Hypothesis

Mitochondrial uncoupling proteins and phylogensis – UCP4 as the ancestral uncoupling protein

Petr Hanák, Petr Ježek*

Department of Membrane Transport Biophysics, No. 375, Institute of Physiology, Academy of Sciences of the Czech Republic, Vídeňská 1083, CZ 14220 Prague, Czech Republic

Received 20 December 2000; revised 31 January 2001; accepted 6 February 2001

First published online 28 March 2001

Edited by Vladimir Skulachev

Abstract We searched for the previously defined uncoupling protein (UCP) signatures [Ježek, P. and Urbánková, E. (2000) IUBMB Life 49, 63–70] in genomes of Drosophila melanogaster, Caenorhabditis elegans, Dictyostelium discoideum, and Arabidopsis thaliana. We identified four UCPs in Drosophila and one in Caenorhabditis or Dictyostelium as close relatives of human UCP4 (BMCP), but distant from UCP1, UCP2, UCP3, and two plant UCPs of Arabidopsis. But the third Arabidopsis UCP is the closest UCP4 relative. This suggests that UCP4 represents the ancestral UCP from which other mammalian and plant UCPs diverged. Speculations on UCP4 participation in apoptosis are thus supported by its early phylogenetic occurrence. © 2001 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Phylogenesis of mitochondrial uncoupling protein; Specific sequence motif; Drosophila melanogaster; Caenorhabditis elegans; Dictyostelium discoideum; Arabidopsis thaliana

1. Introduction

Annotation of fully sequenced genomes also requires a search for the inherent marks, represented by specific sequence motifs common to groups of functionally or phylogenetically related proteins [1]. Recently, a revolution in biosciences has begun by achieving a nearly complete sequence of human genome [2], Drosophila melanogaster, [3], Caenorhabditis elegans, [4] or plant Arabidopsis thaliana [5] genomes. Their detailed annotation (ascribing phenotypes to the revealed genes), namely in the case of human genome, will probably take years, however the first steps have already been taken [6, 7].

Concerning the mitochondrial anion carrier protein (MACP) gene family, its 43 members were found in Drosophila genome [3], 35 members with probably 28 phenotypes were reported in yeast [8–12, 14–16] and 32 in C. elegans [13]. To date 14 genes representing 10 phenotypes were already annotated in yeast [9, 11, 12]. For only five out of 10 phenotypes of yeast MACPs the human orthologs (ADP/ATP-, phosphate, dicarboxylate, citrate, and carnitine carriers) were previously recognized by cloning and reconstitution of the expressed proteins. Oxoglutarate carrier plus five uncoupling proteins most probably do not exist in yeast. MACPs are homologous proteins with threefold sequence repeat (yet imperfect) of about 100 amino acids [8–12, 14–16] forming six trans-membrane α-helices [8, 9, 15, 16]. On the interface between the odd α-helices and the proximal matrix segments, a unique specific sequence, the MACP signature exists [8–16]. These MACP signature sequences could predetermine the end of the odd α-helices and the beginning of the matrix segments in all MACPs [17]. Moreover, intra- and interdomain charge pairing between positive and negative charges of MACP signatures stabilizes two possible carrier conformations [9].

The mitochondrial uncoupling proteins form a distinct subfamily of the MACP gene family [17]. Although their phenotypes are known in detail only for brown adipose tissue-specific UCP1 [16, 18] and some plant UCPs (PUMP from potato [19, 20], tomato [21, 22], A. thaliana (Borecký, J., Ježek, P., et al., unpublished) and several other plant species [23, 24]), it is a general consensus that mammalian ubiquitous UCP2 [25, 26] and skeletal-muscle-specific UCP3 [27, 28] might participate in regulation of body weight (with dysfunctions leading to obesity), in adaptive thermogenic processes including fever and in reduction of the excessive formation of reactive oxygen species (ROS) [29]. Recently revealed brain-specific UCP4 has been suggested to participate in apoptosis in the brain [30]; this speculation could be also extended for another brain-specific UCP, termed BMCP [31] which appeared to be the most distant member within the UCP subfamily [17]. A putative purple nucleotide-binding domain (PNBD) was identified as the first common sequence motif unique for UCPs [32]. Further sequence analysis of UCPs [17] lead us to define their common sequence motifs that do not exist in any other MACPs (namely those with sequences available up to 1999, including the yeast MACP sequences), so-called UCP signatures in the first, second and fourth α-helix and in the second matrix (the last one is not valid for BMCP). Since mutagenesis of the first α-helical [33] and second matrix segment [34] has led to the reduced FA-induced H+-transport, we suggested that the corresponding UCP signatures are prerequisites for putative FA-binding/translocation sites [17]. The second UCP signature and parts of the fourth α-helical and matrix UCP signatures also exist in the yeast dicarboxylate transporter [17], probably representing the carboxyl group reactive domains.

