Human Fumarate Hydratase Is Dual Localized by an Alternative Transcription Initiation Mechanism

Ekaterina Dik1,†, Adi Naamati1,†, Hadar Asraf1, Norbert Lehming2 and Ophry Pines1,2,*

1Department of Microbiology Molecular Genetics, IMRIC, Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem, Israel
2CREATE-NUS-HUJ Program and the Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
*Corresponding author: Ophry Pines, ophry@ekmd.huji.ac.il

Abstract

Fumarate hydratase (FH, fumarase), is a tricarboxylic acid cycle enzyme localized in the mitochondrial matrix. However, a common theme, conserved from yeast to human, is the existence of a large cytosolic population of FH. FH has been shown to function as a tumor suppressor gene and is now implicated in various diseases. We have previously indicated that the cytosolic echoform of FH has a role in the DNA damage response and specifically in the response to DNA double strand breaks. In fact, recently FH has been shown to be involved in histone demethylation. Therefore, it has become important to understand the underlying mechanism of FH dual subcellular location in human cells. We revealed that in human cells, in contrast to yeast, the FH gene encodes two gene products, one containing and one lacking the mitochondrial targeting sequence. On the basis of expression of endogenous wild-type FH and mutant FH cDNAs from plasmids, RT-PCR, RACE to determine the 5′ termini of FH mRNAs, and mass spectrometry of FH products, we show that the mechanism of FH distribution is alternative transcription initiation from a broad promoter. This is contrary to the suggested mechanism for rat liver cells which had claimed alternative translation initiation.

Keywords dual targeting, fumarate hydratase, mitochondria, nucleus, transcription initiation, tumor suppressor

Received 12 November 2015, revised and accepted for publication 29 March 2016, uncorrected manuscript published online 1 April 2016, published online 6 May 2016

Eukaryotic cells are defined by the existence of membrane delimited subcellular compartments. Each compartment contains a unique set of proteins, which execute specific biological processes that are fundamental for the normal performance of the compartment. Thus, protein targeting to a specific location in the cell is an essential process for the proper function and maintenance of the whole cell. In eukaryotic cells, identical proteins can be localized to more than a single subcellular compartment, a phenomenon termed dual targeting. The two populations of the proteins are identical or almost identical and are termed echoforms (1).

Dual targeting can be achieved by various mechanisms. These mechanisms can be divided into two major subgroups, two and single translation product mechanisms. Two translation products can be obtained by two mRNAs from a single gene, either by alternative transcription initiation or by alternative splicing. In the case of a single mRNA, the two translation products can be achieved by alternative translation initiation or alternative translation termination (2,3). In the cases of single translation products, about 10 different molecular mechanisms have been described (3–7).

A common theme conserved throughout evolution is the dual localization of the enzyme fumarase, which in humans is designated FH (fumarate hydratase). FH is distributed
between mitochondria and the cytosol of all eukaryotes. In mitochondria FH is essential for cell respiration as part of the tricarboxylic acid (TCA) cycle. More than a decade ago FH was found to underlie a tumor susceptibility syndrome, hereditary leiomyomatosis and renal cell cancer (HLRCC). This syndrome is characterized by benign cutaneous and uterine leiomyomas, renal cell carcinomas and uterine leiomyosarcomas. A bi-allelic inactivation of FH has been detected in almost all HLRCC tumors, and therefore FH was suggested to function as a tumor suppressor (8,9). We discovered that extra-mitochondrial fumarase is required for protection from double strand breaks, as part of the DNA damage response (DDR), both in yeast and human (10). We proposed that the role of fumarase as a tumor suppressor gene is because of its activity within the DDR. More recently, Jiang et al. have shown that upon DNA damage, FH is required to inhibit histone demethylases which is crucial for efficient DNA damage repair (11). The histone demethylases are one example for a superfamily of enzymes called α-ketoglutarate-dependent dioxygenases. These enzymes utilize α-ketoglutarate as a co-substrate, and are involved in fatty acid metabolism, oxygen sensing, collagen biosynthesis and modulation of the epigenome. Fumaric acid is a competitive inhibitor of such enzymes and can potentially modulate different cellular processes (12,13). In addition, high concentrations of fumarate have been shown to cause succination, which is an irreversible non-enzymatic modification of cysteine residues by fumarate, to form S-(2-succinyl) cysteine (2SC). For instance, chronic succination of glutathione has been associated with persistent oxidative stress and cellular senescence (14). Additional proteins that have been shown to be succinated by fumaric acid are the Kelch-like ECH-associated protein 1 (KEAP1) and mitochondrial aconitase (ACO2) which can have profound effects on cellular metabolism (15–17). Finally, protein succination by fumaric acid has recently been suggested to be involved in development of type 2 diabetes. In this regard, succinated mitochondrial proteins have been suggested as biomarker for type 2 diabetes [(15,18,19) and Rorsman, unpublished data].

