Title
Bioinformatic Analysis of Mitochondrial Pyruvate Carriers

Permalink
https://escholarship.org/uc/item/7j47x40w

Author
Myers, Scott Randall

Publication Date
2015

Peer reviewed|Thesis/dissertation
Bioinformatic Analysis of Mitochondrial Pyruvate Carriers

A Thesis submitted in partial satisfaction of the requirements for the degree
Master of Science

in

Biology

by

Scott Randall Myers

Committee in charge:

Dr. Milton Saier, Chair
Dr. Eric Allen, Co-Chair
Dr. Lakshmi Chilukuri

2015
The Thesis of Scott Randall Myers is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________

________________________________________

Co-Chair

________________________________________

Chair

University of California, San Diego

2015
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Figures</td>
<td>v</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Supplemental Files</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract of the Thesis</td>
<td>ix</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Methods</td>
<td>4</td>
</tr>
<tr>
<td>Results</td>
<td>7</td>
</tr>
<tr>
<td>MPC Topology</td>
<td>7</td>
</tr>
<tr>
<td>Investigation of Possible Relationships with Mitochondrial Carriers</td>
<td>7</td>
</tr>
<tr>
<td>MPCs belong to the TOG Superfamily</td>
<td>8</td>
</tr>
<tr>
<td>Phylogenetic Tree of Representative SWEET and MPC Homologues</td>
<td>10</td>
</tr>
<tr>
<td>Investigation of MPC Fusion Protein</td>
<td>10</td>
</tr>
<tr>
<td>Discussion</td>
<td>13</td>
</tr>
<tr>
<td>Figures</td>
<td>16</td>
</tr>
<tr>
<td>Tables</td>
<td>29</td>
</tr>
<tr>
<td>Supplemental Files</td>
<td>30</td>
</tr>
<tr>
<td>References</td>
<td>32</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1. Average hydropathy, amphipathicity and similarity (AveHAS) plots based on ClustalX alignments for MPC, 7 TMS SWEET and 3 TMS SWEET families........... 16

Figure 2. Sequence alignments do not provide evidence for homology between MPC and MC homologs.......................................................................................... 17

Figure 3. Binary sequence alignments between MPC family homologues and SWEET family homologues................................................................................... 19

Figure 4. Evidence for a conserved sequence motif shared between MPC and SWEET families............................................................................................................ 20

Figure 5. Proposed evolutionary pathway of the ancestral 4 TMS repeat for the MPC and SWEET families along with their relationship......................... 22

Figure 6. A phylogenetic tree based off multiple alignments depicting SWEET and MPC family separation ................................................................. 23

Figure 7. Average hydropathy and amphipathicity plots with predicted TMSs and domains for fusion protein Pma1 .................................................................................. 24

Figure 8. Binary alignments showing N-terminal MPC1 and C-terminal MPC2 domains in the fusion MPC protein, Pma1 ................................................................. 25

Figure 9. Schematic depiction showing the proposed evolutionary pathway for 6-TMS MPC fusion proteins and the pathway for most TOG superfamily proteins ......... 26

Figure 10. Binary alignments displaying N-terminal MPC1 and C-terminal MPC2 domains in the fusion MPC protein, Aan1................................................................. 28
LIST OF TABLES

Table 1. Representative MPC and SWEET/MtN3 family proteins used in the MEME common motif analysis (Figure 4A), the multiple alignment (Figure 4B) and the phylogenetic tree (Figure 5).

Table 2. List of possible MPC fusion proteins used in TMS topology analysis (Figure 7 and Figure 9) and domain identifications (Figure 9 and Figure 10).
LIST OF SUPPLEMENTAL FILES

**Figure S1.** Flowchart schematic showing the Superfamily Principle transitivity rule for establishing homology of distant protein families ................................................................. 30

**Figure S2.** Average residue hydropathy and amphipathicity characteristics with predicted TMSs and domains for fusion protein Pma2 ................................................................. 31
ACKNOWLEDGEMENTS

I would like to acknowledge all those who have come and gone during my 3 years in the Saier Lab. With them I learned and experienced much of what makes me the person I am today.

I would also really like to thank the wonderful Dr. Saier. His patience has never faltered and he has challenged me to develop my own ideas and curiosities. His mentorship has been formative in my life and of those of my peers. His dedication to our future and the preservation of the environment should be inspirational to us all.
ABSTRACT OF THE THESIS

Bioinformatic Analysis of Mitochondrial Pyruvate Carriers

by

Scott Randall Myers

Master of Science in Biology

University of California, San Diego, 2015

Professor Milton Saier, Chair
Professor Eric Allen, Co-Chair

The Mitochondrial Pyruvate Carrier (MPC) family of eukaryotic protein transporters reside in the mitochondrial inner-membrane and are the essential for the 150kDa transport complex of pyruvate into the mitochondrial matrix for cellular respiration. Studies in this paper show evidence that the two primary isoforms of the family, MPC1 and MPC2, both have 3-TMS topologies. Exploration of family homologies resulted in a relationship to the bidirectional sugar transporter (SWEET)
family and by extension the Transporter/Opsin/G-protein (TOG) superfamily. The two families share a 41-residue sequence motif with high sequence conservation. Binary sequence alignments using the program GSAT allowed for confirmation of the 3 TMSs of MPC aligning with the corresponding 3-TMSs of SWEET family proteins. Comparisons to the TOG superfamily show MPC evolved from the N-terminal TMS loss on the 4-TMS TOG precursor protein. Family sequence analysis also resulted in the discovery of four novel fusion proteins consisting of MPC1 and MPC2 domains to form a predicted functional pyruvate transport unit. These 6-TMS fusion proteins arose from a novel evolutionary pathway dissimilar than the other discovered TOG superfamily pathways. MPC familial homology between SWEET and other TOG superfamily proteins may help to elucidate its transport mechanism in future studies.
INTRODUCTION

