THE CHARACTERIZATION, EXPRESSION AND IN SILICO STUDIES ON THE SLC39A13 GENE; IT’S INVOLVEMENT IN BREAST CANCER

NORMAWATI MOHAMAD ZAHARI
Institute of Biological Sciences,
Faculty of Science, University of Malaya,
50603 Kuala Lumpur, Malaysia
normazahari@um.edu.my

TEOH TEOW CHONG
Institute of Biological Sciences,
Faculty of Science, University of Malaya,
50603 Kuala Lumpur, Malaysia
ttchong@um.edu.my

The zinc transporters superfamily is divided into four subfamilies, and SLC39A is one of the subfamilies. The SLC39A subfamily has 9 members. Based on our computer searchers, all 9 sequences each contain 8 transmembranes domains. Since it is related to the zinc transporters superfamily, the SLC39A subfamily may have the same function that is to transport zinc ion. This paper focus on SLC39A13 studies and using the recombinant technology with CHO cells, it is shown that the recombinant protein, pcDNA5/FRT/V5-His-TOPO®-SLC39A13, has 43kD molecular weight. A second study using immunofluorescence technique with MCF-7 cells, it is shown that the recombinant protein expresses intracellularly. Both studies demonstrate that SLC39A13 expresses in breast cancer cells line, therefore the gene has involvement in the development of breast cancer disease. In our computational studies which is divided into two; the homologous study and sequence analysis, both results are supporting our laboratory results. The homologous study using EMBL-EBI and UniProt tools concluded that SLC39A13 is a member to SLC39A subfamily and it is closely related to SLC39A7 member. Although the sequence analysis shows that the molecular weight of SLC39A13 is 38.35kD it is still comparable to our laboratory result. Separately, using Swiss-EMBnet tools, TMpred, has shown that SLC39A13 has 8 transmembranes domains similar to other family members of SLC39A subfamily. Another analysis using EMBL-EBI tools, PPsearch, shows that SLC39A13 has various protein motif such as the protein kinase C phospho, casein kinase II phospho, leucine zipper and ASN-glycosylation sites. These are the useful information that we need when we study its tertiary structure and simulation in the future.

Keywords: Zinc transporter; gene expression; resistant breast cancer cells; in-silico/computational study.

1. Introduction

Breast cancer is the commonest cancer in woman among all races from the age of 15 years in Malaysia. In 2003, statistic recorded the breast cancer as most common cancer
affecting the females and it had account 31.0% of newly diagnosed\(^1\). Breast cancer is caused by an uncontrolled growth of cells; therefore develop the formation of tumor within breast tissue. Many risk factors can lead to breast cancer disease but it is difficult to know how much the factors had contributed to the cancer disease\(^3\). Hence, interestingly in this study is the success in identifying the expression of the zinc transporters families in breast cancer cells line. However, the study will be only focusing on one gene, which is SLC39A13 gene.

General information on zinc ion, it is an important element in cells as it involve numerous in cellular processes. Such as, a cofactor for many enzymes and transcription factors and participates in regulating cells proliferation and division\(^3-4\). Without zinc, cells might die. Nevertheless, level of zinc ion existence in cells needed to control because it can cause an effect in cells. High level of zinc in cells might become toxic to cells, whereas in deficiency of zinc ion can cause stunted growth in cells\(^5\). Therefore, zinc transporters proteins are require in regulating the zinc ion in cells. Zinc transporters proteins categorized to two superfamilies, based on their opposite role in transporting the zinc ion. They are ZnT superfamily, function to transport the zinc extracellularly, while ZIP superfamily is function to transport the zinc intracellularly\(^6\).

ZIP is a designation that stands for Zrt-, Irt-like protein. It is capable to transport various elements other than zinc into cytoplasm\(^7,8\). There are over 86 ZIP members been found and they were grouped into subfamilies\(^8\). SLC39A subfamily is one of the subfamilies of ZIP superfamily. They are predicted to have eight transmembrane domains, and occurrence within domains IV to V an extramembranal metal binding protein in these proteins are highly conserved with histidines\(^10-11\). SLC39A subfamily consists of 14 human sequences, as a result only nine members successfully showing their function\(^6-11\).