While the role of FH in health and disease has been the topic of numerous studies, the distribution mechanism of the enzyme in human cells remains undetermined. In rat liver, the two echoforms are encoded by a single gene and it has been claimed that it is transcribed into a single mRNA which contains two translation initiation codons. The two translation products presumably differ by the presence or absence of the mitochondrial targeting sequence (20–23). In the yeast *Saccharomyces cerevisiae*, fumarase is encoded by the FUM1 gene and is expressed as a single translation product which is distributed between the two compartments by a reverse translocation mechanism (24–29). According to this mechanism all the protein molecules start import into the mitochondria, during which the mitochondrial targeting sequence (MTS) is removed by the mitochondrial processing peptidase (MPP). After the removal of the MTS, a subgroup of the molecules moves back into the cytosol. The driving force of this process is the folding of the protein (1,6,26–28). In Arabidopsis, fumarase is encoded by two highly homologous genes which differ by one containing and one lacking an MTS (30).

In this study we first determined by a metabolic labeling assay that, unlike yeast fumarase, human FH distributes by a two translation product mechanism. Based on the analysis of FH distribution, when expressed from plasmids, we conclude that FH does not distribute by an alternative translation mechanism as suggested for rat liver. Finally, by identifying the 5’ ends of FH mRNAs and by measurements of transcript levels we provide evidence that FH distribution is determined at the level of transcription initiation, meaning that mitochondrial and cytosolic forms of the protein are translated from different mRNAs. This conclusion is also supported by identification of FH N-termini by mass spectrometry (MS) of wild-type and mutant FHs.

**Results**

**FH has two translation products**

In order to study the distribution mechanism of FH in human cells, HeLa and HEK293T cell lines were employed. First we examined whether FH is expressed as a single translation product, as in yeast, or as two translation products, as was suggested for rat liver. Analysis of the FH sequence suggests the existence of an MTS at the N terminus of the protein, which implies that this protein is a substrate of MPP and is processed upon import into mitochondria. We employed a metabolic labeling assay in the absence and presence of the mitochondrial membrane potential uncoupler carbonyl CCCP. Uncoupling of
Figure 1: Human FH has two translation products.
Metabolic labeling with [35S]-methionine of yeast and human (HeLa) cells was carried out in the absence or presence of CCCP (a mitochondrial membrane potential uncoupler which blocks mitochondrial protein import). Fumarase/FH was immunoprecipitated from cell extracts using fumarase/FH antiserum which was then analyzed using PAGE and autoradiography. P, precursor; M, mature.

mitochondrial membrane potential blocks protein import into the mitochondria and under these conditions we can distinguish between the mitochondrial precursor which contains the MTS from the mature processed form of the protein. In the absence of CCCP only a ‘mature’ sized form (lower arrow) of the protein can be detected, in both yeast and human cells (Figure 1, lanes 1 and 3, bottom panels). In the presence of CCCP we observe different results for yeast and human cells. In yeast, we detect only the precursor (upper arrow) that harbors the MTS which is consistent with the production of a single translation product and the reverse mechanism of distribution (Figure 1, lane 2, bottom panel). However, in human cells, in the presence of CCCP, we detect two products, a precursor and a mature sized protein (Figure 1, lane 4, bottom panel). This clearly suggests, as anticipated from the study in rat (20), a two translation product mechanism; one initially long product, harboring the MTS, constitutes the mitochondrial population of the protein and a second shorter product, lacking the localization signal which remains in the cytosol.

The FH distribution mechanism is not at the level of translation initiation
There are various mechanisms by which two translation products can be dual distributed in the cell. Alternative transcription initiation, will give rise to different transcripts, one with and one without the coding sequence for the MTS. Alternative splicing or trans-splicing can remove an MTS and also result in different mRNAs, one with and one without the MTS coding sequence. The mechanism can also be determined at the level of translation; two products as above can be as a result of two translation initiation events on a single mRNA transcript as suggested for rat FH (20).

Examining the FH gene sequence reveals that the first exon encodes two in-frame ATGs, one at the beginning of the exon and the second constitutes the last three nucleotides of this exon (encoding M1 and M44, respectively, Figure 2). Alternative splicing seems an unlikely mechanism for FH distribution, because removal of the first exon would eliminate the two possible translation initiation sites and translation of the protein from any downstream AUG is improbable, because it would eliminate at least 30 highly conserved amino acids, which are crucial for the enzymatic activity of FH (28).

In rat liver, the suggested mechanism for FH distribution is alternative translation initiation from a single mRNA (20). Consequently, based on these reports we first hypothesized that alternative translation initiation, may be the mechanism of distribution of human FH. To approach this question we cloned the ORF of FH fused to a flag tag into a pBabe-puro plasmid following the MO-MuLV 5′LTR as the promoter (32). The resulting plasmid pHF was transformed into HEK293T cells. Expression and subcellular localization of the FH products were determined. In contrast to the endogenous FH, when expressed from the pFH plasmid, harboring only the coding sequence of FH, the protein is completely imported into the mitochondria as detected by subcellular fractionation (Figure 3, compare the two top left panels). Consistent with this conclusion, metabolic labeling in the presence of CCCP shows only the precursor form of the protein is made and processed (Figure 3, compare the two top right panels). Thus in contrast to the endogenous gene, these results indicate that from these plasmids all translation products of FH include the MTS.

