

11281

16



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE FISIOLÓGIA CELULAR
DOCTORADO EN CIENCIAS BIOMÉDICAS

Transferencia de genes mitocondriales al núcleo.
Implicaciones sobre la evolución de las algas clorofíceas
y de los parásitos apicomplejos.

Tesis
que para obtener el grado de
Doctora en Ciencias
presenta
Biól. María Soledad Funes Argüello

TESIS CON
FALLA DE ORIGEN

Director de Tesis: Dr. Diego González Halphen



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Este trabajo se realizó en el laboratorio 325 norte del Departamento de Genética Molecular del Instituto de Fisiología Celular de la Universidad Nacional Autónoma de México, bajo la dirección del Dr. Diego González Halphen.

Parte de este trabajo se realizó en el Departamento de Bioquímica y Farmacología Molecular de la Universidad Thomas Jefferson en Filadelfia, U.S.A. bajo la dirección del Dr. Michael P. King y del Dr. Edgar Davidson.

TESIS CON
FALLA DE ORIGEN

Este trabajo ha sido apoyado por el Consejo Nacional de Ciencia y Tecnología (CONACyT, México) a través del donativo 27754N, por la Dirección General de Asuntos del Personal Académico (DGAPA, UNAM, México) a través del donativo IN202598 y por el Fogarty International Center (National Institutes of Health (NIH), USA) a través del donativo TW01176 (NHLBI, NIH, donativo HL59646).

Durante la realización de los estudios de doctorado se contó con una beca para estudios de posgrado del Consejo Nacional de Ciencia y Tecnología (CONACyT, México) con el número de registro 124821, un complemento de beca de la Dirección General de Estudios de Posgrado (DGEF, UNAM, México) y dos apoyos del Programa de Apoyo a los Estudios de Posgrado (PAEP, UNAM, México) con la clave 102322 y 202370.

TESIS CON
FALLA DE ORIGEN

Agradezco a los miembros del comité tutorial, a la Dra. Marietta Tuena Sangri, al Dr. Angel Zarain Herzberg y al Dr. José de Jesús García Trejo. Por el interés mostrado en el proyecto, por sus críticas y sugerencias que a lo largo de estos años me permitieron aprender cada día un poco mas.

A los miembros del jurado de examen, Dr. Mario Calcagno Montáns, Dr. Guillermo Dávila Ramos, Dr. Diego González Halphen, Dr. David Romero Camarena, Dr. Luis Servín González, Dr. Mario Soberón Chávez, y Dra. Marietta Tuena Sangri. Asi como al Dr. Angel Zarain Herzberg. Les agradezco profundamente el tiempo que dedicaron a la lectura de este manuscrito, y los valiosos comentarios y sugerencias que realizaron al trabajo.

A la Q.B.P. Miriam Vázquez Acevedo por la ayuda técnica brindada durante la realización de este proyecto.

A Juan Manuel Barbosa, Gerardo Coello e Ivett Rosas, miembros de la Unidad de Cómputo del Instituto de Fisiología Celular (UNAM) por su ayuda y apoyo constante.

A María Eugenia Araiza por toda la ayuda administrativa brindada durante el doctorado. Por hacer que la burocracia fuera mas amigable.

Durante la realización de este proyecto, se contó con la atinada ayuda de algunas personas que con su experiencia y entusiasmo permitieron realizar un análisis mas detallado de los resultados obtenidos y enriquecer el trabajo: El Dr. M. Gonzálo Claros de la Universidad de Málaga (España) llevó a cabo los análisis de hidrofobicidad de las proteínas ATP6; la Dra. Susana Magallón del Instituto de Biología (UNAM) los análisis filogenéticos de las proteínas COX II y el Dr. Pascal Herión Instituto de Investigaciones Biomédicas (UNAM) cedió el DNA de *T.gondii*;

TESIS CON
FALLA DE ORIGEN

i. Resumen

Los genomas mitocondriales codifican un conjunto limitado de polipéptidos que son componentes esenciales de los complejos enzimáticos membranales que participan en la fosforilación oxidativa. El presente trabajo aborda el estudio de varios genes mitocondriales que de manera atípica fueron transferidos al núcleo en un grupo reducido de organismos. Dichos genes, en particular *atp6* y *nd4L* de *Chlamydomonas reinhardtii*, sufrieron modificaciones estructurales importantes durante y después de dicha transferencia al núcleo, lo que les permitió adaptarse de manera funcional a su nuevo ambiente: el cambio en el uso de codones, la adquisición de secuencias que codifican para presecuencias mitocondriales, y la adquisición de promotores, señales de poliadenilación y de intrones. Asimismo, se encontró que las proteínas mitocondriales codificadas por los genes nucleares muestran propiedades de hidrofobicidad disminuida, en particular, en aquellas regiones transmembranales que no participan directamente en la función catalítica de la proteína. También se estudió la localización del gen *cox2* en otras algas clorofíceas. Se encontró que este gen se encuentra fragmentado en dos unidades en el alga *Scenedesmus obliquus*: el gen *cox2a* que se encuentra en el genoma mitocondrial, y el gen *cox2b* que fue transferido al núcleo y representa un estado intermedio en la evolución de este grupo de algas. Finalmente, se encontró la presencia de los genes *cox2a* y *cox2b* en los genomas nucleares de algunos parásitos apicomplejos. La presencia de estos genes equivalentes a los encontrados en las algas clorofíceas apoya fuertemente la hipótesis de que estos parásitos se originaron a partir de un evento de endosimbiosis secundaria con un alga miembro de la clase Chlorophyceae, y da por terminado el largo debate que se ha desarrollado en la literatura sobre el origen de este grupo de parásitos.

ii. Abstract

Mitochondrial genomes encode a limited set of polypeptides that are components of the membrane-embedded oligomeric complexes that participate in oxidative phosphorylation. The main goal of the present work was to analyze the atypical transfer to the nucleus of some mitochondrial genes in a particular group of organisms. The transfer of *atp6* and *nd4L* genes to the nucleus in *Chlamydomonas reinhardtii*, was accompanied by structural modifications that allowed the successful expression in the nuclear environment: changes in codon usage, the acquisition of sequences encoding mitochondrial targeting sequences, and the acquisition of nuclear promoters, polyadenylation signals and introns. The mean hydrophobicity of the nucleus-encoded mitochondrial proteins is diminished, particularly in those transmembrane stretches that are thought not to participate directly in the catalytic function of the protein. The localization of the *cox2* gene in other chlorophyceae algae was also studied. This gene is splitted into two different units in *Scenedesmus obliquus*: the *cox2a* gene is located in the mitochondrial genome while the *cox2b* gene has been transferred to the nucleus. This organism represents an intermediary stage in the evolution of this group of algae. Finally, the genes *cox2a* and *cox2b* were found in the nuclear genomes of some apicomplexan parasites. The presence of these genes equivalent to those of the chlorophycean algae strongly supports the hypothesis that these parasites were originated by a secondary endosymbiosis event with a member of the class Chlorophyceae. This observation gives an end to the long debate that has been taken place in the literature about the origin of this group of parasites.

iii. Abreviaturas utilizadas

ADP	difosfato de adenosina
ATP	trifosfato de adenosina
<i>atp6</i>	gen que codifica la subunidad ATP6
ATP6	subunidad <i>a</i> del complejo F ₀ F ₁ ATP sintasa
CYTb	citocromo <i>b</i> del complejo <i>bc</i> ₁
<i>cox2</i>	gen que codifica la subunidad COXII
COXII	subunidad II del complejo citocromo <i>c</i> oxidasa
<i>cox2a</i>	gen que codifica la subunidad COXIIA
COXIIA	subunidad IIA del complejo citocromo <i>c</i> oxidasa
<i>cox2b</i>	gen que codifica la subunidad COXIIB
COXIIB	subunidad IIB del complejo citocromo <i>c</i> oxidasa
<i>cox3</i>	gen que codifica la subunidad COXIII
COXIII	subunidad III del complejo citocromo <i>c</i> oxidasa
Cr -ATP6	subunidad <i>a</i> del complejo F ₀ F ₁ ATP sintasa de <i>C. reinhardtii</i>
DNA	siglas en inglés de ácido desoxirribonucleico
dNTPs	desoxinucleótidos de trifosfato
EST	siglas en inglés de etiqueta de una secuencia expresada
<H>	hidrofobicidad local
<i>mesoH</i>	mesohidrofobicidad
mtDNA	DNA mitocondrial
MTS	siglas en inglés de presecuencia mitocondrial (Mitochondrial Targeting Sequence)
NADH	nicotinamida adenina dinucleótido (forma reducida)
ORF	siglas en inglés de marco de lectura abierto
PCR	siglas en inglés de reacción en cadena de la polimerasa
SDS-PAGE	siglas en inglés de electroforesis en geles de poliacrilamida-dodecil sulfato de sodio
RNA	siglas en inglés de ácido ribonucleico
TIM-TOM	siglas en inglés de translocasas de la membrana interna y externa mitocondrial

Introducción

Las mitocondrias son organelos semiautónomos que se encuentran en la gran mayoría de las células eucariontes. Estos organelos, a pesar de haberse originado como entidades independientes, se encuentran estrechamente involucrados en el metabolismo celular: además de producir una gran cantidad de ATP, en los últimos años se han relacionado con procesos de muerte celular y envejecimiento, desórdenes clínicos e incluso obesidad (Scheffler, 1999).

La producción de energía se realiza fundamentalmente mediante el proceso llamado fosforilación oxidativa, en el cual electrones son transportados a través de una serie de complejos proteicos que constituyen la cadena respiratoria, generando de manera simultánea un gradiente electroquímico de protones que es aprovechado por la F_1F_0 ATP sintasa para sintetizar ATP a partir de ADP y fosfato (Saraste, 1999).

La estructura de las mitocondrias

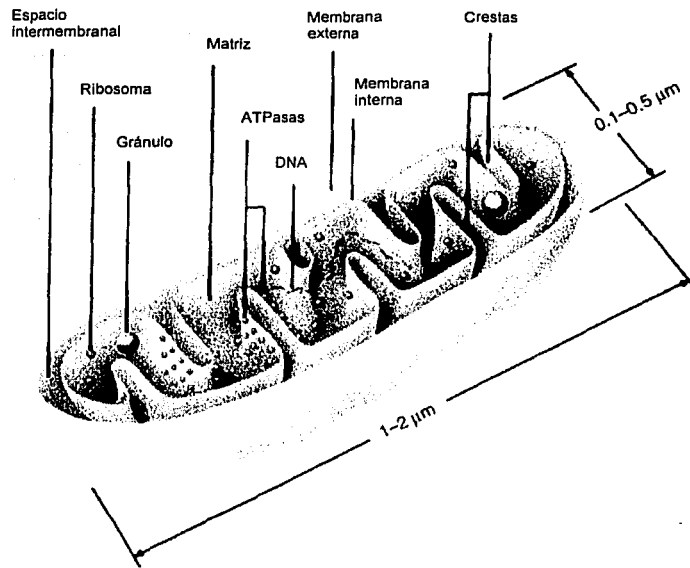
Las mitocondrias son organelos de forma ovalada cuyo tamaño, estructura interna y número por célula varía enormemente en diferentes tipos celulares. Están compuestas por dos membranas, una interna y una externa, que engloban una matriz densa que incluye enzimas del metabolismo intermediario y múltiples copias de un genoma que contiene algunos genes que codifican proteínas de la membrana interna, y algunos elementos necesarios para su traducción como RNAs ribosomales y de transferencia. Los otros cientos de proteínas mitocondriales se encuentran codificadas en el genoma nuclear y son sintetizadas en el citoplasma celular (Figura I.1A) (Frey y Mannella, 2000).

Estudios recientes de tomografía por microscopía electrónica han permitido hacer reconstrucciones tridimensionales de las mitocondrias, mostrando que aunque la membrana interna y las crestas son superficies continuas, éstas últimas no son estructuras estáticas abiertas por completo hacia el espacio intermembranal, sino que mas bien son cisternas polimórficas que usualmente presentan múltiples conexiones tubulares estrechas (de aproximadamente 28 nm de diámetro) hacia la superficie periférica de la membrana interna y hacia otras crestas (Figura I.1B) (Frey y Mannella, 2000).

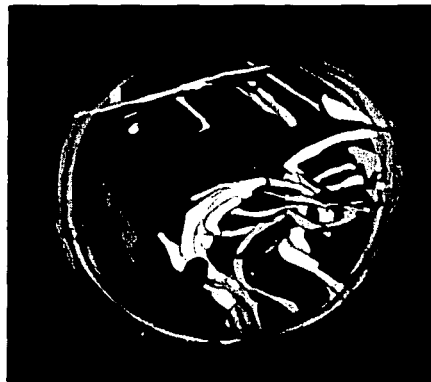
La mitocondria como unidad energética: la cadena respiratoria y la fosforilación oxidativa

La membrana interna de las mitocondrias eucariontes alberga un grupo de complejos proteicos que incluyen a diversos transportadores, a los componentes de la cadena respiratoria mitocondrial (o cadena de transporte de electrones) y a la ATP sintasa.

A



B



TESIS CON
FALLA DE ORIGEN

Figura I.1. Estructura general de las mitocondrias.

A. Visión esquemática de las mitocondrias (Tomada de Saraste, 1999).

B. Reconstrucción tridimensional de una mitocondria (Tomada de Frey y Manella, 2000).

La cadena respiratoria mitocondrial, está compuesta principalmente de proteínas integrales de membrana con grupos prostéticos capaces de aceptar y donar electrones. Cada componente de la cadena puede aceptar electrones del transportador precedente y transferirlos al siguiente en una secuencia específica determinada por los potenciales redox. En la reacción global catalizada se transportan electrones desde el NADH, el succinato u otro donador de electrones primario, a través de flavoproteínas, ubiquinona, proteínas ferrosulfuradas y citocromos, hasta el O₂.

La cadena respiratoria mitocondrial se compone de cuatro complejos enzimáticos: la NADH:ubiquinona oxidorreductasa (complejo I), la succinato:ubiquinona oxidorreductasa (complejo II), la ubiquinona:citocromo *c* oxidorreductasa (complejo III o citocromo *bc₁*) y la citocromo *c* oxidasa (complejo IV). Los complejos I, III y IV funcionan como bombas de protones, utilizando la energía de la transferencia de electrones para realizar trabajo electroquímico y translocar protones desde la matriz mitocondrial hasta el espacio intermembranal. Este trabajo produce y mantiene un gradiente electroquímico de protones a través de la membrana interna mitocondrial. Bajo condiciones fisiológicas normales, el complejo F₁F₀ ATP sintasa (complejo V) utiliza la energía libre almacenada guiando el flujo de protones nuevamente a la matriz mitocondrial para producir ATP a partir de ADP y fosfato. Este conjunto de reacciones se conoce como fosforilación oxidativa (Figura 1.2) (Schultz y Chan, 2001).

El complejo I o NADH:ubiquinona oxidorreductasa, es el complejo de mayor tamaño. En el complejo de mitocondrias de mamíferos se encuentran 43 subunidades diferentes y varios componentes que pueden funcionar como centros de óxido-reducción (flavín mononucleótido, entre siete y nueve centros FeS y ubiquinona fuertemente unida). Este complejo transporta dos electrones desde el NADH, formado durante la glucólisis y el ciclo de los ácidos tricarboxílicos, hasta una molécula de ubiquinona. De manera acoplada al transporte de electrones, sucede una translocación de dos protones por cada electron transferido a través del complejo (Vinogradov, 2001; Schultz y Chan, 2001).

El complejo II o succinato:ubiquinona oxidoreductasa es un componente tanto del ciclo de los ácidos tricarboxílicos como de la cadena respiratoria. El complejo de mitocondrias de mamífero se compone de 4 subunidades, se mantiene anclado a la membrana interna por un citocromo tipo *b*, contiene FAD y varios centros FeS. Este complejo oxida succinato a fumarato y transfiere electrones del succinato a la ubiquinona. Este complejo no es una bomba de protones, pero funciona en el ciclo de Krebs y como punto de entrada de electrones a la cadena respiratoria (Schultz y Chan, 2001; Saraste, 1999).

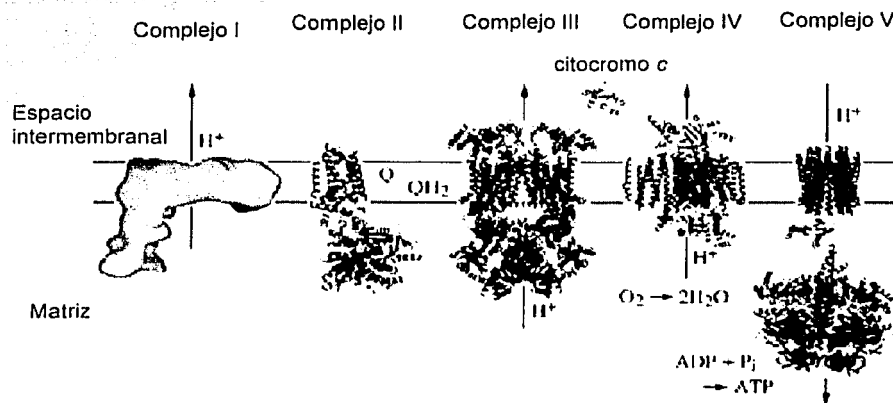


Figura 1.2. Complejos enzimáticos mitocondriales involucrados en la fosforilación oxidativa (Tomado de Schultz y Chan, 2000).

TESIS CON
FALLA DE ORIGEN

El complejo III o citocromo *bc*, transfiere los electrones desde la ubiquinona hasta el citocromo *c* soluble. Este transporte está acoplado a la translocación de protones a través de la membrana interna mediante lo que se conoce como el ciclo Q. El complejo de mamíferos contiene 11 subunidades, pero solo tres de ellas contienen los centros redox: el citocromo *b* (que contiene dos grupos hemo), la proteína Fe-S (que contiene un centro tipo Rieske) y un citocromo *c*, (Saraste, 1999).

El complejo IV o citocromo *c* oxidasa, realiza la última transferencia de electrones de la cadena respiratoria. Recibe los electrones del citocromo *c* soluble y reduce al oxígeno molecular generando agua. De manera acoplada a esta reacción transloca protones hacia el espacio intermembranal. Dentro de su estructura se localizan dos grupos redox asociados, los citocromos tipo *a* y *a₃*; además, se localizan dos centros de cobre (Cu_A y Cu_B) que son de gran importancia para la transferencia de electrones hasta el O_2 (Malatesta y col., 1995).

El empleo de la energía del transporte de electrones permite el bombeo de protones de la matriz mitocondrial al espacio intermembranal por los complejos I, III y IV. Esto genera un gradiente electroquímico de protones que es empleado por la F_0F_1 ATP sintasa (complejo V) para transformar la energía del gradiente electroquímico en energía química de enlace para la síntesis de ATP a partir de ADP y Pi (Mitchell, 1961).

La citocromo *c* oxidasa

La citocromo *c* oxidasa es el complejo terminal de la cadena de transporte de electrones. Este complejo enzimático se encarga de reducir al oxígeno molecular y transformarlo en agua a partir de los electrones que recibe del citocromo *c* soluble presente en el espacio intermembranal. Los cuatro protones que son necesarios para la reducción de una molécula de oxígeno son tomados de la matriz mitocondrial. De manera acoplada, se realiza la translocación de cuatro protones adicionales hasta el espacio intermembranal contribuyendo a la formación del gradiente electroquímico de protones (Schultz y Chan, 2001).

La reacción catalizada puede resumirse como:



Los electrones donados por el citocromo *c* entran al complejo mediante un triptofano conservado que se localiza en la subunidad II (COX II), y son transferidos hasta el oxígeno molecular a través de los cuatro centros de óxido reducción. El centro Cu_A se localiza en el dominio hidrofóbico de COX II y constituye el primer y único sitio aceptor del complejo. El electrón se

transfiere al hemo tipo *a* que se encuentra en la subunidad I (COX I) y posteriormente al centro binuclear de la enzima (compuesto por un hemo tipo a_3 y un Cu_B acoplado). Este centro binuclear se encuentra embebido en la membrana y es el sitio de unión de oxígeno y de formación de la molécula de agua (Ludwig y col., 2001).

El mecanismo a través del cual se lleva a cabo la translocación de protones por la citocromo *c* oxidasa no ha sido descifrado por completo. Hasta el momento se han realizado numerosos estudios cinéticos y de mutagénesis dirigida que parecen descubrir ciertos pasos, sobretodo de cómo se captan los protones de la matriz mitocondrial. Se ha planteado que existen dos canales hidrofílicos en COX I que reciben el nombre de los aminoácidos conservados que se encuentran en la entrada de cada uno de ellos: lisina (canal K) y aspártico (canal D); el canal K conecta la matriz con el centro binuclear y el canal D atravieza por completo el complejo (y por lo tanto la membrana). En el fondo del canal D existe un glutamato conservado que es esencial para el bombeo de protones (Wikström, 1998). Se ha planteado que el canal K permite la entrada de los cuatro protones necesarios para la formación de agua, mientras que el canal D lleva a cabo la translocación de protones a través de la membrana. Se ha observado también que los protones que utilizan el canal D pueden emplearse en cualquiera de los dos procesos, por lo que se ha planteado la necesidad de cierto mecanismo que sea capaz de distinguir a qué proceso están destinados los protones. El camino que siguen los protones y los cambios estructurales que puede estar sufriendo el complejo es motivo de estudio intenso (Ludwig y col., 2001; Saraste, 1999; Shultz y Chan, 2001).

En todos los organismos aerobios (tanto bacterias como eucariontes) la composición mínima de este complejo es de tres subunidades: COX I, COX II y COX III (que presentan 12, 2 y 7 cruces transmembranales respectivamente). A partir de la estructura cristalográfica de los complejos de la bacteria *Paracoccus denitrificans* (Figura 1.3) (Iwata y col., 1995) y de bovino (Tsukihara y col., 1996), se sabe que estas tres subunidades se encuentran bastante conservadas tanto en la secuencia de aminoácidos como en las estructuras terciaria y cuaternaria que adoptan en el complejo. El número de subunidades adicionales que se encuentran presentes en diferentes organismos es muy variable: en bacterias a veces se encuentra una subunidad extra, en levaduras 6 y en mamíferos 10 (Scheffler, 1999).

Como ya se mencionó anteriormente, COX I y COX II son las subunidades directamente responsables de guiar las reacciones de óxido reducción; el papel funcional de COX III no ha sido completamente elucidado, sin embargo se ha relacionado con la biogénesis del complejo (Haltia y col., 1991) y con el transporte del oxígeno molecular hasta el sitio catalítico de COX I (Riistama y col., 1996). El resto de las subunidades, se ha planteado que podrían estar involucradas en la regulación fisiológica del complejo dentro de diferentes estados del metabolismo energético de la



Figura I.3. Estructura cristalográfica del complejo citocromo c oxidasa de *Paracoccus denitrificans* (Iwata, y col., 1995).

TESIS CON
FALLA DE ORIGEN

célula (Ludwig y col., 2001). Así mismo, algunas de estas subunidades supernumerarias son específicas de cada tejido (Kadenbach y col., 1983).

En la mayoría de los organismos eucariontes, las subunidades COX I, COX II y COX III, se encuentran codificadas en el genoma mitocondrial, mientras que el resto de las subunidades son sintetizadas en el citoplasma celular e importadas posteriormente al organelo (Lang y col., 1999).

La F_1F_0 ATP sintasa/ATPasa

La ATP sintasa mitocondrial es un complejo enzimático funcionalmente reversible, capaz de sintetizar ATP utilizando la energía de un gradiente electroquímico de protones y de hidrolizar ATP para bombear protones en sentido inverso (Saraste, 1999).

Esta enzima se conforma por dos dominios principales: una porción transmembranal asimétrica, F_0 , que contiene el canal de protones; y una porción soluble, F_1 , que contiene los tres sitios catalíticos que funcionan cooperativamente en las reacciones de síntesis (Figura 1.4A). La región F_1 del complejo está compuesto por cinco subunidades α , β , γ , δ y ϵ . Las secuencias de las subunidades α y β son homólogas y ambas proteínas tienen un plegamiento similar; en el complejo activo se encuentran 3 subunidades α y 3 subunidades β localizadas de manera alternada a manera de segmentos de una naranja alrededor de un dominio α -hélice que contiene los extremos amino y carboxilo de la subunidad γ (Abrahams y col., 1994; Boyer, 1997). La composición de la región F_0 varía según la especie: el F_0 más simple se encuentra en *Escherichia coli* y está compuesta por tres subunidades: a , b y c (Senior y Wise, 1983); el F_0 más complejo es el de las mitocondrias de mamíferos con 9 subunidades (Boyer, 1997). El nombre de cada una de las subunidades del complejo ha sido objeto de numerosas confusiones, ya que subunidades homólogas han adquirido nombres distintos de acuerdo al organismo en el que se describieron, en la Cuadro 1.1 se ha hecho una correlación entre la nomenclatura de distintos organismos.

La estructura completa de la F_1F_0 ATP sintasa puede ser dividida en dos estructuras mecánicas con diferente función: un rotor y un estabilizador. La porción estabilizadora se forma por el hexámero $\alpha_3\beta_3$ junto con las subunidades a , b_2 y δ ; y el rotor se compone por las subunidades c conectadas a las subunidades γ y ϵ que forman un cuello central (Wang y Oster, 1998). Se piensa que el flujo de protones sucede a través de la interfase de las subunidades a y c de la porción F_0 . La rotación del múltiplo de subunidades c genera una torsión que es transmitida a la región F_1 a través de la subunidad γ . La rotación de γ se realiza en pasos discretos de 120° cada uno (con subpasos de 30° y 90°) y se encuentra acoplada a cambios conformacionales de las subunidades

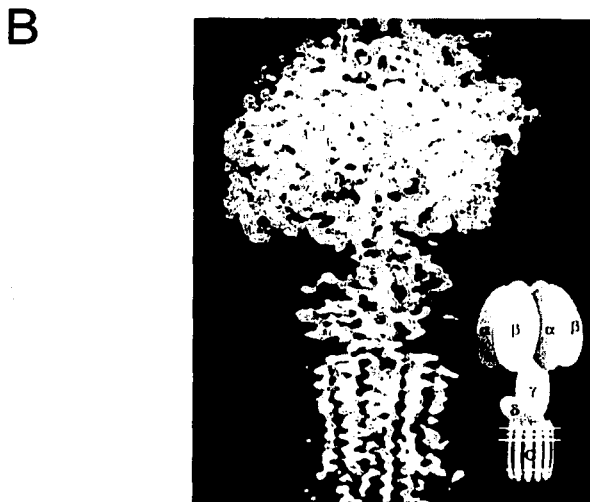
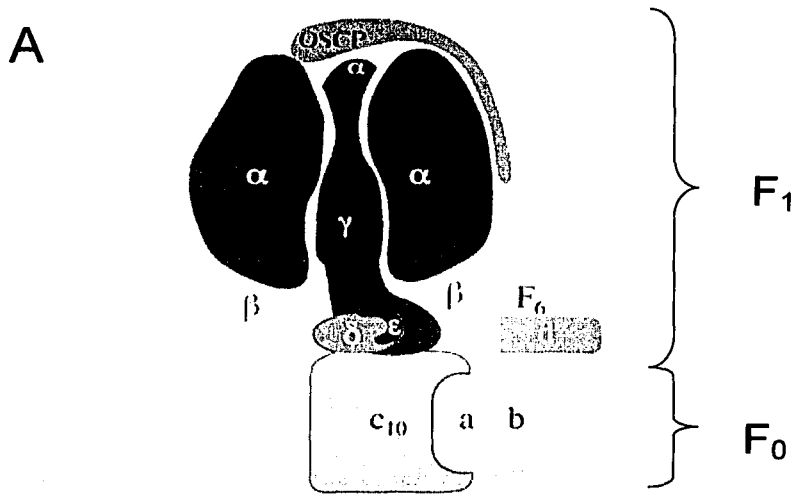


Figura 1.4. Estructura del complejo F₁F₀ ATPasa. (A) Modelo del complejo de mitocondrias de bovino. (B) Estructura cristalográfica parcial del complejo que muestra el sector F₁ y el oligómero de subunidades c del complejo de *Saccharomyces cerevisiae* (Tomados de Stock y col., 1999)

Cuadro I.1.
Comparación y nomenclatura de las subunidades del complejo V de varios organismos
(Modificado de Scheffler, 1999).

Bacterias (<i>E. coli</i>)	Mitocondrias animales	Mitocondrias de Levadura	Cloroplastos (y cianobacterias)
F ₁			
α	α	α	α
β	β	β	β
γ	γ	γ	γ
δ	OSCP	OSCP	δ
ε	δ	δ	ε
--	ε	ε	--
F ₀			
a	a o ATP6	ATP6	a ó IV
b	b	ATP4	b,b' ó I, II
c	c ó ATP9	ATP9	c ó III
--	d	d	--
--	e	e o TIM11	--
--	f	b	--
--	g	g	--
--	--	l	--
--	A6L	ATP8	--
--	F6	--	--

β : el estado catalítico de cada una de las subunidades β y por lo tanto su conformación local, depende del estado en que se encuentren los otros dos sitios y de la posición de γ , aunque puede decirse también que la posición de la subunidad γ depende de la estructura relativa que presenten las subunidades β . Se ha propuesto que este proceso sucede en una serie de tres estados por cada una de las subunidades β que varían conforme van cambiando los substratos unidos: abierto (con una alta afinidad), semicerrado (con una afinidad intermedia) y cerrado (con una baja afinidad). Al analizar la región F_1 completa resulta que todos los cambios sufridos por una de las subunidades β , afectan a las otras dos, por lo que el proceso se realiza a lo largo de seis estados (estos pasos se han descrito durante el proceso de hidrólisis, pero se ha planteado que durante la síntesis de ATP se siguen las mismas fases pero en sentido inverso). La fuerza que guía los cambios conformacionales en las subunidades catalíticas, proviene tanto de la translocación de protones como de la hidrólisis del ATP (Abrahams y col., 1994; Elston y col., 1998; Menz y col., 2001; Wang y Oster, 1998; Yasuda y col., 1998; Yasuda y col., 2001).

El subcomplejo F_0 de la bacteria *Escherichia coli* es el más sencillo que se ha caracterizado, constituyendo lo que parece ser la unidad mínima funcional para este sector. Tiene sólo tres subunidades diferentes con una estequiometría de $a_1b_2c_{9-12}$. Se piensa que el paso de protones de un lado a otro de la membrana sucede en la interfase entre las subunidad a y el conjunto de subunidades c (Jiang y Fillingame, 1998; Vik y col. 2000). La subunidad a es la más hidrofóbica y en *E. coli* se cree que adquiere una conformación de cinco cruces transmembranales en los cuales se localizan residuos de aminoácidos ácidos y básicos que son esenciales para la translocación de protones (Arg210, His245, Glu196, Glu219) (Valiyaveetil y Fillingame, 1998; Wada y col. 1999). La subunidad c es también hidrofóbica, tiene dos cruces transmembranales unidos por un asa hidrofílica. Se cree que, según el organismo y las condiciones fisiológicas en que se encuentre, en cada región F_0 existen entre 9 y 14 subunidades c ordenadas a manera de anillo. Se ha planteado que la translocación de protones está mediada por una protonación y desprotonación del grupo carboxilo del residuo Asp61 que se localiza en el centro del segundo cruce transmembranal de cada subunidad c . Hasta la fecha, sólo ha sido determinada la estructura del oligómero de subunidades c de *Saccharomyces cerevisiae*, mostrando que la región F_1 se encuentra asociado con un anillo de 10 subunidades c (Figura I.4B) (Girvin y col., 1998; Stock y col., 1999). La subunidad b tiene una porción citoplásmica y un dominio transmembranal mediante el cual se mantiene anclado a la membrana; el dominio citoplásmico es esencial para la unión entre la región F_1 y la subunidad a (Dunn y col., 1992; Jiang y Fillingame, 1998; Stock y col., 2000).

