Analyses of frequent and conserved intron positions shed light on the evolution of the mitochondrial carrier family SLC25

Magnus Monné[1,2,*], Antonia Cianciulli[1], Maria A. Panaro[1], Rosa Calvello[1], Anna De Grassi[1], Luigi Palmieri[1,3], Vincenzo Mitolo[1] and Ferdinando Palmieri[1,3,*]

[1]Department of Biosciences, Biotechnology and Environment, University of Bari, Bari, Italy. [2]Department of Sciences, University of Basilicata, Potenza, Italy. [3]CNR Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), Bari, Italy. *Corresponding authors: Ferdinando Palmieri (ferdpalmieri@gmail.com) and Magnus Monné (magnus.monne@unibas.it).

ABSTRACT. **Mitochondrial carriers (MCs) constitute a family of mostly mitochondrial proteins that transport different specific substrates, such as cofactors, nucleotides, amino acids, dicaboxylates and inorganic anions, across the inner membrane. MCs have characteristic triplicated protein sequence repeats that are reflected in the three-fold symmetrical structure of their six-transmembrane α-helical transporter domain. These common sequence features have been used to identify MC genes in various eukaryotic genomes. We have mapped and analyzed the positions of the introns in MCs of highly diversified organisms. The results show that many MCs have introns at the same specific positions within the MC transporter domain and that several of these positions are three-fold symmetric. Moreover, many of these frequently occurring intron positions are particularly common in orthologs of specific MC subfamilies, which transport similar substrates. These findings imply that the present day MCs have partially conserved the gene architectures of ancestral MCs. Based on this reasoning the frequent and conserved intron positions were used to reconstruct a phylogenetic tree that also included evolutionary relationships between distant MC homologs with low sequence similarities. Furthermore, the structural locations of the intron positions suggest that exon shuffling and intron sliding may have contributed to the substrate specificity diversification in the evolution of the MC family.**

RESULTS 1.

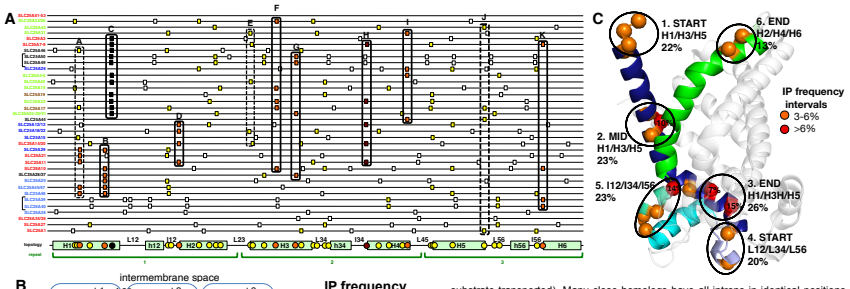

Fig 1. The intron positions (IPs) in the protein sequences of the 53 human SLC25 family members were analyzed in:

A) the primary structure. The IPs (small rectangles coloured according to their frequencies among the human MCs) were mapped onto a sequence alignment of the 53 human MCs (represented as horizontal lines). The names of the MCs are coloured according to their type of substrate: green for nucleotides, red for carboxylates and blue for amino acids, which is based on their contact point II residues (two residues on helix H4 appearing to covariate with the class of substrate transported). Many close homologs have all introns in identical positions and are therefore presented as one MC isoform representative, for example SLC25A4-6 (89-93% sequence identity). The most frequent IPs were found in eleven positions (named A-K) and especially IPs B, C and D, which are mutually exclusive, appear to divide the majority of MCs into three groups. **Conclusion:** many of the frequent IPs seem to be present and combined in specific subsets of MCs.

B) the secondary structure. The frequent IPs (coloured by frequency intervals) of all human MCs were mapped onto the topology model for MCs, which consists of the six transmembrane α-helices H1-H6 and connecting segments oriented in the innermembrane of mitochondria. **Conclusion:** many of the frequent IPs are found in exact or similar positions in the about 100 residue three-fold MC repeat sequences of MCs, at the extremes or centrally in the transmembrane α-helices.