The transcript of the plasmid-expressed-FH could lack important elements of the endogenous FH mRNA. It is well documented that the 5′ and 3′ UTRs affects the choice of the translation initiation codon (33–35). In order to examine the influence of the addition of the 5′ UTR on the distribution of FH from a plasmid, we constructed a plasmid containing an additional 60 nucleotides (nt) upstream to the first AUG as was proposed by the genome browser.
Figure 2: Schematic illustration of the FH gene (not to scale). The FH gene has 10 exons represented by solid black rectangles. Black lines between exons are introns and broken lines represent the upstream and downstream UTR regions. DNA: first and last nucleotides of the first exon (+1 and +132), as well as the location of the last nucleotide (+267) of the second exon, are indicated by arrows pointing up. Protein: The first and second in frame methionines (M1 and M44, respectively) are indicated with respect to their amino acid location by arrows pointing down. The first five amino acids (starting with A45) encoded by the second exon are indicated, the bold and underlined F is the first amino acid that is highly conserved from bacteria to human.

Figure 3: FH does not distribute by an alternative translation initiation mechanism. The schematic illustration of the FH gene (not to scale), on the left, shows the FH ORF (FH), the FH mitochondrial targeting sequence (MTS), the 5′ and 3′ UTRs (60, 39 and 48 bp), the Met1 and M44 methionines and the flag tag. The 5′ and 3′ ends of the transcripts are according to genome browser prediction (60 bp) or experimentally by RACE (39 and 48 bp, respectively). The panels on the right show the analysis of cells harboring wild-type endogenous FH or FH expressed from the indicated plasmids. Cells were subjected to subcellular fractionation (left panels). The anti-Hsp60 and anti-tubulin controls were performed for all samples and the one shown is representative. [35S]-methionine metabolic labeling in the absence or presence of CCCP is shown in the right panels. P, precursor; M, mature; M1 and M2 are the two processed forms of the M44V mutant. Antiserum: anti-FH; anti-flag; anti-tubulin (cytosolic marker); anti-Hsp60 (mitochondrial marker).

(http://genome.ucsc.edu/) for the FH 5′ UTR. The protein expressed from this construct, again, did not distribute and was completely imported into mitochondria (60 FH, Figure 3, third panel).

Because there are only predicted UTRs described for FH, we decided to determine these elements experimentally. For this we performed rapid amplification of cDNA ends (RACE), followed by sequencing of the amplified...
Dik et al.

products. We detected several 5′UTRs of different lengths. We assumed that the longest product contains all the control elements required for distribution. The 5′ UTR of FH that we detected contains 39 nt upstream of the first ATG, whereas, the 3′ UTR was 48 nt long, of which the terminal 16 nt constitute the poly A. We then re-cloned the FH ORF according to the determined UTRs into the same expression vector (Figure 3, p39 FH, fourth panel). Again FH appears as a single translation product (fourth right panel, right lane) and is not dual targeted residing exclusively in mitochondria (fourth left panel). Thus we were unable to restore the distribution of FH by including the UTRs. Worth mentioning is that some of these inserts were also constructed in pTre2 puro or pCDNA3.1 expression vectors under the CMV promoter (Figure S1, Supporting Information). Thus our results were not expression vector dependent.

To further test the hypothesis of alternative translation initiation as the FH distribution mechanism, we constructed the above plasmids with the FH gene harboring specific point mutations and examined the distribution of the corresponding plasmid-expressed-FHs. Mutagenesis of the initiation codons (AUGs) resulting in M1V and M44V (Figure 3). pM1V FH and pM44V FH eliminated the first and second translation initiation codons, respectively. FH M1V produced only the cytosolic echoform of FH, a protein without the mitochondrial targeting sequence as shown by subcellular fractionation and metabolic labeling (Figure 3, pM1V FH, fifth right and left panels, respectively). This result is consistent with the scanning mechanism of translation initiation, according to which, the ribosome scans the 5′ UTR until it reaches the first AUG codon it encounters and starts translating from it (36–38). The second mutation M44V resulted, as above with the wild-type, in full import into mitochondria (Figure 3, M44V FH, sixth panels). Unexpectedly we detected a new protein product (M2 arrow) which has an apparent molecular weight between those of the mature (M1 arrow) and precursor proteins (Figure 3, sixth right panel). This is most probably due to alteration of the MPP recognition site as will be discussed in more detail further on in this manuscript. Finally we constructed a third mutant FH, which contains an insertion of a single nucleotide between the two AUGs, at position 69 of the nucleotide sequence. This insertion results in a frame shift which should not eliminate the ability of the ribosome to start translation from this AUG1, but no expression of FH is expected from this first AUG1 due to the downstream frame shift. This construct allows us to test if the translation from the second AUG is possible in the presence of an intact first AUG. As shown in Figure 3 (pFS FH, seventh panels) under these conditions there is no expression of FH, implying that translation initiation from the first AUG in a sense ‘rules out’ translation initiation from the second. Taken together we conclude that the FH distribution mechanism cannot be explained by two translation products from a single mRNA.

