaaaaaaatattttgactttaactttaaacccaagactaaatctcataattatttccttgtttttcttccttgtagtcatgtcaaatctttcttagcgtgctagtatctaat
      ttaacacatgcctgactgttttactgtacatttcacaaagtttaccgcacgttcttcctttatactttaacaaattacagagtaatagtttcaactaaaattatgaagt
      ttgagtcagctttgttgtttgttccctgtaaagtcctctgatgcattatgatcaaatcaatgcaaatacttttctgactgttttaacatgcctctcatctaccaaactc
      caaaaacatttttatatttggtcttactatttggtacgttctgattatcttgtaagtaaaaatatatatttttgtctaggacaagcaaagatattttatttaaaaacattc
      atattttgccatggtatatgtattgtactctgtaatatattaatttaatgtttctgtctatatttaataaagattaatgaatttagtaaaaaaaaaaaa
```

Figure 2. Nucleotide and deduced amino acid sequence of Na^+/glucose cotransporter in intestine of *Cyprinus carpio* L. The sequence contains a single open reading frame which encodes a protein with 658 amino acids. The complete 5'-untranslated region was 113 nucleotides.

Sequence analysis of *C. Carpios sglt1* gene

The deduced amino acid of carp Sglt1 using EXPASY is composed of 658 amino acids with a molecular weight of approximately 72.9 kDa and the isoelectric point of 6.35. The secondary structure of the deduced Sglt1 amino acid sequence was analyzed to seek potential transmembrane regions using TMHMM Server v. 2.0 (DTU) (Figure 3). Fourteen transmembrane domains were also putatived in this study using SOPMA (Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments) (Figure 4). The *sglt1* gene contains 47.26% of α-helix, 17.17% of extended strand, 2.74% of β-turn and 32.83% of random coil.

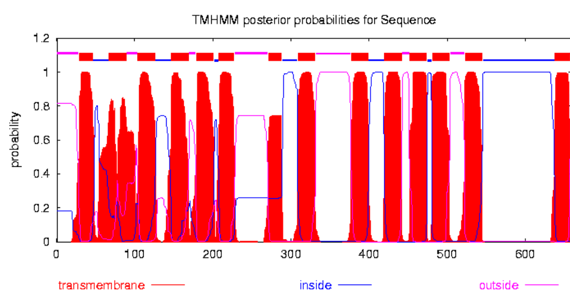

Figure 3. Secondary structure model and 14 transmembrane domains of Sglt1 Predicted by TMHMM Server v. 2.0. The C-terminal and the N-terminal of Sglt1 were out of the cytoplasm

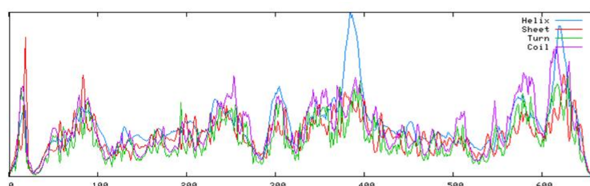

Figure 4. SOPMA result for *sglt1* from *Cyprinus carpio* L. intestine. The *sglt1* gene contains 47.26% of α-helix, 17.17% of extended strand, 2.74% of β-turn and 32.83% of random coil.

SignalP 3.0 Sercer analysis predicted a signal peptide of carp Sglt1 positioned in the amino-terminal (N-terminal) sequence (MGEEYFGFSWVRNENRKNV TIYVNNPADISVIVIYFLVVLAVGIWA) (Figure 5).

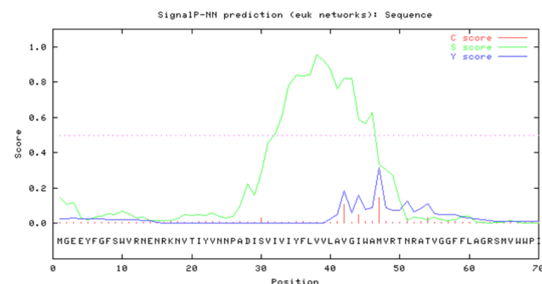

Figure 5. SignalP 3.0 Sercer analysis predicted a signal peptide of carp Sglt1 positioned in the amino-terminal (N-terminal) sequence (MGEEYFGFSWVRNENRKNVTIYVNNPADISVI VIYFLVVLAVGIWA).