Members of the Mitochondrial Pyruvate Carrier (MPC) family (TC# 2.A.105) are found in the mitochondrial inner-membrane and play a central role in cellular metabolic respiration [1]. Cytosolic 3-carbon pyruvate molecules, formed in the cytoplasm via glycolysis or transported into the cell by Monocarboxylate Transporter (MCT) Family members (TC# 2.A.1.13) must journey to the mitochondrial matrix where they feed into the citric acid cycle to drive ATP production for cellular anabolism [2]. To reach the mitochondrial matrix, pyruvate must cross two mitochondrial membranes, first, the outer membrane and second, the inner membrane. The pyruvate dehydrogenase complex (PDC) and pyruvate carboxylase (PC), respectively, convert pyruvate into TCA cycle constituents, acetyl-CoA and oxaloacetate [3]. Pyruvate transport across the outer mitochondrial membrane into the intermembrane space occurs via large nonspecific porins of the VDAC family (TC# 1.B.8) [4], while the recently identified MPC family mediates transport across the inner membrane into the matrix [1, 5]. MPC transport is driven by the proton motive force (pmf). These porters allow strict regulation of mitochondrial oxidative metabolism and catalyze the rate-limiting step in pyruvate catabolism [6, 7].

Mammalian MPCs have 2 primary isoforms, MPC1 and MPC2, that share ~30% sequence identity and are well conserved from yeast to humans [1, 8]. A third isoform, MPC3, is found in yeast with 79% sequence identity to MPC2 and appears to be a result of a more recent gene duplication event [5]. MPC proteins form heterocomplexes involving MPC1 and MPC2 or MPC3 which together join other unidentified proteins to form an approximately 150 kilodalton complex [1, 5]. MPC2 appears to be a crucial
structural subunit as it is required for MPC1 co-localization [5]. *In vivo*, both MPC proteins are necessary for pyruvate transport. However, homodimers of MPC2 can mediate limited transport although homodimers of MPC1 have not been shown to be active. [5]

Members of the MPC family contain a highly conserved motif (hidden Markov model; HMM) that characterizes the MPC domain (Pfam designation: PF03650 or UPF0041) [1]. The domain includes the conserved transmembrane segments (TMSs) of each protein. Most of the sequence variability within these proteins is observed within their C-terminal regions, and mammalian homologs lack an N-terminal extra mitochondrial domain found in many MPC family members [6]. The recent molecular identification of MPCs has resulted in research focused mainly on MPC characterization and physiology. However there is a wealth of sequence information revealing its possible evolutionary origin. Analyses of the topologies of these proteins and specification of their evolutionary origins are reported here. This family belongs to the Transporter/Opsin/G-protein (TOG) superfamily, one of the most functionally diverse transport families found in nature [9]. The TOG superfamily is characterized by an ancestral 4 TMS repeat unit. Phylogenetic trees allow documentation of the relationships of these proteins to others of the superfamily.

This paper focuses on specific relationships to the bidirectional sugar transporter SWEET (MtN3) family (TC# 9.A.53), a established member of the TOG superfamily [9]. The SWEET family includes homologs that exist in eukaryotes as 7 TMS proteins while in prokaryotes they can be full-length 7 TMS proteins although half sized 3 TMS proteins are also present. In relation to the TOG superfamily, the 3 TMS SWEET homologs
contain a single repeat unit and resulted from the loss of the first TMS of the ancestral 4 TMS repeat unit [9]. The 7 TMS SWEET family arose from duplication of the ancestral 4 TMS repeat unit followed by loss of the first TMS [9].
METHODS

Obtaining Homologous Sequences for Study

An initial list of MPC proteins was generated using the Pfam database to obtain proteins exhibiting the conserved MPC (PF03650) domain [10]. This list included a diverse sample of MPC1, MPC2 and MPC3 homologues to be used for whole family comparisons.

To create more extensive lists of homologs from three different transport protein families, query sequences were obtained from the Transporter Classification Database (TCDB) [11]. The query sequences used were (1) MPC2 of Homo sapiens (TC# 2.A.105.1.2, O95563), (2) SLC25A5, a Mitochondrial Carrier (MC) family protein of Homo sapiens (TC# 2.A.29.1.1, P05141) and (3) SWEET11 of Arabidopsis thaliana (TC# 9.A.58.1.13, Q9SMM5). NCBI PSI-BLAST searches were performed on each individual sequence using the Protocol1 program [12]. Two iterations with an e-value cut-off of 0.005 and a maximal number of 1000 hits were used as the parameters for each search. The retrieved sequences were formatted into FASTA files and automatically curated by the CD-HIT program to eliminate redundant and closely similar sequences. An 80% cut off was used so that no two proteins in the final list were more than 80% identical [13]. To remove fragmentary protein sequences, those below 40 aas were removed. Relatively large sequences above 500 aas were also eliminated.
Creating Multiple Alignments

Multiple alignments of homologous proteins were generated using the ClustalX program [14]. These multiple alignments were used to identify and remove probable fragmentary proteins along with proteins that introduced significant unique gaps. ClustalX was also used to generate the .ALN files required for further analysis and to identify areas of sequence conservation between families.