According to Kozak (39) translation initiation can be affected by the context of the AUG. The most important position is −3 nt with respect to the A of the AUG and in fact this is the nucleotide in the human FH. Consistent with this argument, translation initiation from the first AUG of the human FH mRNA is highly efficient without skipping in vivo to the second AUG.

The FH gene has more than one type of transcripts

Once we concluded that a single mRNA mechanism is unlikely to explain FH distribution, we asked whether there is more than a single mRNA for FH in human cells. As mentioned above, when we performed the 5′ RACE reaction we detected more than one sequence. Considering this, we repeated the RACE reaction using a CapFishing™ RACE kit which amplifies only 5′ capped mRNAs, thereby increasing the specificity of the reaction. The reaction products were cloned into an A/T cloning vector. Out of the 90 plasmids extracted and sequenced, 22 were ‘empty’ (contained no sequence), whereas 68 provided FH 5′ sequences. Presented in Figure 4A are sequences that repeated more than once. For each sequence the position of the first nucleotide and the number of times each sequence was found are displayed on x- and y-axes, respectively. Displayed are 11 different 5′ ends of the FH mRNA that encompass the region from 45 nt upstream of the first AUG to 20 nt downstream of the second AUG. One exception is an mRNA starting at position +205.

Not only the context of the sequence around the translation initiation site is important but also the proximity to the cap site. Recognition of the 5′ proximal AUG codon is inefficient when located too close the cap site causing utilization of AUGs further downstream (40). A leader sequence

724

Traffic 2016; 17: 720–732
FH has more than one type transcript of transcripts. A) Graphic representation of all FH transcripts detected by the CAP Fishing™ assay: After amplification of FH mRNA 5′ ends, cDNA products were cloned into an A/T vector and 68 sequences were determined. The illustration bellow represents the location of each mRNA start point within the gene (x-axis) and number of times the sequence was found (y-axis). B) Measurement of transcript levels of FH mRNAs: transcript levels were determined using ddPCR. Primer sets were designed to amplify only transcripts that contain a specific region of the 5′ UTR (blue bar), first exon after the first ATG (after ATG1, red bar) and second exon (after ATG2 green bar). Error bars indicate the Poisson 95% confidence intervals for each copy number determination.

shorter than 15 nt is expected to be too short for proper binding and activity of the ribosome [reviewed by (41)]. To simplify our analysis, we assumed that all transcripts with 5′UTRs longer than 18 nt are translated from the first AUG, whereas for the shorter mRNAs the ribosome does not initiate translation and continues to scan until it reaches the second AUG. The transcript starting much further downstream (nt 205) is not expected to produce an active enzyme and no protein products were detected to support this short translation product (Figure 6), as will be discussed later on.

The detection of two major groups of transcripts, which according to our initial findings, can each generate only one of the enzyme’s echoforms, supports a mechanism of distribution at the level of mRNA generation; transcription. It is important to mention that none of the sequences displayed any evidence for trans-splicing or splicing, such as non-FH sequences or spliced FH sequences. This conclusion is supported using the MS analysis of FH products shown in the following sections.

Additional proof that FH distributes by generation of different mRNAs, came from measuring the level of FH transcripts using three sets of primers, each for a different section of the FH coding sequence. The assumption is that if there are mRNAs with different termini we should detect variations in the levels of the different regions of the FH transcript. For instance, expression from the first AUG is only possible from mRNAs that contain the 5′ UTR region, whereas the expression of the cytosolic echoform is only possible when the first AUG is functionally absent.

We measured transcript levels by digital droplet PCR (ddPCR) (42). This method involves a regular real-time PCR reaction which is distributed across multiple reactions by packing the reaction mix and template within oil droplets. This dilutes the DNA to a level at which some reactions have no template and others that have one or more template copies present. After amplification to the terminal plateau phase of PCR, reactions containing the template of interest yield positive endpoints, which are detected by evagreen (Biorad) staining of the droplets. The number of target DNA molecules present can be calculated from the fraction of positive endpoint reactions using Poisson statistics. As ddPCR relies on a binary endpoint threshold to assign each replicate reaction as either positive or negative (one or zero, respectively), it can tolerate

Figure 4: FH has more than one type transcript of transcripts. A) Graphic representation of all FH transcripts detected by the CAP Fishing™ assay: After amplification of FH mRNA 5’ ends, cDNA products were cloned into an A/T vector and 68 sequences were determined. The illustration below represents the location of each mRNA start point within the gene (x-axis) and number of times the sequence was found (y-axis). B) Measurement of transcript levels of FH mRNAs: transcript levels were determined using ddPCR. Primer sets were designed to amplify only transcripts that contain a specific region of the 5′ UTR (blue bar), first exon after the first ATG (after ATG1, red bar) and second exon (after ATG2 green bar). Error bars indicate the Poisson 95% confidence intervals for each copy number determination.
wide variations in amplification efficiencies without affecting DNA copy number estimation. Because we used three different primer sets, we could avoid the need to consider primer efficiency which by regular real-time PCR can greatly influence the results (42).