Homology and phylo genetic analysis of *sglt1* genes

The *sglt1* cDNA sequence achieved in this study has been submitted to GenBank and assigned the accession No EU328389.1. The reference sequences in the analysis were downloaded from GenBank. The deduced amino acid sequence of *C. carpio sglt1* was 72.3%, 72.1%, 71.4%, 73.7%, 74.3%, 75.0%, 74.0%, 75.2% and 90.7% identical to Human (*Homo sapiens*), Horse (*Equus caballus*), Rabbit (*Oryctolagus cuniculus*), Bovine (*Bos taurus*), Bat (*Rhinolophus ferrumequinum*), Rat (*Rattus norvegicus*), Mouse(*Mus musculus*), Spiny dogfish (*Squalus acanthias*), Zebrafish (*Danio rerio*) respectively, as shown in Table 2. The homology and divergence among the sequences were calculated using the Laser-gene analysis software package (DNAMAN, USA) (Figure 6). The deduced amino acid sequence of Sglt1 in *C. carpio* was the lowest similarity with that in rabbit (71.4%) and the highest similarity with that in *Danio rerio* (90.70%)

4. Discussions

The Na^+/glucose cotransporter (Sglt) are members of the expanded solute carriers SLC5A family and are predominantly expressed in the brush-border membranes of small intestine and proximal convoluted tubule of the kidney (Zhao *et al.*, 2005). Sglt1 as a member of the Na^+-glucose cotransporter family, play an important role in transporting sodium and glucose to maintain the basic physiological metabolism and nutrition requirement (Zhao *et al.*, 2005; Zhou *et al.*, 2003). Sglt1 moves 2 Na^+ ions with each glucose per cycle (Díez-Sampedro and Barcelona, 2010； Sabino-Silva *et al.*, 2010). In present study, *sglt1* was first cloned and characterized in *C. carpio*, and subsequently used to obtain new insights into the molecular mechanism of Na^+-glucose cotransporter.

Table 2. The similarity dataset of Sglt1 amino acid sequences between *C. carpio* and other species

	Car	Zeb	Spi	Mou	Rat	Bat	Bov	Hor	Hum	Rab
Car	100%									
Zeb	90.7%	100%								
Spi	75.2%	75.0%	100%							
Mou	74.0%	75.1%	72.6%	100%						
Rat	75.0%	76.0%	73.5%	95.8%	100%					
Bat	74.3%	74.0%	72.1%	89.9%	90.0%	100%				
Bov	73.7%	74.3%	72.8%	88.0%	86.9%	89.3%	100%			
Hor	72.1%	71.8%	71.5%	86.9%	86.5%	89.4%	86.6%	100%		
Hum	72.3%	72.8%	71.0%	88.0%	87.7%	88.1%	85.8%	87.2%	100%	
Rab	71.4%	72.1%	71.5%	86.0%	86.3%	86.6%	85.0%	84.6%	84.6%	100%

Note: The similarity dataset of Sglt1 amino acid sequences were from *Cyprinus carpio* L. (Car), Zebrafish (Zeb), Spiny dogfish (Spi), Human (Hum), Horse (Hor), Mouse (Mou), Rabbit (Rab), Bovine (Bov), Bat (Bat) and Rat (Rat).