Además del papel esencial de la ATP sintasa en la producción de energía celular, se ha relacionado también con la morfología mitocondrial. Hace pocos años, se describió que el complejo F_1F_0 ATP sintasa de levadura y en mitocondrias de mamífero, existe como un dímero en la

membrana interna mitocondrial, y recientemente se descubrió la necesidad de que las subunidades responsables de esta dimerización estén presentes para que se formen las crestas mitocondriales a partir de la membrana interna (Arnold y col., 1998; Paumard y col., 2002; Schägger y Pfeiffer, 2000).

Las mitocondrias como unidad genética

La teoría del origen endosimbionte de las mitocondrias es la más aceptada sobre el origen de estos plástidos; en ella se propone que las mitocondrias son descendientes directos de un endosimbionte bacteriano que se estableció en el citoplasma de una célula huésped en una etapa temprana de la evolución (Margulis, 1970).

El fenómeno de simbiosis intracelular ha sido estudiado experimentalmente en varios sistemas, y se ha planteado que una endosimbiosis estable incluye cuatro pasos fundamentales: un reconocimiento inicial; la entrada de los simbioses a la célula hospedera; el que los simbioses eviten la digestión por la célula hospedera; y la división consecutiva de ambos elementos involucrados (revisado en Jeon, 1995). Sin embargo, no es claro aún el mecanismo a través del cual se llevó a cabo el proceso de endosimbiosis que dio origen a las mitocondrias actuales. Recientemente se propuso que los eucariontes surgieron a través de una asociación metabólica entre una α -proteobacteria fermentadora que generaba hidrógeno y bióxido de carbono como productos de desecho, y una archaea metanógena anaeróbica y autotrófica cuyo metabolismo dependía estrictamente del hidrógeno, explicando de la manera más convincente hasta ahora las presiones que favorecieron el establecimiento de la endosimbiosis (Martin y Müller, 1998).

En fenómenos de endosimbiosis espontáneas entre amibas y eubacterias que se acumulan en el interior de simbiosomas de diversos tamaños, se ha observado que las amibas se convierten en organismos completamente dependientes de la bacteria endosimbionte en un periodo de aproximadamente 18 meses (alrededor de 200 generaciones). En este tipo de simbiosis, se ha observado que rápidamente comienza un fenómeno de migración de proteínas desde la bacteria endosimbionte hacia el núcleo y citoplasma de la amiba hospedera (Jeon, 1995; Pak y Jeon, 1997). Se ha observado también que la amiba hospedera sufre modificaciones morfológicas en el núcleo inducidas por la entrada de la bacteria; además de la alteración en la expresión de algunos genes de mantenimiento celular que le permiten la sobrevivencia autónoma. La pérdida funcional de estos genes es contrarrestada por la importación de proteínas equivalentes producidas por la bacteria. No se ha observado directamente la migración de material genético desde una bacteria endosimbionte hasta el núcleo de su hospedero, sin embargo, se ha planteado que este debe ser uno de los primeros pasos en la estabilización de la endosimbiosis y que puede ser el mecanismo

responsable de la inactivación de genes en el núcleo hospedero, ya sea por la inserción de secuencias, la interrupción de genes, la inactivación en la expresión de ciertos genes o la estimulación de la delección de otros genes (Choi y col., 1997).

Uno de los elementos que permitieron sustentar la teoría endosimbionte sobre el origen de las mitocondrias fue el hecho de que se descubriera material genético en el interior de las mismas (Nass y Nass, 1963). Se ha propuesto que los genomas mitocondriales son el resultado de una reducción del genoma bacteriano del endosimbionte original. Las secuencias obtenidas de genes mitocondriales han permitido a los investigadores rastrear los orígenes de las mitocondrias hacia un ancestro común en el grupo de las α -proteobacterias; específicamente con miembros de la subdivisión de los rikettsiales a la cual pertenecen parásitos intracelulares obligados como los organismos de los géneros *Rickettsia*, *Anaplasma* y *Ehrlichia* (Lang y col., 1999).

La secuencia completa del genoma de *Rickettsia prowazekii* (Andersson y col. 1998) ha permitido identificar aquellos genes que pueden perderse fácilmente en cualquier organismo cuyo modo de vida sea el de parásito intracelular, ya que las funciones que cumplen pueden ser complementadas por las proteínas del hospedero: los genes involucrados en la biosíntesis de aminoácidos y de nucleótidos, de la glucólisis anaeróbica y de la regulación celular.

La comparación del genoma mitocondrial de *Reclinomonas americana* (el genoma mitocondrial más parecido a una bacteria) con el genoma completo de *Rickettsia prowazekii* (el genoma bacteriano más parecido a un genoma mitocondrial) han permitido identificar aquellos genes que se perdieron durante la evolución del genoma mitocondrial: algunos genes operacionales (aquellos involucrados en la biosíntesis de cofactores, el metabolismo de ácidos grasos y de fosfolípidos, la síntesis de la cubierta celular y de productos necesarios para la división celular) y genes de información (aquellos genes que dirigen la replicación, la transcripción y la traducción). La gran similitud que guardan algunos genes de *Reclinomonas* y *Rickettsia*, así como el arreglo espacial similar de estos genes en los dos genomas, apoyan fuertemente la idea de que un ancestro rikettsial dio origen a las mitocondrias (Andersson y col. 1998).

Los genomas mitocondriales

Los genomas mitocondriales presentan una gran diversidad en cuanto a su organización génica, mecanismos de expresión y los patrones de evolución que siguen en cada grupo de organismos. La mayoría de los genomas mitocondriales caracterizados hasta la fecha, son circulares, de la misma manera que los genomas bacterianos, sin embargo también existen los genomas lineales. El tamaño puede variar enormemente desde un poco menos de 6 kb en

Plasmodium falciparum hasta más de 300 kb en las plantas. Las diferencias tan grandes en los tamaños de las moléculas obedecen fundamentalmente a que en los genomas mitocondriales de plantas se encuentran muchas regiones no codificantes [que incluyen intrones, ORFs cuyo posible producto no se ha relacionado con ninguna función, pseudogenes y segmentos de DNA exógeno (de cloroplasto y nuclear)]. En el caso de las plantas, sólo entre el 10 y el 20 % de la secuencia representa genes estructurales (Lang y col., 1999).

A pesar de la enorme diversidad de los genomas mitocondriales, en general, sólo se encuentran codificadas algunas proteínas de los complejos de la cadena respiratoria y del complejo ATP sintasa (preferentemente aquellos genes que codifican proteínas altamente hidrofóbicas). También se encuentran algunos componentes del sistema de traducción mitocondrial como los RNAs ribosomales (tanto de la subunidad pequeña como de la subunidad grande), algunas de las proteínas ribosomales y en general algunos o todos los RNAs de transferencia. Aunque la contribución del DNA mitocondrial es esencial, la mayor parte de la información para la función y biogénesis mitocondrial se encuentra codificada en el genoma nuclear, por lo tanto, las proteínas y en algunos casos RNAs necesarios se importan desde el citoplasma (Gray y col. 1999).

Evolución de los genomas mitocondriales

A partir del conjunto de 1000 genes o más que debieron encontrarse en el mtDNA del ancestro premitocondrial, la tendencia general en los genomas mitocondriales ha sido a reducir el contenido génico (independientemente de los otros patrones evolutivos que cada grupo de organismos presenta). La mayor parte de ésta pérdida de material genético debe haber sucedido durante el establecimiento del proceso de endosimbiosis (Lang y col., 1999).

La transferencia física de genes desde la mitocondria hasta el núcleo es un proceso que puede ser dividido de manera teórica en varios pasos: en primer lugar, debe existir una disponibilidad del material genético, ya que en principio las unidades genéticas transferibles pueden incluir cualquier región del genoma mitocondrial (se ha propuesto que pueden ser fragmentos de DNA o moléculas de RNA) siempre y cuando se encuentren en número abundante, de manera que su ausencia no repercuta en el funcionamiento celular. La migración del material genético implica la salida de la mitocondria, que se ha planteado como resultado de rompimientos temporales de las membranas mitocondriales. Sin embargo, se ha observado también la presencia de RNAs de transferencia característicos del genoma nuclear en el interior de las mitocondrias, lo que indica que el transporte de ácidos nucleicos a través de las membranas mitocondriales sucede continuamente. El paso físico final del transporte de material genético es la entrada al núcleo, el cual ocurre fácilmente a través de los poros de la membrana nuclear. Una vez que el material genético ha entrado al núcleo, debe insertarse en el genoma nuclear y coordinar su regulación y expresión en el nuevo sitio. Una vez que el gen ha sido activado en el núcleo, el gen mitocondrial

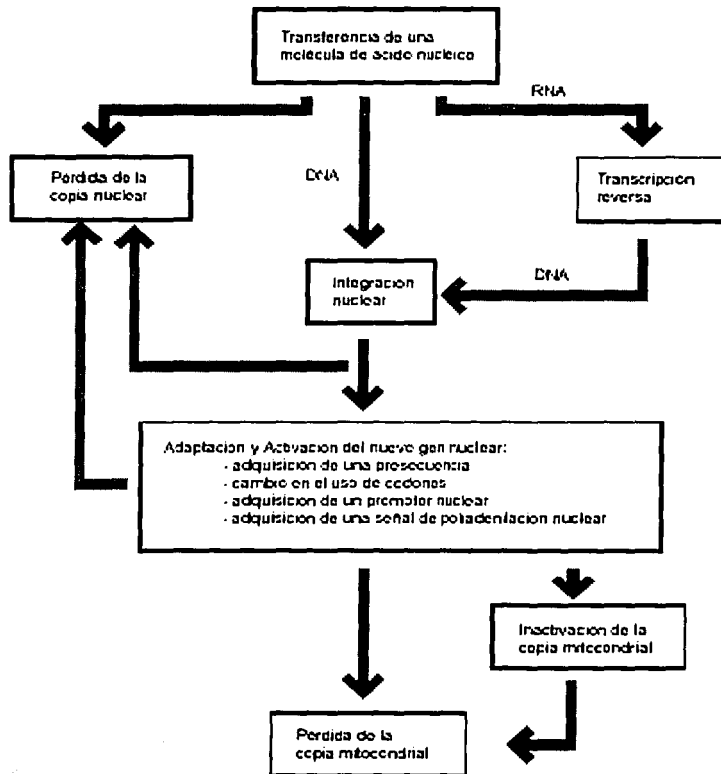


Figura I.5. Esquema general de la secuencia de eventos que suceden durante la transferencia de información genética desde la mitocondria al núcleo (Tomado de Brennicke y col. 1993).

deja de ser necesario y puede ser apagado después de un cierto período de tiempo durante el cual ambos genes son funcionales. Al no ser necesario, el gen mitocondrial va perdiendo algunos fragmentos hasta que finalmente es eliminado del genoma mitocondrial (Figura 1.5) (Brennicke y col., 1993).

Todavía no ha sido posible distinguir cuál es el paso limitante durante la transferencia de genes mitocondriales al núcleo: si la salida del DNA de la mitocondria o su integración en el genoma nuclear. Se han estudiado diversos modelos en los que aparentemente se observan etapas intermedias de la transferencia, lo que permite aclarar poco a poco cuál es el mecanismo más probable y los patrones que rigen este fenómeno.

a) Salida del material genético

El primer paso en el camino de la transferencia de una región del genoma mitocondrial hacia el genoma nuclear es la salida del material genético de la mitocondria. Se ha planteado que esta salida puede ser en forma de moléculas de DNA o de RNA. Las secuencias mitocondriales que se han identificado a lo largo de los genomas nucleares corresponden tanto a genes como a intrones, sugiriendo que en principio, cualquier región del genoma mitocondrial es susceptible de transferirse.

Las unidades genéticas que serán exportadas desde el interior de la mitocondria deben ser prescindibles, es decir que su ausencia funcional temporal no influya en el funcionamiento celular. Esta medida de seguridad se resuelve de manera muy sencilla gracias a la redundancia de organelos que existe en todas las células: en general se localizan varias mitocondrias por célula y cada una de ellas contiene un número variable de genomas. Estas dos características permiten que exista una fuente continua de moléculas de DNA disponibles para la translocación. La transferencia de material genético puede suceder también a partir de los transcritos mitocondriales, la presencia de un número abundante de moléculas de RNA de cualquier región transcrita asegura la abundancia y disponibilidad del material genético. En este tipo de transferencia puede suceder una preselección del material transferible restringiendo el fenómeno a las regiones del genoma mitocondrial que son funcionalmente importantes, además presenta la ventaja de que solo se transfiere la información madura (por ejemplo, se eliminan los intrones) (Brennicke y col., 1993).

En algunos hongos como *Saccharomyces cerevisiae* se ha observado que el fenómeno de transferencia de genes mitocondriales al núcleo sucede en una generación de organismos, por lo que han permitido realizar observaciones directas sobre el fenómeno de escape de DNA desde la mitocondria. Se ha demostrado que en la levadura *Saccharomyces cerevisiae* la transferencia de DNA mitocondrial al núcleo es un fenómeno intracelular dependiente del estado genético del

genoma mitocondrial y del estado metabólico de la célula, pero no de un intermediario de RNA (sin embargo la presencia de genes mitocondriales en el núcleo con modificaciones postranscripcionales indican que aunque no es necesario, el fenómeno de transferencia a partir de una molécula e RNA sí puede llegar a suceder). Las moléculas de DNA mitocondrial se comportan como plásmidos inestables de manera que pueden migrar sin perder su integridad (mientras más grande sea el plásmido más estable es, de manera que podría suceder la transferencia de todo el genoma en un solo evento (Wright y Cummings, 1983; Thorsness y Fox, 1990; Brennicke y col., 1993; Shafer y col., 1999).

En las levaduras, se ha identificado que el escape de DNA mitocondrial puede suceder a través de la formación de una vacuola que contenga el fragmento de DNA; estudios citológicos adicionales han mostrado que existen asociaciones físicas directas entre las membranas nuclear y mitocondrial, así como compartimentos mitocondriales encapsulados en el interior del núcleo (Campbell y Thorsness, 1998). Por otro lado se ha relacionado la transferencia de material genético con factores del medio ambiente que rodea a las células como cambios bruscos de temperatura o medios de cultivo con glicerol (Thorsness y Fox, 1990).

Existen algunos genes nucleares que son importantes en la transferencia de genes mitocondriales y que pueden, potencialmente, promover o suprimir la tasa de escape de DNA. Los supresores pueden actuar a dos niveles diferentes: impidiendo la propagación del DNA mitocondrial que se ha escapado o alterando la estabilidad, concentración o forma del DNA mitocondrial que es capaz de migrar hacia el núcleo (Thorsness y Fox, 1993; Shafer y col., 1999).

La velocidad de escape del DNA mitocondrial al núcleo sucede a una frecuencia de aproximadamente 2×10^{-5} fragmentos por célula por generación; mientras que la tasa a la que migra DNA nuclear hacia la mitocondria es 100,000 veces menor (Wright y Cummings, 1983; Thorsness y Fox, 1990).

Un ejemplo muy interesante es el del hongo *Podospora anserina*, en donde se ha reportado que el envejecimiento celular está relacionado con la migración de fragmentos del genoma mitocondrial hacia el núcleo. Se ha observado que fragmentos discretos de DNA mitocondrial son cortados del genoma y permanecen en el interior de las mitocondrias de individuos envejecidos formando plásmidos que se replican a una velocidad mas acelerada que el genoma mitocondrial de las células. De esta manera, al final de la etapa de envejecimiento los vectores constituyen la mayor parte del material genético presente. Una fracción de los plásmidos formados son transferidos al núcleo e integrados en el genoma nuclear (Wright y Cummings, 1983).

b) Inserción del material genético mitocondrial en el genoma nuclear

Una vez que el fragmento de DNA ha salido del genoma mitocondrial se ha propuesto que puede integrarse al genoma nuclear a través de varios mecanismos, los cuales han sido descritos al analizar la secuencia señal anterior al evento de inserción, la secuencia que se insertó y la secuencia resultante después de la inserción (Sun y Callis, 1993):

i) **Fusión en los extremos:** Terminaciones libres en los cromosomas nucleares pueden ser generados a partir de errores en el metabolismo del DNA y en estos sitios segmentos de DNA ajeno al nuclear pueden integrarse independientemente de sus secuencias terminales. La fusión se desarrollaría ligando los extremos del DNA mitocondrial utilizando una complementariedad limitada de extremos de una sola hebra de DNA (entre 0 y 3 pares de bases) (Blanchard y Schmidt, 1996).

ii) **Recombinación homóloga:** Se ha propuesto que los extremos del DNA mitocondrial que se ha transferido al núcleo existen regiones repetidas que presentan secuencias con un porcentaje alto de similitud con alguna secuencia nuclear, de manera que la inserción del fragmento de DNA exógeno sucede fácilmente por recombinación homóloga (Pichersky y col., 1991).

iii) **Recombinación no homóloga:** En este mecanismo, se propone que la inserción sucede a partir de la ligación de varios fragmentos no contiguos de DNA mitocondrial. Se piensa que fragmentos separados del genoma mitocondrial son ligados antes de la inserción por el mismo mecanismo que es capaz de ligar DNA introducido artificialmente y puede integrarse en uno o varios sitios. Este tipo de inserción sucede independientemente de la presencia de extremos compatibles para la ligación o de que los fragmentos de DNA se encuentren separados. Este mecanismo difiere de la recombinación homóloga ya que no es posible identificar secuencias repetidas o porcentajes importantes de identidad que favorezcan la recombinación (Sun y Callis, 1993).

En estos tres mecanismos se ha propuesto que la integración de los genes mitocondriales en el núcleo debe ser en regiones en las que no interfieran con la síntesis del resto de los genes nucleares, por lo que se ha propuesto que la inserción se favorece en regiones no codificantes como los intrones, regiones adyacentes a los genes nucleares o regiones teloméricas (Blanchard y Schmidt, 1996; Pichersky y Tanksley, 1988; Louis y Haber, 1991).

c) Activación del nuevo gen nuclear

Una vez que el gen originalmente mitocondrial se ha insertado en el genoma nuclear, puede moverse, duplicarse y modificarse de manera que adquiera los elementos necesarios para

que pueda expresarse de manera correcta. Entre las modificaciones que deben incorporarse al gen se incluyen: i) la adquisición de una presecuencia que codifique para un péptido señal que permita a la proteína entrar a la mitocondria una vez que fue sintetizada en el citoplasma; ii) la modificación del uso de codones (si es que era diferente); iii) la adquisición un nuevo promotor, iv) la adquisición de una señal que permita la poliadenilación y v) la adquisición las señales nucleares para la unión con ribosomas. Estas últimas modificaciones le permitirán al gen ser transcrito y traducido bajo los mismos mecanismos que rigen al resto de las proteínas que son sintetizadas en el citoplasma celular (Brennicke y col., 1993).

La proteína mitocondrial sintetizada ahora a partir de un gen nuclear debe ser dirigida de manera correcta hacia el interior de la mitocondria, por lo que debe adquirir una secuencia señal. Se ha planteado que las presecuencias pueden adquirirse como resultado de un reordenamiento de DNA que coloca, al azar, presecuencias mitocondriales potenciales en el extremo 5' de algunos marcos abiertos de lectura (Baker y Schatz, 1987); o por la modificación puntual de los extremos amino terminales de algunas proteínas que pueden generar una secuencia importadora funcional (Vassarotti y col., 1987), o por un mecanismo de duplicación y recombinación de presecuencias ya existentes (Kadowaki y col., 1996).

El éxito de la transferencia, depende de que el gen transferido al genoma nuclear sea activado, si por alguna razón no es posible realizar las modificaciones necesarias el gen nunca se activa en el núcleo y puede permanecer en el genoma nuclear como un pseudogene. Existen evidencias de que en el genoma nuclear de los primates se localizan pseudogenes de *cob* y en el caso particular de los seres humanos, se han identificado numerosos pseudogenes de *cox1* y *cox2*. Dichas secuencias se consideran "fósiles moleculares" que representan genes transferidos de la mitocondria al núcleo que nunca fueron activadas (Collura y Stewart, 1995; Wallace y col., 1997).

d) Pérdida del gen original

Durante un tiempo corto, el gen mitocondrial y el nuclear funcionan simultáneamente, sin embargo el gen mitocondrial ya no es necesario para el desarrollo celular y por lo tanto puede ser apagado y eliminado. La eliminación de un gen mitocondrial no es un proceso que suceda en un solo paso, sino que se trata de un proceso paulatino durante el cual se van perdiendo y editando fragmentos del gen mitocondrial hasta que se pierde por completo y por lo tanto la transcripción del gen desaparece. La eliminación de la copia mitocondrial, también puede suceder a nivel de traducción al introducir a la secuencia un codón de término que impida la síntesis correcta de la proteína en el interior de la mitocondria (Brennicke y col., 1993; Kadowaki y col. 1996).

En los últimos años el estudio de genes que han sido transferidos recientemente a los genomas nucleares, ha permitido identificar pasos intermedios del fenómeno de transferencia:

a) En el grupo de las plantas leguminosas es posible observar la coexistencia de dos copias diferentes del gen *cox2*. En un análisis amplio de diversas leguminosas se determinó que al menos en cinco géneros el gen *cox2* se encuentra presente tanto en el genoma mitocondrial como en el nuclear, que ambas copias son transcritas y correctamente editadas; sin embargo también en este grupo es posible observar pasos intermedios de la transferencia de este gen al núcleo: en el chícharo, se encuentra una copia del gen en la mitocondria y otra en el núcleo, sin embargo la copia nuclear no es funcional. En el caso de la soya, la copia mitocondrial se encuentra completamente intacta y parecería un gen funcional, sin embargo no es posible detectar ningún transcrito en la mitocondria, por lo que no se trata de un gen activo. En cambio, el gen nuclear es el único que se expresa, esto podría representar un paso temprano en el proceso de eliminación de genes redundantes. Finalmente, en el caso del frijol mungo, la copia mitocondrial ha sido alterada por mutaciones y pérdida de algunos de sus fragmentos, de manera que apenas es reconocible. (Covello y Gray 1992; Nugent y Palmer 1991, Adams y col. 1999).

b) El gen *rps12* (que codifica a la proteína ribosomal mitocondrial S12) en *Oenothera* ha sido transferido al genoma nuclear y aproximadamente dos terceras partes de la copia mitocondrial del gen se han eliminado (Grohmann y col. 1992).

c) En el grupo de las plantas angiospermas se ha identificado que el gen *rpl2* ha sufrido diversas modificaciones, y actualmente se puede encontrar en cuatro diferentes versiones: i) el gen completo se encuentra codificado de manera funcional en la mitocondria; ii) el gen completo se encuentra codificado de manera funcional en el núcleo; iii) el gen se encuentra fragmentado en dos, la región 5' se encuentra en la mitocondria y la región 3' se encuentra en el núcleo; iv) el gen se encuentra fragmentado en dos y ambas regiones se encuentran en el núcleo (Adams y col., 2001).

Inserción en la mitocondria de las proteínas sintetizadas en el citoplasma

Las proteínas mitocondriales, se sintetizan casi siempre como preproteínas citoplásmicas en las cuales es posible identificar una extensión en el extremo amino terminal que funciona como secuencia guía. Estas presecuencias pueden ser de tamaños muy diversos y con secuencias primarias muy diferentes; sin embargo, forman generalmente α -hélices anfífilicas ricas en residuos básicos o con cadenas laterales hidroxilo. Las presecuencias son procesadas por una peptidasa de

la matriz mitocondrial, de manera que en las proteínas maduras y funcionales no se encuentra ningún rastro de ellas (Hartl y col. 1989; Roise y Schatz, 1988).

El paso de las proteínas desde el citoplasma hasta su acoplamiento correcto en el interior de la mitocondria involucra una gran cantidad de elementos como chaperoninas citosólicas y de la matriz mitocondrial, receptores localizados en la superficie mitocondrial, proteínas que rodean los canales de paso y proteínas de maduración.

Entre las proteínas citosólicas que interaccionan con los precursores mitocondriales se han identificado a la proteína de choque térmico Hsp70, a un factor de estimulación de importación mitocondrial, a un supuesto factor de unión a presecuencias, y otros factores que se han identificado por su actividad estimuladora de importación de proteínas mitocondriales, pero que todavía no han sido completamente purificados y caracterizados. La Hsp70 y el factor de estimulación de importación mitocondrial son enzimas dependientes de ATP capaces de mantener cierta estructura secundaria de los precursores evitando que se agreguen adquiriendo conformaciones que serían incapaces de importarse. El factor promotor de importación mitocondrial es específico para presecuencias, mientras que Hsp70 parece unirse a regiones hidrofóbicas expuestas de las preproteínas (Mihara y Omura, 1996; Voos y col., 1999).

Los elementos involucrados en el paso de las proteínas mitocondriales a través de las membranas mitocondriales se identifican por las siglas TIM o TOM por las iniciales en inglés de translocasa de la membrana mitocondrial interna o externa respectivamente; cada una de las diferentes proteínas de este tipo se identifica por su masa molecular aparente en kilodaltones indicado después de las siglas. Los mecanismos que llevan al paso de proteínas a través de los complejos TOM son diferentes a los que se desarrollan en los complejos TIM; ambos complejos pueden funcionar de manera independiente, y por lo menos un grupo de proteínas localizado en el espacio intermembranal parece necesitar solamente de los complejos TOM para llegar a su destino final. Sin embargo, se cree que sin la presencia de los complejos TIM la capacidad translocadora de TOM se limita a la inserción y transporte de las presecuencias, además de que la eficiencia de ambos complejos aumenta cuando se encuentran acoplados (Figura 1.6) (Lill y col., 1992; Voos y col., 1999).

Las preproteínas se unen a la superficie mitocondrial a través de interacciones con proteínas receptoras que se encuentran ancladas en la membrana externa y se exponen hacia el citosol, estos receptores incluyen a las proteínas Tom70, Tom37, Tom22 y Tom20. La unión inicial y la presentación de las proteínas precursoras al complejo importador puede también depender de la interacción de proteínas de tipo chaperoninas con algunos componentes de los receptores, de la liberación de las chaperoninas, o de la interacción directa entre la presecuencia y la bicapa

lipídica. Los complejos proteínicos Tom reconocen las secuencias señal y catalizan la formación de los intermediarios necesarios para la transferencia de las proteínas al interior o del otro lado de la membrana externa mitocondrial; además interactúan de manera dinámica con la membrana interna (Kiebler y col. 1990; Segui-Real y col. 1993).

Las proteínas atraviesan la membrana externa en contacto con Tom40, una proteína que se encuentra casi totalmente embebida en la membrana y se piensa que se encuentra formando un poro general de inserción junto con otros receptores y proteínas no caracterizadas.

La inserción de las presecuencias en la membrana interna requiere, en la mayoría de los casos, que exista un potencial electroquímico $\Delta\psi$, probablemente para promover un movimiento electroforético de cargas positivas a través del aparato importador. Para cruzar la membrana interna, las proteínas precursoras, se asocian con un complejo que contiene a Tim17, Tim44 y Tim23, que se han propuesto como proteínas integrales de membrana (Martin y col., 1991).

Para que se complete el paso de las proteínas precursoras a la matriz mitocondrial, se requiere de la proteína de choque térmico mitocondrial 70 (mtHsp70) y de la hidrólisis de ATP. Se piensa que ciclos de unión de mtHsp70 con la proteína precursora y con Tim44 que se encuentra asociada a la membrana interna, funcionan como un pistón que dirige el movimiento unidireccional de las preproteínas hasta la matriz mitocondrial (Ungermann y col., 1994).

Finalmente, una peptidasa de matriz mitocondrial cataliza la remoción proteolítica de la presecuencia mitocondrial. Las proteínas de matriz Hsp60, Hsp10 y la ciclofilina 20 promueven la maduración final de las proteínas, que incluyen el plegamiento y ensamblaje de la proteína funcional.

Las proteínas cuya localización funcional se restringe a la membrana externa mitocondrial, suelen no presentar presecuencia y al realizar el reconocimiento de las proteínas Tom son localizadas en el interior de la membrana o en el espacio intermembranal. Las proteínas de la membrana interna siempre presentan presecuencias y pueden ser incorporadas directamente a la membrana interna o ser trasladadas hasta la matriz mitocondrial por las proteínas Tim para después acoplarse a los complejos funcionales en la membrana interna (Millar y Shore, 1996).

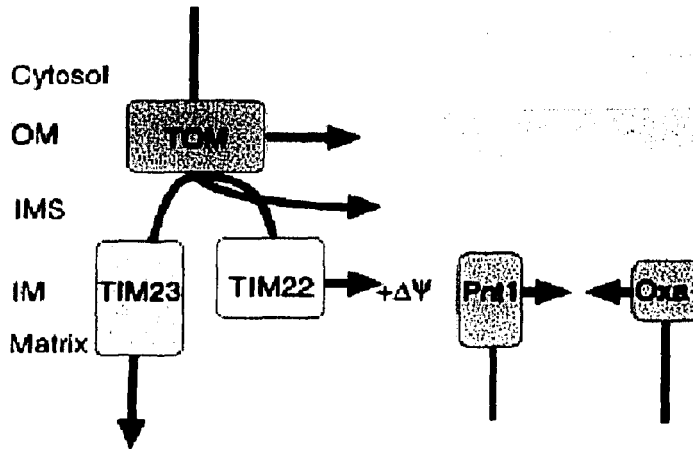


Figura I.6. Vías de importación de proteínas a la mitocondria. Las proteínas mitocondriales son importadas desde el citoplasma a través del complejo TOM, y posteriormente, de acuerdo al destino permanecen en la membrana externa, en el espacio intermembranal (IMS) o se unen a los complejos translocadores de la membrana interna (TIM). Las proteínas con MTSs típicas se unen al complejo TIM23, mientras que aquellas que no la tienen se unen a TIM22. Las proteínas sintetizadas en la matriz mitocondrial se exportan por los complejos Pnt1 y Oxa1 (Tomado de Köeler,

Organismos estudiados

A lo largo de este estudio se analizaron varios miembros del grupo de los protistas en particular parásitos, que no se encuentran estrechamente relacionados filogenéticamente con las algas clorofceas. Sin embargo, descubrir que algunos de los genes nucleares que codifican proteínas mitocondriales comparten características similares, revaloramos la situación en la que se encuentran y la necesidad de que compartan un espacio en este escrito (Figura I.7).