Topological Analysis

The Web-based Hydropathy, Amphipathicity and Topology (WHAT) program was used to analyze single sequence topologies [15]. WHAT is in part based on the HMMTop program which graphically displays the locations and magnitudes of α-helical transmembrane segments (TMSs) based on individual amino acyl residue hydropathy values and Hidden Markov Models [16]. A similar program, AveHAS, uses a multiple alignment generated from ClustalX to depict the average hydropathy, amphipathicity, and similarity plots for the included proteins [17]. This allowed more accurate evaluation of protein family topologies.

Establishing Homology

Preliminary investigation into family homology was conducted within and between MPC, MC and SWEET/MtN3 family members. MPC protein relations to the MC and SWEET proteins were screened using the Protocol2 program and the Targeted Smith-Waterman Search (TSSearch) algorithm to generate comparison scores expressed
in standard deviations (S.D.) [12]. Using FASTA files, the TSSearch program generates a single local alignment for each subject and target pair for comparison [12].

High-scoring protein pairs were then analyzed by binary alignment using the Global Sequence Alignment Tool (GSAT) with 2,000 to 20,000 random shuffles. The program uses the Needleman-Wunsch algorithm to calculate accurate comparison scores expressed in S.D. [12]. Only scores of 12 S.D. or higher were considered statistically sufficient to argue in favor of homology.

Further evidence for familial relationships was obtained using the Multiple Em for Motif Elicitation (MEME) program for motif-based sequence analysis [18]. The program was used to identify areas with the greatest amino acid conservation within and between families. A list of 10 prototypical MPC and SWEET/MtN3 family proteins was used to identify common motifs (Table 1).

Proteins can be shown to be homologous by taking advantage of the Superfamily Principle transitivity rule [19]. This principle states that if protein A can be shown to be related to protein B, and protein B can be shown to be related to protein C, then proteins A and C must be related. In this study, proteins from the distinct groups (MPC, protein A; and SWEET/MtN3, protein D) were used to retrieve homologs (proteins B and C, respectively) for sequence similarity comparisons (Figure S1).
RESULTS

MPC Topology

The 3 yeast MPC paralogs, MPC 1, 2 and 3, were each predicted to have 3 transmembrane α-helical spanners (TMSs) [1], yet it remained unclear whether all proteins of the family share the same topology. Using the WHAT and HMMTop programs on individual proteins gave inconclusive TMS predictions ranging from 0-3. The problem was that MPCs are relatively small, ranging in size between ~90 and 140 aas, and the putative TMSs exhibit relatively weak degrees of hydrophobicity. In order to better identify the actual number of TMSs in homologues, a larger sample size is required. As aforementioned in the Methods section, a list of 433 homologues of human MPC2 was generated using Protocol1 and aligned with ClustalX. The AveHAS program was used to analyze and average the hydropathy values. The graphical display revealed 3 clear peaks of hydropathy (Figure 1A). Each peak displayed a high degree of similarity, suggesting that most MPC proteins have a 3 TMS topology. Interestingly, the first TMS contains the highest degree of conservation but is less hydrophobic.

Investigation of possible relationships with Mitochondrial Carriers

Members of the Mitochondrial Carrier (MC) family (TC# 2.A.29) have long been thought to be responsible for the transport of pyruvate into mitochondria [20, 21]. The MC family remains partially uncharacterized, and the possibility existed that MPC family members could be half sized members of the MC family. A potential association was
explored using the Protocol2 program to compare 196 MC homologs to 433 MPC proteins [12]. The top matches were evaluated in GSAT, and the best comparison scores obtained were 6.5, 6.1 and 3.1 S.D. (Figure 2), far below the values required to establish homology. The low scores reflected poor alignments of the observed transmembrane segments, requiring the inclusion of many gaps. Thus, we could not obtain statistically significant evidence to suggest that these two protein families are related. In fact, the evolutionary pathways taken by members of these two families differ (see Discussion). However these experiments provided an excellent control for the homology studies reported below.

**MPCs belong to the TOG Superfamily**

Attempts to expand the MPC family using NCBI’s Psi-BLAST methods yielded homologues suggesting a relationship to the bidirectional sugar transporter SWEET family (TC# 2.A.123; PF03083). Homologous sequence acquisition with the Protocol1 program resulted in 503 SWEET (MtN3 and SWEET-like) proteins that were subsequently used for comparative studies. The number of TMSs observed for the SWEET family proteins was estimated using the AveHAS program. Results showed a sample of 14 eukaryotic SWEET homologs containing 7 TMSs (Figure 1B) while a sample of 4 prokaryotic SWEET homologs showed 3 TMSs (Figure 1C). Interestingly, the first TMS of the prokaryotic 3 TMS SWEET homologs shows the highest degree of identity as is true for MPC family homologs.

A relationship was explored using the Protocol2 program to establish closely related members of the MPC and SWEET families. Comparison scores were generated
from binary alignments for the best scoring protein pairs using the GSAT program with
20,000 random shuffles. Binary alignment for TMSs 2-3 of MPC homologs compared to
TMSs 6-7 of SWEET homologs gave the best scores of 15.8 and 15.2 standard deviations
(Figure 3).