```
Car   MGEEYFGFSWVRNENRKN.VTIYVNNPADISVIVIYFLVVLAVGIWAMVRTNRATVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Zeb   MGADYFGFSLLRNDVKRN.TTIYVNNPADISVIVIYFLVVLAVGVWAMVRTNRATVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Spi   MESSTVPAGSATTPAGGSPGSSAVNNPADISVIVIYFLVVLAVGIWAMVRTNRGTVGGFFLAGRSMVWWPIGASLFASNIGSGH   84
Mou   MDSSTLSPAVTATDAPIP.SYERIRNAADISVIVIYFVVVMAVGLWAMFSTNRGTVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Rat   MDSSTLSPAVTATDAPIQ.SYERIRNAADISVIVIYFVVVMAVGLWAMFSTNRGTVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Hum   MDSSTWSPKTTAVTRPVE.THELIRNAADISIIVIYFVVVMAVGLWAMFSTNRGTVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Bat   MDSSTWSPSTPVTAAPLA.SHERIRNAADISVIVIYFVVVMAVGLWAMFSTNRGTVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Hor   MDSSTWSPGTATTAAPLA.SYERIRNAADISVIVIYFVVVMAVGLWAMFSTNRGTVGGFFLAGRNMVWWPIGASLFASNIGSGH   83
Bov   MGADYFGFSLLRNDVKRN.TTIYVNNPADISVIVIYFLVVLAVGVWAMVRTNRATVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Rab   MDSSTLSPLTTSTAAPLE.SYERIRNAADISVIVIYFLVVMAVGLWAMFSTNRGTVGGFFLAGRSMVWWPIGASLFASNIGSGH   83
Consenmus                      n adis iviyf vv avg wam  tnr tvggfflagr mvwwpigaslfasnigsgh

Car   FVGIAGTGAAAGIAIGGFEWNALVVVIILGWLFVPIYIKAGIVTMPEYLKKRFGGQRIRIYLSVLSLFLYVLTKISADMFAGAI   167
Zeb   FVGIAGTGAAAGIAIGGFEWNALVVVIILGWLFVPIYIKAGVVTMPEYLKKRFGGQRIRIYLSVLSLCLYVFTKISADMFAGAI   167
Spi   FVGLAGNGAAGGIATGGFEWNALVIVIILGWLFVPIYIRAGVVTMPEYLKRRFGGNRIRIYLSVLSLCLYIFTKISADMFSGAI   168
Mou   FVGLAGTGAAAGIAMGGFEWNALVLVVVLGWIFVPIYIKAGVVTMPEYLRKRFGGKRIQIYLSVLSLLLYIFTKISADIFSGAI   167
Rat   FVGLAGTGAAAGIAMGGFEWNALVFVVVLGWLFVPIYIKAGVVTMPEYLRKRFGGKRIQIYLSVLSLLLYIFTKISADIFSGAI   167
Hum   FVGLAGTGAASGIAIGGFEWNALVLVVVLGWLFVPIYIKAGVVTMPEYLRKRFGGQRIQVYLSLLSLLLYIFTKISADIFSGAI   167
Bat   FVGLAGTGAAAGIATGGFEWNALILVVVLGWVFVPIYIKAGVVTMPEYLRKRFGGKRIQIYLSILSLLLYIFTKISADIFSGAI   167