In this work, we used knowledge of UCP signatures [17] to search for uncoupling proteins or their closest relatives in the available genome databases. We concluded that UCP4 could
## Analysis of UCP signatures and PBDN motifs in predicted uncoupling proteins of D. melanogaster, C. elegans, D. discoideum and A. thaliana

The first column indicates a UCP-type; the second column lists species; the third one shows the suggested abbreviations; and the fourth column displays one of their accession numbers. The previously defined UCP signatures [17] are written at the top below the headings; their extensions defined in this work are listed below them in bold in a dotted background. The UCP-specific sequence ('signature' for all UCPs but BMCP) in the second matrix segment is shown with preceding residues to illustrate distribution of positive charges in this segment. BMCP-specific features are indicated by the dotted background; PUMP-specific features by white fonts in the black background; UCP4-specific features by 45° hatching and UCP4 and BMCP common features by 135° hatching. The gray background represents features common to more UCPs. Question marks indicate irregularities in UCP signatures of the predicted D. discoideum UCP4. Symbols represent the following items: n, neutral non-aromatic residue including M; f, aromatic residue; '+' or '−', positive or negative charged residues, respectively; OH stands for S or T; NH stands for N or Q; and, p represents S, T, N and Q. Stars depict the residues well conserved in the MACP family members (up to 10 exceptions); exclamation marks refer to the ‘quite conserved’ residues [17]. The trans-membrane regions are depicted according to Kligenberg [16]. Although various lengths of β-helices were predicted [8], we prefer the model with shorter odd and longer even helices [16] since even its longer helices form structures smaller in length than the ~40 Å membrane width [17].

### UCP4 & BMCP feature

<table>
<thead>
<tr>
<th>Original signatures</th>
<th>2nd helix UCP signature</th>
<th>2nd matrix UCP signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACG - n</td>
<td>G</td>
<td>+</td>
</tr>
</tbody>
</table>

### BMCP feature

<table>
<thead>
<tr>
<th>Original signatures</th>
<th>2nd helix UCP signature</th>
<th>2nd matrix UCP signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACG - S</td>
<td>G</td>
<td>+</td>
</tr>
</tbody>
</table>

### Feature common to more UCP types

<table>
<thead>
<tr>
<th>Original signatures</th>
<th>2nd helix UCP signature</th>
<th>2nd matrix UCP signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACG - S</td>
<td>G</td>
<td>+</td>
</tr>
</tbody>
</table>

---

**Fig. 1.** Analysis of UCP signatures and PBDN motifs in predicted uncoupling proteins of *D. melanogaster, C. elegans, D. discoideum* and *A. thaliana*. The first column indicates a UCP-type; the second column lists species; the third one shows the suggested abbreviations; and the fourth column displays one of their accession numbers. The previously defined UCP signatures [17] are written at the top below the headings; their extensions defined in this work are listed below them in bold in a dotted background. The UCP-specific sequence (‘signature’ for all UCPs but BMCP) in the second matrix segment is shown with preceding residues to illustrate distribution of positive charges in this segment. BMCP-specific features are indicated by the dotted background; PUMP-specific features by white fonts in the black background; UCP4-specific features by 45° hatching and UCP4 and BMCP common features by 135° hatching. The gray background represents features common to more UCPs. Question marks indicate irregularities in UCP signatures of the predicted *D. discoideum* UCP4. Symbols represent the following items: n, neutral non-aromatic residue including M; f, aromatic residue; ‘+’ or ‘−’, positive or negative charged residues, respectively; OH stands for S or T; NH stands for N or Q; and, p represents S, T, N and Q. Stars depict the residues well conserved in the MACP family members (up to 10 exceptions); exclamation marks refer to the ‘quite conserved’ residues [17]. The trans-membrane regions are depicted according to Kligenberg [16]. Although various lengths of β-helices were predicted [8], we prefer the model with shorter odd and longer even helices [16] since even its longer helices form structures smaller in length than the ~40 Å membrane width [17].
represent an ancestral uncoupling protein from which the other UCPs diverged during phylogenesis.