Three sets of primers were applied which were designed to recognize either 5'UTR, the first exon in proximity to the second AUG, or the second exon. The assay was repeated for two different cell lines, HeLa and HEK293T. As shown in Figure 4B, in both cell lines tested, the level of transcripts harboring the first AUG (5’UTR primer set) are significantly lower when compared with the level of transcripts harboring the first exon (after AUG1). This indicates that there is a subgroup of transcripts which contain the second AUG but not the first one. These results are consistent with the RACE results, demonstrating that there are two types of mRNAs for FH, each of which can give rise to one of the protein’s echoforms. Worth mentioning is the fact that we see a third small group of transcripts lacking both AUGs whose biological relevance is unclear (‘205’, Figure 4A).

**FH, expressed from its original promoter, distributes by an alternative transcription initiation mechanism**

The proposed distribution mechanism at the transcription level predicts that reconstitution of FH distribution from a plasmid should occur by adding the genetic elements that control its transcription. We constructed a plasmid containing 1000bp upstream to the FH ORF. As shown in Figure 5B (upper panel, lanes 1–3), the presence of the promoter region resulted in FH distribution between the mitochondria and cytosol of the ‘p1000FH’ plasmid-encoded-FH. Metabolic labeling detected two translation products, as previously shown for the endogenous protein (Figure 5C, p1000FH, lanes 3 and 4), indicating distribution by a mechanism of alternative transcription initiation. According to this mechanism, each echoform is expected to be translated from a different transcript. To test this model, we inserted a nucleotide between the two AUG’s thereby creating a frame shift mutation which should only disrupt the expression from the transcript harboring the first AUG (of the mitochondrial echoform). In fact that is what we found, and unlike the FS mutant, without the FH promoter, Figure 3; we were able to detect a FH protein product from the 1000 FS plasmid. As predicted this FH product is restricted to the cytosolic fraction (Figure 5B, lanes 4–6) and metabolic labeling detects a single product (Figure 5B, lanes 5 and 6). Additional verification of these results is a M44V mutant containing the promoter region, which as expected, expresses only the mitochondrial protein (Figure S3). Taken together, our results indicate that FH distribution is determined at the level of transcription initiation.

**Detection of the FH protein products**

Up to this point in this study we examined the products of the FH gene on denaturing polyacrylamide gels using western blot or following labeling and immunoprecipitation. To obtain more precise information, we employed MS with the intention of determining the N termini of the different FH protein products. We isolated the endogenous enzyme by immunoprecipitation and determined its N terminus after extracting the purified protein from an acrylamide gel. A single N terminus was detected, ASQNSF (Figure 6, endogenous FH).

The fact that we detected a single amino terminal peptide starting immediately after the second methionine was surprising. The ASQNSF sequence result is consistent with removal of the MTS by MPP between amino acid M44 and A45 as predicted by MitoProt (43). However, we expected to detect a second N terminus, of the cytosolic enzyme, starting with a methionine, MASQNSF, because this cytosolic echoform is expressed from the second in frame AUG. At this stage we assumed that a Met aminopeptidase (MetAP) could remove the first methionine. In this regard MetAPs have been shown to require a short side chain amino acid at the position immediately following the N-terminal methionine (P1). Specifically for both MetAP1 and MetAP2 alanine is preferred (44,45) which in fact is the amino acid following M44.

To gain support for the above notions we decided to look at FH mutants that were referred to, in previous sections. These mutant FHs were tagged with flag and overexpressed from plasmids, purified and the samples were subjected to N terminal labeling by dimethylation (31). This procedure allows one to distinguish between the naturally occurring N terminal in vivo, with those that may occur during the preparation of the samples for the MS analysis. Translation of M1V (the first AUG mutant), whose translation initiates
Figure 5: FH distributes by an alternative transcription initiation mechanism. A) Schematic illustration of the indicated plasmids. Bold black arrows represent the two translation initiation codons. MTS, FH, Met1, Met44 and flag are as indicated in the legend to Figure 3. ‘Promoter’ indicates a 1000 bp FH sequence upstream of ATG1 (encoding Met1). The +C with the thin arrow indicates insertion of a cytosine within the MTS encoding DNA sequence which creates a frame shift (FS) of translation initiating from Met1 (ATG1). B) Subcellular fractionation of HEK293T cells harboring the indicated plasmids. The total (T), cytosolic (C), and mitochondrial (M) fractions were analyzed using western blotting using the indicated antibodies. Controls: tubulin, cytosolic marker; Hsp60, mitochondrial marker. C) Metabolic labeling ([35S]-methionine) of HEK293T cells in the absence or presence of CCCP, with or without the indicated plasmids. FH was immunoprecipitated with the indicated antibodies followed using PAGE and autoradiography.

