Review–Hypothesis

Why mitochondria need a genome

Gunnar von Heijne

Research Group for Theoretical Biophysics, Department of Theoretical Physics, Royal Institute of Technology, S-100 44 Stockholm, Sweden

Received 2 December 1985; revised version received 10 January 1986

The evolution of the mitochondrial genome towards the compact organization found in the higher eukaryotes is discussed. It is suggested that the machinery for co-translational protein export across the endoplasmic reticulum membrane sets strict limits on the kinds of protein-coding genes that can be successfully transferred from the mitochondrial to the nuclear genome. This hypothesis is in perfect agreement with the pattern of mitochondrially vs nuclearly encoded mitochondrial proteins found in species such as man, mouse, and *Xenopus*.

Mitochondria *Evolution* *Protein export*

1. INTRODUCTION

The mitochondrial genome of higher eukaryotes is extremely economically organized [1–3]. It makes use of only 22 tRNAs. It codes for no more than 13 proteins, the overwhelming majority of mitochondrial proteins being encoded in the cell nucleus and imported into the mitochondrion in a predominantly post-translational fashion [4]. There are no introns or 'spacers' between the genes, and a few genes even overlap. These observations indicate that the mitochondrial genome in these organisms has been stripped of all elements that could be either moved to the nucleus or dispensed with altogether; but why do some genes still remain? Here, I show that only one assumption is needed to account for the existence and composition of the present-day mitochondrial genome, namely that the mechanism of post-translational mitochondrial protein import evolved only *after* the signal peptide-dependent machinery for co-translational protein export from the cell had come into existence.

2. PROTEIN SECRETION AND MEMBRANE PROTEIN BIOGENESIS

According to the currently accepted model for protein export in eukaryotic cells [5], export through the endoplasmic reticulum is initiated co-translationally as soon as a hydrophobic leader peptide [6] (or signal sequence) emerges from the ribosome. This leader peptide may subsequently be cleaved from the mature chain by a leader peptidase (thus producing either a secreted soluble protein or an internally anchored membrane protein), or may resist cleavage to yield an N-terminally anchored protein. In fact, it has been shown that an N-terminal transmembrane segment can initiate export and permanently anchor an unrelated protein to the membrane when fused to its N-terminus [7]. A sufficiently long *internal* hydrophobic stretch apparently can also initiate export, resulting in a protein with a large cytosolic N-terminal domain [19,20]. Such internal stretches have also been shown to initiate export in the normal way when artificially moved to the N-terminus of the protein [21]. A hydrophobic segment no further than 70–90 residues from the C-terminus, however, is not expected to activate the co-translational export machinery, since the whole protein will be finished by the time such a segment emerges from the ribosome. Finally, the only difference between an N-terminal, cleavable signal peptide and a (N-terminal or internal) permanent anchor segment found so far is the length of the

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

Table 1
Hydrophobicity analysis of mitochondrially and nuclearly encoded proteins