Hor   FVGLAGSGAASGIAMGGFEWNALILVVVLGWVFVPIYIKAGVVTMPEYLKKRFGGQRIQVYLSLLSLLLYIFTKISADIFSGAI   167
Bov   FVGIAGTGAAAGIAIGGFEWNALVVVIILGWLFVPIYIKAGVVTMPEYLKKRFGGQRIRIYLSVLSLCLYVFTKISADMFAGAI   167
Rab   FVGLAGTGAASGIATGGFEWNALIMVVVLGWVFVPIYIRAGVVTMPEYLQKRFGGKRIQIYLSILSLLLYIFTKISADIFSGAI   167
Consenfvg ag gaa gia ggfewnal  v  lgw fvpiyi ag vtmpeyl  rfgg ri  yls lsl ly  tkisad f gai

Car   FINQALGLNIYLAVIILLLITALYTVTGGLAAVIYTDTLQTIIMVVGSFILMGFAFNEVGGYENFKDRYMTAIPSVVG...VNI   248
Zeb   FINQALGLNIYLAVVILLLITALYTVTGGLAAVIYTDTLQTIIMVVGSFILMGFAFNKVGGYENFRDSYMTAIPSVIG...TNI   248
Spi   FIREALGLNLYVAVVILLSITALYTVTGGLAAVIYTDTLQTIIMVVGSFILMGFAFNEVGGYDAMKEKYMNAITSQTTDPILNI   252
Mou   FINLALGLDIYLAIFILLAITALYTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFNEVGGYEAFMDKYMKAIPTKVSNGNFTA   251
Rat   FINLALGLDIYLAIFILLAITALYTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFREVGGYEAFMDKYMKAIPTLVSDGNITV   251
Hum   FINLALGLNLYLAIFLLLAITALYTITGGLAAVIYTDTLQTVIMLVGSLILTGFAFHEVGGYDAFMEKYMKAIPTIVSDGNTTF   251
Bat   FINLALGLDLYLAIFLLLAITGIYTITGGLAAVIYTDTLQTAIMLVGSFILTGFAFHEVGGYEAFMAKYMKAIPTVVSDGNITI   251
Hor   FINMALGLDLYLAIILLLAVTGIYTITGGLAAVIYTDTLQTAIMLVGSAILTGFAFNEVGGYEAFMEKYMKAIPTMISDGNITI   251
Bov   FINQALGLNIYLAVVILLLITALYTVTGGLAAVIYTDTLQTIIMVVGSFILMGFAFNKVGGYENFRDSYMTAIPSVIG...TNI   248
Rab   FIQLTLGLDIYVAIIILLVITGLYTITGGLAAVIYTDTLQTAIMMVGSVILTGFAFHEVGGYEAFTEKYMRAIPSQISYGNTSI   251
Consenfis  lgl  y a   ll  t  yt tgglaaviytdtlqt im vgs il gfaf  vggy      ym ai

Car   SEECYTPRADSFHIFRDPLSGDLPWPGLIFGLTIQAGWYWCTDQVIVQRCLSAKSLSHVKAGCILCGYLKLLPMFLMVFPGMIS   332
Zeb   SESCYTPRADSFHIFRDPINGDLPWPGLLFGLTIQAGWYWCTDQVIVQRCLSAKNLSHVKAGCILCGYLKLLPMFLMVFPGMIS   332
Spi   SAKCYTPRPDSFHLFRNPLTADLPWPGLIFGLSILAIWYWCTDQVIVQRCLSAKNMSHVKAGCILCGYLKLLPMYLMVFPGMVS   336
Mou   KEECYTPRADSFHIFRDPITGDMPWPGLIFGLAILALWYWCTDQVIVQRCLSAKNMSHVKAGCTLCGYLKLLPMFLMVMPGMIS   335