Further evidence for homology was obtained by analyzing a common motif
shared by members of the SWEET and MPC families. A multiple alignment was
generated using the ClustalX program with 10 sequences which contained five
representative 7 TMS SWEET family homologs and five representative MPC homologs.
The multiple alignment displayed a region of 82 residues with striking sequence
similarity and only 2 introduced gaps. Within this alignment, a cluster of 41 residues
showed the strongest degree of similarity with no introduced gaps (Figure 4B&C).
These positions included TMSs 2-3 of the MPC family and 6-7 of the SWEET family.
Confirmation was obtained via an independent motif sequence analysis on the same 10
sequences using the MEME program. A single 41-residues motif was discovered that
was shared among all 10 sequences and matched the region of highest similarity observed
within the multiple alignment (Figure 4A).

The 3 MPC TMSs in relation to the corresponding ancestral TOG 4 TMS repeat
unit was next examined (Figure 5). The 3 TMS SWEET family members arose by the
loss of the first TMS, resulting in TMSs 1-3 aligning with TMSs 2-4 of the ancestral
repeat unit. To explore the relationship between the MPC homologs and the 3 TMS
SWEET homologs, binary alignments were generated with the GSAT program. The two
alignments that gave the best comparison scores yielded only 8 S.D. Homology between
the 7 TMS SWEET and the 3 TMS SWEET family proteins was established when
comparing TMSs 5-7 and 2-4, respectively. A binary alignment using the program GSAT between a 3 TMS SWEET protein homolog (gi 575458051, Gemmatimonadetes bacterium) and a 7 TMS SWEET protein homolog (gi 470122975, Fragaria vesca) gave a comparison score of 33 standard deviations. With the evidence of TMSs 1-3 of the MPC family corresponding to 5-7 of the 7 TMS SWEET families and the comparison scores obtained for the alignments shown in Figure 3, it could be deduced that the 3 TMSs of MPC proteins are homologous to TMSs 2-4 of the TOG superfamily repeat unit. The relationships proposed for the TOG superfamily repeat unit, the SWEET family protein structure and the MPC 3 TMS proteins are presented schematically in Figure 5.

Phylogenetic Tree of Representative SWEET and MPC Homologues

With homology established between members of the SWEET and MPC families, a phylogenetic was drawn using the representative proteins listed in Table 1. The tree shows the MPC proteins clustering together, separately from the clustered SWEET proteins as expected (Figure 6).

Investigation of MPC Fusion Proteins

Within the list of MPC homologs obtained via NCBI’s BLAST and Pfam’s clan system, a few sequences were found to be longer than the expected range of 90-130 residues found in most MPC proteins. These longer proteins were examined with Pfam’s architecture search, which revealed 14 sequences of interest that contained 2 MPC domains. Among the 14 sequences of interest, 8 were identified as being artificial, caused by probable sequence errors. Two proteins from Arabidopsis thaliana both
contained 4 TMSs with an appearance of an MPC2 domain plus one extra predicted TMS within the N-terminal domain. Upon analysis, the extra TMS spliced in at residue 51-69 was a published result of an isoform due to alternative splicing of MPC3 with an unknown function.

The remaining 4 sequences in the list contained an appearance of possible fusion proteins (Table 2). Two protein sequences from *Perkinsus marinus*, Pma1 (gi 294954170) and Pma2 (gi 294954172) showed the same topology of 6 predicted TMSs regions with a 50 residue hydrophilic connection between the two MPC conserved domains (Figure 7). It was considered that this protein arose from a recent evolutionary duplication of the 3-TMS MPC protein. To test this hypothesis, Pma1 was split into its two halves to examine each domain independently. BLAST searches of the N-terminal domain brought up MPC1 homologs while the C-terminal domain gave MPC2 homologs as the top hits. A binary alignment between the N-terminal Pma1 domain and the yeast MPC1 (gi 1723875) gave a comparison score of 32.6 S.D. (Figure 8A). As a control the N-terminal domain was also compared to the yeast MPC2 (gi 731736), which gave a Z-score of only 15.2 S.D. A binary alignment of the C-terminal domain was compared to yeast MPC2 giving a comparison score of 32.4 S.D. with a control comparison to yeast MPC1 giving a comparison score of 11.5 S.D (Figure 8B). The other sequence (Pma2) showed a similar topology and the same domains (Figure S2). Thus the 6-TMS MPC protein arose not by a duplication of single MPC isoform but by fusion of MPC1 and MPC2 isoforms. This result is not entirely surprising as MPC1 and MPC2 are components of a functional heterodimer [1]. The proposed pathway for the appearance of this fusion is presented in Figure 9A which contrasts this pathway with the dissimilar
pathway taken for the appearance of most 7-TMS TOG superfamily proteins (Figure 9B).