Protein	No. of membrane segments	Signal sequence	N–C distance
Mitochondrial genes:			
Xenopus			
Cytochrome *c* oxidase I	8–12	56	499
Cytochrome *c* oxidase II	2– 3	27	202
Cytochrome *c* oxidase III	3– 6	20	225
Cytochrome *b*	6–10	46	345
ATPase 6	6– 6	9	213
URF1	6– 9	11	313
URF2	4– 9	205	167
URF3	3– 3	1	109
URF4	4–10	4	400
URF4L	2– 3	1	95
URF5	14–15	9	600
URF6	4– 4	1	170
URFA6L	1– 1	6	46
Yeast			
ATPase 8	1– 1	13	30
ATPase 9	2– 2	14	60
Hypothetical protein B-386	0– 0	–	–
Ribosomal protein var1	0– 0	–	–
Cytochrome *b* maturase (intron)	3– 6	32	389
oxi3 intron 1 protein	1– 3	14	820
oxi3 intron 2 protein	0– 3	31	762
oxi3 intron 3 protein	0– 0	17	–
oxi3 intron 4 protein	3– 6	22	455
cob intron 4 protein	1– 0	258	–
21 S RNA intron protein	0– 0	–	–
Nuclear genes:			
Yeast MSS 51	0– 0	–	–
Yeast porin	0– 0	–	–
Yeast 70 kDa protein	0– 0	–	–
Yeast cytochrome c_1	1– 0	–	–
Yeast cytochrome *c* oxidase IV	0– 0	–	–
Yeast cytochrome *c* oxidase V	1– 1	98	58
Yeast cytochrome *c* oxidase VI	0– 0	–	–
Yeast cytochrome *c* oxidase VIII	1– 1	27 + ts	20
Yeast manganese superoxide dismutase	0– 0	–	–
Yeast cytochrome *c* peroxidase	0– 0	–	–
Yeast cytochrome *c* reductase, 14 kDa	0– 0	–	–
Yeast cytochrome *c* reductase, 17 kDa	0– 0	–	–
Yeast citrate synthase	1– 0	131	(48)
Yeast cytochrome *c*	0– 0	–	–
Yeast EF-Tu	0– 0	–	–
Yeast ilv1 protein	1– 4	393	(346)
Neurospora ATPase 9	2– 2	91	59
Neurospora Rieske 2Fe-2S protein	0– 0	68	–
Nicotiana ATPase β	0– 0	–	–
Rat carbamoyl-phosphate synthetase I	0– 0	–	–
Rat ornithine transcarbamylase	0– 0	334	–
Pig malate dehydrogenase	0– 0	–	–
Pig aspartate transferase	0– 0	378	–
Pig citrate synthase	0– 0	–	–
Pig 3-hydroxyacyl-CoA dehydrogenase	0– 0	–	–
Bovine ATPase β-chain	0– 0	–	–
Bovine ATPase inhibitor	0– 0	–	–
Bovine cytochrome P450 (SCC)	0– 0	–	–
Bovine ATPase 9	2– 2	58 + ts	21
Bovine coupling factor 6	0– 0	–	–
Bovine cytochrome *c* oxidase IV	1– 1	102	67
Bovine cytochrome *c* oxidase V	0– 0	–	–
Bovine cytochrome *c* oxidase VII	0– 0	–	–
Bovine uncoupling protein	2– 4	215 + ts	90
Bovine ADP/ATP carrier protein	1– 4	281 + ts	16
Bovine adrenodoxin	0– 0	–	–
Bovine rhodanese	0– 0	–	–

hydrophobic stretch: it ranges approximately from 7 to 15 residues in the former and from 15 to 25 residues in the latter [6,12]. There appear to be no strong constraints on the sequence of the hydrophobic stretch as long as the length and minimum hydrophobicity requirements are met [6], but internal signal peptide-like segments that are not long enough to qualify as transmembrane helices apparently cannot initiate export [13].

These properties of the present-day co-translational protein export machinery suggest that if this machinery was already in existence when the first mitochondria appeared, only those mitochondrial genes whose protein products did *not* contain any potential hydrophobic export signals could have been successfully moved into the nucleus.

3. ANALYSIS OF MITOCHONDRIAL PROTEINS

To test this hypothesis, 24 mitochondrially and 37 nuclearly encoded mitochondrial proteins were extracted from the National Biomedical Research Foundation protein sequence data bank (version 6.0, September 1985) and searched for the existence of hydrophobic regions that could serve as targeting signals for export across the endoplasmic reticulum. As shown in table 1, *all* of the mitochondrially encoded proteins from *Xenopus* and most of those from yeast, but only a few of the nuclearly encoded ones are predicted to be intrinsic membrane proteins with long, hydrophobic membrane-spanning regions. Moreover, the few imported membrane proteins found have their hydrophobic regions located within ~90 residues of their C-termini, whereas the mitochondrially encoded proteins are predicted to have membrane-spanning segments all through their sequences. Signal sequence-like segments are also present near the N-terminus of most of the mitochondrially encoded proteins, but are only found far from it in the nuclearly encoded group.