Rat   KEECYTPRADSFHIFRDPITGDMPWPGLIFGLSILALWYWCTDQVIVQRCLSAKNMSHVKAGCTLCGYLKLLPMFLMVMPGMIS   335
Hum   QEKCYTPRADSFHIFRDPLTGDLPWPGFIFGMSILTLWYWCTDQVIVQRCLSAKNMSHVKGGCILCGYLKLMPMFIMVMPGMIS   335
Bat   KKECYTPRDDSFHIFRDPLKGDLPWPGLIFGLSILALWYWCTDQVIVQRCLSAKNMSHVKAGCIMCGYLKLLPMFLMVMPGMIS   335
Hor   KEECYTPRADSFHIFRDPLKGDLPWPGLIFGLFILTLWYWCTDQVIVQRCLSAKNMSHVKGGCIMCGYLKLLPMFLMVMTGMVS   335
Bov   SESCYTPRADSFHIFRDPINGDLPWPGLLFGLTIQAGWYWCTDQVIVQRCLSAKNLSHVKAGCILCGYLKLLPMFLMVFPGMIS   332
Rab   PQKCYTPREDAFHIFRDAITGDIPWPGLVFGMSILTLWYWCTDQVIVQRCLSAKNLSHVKAGCILCGYLKVMPMFLIVMMGMVS   335
Consensuscytpr d fh fr     d pwpg  fg  i   wywctdqvivqrclsak  shvk gc  cgylk  pm   v  gm s
```

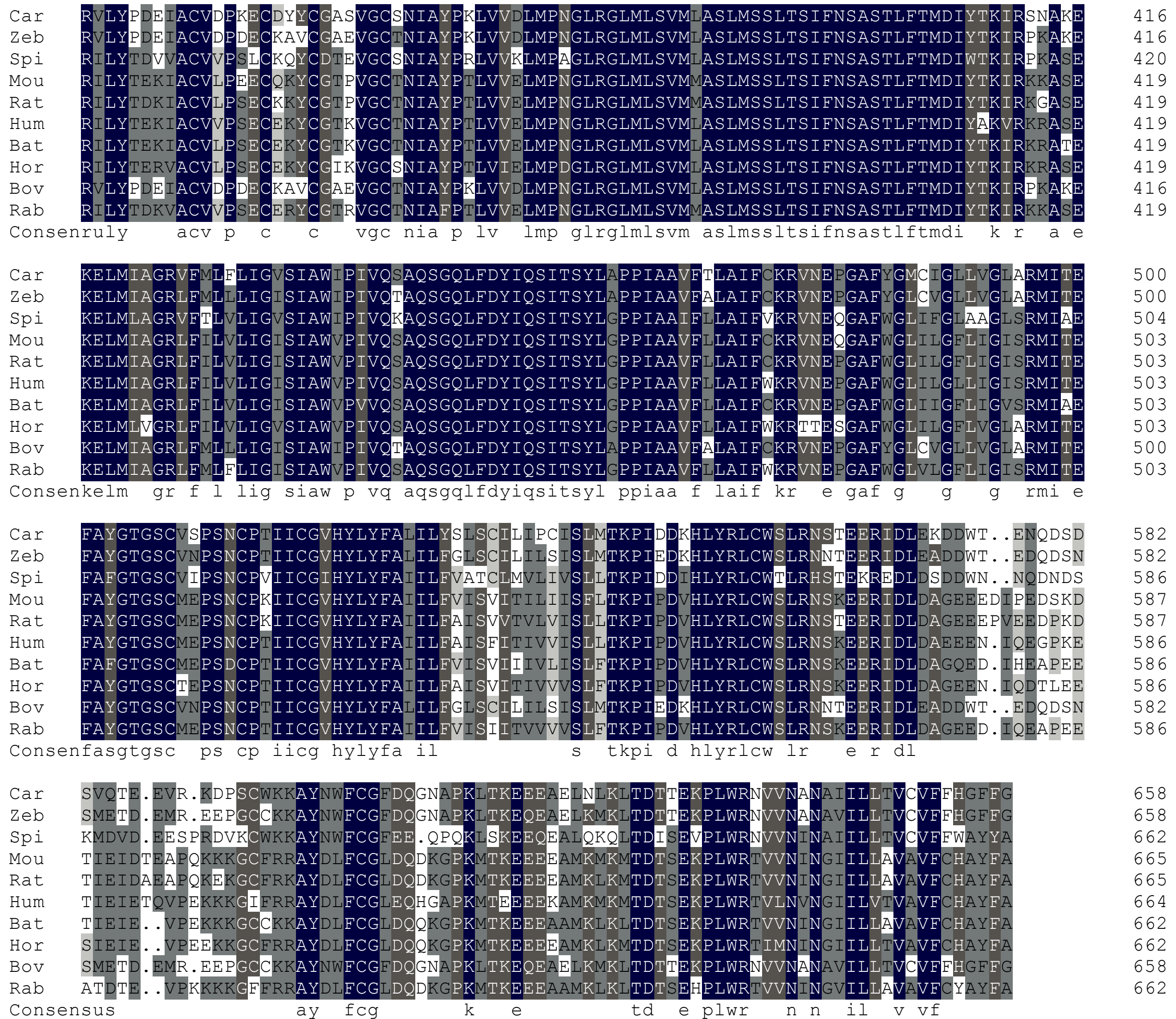

Figure 6. Alignment of deduced amino acid sequences of the Sglt1 subunit from the *Cyprinus carpio* L. (Car), Zebrafish (Zeb), Spiny dogfish (Spi), Human (Hum), Horse (Hor), Mouse (Mou), Rabbit (Rab), Bovine (Bov), Bat (Bat) and Rat (Rat). These protein sequences were aligned using the Clustal program. Identical amino acids are shown on a black background, ≥75% similar amino acids on a red background and ≥50% similar amino acids on a green background.

Totally, based on the characteristic analysis of the transmembrane and cytoplasmic domains, it was indicated to be the highest conservative between *C. carpio* and other species. For the Sglt1 subunit of the *C. carpio* it was high similar with that of Zebrafish, as well as the homologues of Spiny dogfish. On the other hand, phylogenetic trees were constructed using MEGA4.0 (Fig. 7). The nucleotide sequences of *sglt1* gene from several species were classified into two major groups. The Sglt1 subunits of other mammal were clustered into one group. The Sglt1 subunit of *C. carpio*, Zebrafish and Spiny dogfish were clustered into another group. The homology of Sglt1 was highest between *C. carpio* and Zebrafish.