Las “algas verdes” (phylum Chlorophyta)

Actualmente se considera que las plantas terrestres y las algas verdes constituyen un grupo monofilético conocido como Viridiplantae o Clorobionta (Moreira y col., 2000). Todos los organismos incluidos en este grupo se caracterizan por ser organismos eucariontes que tienen cloroplastos con una doble membrana, tilacoides apilados, clorofilas tipo *a* y *b* y la presencia de almidón como polisacárido de reserva (a excepción de los miembros que han perdido la capacidad fotosintética de manera secundaria). La filogenia interna de este grupo se encuentra en constante revisión considerando fundamentalmente evidencias moleculares que coinciden con observaciones morfológicas (Friedl, 1997). De manera general, este grupo se ha dividido en dos grandes categorías: Streptophyta y Chlorophyta. Dentro del phylum Streptophyta se incluyen a las plantas terrestres y a las algas verdes más cercano evolutivamente a ellas (las Charophyceae), mientras que dentro del phylum Chlorophyta se incluye al resto de las algas verdes (Turmel y col., 2002). El grupo Chlorophyta se subdivide en por lo menos cuatro linajes: Trebouxiophyceae (*P. wickerhamii* *wickerhamii*), Chlorophyceae [*Chlamydomonas reinhardtii*, *Chlamydomonas eugametos* (ahora *Chlamydomonas moewusii*, Gerloff), *Scenedesmus obliquus*], Ulvophyceae (*Caulerpa taxifolia*) y Prasinophyceae (*Nephroselmis olivacea*) (Figura I.8) (Friedl, 1997).

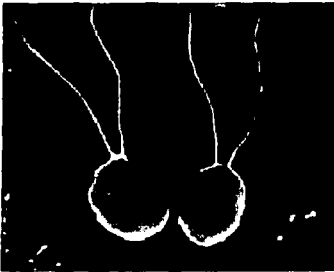
Los “apicomplejos” (phylum Apicomplexa)

Los miembros del phylum Apicomplexa son protistas endoparásitos que se definen por tener un ciclo de vida que comienza con una forma infectiva móvil (zoito) que contiene un complejo apical característico que le da el nombre al phylum. En algunos miembros de este phylum se encuentra un plástido reminiscente con su propio material genético. Dentro de este grupo, se encuentran numerosos agentes causales de enfermedades importantes tanto en seres humanos como en animales domésticos. En general, su ciclo de vida tiene tres estadios: i) una etapa de crecimiento durante la cual se lleva a cabo la infección a la célula huésped; ii) una etapa sexual durante la cual se lleva a cabo la producción de gametos y fertilización para formar cigotos localizados dentro de oocitos; y iii) una etapa de esporogénesis durante la cual suceden varias

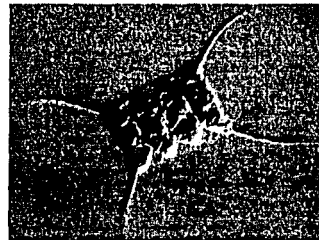
divisiones dentro de los oocitos para formar esporozoitos, que son las nuevas formas infectivas (Vivier y Desportes, 1990; Ellis y col. 1998).

Usualmente se considera que dentro del grupo de los apicomplejos se encuentran cuatro grupos bien definidos (de acuerdo a características biológicas como las especificidades al hospedero y al tejido que infectan): los coccidios, los gregarinos, los haemosporidios y los piroplasmas. Sin embargo, el arreglo taxonómico apropiado de estos grupos está en discusión, y en los últimos años al menos siete arreglos diferentes han sido propuestos (Ellis y col. 1998).

Chlorophyta



Chlamydomonas reinhardtii



Scenedesmus obliquus

Apicomplexa



Toxoplasma gondii



Plasmodium falciparum

Figura I.7. Organismos estudiados

TESIS CON
FALLA DE ORIGEN

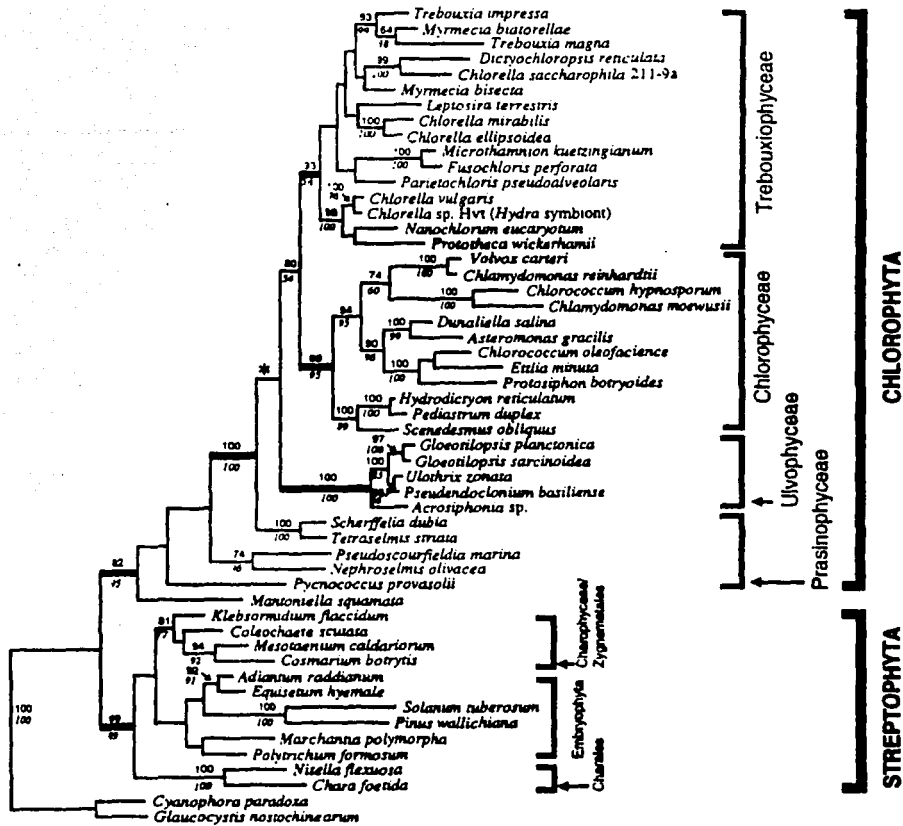


Figura I.8. Relaciones filogenéticas del grupo Viridiplantae (Tomado de Friedl, 1997).

Planteamiento del Problema

Los genomas mitocondriales representan una reminiscencia de lo que fue el genoma de la bacteria que les dio origen y en general, contienen genes que codifican un pequeño número de proteínas, predominantemente hidrofóbicas, que participan en el proceso de fosforilación oxidativa en la membrana interna mitocondrial. La mayoría de los genes que se encontraban en el simbiote eubacteriano han desaparecido de los genomas mitocondriales, algunos de ellos han sido transferidos a los genomas nucleares y otros simplemente se han perdido. Es posible plantear un patrón general de modificaciones que deben experimentar los genes mitocondriales al transferirse al genoma nuclear, de manera que se conviertan en genes activos que codifiquen productos capaces de migrar hasta la mitocondria e insertarse correctamente en los complejos funcionales. Sin embargo, los mecanismos que han seguido estos genes a lo largo de la evolución para ser transferidos al núcleo y los mecanismos con los cuales fueron transferidos, así como el porqué un particular grupo de genes continúa en los genomas mitocondriales de la mayor parte de los organismos eucariontes, siguen siendo tema de numerosas investigaciones y debates.

En este trabajo se han analizado algunos genes nucleares que codifican proteínas de la cadena respiratoria mitocondrial en algunos protozoarios unicelulares, como las algas *Chlamydomonas reinhardtii* y *Scenedesmus obliquus* y los parásitos *Toxoplasma gondii*, *Plasmodium falciparum*, *Plasmodium yoelii* y *Theileria parva*, que en la mayor parte de los organismos estudiados hasta la fecha se han conservado en los genomas mitocondriales.

El análisis de los genes *atp6*, *cox2a*, *cox2b* y *nad4L* en estos organismos ha permitido plantear estrategias específicas de transferencia para cada uno de estos genes que se encuentran de manera única en el genoma nuclear y al mismo tiempo, ha permitido relacionar evolutivamente a estos organismos, que pertenecen a grupos filogenéticos alejados pero que comparten ciertas características particulares.

Objetivos

Este trabajo tiene como objetivo general entender un poco más el proceso de transferencia de genes mitocondriales al núcleo en las algas clorofíceas.

Para cumplir con las características de este proyecto, se plantearon algunos objetivos particulares para cada uno de los genes a estudiar:

1. Caracterizar los genes *atp6* y *nd4L* de *C. reinhardtii*.

a) Encontrar la secuencia genómica y de traducción del gen *atp6* de *C. reinhardtii*. Demostrar que se trata de un gen nuclear en este organismo. Analizar las características de este gen como la presencia de intrones, las regiones 5' y 3' no traducidas, la presencia de señales de poliadenilación y el uso de codones. Localizar la subunidad ATP6 en el complejo F₁F₀ ATP sintasa de *C. reinhardtii*. Determinar la secuencia aminoterminal de la proteína madura. Analizar las posibles modificaciones estructurales y de hidrofobicidad que se produjeron en esta subunidad para ser transferida desde el citoplasma hasta la mitocondria en este organismo.

b) Encontrar la secuencia genómica y de traducción del gen *nd4L* de *C. reinhardtii*. Demostrar que se trata de un gen nuclear. Analizar las características de este gen como la presencia de intrones, las regiones 5' y 3' no traducidas, la presencia de señales de poliadenilación y el uso de codones.

Analizar algunas características de los genes estudiados y de las proteínas que codifican que permiten que se encuentren codificados en el genoma nuclear de las algas chlamydomonadaceas.

Los resultados que cubren este objetivo se encuentran en el capítulo de Resultados I.

2. Caracterizar el gen *cox2b* de *S. obliquus*.

Encontrar la secuencia genómica del gen *cox2b* de *S. obliquus*. Analizar las características de este gen como la presencia de intrones, las regiones 5' y 3' no traducidas, la presencia de señales de poliadenilación y el uso de codones.

Analizar las implicaciones de la presencia de este gen nuclear en el proceso de evolución de los genomas mitocondriales de las algas clorofitas.

Los resultados que cubren este objetivo se encuentran en el capítulo de Resultados II.

3. Caracterizar los genes *cox2a* y *cox2b* de *T. gondii*. Analizar la relación que existe entre los genes *cox2a*, *cox2b* de algunos organismos apicomplejos y los genes de las algas chlamydomonadaceas.

Durante el desarrollo del presente trabajo surgió un objetivo que no estaba planteado en el proyecto original, sin embargo ha resultado ser de gran interés. Este último objetivo fue determinar la secuencia genómica de los genes *cox2a* y *cox2b* de *T. gondii* y analizar las posibles relaciones filogenéticas que existen entre los apicomplejos y las algas clorofíceas.

Los resultados que cubren este objetivo se encuentran en el capítulo de Resultados III.

Resultados

A lo largo de este trabajo se clonaron y caracterizaron genes que codifican distintas proteínas de la cadena respiratoria mitocondrial en algunos protozoarios. Todos los genes estudiados codifican proteínas hidrofóbicas y en la gran mayoría de los organismos examinados hasta la fecha se encuentran en los genomas mitocondriales (mientras que en los organismos en cuestión los genes han sido transferidos al genoma nuclear y la copia mitocondrial se ha eliminado por completo).

A partir de la secuencia de estos genes es posible identificar las modificaciones que han sufrido para poder ser entidades que residen en el genoma nuclear de manera que puedan ser expresados y regulados en este ambiente. Además, el análisis de la secuencia deducida de aminoácidos de la proteína que codifican, ha permitido determinar algunos de los cambios que han experimentado para poder ser transferidas al interior de la mitocondria una vez que han sido sintetizadas en el citoplasma celular.

Aunque todos los genes examinados presentan algunas características "típicas" de los genes que han sido transferidos al genoma nuclear (como el cambio en el uso de codones y la adquisición de señales de poliadenilación nucleares), ha sido posible identificar ciertos elementos que son particulares para este grupo de genes, como la adquisición de presecuencias relativamente largas. Además, cada uno de los genes estudiados es un caso especial, cada uno ha desarrollado estrategias únicas que suponemos confieren las características necesarias para que estas proteínas hidrofóbicas viajen hasta el interior de las mitocondrias y se inserten en los complejos respiratorios para conformar unidades funcionales.

Los siguientes capítulos, detallan las características de los genes *atp6* y *nd4L* de *C. reinhardtii*, *cox2b* de *S. obliquus*, y *cox2a* y *cox2b* de *T. gondii*.

Todas las secuencias obtenidas en este trabajo se depositaron en la base de datos DDBJ/GenBankTM/EBI Data Bank con los siguientes números de acceso:

AF41119	secuencia genómica del gen <i>atp6</i> de <i>C. reinhardtii</i> .
AF411921	secuencia de traducción (cDNA) del gen <i>atp6</i> de <i>C. reinhardtii</i> .
AF499003	secuencia genómica del gen <i>cox2b</i> de <i>S. obliquus</i> .
AF503914	secuencia genómica del gen <i>cox2a</i> de <i>T. gondii</i> .
AF499004	secuencia genómica del gen <i>cox2b</i> de <i>T. gondii</i> .
	secuencia genómica y de traducción del gen <i>nd4L</i> de <i>C. reinhardtii</i>

Resultados 1:
Los genes *atp6* y *nd4L* de *C. reinhardtii*

El alga verde unicelular *Chlamydomonas reinhardtii*, es una clorofita biflagelada que ha sido ampliamente utilizada como modelo experimental en el estudio de la fotosíntesis, la biogénesis en organelos y muchos otros aspectos de la biología celular (Harris, 1998). El genoma mitocondrial de estos organismos es una molécula lineal de 15.8 kb muy compacta. Se encuentran genes que codifican para 8 proteínas que incluyen al citocromo *b* del complejo *bc₁*, *cox1* del complejo IV, 5 subunidades de la NADH deshidrogenasa y una posible transcriptasa reversa. Codifica también para 3 RNAs de transferencia y para 2 RNAs ribosomales fragmentados en módulos. Las regiones intergénicas son muy pequeñas y no se presenta ningún intrón. Al contrario de lo que se observa en la mayoría de los organismos eucariontes estudiados hasta la fecha, en este genoma están ausentes todos los genes que codifican para el complejo F_1F_0 ATP sintasa, incluyendo los genes que codifican a las subunidades ATP6 y ATP8, varios genes que codifican a subunidades hidrofóbicas del complejo I como NAD3 y NAD4L y los genes que codifican para las subunidades II y III (COX II y COX III) del complejo citocromo *c* oxidasa (Figura V.1) (Michaelis y col., 1990).

V.1 El gen *atp6* de *C. reinhardtii*

En la mayoría de los organismos eucariontes todas las subunidades que componen la región F_1 y algunas de las subunidades de la región F_0 del complejo F_1F_0 ATP sintasa se encuentran codificadas en el genoma nuclear, se sintetizan en el citoplasma y finalmente se transportan al interior de la mitocondria para formar el complejo funcional. En general en los genomas mitocondriales, solamente se encuentran los genes *atp6*, *atp8* y *atp9*. Estos genes codifican proteínas hidrofóbicas con varios cruces transmembranales que constituyen parte de la región F_0 .

En este trabajo, se obtuvieron las secuencias genómica y de traducción del gen *atp6* de *C. reinhardtii*, se confirmó que este gen está localizado en el genoma nuclear y que la proteína que codifica forma parte del complejo F_1F_0 ATP sintasa mitocondrial en este organismo. También se analizaron las características de la proteína que le permiten estar codificada en el núcleo y ser importada correctamente a la mitocondria sin alterar las regiones que son importantes para el funcionamiento del complejo.

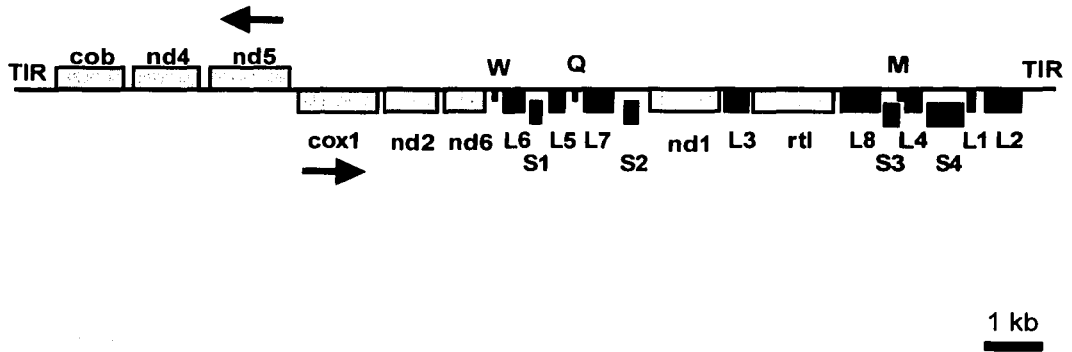


Figura V.1. Genoma mitocondrial de *C. reinhardtii*.

En gris se muestran los genes que codifican proteínas mitocondriales, en negro los que codifican rRNAs, en los extremos se muestran las regiones invertidas (TIR), las flechas muestran el sentido de la transcripción de los genes indicados, y los tres tRNAs que se localizan en este genoma están señalados por el aminoácido que reconocen (W, Q y M).

Secuencia genómica y de traducción del gen *atp6* de *C. reinhardtii*

En febrero del año 2000 se liberó una base de datos con más de 3500 ESTs de *C. reinhardtii* (Asamizu y col., 1999). En esta colección de secuencias parciales de cDNA, se identificó una clona (AV394701) que codificaba para un marco de lectura abierto que presentaba un alto porcentaje de similitud con el extremo carboxilo algunas secuencias reportadas de ATP6.

La secuencia identificada se utilizó en un análisis de BLAST a nivel de nucleótidos contra la misma base de datos. Este examen de la base de datos, permitió identificar otras dos secuencias parciales cuyo extremo 3' corresponde al extremo 5' de la clona AV394701 (posible cDNA completo de ATP6). Estas dos clonas (AV395475 y AV388269) son casi idénticas entre ellas (una de ellas se sobrelapa más que la otra con la clona AV394701), y la secuencia de aminoácidos que se puede deducir de ellas no presenta similitud con ninguna proteína descrita en la base de datos. A pesar de que la secuencia de estas dos clonas no se parecía a nada, se consideró que tal vez no se trataba de un artefacto o error en la clonación de fragmentos de cDNA sino que en realidad eran secuencias que al unir las formaban parte de un gen completo o casi completo.

Se diseñaron oligodesoxinucleótidos en los extremos de la secuencia de las clonas AV395475 y AV394701 y se emplearon en una reacción de PCR utilizando una biblioteca de cDNA clonada en el fago λ gt10 como templado (Franzén y Falk, 1992). Se obtuvo un producto amplificado de 822 pb. La secuencia de este producto, confirmó que las clonas identificadas en la base de datos son complementarias y corresponden a un fragmento del gen *atp6*. Las secuencias de los extremos 5' y 3' del cDNA se obtuvieron por reacciones de RACE-PCR.

La secuencia completa de traducción del gen *atp6* de *C. reinhardtii* es de 2349 pb que incluyen 1014 pb de un marco de lectura abierto flanqueados por 27 pb que podrían corresponder a la región 5' no traducida y una región 3' no traducida de 1299 pb (Figura V.2).

Los primeros 27 pb incluyen la secuencia anterior al primer codón ATG y que corresponde a la primera metionina del ORF identificado como ATP6. Se ha planteado que la secuencia (A/C) A (A/C) (A/C) ATG (G/C) corresponde al consenso que se encuentra alrededor del sitio de inicio de traducción de los genes nucleares en *C. reinhardtii* (Silflow, 1998). Al revisar la secuencia que se encuentra rodeando al codón identificado como la posible Met inicial, se encontró que es AACC ATG G, y que por lo tanto coincide con la secuencia consenso reportada. El ORF se detiene con el codón de término TAA, que corresponde al codón utilizado por alrededor del 70% de los genes nucleares de *C. reinhardtii* descritos (Figura V.2).

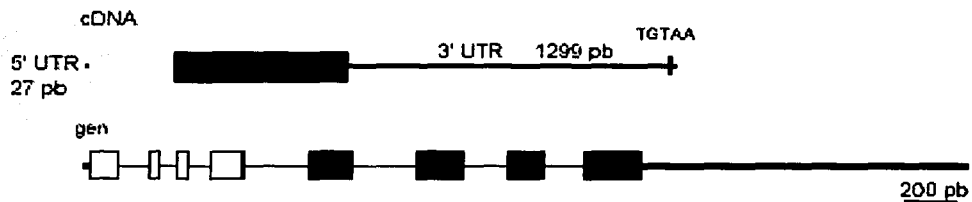


Figura V.2. Organización del gen *atp6* de *C. reinhardtii*.

Las regiones codificantes se encuentran señaladas como cajas; las regiones no traducidas se encuentran como líneas gruesas, y los intrones como líneas delgadas. La región correspondiente a la MTS se muestra en gris y la región correspondiente a la proteína madura en negro.

TESIS CON
FALLA DE ORIGEN

En general, el genoma nuclear de *C. reinhardtii* presenta un alto contenido de G-C, y el gen *atp6* no es la excepción: la secuencia de traducción contiene un porcentaje de G-C de 67.4%.

Con base en la secuencia de traducción obtenida, se diseñaron oligodesoxinucleótidos específicos para amplificar por PCR la secuencia genómica de *atp6*. La secuencia completa se obtuvo al sobrelapar la secuencia obtenida en tres diferentes reacciones utilizando pares de oligodesoxinucleótidos diferentes.

La secuencia genómica obtenida equivale a la secuencia de traducción (es decir, incluye exclusivamente el segmento genómico desde la región 5' UTR hasta el sitio donde comienza la poliadenilación en el extremo 3' UTR) y tiene un tamaño de 3577 pb.

El porcentaje de G-C de la secuencia genómica es prácticamente igual al de la secuencia de traducción: 65.3%.

Otra característica común de los genes nucleares de *C. reinhardtii* es la presencia de un promedio de cuatro intrones por cada kb de DNA codificante. En el gen *atp6* se encuentran siete intrones, representando un promedio relativamente bajo de 2.9 intrones por kb de DNA codificante. Los intrones del gen *atp6* tienen 115, 72, 89, 264, 263, 186 y 161 pb, reflejando un promedio de 164 pb, que resulta también pequeño al compararlo con el tamaño medio de los intrones en *C. reinhardtii* (219 pb). Los sitios de corte de los siete intrones coinciden con el consenso de los organismos eucariontes para este sitio: 5'- GT-AG -3'. En los intrones 3 y 7 se identificó la posible señal de ramificación NCTAG a 74 y 27 pb respectivamente del sitio 3' de corte.

El gen *atp6* se encuentra localizado en el genoma nuclear de *C. reinhardtii*

Para confirmar que el gen *atp6* se encuentra en el genoma nuclear de *C. reinhardtii*, se realizó un análisis de tipo Southern blot: DNA total de *C. reinhardtii* se sometió a una electroforesis horizontal en un gel de agarosa al 0.7%, separando el DNA mitocondrial como una banda discreta que corre por debajo de una banda principal que contiene el DNA nuclear y el DNA del cloroplasto. El DNA de estos geles fue transferido a una membrana de nylon y sujeto a un análisis de hibridación utilizando un juego de sondas de origen mitocondrial y un juego de sondas de origen nuclear. La banda de menor tamaño mostró una reacción positiva con las tres sondas mitocondriales utilizadas, correspondientes a los genes *cob1*, *nad2* y *cox1*; mientras que el DNA nuclear muestra una reacción positiva con sondas de origen nuclear como *Cyc*, *AtpB* y *cox3*. Al utilizar el fragmento del gen *atp6* de 822 pb como sonda se observó una reacción clara y única con el DNA nuclear (Figura V.3).

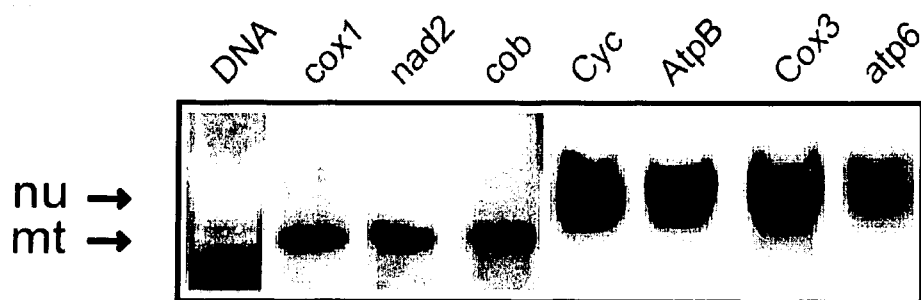


Figura V.3. El gen *atp6* de *C. reinhardtii* se encuentra localizado en el genoma nuclear.

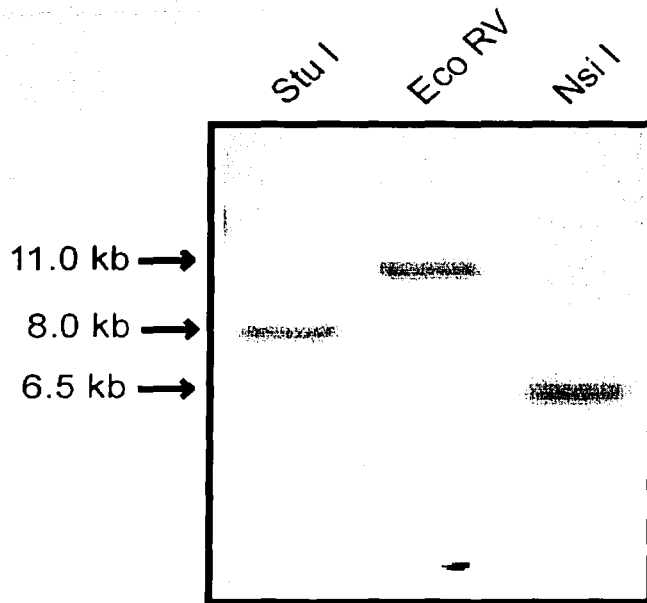


Figura V.4. El gen *atp6* se encuentra como copia única en el genoma nuclear de *C. reinhardtii*.

Para determinar si el gen *atp6* se encuentra como copia única en el genoma de *C. reinhardtii*, se realizó otro experimento tipo Southern blot digiriendo DNA total de *C. reinhardtii* con diferentes enzimas de restricción que no cortaran la secuencia genómica de *atp6* y se transfirieron los fragmentos producidos a una membrana de nylon; se hizo una hibridación utilizando como sonda el producto de PCR de 822 pb que contiene un fragmento del gen *atp6*. Se observó una reacción positiva en una región única al digerir el DNA con enzimas diferentes de manera independiente (Figura V.4).

El patrón de uso de codones del gen *atp6* de *C. reinhardtii* presenta una clara tendencia a favorecer el uso de trinucleótidos que presentan una G o C en la tercera posición. Esta característica es típica del uso de codones nuclear y diferente al uso de codones mitocondrial o de cloroplasto en las algas chlamydomonadaceas (Cuadro V.1).

Dentro de la secuencia de 1299 pb correspondiente al extremo 3' no traducido, se identificaron dos posibles señales de poliadenilación: una secuencia TGTA A y una secuencia TGTAG, localizadas 15 y 1048 pb respectivamente antes del primer sitio de poliadenilación observado. La primera es la señal que se utiliza en la mayor parte de los genes nucleares de *C. reinhardtii* reportados, y la segunda ha sido identificada como una variación funcional de la primera en algunos genes (Silflow, 1998). Para verificar si ambas señales eran funcionales o descartar alguna de ellas, se realizó un experimento tipo Northern utilizando una preparación de RNA total y el fragmento de 822 pb de la secuencia de traducción del gen obtenido por PCR como sonda. Se observó una señal única de alrededor de 2.4 kb que sugiere que sólo la señal TGTA A está funcionando en el caso de este gen (Figura V.5).

La subunidad ATP6 se encuentra presente en el complejo F₁F₀ ATP sintasa de *C. reinhardtii*.

A pesar de que se había demostrado que el gen *atp6* se expresa en *C. reinhardtii*, había que demostrar que su producto proteico formaba parte del complejo F₁F₀ ATP sintasa mitocondrial en *C. reinhardtii*.

El complejo F₀F₁ ATP sintasa de *C. reinhardtii* se identificó en geles azules nativos por su posición relativa. Después de realizar un SDS-PAGE fue posible identificar 12 polipéptidos diferentes con una masa molecular aparente de 61.0, 51.9, 40.2, 30.1, 27.8, 26.3, 17.8, 15.5, 13.0, 10.8, 9.1 y 7.8 kDa respectivamente. 11 de los 12 polipéptidos observados, fueron sometidos a una secuenciación del extremo aminoterminal: las tres subunidades de mayor masa se identificaron como las subunidades α , β y γ al comparar las secuencias y la disposición en el patrón

Cuadro V.1.
 Uso de codones mitocondrial y nuclear de *C. reinhardtii* comparado
 con el de los genes *atp6* y *nd4L*.

Res	Codon	Mito	Nuc	<i>atp6</i>	<i>nd4L</i>	Res	Codon	Mito	Nuc	<i>atp6</i>	<i>nd4L</i>
Gly	GGG	0	8	8	0	Trp	TGG	100	100	100	100
Gly	GGA	4	4	5	6	End	TGA	0	13	0	0
Gly	GGT	78	13	24	12	Cys	TGT	37	11	0	0
Gly	GGC	18	75	62	82	Cys	TGC	63	89	100	100
Glu	GAG	99	96	100	100	End	TAG	19	18	0	100
Glu	GAA	1	4	0	0	End	TAA	81	64	100	0
Asp	GAT	56	15	60	0	Tyr	TAT	34	10	17	0
Asp	GAC	44	85	40	100	Tyr	TAC	66	90	83	100
Val	GTG	4	66	71	83	Leu	TIG	57	3	5	0
Val	GTA	47	3	0	0	Leu	TTA	0	2	0	0
Val	GTT	32	8	8	0	Phe	TTT	37	14	0	8
Val	GTC	17	23	21	17	Phe	TTC	63	86	100	92
Ala	GCG	2	28	24	30	Ser	TCG	0	28	32	31
Ala	GCA	2	6	0	13	Ser	TCA	0	4	0	6
Ala	GCT	65	16	26	27	Ser	TCT	29	8	10	19
Ala	GCC	31	50	50	30	Ser	TCC	8	27	26	13
Arg	AGG	0	4	9	33	Arg	CGG	0	14	0	17
Arg	AGA	0	1	0	0	Arg	CGA	6	3	0	9
Ser	AGT	35	3	0	6	Arg	CGT	67	10	36	17
Ser	AGC	28	30	32	25	Arg	CGC	27	69	55	25
Lys	AAG	13	96	100	100	Gln	CAG	28	90	94	88
Lys	AAA	87	4	0	0	Gln	CAA	72	10	6	12
Asn	AAT	16	7	10	12	His	CAT	28	12	20	0
Asn	AAC	84	93	90	88	His	CAC	72	89	80	100
Met	ATG	100	100	100	100	Leu	CTG	4	74	83	59
Ile	ATA	1	3	0	0	Leu	CTA	31	3	0	0
Ile	ATT	66	24	30	36	Leu	CTT	8	5	2	12
Ile	ATC	33	73	70	64	Leu	CTC	0	13	10	30
Thr	ACG	0	24	14	35	Pro	CCG	2	26	11	13
Thr	ACA	1	6	0	20	Pro	CCA	78	6	5	27
Thr	ACT	65	11	5	0	Pro	CTT	16	13	21	13
Thr	ACC	34	59	81	45	Pro	CCC	4	55	63	47

En gris se indican las diferencias claras entre el uso de codones.
 Los valores son porcentuales.