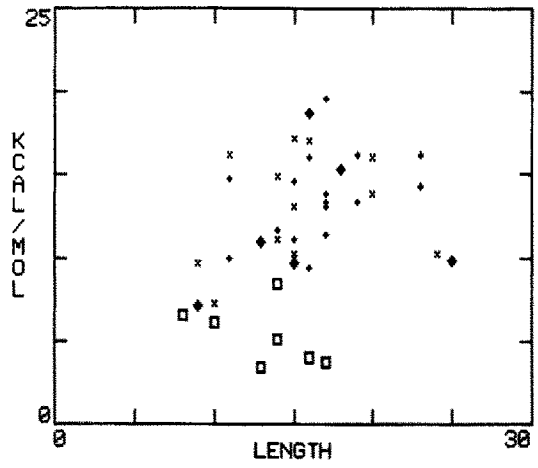

Fig.1. Length vs total hydrophobicity [12] (in kcal/mol) of the most hydrophobic predicted transmembrane segment in the C-terminal (N–C distance ≦ 90 residues) and N-terminal regions of each of the mitochondrially and nuclearly encoded proteins in table 1. (×) Mitochondrial C-terminal segments, (+) mitochondrial N-terminal segments, (♦) nuclear C-terminal segments, (□) nuclear N-terminal segments. Five nuclearly encoded proteins [14–18] with suggested N-terminal transmembrane segments not so predicted by the methods employed here have also been included in the last group.

These findings are further illustrated in fig.1, where the length and total hydrophobicity of the most hydrophobic C-terminal (N–C distance ≦ 90 residues) and N-terminal (N–C distance > 90 residues) segments have been plotted for each protein in the mitochondrial and nuclear groups. It is clear that long hydrophobic segments are found throughout the mitochondrial sequences, but only in the C-terminal region of the nuclear ones.

Thus, *none* of the mitochondrially encoded *Xenopus* proteins in table 1 would be expected to be able to escape co-translational routing into the export pathway were they to be made by cytosolic ribosomes, the only possible exception being the URFA6L protein which probably is too short to be

The mitochondrially encoded proteins are from the complete *Xenopus* genome [1] and from yeast (not including proteins homologous to *Xenopus*). Hydrophobicity analysis was carried out according to Kyte and Doolittle [8] with a span length of 19 residues and the requirement that ⟨HI⟩ ≧ 1.6 for predicting membrane segments (first entry), and Eisenberg et al. [11] (span length = 21, second entry). Putative signal sequence-like segments were searched for by looking for the most N-terminal segment with a *total* hydrophobicity ≧ 7 kcal/mol for span lengths in the interval 7–16 residues as described elsewhere [12] (+ts indicates that the N-terminal mitochondrial targeting sequence is unknown). The N–C distance is the number of residues between the first residue of the most N-terminal of the predicted membrane segments that is also predicted to be a signal sequence, and the C-terminus (parentheses indicate that the protein is probably erroneously predicted to be a transmembrane protein)

co-translationally exported. Interestingly, this protein overlaps the coding region for the ATPase subunit 6 by 10–46 nucleotides in the *Xenopus*, mouse, bovine, and human mitochondrial genomes [1]; this overlap may effectively 'lock' the URFA6L gene to the mitochondrion.

Although the yeast genome is less streamlined than those of the higher eukaryotes, most of the mitochondrially encoded yeast proteins similarly seem to be permanently confined to the mitochondrial genome by virtue of their hydrophobic segments; this is even the case for many of the intron-encoded proteins (table 1).