The MS analysis of plasmid-expressed-unmutated FH gave both the ASQNSF sequence which is product of MPP processing and in addition at very low levels, a second N terminus ASAPGL (starting at amino acid A23, Figure 6, p39 FH, second row). This was a rare peptide in the sample based on both the frequency of the peptides detected starting at this position (A23), and semiquantitation that was calculated by the peak area of each peptide.

The second AUG mutant M44V produced the ASAPGL N terminus as the more frequent peptide, along with significant amounts of the ASQNSF peptide. These results indicate that by mutating M44 we affected the MPP processing. This is not surprising because the MPP cleavage site is between M44 and A45, and replacing M44 with a valine, changes the MPP recognition site. This change of the MPP cleavage site allows a less preferred recognition site (between L22 and A23) to be processed by MPP. The detection of high amounts of the longer form of FH (ASAPGL) in this mutant, can also explain the novel high-molecular band detected on SDS polyacrylamide gels for this mutant (M2 band, Figure 3). In fact, we confirmed that the M1 and M2 bands (Figure 3) of FH have the ASQNSF and ASAPGL N-termini, respectively. Separate MS analysis was carried out for each band.

Figure 6: Determination of the FH protein N termini and MTS. Endogenous genome encoded FH was purified from HEK293T cell lysates by immunoprecipitation. The N terminus was determined using mass spectrometry (MS) analysis. Plasmid encoded FHs were purified from HEK293T cells and were subjected to N terminal labeling by dimethylation followed using MS analysis. Antisera: anti FH, anti-flag.

only from the second AUG, resulted mainly the ASQNSF N-terminus (Figure 6, M1V FH, third row). This result supports our conclusion that the MASQNSF terminus is processed by MetAPs. The MASQNSF peptide was also detected for this mutant but at lower levels, which can be explained by the plasmid-encoded-M1V-FH mutant overexpression.
In all of the samples no shorter N termini were detected, indicating that no AUG downstream to the second AUG, is functional (Table S1). These MS results are in agreement with our proposed mechanism of FH distribution. These results also provide insight into events of processing, such as MPP cleavage and Met-aminopeptidase cleavage.

**Discussion**

Here we explore the distribution mechanism of FH in human cells. We show that the FH distribution mechanism is at the level of transcription initiation, where for each echoform separate mRNAs are made containing or lacking a first functional AUG. The data supporting this conclusion are as follows (i) The FH ORF expressed from a plasmid with a defined promoter (Mo MuLV 5’LTR or CMV promoter) initiating upstream of the first ATG, produces only the mitochondrial protein. (ii) There is no expression from the second AUG when the first AUG is present on the mRNA (a frameshift mutation between the ATGs abolishes all expression). (iii) A variety of transcripts were detected both by ddPCR and RACE (both upstream and downstream of the first ATG). (iv) The FH ORF expressed from a plasmid containing the promoter elements of FH (1000 bp upstream of the first ATG) expressed the two echoforms (p1000-FH). (v) The presence of the first AUG in this construct did not interfere with the expression of the second shorter echoform whose transcription initiates from the second ATG (p1000-FS).

In addition, the different N-termini that we detected by MS support our model of an alternative transcription initiation mechanism of dual targeting. According to our analysis MPP cleaves between M44 and A45, whereas translation of the shorter cytosolic echoform starts at M44. The MS analysis provides insight into the processing events which FH undergoes. The molecular size of mitochondrial and cytosolic FH echoforms, as can be detected using western blot and MS analysis, are identical. MS analysis of the M1V mutant confirms the identity of the N-terminus of the two echoforms which both start with A45. Without going into details here, the data indicate that the N-terminal methionine of the shorter cytosolic echoform is removed by methionine aminopeptidases. It is important to note that no other N-terminal sequences were detected which could support alternative splicing or translation from downstream AUGs. Another point worth making is that previous reports expressing FH from plasmids should be revisited with regard to conclusions drawn, because we expect that such expressed FHs are completely localized to the mitochondria.

Our findings imply that the FH promoter region contains the crucial information for FH dual targeting. The core promoter is defined as the DNA segment of 50–100 bp within which transcription initiates (46). This region is required for recruitment of the transcription apparatus and can be thought of as the priming stage for transcription initiation. Promoters can be classified with respect to the distribution of the transcription start sites (TSSs) they use. ‘Sharp’ class promoters use only one or a few consecutive nucleotides as TSSs, resulting in a single-peak TSS distribution. These promoters often have TATA and/or initiator boxes and are correlated with tissue-specific genes (47). A second group are 'broad' class promoters. For these, transcription can initiate over a ~150 bp region, resulting in a population of mRNAs that have different lengths but usually the same protein-coding content. Broad promoters are often TATA-less and CpG-island-enriched and are associated with housekeeping genes. Nevertheless, promoters containing both TATA boxes and CpG islands, as well as promoters without any of these elements, have been reported (48,49).