The major member of Sglt1 family have been successfully cloned and sequenced (Hediger *et al.*, 1987; Pajor *et al.*, 1992; Kwon *et al.*, 1992; Kong *et al.*, 1993) in the intestine of human (Hediger *et al.*, 1989), rat (Lee *et al.*, 1994; Aoshima *et al.*, 1997), mouse (Tabatabai *et al.*, 2001), rabbit (Hediger *et al.*, 1987; Morrison *et al.*, 1991), bovine (Zhao *et al.*, 1999), and zebrafish, etc. The results of comparison analysis showed that

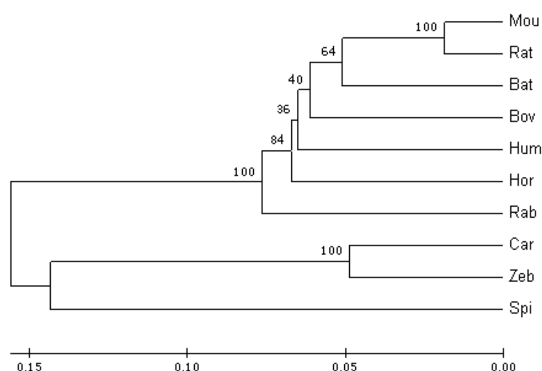

Figure 7. The Phylogenetic relationship of fish Sglt1 and its orthologues. A molecular phylogenetic tree of Sglt1 was generated based on the alignment of the amino acid sequences by MEGA4. The accession numbers for the sequences are as follows: Hum, *Homo sapiens* (AAA60320); Hor, *Equus caballus* (NP001075341); Rab, *Oryctolagus cuniculus* (CAA39040); Bov, *Bos Taurus* (AAM34274); Bat, *Rhinolophus ferrumequinum* (ACC68880), Rat, *Rattus norvegicus* (BAA03676); Mou, *Mus musculus* (AAF17249); Spi, *Squalus acanthias* (CAJ75582), Zeb, *Danio rerio* (NP956975).

the homologous sequences of other fishes were more than 90% and among different animals were more than 70% (Table II). It is showed that *sglt1* gene is the high conserved sequence.

Transporters with a C-terminal extension (e.g. hSglt1) were proposed to have an additional 14th TM (Hirsh *et al.*, 1998). Information on tertiary interactions has recently been gained by chemical cross-linking of splits of the sodium-galactose transporter in *Vibrio parahaemolyticus* (vSglt) (Reizer *et al.*, 1994). However, little is known about the tertiary status of other members of the SSSF. The 11~15 putative transmembrane domains (TMs) in a-helical conformation of SSSF proteins were the average hydropathy plot (Turk *et al.*, 1997). For PutP, it contained 13 TMs with the N-terminus located on the periplasmic side of the membrane and the C-terminus facing the cytoplasm.

To gain structural insight into the mechanistic details, the structure of vSglt were solved in the presence of Na^+ and galactose (Faham *et al.*, 2008). The 3-D protein models in this study were predicted by comparative protein modeling program SWISS-MODEL (Fig. 8). The transporter has two charged regions in common at residues 130-140 and 408-420. It was further confirmed from the similarity between sequences, when compared with the secondary structural elements such as the occurrence of a-helix in front of transmembrane regions 4. The structural core is involved in the inverted repeats of 5 TM helices (TM2-TM6 and TM7-TM11). It was suggested that glucose might be bound in the center of the core, and a Na^+-binding site might be located at the intersection of TM2 and TM9. Results of the study are in line with the result of Sodium-binding site and galactose-binding site by Faham *et al.* (Faham *et al.*, 2008).