TESIS CON
 FALLA DE ORIGEN

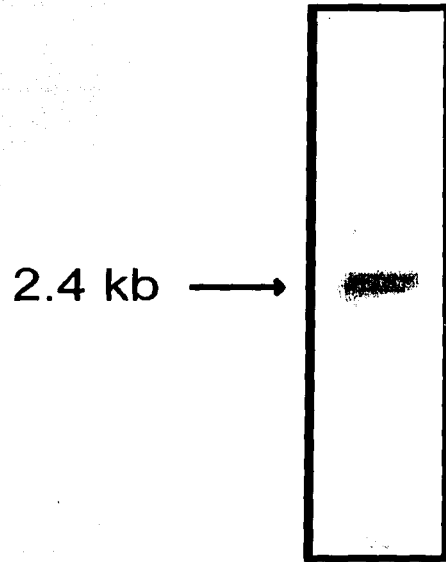


Figura V.5. El gen *atp6* produce un solo mensajero en *C. reinhardtii*

TESIS CON
FALLA DE ORIGEN

electroforético de la preparación obtenida en este trabajo y la reportada previamente por Nurani y Franzén (1996). La subunidad de 26.3 kDa se identificó como la subunidad δ por la similitud de la secuencia amino terminal obtenida y otras subunidades δ reportadas; el polipéptido cuya masa molecular aparente es de 15.5 kDa se identificó como la proteína codificada por el gen *atp6* ya que los 15 aminoácidos determinados coinciden exactamente con los residuos 108 a 122 de la secuencia deducida de aminoácidos codificada por el gen *atp6*. La identidad de los 7 polipéptidos restantes no muestra ninguna similitud clara con subunidades del complejo F_1F_0 ATP sintasa reportadas en las bases de datos, por lo que la identidad de estas subunidades debe todavía ser esclarecida (Figura V.6).

Al identificar la secuencia amino terminal de la subunidad Cr-ATP6, fue posible predecir que la subunidad madura se compone de 233 residuos de aminoácidos con una masa molecular de 24,577 Da.

En el ORF predicho por la secuencia de nucleótidos, antes del inicio de la proteína madura se encuentran siete diferentes codones ATG y en principio, cualquiera podría corresponder al inicio del péptido señal. Las diferentes metioninas predicen presecuencias de 107, 35, 30, 26, 12, 3 y 1 aminoácidos. Se ha reportado que la secuencia que rodea los codones de inicio y de término influye en la eficiencia de traducción en *C. reinhardtii*. La secuencia que rodea el codón correspondiente a la primera metionina AACCATG corresponde al sitio de inicio de la traducción consenso reportado (A/C) A (A/C) (A/C) ATG (G/C) para *C. reinhardtii* (Silflow, 1998). De acuerdo con esta secuencia consenso la preproteína de Cr-ATP6 debe tener un péptido señal de 107 residuos, la cual debe dirigir y permitir la importación de esta proteína a la mitocondria (Figura V.1).

La secuencia deducida de aminoácidos predice una proteína madura de 233 residuos con una masa molecular de 24,577 Da.

La secuencia completa de aminoácidos de Cr-ATP6 se comparó con la secuencia deducida de aminoácidos del gen mitocondrial *atp6* de *Scenedesmus obliquus*, mostrando una identidad del 46% y una similitud del 59%.

Importabilidad de Cr-ATP6 a la mitocondria

Como se mencionó anteriormente, en los organismos eucariontes la mayor parte de las proteínas mitocondriales se encuentran codificadas en el genoma nuclear, y en los genomas mitocondriales se ha conservado un número limitado de genes que codifican proteínas hidrofóbicas que funcionan en la membrana interna mitocondrial (von Heijne, 1986). La naturaleza hidrofóbica

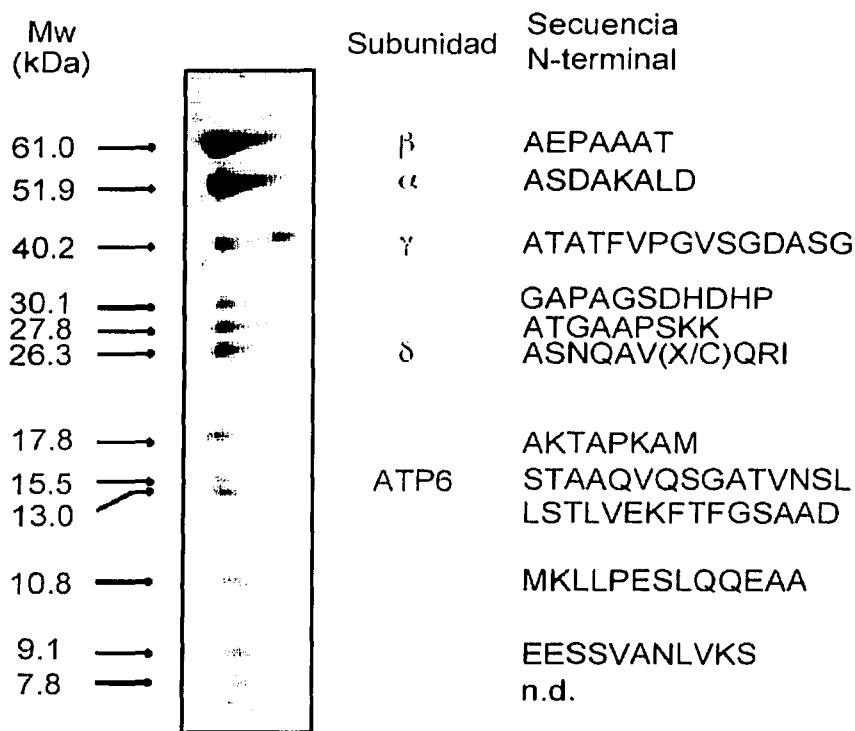


Figura V.6. Composición peptídica del complejo F_1F_0 ATPasa de *C. reinhardtii*. Se muestran las secuencias amino terminal de las subunidades analizadas y la identidad de aquellas subunidades determinadas.

TESIS CON
FALLA DE ORIGEN

de las proteínas que permanecen en los genomas mitocondriales repercute directamente con la capacidad de viajar a través del citoplasma, cruzar las dos membranas mitocondriales y posteriormente insertarse en la membrana interna para acoplarse al complejo proteico al que pertenecen y funcionar adecuadamente.

La importabilidad de una proteína al interior de la mitocondria está determinada al menos por su hidrofobicidad (que de manera burda puede verse como el número de cruces transmembranales) y por la MTS que contenga los elementos necesarios para que la proteína se dirija hacia la mitocondria y hacia ningún otro destino celular.

Para que el gen *atp6* fuera importado exitosamente al genoma nuclear, la proteína codificada debió sufrir previamente o de manera simultánea una fuerte disminución en la hidrofobicidad promedio que facilitara el paso de la proteína a través de las membranas mitocondriales y que permitiera su posterior acoplamiento en el complejo funcional. También debió adquirir una MTS que dirigiera el correcto desplazamiento de la proteína a través del citoplasma celular hasta la mitocondria y la inserción en el complejo F_1F_0 ATP sintasa.

Presecuencia mitocondrial en Cr-ATP6

Tomando el primer codón ATG como aquel que codifica para la Met inicial de la preproteína, y el inicio de la proteína madura detectado por la secuenciación directa del polipéptido correspondiente, Cr-ATP6 presenta una MTS de 107 residuos de aminoácidos de longitud.

Esta presecuencia predicha, se caracteriza por tener un número relativamente bajo de residuos ácidos, además de que los residuos 2-24 pueden estar formando el dominio anfipático necesario para la importación de la proteína hasta la mitocondria. Muy probablemente, la secuencia que codifica a la presecuencia se añadió al gen cuando este ya se encontraba en el genoma nuclear, y tal vez la Met⁻¹ que se encuentra justo antes del sitio de corte de la MTS puede haber sido la Met inicial de la proteína mitocondrial original.

Estructura de Cr-ATP6

Utilizando los programas TopPred II (Claros y von Heijne, 1994), el servidor de red PredictProtein (www.es.embnnet.org/Services/MolBio/PredictProtein; Rost, 1996) y el servidor ExPASy de Biología Molecular (www.expasy.ch) se predijo la estructura secundaria de Cr-ATP6. Los límites de cada uno de los cruces transmembranales se determinó de manera más precisa alineando todas las secuencias de aminoácidos de las subunidades ATP6 reportadas hasta la fecha y comparando las estructuras predichas de algunas de ellas con los resultados

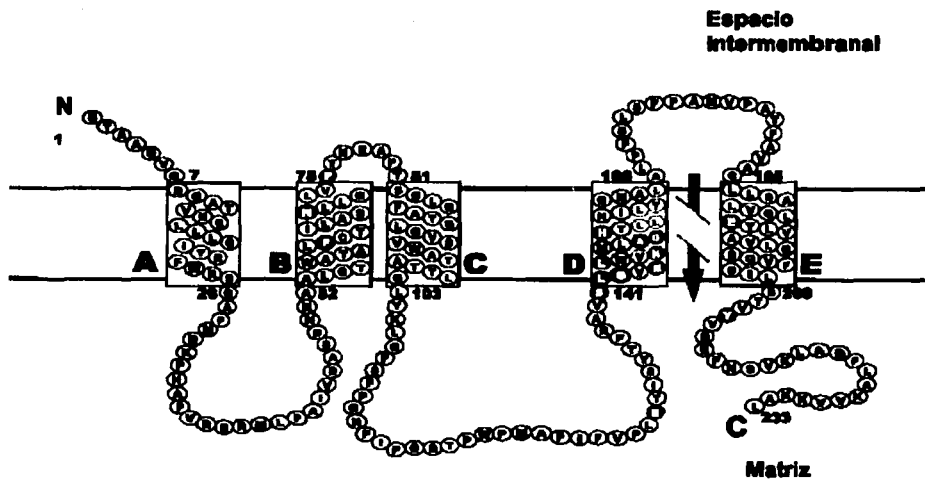


Figura V.7. Modelo de hélices transmembranales de la subunidad Cr-ATP6. Los residuos conservados se muestran como círculos negros. Los cruces transmembranales D y E son los que participan en la translocación de protones a través de dos semicanales formados por estos cruces y que se representan en este esquema como una flecha interrumpida.

experimentales que han permitido determinar la estructura secundaria de la subunidad a en *E. coli* (Valiyaveetil y Fillingame, 1998; Wada y col., 1999).

Estos análisis sugirieron una estructura de cinco cruces transmembranales (nombrados A, B, C, D y E respectivamente) y cuatro asas que los unen (nombradas a-b, b-c, c-d y e-f). Los siete residuos que corresponden al extremo amino terminal de la subunidad se predice que se encuentran localizados en el espacio intermembranal; mientras que los 25 residuos del extremo carboxilo terminal se encuentran expuestos hacia la matriz mitocondrial (Figura V.7).

En la secuencia deducida de aminoácidos fue posible identificar algunos residuos conservados en las proteínas ATP6 de todos los organismos reportados hasta la fecha. La función precisa de cada uno de estos residuos no ha sido determinada, sin embargo se han identificado algunos que están directamente involucrados en la translocación de protones, algunos de ellos formando parte de los semicanales por los cuales se guía a los protones a través de la membrana, y otros como la Arg145 involucrados directamente en la protonación y desprotonación del oligómero de subunidades c.

Disminución en la hidrofobicidad promedio de Cr-ATP6

La capacidad de importación de una proteína a la mitocondria puede ser predicha en función de la hidrofobicidad promedio más alta de un segmento de entre 60 y 80 residuos (mesohidrofobicidad) junto con la máxima hidrofobicidad de los posibles cruces transmembranales. Esta relación se puede representar en una gráfica de mesohidrofobicidad *versus* la hidrofobicidad máxima local (<H>), donde además es posible localizar una "frontera" entre las proteínas que potencialmente pueden y las que no pueden ser importadas a la mitocondria (Claros y col., 1995). En la figura V.8, se muestra una gráfica donde se comparan estos parámetros de Cr-ATP6 con otras subunidades ATP6 que se encuentran codificadas en genomas mitocondriales. Es posible observar que ambos parámetros se encuentran disminuidos, por lo que la proteína de *C. reinhardtii* puede potencialmente ser importada al interior de la mitocondria. De acuerdo a este análisis otras ATP6 que podrían potencialmente ser importadas al interior de la mitocondria son aquellas de *Oryza sativa*, *Scenedesmus obliquus* y *Glycine max*. En otras palabras, sus genes mitocondriales respectivos se encuentran "listos para ser transferidos al núcleo".

Al realizar un análisis de hidrofobicidad de cada uno de los cruces transmembranales de Cr-ATP6, se observó que si se compara cada uno con la hidrofobicidad promedio de los cruces equivalentes de otros organismos, existe una notable disminución en la hidrofobicidad de los cruces A, B y C. Estos cruces son aquellos cuya secuencia de aminoácidos se encuentra menos

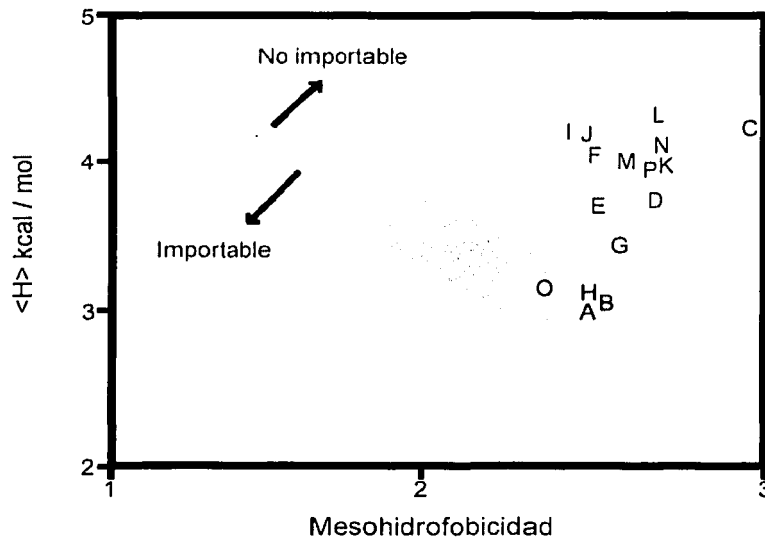


Figura V.8. Gráfica de mesohidrofobicidad contra hidrofobicidad local máxima de algunas subunidades ATP6. *C. reinhardtii* (A), *S. Obliquus* (B), *P. minor* (C), *H. annuus* (D), *P. subcordiformis* (E), *Z. mays* (F), *P. wickerhamii* (G), *O. sativa* (H), *S. tuberosum* (I), *N. tabacum* (J), *B. napus* (K), *T. aestivum* (L), *M. polymorpha* (M), *B. vulgaris* (N), *G. max* (O), *A. thaliana* (P).

TESIS CON
FALLA DE ORIGEN

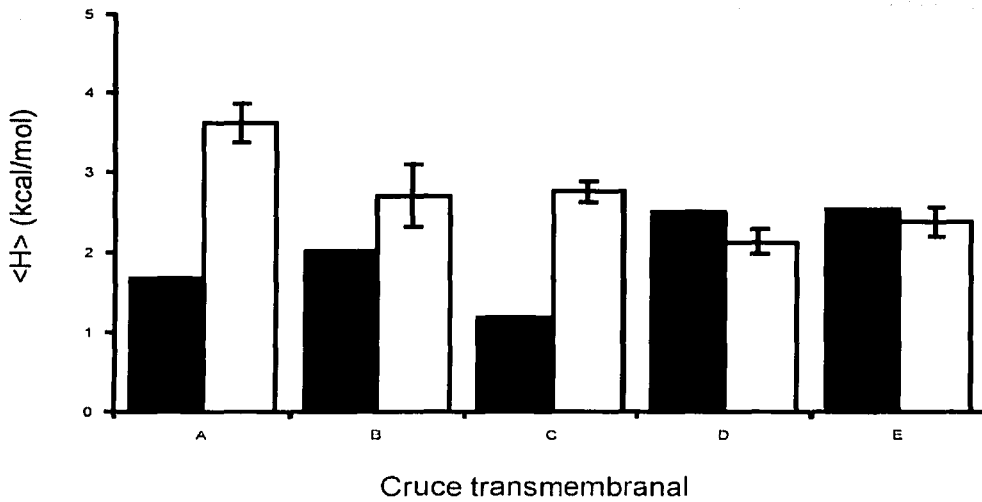


Figura V.9. Análisis de la hidrofobicidad de cada uno de los cruces transmembranales de ATP6. Los valores de hidrofobicidad de los cinco cruces transmembranales de Cr-ATP6 (barras negras) se comparan con el valor promedio de hidrofobicidad de 21 subunidades ATP6 codificadas en genomas mitocondriales (barras blancas).

TESIS CON
FALLA DE ORIGEN

conservada y se cree que no participan directamente en el proceso de translocación de protones. Al contrario, la hidrofobicidad de los cruces D y E, donde se encuentran los residuos conservados involucrados en la translocación de protones, no presentan diferencias significativas. La disminución en la hidrofobicidad del cruce A es tan pronunciada que en algunos análisis realizados se observó que este segmento de la secuencia no era en realidad un cruce transmembranal. Cr-ATP6 podría haber disminuido tanto su hidrofobicidad que de manera atípica tendría solo cuatro cruces transmembranales y no cinco (Figura V.9). La determinación bioquímica de la topología de Cr-ATP6 representa un estudio interesante por realizarse.

V.II El gen *nd4L* de *C. reinhardtii*

El complejo I de la cadena respiratoria mitocondrial cataliza la transferencia de dos electrones desde el NADH hasta la ubiquinona de manera acoplada a la translocación de cuatro a cinco protones hacia el espacio intermembranal. Hasta el momento no ha sido posible determinar la estructura cristalográfica de este complejo, solamente se han construido mapas de baja resolución con base en estudios de microscopía electrónica. El complejo tiene forma de "L", uno de los brazos se encuentra embebido en la membrana y el otro se encuentra en la matriz mitocondrial. El complejo de bovino se compone de al menos 42 subunidades, incluyendo siete proteínas hidrofóbicas que se encuentran codificadas en el genoma mitocondrial; el complejo de organismos procariontes puede ser de 13 o 14 subunidades. Al igual que en el caso de otros complejos respiratorios, se cree que las subunidades supernumerarias que se encuentran en los organismos eucariontes están involucradas en la estabilización y ensamblaje del complejo (Sazanov y col., 2000; Scheffler, 1999).

NAD4L es una de las subunidades hidrofóbicas altamente conservadas en los complejos I. Se encuentra interactuando en la membrana con las subunidades NAD1, NAD2 y NAD3 en la región del complejo I_g que corresponde a la región membranal que une a los dos brazos (Sazanov y col., 2000).

La secuencia de traducción del gen *nd4L* de *C. reinhardtii* es de 1647 pb y la secuencia genómica correspondiente es exactamente igual, es decir no presenta intrones ni ediciones transcripcionales. Esta secuencia se compone de una región 5' no traducida de 45 pb, un ORF de 681 pb y una región 3' no traducida de 921 pb dentro de la cual se localiza la señal de poliadenilación típicamente nuclear TGTA. El ORF predice una proteína de 227 residuos de aminoácidos con una masa molecular de 24,226 Da (Figura V.10).

La mayor parte de las características de este gen coinciden con todas las características de los genes nucleares de *C. reinhardtii*: el uso de codones (Tabla V.1), la presencia de una señal



Figura V.10. Organización del gen *nd4L* de *C. reinhardtii*. El ORF se encuentra señalado como una caja, en gris se muestra la región de la probable MTS y en negro la proteína madura.

TESIS CON
FALLA DE ORIGEN

de poliadenilación y la adquisición de una MTS; sin embargo, es extraño que no presente ningún intrón, siendo que los genes nucleares suelen tener una gran cantidad de intrones (Silflow, 1998).

El tamaño de esta proteína es muy grande si se compara con las subunidades NAD4L mitocondriales reportadas, incluyendo las de otras clorofitas como *Prototheca wickerhamii*, *Pedinomonas minor*, *Nephroselmis olivacea* y *Scenedesmus obliquus* (que tienen cada una un tamaño de 100, 96, 100 y 100 aminoácidos respectivamente). Esta diferencia en tamaños se debe a la adquisición de la MTS al transferir el gen al núcleo. Aunque la MTS de esta proteína no ha podido ser establecida con precisión, ya que no se cuenta con la secuencia amino terminal de la proteína madura, a través de comparaciones con las subunidades de otras clorofitas, es posible predecir que se trata de una presecuencia de alrededor de 133 residuos. La presencia de MTSs largas en *C. reinhardtii* es una característica que parece involucrar a todas las proteínas como ATP6, COX III, COX IIA y NAD4L que normalmente se encuentran codificadas por genes mitocondriales y en este organismo se encuentran en el genoma nuclear; sin embargo, no deja de sorprender que para importar una proteína de alrededor de 100 aa se requiera una MTS de la misma longitud.

Al hacer un análisis de hidrofobicidad de la región de 94 aminoácidos que podría corresponder a la proteína madura, es posible distinguir la presencia de dos cruces transmembranales que corresponderían a los predichos en estudios realizados en subunidades NAD4L de otros organismos (Figura V.10) (Le y col., 2001).

En esta secuencia, es posible localizar cuatro residuos de Leu que están conservados en la mayoría de las subunidades reportadas hasta la fecha, incluyendo las equivalentes en bacterias. Estos residuos de Leu están separados entre sí por 7, 6 y 5 residuos no conservados respectivamente. Este arreglo corresponde a una probable estructura de zipper de leucina, en donde de manera característica se localizan alrededor de cuatro residuos de Leu dispuestos cada 7 residuos. La presencia de este tipo de arreglo podría indicar que la subunidad NAD4L además de tener un papel en el complejo I mitocondrial, podría estar involucrada en otro tipo de procesos biológicos como la transcripción y/o activación de la replicación, o como transmisores de señales en la comunicación con el núcleo y con otros organelos celulares (Le y col., 2001).

Resultados 2:
El gen *cox2b* de *S. obliquus*

VI. RESULTADOS 2: El gen *cox2b* de *Scenedesmus obliquus*

El grupo de las algas clorofitas, de acuerdo a las características generales de sus genomas mitocondriales, ha sido dividido en dos grandes grupos: uno denominado como "ancestral" y que incluye a organismos como *Prototheca wickerhamii* y otro denominado "reducido derivado" en el que se incluyen los organismos del género *Chlamydomonas*. En los genomas ancestrales, se han conservado ciertas características similares al endosimbionte bacteriano original como la presencia de un rRNA 5S, un juego casi completo de tRNAs, unos cuantos intrones y conjuntos de genes arreglados a manera de operones. En cambio, en los genomas tipo *Chlamydomonas*, se han perdido una enorme cantidad de genes, se ha modificado la organización y estructura de los rRNAs y en algunos casos se ha modificado el código genético que utilizan (Nedelcu y col., 2000; Turmel y col., 1999).

Hasta la fecha, han sido secuenciados por completo los genomas mitocondriales de ocho clorofitas: tres de ellos corresponden al grupo ancestral (*Prototheca wickerhamii*, *Nephroselmis olivacea* y *Mesostigma viride*), cuatro al grupo derivado reducido (*C. reinhardtii*, *Chlamydomonas eugametos*, *Chlorogonium elongatum* y *Pedinomonas minor*) y uno ha sido descrito como estado intermedio entre los dos anteriores (*Scenedesmus obliquus*) (Denovan-Wright y col., 1998; Kroymann y Zetsche, 1998; Kück y col., 2000; Michaelis y col., 1990; Nedelcu y col., 2000; Turmel y col., 1999; Turmel y col., 2002; Wolff y col., 1994).

El genoma mitocondrial de *S. obliquus* es una molécula circular de cerca de 43 kb, contiene 42 genes conservados y cinco ORFs, todos presentes como copia única. Los genes conservados corresponden a: rRNAs de las subunidades grande y pequeña fragmentados en módulos, 27 tRNAs y 13 genes que codifican proteínas de la cadena respiratoria. Los genes reportados en estos trabajos aparecieron anotados como *nd1*, *nd2*, *nd3*, *nd4*, *nd4L*, *nd5*, *nd6*, *cob*, *cox1*, *cox2*, *cox3*, *atp6* y *atp9* (Kück y col., 2000; Nedelcu y col., 2000). El gen *cox2* reportado en estos trabajos fue identificado por Nedelcu y col. (2000) como un pseudogene, ya que la secuencia deducida de aminoácidos es aproximadamente 100 residuos mas corta que otras subunidades reportadas, además de que la región faltante incluye motivos altamente conservados que son indispensables para la función del complejo proteico.

Como resultado del trabajo en el laboratorio, se demostró que en las algas *C. reinhardtii* y *Polytomella* sp. el gen *cox2* se encuentra fragmentado en dos genes independientes (*cox2a* y *cox2b*) que se localizan en el genoma nuclear (Pérez-Martínez y col., 2001., Apéndice II). Al realizar un análisis de comparación entre los dos genes descritos y el gen *cox2* encontrado en *S. obliquus*, se observó que la proteína COX II predicha tiene un alto porcentaje de similitud con COX IIA de *C. reinhardtii* y *Polytomella* sp. En este trabajo, se propuso que el gen *cox2* encontrado en el

genoma de *S. obliquus* no es un pseudogene, sino que en realidad se trata de un gen *cox2a* activo y que el gen *cox2b* ha sido transferido al núcleo en estos organismos.

Utilizando como sonda el gen *cox2b* de *C. reinhardtii*, se realizó una búsqueda exhaustiva en una biblioteca genómica de *S. obliquus* (donada por la Dra. Sabeeha Merchant, Universidad de California en Los Angeles, USA (Quinn y Merchant, 1999)). Se obtuvo una clona positiva que contenía un fragmento del gen *cox2b*.

La secuencia genómica parcial del gen *cox2b* de *S. obliquus* obtenida en este trabajo, es de 2014 pb. Dentro de esta secuencia se puede predecir un ORF de 318 pb que permite deducir una proteína de 105 residuos. Esta secuencia incluye la región de COX IIB homóloga a la región carboxilo de las proteínas COX II ortodoxas. Es posible predecir que en el gen debe estar también codificada una región equivalente a la extensión amino terminal de COX IIB que se encuentra en *C. reinhardtii* y *Polytomella* sp., sin embargo este hecho debe ser confirmado en futuros experimentos.

Si comparamos la secuencia deducida de aminoácidos de los genes *cox2a* y *cox2b* de *C. reinhardtii* y *Polytomella* sp. con las estructuras cristalográficas reportadas, la subunidad COX IIA corresponde a la región más hidrofóbica de la proteína madura, incluyendo a los dos cruces transmembranales, mientras que COX IIB corresponde a la porción hidrofílica (Pérez-Martínez y col. 2001).

El hecho de que en *S. obliquus* el gen *cox2* haya sido fragmentado en *cox2a* y *cox2b* y que uno de los genes haya sido enviado al genoma nuclear, es un argumento más a favor de que el genoma mitocondrial de *S. obliquus* podría representar un estado intermedio en la evolución de los genomas mitocondriales de las algas clorofitas.

A partir de los genes encontrados, es posible deducir algunos eventos en las historias evolutivas de este grupo de algas clorofitas: en el ancestro común de *C. reinhardtii*, *Polytomella* sp. y *S. obliquus* (pero ya no de organismos como *P. wickerhamii*) debió existir forzosamente un genoma mitocondrial en el que el gen *cox2* se encontrara fragmentado en *cox2a* y *cox2b*. En cuanto al orden de transferencia de los genes, una opción sería que el gen *cox2b* fue transferido al núcleo en este ancestro común. Más tarde los dos grupos se separaron y una vez que se encontraron independientes, en el grupo de *Chlamydomonas* se transfirió el gen *cox2a*, mientras que en el grupo de *Scenedesmus* este gen se retuvo en el organelo. Este hecho se vería favorecido ya que *cox2b* codifica una proteína hidrofílica y por lo tanto la importación del producto peptídico no resultaría complicada. El que el gen *cox2a* todavía se conserve en el genoma mitocondrial en este organismo, puede ser resultado de que el código genético utilizado en este

genoma mitocondrial es diferente al código genético universal (Nedelcu y col., 2000, Kück y col., 2000). Es posible entonces que aunque el genoma de *S. obliquus* represente un estado intermedio de evolución, ésta haya quedado "congelada" en este organismo, ya que transferir ahora *cox2a* al núcleo implica un nuevo reto: el cambio en el código genético (Figura VI.1).

Otra opción sería que la transferencia de ambos genes haya sucedido de manera independiente después de que los grupos se separaron, esta versión ha sido propuesta por Watanabe y Ohama (2001) para el caso concreto de *C. reinhardtii*, en función del número de intrones presentes en cada uno de los genes. Partiendo de que los genes originales mitocondriales no tenían intrones, el que *cox2a* presente una cantidad mayor de intrones por kb de secuencia comparada con el gen *cox2b*, indicaría que el primer gen en ser transferido al núcleo debió ser *cox2a*. Esta secuencia de eventos no correspondería al caso de *S. obliquus*, ya que en este organismo el gen *cox2b* es el único que ha sido transferido al núcleo, por lo tanto esto sugeriría que ambos linajes se separaron antes de que ocurriera la transferencia al núcleo de cualquiera de los dos genes.

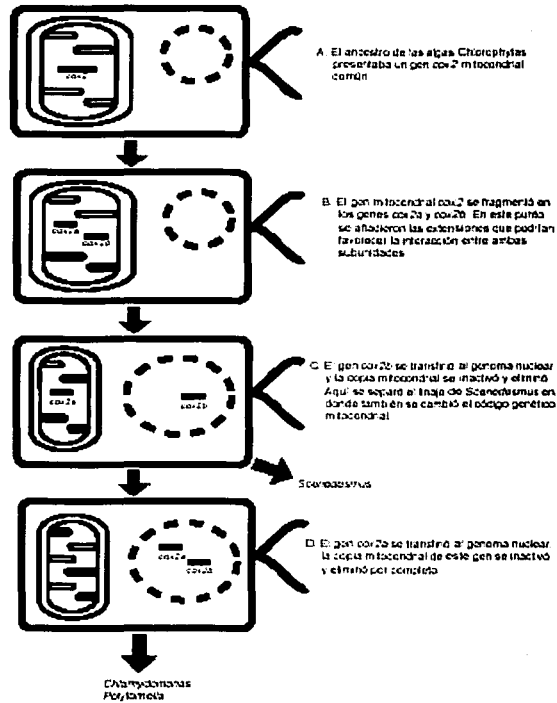


Figura VI.1. Hipótesis sobre la historia evolutiva del gen *cox2* en las algas clorofitas

Resultados 3:
Los genes *cox2a* y *cox2b* de *T. gondii* y otros apicomplejos

La subunidad COX II es uno de los componentes esenciales del complejo citocromo c oxidasa tanto de bacterias como de mitocondrias eucariontes. Es una proteína compuesta por tres segmentos: una región pequeña amino terminal que se localiza en el espacio intermembranal, dos cruces transmembranales y una región carboxilo terminal hidrofílica localizada también en el espacio intermembranal. Ésta subunidad transfiere los electrones del citocromo c hasta los centros $a a_3$ de la subunidad I que se encuentra embebida en la membrana interna mitocondrial (Figura VII.1) (Tsukihara y col., 1996).