The other 2 similar fusional events are seen in Aan1 (gi 676392207, *Aureococcus anophagefferens*) and Pso1 (gi 695429525, *Phytophthora sojae*). Here, however, the results are not as statistically separated. In the 6 predicted TMSs of Aan1, the 3 TMSs of the N-terminal Domain is most identified as MPC1-like while the final 3 TMS of the C-terminal Domain is most identified as MPC3-like. The MPC1-like similarity of the N-terminal domain is given by a comparison score of 19.0 S.D. when conducting a binary alignment with yeast MPC1 (Figure 10A). As control, the binary alignment with yeast MPC2 resulted in 14.5 S.D. The MPC2-like similarity of the C-terminal domain is given by a comparison score of 22.0 S.D. with yeast MPC2 while the control with yeast MPC1 gave a comparison score of 11.6 S.D. (Figure 10B). Finally the results for Pso1 showed statistically insignificant scores under 12 S.D. for the split domains for both MPC1 and MPC2 homologs.
DISCUSSION

In this report we analyze members of the recently discovered Mitochondrial Pyruvate Carrier (MPC) family. These proteins are found exclusively in eukaryotes, primarily within mitochondria. The majority has a simple 3 TMS topology that falls into two groups, MPC1 and MPC2, both of which are present in most organisms. A third parologue, MPC3, in *Saccharomyces cerevisiae* and other yeasts most closely resembles MPC2 and is probably derived from a common ancestor of these two paralogues which separated from MPC1 by a more recent gene duplication event that gave rise to MPC2 and MPC3.

A complete MPC is believed to consist of both MPC1 and MPC2, suggesting that these systems in most organisms are heteromultimer. We proposed that one or more MPC1(s) plus a one or more MPC2(s) comprise the basic transporter unit. This postulate is consistent with the observation that a few of the homologues are apparent fusion proteins containing an N-terminal MPC1-like domain a C-terminal MPC2 like domain.

Previous discussion about MPC’s origins and familial relations postulated a possible connection to the mitochondrial carrier (MC) family of proteins [12]. Using available tools to compare their sequences, however, did not result in statistically significant comparison scores. In fact, comparisons with the MC family were used as an excellent control for comparisons to members of the SWEET family, which did prove to be homologous to MPCs.

Comparison of MPC proteins with SWEET proteins revealed that these two families are related as suggested previously [22]. Because the SWEET family is a constituent of the large and functionally diverse TOG superfamily [9], we can propose
the pathway by which the MPCs evolved. This proposal is present in Figure 5. The TOG superfamily is believed to have arisen from a 4-TMS precursor that duplicated into an 8-TMS precursor and subsequent derivatives evolved through loss or gain of TMSs [9]. In the SWEET family, the N-terminal TMS of the 8 TMS precursor protein is missing in both the 7 and 3 TMS varieties. Through the observed high degrees of sequence relatedness found within TMS regions of MPC versus SWEET alignments, we obtain good statistical evidence to support the conclusion that MPC also evolved by a loss of the N-terminal TMS region from the 4-TMS evolutionary precursor. This documented relatedness further implies that both SWEET, particularly in the 3-TMS form, and MPC may transport using similar or identical mechanisms. Since high-resolution 3-d structural data on 3-TMS SWEET family proteins provided convincing evidence for a carrier mechanism, as opposed to a simple channel mechanism [22], the same can be concluded for the members of the MPC family. Existence of the 3-TMS MPC as a heterodimer in its functional form may also imply that the carrier properties implied for 3- and 7-TMS SWEET proteins and some of their larger TOG superfamily relatives may apply to both 3-TMS MPC proteins and 6-TMS fusion proteins. Further studies will be required to confirm these postulates.

The appearance of MPC1 and MPC2 fusional events display a novel pathway to 6-TMS proteins in the TOG superfamily. These events appear to have occurred after the evolution of the 3-TMS MPC proteins. The proposed pathway demonstrated in Figure 9A presents the 4-TMS TOG precursor protein losing the N-terminal TMS to form the 3-TMS MPC protein. Further sequence divergence resulted in the development of the isoforms, MPC1 and MPC2, fusing in some species to form a 6-TMS protein to replace
the heterodimer transport unit present in most organisms. Though few examples have been found thus far, their discovery strengthens the possibility that there may be more fusion events, which may help to elucidate the unknown associated proteins that make up the fully functional 150 kDa MPC protein complex.

Though many of the exact transport details of the MPC complex remain predominantly an enigma, many mysteries are coming to light from their sequence data. Association of the MPC family to the TOG superfamily provides the framework for its evolution. The prokaryotic proteins of the SWEET family may provide the essential comparative link to the prokaryotic evolution of the MPC family, allowing for future directions of study.
Figure 1. Average hydropathy, amphipathicity and similarity (AveHAS) plots based on ClustalX alignments for MPC, 7 TMS SWEET and 3 TMS SWEET families. Average hydropathy (dark lines, top), amphipathicity (light lines, top) and similarity (dotted lines, bottom) plotted for members of the MPC and SWEET families generated using the AveHAS program. Vertical lines (bottom) indicate positions of the TMSs. (A) Results for 270 eukaryotic MPC family homologs showing 3 TMSs. The conserved MPC domain (PF003650) contains all 3 TMSs at positions 90-180. (B) Results for 14 eukaryotic SWEET family homologs showing 7 TMSs. (C) Results for 4 prokaryotic SWEET family homologs showing 3 TMSs. Note the similarities observed between A and C, showing that in both sets of proteins, TMS1 is best conserved but is less hydrophobic.
Figure 2. Sequence alignments do not provide evidence for homology between MPC and MC homologs. (A) GSAT alignment (2000 random shuffles) of TMS 1-3 of MPC homolog Dpe2 (*Drosophila persimilis*, gi 195147502; 3 TMS) with TMS 1 of MC homolog Ssp1 (*Salpingoeca* sp. ATCC 50818, gi 514700572; 5 TMS) gave a comparison score of 6.5 SD with sequences showing 25.5 % identity and 41.6 % similarity. (B) GSAT alignment (2000 random shuffles) of TMS 1-3 of MPC homolog Mpu1 (*Mustela putorius furo*, gi 511866909; 3 TMS) with TMS 1 of MC homolog Pte3 (*Paramecium tetraurelia*, gi 145496860; 6 TMS) gave a comparison score of 6.1 SD with sequences showing 19.2 % identity and 37.7 % similarity. (C) GSAT alignment (2000 random shuffles) of TMS 1-3 of MPC homolog Pkn1 (*Plasmodium knowlesi* gi 221059027; 3 TMS) with TMS 1-3 of MC homolog Hph1 (*Hanusia phi*, gi 157168250; 4 TMS) gave a comparison score of 3.1 with sequences showing 22.2 % identity and 41.4 % similarity. These three alignments gave the best scores of the alignments examined.
A