C) the tertiary structure. The IPs of the three repeats of all human MCs (frequency percentage intervals indicated) were mapped on the first repeat of the structure of the bovine ADP/ATP carrier, in which the six transmembrane α-helices form a three-fold pseudosymmetric bundle confining a central pore for substrate binding and translocation. **Conclusion:** IPs are frequently aggregated in six locations (encircled) in the 3D-structure of MCs and their specific positions are often three-fold symmetry related.

RESULTS 2.

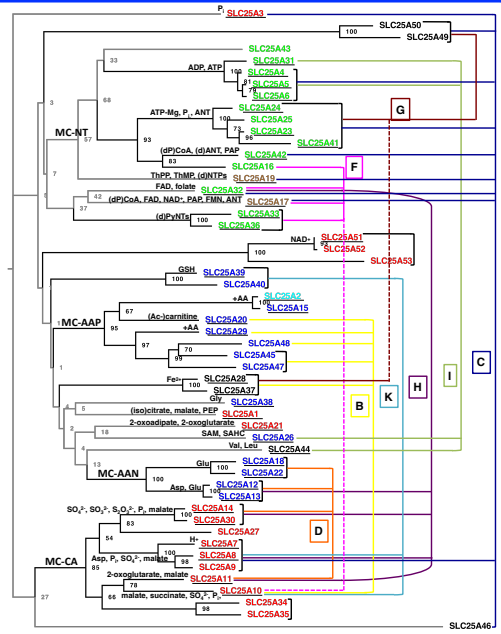

Fig. 2 The frequent IPs were mapped onto a phylogenetic tree of the human MCs. The phylogenetic tree of the 53 human MCs was constructed by using PhyML v3.1 from a multiple-sequence alignment with ClustalO in Seaview4 and drawn in FigTree v1.4.2. Bootstrap values for 1000 replicates are reported on the nodes, that are connected by lines indicating sequences exhibiting less than 30% sequence identity (in grey) and at least 30% sequence identity (in black). The names of MCs are coloured according to their type of substrate (green for nucleotides, red for carboxylates and blue for amino acids) and those with known substrates are underlined (their main substrates indicated at the nodes). Four main phylogenetic clusters of MCs had significant bootstrap values and are indicated: MCs for nucleotides (MC-NT), positively charged amino acids (MC-AAP), negatively charged amino acids (MC-AAN) and carboxylates (MC-CA). The MCs that share the frequent IPs B, C, D, F, G, H, I and K are linked together by coloured lines on the right. Abbreviations: +AA, positively charged amino acids; (d)ANT, (deoxy-)adenine nucleotides; (d)NTPs, (deoxy-)nucleotides; (dP)CoA, (dephospho-)coenzyme A; (d)PyNTs, (deoxy-)pyrimidine nucleotides; GSH, glutathione; PAP, adenosine 3',5'-diphosphate; PEP, phosphoenolpyruvate; P_i, phosphate; SAHC, S-adenosylhomocysteine; SAM, S-adenosylmethionine; ThMP, thiamine monophosphate; ThPP, thiamine pyrophosphate. **Conclusion:** the frequent IPs B, C, D and F appear especially associated with MCs for positively charged amino acids, nucleotides, carboxylates/negatively charged amino acids and nucleotide/adenine nucleotide cofactors, respectively.

RESULTS 3.