En la mayor parte de los organismos eucariontes está codificada por un sólo gen que se localiza en el genoma mitocondrial (*cox2*). Sin embargo, se han identificado tres grupos de organismos en los que este gen no se encuentra en el mtDNA: i) algunas plantas leguminosas (Adams y col., 1999); ii) algunas algas de la clase chlorophyceae (Pérez-Martínez y col., 2001); y iii) los parásitos apicomplejos (revisado en Gray y col., 1999).

El caso particular de las algas de la clase Chlorophyceae ha resultado ser un evento extraordinario en la evolución: en *C. reinhardtii* y *Polytomella* sp., COX II está codificada como dos genes nucleares independientes, y de hecho cada uno de estos genes produce un polipéptido independiente (COX IIA y COX IIB). El gen *cox2a* codifica la subunidad COX IIA que corresponde a la región amino terminal y porción membranal de las subunidades COX II comunes; y el gen *cox2b* codifica a la subunidad COX IIB que corresponde a la región soluble de las COX II comunes y que contiene el sitio de unión a cobre. Ambas subunidades son importadas por la mitocondria y forman parte del complejo IV funcional en estos organismos. Se cree que estas dos subunidades interactúan entre sí a través de extensiones que se encuentran en el extremo carboxilo de COX IIA y amino de COX IIB para acoplarse de manera similar a como sería una COX II típica. Ambas extensiones se encuentran presentes en las proteínas maduras y no presentan similitud con ninguna región de otras subunidades COX II reportadas (Figura VII.2) (Pérez-Martínez y col., 2001).

El hecho de que el gen ortodoxo mitocondrial *cox2* esté ausente en el grupo de los apicomplejos no resulta del todo sorprendente, ya que los genomas mitocondriales de estos organismos parásitos constituyen el grupo de mtDNAs más compactos que se han caracterizado hasta la fecha. En varias especies del género *Plasmodium* los genes mitocondriales se encuentran en una unidad lineal de un poco menos de 6 kb que se encuentra repetida múltiples veces y estas copias están concatenadas unas con otras. En los organismos del género *Theileria*, las unidades lineales se encuentran de manera independiente, y presentan secuencias repetidas en los extremos. Los genomas secuenciados hasta la fecha de éstos dos géneros contienen tres genes que codifican proteínas de la cadena respiratoria mitocondrial (CYTb, COX I y COX III) y genes que

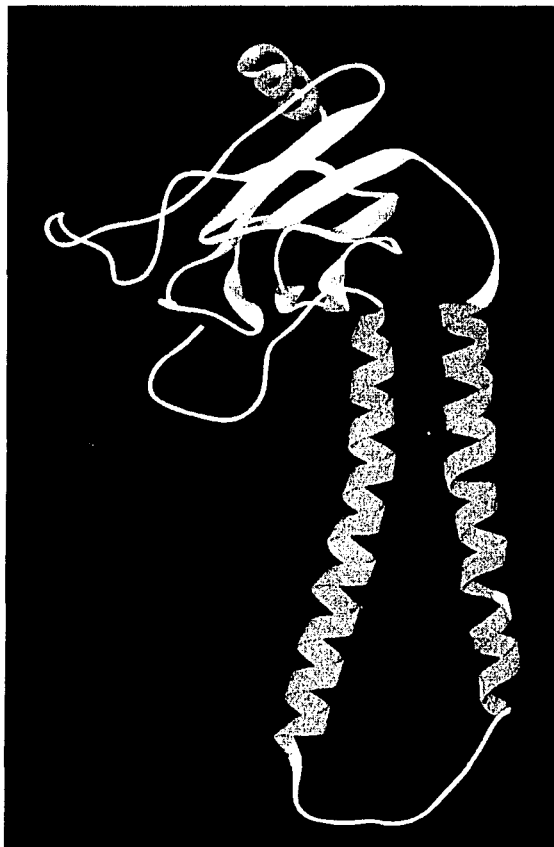


Figura VII.1. Estructura cristalográfica de la subunidad COX II.

TESIS CON
FALLA DE ORIGEN

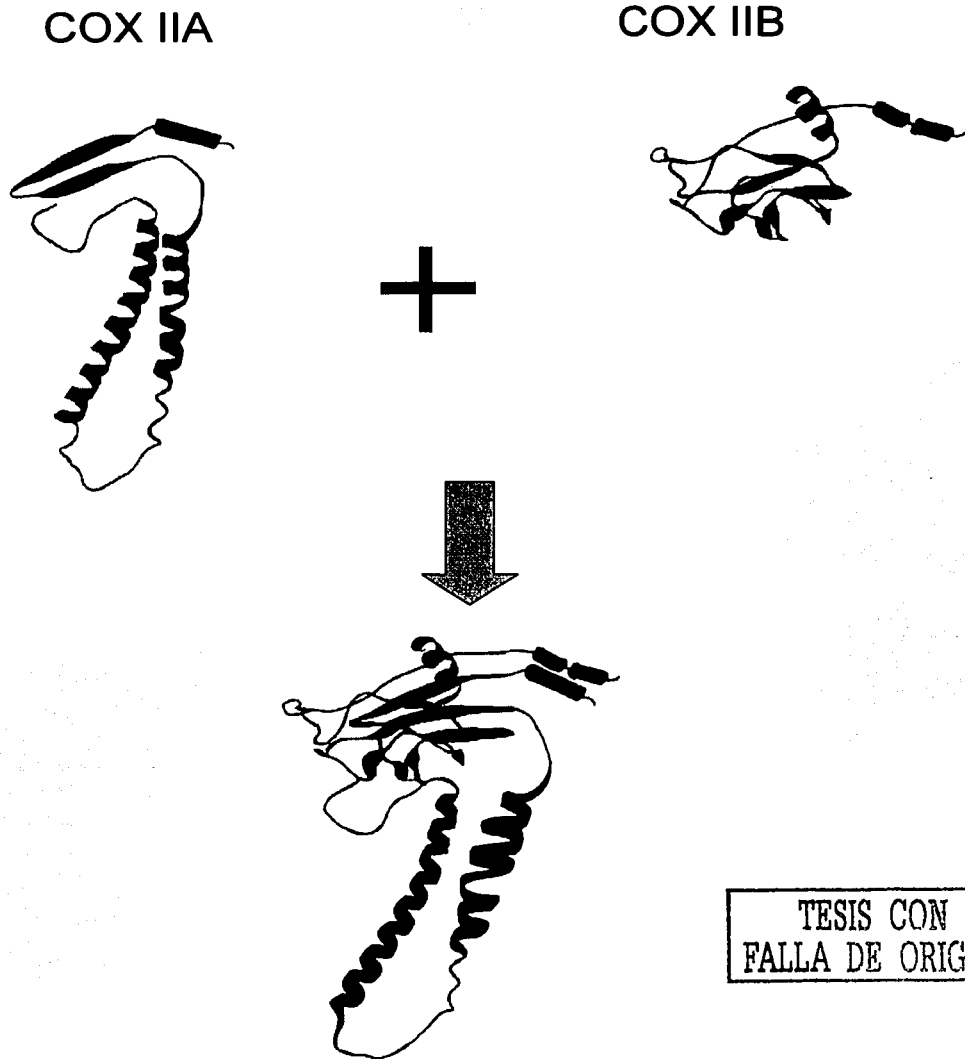


Figura VII.2 Modelo propuesto para la interacción de las subunidades COX IIA y COX IIB de *Polytomella* sp. (Tomado de Pérez-Martínez y col., 2001)

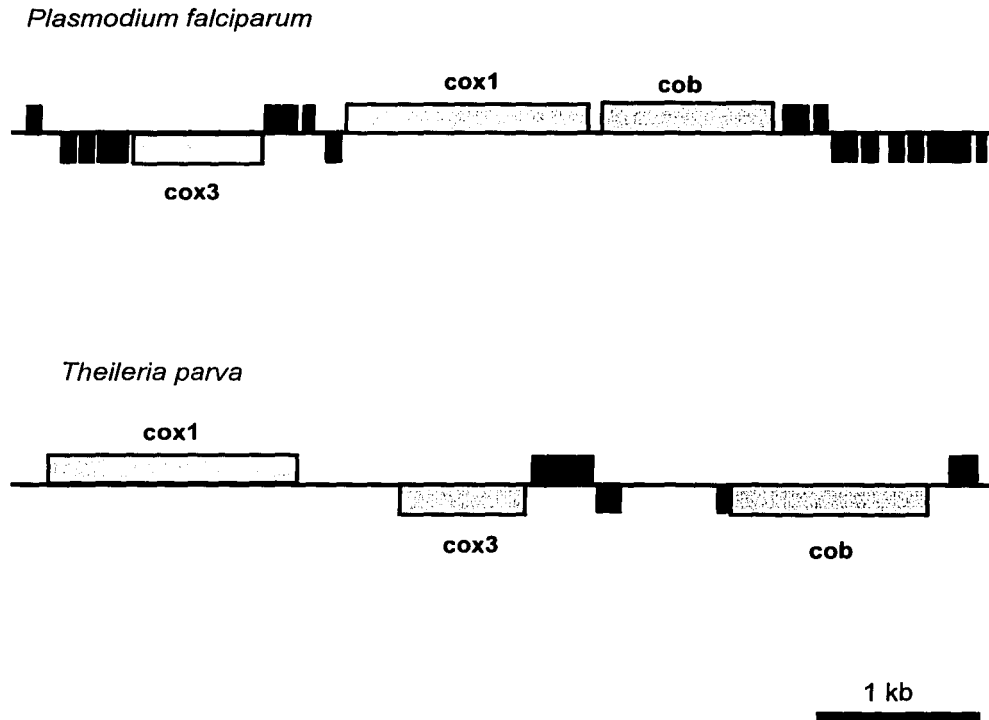
codifican rRNAs ribosomales fragmentados en módulos. En *P. falciparum* se encontraron tres ORFs adicionales que no están en otras especies del mismo género y que no presentan similitud con otros genes reportados. El mtDNA de *Toxoplasma gondii* ha resultado más difícil de caracterizar que el de sus parientes cercanos, ya que se han encontrado numerosos fragmentos de secuencias de origen mitocondrial flanqueadas por secuencias repetidas en el genoma nuclear de este organismo (Figura VII.3) (revisado en Feagin, 2000).

Los genes *cox2a* y *cox2b* de *T.gondii*

Realizando un análisis de las bases de datos se identificó que una de las proteínas con las que COX IIB de *Polytomella* sp. mostraba un alto porcentaje de similitud, era una proteína reportada como COX II de *T.gondii*. El gen reportado que codifica esta proteína (AAF07939) está anotado en las bases de datos como el gen *cox2* mitocondrial de este organismo. Analizando con un poco de detalle la secuencia deducida de aminoácidos, fue evidente que de la proteína reportada sólo la región amino terminal presentaba similitud con COX IIB de *Polytomella* sp., mientras que la región carboxilo terminal no presentaba similitud con ninguna otra subunidad COX II (a pesar de ser de un tamaño similar a otras subunidades típicas). La secuencia de aminoácidos fue deducida en este caso utilizando el código genético mitocondrial reportado para musgos, protistas y celenterados; sin embargo, cuando se obtuvo la secuencia correspondiente con el código genético universal, se identificó que el marco de lectura reportado se detiene en una posición equivalente al final de la proteína COX IIB de *Polytomella* sp. y de *C. reinhardtii*. Se planteó entonces que tal vez no se trataba de un gen mitocondrial, sino que en realidad se trataba del gen *cox2b* nuclear de *T.gondii*.

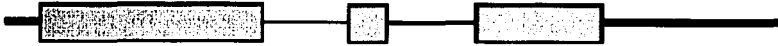
Utilizando como base los genes *cox2a* y *cox2b* de *Polytomella* sp. se realizaron búsquedas en las bases de datos de ESTs de parásitos y en particular de *T.gondii* que hasta ese momento se encontraban disponibles (<http://ParaDB.cis.upenn.edu/toxo/>). Se logró identificar la secuencia completa de traducción de ambos genes y con base en estas secuencias se diseñaron oligodesoxinucleótidos para amplificar por PCR los genes completos.

El gen *cox2a* de *T.gondii* es de 2454 pb, corresponde a la secuencia genómica que incluye todas las ESTs reportadas de esta región hasta la fecha. La secuencia incluye 122 pb de la región 5' no traducida, la secuencia entre el nucleótido 123 y el 1870 que corresponden al marco de lectura abierto uniendo las secuencias entre las posiciones 123-779, 1065-1182 y 1746-1870; y la región 3' no traducida que incluye desde el nucleótido 1870 hasta el 2454 donde termina la clona (Figura VII.4).



TESIS CON
FALLA DE ORIGEN

Figura VII.3. Genomas mitocondriales de los apicomplexos *Plasmodium falciparum* y *Theileria parva*. En gris se muestran los genes que codifican proteínas mitocondriales y en negro los que codifican rRNAs.

cox2a*cox2b*

200 pb



Figura VII.4. Organización de los genes *cox2a* y *cox2b* de *T. gondii*. Las regiones codificantes se encuentran señaladas como cajas; las regiones no traducidas se encuentran como líneas gruesas, y los intrones como líneas delgadas.

La secuencia del gen *cox2b* de *T.gondii* es de 1348 pb: incluye la región 5' no traducida de 60 pb, la secuencia que codifica para la proteína y una región 3' no traducida de 713 pb. Al comparar la secuencia genómica obtenida con las bases de datos de ESTs (que son producto de cDNAs) y el ORF completo que codifica para la subunidad, fue posible identificar la presencia de un intrón de 56 pb entre las posiciones 199 y 254 pb (Figura VII.4).

Una vez que se tuvieron los genes completos, se realizó una búsqueda en las bases de datos de los genomas de otros apicomplejos como *P. falciparum*, *P. yoelii* y *T. parva* (<http://plasmodb.org>; <http://www.tigr.org/tdb/edb2/pya1/htmls/> y <http://www.tigr.org/tdb/e2k1/tpa1/> respectivamente). Los genes *cox2a* y *cox2b* se identificaron fácilmente en regiones no contiguas en todos los organismos, de hecho en el caso del genoma de *P. falciparum*, fue posible determinar que cada uno de los genes se encuentra en cromosomas diferentes (*cox2a* se encuentra en el cromosoma 13 mientras que *cox2b* se encuentra en el cromosoma 14). Esto permite plantear que la existencia de estos genes es común a todos los apicomplejos.

COX IIA

En COX IIA de *Polytomella* sp., se identificó experimentalmente la secuencia amino terminal de la proteína madura y se planteó el sitio equivalente en *C. reinhardtii* mediante análisis de similitud entre ambas secuencias (Pérez-Martínez y col. 2001). En el caso de los apicomplejos analizados, la extensión en el extremo amino terminal es claramente identificable si se compara con las proteínas codificadas en la mitocondria (como en el caso de COX IIA de *S. obliquus*), sin embargo la similitud a nivel de estructura primaria no es significativa. Esta extensión amino terminal, podría estar actuando como MTS tanto en el caso de las clorofíceas como de los apicomplejos. Como se mencionó anteriormente, lo importante para que una MTS sea funcional es el tipo de residuos que la constituyen y la estructura secundaria que puedan formar. Por ésta razón no resulta sorprendente que a nivel de estructura primaria las secuencias no presenten un valor significativo de similitud.

En general, las proteínas COX II de estos organismos no se encuentran altamente conservadas, a excepción de la región aromática altamente conservada GRQWYWS que se encuentra muy cerca del extremo carboxilo terminal. El caso de *Theileria parva* parecería único ya que presenta la secuencia GHQWYWS utilizando una His en lugar de la Arg que se encuentra en todos los organismos caracterizados hasta la fecha; sin embargo podría tratarse de un error en la secuencia reportada, ya que se trata de resultados preliminares del proyecto de secuencia del genoma completo de este organismo (Figura VII.5).


```

P. yoo111      MFNHL-----SKLYFG-----
T. parva      MLYVYGI-----TRLLFGSS-----
T. gondii      MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
Polytomella sp. MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
C. salinarum11 MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
S. zeliqoux    MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-

P. yoo111      -----LNNINQUTSNNIYVANSKRIYVFPF---FYVHEFTYTKKIDINN---IVYDI
T. parva      -LLNITKPTFCQHFGLYFNHYVQRHIFTSRLLI---FTTKEFEANINSTDEESVNH
T. gondii      SCQGRPAQVFPFANIVYVPLSAHTFENINPLSSFLRFYGHLEFPPKRFESGLR-GYVSL
Polytomella sp. MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
C. salinarum11 MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
S. zeliqoux    MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-

P. yoo111      HIKTYAEN-----NHGNGEN--TEGENTHVDHNNHNPICHL
T. parva      FNQKQFTVWSETPCGEYNGTGLEKPKVQGFEDLPSIEGKPI-QGNEFVNHLEIDPRCHL
T. gondii      MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
Polytomella sp. MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
C. salinarum11 MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
S. zeliqoux    MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-

P. yoo111      TEDSTRFPEYI FTYVYRQVGMNI FNNIILN YNNINPHEINLEI FVWVQVNI FVH
T. parva      REDGRPREKTAQDQFVYVATVTRKSRKEDIANRANVYVETVIEEELVRLIYVYFY
T. gondii      HEDYTHRELQIAE I FVGEHLSAEI FVAGVNSRKHEDVETVTRHLEFLELVH I N I FVH
Polytomella sp. MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
C. salinarum11 MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
S. zeliqoux    MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-

P. yoo111      SRILKAGNHHNLENEPLFVW--FPLITETIATFQPHLSHLEGIN-VYTRIVEVIG
T. parva      HENGLHAGDHEKALGKAPCFVW--FPLITETIATFQPHLSHLEGIN-VYTRIVEVIG
T. gondii      FVHRESEVNHDLI FVAGSPLFVW--FPLITETIATFQPHLSHLEGIN-VYTRIVEVIG
Polytomella sp. MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
C. salinarum11 MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-
S. zeliqoux    MLYVYASRELAWVFGVGTLPENTFLFLRRLLIIPASSFQSLPFG-SKLYPIR-

P. yoo111      RGNWYVEVESHVDEDE-----
T. parva      RGNWYVEVESHV-----
T. gondii      RGNWYVEVESHVDEDE-----
Polytomella sp. RGNWYVEVESHVDEDE--DADRVAIAEKITK
C. salinarum11 RGNWYVEVESHVDEDE--DADRVAIAEKITK
S. zeliqoux    RGNWYVEVESHVDEDE--DADRVAIAEKITK

```

Figura VII.5. Comparación entre la secuencia de COX IIA de algunos apicomplexos y algunas clorofitas. Los asteriscos señalan residuos idénticos y los dos puntos residuos conservados. La secuencia amino-terminal de la proteína madura de *Polytomella sp.* se encuentra encerrada en una caja, la secuencia que correspondría a la MTS se encuentra sombreada, y en negritas la secuencia correspondiente a la extensión carboxilo terminal de COX IIA. Los triángulos invertidos indican la posición de los intrones.

TESIS CON
FALLA DE ORIGEN

C. reinhardtii, *Polytomella* sp. y *S. obliquus* presentan también una extensión en el extremo carboxilo terminal que se ha propuesto que permite la interacción entre COX IIA y COX IIB para conformar el complejo final. Esta extensión podría estar presente (aunque de manera reducida) en *T. gondii* y *P. falciparum*, mientras que parece estar ausente en *T. parva*.

COX IIB

Como se dijo anteriormente, la proteína COX IIB corresponde a la región soluble de las comunes COX II, aquí se encuentra la secuencia GQCSE(I/L/M)CG que une los átomos de cobre responsables de la transferencia de electrones. Esta es la zona que tiene una mayor conservación a nivel de estructura primaria en todos los organismos, y se observa claramente la gran similitud que existe entre las proteínas de las algas clorofíceas y las de los apicomplejos.

El alto nivel de similitud entre los apicomplejos y las algas clorofíceas se observa fundamentalmente en la región que corresponde a la carboxilo de una COX II común. La extensión en el extremo amino terminal de esta subunidad presenta muy pocos residuos conservados entre los dos grupos de organismos, sin embargo la sola presencia de esta extensión para poder interactuar con COX IIA se puede considerar como un elemento importante al tratar de determinar la historia evolutiva de estos organismos (Figura VII.6).

Aunque todavía es necesario realizar estudios de importación y ensamblaje en los apicomplejos, es posible suponer que las dos subunidades COX II están siendo sintetizadas en el citoplasma celular y dirigidas específicamente hacia la mitocondria (COX IIA gracias a una MTS y COX IIB posiblemente por la presencia de algún tipo de secuencia de reconocimiento interna). Una vez en el organelo, COX IIA debe insertarse en la membrana interna y acoplarse a la subunidad COX I (que es de origen mitocondrial), mientras que COX IIB debe mantenerse en el espacio intermembranal para acoplarse a COX IIA. Entonces, el que la secuencia de aminoácidos de COX IIA se encuentre menos conservada entre los apicomplejos y las clorofíceas no es del todo extraño, ya que esta es la región que tiene que interactuar con la subunidad COX I cuyo origen evolutivo se encuentra en las mitocondrias del ancestro apicomplejo. Durante la evolución, debieron favorecerse las mutaciones que permitieran la interacción adecuada entre estas dos subunidades de origen distinto.

Sería importante que se realizaran estudios para determinar sin ambigüedades cómo es que interactúan las subunidades COX IIA y COX IIB entre sí. Por ejemplo si es que la extensión carboxilo de COX IIA tiene verdaderamente un papel funcional en la formación del complejo.

¿Existe alguna relación entre los Apicomplejos y las Clorofitas?

Los parásitos apicomplejos contienen un plástido rodeado por cuatro membranas: el apicoplasto (Köhler y col., 1997). En el interior de este plástido, existe un genoma circular de entre 30 y 35 kb que se caracteriza por tener un alto contenido de A-T (alrededor del 86%) y los genes que codifican se encuentran dispuestos de manera muy compacta (con regiones intergénicas muy pequeñas); codifica genes para rRNAs de las subunidades pequeña y grande, 25 tipos de tRNAs, proteínas ribosomales y RNA polimerasas (Wilson y col., 1996). La secuencia de este genoma se ha analizado con mucho detalle, fundamentalmente tratando de determinar el origen de este organelo, y hasta ahora lo único que se puede afirmar con certeza es que se trata de un organelo derivado de un cloroplasto. El que el apicoplasto se encuentre rodeado por cuatro membranas (y no por dos como las mitocondrias y los cloroplastos), ha permitido imaginar que se trata de un organelo adquirido por un proceso de endosimbiosis secundaria entre el ancestro de los apicoplastos y un alga unicelular: las dos membranas internas del apicoplasto corresponderían a las membranas del cloroplasto original, la tercera membrana correspondería a la membrana del alga y finalmente la membrana externa provendría de la vacuola formada durante la "ingestión" del alga (Köhler y col., 1997).

Qué tipo de alga fue la que estuvo involucrada en el fenómeno endosimbiótico, todavía es un tema de numerosos debates. Hasta ahora se ha propuesto que pudo haberse originado a partir de dos tipos diferentes de algas: un alga verde (Chlorophyta) o un alga roja (Rhodophyta). El análisis de la estructura de estos organismos, junto con estudios cladísticos a nivel de DNA de genes nucleares cuyos productos actúan en los plástidos y de la gliceraldeído-3-fosfato deshidrogenasa, sugieren un origen común entre los plástidos localizados en los apicomplejos y en los dinoflagelados, y por lo tanto se planteó su origen en las algas rojas (Figura VII.7) (Blanchard y Hicks, 1999; Fast y col., 2001). El origen de los apicoplastos a partir de un alga clorofita fue propuesto con base en la similitud que muestran los genes *tufA* localizados en el genoma del apicoplasto y en los cloroplastos de las clorofitas (Köhler y col., 1997). Las filogenias basadas en ciertos marcadores moleculares como la glucosa-6-fosfato isomerasa y la enolasa, muestran una estrecha relación entre los apicomplejos y las proteínas de plantas (Dzierszinski y col., 1999); y relaciones filogenéticas establecidas utilizando la secuencias de la vía de las pentosas de fosfato (como la 6-fosfogluconato deshidrogenasa) muestran una relación entre los apicomplejos y las cianobacterias (Andersson and Roger, 2002). El vínculo con las plantas, podría apoyar la hipótesis de que el origen se remonta al grupo de las clorofitas, ya que el origen de las plantas terrestres se relaciona con este grupo de algas.

Como se puede observar, determinar la posición precisa en la que se encuentran en los análisis filogenéticos ha resultado muy complicado, y los análisis elaborados con genes

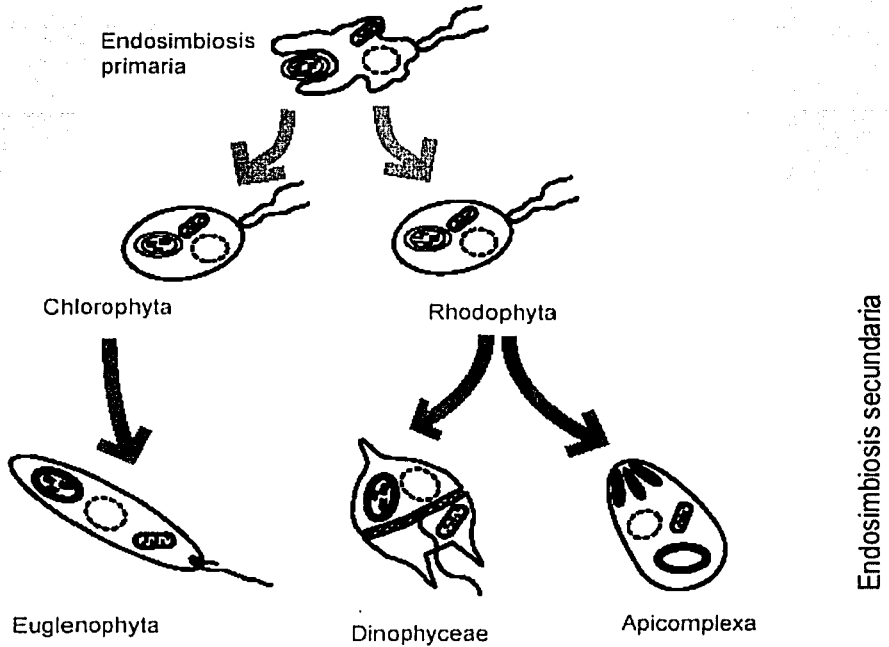


Figura VII.7. Evolución de algunos organismos eucariotes y sus plástidos (Modificado de Moreira y Philippe (2001))

TESIS CON
FALLA DE ORIGEN

individuales produce resultados cofusos y contradictorios; una de las razones por la que esto podría estar sucediendo, es que los genes derivados del apicoplasto presenten tasas muy altas de evolución (Moreira y Philippe, 2001). Blanchard y Hicks (1999), también plantearon que el uso de secuencias primarias tanto de DNA como de aminoácidos, son elementos que dificultan la determinación del origen exacto del apicoplasto que por lo tanto asocian a los apicomplejos con diversos grupos como los organismos del género *Euglena* o los dinoflagelados, sin poder resolver con exactitud cómo se originaron; el uso de la presencia de genes como *cox2* que sufrieron modificaciones únicas y características de un cierto tipo de organismos podría ayudar a esclarecer el origen del alga endosimbionte y por lo tanto del apicoplasto. El análisis realizado en este trabajo con base en los genes nucleares *cox2a* y *cox2b*, ha permitido identificar que el alga que estuvo involucrada en la endosimbiosis secundaria en estos organismos fue una Chlorophyta, en particular un miembro de la clase Chlorophyceae.

En un análisis filogenético entre las secuencias disponibles de COX IIA y COX IIB y algunas secuencias de COX II convencionales codificadas en genomas mitocondriales, los organismos apicomplejos se agrupan con las algas clorofitas. Por otro lado, si se realiza un análisis equivalente con las secuencias del gen *cytb* característicamente mitocondrial, los apicomplejos se agrupan con las secuencias de animales y hongos, una distribución similar a las obtenidas en los análisis clásicos (Degli-Esposti y col., 1993). Esto confirma que los genes *cox2a* y *cox2b* apicomplejos tienen un origen común con los genes equivalentes de las algas clorofíceas, y que no son resultado de una fisión independiente del gen *cox2* mitocondrial.

La presencia los genes *cox2a* y *cox2b* en el genoma nuclear de los apicomplejos (junto con el análisis filogenético realizado) sugiere que originalmente eran los genes que se encontraban en el alga clorofícea con la que el ancestro apicomplejo realizó el fenómeno de endosimbiosis. De manera más particular, podemos restringir aún más el grupo a la clase Chlorophyceae y no a las clases Trebouxiophyceae, Ulvophyceae, o Prasinophyceae (*sensu* Friedl, 1997), ya que en caso de *Prototheca wickerhamii* (miembro de la clase Trebouxiophyceae) o *Nephroselmis olivacea* (miembro de la clase Prasinophyceae) el gen *cox2* es un solo gen codificado en el genoma mitocondrial (Figura VII.8).

A partir de este hecho, es posible plantear varias posibilidades: i) que los genes se encontraran fragmentados en la mitocondria de la clorofícea y que se transfirieran horizontalmente desde el organelo hasta el núcleo del apicomplejo; ii) que el gen *cox2b* se encontrara en el núcleo del alga y *cox2a* todavía en el genoma mitocondrial (como en el caso de *S. obliquus* (ver resultados II) y que se transfirieran desde diferentes organelos hasta el núcleo apicomplejo; y iii) que los genes se encontraran ya localizados en el genoma nuclear de la clorofícea, con todas las

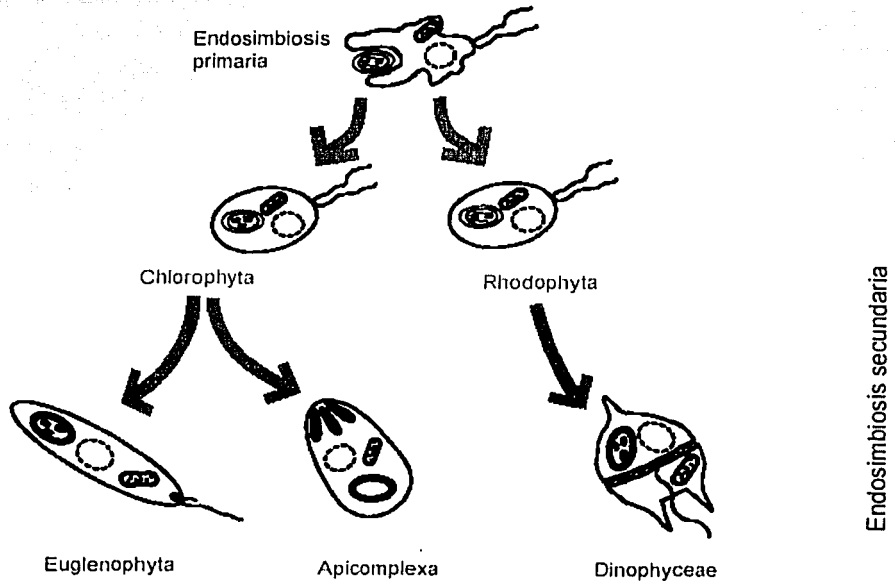


Figura VII.8. Hipótesis sobre el origen de los organismos apicomplexos por endosimbiosis secundaria con un miembro de la clase Chlorophyta.

modificaciones necesarias para que fueran genes activos en el ambiente nuclear y produjeran proteínas capaces de conformar el complejo COX funcional en la membrana interna mitocondrial.