Dpe2 10 KLYNAIILSVDKIVPQAQPLQSPAGPKTVF\ vffwafpkwslvaagldt 60
   || : : :::: | | :: | : : : : | : |
Ssp1  1 MLYNQVNIR-ERL----AA------APAGK:
   | | : : : : | | | | : | : | : |
Dpe2 61 LSRPAQNISL---NQCGSLAANTGLIWSRYS
Ssp1  41 AAAPIERIKLLLQNQ-GEMLKLTGSLDRPSGGM-DCFGRVIKSEG
Dpe2 107 QSFLVVKHLRAWNEKARNTVFKHPYPI--WSGD-DW 140
   | : | : | | | : | : | : |
Ssp1  88 WRGNAANVLRYFTPTQLANFPAFDFIKALINF

B

Mpu1 1 MAMVAALWRSARDYMKTKEFRNYLASTHFWGFPVANYGL
   : | : : : | | : | : | : |
Pte3  29 NSLVKE--ESKITIINSINMKKWHQFRIKGNQ
       : | : | : | : | | : |
Mpu1 51 EIISGRMKVALI-FYLM\ MHFYARVQPR----NLLLFVCH-STNVLAR
   : | : | : | : | : | : |
Pte3  78 MQEGWKRIKIAFFQGFKNKQCKCNKIEGNKRI
       : | | : : | : | : | |
Mpu1 94 SVQLSRYLNYYC\ CGGASAASAVGTTAATPDPSTCPVPSVDLDNDDC
   : : : : | | : | : |
Pte3 128 FIQMADFLRDFLIGGVSAAVSKTAVAPIERK
       : : : : : : | |