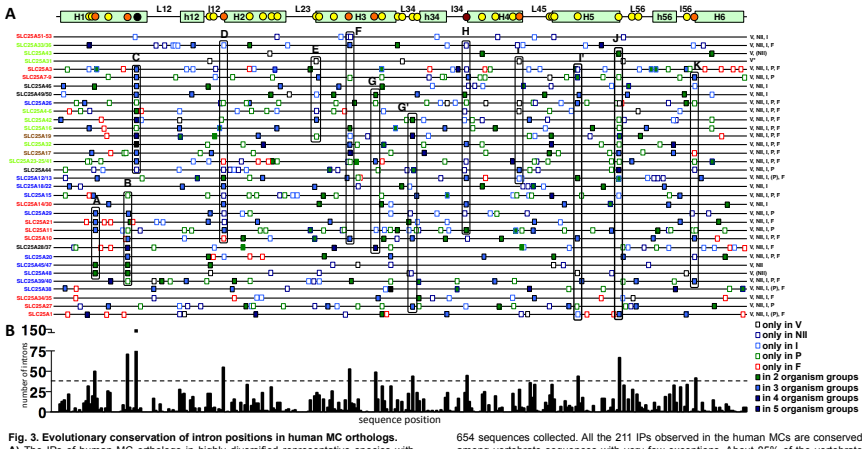

Fig. 3. Evolutionary conservation of intron positions in human MC orthologs.

A) The IPs of human MC orthologs in highly diversified representative species with confidently annotated genomes were mapped onto the human sequences to investigate their frequency and conservation. The MC orthologs included were from several species of five organism groups: vertebrates (**V**, *H. sapiens*, *Gallus gallus*, *Xenopus tropicalis* and *Danio rerio*), non-insect invertebrates (**NII**, *Caenorhabditis elegans*, *Aplysia californica*, *Ciona intestinalis* and *Acanthaster planci*), insects (**I**, *Drosophila melanogaster*, *Bemisia tabaci*, *Spodoptera frugiperda*, *Acromyrmex echinatior*, *Cryptotermes secundus* and *Dendroctonus ponderosae*), plants (**P**, *Arabidopsis thaliana*, *Oryza sativa*, *Physcomitrium patens* and *Chlamydomonas reinhardtii*) and fungi (**F**, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Neurospora crassa* and *Aspergillus fumigatus*). In total 2883 introns were found in the 654 sequences collected. All the 211 IPs observed in the human MCs are conserved among vertebrate sequences with very few exceptions. About 85% of the vertebrate IPs were found in at least one non-insect invertebrates, 57% in insects, 9% in plants and 3% in fungi. Moreover, additional IPs not found in vertebrates were identified exclusively in the other organism groups. **B**) The number of IPs in each position were counted in all MC orthologs. The dashed line indicates a threshold, which was chosen as four times higher (36) than the average number of introns (9) per position. Most of the frequent IPs observed in human MCs had high frequency also among their orthologs (except IPs E and I), and G' and I' could be added. **Conclusions**: i) the human IPs are absolutely conserved among vertebrates, ii) the majority of them are found in other animals, and iii) some of them are found in plants and fungi.

RESULTS 4.