Existen varios elementos que podrían ser indicadores que permitan creer más en alguna de las posibilidades anteriores: a) la presencia de una MTS conservada; b) la presencia de las extensiones que permiten la interacción entre COX IIA y COX IIB; y c) la presencia de intrones conservados tanto en posición como en fase (Cuadro VII.1). En este caso, la MTS de COX IIA no se encuentra conservada en secuencia, aunque las variaciones entre los diversos organismos no necesariamente indican que la transferencia desde la mitocondria fue un fenómeno independiente, ya que como se sabe, las presiones de selección que actúan sobre este tipo de secuencias son más sobre la estructura secundaria que puedan formar que sobre la secuencia de aminoácidos. La existencia de las extensiones carboxilo de COX IIA y amino de COX IIB tampoco es un factor determinante, ya que estas extensiones debieron existir antes de la transferencia al núcleo. Esto se debe a que el gen *cox2* debió fragmentarse en el genoma mitocondrial, como podemos deducir del caso de *S. obliquus*, en el que *cox2a* se localiza en el genoma mitocondrial mientras que *cox2b* ha sido transferido al genoma nuclear. El único elemento que resta para este análisis sería la presencia de dos intrones conservados tanto en posición como en fase entre los genes *cox2a* de *T.gondii*, *Polytomella* sp. y *C. reinhardtii*; esto podría sugerir que al menos el gen *cox2a* fue transferido desde el núcleo del alga cloroficea hasta el núcleo del apicomplejo. Si partimos de la idea de que *S. obliquus* es un paso intermedio de transferencia, podemos suponer que *cox2b* se transfirió al núcleo antes que *cox2a*, por lo que si *cox2a* ya se encontraba en el núcleo, entonces *cox2b* también debía estar en el núcleo del alga endosimbionte y por lo tanto ambos genes fueron transferidos desde el núcleo del alga cloroficea hasta el núcleo del apicomplejo. Lo anterior implica una transferencia de genes de la mitocondria al núcleo y de núcleo a núcleo (Figura VII.9).

Con base en todo lo anterior, es posible plantear una probable historia evolutiva del gen *cox2* en este grupo de organismos partiendo del hecho de que la fragmentación del gen *cox2* es un evento extremadamente raro en la evolución que sucedió (hasta el momento) una sola vez, por lo tanto todos los organismos que presenten los genes *cox2a* y *cox2b* debieron originarse de un ancestro común:

- i) *El ancestro de las chlamydomonádaceas adquirió una inserción en su gen mitocondrial cox2. La inserción pudo haber ocurrido de manera similar a las que se presentan en otros organismos, por ejemplo, Cafeteria roenbergensis que tiene una inserción de 84 pb o Streptomyces coelicor que presenta una inserción de 90 pb; el caso extremo es el de Pylaiella littoralis que contiene una inserción en fase de 3018 pb, produciendo un transcrito que potencialmente codifica una proteína de 1282 residuos en lugar de los 260-280 que se encuentran en una COX II común (Oudot-Le Secq y col., 2001). Estas inserciones dividen*

Cuadro VII.1.

Intrones de los genes *cox2a* y *cox2b* de *T. gondii*, *C. reinhardtii* y *Polytomella* sp.*cox2a*

	Intrón n°	Fase	sitio de corte 5'	sitio de corte 3'
<i>T. gondii</i>	1	0	CAA/GTACGTC	TTTTCAG/AAA
	2	1	CAA/GTAAGAG	TGCGCAG/TTC
<i>Polytomella</i> sp.	1	0	CAG/GTTAATA	TTTATAA/TGC
	2	0	CAG/GTAATTG	CTCACAG/GCC
	3	0	CAG/GTCCGGG	TGTAAGA/ATT
	4	1	TGG/GTAAATA	CAATAG/GCT
	5	0	AAG/GTAAAGC	CAATAG/GCC
	6	0	CAG/GTAAATA	TEGGTAG/ATC
<i>C. reinhardtii</i>	1	0	CAG/GTCCGAC	CTTGCAG/AGC
	2	0	CAG/GTACAGG	ATTGCAG/CAA
	3	0	CAG/GTCCGAA	ATTTAG/GCT
	4	1	GCA/GTAAATC	CGTGCAG/GCG
	5	0	AAG/GTGGGCA	CGTGCAG/GCC
	6	0	CAG/GTCCGCL	TGGACAG/GCT
	7	0	CAG/GTCCGGC	AGCCAG/ATT
	8	1	CAG/GTATCAT	CTTGTAG/GCT

cox2b

	Intrón n°	Fase	sitio de corte 5'	sitio de corte 3'
<i>T. gondii</i>	1	0	CAG/TCACGAC	GCCCCAG/AAA
<i>C. reinhardtii</i>	1	2	CAG/GTCAGTT	ATTGCAG/GTA
<i>S. obliquus</i>	1	0	CAG/GTGGGGG	GCGTCAG/GTG
	2	0	CAG/GTAGGGG	GCTGCAG/AAC

TESIS CON
FALLA DE ORIGEN

a los genes *cox2* mitocondriales en dos regiones que corresponden a los genes *cox2a* y *cox2b* de *P. falciparum*, *P. yoelii*, *T. parva*, *T. gondii*, *C. reinhardtii*, *Polytomella* sp. y *S. obliquus*.

- ii) *El gen mitocondrial cox2 fue fragmentado en dos.* La fragmentación de genes que codifican proteínas mitocondriales y que son transcritos de manera independiente, es un fenómeno relativamente raro, sin embargo se han descrito otros casos: los genes *nad1* de *Tetrahymena pyriformis* y *Paramecium aurelia* (Edqvist y col., 2000), un gen mitocondrial de semilla de nabo que codifica una proteína homóloga a la Ccl1 de bacterias (Handa y col., 1996) o el gen *rp12* de angiospermas (Adams y col., 2001). En el caso de las clorofitas, los genes *cox2a* y *cox2b* retuvieron extensiones en los extremos 5' y 3' respectivamente, que probablemente derivaron de la inserción nucleotídica que dividió al gen *cox2* original. Estas extensiones de nucleótidos codifican extensiones en las proteínas COX IIA y COX IIB de *C. reinhardtii*, *Polytomella* sp. y *S. obliquus* (Pérez-Martínez y col., 2001; Nedelcu y col., 2000; Kück y col., 2000.)
- iii) *Los genes mitocondriales cox2a y cox2b fueron transferidos de manera independiente al núcleo.* Los genes *cox2a* y *cox2b* de *C. reinhardtii* y *Polytomella* sp. son claramente genes nucleares y presentan todas las características típicas de este tipo de genes (Pérez-Martínez y col., 2001). La secuencia parcial de *cox2b* de *S. obliquus* presenta también características nucleares, incluyendo el código genético universal y la presencia de tres intrones. Como se vio en el capítulo anterior, *S. obliquus* podría representar un estado intermedio en el proceso de transferencia de estos genes hacia el núcleo; y podría considerarse una prueba viviente de que los genes se transfirieron de manera independiente.
- iv) *Un ancestro de las clorofíceas actuales, que incluía los genes cox2a y cox2b en el genoma nuclear fue atrapado por un ancestro de los apicomplejos.* Durante este proceso de endosimbiosis secundaria, ambos genes fueron transferidos de manera horizontal desde el núcleo del alga hasta el núcleo del apicomplejo. Actualmente, al menos en *T. gondii*, *P. falciparum*, *P. yoelii*, y *T. parva* estos genes se expresan funcionalmente.
- v) *La adquisición de genes nucleares funcionales que codificaran a las subunidades COX IIA y COX IIB, proteínas capaces de importarse a la mitocondria, permitieron la inactivación y eventual eliminación del gen cox2 mitocondrial presente en las mitocondrias del ancestro apicomplejo.* Como se mencionó anteriormente, en los genomas mitocondriales de los apicomplejos no se ha identificado la presencia del gen *cox2*. La adquisición de los genes *cox2a* y *cox2b* funcionales nucleares y sus productos peptídicos, que originalmente provienen del alga clorofícea, eventualmente reemplazaron la función del gen mitocondrial *cox2* del ancestro apicomplejo, el cual eventualmente fue eliminado del genoma mitocondrial original del apicomplejo.

La fragmentación del gen *cox2* y por lo tanto la existencia de las subunidades COX IIA y COX IIB es un evento raro en la evolución, por lo que el hecho de que los organismos apicomplejos presenten este tipo de conformación, permite suponer que en el evento de endosimbiosis secundaria que dio origen a los apicomplejos actuales, estuvieron involucradas las algas clorofitas, de manera específica un grupo de la clase Chloroficeae donde los genes *cox2a* y *cox2b* ya se encontraban en el genoma nuclear. Durante el evento de intercambio genético entre el organismo endosimbionte y el organismo apicomplejo, estos genes debieron migrar desde el núcleo algal hasta el núcleo apicomplejo. Como ya se trataba de genes capaces de funcionar activamente en el núcleo y sus productos ser dirigidos específicamente hasta la mitocondria, el proceso posterior de eliminación del gen mitocondrial *cox2* debió ser enormemente favorecido, dejando como únicos genes funcionales a los genes nucleares fragmentados recién adquiridos.

Es probable que el gen *atp6* que tampoco se encuentra en el genoma mitocondrial de los apicomplejos, se haya enviado al genoma nuclear, puede ser que este gen también sea de origen cloroficeo, sin embargo los análisis realizados en las bases de datos no permitieron localizar ninguna secuencia con una similitud significativa. ¿Cuántos de los genes mitocondriales (que ya habían sido transferidos al núcleo) fueron adquiridos por los apicomplejos? Esta es una pregunta que debe ser contestada por experimentos futuros.

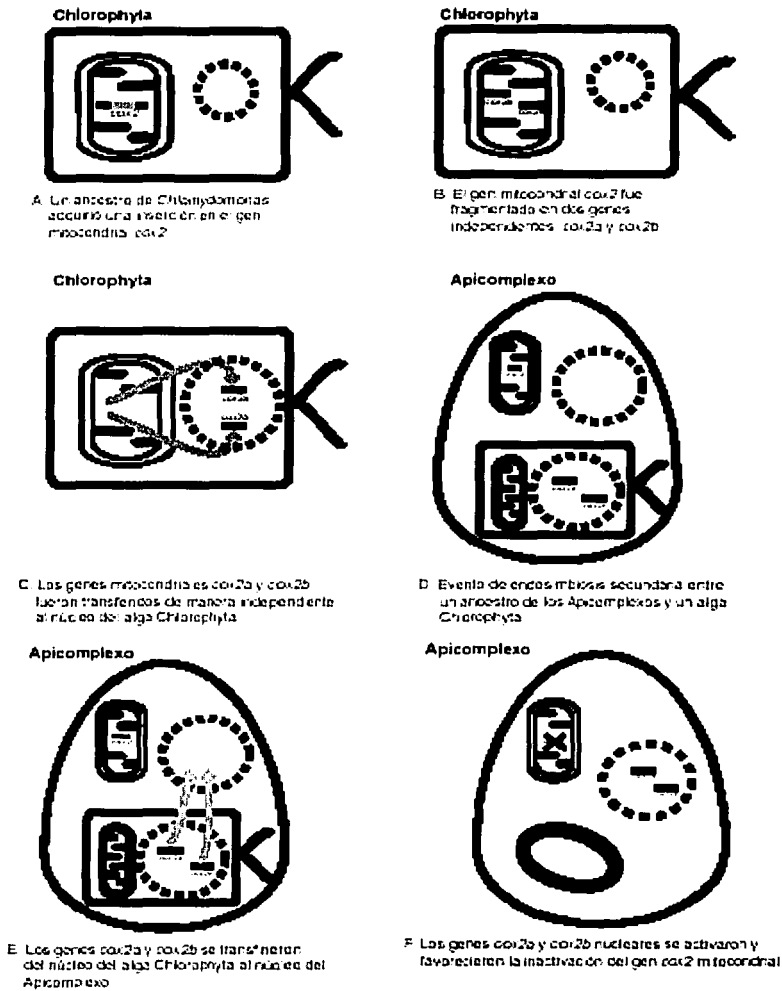


Figura VII.9. Hipótesis sobre la adquisición de los genes *cox2a* y *cox2b* en los Apicomplexos

Discusión

Actualmente, es bastante aceptado que las mitocondrias se originaron por un evento de endosimbiosis único en la evolución entre α -proteobacterias de vida libre (relacionadas con los miembros actuales del género *Rickettsia*) y un hospedero que podría haber sido un miembro de las arqueobacterias o bien un eucarionte sin mitocondrias. Se cree también, que una vez que las mitocondrias se establecieron definitivamente como organelos, comenzó un fenómeno de transferencia de material genético hacia el núcleo celular primero de forma masiva y más tarde de forma gradual. Esta transferencia parece ser un fenómeno que continúa y potencialmente todas las regiones codificantes serán enviadas al núcleo.

A medida que ha ido aumentando el análisis de los genomas mitocondriales, de los genomas nucleares y de genes aislados en especies filogenéticamente emparentadas, se han encontrado genes en el núcleo que representan evidencias de que la transferencia es un proceso que todavía sucede. Fundamentalmente se han reportado casos en plantas y en algas verdes (Adams y col., 2000; Adams y col., 2001; Covello and Gray, 1992; Figueroa y col., 1999a; Figueroa y col., 1999b; Grohmann y col., 1992; Kadowaki y col., 1996; Kobayashi y col., 1997; Kubo y col., 1999; Kubo y col., 2000; Nugent and Palmer, 1991; Pérez-Martínez y col. 2000; Pérez-Martínez y col. 2001; Sanchez y col., 1996; Wischmann and Schuster, 1995.)

Sin embargo, en algunos grupos de organismos la transferencia de genes al núcleo y su activación funcional parece haber sido detenida en algún momento de la evolución. Un ejemplo de este fenómeno, es el caso de los genomas mitocondriales de los vertebrados, ya que el contenido génico es prácticamente constante dentro del grupo (Boore, 1999). A pesar de esto, hay evidencias de una gran cantidad de "intentos" de transferir genes al núcleo, ya que copias de genes mitocondriales se localizan actualmente en los genomas nucleares como pseudogenes (Wallace y col., 1997; Bensasson y col., 2001).

Una de las razones por las cuales la transferencia de genes mitocondriales al núcleo se ha detenido en los animales y no en las plantas, podría relacionarse con la diferencia que existe entre el código genético de las mitocondrias animales y el código genético universal que se emplea en el núcleo celular. En las plantas no existe esta diferencia (Adams y col. 2000). El caso del gen *cox2b* de *S. obliquus* analizado en este trabajo podría también ser un ejemplo de esto, ya que aunque potencialmente el gen *cox2a* podría haber sido también transferido al genoma nuclear, permaneció en la mitocondria posiblemente debido a que en algún punto intermedio en el proceso de transferencia de estos genes, la mitocondria cambió el código genético utilizado, estableciendo un obstáculo más en el proceso de transferencia de genes al núcleo.

Es evidente que la transferencia de genes desde la mitocondria hasta el núcleo no es un proceso sencillo y requiere de grandes cambios funcionales en el interior de la célula. Este hecho hace que de manera inmediata se postulen dos preguntas fundamentales: i) ¿Porqué la células se han tomado todo el trabajo de enviar los genes al núcleo? ii) ¿Qué ha hecho que en los genomas mitocondriales permanezcan ciertos genes? Actualmente, estas dos preguntas están lejos de contestadas de una manera completamente certera, por el contrario, a medida que aumenta el conocimiento acerca de estos organelos y de su relación con el núcleo y el ciclo celular, parecería que la cantidad de respuestas posibles es cada vez más amplia y resulta difícil pensar que una sola de las presiones evolutivas involucradas actúa por sí sola.

¿Porqué un gen se transfiere al genoma nuclear?

Una de las hipótesis planteadas parte del hecho de que en los organelos suceden reacciones de óxido reducción que pueden incrementar la tasa de mutación genética inducida por la presencia de radicales libres, por lo que la información genética podría conservarse de mejor manera estando en un ambiente como el nuclear (Allen y Raven, 1996). Sin embargo, este hecho no se cumple en hongos o en plantas donde la tasa de mutación en los genes nucleares es más alta que en los mitocondriales (Wolfe y col., 1987; Palmer y Herbon, 1988; Lynch, 1997).

Otra explicación es que como los organelos fueron alguna vez bacterias de vida libre, al establecerse la endosimbiosis, perdieron algunos atributos que los caracterizaba como individuos independientes, una de ellas fue que probablemente se transformaron en organismos exclusivamente asexuales que solo se dividían por fisión binaria y por lo tanto cada uno de los organelos vendría siendo un clon del organelo original. Cuando no existe reproducción sexual en un grupo de organismos, no existe recombinación genética y por lo tanto, la producción de mutaciones letales o de pérdidas en el genoma sucederían de manera irreversible. Eventualmente la acumulación de estas mutaciones llevaría a la extinción de la población de organelos. Cuando un gen es transferido al núcleo, se mueve de un ambiente predominantemente asexual hacia un genoma predominantemente sexual, reestabliéndose la recombinación y liberando al gen del peligro de ser eliminado permanentemente (Martin y Herrmann, 1998).

¿Porqué hay genes que se conservan en el genoma mitocondrial?

Como se mencionó anteriormente, en principio las mismas presiones que han hecho que la mayor parte de los genes mitocondriales hayan migrado al núcleo deben estar actuando sobre los genes que permanecen en las mitocondrias y potencialmente todos los genes mitocondriales deberían eventualmente migrar hacia el núcleo. Sin embargo, todas las mitocondrias estudiadas presentan algún tipo de genoma y el tipo de genes que codifican suele ser más o menos el mismo.

Así, de la misma manera como se han planteado hipótesis acerca de porqué los genes originalmente mitocondriales se han establecido de manera permanente en el genoma nuclear, se han planteado numerosas hipótesis acerca de porqué en los genomas mitocondriales todavía se conserva un conjunto limitado de genes.

Una posible explicación ha sido que en las mitocondrias de algunos organismos se utiliza un código genético diferente al código genético que se emplea en el genoma nuclear. Esta heterogeneidad de lenguajes, podría evitar la expresión de los genes mitocondriales transferidos al núcleo. Como se mencionó anteriormente, éste fenómeno ha sucedido de manera independiente en diferentes grupos de seres vivos, y lo que parece es que detiene de alguna manera la transferencia de genes mitocondriales al núcleo, pero no explica porqué cierto grupo de genes permaneció en los genomas mitocondriales durante la migración masiva de genes hacia el núcleo cuando el fenómeno de endosimbiosis apenas comenzaba. Otro proceso que también podría entorpecer la migración de genes mitocondriales al núcleo, es que algunos genes mitocondriales, especialmente en plantas, han acumulado patrones complejos de procesamiento, como la edición del mRNA, que haría imposible el procesamiento del gen en un ambiente nuclear, sin embargo, este problema se solucionaría si el vehículo intermediario entre el genoma mitocondrial y el genoma nuclear fuera la molécula ya editada de mRNA (Wischmann y Schuster, 1995).

Una de las características que se ha mencionado varias veces a lo largo de este trabajo, es que las proteínas codificadas en los mtDNAs son generalmente aquellas que funcionan en la membrana interna mitocondrial y que se caracterizan por ser hidrofóbicas. Se ha propuesto que este fenómeno se ve favorecido porque la síntesis citoplásmica de éste tipo de proteínas podría ocasionar que no se dirigieran al compartimento celular adecuado. Además, al tratarse de polipéptidos hidrofóbicos podría ocurrir que el transporte a través del sistema de membranas mitocondrial resultara imposible (von Heijne, 1986). Se ha propuesto que las proteínas altamente hidrofóbicas, que presentan cuatro o más cruces transmembranales no pueden ser importadas correctamente a la mitocondria (Claros y col., 1995); por lo que estas proteínas deben ser sintetizadas en el interior del organelo para ser insertadas a la membrana interna mitocondrial y adquirir la conformación topológica necesaria para su acoplamiento correcto al complejo respiratorio al que pertenecen. Hay dos ejemplos de proteínas que se han conservado en los genomas mitocondriales de todos los organismos caracterizados hasta la fecha: el gen *cob* que codifica una proteína de ocho cruces transmembranales (Xia y col., 1997) y el gen *cox1* que codifica la subunidad I de la citocromo *c* oxidasa que tiene doce cruces transmembranales (Tsukihara y col., 1996).

Algunas de las proteínas mitocondriales de *C. reinhardtii* han mostrado que la transferencia exitosa al núcleo de los genes que las codifican involucra una gran cantidad de elementos y que

aunque se pueden plantear patrones generales, cada uno de los genes es producto de una serie de modificaciones particulares:

i) ¿Cómo se exportó el material genético desde la mitocondria hasta el núcleo?

Como se mencionó anteriormente, en algunas plantas parece ser que la transferencia de material genético sucedió a través de un intermediario de RNA, la molécula de RNA ya editada debió sufrir un paso obligado de transcripción reversa antes de su inserción en el genoma nuclear. En el caso de *C. reinhardtii*, no es posible determinar si los genes se transfirieron en forma de DNA o de RNA ya que hasta el momento, no se ha encontrado ninguna evidencia de una edición en los genes mitocondriales y tampoco se ha encontrado ninguna actividad de transcriptasa reversa.

ii) Integración al genoma nuclear.

La integración al genoma nuclear, probablemente ocurrió en una región no codificante a través de mecanismos como fusión en los extremos (Blanchard y Schmidth, 1996) o por recombinación no homóloga (Sun y Callis, 1993). El fenómeno de recombinación homóloga (Pichersky y col., 1991) podría descartarse al menos para este grupo particular de genes en este grupo de algas, ya que en general, los genomas mitocondriales de las clorofíceas no presentarían regiones repetidas.

iii) Adquisición de una MTS.

Las presecuencias de los genes *atp6* y *nd4L* no presentan similitudes con otras proteínas reportadas en las bases de datos, así que probablemente no derivaron de las presecuencias de otro gen, como se ha reportado en otros casos (Adams y col., 2000; Adams y col., 2001; Figueroa y col., 1999; Kubo y col., 1999). Este hecho no resultaría sorprendente, ya que como se ha visto, parecería que estos genes característicamente hidrofóbicos necesitan MTSs especialmente largas si se comparan con otras proteínas mitocondriales también de *C. reinhardtii*. Además, se ha reportado que en el caso de las proteínas COX II nucleares que se encuentran en el grupo de las leguminosas, la presecuencia no puede ser substituida por otra presecuencia mitocondrial, mientras que sí son intercambiables en otras proteínas como la presecuencia de la oxidasa alterna que es capaz de dirigir otras proteínas hacia la mitocondria como la glutation reductasa o la ferroquelatasa (Sjöling y Glaser, 1998; Daley y col., 2002). Se ha propuesto que cualquier genoma nuclear contiene un número de MTSs latentes (Kobayashi y col., 1997); además de que aproximadamente un 2.5% de clones de *E. coli* generadas al azar mostraron la capacidad de dirigirse hacia la mitocondria (Baker y Schatz, 1987); por lo que podría decirse de alguna manera que las secuencias que ahora funcionan como MTSs en estos genes estaban ya presentes en el

genoma "esperando" un gen que pudieran enviar a la mitocondria. Es importante recordar que en general, para que una secuencia pueda funcionar como MTS más que la estructura primaria, lo importante es la estructura secundaria que puedan adquirir: son regiones ricas en residuos cargados positivamente, sin residuos cargados negativamente y tienen la capacidad de formar *α*-hélices anfifílicas. La estructura anfifílica es importante para el reconocimiento y unión con los receptores localizados en la membrana externa mitocondrial, y la carga neta positiva puede ser necesaria para el proceso de importación a través de la membrana interna mitocondrial (Emanuelsson y von Heijne, 2001). Las MTSs identificadas en las proteínas ATP6 y NAD4L (reportadas en este trabajo) y COX III y COX IIA de otros resultados obtenidos en el laboratorio en *C. reinhardtii*, muestran que son particularmente largas, de 107, 133, 119 y 143 residuos de aminoácidos respectivamente ricas en residuos cargados positivamente, especialmente alaninas, prolinas y argininas. En *S. cerevisiae* se ha observado que la duplicación de las MTSs favorece la importación *in vivo* e *in vitro* de proteínas hidrofóbicas a la mitocondria (Galanis y col., 1991), tal vez porque mejora de alguna manera la interacción entre el precursor y la maquinaria de importación (Claros y col., 1996). Otra opción, sería que una MTS larga afecte de alguna manera el plegamiento de la proteína aumentando su capacidad de importación (Claros y col., 1995). En general, las MTSs tienen dos sitios de edición, sin embargo, en el caso de la proteína COX II que se encuentra codificada en el núcleo de *Glycine max*, se ha demostrado que también tiene una presecuencia relativamente larga (de 136) y que el fenómeno migración hasta la mitocondria involucra tres segmentos diferentes de la secuencia: el primer segmento debe dirigir a la proteína hasta la mitocondria; la segunda sección guía a la proteína para que se introduzca completamente en la mitocondria y finalmente el tercer segmento lleva a la proteína hasta el anclaje correcto que debe tener para acoplarse al complejo funcional. Cada una de las regiones de la MTS es eliminado después de localizar a la proteína (Daley y col., 2002). Es probable que en el caso de las proteínas estudiadas en *C. reinhardtii*, también exista un mecanismo de translocación de las proteínas hidrofóbicas similar, ya que por ejemplo, la región de la presecuencia que es identificada *in silico* como MTS es aquella que corresponde a los primeros residuos.

iv) Adquisición de elementos nucleares como intrones, promotores y sitios de unión a ribosomas.

La mayor parte de los genes nucleares en *C. reinhardtii* tienen intrones, con un promedio de cuatro intrones por cada kb de secuencia codificante (Silflow, 1998). En el caso del gen *atp6* este fenómeno de adquisición de intrones es muy evidente, sin embargo en el caso del gen *nd4L* no se identificó ningún intrón. Los intrones de *atp6* muestran los sitios consenso de edición: la secuencia GT en el extremo 5' del intrón y AG en el extremo 3' del intrón.

v) *Adquisición de una señal de poliadenilación nuclear.*

En *C. reinhardtii* la señal de poliadenilación más común en los genes nucleares es TGTA cuando se localiza entre 10 y 20 pb antes del sitio de poliadenilación. Esta señal fue identificada tanto en el gen *atp6* como en *nd4L*.

vi) *Cambio en el uso de codones.*

C. reinhardtii utiliza el código genético universal tanto en la mitocondria como en el núcleo, hecho que facilita la translocación de genes mitocondriales al núcleo. Sin embargo, el uso preferencial de codones utilizado en el núcleo es radicalmente distinto al uso de codones mitocondrial. El genoma nuclear tiene un alto contenido de G-C y por lo tanto los codones más utilizados son aquellos que presentan una G o una C en la tercera posición. En *S. obliquus*, el código genético mitocondrial si es diferente, por lo que además este organismo la transferencia de genes parece haberse detenido al igual que en otros grupos de organismos que lo han modificado.

vii) *Inactivación y pérdida de la copia mitocondrial de los genes.*

Hasta ahora en el grupo de las algas clorofíceas, no ha sido posible identificar ninguna etapa intermedia de inactivación de los genes mitocondriales, es decir, no hay ningún ejemplo en el que se haya encontrado que la copia mitocondrial de un gen transferido al genoma nuclear se encuentre modificado en el genoma mitocondrial de tal manera que su expresión sea imposible. Cuando se publicó el genoma mitocondrial de *S. obliquus*, Nedelcu y col. (2000) plantearon que el gen que presentaba similitud con el gen *cox2* se trataba de un pseudogen y que el gen *cox2* debía encontrarse en el núcleo, de manera similar a como sucede en el grupo de las leguminosas; sin embargo, al conocer el caso de fragmentación génica que se observa en *C. reinhardtii* y *Polytomella* sp. (Pérez-Martínez y col., 2001) fue posible concluir que el gen que se encuentra en la mitocondria es en realidad *cox2a* y que el gen *cox2b* fue transferido al núcleo. Esto no es entonces un ejemplo de un paso intermedio de la transferencia de un gen al núcleo, porque la transferencia de *cox2a* y *cox2b* sucedió de manera independiente. Al hacer una reconstrucción de la historia de este grupo, podría plantearse que el gen *cox2* se fragmentó en la mitocondria, a continuación se transfirió el gen *cox2b* y finalmente el gen *cox2a*. La transferencia del gen *cox2b* debe haber sido antes de que la línea de organismos como *C. reinhardtii* y *Polytomella* sp. se separara de *S. obliquus*. En cuanto se separaron ambos linajes, *S. obliquus* cambió el código genético mitocondrial imposibilitando la posterior transferencia de *cox2a*. Por supuesto, sería

necesario realizar estudios para confirmar que el gen *cox2a* mitocondrial de *S. obliquus* es funcional y si no se encuentra alguna copia de este gen en el genoma nuclear.

viii) *Disminución en la hidrofobicidad de las proteínas codificadas*

Existen dos ejemplos de proteínas que se encuentran codificadas en todos los genomas mitocondriales caracterizados: el citocromo *b* que tiene una estructura de ocho cruces transmembranales (codificado por el gen *cob*) y la subunidad I del complejo citocromo *c* oxidasa que tiene doce cruces transmembranales (Tsukihara y col. 1996) (codificada por el gen *cox1*). En *S. cerevisiae* se han realizado estudios *in vivo* intentando dirigir hacia la mitocondria construcciones citoplásmicas de longitudes variables del citocromo *b*; mostrando que la importabilidad de los polipéptidos no se encuentra forzosamente relacionada al número de cruces transmembranales que posea (Claros y col., 1995). Estos estudios sugirieron que el promedio máximo de hidrofobicidad de una cadena de entre 60 y 80 residuos de aminoácidos (*mesoH*) y la hidrofobicidad máxima de las regiones transmembranales, son indicadores de la facilidad con la que una proteína puede ser importada a la mitocondria. Utilizando estos parámetros para analizar los genes *cox2*, *cox3* y *atp6* mitocondriales de las algas clorofitas reportadas, codifican proteínas con características que impedirían su importación hasta la mitocondria si se encontraran codificadas en el genoma nuclear. Las proteínas ATP6, NAD4L, COX III, COX IIA y COX IIB de *C. reinhardtii* en cambio, tienen una hidrofobicidad reducida que podría permitir su importación a la mitocondria. Los análisis de hidropatía realizados en cada uno de los cruces transmembranales de ATP6 mostraron que la reducción de la hidrofobicidad sucede preferentemente en aquellos que no están involucrados en la función de la subunidad (este fenómeno también se observa en la proteína COX III (Pérez-Martínez y col., 2000; Funes y col. *en prensa* (Apéndice VI)).