C

Pkn1 1 MGLIKSIFYPNIIIPRIKNEIK---GYNI\ QSVKKAVSDTGILTIH-----43
   | : | | : | : : | : : |
Hph1 97 FNFAFKDFFKSIFPKYNYQKTQFPP\ FGVNMA-SGGLAGAASLLIVY
     | : | | : | | | : |
Pkn1 44 -----FWAPTFKWSISL-LN-IVD-----INRDP\ KLLSLPOQQFAICMTG
     | | | : | : : | : |
Hph1 146 ARTRLGADVGKSGREFMNGLVDCVMKTA-
     : : | : | : | | | |
Pkn1 85 RFA\ Y------VI\ KPNYO\ LTTINF\ IMSLTA\ LYYA\N\ K
    | | : | : | : | : | |
Hph1 197 RGAYFGFYSDAVGVFK\ P---VLT-F\ KLFIA-Q\ A\ VASSGIAS\ PFDTVRR
     | : | : | : | : | : |
Pkn1 126 NGGIGEKKEGAQ 137
     : : | |
Hph1 244 RLMM---QSGAK 250
Figure 3. Binary sequence alignments between MPC family homologues and SWEET family homologues. (A) GSAT alignment (20,000 random shuffles) of TMS 2-3 of MPC homolog Spo1 (*Schizosaccharomyces pombe*, gi 19113758; 3 TMS) with TMS 6-7 of the SWEET16 homolog Tur21 (*Triticum urartu*, gi 474116864; 7 TMS) gave a comparison score of 15.8 SD with sequences showing 31.9 % identity and 51.4 % similarity. (B) GSAT alignment (20,000 random shuffles) of TMS 2-3 of MPC homolog Ptr1 (*Pyrenophora tritici-repentis*, gi 189199196; 3 TMS) with TMS 6-7 of SWEET homolog Vvi10 (*Vitis vinifera*, gi 147815543; 7 TMS) gave a comparison score of 15.2 SD with the sequences showing 28.4 % identity and 51.4 % similarity.
Figure 4. Evidence for a conserved sequence motif shared between MPC and SWEET families. (A) The 41-residue most conserved motif shared by the MPC and SWEET families as revealed using the MEME program. The statistical score for the MEME motif shown was e^{-109} (see http://meme.nbcr.net). (B) A multiple alignment created using the ClustalX program showing the corresponding conserved region. The alignment was generated with the top five sequences being MPCs (MPC3, gi 1723760; MPC2, gi 731736; and 3 proteins with PF03650 domains, gi 19113758, gi 406861420, gi 169609298) as well as the five bottom sequence being SWEET homologues (proteins containing the MtN3 domain, gi 388501750; gi 75206789; gi 474116864; gi 75220431; gi 514809011). (C) Histogram depicting the relative level of identity at each residue position.
Figure 5. Schematic depiction showing the proposed evolutionary relationships between eukaryotic MPC family members (top), eukaryotic or prokaryotic 7 TMS SWEET family members (middle) and prokaryotic 3 TMS SWEET family members (bottom). The 4 TMS ancestral precursor lost the first TMS to form the 3 TMS MPC and SWEET family members. The 7 TMS SWEET family proteins evolved by duplication of the 4 TMS repeat unit followed by loss of the first TMS [9]. The common 3 TMSs are shown aligned (gray box) with the corresponding best binary alignment comparison scores expressed in standard deviations (SD) as determined using the GSAT program with 20,000 random shuffles.
Figure 6. A phylogenetic tree based off multiple alignments depicting SWEET and MPC family separation. The tree was generated using the ClustalX/TreeView program for the ten prototypical proteins, 5 MPC and 5 SWEET, listed in Table 1.
Figure 7. Average hydropathy and amphipathicity plots with predicted TMSs and domains for fusion protein Pma1. Plots of hydropathy (dark line) and amphipathicity (light line) for Pma1 (gi 294954170, Perkinsus marinus) generated by the WHAT program. The vertical bars indicate presence of a TMS as predicted by the program HMMTop. The N-terminal 3 TMS domain is MPC1-like and the C-terminal 3 TMS domain is MPC2-like as indicated in Figure 8.
Figure 8. Binary alignments showing N-terminal MPC1 and C-terminal MPC2 domains in the fusion MPC protein, Pma1. Binary sequence alignments generated between the N- and C-terminal domains of Pma1 (gi 294954170, Perkinsus marinus) and MPC proteins in yeast. (A) GSAT alignment (20,000 random shuffles) of TMS 1-2-3 of the N-terminal domain of Pma1 with TMS 1-2-3 of the yeast MPC1 homolog (gi 330443578, Saccharomyces cerevisiae) gave a comparison score of 33.4 S.D. A control between TMS 1-2-3 of Pma1 with TMS 1-2-3 of the yeast MPC2 homolog (gi 330443590, Saccharomyces cerevisiae) gave a comparison score of 15.2 S.D. (not shown). (B) GSAT alignment (20,000 random shuffles) of TMS 4-5-6 of the C-terminal domain of Pma1 with TMS 1-2-3 of the yeast MPC2 homolog gave a comparison score of 32.4 S.D. A control between TMS 4-5-6 of Pma1 with TMS 1-2-3 of the yeast homolog MPC1 gave a comparison score of 11.5 S.D. (not shown).
Figure 9. Schematic depiction showing the proposed evolutionary pathway for 6-TMS MPC fusion proteins and the major pathway for most TOG superfamily proteins. (A) The TOG superfamily 4-TMS evolutionary precursor lost the N-terminal TMS to give rise to the 3-TMS MPC family. Two isoforms, MPC1 and MPC2 fused to form the observed 6-TMS MPC fusion proteins consisting of with MPC1 TMSs 2-3-4 in the N-terminal domain and MPC2 TMSs 2-3-4 in the C-terminal domain. (B) The TOG superfamily 4-TMS evolutionary precursor duplicated to form the 8-TMS precursor followed by subsequent loss of the N-terminal TMS (7-TMS SWEET family) or C-terminal TMS (7-TMS NiCoT family). A later loss of the N-terminal TMS in the 7-TMS NiCoT family gave rise to the 6-TMS NiCoT family of dissimilar makeup then the 6-TMS MPC fusion proteins.
A

(1 2 3 4) \rightarrow \begin{array}{c}
2 3 4 \\
3-TMS MPC
\end{array}
\rightarrow \begin{array}{c}
2 3 4 2 3 4 \\
6-TMS MPC Fusion
\end{array}

TOG Superfamily 4-TMS
Evolutionary Precursor

B

(1 2 3 4) \rightarrow \begin{array}{c}
1 2 3 4 1 2 3 4 \\
7-TMS TOG Superfamily (SWEET)
\end{array}
\rightarrow \begin{array}{c}
2 3 4 1 2 3 4 \\
7-TMS TOG Superfamily (NiCoT)
\end{array}
\rightarrow \begin{array}{c}
2 3 4 1 2 3 4 \\
6-TMS TOG Superfamily (NiCoT)
\end{array}

TOG Superfamily 4-TMS
Evolutionary Precursor
Figure 10. Binary alignments displaying N-terminal MPC1 and C-terminal MPC2 domains in the fusion MPC protein, Aan1. Binary sequence alignments generated between the N- and C-terminal domains of Aan1 (gi 676392207, *Aureococcus anophagefferens*) and MPC1 and MPC2 proteins in yeast. (A) GSAT alignment (20,000 random shuffles) of TMS 1-2-3 of Aan1 with TMS 1-2-3 of yeast MPC1 homolog (gi 330443578, *Saccharomyces cerevisiae*) gave a comparison score of 19.0 S.D. compared to control of yeast MPC2 homolog (gi 330443590, *Saccharomyces cerevisiae*) of 14.5 S.D. (not shown). (B) GSAT alignment (20,000 random shuffles) of TMS 4-5-6 of Aan1 with TMS 1-2-3 of yeast MPC2 homolog gave a comparison score of 22.0 S.D. compared to a control with the yeast MPC1 homolog which gave a comparison score of 11.6 S.D. (not shown).
### Table 1.