2. Materials and methods

We have referred to accessible databases such as Fly Base (www.fruitfly.org/blast/); DicytDB (dicyt.sds.edu/); C. elegans Blast Server (www.sanger.ac.uk/Project/C elegans/blast server.shtml); or through SRK6 hub (sr6s.ebi.ac.uk) to various commonly used protein or nucleic acid databases. The core reference UCP sequences were those of hamster UCP1 (P04575 in SwissProt, SP, if not stated otherwise) and human UCP1 (SP P25874); human UCP2 (GenBank accession number U76367; SP P55851); human UCP3 (SP P55916); human UCP4 (NM040277.1 or SP O95587); BMCP1 (AL035423.4 or SP O95258) and PUMPs (U763120) [35] and UCP1 [36] (EMBL AJ239838) and UCP2 [37] (EMBL AB021706). Two fish UCP2 sequences [38] and skunk cabbage (Symplocarpus foetidus) UCPs (PUMP) [39] (EMBL AB024733) were also taken into account. For invertebrate and other newly annotated sequences see Section 3. Our computer alignment was performed using the Clustal method (Megalign program of the Lasergene 99 sequence analysis system), applying the Dayhoff PAM 250 matrix. Function ‘percent similarity’ of the Megalign program was employed to assess similarities between various segments of sequences. In the case of final phylogenetic tree construction from already annotated sequences the Jotun Hein method was also employed.

3. Results

3.1. Possible uncoupling proteins in D. melanogaster

A protein AE003506 (accession number in GADFly CG6492) clustered as the closest protein to human UCP4 (52.3% homology) in a constructed homology-based phylogenetic tree (not shown) of all 43 MACP members found in the constructed homology-based phylogenetic tree of all 43 MACP members found in the (52.3% homology) in a constructed homology-based phylogenetic tree (not shown) of all 43 MACP members found in the Drosophila together with human MACP carriers of known phenotype and all five human UCPs (known up to 1999), potato PUMP (StUCP), two other Arabidopsis PUMPs, plus all known yeast MACPs. In a cluster together with the pair of AE003506 and hUCP4, another pair of Drosophila MACPs was located – AE003487 (i.e. GADFly CG18340) and AE003612 (GADFly CG9064). The closest to this quartet was located ^ AE003487 (i.e. GADFly CG18340) and AE003612 (GADFly CG9064). The closest to this quartet was located ^ AE003487 (i.e. GADFly CG18340) and AE003612 (GADFly CG9064).

In support of a working hypothesis that the Drosophila MACPs are uncoupling proteins, we performed their sequence alignment with human UCPs, two known fish UCP2 [38] and five PUMP sequences, that fit with Klingen’s model of trans-membrane folding (shorter odd and longer even helices) and inspected whether these putative Drosophila UCPs contain the previously identified UCP signatures. As Fig. 1 (and supplementary material on the web, http://www.elsevier.nl/PII/S0014579301023389) shows in detail, all four inspected Drosophila proteins contain all \( \alpha \)-helical UCP signatures in somewhat more general form as well as the PNBD motif defined previously by Bouillaud et al. [32], that also required an extended definition.

Our signature analysis has shown that all Drosophila UCP-like proteins contain the PNBD motif of the UCP4-type, except of the CG7314 protein which contains the PNBD motif of a type existing in BMCP and other UCPs: it contains asparagine as the second residue after proline and phenylalanine as the fourth residue before the negative charge. Similarly to BMCP, the CG7314 protein lacked a unique sequence motif common to all human UCPs but BMCP. This motif resides in the second matrix segment. That is why, in accordance with the obtained phylogenetic tree, we consider CG7314 (AE003544) to be the Drosophila BMCP analog. Its first \( \alpha \)-helical UCP signature contains serine as the first residue (other UCPs have alanine) that appears to switch its position with the second residue that always contains an SH, OH or NH2 group, but isoleucine is here in BMCPs (Fig. 1). A BMCP feature is also represented by the existence of phenylalanine after the negative charge (Fig. 1). BMCP-specific features can be also found in the second \( \alpha \)-helical UCP signature (neutral leucine as the third residue) and the fourth \( \alpha \)-helical UCP signature (threonine after proline; neutral residue as the 10th residue; and the second proline as the last but one residue).