The FH promoter region contains a number of potential control elements. Of interest is a 639 nt sequence that starts at position –90 with regard to the first ATG, and includes the first exon and part of the first intron. This sequence has a high GC content and contains 76 CGs, thus it is considered a CpG island. CpG islands are genomic stretches in which CG dinucleotides are overrepresented at significantly higher levels than is typical for the genome as a whole. CpG’s in the promoter regions are unmethylated CpGs, in contrast to methylated CpG rich regions which are associated with gene silencing (50). The existence of a CpG island in the promoter of the FH gene, is consistent with our findings of multiple TSSs for FH, suggesting a broad class promoter. A few potential TATA-like regions were identified using a prediction analysis by Eukaryotic Promoter Database (EPD) (http://epd.vital-it.ch/). However, these TATA boxes are located more than 50 nt upstream to the longest 5’UTR.
A schematic model of the FH dual distribution mechanisms summarized in Figure 7. The broad promoter (A) leads to the transcription of different mRNAs which can be divided into two groups, one with at least 15 nt upstream to the first AUG indicated by orange arrows (B) that are translated from the first AUG of the MTS. The second group of mRNAs lacks the first AUG or lacks sufficient length of 5′ UTR to be translated from the first AUG (less than 15 nt). These mRNAs are translated from the second AUG without the MTS and the protein products remain in the cytosol (C).

The broad promoter of FH is expected for a housekeeping gene which is required for respiration. The FH gene represents unique use of a broad promoter, which allows the generation of multiple mRNAs whose protein products are differentially localized. This mechanism of FH dual targeting in human, differs from suggested mechanisms of this highly conserved enzyme in rat liver (translation initiation), yeast (reverse translocation) and Arabidopsis (two genes). Although the mechanism of dual targeting may differ, the dual function of these enzymes appears to be conserved (10). These findings are consistent with our hypothesis that the evolutionary driving force for protein dual targeting, is the function of each echoform in its specific subcellular compartment, regardless of the mechanism by which the proteins are targeted to their destination (53). Consistent with this line of thought is the extensive diversity of mitochondrial targeting sequences (Figure S2) while the fumarase structural protein sequences are highly conserved.

Many transcription factors (TFs) have been reported to be associated with the FH upstream sequence. According to the USCS (University of California Santa Cruz) genome browser (https://genome.ucsc.edu/) more than 70 different TFs were detected in different ChIP assays for the 1000 nt region upstream to FH. Future experiments will map this region at a higher resolution in order to understand the elements controlling the transcription of FH in human, and whether we can distinguish between different transcription initiation and control elements under different growth or stress conditions (e.g. DNA damage).

In the recent years FH is the focus of an intensive research along with other metabolic enzymes such as succinate dehydrogenase (SDH) and isocitrate dehydrogenase (IDH). It is becoming clear that the metabolites generated by these enzymes can modulate the activity of various enzymatic processes in the cell a are strongly related to cancer prevention and development (54). Specifically, FH
(fumarase) and fumarate have been implicated in various cellular processes related to diseases such as oxidative stress leading to mitochondrial dysfunction, succination leading to diabetes and DNA damage and genome instability leading to cancer (10–12,15–19). Thus, our contribution to the understanding of gene expression and dual subcellular localization of this enzyme is important for future studies.

Materials and Methods

Cell lines and growth conditions

HEK293T/17 [American Type Culture Collection (ATCC): CRL-11268], HeLa (ATCC: CCL-2) cell lines were purchased from ATCC and grown according to ATCC specifications in Dulbecco’s Modified Eagle Medium (DMEM; D5671 sigma) with 10% FBS European grade (040071A Biological industries).

Plasmids

Human FH was amplified by polymerase chain reaction (PCR) from human cDNA and a flag tag was added on the reverse primer. Untranslated regions (UTRs) were also added on the primers. The 60 nt long 5′ UTR was fused by a two-step PCR. The different fragments were inserted into three different plasmids pBabe-puro, pTre2-puro and pCDNA3 as indicated. All fragments inserted into pBabe-puro are between BamHI and SalI restriction sites. Fragments were inserted into pTre2-puro between Nhel and ClaI restriction sites.

In order to clone the promoter site of FH preceding to the open reading frame (ORF), the cytomegalovirus (CMV) promoter region of pCDNA3 plasmid was removed by restriction with NruI and HindIII restriction enzymes, followed by a Klenow reaction and the purified plasmid was ligated. A two steps PCR amplified fragment including 1000 bp upstream of FH was fused to FH encoding ORF these fragments were cloned into the promoterless pCDNA3 using BamHI and XhoI restriction sites.