Conclusiones

1. Los genes que codifican para las subunidades mitocondriales ATP6 y NAD4L en *C. reinhardtii* se encuentran en el genoma nuclear.

Los genes *atp6* y *nd4L* se encuentran codificados en el núcleo de *C. reinhardtii* y presentan características típicas de este tipo de genes: un uso de codones nuclear, una señal de poliadenilación típica de genes nucleares y predicen una secuencia que podría funcionar como presecuencia mitocondrial. Además el gen *atp6* presenta otra característica común en los genes nucleares de esta alga: la presencia de intrones.

2. La subunidad ATP6 está presente en el complejo enzimático.

La subunidad ATP6 de *C. reinhardtii* se identificó en geles de acrilamida con una masa molecular aparente de 15.5 kDa. Esta subunidad se caracterizó al obtener la secuencia amino terminal.

3. La subunidad ATP6 de *C. reinhardtii* es una proteína con propiedades físicas que facilitarían su integración en la membrana interna mitocondrial y en el complejo enzimático activo.

El gen *atp6* predice una proteína con una hidrofobicidad general disminuida respecto a sus equivalentes sintetizadas en la mitocondria. Esta disminución se ve restringida a las regiones de la subunidad que no comprometen el funcionamiento del complejo enzimático. Además, la secuencia deducida del gen predice una presecuencia que podría ser la responsable de dirigir a la preproteína hasta la mitocondria.

4. En cuanto a la evolución de la transferencia al núcleo de los genes *cox2* en las algas clorofitas, *S. obliquus* representa un estado intermedio de evolución.

En el alga clorofita *S. obliquus* el gen *cox2* ha sido fragmentado al igual que en otros miembros de la misma clase como *C. reinhardtii*, sin embargo el gen *cox2a* permanece en el genoma mitocondrial y el gen *cox2b* ha sido enviado al genoma nuclear. En este trabajo se obtuvo la secuencia parcial del gen *cox2b* encontrando que es un gen característicamente nuclear que utiliza el código genético universal y cuya secuencia codificante está interrumpida por algunos intrones.

5. La subunidad II del complejo citocromo c oxidasa de los organismos apicomplejos está codificada por dos genes nucleares adquiridos por transferencia horizontal de un alga clorofita.

Los organismos apicomplejos como los conocemos actualmente, se originaron después de un evento de endosimbiosis secundaria con un alga. En este trabajo se caracterizaron los genes nucleares *cox2a* y *cox2b* de *T. gondii*. Estos genes también se identificaron en otros apicomplejos como *P. falciparum*, *P. yoelii* y *Theileria parva*. La presencia de los genes nucleares *cox2a* y *cox2b*

en los organismos apicomplejos y su alta similitud estructural con los mismos genes de algas como *C. reinhardtii*, *Polytomella* sp. y *S. obliquus* ha permitido esclarecer que el evento de endosimbiosis fue con un alga clorofícea.

Estudios Pendientes

A lo largo de este trabajo se han identificado algunos genes que codifican proteínas mitocondriales hidrofóbicas en dos grupos de organismos: i) en *C. reinhardtii* (y potencialmente en otras clorofíceas) los genes *atp6* y *nd4L* se encuentran en el genoma nuclear y ii) en el grupo de los apicomplejos los genes *cox2a* y *cox2b* son nucleares y se adquirieron por un fenómeno de transferencia horizontal entre el núcleo de un alga clorofícea con la que realizaron un proceso de endosimbiosis secundaria.

El hecho de contar con la secuencia de los genes es sólo la puerta que abre el camino para numerosos estudios:

1. Estudiar el proceso de regulación de los genes mitocondriales.

De alguna manera los genes que codifican productos que funcionan en la mitocondria (proteínas o RNAs) deben estar regulados para responder a las necesidades mitocondriales y por lo tanto celulares. Una opción sería que los genes que se transfirieron de manera masiva después de la endosimbiosis, lo hayan hecho a manera de bloques que adquirieron secuencias regulatorias particulares para cada bloque, facilitando el proceso de regulación de todos los genes. Sin embargo, aunque esto fuera cierto, los genes que se transfirieron posteriormente, como *atp6*, *cox2a*, *cox2b* o *nd4L*, claramente lo hicieron como genes individuales y no como un gran fragmento de genoma mitocondrial. Sería muy interesante poder detallar los mecanismos de regulación genética mitocondrial, en particular de aquellos genes que codifican proteínas de la cadena respiratoria, ya que además éstos deben estar regulados de manera coordinada con los genes que permanecen en el genoma mitocondrial.

2. Realizar estudios de los mecanismos de importación a la mitocondria utilizados por las proteínas codificadas por genes nucleares.

Aunque sea posible suponer qué tipo de mecanismos seguirán estas proteínas para poder insertarse correctamente en la membrana interna mitocondrial y acoplarse en el complejo funcional, sería importante disectar la presecuencia, para ver qué regiones son importantes en estas presecuencias tan largas, si existe un mecanismo de inserción especial como en el caso de las leguminosas, o que región de la presecuencia guía a la proteína hasta qué paso específico de la importación.

En el caso de los apicomplejos, sería muy importante determinar qué mecanismos son los que le permiten a las subunidades COX IIA y COX IIB que interactúan entre sí y con las otras subunidades del complejo, ya que las subunidades COX II tienen un origen evolutivo diferente a COX I y COX III.

Perspectivas

Expresión alotópica de genes mitocondriales y su aplicación en terapias génicas mitocondriales humanas.

El genoma mitocondrial humano es una molécula pequeña, circular, de doble cadena que contiene 37 genes. 24 de estos genes son parte de la maquinaria de traducción mitocondrial (2 rRNAs y 22 tRNAs) y 13 codifican subunidades de la cadena respiratoria (siete subunidades del complejo I, una subunidad del complejo III, tres subunidades del complejo IV y dos subunidades del complejo V) (Anderson y col., 1981).

Las primeras mutaciones patógenas del genoma mitocondrial humano, fueron descritas en los trabajos de Holt y col. (1988) y de Wallace y col. (1988). Desde entonces, se han descrito más de cien mutaciones patogénicas en el genoma humano (DiMauro y Andreu, 2000). Desde el punto de vista genético, estas mutaciones pueden caer en tres categorías: i) mutaciones en genes que codifican proteínas; ii) mutaciones en genes de RNA estructurales como tRNAs y rRNAs y que por lo tanto afectan todo el proceso de síntesis de proteínas en el organelo; y iii) rearrreglos del genoma a gran escala como deleciones o duplicaciones. Estas modificaciones pueden producir una gran cantidad de condiciones clínicas y patológicas.

Las mutaciones patogénicas en los genes que codifican proteínas se han caracterizado fundamentalmente en dos enfermedades: Neuropatía Óptica Hereditaria de Leber (LHON por sus siglas en inglés) y NARP/MILS (NARP por las siglas en inglés de debilidad muscular neurogénica, ataxia y retinitis pigmentosa y MILS por las siglas en inglés para el síndrome de Leigh heredado por vía materna). La enfermedad de LHON ha sido asociado con mutaciones en el complejo I; y el síndrome NARP/MILS se ha asociado fundamentalmente con mutaciones en la subunidad ATP6 (DiMauro y Schon, 2001).

Además de las alteraciones mitocondriales producidas por modificaciones en el genoma mitocondrial, recientemente, se han detectado un gran número de síndromes neurológicos causados por alteraciones en el funcionamiento mitocondrial y que tienen su origen en genes nucleares. Las alteraciones nucleares que son responsables de desórdenes mitocondriales, se han agrupado en cuatro grupos: i) alteraciones en componentes estructurales de la cadena respiratoria; ii) alteraciones en factores involucrados en el ensamblaje de los complejos respiratorios; iii) alteraciones en los factores involucrados en vías metabólicas que influyen en la biogénesis mitocondrial (incluyendo a los complejos respiratorios); y iv) alteraciones en los factores que controlan la estabilidad del mtDNA (Zeviani y Klopstock, 2001).

Hasta el momento no existe ningún tipo de terapia disponible para los desórdenes mitocondriales, ya sean de origen mitocondrial o de origen nuclear. Sin embargo, es posible realizar una serie de procedimientos como mejoras en la dieta, ciertas correcciones quirúrgicas y tratamientos con sustituciones de los productos de las vías alteradas, que pueden disminuir problemas específicos y mejorar la calidad de vida de los pacientes. Una de las razones de este problema médico ha sido la heterogeneidad de cada uno de los síndromes (Zeviani y Klopstock, 2001).

Se han planteado varias estrategias para desarrollar terapias génicas que puedan corregir las enfermedades mitocondriales (Collombet y Coutelle, 1998; deGrey, 2000; DiMauro y col., 2000). Recientemente se han comenzado a implementar algunos modelos experimentales de terapia génica que eventualmente permitirían restaurar las funciones mitocondriales en los pacientes. Estas estrategias incluyen el introducir genes modificados o productos génicos a la mitocondria a través de la maquinaria de importación proteica (Kolesnikova y col., 2000); o bien por la inhibición del mtDNA alterado por complejos de ácidos nucleicos y proteínas antígenómicas dirigidos a secuencias específicas (Taylor y col., 2000).

Otro tipo de terapia génica es el colocar una copia "normal" del gen o genes alterados en el núcleo celular y dirigir los productos que se sintetizan en el citoplasma hacia la mitocondria. Esta activación funcional de un gen en un compartimento celular diferente a su lugar de origen se conoce como expresión alotópica. La expresión alotópica de genes mitocondriales ha sido desarrollada con éxito en *S. cerevisiae* para reparar defectos en la RNA madurada de la matriz mitocondrial b14 (Banroques y col., 1986; Banroques y col., 1987); para estudiar las funciones de la subunidad ribosomal mitocondrial VARI (Sanchirico y col., 1995); y con la subunidad ATP8 (equivalente a la subunidad A6L en mitocondrias de mamíferos) (Gray y col., 1996). La secuencia del gen *atp8* fue modificada para la expresión nuclear y se le añadió artificialmente la secuencia de una MTS. El gen se expresó correctamente y el producto polipeptídico sintetizado en el citoplasma fue importado a la membrana interna mitocondrial y ensamblada en el complejo F_1F_0 ATP sintasa.

De especial interés para este trabajo, ha sido el hecho de el grupo de Manfredi y col. (2002) ha publicado un modelo experimental de terapia génica en células humanas con mutaciones en el gen *atp6* y que por lo tanto no pueden realizar la síntesis de ATP de manera correcta. En este trabajo se rediseñó el gen mitocondrial humano para que utilizara el código genético universal, se le añadieron dos presecuencias mitocondriales diferentes (una de 25 residuos de la subunidad 8 de la citocromo c oxidasa y otra de 61 residuos de una isoforma humana de la subunidad c de la ATPasa) y se insertó la construcción en el genoma nuclear. El grupo observó que en este modelo celular se rescata la deficiencia en la síntesis de ATP, sin

embargo la eficiencia de importación de la subunidad es muy baja con cualquiera de las dos MTSs utilizadas. Una posible modificación que podría realizarse para tratar de mejorar la eficiencia de importación, sería la utilización de presecuencias largas, como la encontrada en la subunidad ATP6 de *C. reinhardtii*.

La expresión alotópica de genes mitocondriales en el núcleo para propósitos de terapia génica debe considerar los cambios apropiados en el uso de codones o en el código genético y la adición de una presecuencia mitocondrial capaz de dirigir el producto peptídico sintetizado hacia la mitocondria y su inserción funcional. Estos dos elementos mostraron no ser suficientes para la expresión alotópica. Tal vez la disminución en la hidrofobicidad promedio de la proteína sintetizada, fundamentalmente en las regiones transmembranales que no son fundamentales para la función de la proteína, facilitaría la importación de éstas proteínas altamente hidrofóbicas. Tomar el caso de *C. reinhardtii* como base para decidir en qué regiones deberían hacerse las modificaciones podría facilitar enormemente el trabajo ya que en este organismo esta disminución se ha hecho de manera natural sin alterar el funcionamiento del organismo. Sin embargo, sería ideal conocer con más detalle la topología y función de las proteínas de interés y un conocimiento más profundo de los mecanismos de acción de las proteínas coexpresadas.

Posibles aplicaciones para terapias contra los apicomplejos

Dentro del grupo de los apicomplejos se ubican protistas endoparásitos, algunos de los cuales son agentes causales de enfermedades en el ser humano o en especies de importancia económica. En los últimos años el estudio de los apicomplejos ha sido enormemente fomentado, prueba de esto es que actualmente está secuenciado completamente el genoma de *P. falciparum*, el de *T. gondii* se encuentra ya en una fase avanzada y en etapas tempranas de secuenciación se están analizando los genomas de otras especies como *P. yoelii*, *Theileria parva* y *Theileria annulata*. La meta final de realizar tantos estudios en estos organismos es detener las enfermedades que producen: por ejemplo la toxoplasmosis es actualmente una de las principales causas de muerte entre los enfermos de SIDA (Carruthers, 2002).

Conocer la secuencia completa del genoma de un organismo, nos abre la puerta a los más íntimos secretos de su funcionamiento, sin embargo, actualmente hay una enorme cantidad de regiones de los genomas para las que todavía no conocemos su función o su importancia. Por lo tanto el aprovechamiento médico de esta serie de conocimientos, todavía está limitado a que podamos desentrañar cuáles son las vías metabólicas que son exclusivas de estos parásitos, o qué elementos podrían alterarse sin que esto repercuta en el metabolismo del ser humano.

El hecho de conocer elementos específicos de estos organismos en proteínas para las cuáles ya se conoce la función que desempeñan y el papel fundamental que juegan dentro del metabolismo celular, podría permitir el desarrollo de estrategias específicas de ataque contra estos parásitos.

En el caso específico de los resultados obtenidos en este trabajo, el hecho de saber que la subunidad COX II en realidad se compone de una subunidad COX IIA y una COX IIB podría permitir desarrollar algún tipo de estrategia dirigida exclusivamente hacia cualquiera de estas subunidades que no altere el funcionamiento de una subunidad COX II ortodoxa como la que se encuentra en las mitocondrias humanas. Una opción sería desarrollar terapias con RNAs mensajeros antisentido de *cox2a* o de *cox2b* que en principio no alterarían la expresión del gen *cox2* mitocondrial humano; otra opción sería desarrollar péptidos específicos que se asociaran con las extensiones de COX IIA y COX IIB y por lo tanto eviten que estas dos proteínas interactúen entre sí impidiendo el ensamblaje del complejo funcional; una tercera opción sería modificar de alguna manera la entrada de COX IIA o COX IIB a la mitocondria. COX IIB sería un blanco mucho más sencillo y eficaz, ya que esta subunidad es la directamente involucrada en el paso de electrones entre el citocromo *c* y los centros de reacción de la subunidad COX I; el bloquear solamente el funcionamiento de COX IIA podría favorecer el que la subunidad COX IIB modifique su campo de acción e interactúe con otra subunidad. El hecho de evitar que COX IIA o COX IIB se ensamblen de manera funcional en el complejo citocromo *c* oxidasa de los apicomplejos, repercutiría directamente en el proceso de la cadena respiratoria mitocondrial orillando a estos organismos patógenos a la muerte.

Por supuesto, para poder pensar de una manera más real en alguna de estas posibilidades es necesario realizar antes toda una serie de estudios de expresión de los genes, síntesis e importación de las subunidades para determinar qué vía podría ser la más específica y al mismo tiempo la más eficaz.

Referencias Bibliográficas

- Abrahams, J.P., Leslie, A.G., Lutter, R. and Walker, J.E. (1994)** Structure at 2.8 Å resolution of F_1 -ATPase from bovine heart mitochondria. *Nature*, **370**, 621-628.
- Adams, K.L., Song, K., Roessler, P.G., Nugent, J.M., Doyle, J.L., Doyle, J.J. and Palmer, J.D. (1999)** Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. *Proc Natl Acad Sci U S A*, **96**, 13863-13868.
- Adams, K.L., Daley, D.O., Qiu, Y.L., Whelan, J. and Palmer, J.D. (2000)** Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature*, **408**, 354-357.
- Adams, K.L., Ong, H.C. and Palmer, J.D. (2001)** Mitochondrial gene transfer in pieces: fission of the ribosomal protein gene *rpl2* and partial or complete gene transfer to the nucleus. *Mol Biol Evol*, **18**, 2289-2297.
- Allen, J.F. and Raven, J.A. (1996)** Free-radical-induced mutation vs redox regulation: costs and benefits of genes in organelles. *J Mol Evol*, **42**, 482-492.
- Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R. and Young, I.G. (1981)** Sequence and organization of the human mitochondrial genome. *Nature*, **290**, 457-465.
- Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H. and Kurland, C.G. (1998)** The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature*, **396**, 133-140.
- Andersson, J.O. and Roger, A.J. (2002)** A cyanobacterial gene in nonphotosynthetic protists--an early chloroplast acquisition in eukaryotes? *Curr Biol*, **12**, 115-119.
- Arnold, I., Pfeiffer, K., Neupert, W., Stuart, R.A. and Schagger, H. (1998)** Yeast mitochondrial F_1F_0 -ATP synthase exists as a dimer: identification of three dimer-specific subunits. *Embo J*, **17**, 7170-7178.
- Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H. and Tabata, S. (1999)** A large scale structural analysis of cDNAs in a unicellular green alga, *Chlamydomonas reinhardtii*. I. Generation of 3433 non-redundant expressed sequence tags. *DNA Res*, **6**, 369-373.
- Atteia, A., Dreyfus, G. y González-Halphen, D. (1997)** Characterization of the alpha and beta-subunits of the F_0F_1 -ATPase from the alga *Polytomella* spp., a colorless relative of *Chlamydomonas reinhardtii*. *Biochim Biophys Acta*, **1320**, 275-284.
- Baker, A. and Schatz, G. (1987)** Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. *Proc Natl Acad Sci U S A*, **84**, 3117-3121.
- Banroques, J., Delahodde, A. and Jacq, C. (1986)** A mitochondrial RNA maturase gene

- transferred to the yeast nucleus can control mitochondrial mRNA splicing. *Cell*, **46**, 837-844.
- Banroques, J., Perea, J. and Jacq, C. (1987)** Efficient splicing of two yeast mitochondrial introns controlled by a nuclear-encoded maturase. *Embo J*, **6**, 1085-1091.
- Bensasson, D., Petrov, D.A., Zhang, D.X., Hartl, D.L. and Hewitt, G.M. (2001)** Genomic gigantism: DNA loss is slow in mountain grasshoppers. *Mol Biol Evol*, **18**, 246-253.
- Bischoff, H. and Bold, H. C. (1963)** Phycological studies IV. Some soil algae from Enchanted Rock and related algal species. *Univ. Tex. Publ. No. 6318*, Austin, TX, USA.
- Blanchard, J.L. and Schmidt, G.W. (1996)** Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol Biol Evol*, **13**, 893.
- Blanchard, J.L. and Hicks, J.S. (1999)** The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage. *J Eukaryot Microbiol*, **46**, 367-375.
- Boore, J.L. (1999)** Animal mitochondrial genomes. *Nucleic Acids Res*, **27**, 1767-1780.
- Boyer, P.D. (1997)** The ATP synthase: a splendid molecular machine. *Annu Rev Biochem*, **66**, 717-749.
- Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. and Schuster, W. (1993)** The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett*, **325**, 140-145.
- Campbell, C.L. and Thorsness, P.E. (1998)** Escape of mitochondrial DNA to the nucleus in *yme1* yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. *J Cell Sci*, **111**, 2455-2464.
- Carruthers, V.B. (2002)** Host cell invasion by the opportunistic pathogen *Toxoplasma gondii*. *Acta Trop*, **81**, 111-122.
- Choi, J.Y., Lee, T.W., Jeon, K.W. and Ahn, T.I. (1997)** Evidence for symbiont-induced alteration of a host's gene expression: irreversible loss of SAM synthetase from *Amoeba proteus*. *J Eukaryot Microbiol*, **44**, 412-419.
- Claros, M.G. and von Heijne, G. (1994)** TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci*, **10**, 685-686.
- Claros, M.G. (1995)** MitoProt, a Macintosh application for studying mitochondrial proteins. *Comput Appl Biosci*, **11**, 441-447.
- Claros, M.G., Perea, J., Shu, Y., Samatey, F.A., Popot, J.L. and Jacq, C. (1995)** Limitations to in vivo import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome *b*. *Eur J Biochem*, **228**, 762-771.
- Claros, M.G., Perea, J. and Jacq, C. (1996)** Allotopic expression of yeast mitochondrial maturase to study mitochondrial import of hydrophobic proteins. *Methods Enzymol*, **264**, 389-403.
- Collura, R.V. and Stewart, C.B. (1995)** Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature*, **378**, 485-489.

- Covello, P.S. and Gray, M.W. (1992)** Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome *c* oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. *Embo J*, **11**, 3815-3820.
- Daley, D.O., Adams, K.L., Clifton, R., Qualmann, S., Millar, A.H., Palmer, J.D., Pratje, E. and Whelan, J. (2002)** Gene transfer from mitochondrion to nucleus: novel mechanisms for gene activation from *Cox2*. *Plant J*, **30**, 11-21.
- Degli-Esposti, M., De Vries, S., Crimi, M., Ghelli, A., Patarnello, T., y Meyer, A. (1993)** Mitochondrial cytochrome *b*: evolution and structure of the protein. *Biochim Biophys Acta*, **1143**, 243:271.
- Denovan-Wright, E.M., Nedelcu, A.M. and Lee, R.W. (1998)** Complete sequence of the mitochondrial DNA of *Chlamydomonas eugametos*. *Plant Mol Biol*, **36**, 285-295.
- DiMauro, S. and Andreu, A.L. (2000)** Mutations in mtDNA: are we scraping the bottom of the barrel? *Brain Pathol*, **10**, 431-441.
- DiMauro, S. and Schon, E.A. (2001)** Mitochondrial DNA mutations in human disease. *Am J Med Genet*, **106**, 18-26.
- Dunn, S.D., Bi, Y. and Revington, M. (2000)** A re-examination of the structural and functional consequences of mutation of alanine-128 of the *b* subunit of *Escherichia coli* ATP synthase to aspartic acid. *Biochim Biophys Acta*, **1459**, 521-527.
- Dzierszinski, F., Popescu, O., Toursel, C., Slomianny, C., Yahiaoui, B. and Tomavo, S. (1999)** The protozoan parasite *Toxoplasma gondii* expresses two functional plant- like glycolytic enzymes. Implications for evolutionary origin of apicomplexans. *J Biol Chem*, **274**, 24888-24895.
- Edqvist, J., Burger, G. and Gray, M.W. (2000)** Expression of mitochondrial protein-coding genes in *Tetrahymena pyriformis*. *J Mol Biol*, **297**, 381-393.
- Ellis, J.T., Morrison, D.A. y Jeffries, A.C. (1998)** The phylum Apicomplexa : an update on the molecular phylogeny. In *Evolutionary relationships among protozoa* (eds. G.H. Coombs, K. Vickerman, M.A. Sleight y A. Warren) Kluwer Academic Publishers, Dordrecht, The Netherlands. pp 255-274.
- Elston, T., Wang, H. and Oster, G. (1998)** Energy transduction in ATP synthase. *Nature*, **391**, 510-513.
- Emanuelsson, O. and von Heijne, G. (2001)** Prediction of organellar targeting signals. *Biochim Biophys Acta*, **1541**, 114-119.
- Eriksson, M., Gardeström, P. y Samuelsson, G. (1995)** Isolation, Purification, and Characterization of Mitochondria from *Chlamydomonas reinhardtii*. *Plant Physiol* **107**, 479-483,
- Fast, N.M., Kissinger, J.C., Roos, D.S. and Keeling, P.J. (2001)** Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids.

- Mol Biol Evol*, **18**, 418-426.
- Feagin, J.E. (2000)** Mitochondrial genome diversity in parasites. *Int J Parasitol*, **30**, 371-390.
- Figuroa, P., Gomez, I., Carmona, R., Holuigue, L., Araya, A. and Jordana, X. (1999)** The gene for mitochondrial ribosomal protein S14 has been transferred to the nucleus in *Arabidopsis thaliana*. *Mol Gen Genet*, **262**, 139-144.
- Figuroa, P., Gomez, I., Holuigue, L., Araya, A. and Jordana, X. (1999)** Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J*, **18**, 601-609.
- Franzen, L.G. and Falk, G. (1992)** Nucleotide sequence of cDNA clones encoding the beta subunit of mitochondrial ATP synthase from the green alga *Chlamydomonas reinhardtii*: the precursor protein encoded by the cDNA contains both an N-terminal presequence and a C-terminal extension. *Plant Mol Biol*, **19**, 771-780.
- Frey, T.G. and Mannella, C.A. (2000)** The internal structure of mitochondria. *Trends Biochem Sci*, **25**, 319-324.
- Friedl, T. (1997)** The Evolution of Green Algae. *Pl Syst Evol*, **11**, 87-101.
- Frohman, M.A. (1993)** Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol*, **218**, 340-356.
- Galanis, M., Devenish, R.J. and Nagley, P. (1991)** Duplication of leader sequence for protein targeting to mitochondria leads to increased import efficiency. *FEBS Lett*, **282**, 425-430.
- Girvin, M.E., Rastogi, V.K., Abildgaard, F., Markley, J.L. and Fillingame, R.H. (1998)** Solution structure of the transmembrane H⁺-transporting subunit c of the F₁F₀ ATP synthase. *Biochemistry*, **37**, 8817-8824.
- Gray, R.E., Law, R.H., Devenish, R.J. and Nagley, P. (1996)** Allotopic expression of mitochondrial ATP synthase genes in nucleus of *Saccharomyces cerevisiae*. *Methods Enzymol*, **264**, 369-389.
- Gray, M.W. (1999)** Evolution of organellar genomes. *Curr Opin Genet Dev*, **9**, 678-687.
- Gorman, D.S. y Levine, R.P. (1965)** Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA*, **54**, 1665-1669.
- Grohmann, L., Brennicke, A. and Schuster, W. (1992)** The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. *Nucleic Acids Res*, **20**, 5641-5646.
- Haltia, T., Saraste, M. and Wikstrom, M. (1991)** Subunit III of cytochrome c oxidase is not involved in proton translocation: a site-directed mutagenesis study. *Embo J*, **10**, 2015-2021.
- Handa, H., Bonnard, G. and Grienenberger, J.M. (1996)** The rapeseed mitochondrial gene encoding a homologue of the bacterial protein Ccl1 is divided into two independently transcribed reading frames. *Mol Gen Genet*, **252**, 292-302.

- Harris, E.H. (1988)** The *Chlamydomonas* sourcebook. Academic Press. San Diego, CA, USA. pp 76-80.
- Harris, E.H. (1998)** Introduction to *Chlamydomonas*. In The Molecular Biology of Chloroplasts and Mitochondria in *Chlamydomonas* (eds. J.D. Rochaix, M. Goldshmidt-Clermont y S. Merchant) Kluwer Academic Publishers, Dordrecht, The Netherlands. pp 1-11.
- Hartl, F.U., Pfanner, N., Nicholson, D.W. and Neupert, W. (1989)** Mitochondrial protein import. *Biochim Biophys Acta*, **988**, 1-45.
- Holt, I.J., Harding, A.E. and Morgan-Hughes, J.A. (1988)** Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature*, **331**, 717-719.
- Iwata, S., Ostermeier, C., Ludwig, B. and Michel, H. (1995)** Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*. *Nature*, **376**, 660-669.
- Jeon, K.W. (1995)** The large, free-living amoebae: wonderful cells for biological studies. *J Eukaryot Microbiol*, **42**, 1-7.
- Jiang, W. and Fillingame, R.H. (1998)** Interacting helical faces of subunits *a* and *c* in the F₁F₀ ATP synthase of *Escherichia coli* defined by disulfide cross-linking. *Proc Natl Acad Sci U S A*, **95**, 6607-6612.
- Kadenbach, B., Jarausch, J., Hartmann, R. and Merle, P. (1983)** Separation of mammalian cytochrome *c* oxidase into 13 polypeptides by a sodium dodecyl sulfate-gel electrophoretic procedure. *Anal Biochem*, **129**, 517-521.
- Kadowaki, K., Kubo, N., Ozawa, K. and Hirai, A. (1996)** Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing targeting signals. *Embo J*, **15**, 6652-6661.
- Kiebler, M., Pfaller, R., Sollner, T., Griffiths, G., Horstmann, H., Pfanner, N. and Neupert, W. (1990)** Identification of a mitochondrial receptor complex required for recognition and membrane insertion of precursor proteins. *Nature*, **348**, 610-616.
- Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A. and Ohyama, K. (1997)** Interorganellar gene transfer in bryophytes: the functional *nad7* gene is nuclear encoded in *Marchantia polymorpha*. *Mol Gen Genet*, **256**, 589-592.
- Köhler, C.M. (2000)** Protein translocation pathways of the mitochondrion. *FEBS lett*, **476**, 27-31.
- Köhler, S., Delwiche, C.F., Denny, P.W., Tilney, L.G., Webster, P., Wilson, R.J., Palmer, J.D. and Roos, D.S. (1997)** A plastid of probable green algal origin in Apicomplexan parasites. *Science*, **275**, 1485-1489.
- Kolesnikova, O.A., Entelis, N.S., Mireau, H., Fox, T.D., Martin, R.P. and Tarassov, I.A. (2000)** Suppression of mutations in mitochondrial DNA by tRNAs imported from the cytoplasm. *Science*, **289**, 1931-1933.
- Kroymann, J. and Zetsche, K. (1998)** The mitochondrial genome of *Chlorogonium elongatum* inferred from the complete sequence. *J Mol Evol*, **47**, 431-440.