SLC39A13 is a gene that belongs to the SLC39A subfamily\(^6-10-11\). Length for SLC39A13 protein sequence is 364 amino acid residues. SLC39A13 is showing a close relationship with other members from SLC39A subfamily members, SLC39A7 sequence\(^11-13\). It has been found, SLC39A7 had an effect on EGF/IGF signaling and tamoxifen resistancy of breast cancer cells\(^11-12\). This may shows similarity in protein sequences between SLC39A13 and SLC39A7. Hence, the objectives are to study the characterization and expression of SLC39A13 using laboratory experiment and in-silico studies and also to study the molecular characteristics for SLC39A13.

2. Methodology

Materials in this study are divided into two, the laboratory experiments and computational studies. Together, these studies are used to compare each other result.
2.1. **Laboratory experiments**

The laboratory experiment involved in cloning the SLC39A13 gene, accession number NM_152264, into pcDNA5/FRT/V5-His TOPO® TA vector and CHO cells as the host. The gene was prepared by using PCR technique to multiply its amount via the PCR program: Denaturation 94°C, Annealing 57°C and Extension 72°C parameters for 26 cycles.

Western blot analysis is to identify the SLC39A13 recombinant protein in CHO cells, using V5-antibody (Invitrogen) which is available commercially and the dylution factor is 1:5000. The recombinant plasmid was sent to sequence service provider to confirm its nucleotides are in the order as in the Genbank.

The immunofluorescence studies involving 2 types of cells; CHO cells and MCF-7 treated with Tamoxifen (anti-hormonal drug against breast cancer) to become resistant. Experiment using the CHO cells is using the Alexa Fluor 488-conjugated antimouse and Alexa Fluor 594-conjugated antirabbit probes, both from Molecular Probes, used in 1:5000 dylution. The Texas Red Phalloidin (Molecular Probes) reagent is to detect F-actin. Vectorshield containing 1.5 µg/ml of DAPI is used to counterstain the cell nuclei. To conduct the zinc assay experiment, the method uses 50 µM of Newport Green™ diacetate dye.

2.2. **Computational studies**

In computational study, two different homologous studies on SLC39A13 were done separately using Uniprot tools, BlastP to find the similarity sequence to SLC39A13 protein sequence. The sequences were seperated into 2 groups of protein sequences. First group of sequences is to do the homologous study of SLC39A13 with other 8 members from SLC39A subfamily and second group of sequences is for a homologous study of SLC39A13 protein sequences amongst the 7 organism. Uniprot tools, Align, is used to align the groups of sequences separately and build the phylogenetic trees. The accession number for both groups of sequences were recorded as a reference for our next study on sequence analysis (Table 1 & Table 2).

<table>
<thead>
<tr>
<th>Human Gene Name</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLC39A7</td>
<td>NP_001070984.1</td>
</tr>
<tr>
<td>SLC39A10</td>
<td>NP_001120729.1</td>
</tr>
<tr>
<td>SLC39A6</td>
<td>NP_036451.3</td>
</tr>
<tr>
<td>SLC39A14</td>
<td>NP_056174.2</td>
</tr>
<tr>
<td>SLC39A4</td>
<td>NP_570901.2</td>
</tr>
<tr>
<td>SLC39A5</td>
<td>NP_071437.3</td>
</tr>
<tr>
<td>SLC39A8</td>
<td>NP_775867.2</td>
</tr>
<tr>
<td>SLC39A12</td>
<td>NP_689938.2</td>
</tr>
<tr>
<td>SLC39A13</td>
<td>NP_689477.2</td>
</tr>
</tbody>
</table>

*Note: First group of sequences SLC39A family members and accession number for protein sequence were collected from NCBI database.*
Table 2. Organism Accession Number for SLC39A13 Protein Sequence

<table>
<thead>
<tr>
<th>Organism</th>
<th>Accession Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homo Sapiens</td>
<td>Q96H72-2</td>
</tr>
<tr>
<td>Mus Musculus</td>
<td>Q8BZH0</td>
</tr>
<tr>
<td>Pongo abelii</td>
<td>Q5R616</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>Q2M1K6</td>
</tr>
<tr>
<td>Bos taurus</td>
<td>A5DZH1</td>
</tr>
<tr>
<td>Gallus gallus</td>
<td>Q5Z120</td>
</tr>
<tr>
<td>Danio rerio</td>
<td>Q8AW42</td>
</tr>
</tbody>
</table>

Note: Second group of sequences. Study between SLC39A13 sequence from human and other organism were collected from Swiss-Prot database.