Metabolic labeling of yeast cells

Wild-type yeast strain (BY4741) was culture to logarithmic phase in galactose medium. The cells were harvested and labeled with 40 μCi/mL [S35] methionine and further incubated for 3 h. Cells were then harvested and lysed in 400 μL RIPA (Radioimmunoprecipitation assay) buffer (50 mM Tris–HCl pH 7.4, 1% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA and protease inhibitor cocktail) for 30 min on ice. Cell debris was pelleted using centrifugation at 16,000 × g for 15 min. The supernatant was transferred to new tube and antibodies and/or beads were added for overnight incubation at 4°C, as follows: Flag tagged proteins were precipitated with 15 μL anti-Flag affinity gel (Sigma A2220), anti-FH (AB110286) and anti-Hsp60 antibody were diluted 1 of 100 and precipitated with 10 μL protein A conjugated magnetic beads (Life technology 10002D). Beads were washed with RIPA buffer, elution was done with 1× Laemmli sample buffer. Samples were analyzed using SDS-PAGE and autoradiography.

Cell fractionation

A total of 3 × 10 cm plates were grown to 70% of confluence, the media was removed and cells were washed twice with ice cold PBS. Cells were gently scraped and gathered into a single tube and spun down at 600 × g for 10 min. The cell pellet was washed once in 1 mL of MIB (200 mM mannitol, 70 mM sucrose, 1 mM EGTA, 10 mM HEPES-KOH, pH 7.4, 1 mM DTT and a protease inhibitor cocktail) and spun down at 600 × g for 5 min. The cell pellet was resuspended in 1 mL of MIB and incubated on ice for 5 min and then homogenized with 30 strokes of the Dounce homogenizer. The homogenate was spun down at 600 × g for 10 min in a pre-cooled refrigerated centrifuge (4°C). The supernatants were spun down again and a total (7) sample was taken from the supernatant before the next step. The supernatant was spun at 10,000 × g for 10 min at 4°C (the pellet contains mitochondria while the supernatant contains cytosol and microsomes). A cytosol sample was taken from the supernatant and the pellet was resuspended with 150 μL MIB. All samples where boiled in Laemmli sample buffer.

Digital droplet PCR

The ddPCR reaction was performed with 1 ng cDNA synthesized from HeLa or HEK293T cells, 1 μmol primer mix and QX200 ddPCR evagreen supermix (Bio-Rad 186-4036). The reaction was performed in the Biorad QX200 ddPCR system. Results were analyzed using the QuantaSoft software.

Primer sets that were used for amplification are: 5′ UTR: Forward: CAGCACCAGTGACGAG. Reverse: GACCTTTAGTCACCAAAAG. Exon 1: Forward: GTGGCCCTCGTTTTTG. Reverse: AAAGCGGTGT-TACCTCAGCG. Exon 2: Forward: CTTTGTGAACTAAGGTGC. Reverse: CTCATCGCTGCTTTGATT.

Race and Cap fishing assay

Smarter Race by Clontech and Seegene Cap fishing assay were performed according to manufacturer protocol.
**Protein purification for MS**

HEK293T cells were transfected with flag tagged plasmid. After 48 h of transfection cells were harvested and lysed with Hepes-RIPA buffer (50 mM Hepes pH 7.4, 1% NP-40, 0.25% sodium deoxycholate, 150 mM NaCl, 1 mM EDTA and protease inhibitor cocktail). Purification preceded according to Sigma protocol for flag tagged proteins using M2 flag affinity gel. Elution was carried out with 100 μL 0.1 M glycine HCl, pH 3.5 which was neutralized by adding 10 μL of 0.5 M Tris–HCl, pH 7.4, with 1.5 M NaCl. Eluted proteins were subjected to N-terminus labeling by dimethylation (31) and then were analyzed using MS.

**Supporting Information**

Additional Supporting Information may be found in the online version of this article:

**Figure S1:** Subcellular fractionation of FH-flag expressed from pTRE2-puro or pCDNA3 plasmids. Fractions were analyzed by PAGE followed using Western blot analysis. Anti-Hsp60, mitochondrial marker; anti-tub (tubulin), cytosolic marker.

**Figure S2:** Multiple alignment of the N-terminal amino acid sequences of fumarases from different organisms. Methionines (in bold) are colored as follows: blue, first methionine upstream of a mitochondrial targeting sequence (MTS); green, putative methionine for translation initiation of fumarases lacking a MTS; red, amino acids that are highly conserved in all organisms and which identify where the structural fumarase enzyme sequence starts.

**Figure S3:** Subcellular fractionation of p1000M44V. Cells harboring the designated plasmids were subjected to subcellular fractionation. Total (T), cytosol (C) and mitochondrial (M) fractions were analyzed using Western blot. Anti-Hsp60, mitochondrial marker; anti-tub (tubulin), cytosolic marker, anti-FH (fumarase); anti-flag.

**Table S1:** Mass spectrometry results of the N’-terminal determination of FH and the different mutants used in Figure 6.

**References**

20. Tuboi S, Suzuki T, Sato M, Yoshida T. Rat liver mitochondrial and cytosolic fumarases with identical amino acid sequences are encoded...
35. Park J, Sharma N, Cutting GR. Melanocortin 3 receptor has a 5’ exon that directs translation of apically localized protein from the second in-frame ATG. Mol Endocrinol 2014;28:1547–1557.