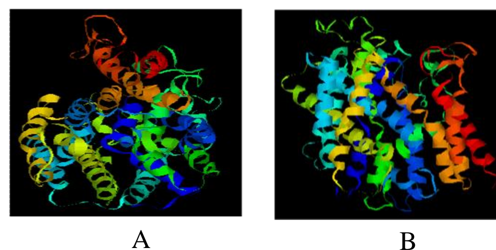

Figure 8. The predicted 3-D structure of Sglt1 in intestine of *Cyprinus carpio* L. A, Structure of Sglt1 viewed in the membrane plane. B, Structure of Sglt1 viewed from the intracellular side.

Acknowledgements:

This study was supported by Program Science & Technology Innovation Talents in Universities of Henan Province (2010HASTIT020).

Corresponding Author:
Dr. Guoxing Nie
College of Life Science
Henan Normal University
Xinxiang, Henan 453007, P R .China
E-mail: niegx@htu.cn

References

1. Reizer J., Reizer A.,Saier M.H.Jr.. A functional superfamily of sodium/solute symporters. Biochim Biophys Acta, 1994;1197(2):133-166.
2. Frank Spiegelhalter, Erhard Bremer. Osmoregulation of the opuE proline transport gene from Bacillus subtilis: contributions of the sigma A- and sigma B-dependent stress-responsive promoters. Molecular Microbiology, 1998;29(1):285-296.
3. Schwan W.R., Coulter S.N., Ng E.Y., Langhorne M.H., Ritchie H.D., Brody L.L., Westbrock-Wadman S., Bayer A.S., Folger K.R., Stover C.K.. Identification and characterization of the PutP proline permease that contributes to in vivo survival of Staphylococcus aureus in animal models. Infect Immnu., 1998;66(2):567-572.
4. Zeuthen T., Meinild A.K., Loo D.D., Wright E.M., Klaerke D.A.. Isotonic transport by the Na^+-glucose cotransporter $SGLT_1$ from humans

and rabbit. Journal of Physiology, 2001;531:631-644.

5. Hirayama B.A., Loo D.D., Wright E.M.. Cation Effects on Protein Conformation and Transport in the Na^+/glucose Cotransporter. J Biol. Chem., 1997;4:2110-2115.
6. Eskandari S., Loo D.D., Dai G., Levy O., Wright E.M., Carrasco N. Thyroid Na^+/I^- Symporter. J Biol. Chem, 1997;43:27230-27238.
7. Wegener C., Tebbe S., Steinhoff H.J., Jung H.. Spin Lableling Analysis of Structure and Dynamics of the Na^+/Proline Transporter of *Escherichia coli*÷. Biochemistry, 2000;39:4831-4837.
8. Martin M.G., Turk E., Lostao M.P., Kerner C., Wright E.M.. Defects in Na^+/glucose cotransporter ($SGLT_1$) trafficking and function cause glucose-galactose malabsorption. Nature Genetics. 1996;12(2):216-220.
9. Ernest M. Wright, Donald D. F. Loo, Bruce A. Hirayama, and Eric Turk. Surprising Versatility of Na^+/Glucose Cotransporters: SLC5. Physiology (Bethesda). 2004;19:370-376.
10. Wright E.M., Turk E.. The sodium/glucose cotransport family SLC5. Pfluegers Arch, 2004;447(5):510-518.
11. Turk E., Wright E.M.. Membrane topology motifs in the SGLT cotransporter family. J Membr Biol, 1997;159(1):1-20.
12. Balen D., Ljubojevic M., Breljak D., Brzica H., Zlender V., Koepsell H., Sabolic I. Revised immunolocalization of the Na^+-D-glucose cotransporter $SGLT_1$ in rat organs with an improved antibody. Arn J Physiol Cell Physiol, 2008;295(2):C475-C489.
13. Hirsh A.J., Cheeseman C. I. Cholecystokinin decreases intestinal hexose absorption by a parallel reduction in $SGLT_1$ abundance in the brush-border membrane. J Biol Chem, 1998;273(23):14545-14549.
14. Ikeda T.S., Hwang E.S., Coady M.J., Hirayama B.A., Hediger M.A., Wright E.M.. Characterization of a Na^+/glucose cotransporter cloned from rabbit small intestine. J Membr Biol, 1989;110(1):87-95.