Sequence analysis is a second computational study on SLC39A13 protein sequence by analyzing the protein sequence to identify its characterization. SLC39A13 accession number is referred to the homologous study. 3 analysis using 4 online bioinformatics tools to analyze the SLC39A13 protein sequence; ExPasy tools, Compute pi/Mw is used in finding the molecular weight for SLC39A13 protein sequence, TMPred used in predicting the transmembrane region for SLC39A13 protein sequence, and the output graph produced from TMPred were transformed into a schematic diagram using TOPO2, and EMBL-EBI tools, PPSearch used in predicting the protein motifs appearance in SLC39A13 protein sequence.

3. Results

3.1. Laboratory experiment

3.1.1. Using CHO cells (I)

Results from the laboratory experiment showed that the PCR technique is successful to produce 1094bp of SLC39A13 plasmid and it is also shown that the recombinant protein produces 43kD of molecular weight in the expression study using CHO cells (Figure 1).

![Figure 1](image-url)

Figure 1. The left-hand side picture is showing the gel electrophoresis of SLC39A13 with a control. The right-hand side picture is showing the Western blot membrane of SLC39A13 recombinant protein with SLC39A12 and SLC39A5 recombinant proteins as controls.
3.1.2. Using CHO cells (II)

The recombinant protein of pcDNA5/FRT/V5-His-TOPO®-SLC39A13 was transfected in CHO cells and the recombinant protein was identified using Alexa Flour 488-conjugated antimouse probe. It is shown that the recombinant protein is found in the cells that are not permeabilised treatment (Figure 2A). The recombinant protein is mostly located on the plasma membrane of the cells as shown in green color (Figure 2A). Separately, the recombinant protein was identified using Alexa Flour 594-conjugated antirabbit probe (Figure 2B) but the recombinant protein is mostly located surrounding the area of endoplasmic reticulum, as shown in red color (Figure 2B). Both pictures of Figure 2A and Figure 2B, has nuclei stain with DAPI as shown with blue color while only in Figure 2A, the F-actin is identified using Texas Red Phalloidin as shown with red color.

![Figure 2](image)

**Figure 2.** A) SLC39A13 is stain with Alexa Flour 488-conjugated antimouse probe and mostly located on the plasma membrane of non-permeabilised treated CHO cells. B) SLC39A13 recombinant protein is stained with Alexa Flour 594-conjugated antirabbit probe suggestive of endoplasmic reticulum staining.

3.1.3. Using MCF-7 Taxoxifen resistant cells

In this experiment using MCF-7 Tamoxifen resistant cells, the SLC39A13 recombinant protein was probe using Alexa Flour 594-conjugated antirabbit probe (Figure 3A). The procedure also involves using 50 µM of Newport Green™ diacetate dye which produces a green color fluorescence image (Figure 3B). Upon merging both images in Figure 3A and Figure 3B, produced a third image which showed brighter green color after the effect of the red color (Figure 3C). Also in Figure 3C, a cell showed brighter green color but seems to be appearing in “a dot format” (see red arrow in Figure 3C). Both picture, Figure 3A and Figure 3C have nuclei stained with DAPI as shown with blue color.
3.2. Computational studies

3.2.1. Homologous study

Result for computational study on homologous sequence shows SLC39A13 protein sequence is closely related to SLC39A7, which is also a member from SLC39A subfamily. The SLC39A13 nodes and SLC39A7 nodes are in the same cluster, shown in the below phylogenetic tree of SLC39A subfamily (Figure 4).
The Characterization, Expression and in Silico Studies on the SLC39A13 Gene

The second phylogenetic tree is showing the result of homologous study on SLC39A13 protein sequences between 7 living organisms. From the result it shows that the Homo sapiens and Pongo abelii are clustered together because they are more identical to each other. Secondly, the Mus Musculus and Rattus norvegicus are also produced by the same branched suggesting to be closely identical to each other. Nevertheless, all these organisms belong to the same main root signifying that the SLC39A13 genes in them are almost identical to each other.

3.2.2. Sequence analysis (I)

The SLC39A13 protein sequence analysis is to analyze it molecular weight using ExPasy proteomic server tools, Compute pI/Mw and the result shows that the molecular weight for SLC39A13 is 38.35 kD.
The second analysis on SLC39A13 protein sequence is to analyze the sequence by using Swiss-EMBnet tools, TMPred and it is found that there are 8 transmembrane domains in the prediction (Figure 6A).

Figure 6A. The graph is showing the transmembrane domains for SLC39A13 protein sequence. The red line indicates the threshold limit for the transmembrane domains which identify regions above 500 scores are considered a transmembrane domain. There are 8 peaks that are above the threshold value therefore SLC39A13 protein sequence is predicted to have 8 transmembrane domains.