4. CONCLUSION

The need for a minimal mitochondrial genome in the higher eukaryotes and the fact that only highly hydrophobic proteins are retained in this genome can be explained if it is assumed that any gene successfully transferred from the mitochondrial to the nuclear genome during evolution had to code for a protein that could escape a pre-existing co-translational export mechanism. Experimentally, this means that fusing a mitochondrial targeting sequence onto a mitochondrially encoded protein and transferring the resulting gene to the nucleus would give rise to a protein that might be possible to import into mitochondria in vitro but not in vivo (since even markedly hydrophobic proteins such as the ATPase subunit 9 from *Neurospora* can be post-translationally imported, there may be no strong sequence constraints on import once a protein gets as far as the mitochondrial outer membrane). Conversely, tagging a sufficiently long C-terminal extension on to a nuclearly encoded mitochondrial membrane protein such as cytochrome *c* oxidase subunit VIII would be predicted to misroute it into the export pathway (had mitochondrial targeting sequences been C-terminal rather than N-terminal, the gene for the coxVIII protein would presumably still be found in the mitochondrion).

Finally, if, as seems to be the case, there is much to be gained by the cell from reducing the size of the mitochondrial genome, the need for this genome only because nuclearly encoded mitochondrial proteins have to be compatible with a pre-existing co-translational export machinery is perhaps the best example yet of a lack of global optimization of cellular function at the molecular level.

ACKNOWLEDGEMENT

This work was supported by a grant from the Swedish Natural Sciences Research Council.

REFERENCES

[1] Roe, B.A., Ma, D.-P., Wilson, R.K. and Wong, J.F.-H. (1985) J. Biol. Chem. 260, 9759–9774.
[2] Anderson, S., Bankier, A.T., Barrell, B.G., De Bruijin, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) Nature 290, 457–465.
[3] Anderson, S., De Bruijin, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.C. (1982) J. Mol. Biol. 156, 683–717.
[4] Reid, G.A. (1985) Curr. Top. Membranes Transp. 24, 295–336.
[5] Wickner, W.T. and Lodish, H.F. (1985) Science 230, 400–407.
[6] Von Heijne, G. (1985) J. Mol. Biol. 184, 99–105.
[7] Bos, T.J., Davis, A.R. and Nayak, D.P. (1984) Proc. Natl. Acad. Sci. USA 81, 2337–2341.
[8] Kyte, J. and Doolittle, R.F. (1982) J. Mol. Biol. 157, 105–132.
[9] Shaw, M.W., Lamb, R.A., Erickson, B.W., Briedis, D.J. and Choppin, P.W. (1982) Proc. Natl. Acad. Sci. USA 79, 6817–6821.
[10] Strubin, M., Mach, B. and Long, E.O. (1984) EMBO J. 3, 869–872.
[11] Eisenberg, D., Schwarz, E., Komaromy, M. and Wall, R. (1984) J. Mol. Biol. 179, 125–142.
[12] Von Heijne, G. (1986) J. Mol. Biol., in press.
[13] Wirth, D.F., Lodish, H.F. and Robbins, P.W. (1979) J. Cell Biol. 81, 154–162.
[14] Harnisch, U., Weiss, H. and Sebald, W. (1985) Eur. J. Biochem. 149, 95–99.
[15] Smeekens, S., De Groot, M., Van Binsbergen, J. and Weisbeek, P. (1985) Nature 317, 456–458.
[16] Sadler, I., Suda, K., Schatz, G., Kaudewitz, F. and Haid, A. (1984) EMBO J. 3, 2137–2143.
[17] Kaput, J., Goltz, S. and Blobel, G. (1982) J. Biol. Chem. 257, 15054–15058.
[18] Hase, T., Riezman, H., Suda, K. and Schatz, G. (1983) EMBO J. 2, 2169–2172.
[19] Kopito, R.R. and Lodish, H.F. (1985) Nature 316, 234–238.
[20] McClelland, A., Kuhn, L.C. and Ruddle, F.H. (1984) Cell 39, 267–274.
[21] Friedlander, M. and Blobel, G. (1985) Nature 318, 338–343.