15. Ikumi Y., Kida T., Sakuma S., Yamashita S., Akashi M.. Polymer-phloridzin conjugates as an anti-diabetic drug that Inhibits glucose absorption through the Na^+/glucose cotransporter ($SGLT_1$) in the small intestine. J Control Release, 2008;125(1):42-49.
16. Sabino-Silva R., Mori R. C., David-Silva A., Okamoto M. M., Freitas H. S., Machado U. F., The Na^+/glucose cotransporters: from genes to therapy. Braz. J. Med. Res. 2010;43,1019-1026.
17. Ryuichi Moriya, Takashi Shirakura, Junko Ito, Satoshi Mashiko and Toru Seo, Activation of sodium-glucose cotransporter 1 ameliorates hyperglycemia by mediating incretin secretion in mice. Am J. Physiol Endocrinol Metab 2009;297,E1358-E1365.
18. Salem Faham, Akira Watanabe, Gabriel Mercado Besserer, Duilio Cascio, Alexandre Specht, Bruce A. Hirayama, Ernest M. Wright, Jeff Abramson. The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na^+/sugar symport. Science 2008;321,810-814.
19. F. Q. Zhao, T. B. McFadden, E. H. Wall, B. Dong, Y.-C. Zheng. Cloning and Expression of Bovine Sodium/Glucose Cotransporter $SGLT_2$. J. Dairy Sci. 2005;88,2738-2748.
20. Lubing Zhou, Ellen V. Cryan, Michael R. D′Andrea, Sranley Belkowski, Bruce R. Conway, Keith T. Demarest. Human Cardiomyocytes Express High Level of Na^+/Glucose Cotransporter 1 ($SGLT_1$). Journal of Cellular Biochemistry 2003;90,339-346.
21. Ana Díez-Sampedro and Stephanie Barcelona. Sugar binding residue affects apparent Na^+ affinity and transport stoichiometry in mouse sodium/glucose cotransporter type 3B. J. Biol Chem. 2010; In press.
22. Hediger M.A., Coady M.J., Ikeda T.S., Wright E.M.. Expression cloning and cDNA sequencing of the Na^+/glucose co-transporter. Nature, 1987;330:379-381.
23. Pajor A.M., Wright E.M.. Cloning and functional expression of a mammalian Na^+/nucleoside cotransporter. A member of the SGLT family. J Biol Chem., 1992;267:3557-3560.
24. Kwon H.M., Yamauchi A., Uchida S., Preston A.S., Garcia-Perez A., Burg M.B., Handler J.S.. Cloning of the cDNa for a Na^+/myo-inositol cotransporter, a hypertonicity stress protein. J Biol Chem., 1992;267(9):6297-6301.
25. Kong C.T., Yet S.F., Lever J.E.. Cloning and expression of a mammalian Na^+/amino acid cotransporter with sequence similarity to Na^+/glucose cotransporters. J Biol Chem., 1993;268(3):1509-1512.
26. Hediger M.A., Turk E., Wright E.M.. Homology of the human intestinal Na^+/glucose and Escherichia coli Na^+/proline cotransporters. Proc Natl Acad Sci USA, 1989;86(15):5748-5752.
27. Lee W.S., Kanai Y., Wells R.G., Hediger M.A.. The high affinity Na^+/glucose cotransporter. Re-evaluation of function and distribution of expression. J Biol Chem., 1994;269(16):12032-12039.

28. Aoshima H., Yokoyama T., Tanizaki J., Yamada M.. The sugar specificity of Na^{+}/glucose cotransporter from rat jejunum. Biosci Biotechnol Biochem, 1997;61(6):979-983.
29. Tabatabai N.M., Blumenthal S.S., Lewand D.L., Petering D. Differential regulation of mouse kidney sodium-dependent transporters mRNA by cadmium. Toxicology and Applied Pharmacology, 2001;177(3):163-173.
30. Morrison A.I., Panayotova-Heiermann M., Feigl G., Schölermann B., Kinne R.K.. Sequence comparison of the sodium-D-glucose cotransport systems in rabbit renal and intestinal epithelia. Biochim Biophys Acta. 1991;1089(1):121-123.
31. F. Q. Zhao, Okine E. K., Kennelly J. J.. Glucose transporter gene expression in bovine mammary gland. J Anim Sci. 1999;77,2517-2522.

2/3/2011