Representative MPC and SWEET/MtN3 family proteins used in the MEME common motif analysis (Figure 4A), the multiple alignment (Figure 4B) and the phylogenetic tree (Figure 5). Proteins are listed with their GenBank identification number (GI #), protein family type, predicted topology, species, size, and abbreviation.

<table>
<thead>
<tr>
<th>GI #</th>
<th>Family</th>
<th># TMSs</th>
<th>Species</th>
<th>Size</th>
<th>Abb’n</th>
</tr>
</thead>
<tbody>
<tr>
<td>1723760</td>
<td>MPC</td>
<td>3</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>146</td>
<td>Sce1</td>
</tr>
<tr>
<td>731736</td>
<td>MPC</td>
<td>3</td>
<td><em>Saccharomyces cerevisiae</em></td>
<td>129</td>
<td>Sce2</td>
</tr>
<tr>
<td>19113758</td>
<td>MPC</td>
<td>3</td>
<td><em>Schizosaccharomyces pombe</em></td>
<td>118</td>
<td>Spo1</td>
</tr>
<tr>
<td>406861420</td>
<td>MPC</td>
<td>3</td>
<td><em>Marssonina brunnea</em></td>
<td>189</td>
<td>Mbr1</td>
</tr>
<tr>
<td>169609298</td>
<td>MPC</td>
<td>3</td>
<td><em>Phaeosphaeria nodorum</em></td>
<td>161</td>
<td>Pno1</td>
</tr>
<tr>
<td>388501750</td>
<td>SWEET/MtN3</td>
<td>3</td>
<td><em>Medicago truncatula</em></td>
<td>147</td>
<td>Mtr1</td>
</tr>
<tr>
<td>75206789</td>
<td>SWEET/MtN3</td>
<td>7</td>
<td><em>Arabidopsis thaliana</em></td>
<td>289</td>
<td>Ath1</td>
</tr>
<tr>
<td>474116864</td>
<td>SWEET/MtN3</td>
<td>7</td>
<td><em>Triticum urartu</em></td>
<td>351</td>
<td>Tur1</td>
</tr>
<tr>
<td>75220431</td>
<td>SWEET/MtN3</td>
<td>7</td>
<td><em>Medicago truncatula</em></td>
<td>268</td>
<td>Mtr2</td>
</tr>
<tr>
<td>514809011</td>
<td>SWEET/MtN3</td>
<td>7</td>
<td><em>Setaria italica</em></td>
<td>296</td>
<td>Sit1</td>
</tr>
</tbody>
</table>

1. Size is expressed in numbers of amino acyl residues
2. The abbreviations (Abb’n) are those used in Figures 4 and 6

### Table 2.

List of possible MPC fusion proteins used in TMS topology analysis (Figure 7 and Figure 9) and domain identifications (Figure 9 and Figure 10). Proteins are listed with their GenBank identification number (GI #), species, predicted topology, size, and abbreviation.

<table>
<thead>
<tr>
<th>GI#</th>
<th>Species</th>
<th>TMSs</th>
<th>Size</th>
<th>Abb’n</th>
</tr>
</thead>
<tbody>
<tr>
<td>294954170</td>
<td><em>Perkinsus marinus</em> ATCC 50983</td>
<td>6</td>
<td>267</td>
<td>Pma1</td>
</tr>
<tr>
<td>294954172</td>
<td><em>Perkinsus marinus</em> ATCC 50983</td>
<td>6</td>
<td>349</td>
<td>Pma2</td>
</tr>
<tr>
<td>676392207</td>
<td><em>Aureococcus anophagefferens</em></td>
<td>6</td>
<td>268</td>
<td>Aan1</td>
</tr>
<tr>
<td>695429525</td>
<td><em>Phytophthora sojae</em></td>
<td>6</td>
<td>319</td>
<td>Pso1</td>
</tr>
</tbody>
</table>

1. Size is expressed in numbers of amino acyl residues
2. The abbreviations (Abb’n) are those used in Figure 7, 8, 9 and 10
Figure S1. Flowchart schematic showing the Superfamily Principle transitivity rule for establishing homology of distant protein families. Protein A can be shown to be related to Protein D by using Protocol1 (PSI-BLAST) to find homologous proteins Protein B and Protein C, respectively. Protocol2 (GSAT) gives comparison scores in standard deviations (S.D.) between Protein B and Protein C to establish homology. A sequence alignment with 60 aas or greater that gives 12 S.D. is required to establish protein homology.
Figure S2. Average residue hydropathy and amphipathicity characteristics with predicted TMSs and domains for fusion protein Pma2. Plots of hydropathy (dark line) and amphipathicity (light line) for Pma2 (gi 294954172, Perkinsus marinus) generated by the WHAT program. The vertical bars indicate presence of a TMS as predicted by the program HMMTop. The N-terminal 3 TMS domain is MPC1-like and the C-terminal 3 TMS domain is MPC2-like as indicated by GSAT alignment comparison scores.
REFERENCES


8. Halestrap AP: The mitochondrial pyruvate carrier: has it been unearthed at last? Cell Metab 2012, 16(2):141-143.