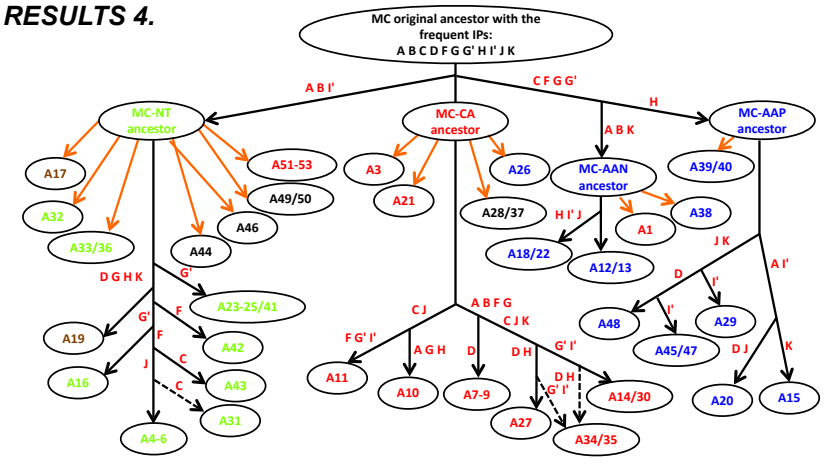

Fig. 4. Reconstructed phylogenetic tree based on the frequent and conserved IPs. The presence of the same frequent and conserved IPs in diverse sets of MCs from various clusters of the phylogenetic tree (**Figs. 1-3**) indicates that these IPs were present early on in evolution before the diversification of this family of proteins. Based on this observation, which is coherent with a scenario where the MC original ancestor contained the total or nearly total set of frequent IPs followed by subsequent partial loss of introns, a phylogenetic tree can be reconstructed through minimization of repeated intron-loss events in the branches. The nodes indicate the MC cluster ancestors and the ancestor for the SLCA25 subfamily orthologs (indicated with A and the specific SLCA25 number) of the nucleotide (green), carboxylate (red) and amino acid (blue) classes based on contact point II residues. Black arrows indicate loss of the frequent IPs indicated in red. Dashed black lines indicate alternative links. Orange arrows indicate massive loss of frequent IPs (not specified). The orange arrows originating from MC-CA ancestor could just as well have originated from the MC original ancestor. **Conclusion:** at variance with the phylogenetic tree (**Fig. 2**), this IP-reconstructed tree indicates that i) SLC25A44, SLC25A46, SLC25A49/50 and SLC25A51-53 are related to the nucleotide and nucleotide-associated carriers; ii) SLC25A3, SLC25A21, SLC25A26 and SLC25A28/37 are derived from the MC-CA cluster ancestor or the original MC ancestor; iii) SLC25A1 and SLC25A38 have come from the MC-AAN cluster ancestor; and iv) SLC25A39/40 is related to the MC-AAP cluster.

DISCUSSION.

In this study the positions of introns in the MC superfamily genes and their relationship to protein structure, function and evolution have been analyzed. The main findings are that many of the frequent and conserved IPs in MCs are found i) at specific positions centrally and towards the extremities of the transmembrane α-helices that are related to the three-fold symmetry of MC sequences and structure; ii) in specific subfamilies or groups of subfamilies with similar substrates or class of substrates. Based on these observations drawn from the IPs in MC sequences of various organisms living today, we would like to speculate on the gene architecture and the evolution of the MC superfamily members:

- Our results point towards the possibility that the ancestral MC gene contained all the frequent and conserved IPs (A, B, C, D, F, G, G', H, I', J and K), or at least IP D, H, J and K, which are found in genes of virtually all main MC clusters. This would mean that the architecture of the original MC gene is partially conserved in the present genes of MCs.

- The regular arrangement of the IPs in the partially conserved gene structure of MC genes suggests that the repositioning of IPs and exon shuffling might have contributed to the diversification of the substrate specificity in MC superfamily members.

- The fact that the four frequent and conserved IPs A, B, C and D are found mostly in the orthologs of MC subfamilies with similar substrates reflects evolutionary relationships that may help in the current attempts to identify the substrates of MCs with unknown transport function, especially of those that have very low sequence identity with any already biochemically characterized carrier or unclear contact point II residue classification.

- Future studies are needed to analyze the intron sequences at the frequent and conserved IPs of MCs to associate them with regulatory elements, potential functions and other genes.

Acknowledgements: Research in the authors' laboratories was supported by (i) the Italian Human ProteomeNet no. RBRN07BMCT_009 from the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), (ii) PRIN 2017 (2017PAB8EM) and PRIN 2020 (2020RRJP5L) grants (MIUR), and (iii) Centre of Excellence on Comparative Genomics (CEGBA).

The results resented in this poster have been published in: M. Monné, A. Cianciulli, M.A. Panaro, R. Calvello, A. De Grassi, L. Palmieri, V. Mitolo and F. Palmieri. New insights into the evolution and gene structure of the mitochondrial carrier family unveiled by analyzing the frequent and conserved intron positions. Mol Biol Evol. (2023) 40:msad051.