From the result of TMPred search (Figure 6A), the data was transformed into another representation of the data into a schematic diagram. Subsequently, by using the Transmembrane Protein Display Page, TOPO2, shows that the SLC39A13 protein sequence has a long N-terminus and short C-terminus. Secondly, it is observed that there are 8 transmembrane domains passing through the plasma membrane (Figure 6B).

Figure 6B. A schematic diagram showing a linear format of amino acid arrangement in SLC39A13 protein sequence. This diagram shows that the SLC39A13 molecule has 8 transmembrane domains with long N-terminus and short C-terminus. It is also shown the HEXPHEX motif is located inside cytoplasm and it is between transmembrane V and VI.
To find the proteins motif, the analysis is using EMBL-EBI tools, PPSearch, and the results show that there are 5 protein motifs prediction for SLC39A13 protein sequence. Interestingly, the frequencies of occurrences of each pattern are different. Highest frequency found is the myristilation having 12 sites occurrences, while protein kinase C phospho has 3 sites; bearing the sequences of \{(S)erine, (I)soleucine, (K)lysine (site 1)\}, \{(S)erine, (K)lysine, (K)lysine (site 2)\} and \{(S)erine, (P)proline, Lysine (site 3)\}. The ASN-glycosylation, casein kinase II phospho site and leucine zipper has the lowest frequency by having one site each. The ASN-glycosylation sequence is \{(N)asparagine, (F)phenylalanine, (T)threonine, (H)histidine\}. As for casein kinase II phosphor sequence is \{(K)lysine, (E)glutamic acid, (E)glutamic acid\}. For leucine zipper sequence is \{(L)leucine, (A)alanine, (V)valine, (A)alanine, (S)erine, (F)phenylalanine, (L)leucine, (V)valine, (S)erine, (K)lysine, (K)lysine, (I)soleucine, (G)lysine, (L)lysine, (L)lysine, (T)threonine, (T)threonine, (M)methionine\}. This information is compiled into a diagram shown in Figure 7.

![Figure 7. A schematic diagram showing all the binding patterns sites located at each distinct site of SLC39A13 protein sequence.](image_url)

4. Discussion

4.1. Laboratory experiment

The laboratory experiment is to find the molecular weight of SLC39A13 and the result shows it has 43kD molecular weight, whereas the computational studies showed that the molecular weight is 38.35kD only. Both of these results are considered comparable although the formal molecular weight is more than the latter value.

The SLC39A13 recombinant protein is expressed on the plasma membrane in an experiment using CHO cells with immunofluorescence technique. The CHO cells were
treated in both permeabilised and non-permeabilised condition but the SLC39A13 recombinant protein expressed in the non-permeabilised condition and located on the plasma membrane.

In the pilot study of zinc experiment using MCF-7 Tamoxifen resistant cells, it is shown that the SLC39A13 recombinant protein is suggested to be located on the same site where the Newport Green™ diacetate, a dye for zinc ion, is observed.

4.2. Computational studies

From the phylogenetic tree that was produced, shows the SLC39A13 protein sequence is closely related to SLC39A7, which is also a member from the SLC39A subfamily. As both SLC39A13 and SLC39A7 appear within the same cluster in phylogenetic tree, it is predicted that SLC39A13 may have similar function to SLC39A7.

The phylogenetic tree is a homologous study on SLC39A13 protein sequence amongst other 7 organism and they are Homo sapiens, Mus musculus, Danio rerio, Rattus norvegicus, Bos taurus, Gallus gallus and Pongo abelii. Closest node to Homo sapiens is Pongo abelii while the second organism that closest to Homo sapiens are Rattus norvegicus and Mus musculus. Hence, it is suggested that Rattus norvegicus and Mus musculus may be useful as a model for our laboratory experiment.

There are 8 transmembrane region been predicted in SLC39A13 protein sequence using Swiss-EMBnet tools, TMpred. Related to this result, the 8 transmembrane domains predicted is similar feature to the other members of SLC39A subfamily.

5. Conclusion

This research has found new biology data of SLC39A13 molecule and this was supported by our in-silico/computational searches as shown above. Nevertheless, the biology data need to be converted into physical biology where advanced studies can be done using the tertiary structure model using computer software and allowed us to do simulation studies. The above studies and data collected are the foundation of understanding the SLC39A13 molecule for our future work.

References