Both, the phylogenetic tree as well as UCP signature analysis ascribe the CG6492 protein (AE003506) as the Drosophila UCP4 analog (thereafter termed DmUCP4A). It contains all three \( \alpha \)-helical UCP signatures, the PNBD motif, and characteristic arginine in the second matrix segment (Fig. 1). It does not contain a positive charge pair as UCP2, UCP3, or PUMPs (or such pair with inserted proline as in UCP1; or proline plus one residue as in hUCP4) in the second matrix segment. Instead it contains the positive charge triplet similarly as human UCP4. Also characters of UCP signatures in CG18340 (AE003487) and CG9064 suggest that these proteins are the closest UCP4 relatives – hereafter termed DmUCP4B and DmUCP4C. They possess the characteristic arginine in the second matrix segment and the first \( \alpha \)-helical UCP signature lacking threonine before the last aromatic residue (neutral in PUMPs). However, its function (if essential) could be substituted by the existing cysteine in the preceding position in DmUCP4C and perhaps by the third serine in the signature in DmUCP4B. Nevertheless, the essential negative residue, the neutralization of which leads to the lack of uncoupling function [33], is present in both of them.

The annotation of DmUCP4A, B, C as the closest human UCP4 analogs is also supported by the other UCP4-specific features present. Thus, after the first positive charge (arginine) in the second \( \alpha \)-helical UCP signature, there is histidine (except of CG9064) characteristic for UCP4, whereas other UCPs contain asparagine (Fig. 1). Also serine, three residues apart, belongs to the UCP4 features. Since CG9064 (DmUCP4C) does not have these features, but does not contain the BMCP-specific proline at the end (last but one) of the fourth \( \alpha \)-helical UCP signature (Fig. 1), it represents a transition between the UCP4 and BMCP proteins. In conclusion, we predict that the considered Drosophila proteins could function as uncoupling proteins, since they are the closest UCP4 (or BMCP) relatives among all Drosophila MACPs. Consequently, we have extended definitions of the originally written UCP signatures [17] and PNBD [32] to define the extended UCP signatures which include the new features of Drosophila (and other invertebrate or fish) proteins. The new UCP signatures and PNBD are defined in Fig. 1.

3.2. Possible uncoupling proteins in other invertebrates and in Protista

Inspecting other available genomes for invertebrates, we found a protein AF 003384 (from a cosmid clone K07B1) from C. elegans as the closest human UCP4 analog (Fig. 1,
44.3% homology). It contains all three α-helical UCP signatures and PNBD with all UCP4 features described above; a BMCP-like feature in the first α-helical UCP signature (neutral residue in its second position); and RYTG sequence (the same as in hUCP1) as a part of the second matrix UCP signature. The positive charge (lysine) is three residues apart from another one (arginine), not two as in hUCP4. Its PNBD motif contains a pair of negative charges (aspartate), exceptional among UCPs.

In the genome of an amoeba living in soil, Dictyostelium discoideum, that is ascribed to Protista, but formerly in Fungi, we found another predicted uncoupling-like protein. Its DNA fragments were composed of clones JAXx47407.1 and II-CAP1D15012, but the sequence is still incomplete. That is why we cannot say whether it contains the first α-helical UCP signature and the first MACP family signature, since this part of the sequence is missing. This protein contains mixed features of human UCP4 and BMCP (35.6% homology to both of them). It contains the BMCP-specific proline in the fourth α-helical UCP signature, but like UCP4 it contains nearly the whole second matrix UCP signature (there is the exceptional serine at the sixth position after the central positive charge). Because of this and because of the serine existing one residue apart from proline in the PNBD domain, we ascribed this protein to the UCP4 class of UCPs. However, the phylogenetic tree constructed for all 19 proteins of Fig. 1 placed this protein as a separate branch of the BMCP cluster (Fig. 2). The available sequence of this protein terminates in the middle of the PNBD domain, but contains the most critical residues in this part. Nevertheless, the available sequence of this predicted UCP does not contain the second α-helical UCP signature (there is an exceptional negative charge after arginine) or the fourth α-helical UCP signature (no negative charge). Hence, when mutational analysis would assign these residues as essential for the uncoupling function, one should exclude the possibility that this protein is an uncoupling protein. At the present state of knowledge it is a possible candidate.

3.3. Possible uncoupling proteins in A. thaliana

Inspecting the plant A. thaliana genome databases we found altogether three distinct sequences that could be annotated as uncoupling-like proteins. The first two sequences coincide with the originally reported PUMP1 (AtPUMP1) cloned from a flower gene library of A. thaliana [36] and with PUMP2 (AtPUCP2) described by Watanabe et al. [37]. However, the third protein, AC007576.24 (or F7A19.22) contains features characteristic for human UCP4 (Fig. 1) and is rather distant from the above Arabidopsis PUMPs and from potato and Siumlocarpus PUMPs. This protein (hereafter called AtUCP4) might therefore represent an evolutionary link to the most ancestral UCP4-like protein. It does not contain a specific PUMP sequence MGDS which follows the third MACP family signature. As in UCP4, the last residue of the first α-helical UCP signature in AtUCP4 is phenylalanine; the residue following arginine in the second α-helical UCP signature is histidine and three residues further there is threonine (serine in hUCP4); and it also contains a PNBD motif of the UCP4-type. The latter is characteristic (Fig. 1) by proline preceding tryptophan, by phenylalanine three residues after this tryptophan and by lysine (instead of asparagine) following glutamate. In the resulting overall phylogenetic tree (Jotun Hein method, Fig. 2) Arabidopsis PUMP1 clusters together with the ‘classic’ potato PUMP1. When using the Clustal method, another PUMP also clustered together with these PUMP1 proteins – skunk cabbage PUMP (S. foetidus) SUC-Pa [39], hitherto called tentatively S/PUMP1. With the Jotun Hein method it clusters rather with the separate PUMP branch, to which a close relative, AtPUMP2, belongs. Nevertheless with both methods AtUCP4 clusters with human UCP4 and other UCP4-related proteins.

PUMPs but not AtUCP4 have some specificities in their UCP signatures (Fig. 1). Thus in the first α-helical UCP signature, the last residue is a neutral one (preceding the first proline of the MACP family signature). In the second α-helical UCP signature, PUMPs contain histidine preceding the conserved arginine and PUMP-specific cysteine, one residue apart from this arginine. In the fourth α-helical UCP signature, there is PUMP-specific asparagine following the first arginine. Their second matrix UCP signature resembles that one found in human UCP2 and UCP3. PNBD domains in PUMPs are similar to other UCPs with exception of the PUMP-specific neutral residue substituting aromatic residues in other UCPs that precedes the negative charge (probably the first residue of the fourth cytosolic segment). Availability of the other plant genomes will either confirm or extend the above described and predicted features of PUMPs.

4. Discussion

We have demonstrated that the previously defined UCP-specific sequence motifs, UCP signatures [17] and the PNBD motif [32], are valid in slightly extended forms for some MACP family members in Protista, invertebrates and A. thaliana. Consequently, we predict that these proteins represent the newly identified uncoupling proteins. Thus, we have shown the usefulness of the definition of UCP signatures (or PNBD) in searching for the UCPn orthologs in various genomes and we might speculate that similar unique specific
sequence motifs could be found for the other proteins (phenotypes) of the MACP gene family.

As the main conclusion from the obtained analysis we suggest that UCP4 most probably represents the ancestral UCP-type from which the other invertebrate, mammalian and PUMPs diverged. Speculations on UCP4 [30] or BMCP par- type from which the other invertebrate, mammalian and PUMPs diverged. Speculations on UCP4 [30] or BMCP par-

Acknowledgements: This work has been supported by a Grant of the Academy of Science Internal Grant Agency No. A5011106. Fig. 1 and annex figure design by Eva Urbánková and Fig. 2 design by Jan Jezek is gratefully acknowledged.

References