- Kubo, N., Harada, K., Hirai, A. and Kadowaki, K. (1999)** A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc Natl Acad Sci U S A*, **96**, 9207-9211.
- Kück, U., Jekosch, K. and Holzamer, P. (2000)** DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon. *Gene*, **253**, 13-18.
- Lang, B.F., Gray, M.W. and Burger, G. (1999)** Mitochondrial genome evolution and the origin of eukaryotes. *Annu Rev Genet*, **33**, 351-397.
- Le, T.H., Blair, D. y McManus, D.P. (2001)** A leucine zipper protein of mitochondrial origin. *Biochim Biophys Acta*. **1546**(2), 435-443
- Louis, E.J. and Haber, J.E. (1991)** Evolutionarily recent transfer of a group I mitochondrial intron to telomere regions in *Saccharomyces cerevisiae*. *Curr Genet*, **20**, 411-415.
- Lowry, O.H., Rosenbrough, N.J., Farr, A.L., y Randall, R. (1951)** Protein measurement with the Folin Phenol reagent. *J Biol Chem*, **193**, 265-275.
- Ludwig, B., Bender, E., Arnold, S., Huttemann, M., Lee, I. and Kadenbach, B. (2001)** Cytochrome *c* oxidase and the regulation of oxidative phosphorylation. *ChemBiochem*, **2**, 392-403.
- Lynch, M. (1997)** Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol Biol Evol*, **14**, 914-925.
- Malatesta, F., Antonini, G., Sarti, P. y Brunori, M. (1995)** Structure and function of a molecular machine: cytochrome *c* oxidase. *Biophys. Chem.* **54**:1-33
- Manfredi, G., Fu, J., Ojaimi, J., Sadlock, J.E., Kwong, J.Q., Guy, J. and Schon, E.A. (2002)** Rescue of a deficiency in ATP synthesis by transfer of MTATP6, a mitochondrial DNA-encoded gene, to the nucleus. *Nat Genet*, **30**, 394-399.
- Margulis, L. (1970)** The origin of Eucaryotic cells. Yale University Press. New Haven, USA.
- Markwell, M.A.K., Haas, S.M., Bieber, L.L. y Tolbert, N.E. (1978)** A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal Biochem*, **87**, 206-210.
- Martin, J., Mahlke, K. y Pfanner, N. (1991)** Role of an energized inner membrane in mitochondrial protein import. Delta psi drives the movement of presequences. *J Biol Chem* **266**, 18051-18057
- Martin, W. and Muller, M. (1998)** The hydrogen hypothesis for the first eukaryote. *Nature*, **392**, 37-41.
- Martin, W. and Herrmann, R.G. (1998)** Gene transfer from organelles to the nucleus: how much, what happens, and Why? *Plant Physiol*, **118**, 9-17.

- Menz, R.I., Walker, J.E. and Leslie, A.G. (2001)** Structure of bovine mitochondrial F₁-ATPase with nucleotide bound to all three catalytic sites: implications for the mechanism of rotary catalysis. *Cell*, **106**, 331-341.
- Michaelis, G., Vahrenholz, C. and Pratje, E. (1990)** Mitochondrial DNA of *Chlamydomonas reinhardtii*: the gene for apocytochrome *b* and the complete functional map of the 15.8 kb DNA. *Mol Gen Genet*, **223**, 211-216.
- Mihara, K. and Omura, T. (1996)** Cytosolic factors in mitochondrial protein import. *Experientia*, **52**, 1063-1068.
- Millar, D.G. y Shore, G.C. (1996)** Signal anchor sequence insertion into the outer mitochondrial membrane. Comparison with porin and the matrix protein targeting pathway. *J Biol Chem*, **271**, 25823-25829.
- Mitchell, P. (1961)** Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type mechanism. *Nature* (London), **191**, 144-148.
- Moreira, D., Le Guyader, H. and Philippe, H. (2000)** The origin of red algae and the evolution of chloroplasts. *Nature*, **405**, 69-72.
- Moreira, D. and Philippe, H. (2001)** Sure facts and open questions about the origin and evolution of photosynthetic plastids. *Res Microbiol*, **152**, 771-780.
- Nass, S. y Nass, M.M.K. (1963)** Ultramitochondrial fibers with DNA characteristics. *J. Cell. Biol.* 19: 593-629.
- Nedelcu, A.M., Lee, R.W., Lemieux, C., Gray, M.W. and Burger, G. (2000)** The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res*, **10**, 819-831.
- Nugent, J.M. and Palmer, J.D. (1991)** RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell*, **66**, 473-481.
- Oudot-Le Secq, M.P., Fontaine, J.M., Rousvoal, S., Kloareg, B. and Loiseaux-De Goer, S. (2001)** The complete sequence of a brown algal mitochondrial genome, the ectocarpale *Pylaiella littoralis* (L.) Kjellm. *J Mol Evol*, **53**, 80-88.
- Pak, J.W. and Jeon, K.W. (1997)** A symbiont-produced protein and bacterial symbiosis in *Amoeba proteus*. *J Eukaryot Microbiol*, **44**, 614-619.
- Palmer, J.D. and Herbon, L.A. (1988)** Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J Mol Evol*, **28**, 87-97.
- Paumard, P., Vaillier, J., Couлары, B., Schaeffer, J., Soubannier, V., Mueller, D.M., Brethes, D., di Rago, J.P. and Velours, J. (2002)** The ATP synthase is involved in generating mitochondrial cristae morphology. *Embo J*, **21**, 221-230.
- Perez-Martinez, X., Vazquez-Acevedo, M., Tolkunova, E., Funes, S., Claros, M.G., Davidson, E., King, M.P. and Gonzalez-Halphen, D. (2000)** Unusual location of a mitochondrial gene.

- Subunit III of cytochrome *c* oxidase is encoded in the nucleus of Chlamydomonad algae. *J Biol Chem*, **275**, 30144-30152.
- Perez-Martinez, X., Antaramian, A., Vazquez-Acevedo, M., Funes, S., Tolkunova, E., d'Alayer, J., Claros, M.G., Davidson, E., King, M.P. and Gonzalez-Halphen, D. (2001)** Subunit II of cytochrome *c* oxidase in Chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem*, **276**, 11302-11309.
- Pichersky, E. and Tanksley, S.D. (1988)** Chloroplast DNA sequences integrated into an intron of a tomato nuclear gene. *Mol Gen Genet*, **215**, 65-68.
- Pichersky, E., Logsdon, J.M., Jr., McGrath, J.M. and Stasys, R.A. (1991)** Fragments of plastid DNA in the nuclear genome of tomato: prevalence, chromosomal location, and possible mechanism of integration. *Mol Gen Genet*, **225**, 453-8.
- Quinn, J.M., Merchant, S. (1999)** Adaptation of *Scenedesmus obliquus* (Chlorophyceae) to Copper-Deficiency: Transcriptional Regulation of *Pcy1* but not *Cpx1*. *J. Phycol.* **35**:1253-1263.
- Riistama, S., Puustinen, A., Garcia-Horsman, A., Iwata, S., Michel, H. and Wikstrom, M. (1996)** Channelling of dioxygen into the respiratory enzyme. *Biochim Biophys Acta*, **1275**, 1-4.
- Roise, D. and Schatz, G. (1988)** Mitochondrial presequences. *J Biol Chem*, **263**, 4509-11.
- Rost, B. (1996)** PHD: predicting one-dimensional protein structure by profile-based neural networks. *Methods Enzymol*, **266**, 525-539.
- Sanchez, H., Fester, T., Kloska, S., Schroder, W. and Schuster, W. (1996)** Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in *Arabidopsis* mitochondria. *Embo J*, **15**, 2138-2149.
- Sambrook, J., Fritsch, E.F. y Maniatis, T. (1989)** Molecular Cloning: A Laboratory Manual. 2nd Edition, Cold Spring Harbor, New York, USA.
- Sanchirico, M., Tzellas, A., Fox, T.D., Conrad-Webb, H., Periman, P.S. and Mason, T.L. (1995)** Relocation of the unusual VAR1 gene from the mitochondrion to the nucleus. *Biochem Cell Biol*, **73**, 987-995.
- Saraste, M. (1999)** Oxidative phosphorylation at the *fin de siècle*. *Science*, **283**, 1488-1493.
- Sazanov, L.A., Peak-Chew, S.Y., Fearnley, I.M. and Walker, J.E. (2000)** Resolution of the membrane domain of bovine complex I into subcomplexes: implications for the structural organization of the enzyme. *Biochemistry*, **39**, 7229-7235.
- Schägger, H. (1994)** Native Gel Electrophoresis. In A Practical Guide to Membrane Protein Purification (eds. G. von Jagow y H. Schägger) Academic Press Inc. San Diego, CA, USA, pp 81-104.
- Schägger, H. y von Jagow, G. (1987)** Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal Biochem*, **166**, 368-379.
- Schägger, H. and Pfeiffer, K. (2000)** Supercomplexes in the respiratory chains of yeast and

- mammalian mitochondria. *Embo J*, **19**, 1777-1783.
- Scheffler, I.E. (1999)** *Mitochondria*. Wiley-Liss, U.S.A.
- Schultz, B.E. and Chan, S.I. (2001)** Structures and proton-pumping strategies of mitochondrial respiratory enzymes. *Annu Rev Biophys Biomol Struct*, **30**, 23-65.
- Segui-Real, B., Kispal, G., Lill, R. and Neupert, W. (1993)** Functional independence of the protein translocation machineries in mitochondrial outer and inner membranes: passage of preproteins through the intermembrane space. *Embo J*, **12**, 2211-2218.
- Senior, A.E. and Wise, J.G. (1983)** The proton-ATPase of bacteria and mitochondria. *J Membr Biol*, **73**, 105-124.
- Shafer, K.S., Hanekamp, T., White, K.H. and Thorsness, P.E. (1999)** Mechanisms of mitochondrial DNA escape to the nucleus in the yeast *Saccharomyces cerevisiae*. *Curr Genet*, **36**, 183-194.
- Silflow, C. (1998)** Organization of the Nuclear Genome. In *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas* (eds. J.D. Rochaix, M. Goldshmidt-Clermont y S. Merchant) Kluwer Academic Publishers, Dordrecht, The Netherlands. pp 25-40.
- Sjöling, S. and Glaser, E. (1998)** Mitochondrial targeting peptides in plants. *Trends Plant Sci*, **3**, 136-140.
- Stock, D., Leslie, A.G. and Walker, J.E. (1999)** Molecular architecture of the rotary motor in ATP synthase. *Science*, **286**, 1700-1705.
- Stock, D., Gibbons, C., Arechaga, I., Leslie, A.G. and Walker, J.E. (2000)** The rotary mechanism of ATP synthase. *Curr Opin Struct Biol*, **10**, 672-679.
- Sun, C.W. and Callis, J. (1993)** Recent stable insertion of mitochondrial DNA into an *Arabidopsis* polyubiquitin gene by nonhomologous recombination. *Plant Cell*, **5**, 97-107.
- Taylor, R.W., Wardell, T.M., Lightowlers, R.N. and Turnbull, D.M. (2000)** Molecular basis for treatment of mitochondrial myopathies. *Neurol Sci*, **21**, S909-S912.
- Thorsness, P.E. and Fox, T.D. (1990)** Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature*, **346**, 376-379.
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. and Yoshikawa, S. (1996)** The whole structure of the 13-subunit oxidized cytochrome c oxidase at 2.8 Å. *Science*, **272**, 1136-1144.
- Turmel, M., Lemieux, C., Burger, G., Lang, B.F., Otis, C., Plante, I. and Gray, M.W. (1999)** The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. *Plant Cell*, **11**, 1717-1730.
- Turmel, M., Otis, C. and Lemieux, C. (2002)** The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Mol Biol Evol*, **19**, 24-38.
- Ungermann, C., Neupert, W. and Cyr, D.M. (1994)** The role of Hsp70 in conferring

unidirectionality on protein translocation into mitochondria. *Science*, **266**, 1250-1253.

- Valiyaveetil, F.I. and Fillingame, R.H. (1998)** Transmembrane topography of subunit *a* in the *Escherichia coli* F₁F₀ ATP synthase. *J Biol Chem*, **273**, 16241-16247.
- Vassarotti, A., Stroud, R. and Douglas, M. (1987)** Independent mutations at the amino terminus of a protein act as surrogate signals for mitochondrial import. *Embo J*, **6**, 705-711.
- Vik, S.B., Long, J.C., Wada, T. and Zhang, D. (2000)** A model for the structure of subunit *a* of the *Escherichia coli* ATP synthase and its role in proton translocation. *Biochim Biophys Acta*, **1458**, 457-466.
- Vinogradov, A.D. (2001)** Respiratory complex I: structure, redox components, and possible mechanisms of energy transduction. *Biochemistry (Mosc)*, **66**, 1086-1097.
- Vivier, E. y Desportes, I. (1990)** Phylum Apicomplexa. In Handbook of Protoctista (eds. L. Margulis, J.O. Corliss, M. Melkonian y D.J. Chapman) Jones and Bartlett. USA. pp 549-573.
- von Heijne, G. (1986)** Why mitochondria need a genome. *FEBS lett.* **198**, 1-4.
- Voos, W., Martin, H., Krimmer, T. and Pfanner, N. (1999)** Mechanisms of protein translocation into mitochondria. *Biochim Biophys Acta*, **1422**, 235-254.
- Wada, T., Long, J.C., Zhang, D. and Vik, S.B. (1999)** A novel labeling approach supports the five-transmembrane model of subunit *a* of the *Escherichia coli* ATP synthase. *J Biol Chem*, **274**, 17353-17357.
- Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J., 2nd and Nikoskelainen, E.K. (1988)** Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. *Science*, **242**, 1427-1430.
- Wallace, D.C., Stugard, C., Murdock, D., Schurr, T. and Brown, M.D. (1997)** Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. *Proc Natl Acad Sci U S A*, **94**, 14900-14905.
- Wang, H. and Oster, G. (1998)** Energy transduction in the F₁ motor of ATP synthase. *Nature*, **396**, 279-282.
- Watanabe, K.I. and Ohama, T. (2001)** Regular spliceosomal introns are invasive in *Chlamydomonas reinhardtii*: 15 introns in the recently relocated mitochondrial *cox2* and *cox3* genes. *J Mol Evol*, **53**, 333-339.
- Wikstrom, M. (1998)** Cytochrome oxidase: structure and mechanism. Foreword. *J Bioenerg Biomembr*, **30**, 3-5.
- Wilson, R.J., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W. and Williamson, D.H. (1996)** Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *J Mol Biol*, **261**, 155-72.
- Wischmann, C. and Schuster, W. (1995)** Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. *FEBS Lett*, **374**, 152-156.

- Wolfe, K.H., Li, W.H. and Sharp, P.M. (1987)** Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*, **84**, 9054-9058.
- Wolff, G., Plante, I., Lang, B.F., Kuck, U. and Burger, G. (1994)** Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. Gene content and genome organization. *J Mol Biol*, **237**, 75-86.
- Wright, R.M. and Cummings, D.J. (1983)** Integration of mitochondrial gene sequences within the nuclear genome during senescence in a fungus. *Nature*, **302**, 86-88.
- Xia, D., Yu, C.A., Kim, H., Xia, J.Z., Kachurin, A.M., Zhang, L., Yu, L. and Deisenhofer, J. (1997)** Crystal structure of the cytochrome *bc₁* complex from bovine heart mitochondria. *Science*, **277**, 60-66.
- Yasuda, R., Noji, H., Kinosita, K., Jr. and Yoshida, M. (1998)** F₁-ATPase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell*, **93**, 1117-1124.
- Yasuda, R., Noji, H., Yoshida, M., Kinosita, K., Jr. and Itoh, H. (2001)** Resolution of distinct rotational substeps by submillisecond kinetic analysis of F₁-ATPase. *Nature*, **410**, 898-904.
- Zeviani, M. and Klopstock, T. (2001)** Mitochondrial disorders. *Curr Opin Neurol*, **14**, 553-560.

Apéndice I.

Artículo publicado

**The typically mitochondrial DNA-encoded ATP6 subunit of the
F₁F₀ ATPase is encoded by a nuclear gene in
Chlamydomonas reinhardtii.**

The Typically Mitochondrial DNA-encoded ATP6 Subunit of the F₁F₀-ATPase Is Encoded by a Nuclear Gene in *Chlamydomonas reinhardtii**

Received for publication, October 16, 2001, and in revised form, December 11, 2001
Published, JBC Papers in Press, December 14, 2001, DOI 10.1074/jbc.M109993200

Soledad Funes†§, Edgar Davidson¶, M. Gonzalo Claros||, Robert van Lis††**,
Xochitl Pérez-Martínez‡ ‡‡, Miriam Vázquez-Acevedo‡, Michael P. King¶,
and Diego González-Halphen‡§§

From the †Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, 04510 México D.F., Mexico, the ‡Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, and the ||Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, Málaga E-29071, Spain

The *atp6* gene, encoding the ATP6 subunit of F₁F₀-ATP synthase, has thus far been found only as an mtDNA-encoded gene. However, *atp6* is absent from mtDNAs of some species, including that of *Chlamydomonas reinhardtii*. Analysis of *C. reinhardtii* expressed sequence tags revealed three overlapping sequences that encoded a protein with similarity to ATP6 proteins. PCR and 5'- and 3'-RACE were used to obtain the complete cDNA and genomic sequences of *C. reinhardtii atp6*. The *atp6* gene exhibited characteristics of a nucleus-encoded gene: Southern hybridization signals consistent with nuclear localization, the presence of introns, and a codon usage and a polyadenylation signal typical of nuclear genes. The corresponding ATP6 protein was confirmed as a subunit of the mitochondrial F₁F₀-ATP synthase from *C. reinhardtii* by N-terminal sequencing. The predicted ATP6 polypeptide has a 107-amino acid cleavable mitochondrial targeting sequence. The mean hydrophobicity of the protein is decreased in those transmembrane regions that are predicted not to participate directly in proton translocation or in intersubunit contacts with the multimeric ring of *c* subunits. This is the first example of a mitochondrial protein with more than two transmembrane stretches, directly involved in proton translocation, that is nucleus-encoded.

The F₁F₀-ATP synthase/ATPase (EC 3.6.1.3) is present in the three domains of life: archaea, prokarya, and eukarya. This membrane-bound complex catalyzes ATP synthesis using the electrochemical gradient generated by light-driven or redox-driven electron transfer chains (1). Two main structural domains constitute this oligomeric protein, the membrane-bound sector F₀ involved in proton translocation, and the extrinsic domain F₁ that catalyzes the synthesis of ATP. The F₁ domain contains five subunits in a 3 α /3 β /1 γ /1 δ /1 ϵ stoichiometry (2). The structure of the F₁ sector of the bovine enzyme has been determined crystallographically (3). The F₀ portion of the ATP synthase is less well characterized, due to its highly hydrophobic nature. The structure of a subcomplex of the *Saccharomyces cerevisiae* ATP synthase shows that 10 *c* subunits are arranged around a central stalk (4). In addition, one *a* subunit, also known as ATP6, is thought to interact with the multimeric ring of *c* subunits, translocating protons from the intermembrane space to the mitochondrial matrix. Subunit *a* is predicted to contain two hemichannels that are an obligate route for protons during ATP synthesis driven by the chemiosmotic gradient (5, 6).

In eukaryotic organisms, all of the subunits of the F₁ and a subset of the F₀ portion of the ATP synthase are nucleus-encoded and cytoplasmically synthesized. In most organisms, only two or three of the hydrophobic components of the F₀ sector are mtDNA¹-encoded. The genes of the F₀ sector that are usually found in the mitochondrial genome are *atp6*, *atp8*, and *atp9* (encoding ATP6 or subunit *a*, ATP8 or A6L1, and ATP9 or subunit *c*, respectively). These genes encode highly hydrophobic polypeptides with multiple, putative transmembrane regions (five for ATP6, two for ATP8, and two for ATP9).

The 15.8-kb linear mitochondrial genome of the green alga *C. reinhardtii* lacks several genes that are usually present in mitochondrial genomes, including *nad3*, *nad4L*, *cox2*, *cox3*, *atp6*, *atp8*, and *atp9* (7, 8). These seven genes encode essential components of oxidative phosphorylation complexes and are also absent from the mitochondrial genomes of the related green algae *Chlamydomonas eugametos* (9) and *Chlorogonium elongatum* (10). The genes *cox2a*, *cox2b*, and *cox3*, which encode subunits IIa, IIb, and III of cytochrome *c* oxidase, have been transferred to the nucleus in at least two members of the family

* This work was supported by the Fogarty International Center at the National Institutes of Health (NIH) Grant TW01176; NHLBI, NIH, Grant HL59646; Consejo Nacional de Ciencia y Tecnología (CONACYT) Grant 27754N; and Dirección General de Estudios de Personal Académico, Universidad Nacional Autónoma de México (UNAM), Grant IN202598. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequences reported in this paper have been submitted to the DDBJ/GenBank/EMBL Data Bank with accession numbers AF411119 (*atp6* cDNA sequence from *C. reinhardtii*) and AF411921 (*atp6* genomic sequence from *C. reinhardtii*).

§ Recipient of a fellowship from CONACYT and Dirección General de Posgrado (UNAM) and a graduate student grant from Programa de Apoyo a los Estudios de Posgrado (UNAM).

** Recipient of a fellowship from Dirección General de Posgrado (UNAM).

†† Recipient of a fellowship from CONACYT and Dirección General de Posgrado (UNAM) and a graduate student grant from Programa de Apoyo a los Estudios de Posgrado (UNAM).

§§ To whom correspondence should be addressed: Dept. de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, 04510 México D.F., Mexico. Tel.: 52-55-5622-5620; Fax: 52-55-5622-5611; E-mail: dhalphen@ifsiol.unam.mx.

¹ The abbreviations used are: mtDNA, mitochondrial DNA; Cr-ATP6, *C. reinhardtii* subunit ATP6 of mitochondrial F₁F₀-ATP synthase; EST, expressed sequence tag; <H>, local hydrophobicity; MTS, mitochondrial targeting sequence; nt, nucleotide(s); Tricine, N-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

Chlamydomonadaceae (11, 12). This gene transfer from the mitochondria to the nucleus was accompanied by several structural changes in these genes. The proteins contain mitochondrial targeting sequences (MTS) and diminished overall mesohydrophobicity that allow the import and assembly of these proteins in the inner mitochondrial membrane.

In this work, the cDNA of the *atp6* gene from *C. reinhardtii* was cloned and sequenced, and the corresponding genomic sequence was obtained. We show that *atp6* is localized in the nuclear genome. The ATP6 polypeptide encoded by this *atp6*, named Cr-ATP6, is homologous to all known ATP6 proteins and is shown to be a constituent of the mitochondrial ATP synthase of *C. reinhardtii*. This is the first description of an *atp6* gene that resides in the nuclear genome and the first example of a nuclear gene that encodes a mitochondrial protein that exhibits structural features characteristic of a proton channel. We demonstrate that the largest decrease in hydrophobicity of Cr-ATP6 occurs in those transmembrane regions that do not participate directly in proton translocation and that are not predicted to interact with adjacent subunits. This characterization of a nuclear version of a gene that is normally present in mitochondrial genomes will facilitate the allelotopic expression of mtDNA-encoded genes and its future application to human mitochondrial gene therapy.

EXPERIMENTAL PROCEDURES

Strain and Culture Conditions—Cell wall-less *C. reinhardtii* strain CW15 was grown in TAP medium (13) with 1% sorbitol (Sigma) under continuous light, with agitation at 100 rpm. The cells were harvested at the late exponential phase of growth.

Nucleic Acid Preparation and Sequencing—Total DNA from *C. reinhardtii* was obtained as previously described (11) or, alternatively, using the DNeasy Plant Mini Kit (Qiagen). Total RNA from *C. reinhardtii* was obtained using the RNeasy Mini Kit (Qiagen). All standard molecular biology techniques were as described (14). Sequencing was performed by the Kimmel Cancer Center DNA Sequencing Facility at Thomas Jefferson University.

Cloning of the cDNA of the Gene *atp6* from *C. reinhardtii*—We identified an EST clone (AV394701) (15) as a fragment of the *atp6* gene from *C. reinhardtii*, because the deduced amino acid sequence at the 3' end of the EST exhibited high similarity with other ATP6 subunits. The 5' end showed sequence similarity to two additional clones (AV388269 and AV395475) (15). Based on these EST sequences, two deoxyoligonucleotides were designed (5'-GAGGGTGTTCGCCCTCTTGG-3' and 5'-CGAAGAACGACAGCGAGAAGG-3') and used to amplify a PCR product of 822 nt, containing a portion of the *atp6* cDNA, using a *C. reinhardtii* cDNA library (16) as template. The sequences at the 5' and 3' ends of the cDNA were obtained following RACE PCR (17). The primers used were as follows: 5' RACE, forward (oligo(dT)/adapter) and reverse (5'-CGCCAGAAGCGGTAGATGCC-3'); for nested PCR, forward (oligo adapter) and reverse (5'-GCAGCGAATGGCACCATCG-3'); 3' RACE, forward (5'-CTACGTGTGGCGAGTTCACAAGC-3') and reverse (oligo(dT)/adapter); for nested PCR, forward (5'-GTGGTCAA-GAAGCGCTGFAAGC-3') and reverse (oligo adapter). The sequences of the dT/adapter and adapter deoxyoligonucleotides were described previously (11). The first cDNA strand for these reactions was obtained using the kit Omniscript (Qiagen) with a specific primer (5'-GAG AAG CCC AGC TTG TAC AGA CC-3') to obtain the 5' end and oligo(dT) to obtain the 3' end of the cDNA.

Cloning of the *atp6* Gene from *C. reinhardtii*—Three pairs of deoxyoligonucleotides were designed using the cDNA sequence to amplify the genomic sequence for the *atp6* gene (5'-AGACGAAGAAT-ATAGATTGG-3' and 5'-CGCCAGAAGCGGTAGATCCC-3'; 5'-CATTC-GCTGCCAGGAGGGC-3' and 5'-GCGAAGAACGACAGCGAGAAGG-3'; 5'-CCTCAACTTGTCTGGGTCTGGTG-3' and 5'-ACGAAGCTTA-CAGTCTTCTC-3'). The PCRs were performed with the *Pfu* Turbo DNA polymerase (Stratagene). For PCR amplification, samples were denatured for 5 min at 94 °C; subjected to 10 cycles of 10 s of denaturation at 94 °C, 1 min of annealing at 55 °C, and 4 min of extension at 68 °C; and subjected to 25 cycles of 10 s of denaturation at 94 °C, 1 min of 55 °C annealing, and 4 min (plus 10 s each cycle) of 68 °C extension. Three different overlapping PCR products of 959, 1973, and 2030 nt were obtained and cloned into pGEM-T Easy Vector (Promega) after the addition of terminal Adenines.

Mitochondrial Protein Analysis—Mitochondria from *C. reinhardtii* were isolated as described (18). The final mitochondrial pellet was resuspended in 35 ml of Percoll dilution buffer with a protease inhibitor mixture (0.5 mM phenylmethylsulfonyl fluoride, 1 mM benzamide, and 5 mM ϵ -aminocaproic acid). Mitochondria were diluted to a final concentration of 5 mg of protein/ml, solubilized in 1% dodecylmaloside, and subjected to blue native electrophoresis using a gradient of 5–12% acrylamide (19). Respiratory chain complexes were separated in the second dimension by Tricine-SDS-PAGE (12% acrylamide) (20). Apparent molecular masses were calculated using prestained molecular weight markers (BenchMark Prestained Protein Ladder; Invitrogen). Protein concentrations were determined according to Markwell *et al.* (21). The isolation of polypeptides for N-terminal sequencing was done as previously described (22). N-terminal sequencing was carried out by Dr. J. d'Alayer on an Applied Biosystems Sequencer at the Laboratoire de Microséquence des Protéines, Institut Pasteur, Paris.

Sequence Analysis in Silico—ATP6 sequences used in this work for mesohydrophobicity analysis were obtained from the SwissProt and TrEMBL data bases at the European Bioinformatics Institute (available on the World Wide Web at srp.ebi.ac.uk). The accession numbers are as follows: Q31720 (*Brassica napus*); P07925, Q36271 (*Zea mays*); P26853 (*Marchantia polymorpha*); P05500 (*Oenothera biennis*); P05499 (*Nicotiana tabacum*); Q04654 (*Vicia faba*); P20599 (*Triticum aestivum*); P92547, P93298 (*Arabidopsis thaliana*); Q21786 (*Oryza sativa*); Q36513 (*Platymonas subcordiformis*); Q35748 (*Raphanus sativus*); Q35781 (*Sorghum bicolor*); Q36376, Q36379 (*Helianthus annuus*); Q37624 (*Prototheca wickerhamii*); Q36730 (*Petunia parviflora*); Q34008, Q34004, Q9XP33, Q9XP114, Q9TGM2 (*Beta vulgaris*); O79682 (*Glycine max*); Q9ZY27 (*Pedinomonas minor*); Q9XL97 (*Solanum tuberosum*); Q9TC92 (*Nephroselmis olivacea*); Q9MGKO (*Scenedesmus obliquus*). Alignments were carried out with the ClustalX program (23) with default parameters. Protein weight matrix was Gonnet, although results obtained with Blossum and Pam were similar. The hydrophilic gap penalty was enabled.

Mitochondrial targeting sequence analysis utilized the program MitProt II (24). The same program was used to calculate the segments with high local hydrophobicity (<H>) in a distance comprising 13–17 amino acids. The mesohydrophobicity was estimated by scanning each sequence for a maximum average hydrophobicity measured in windows from 60 to 80 amino acids and averaging the values. We used several hydrophobicity scales to reduce the possibility of bias.

Protein transmembrane regions and secondary structure were predicted using the program TopPred II (25), the PredictProtein Web server (www.es.emblnet.org/Services/MolBio/PredictProtein/) (26), and the ExpASy Molecular Biology Server (www.expasy.ch).

Data used to analyze the *atp6* gene codon usage were obtained from the Codon Usage Database (www.kazusa.or.jp/codon/) with accession numbers [gbpln]:237 for nuclear genes and [gbpln]:21 for mitochondrial genes of *C. reinhardtii*.

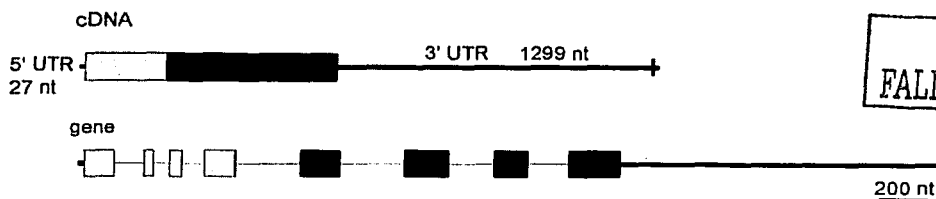
RESULTS

Characterization of the *atp6* cDNA from *C. reinhardtii*—We identified a *C. reinhardtii* EST sequence as a partial fragment of an *atp6* cDNA by the similarity of the predicted translation product with known plant ATP6 proteins. This sequence was used to identify two additional, partially overlapping, EST sequences as potential components of a full-length *atp6* cDNA, although portions of the predicted polypeptide lacked similarity with known ATP6 proteins. Based on these EST sequences, deoxyoligonucleotides were designed and used for PCR amplification. A PCR product of 822 nt was obtained using a *C. reinhardtii* cDNA library as template. The sequence of the amplified product confirmed the co-linearity of the EST sequence fragments. The 5' and 3' ends of the cDNA were obtained by RACE PCR (17) using cDNA made from *C. reinhardtii* total RNA. The full-length cDNA was PCR-amplified as two overlapping fragments and sequenced, and a length of 2349 nt was obtained for the *atp6* cDNA.

The *C. reinhardtii atp6* cDNA contains a 5'-untranslated region of 27 nt followed by an open reading frame of 1014 nt (Fig. 1A). The sequence flanking the proposed open reading frame initiating methionine codon, AACCATCG, is a consensus translation initiation site (AC)A(A/C)(A/C)ATC(G/C) for *C. reinhardtii* (27). The TAA stop codon corresponds to the one

TESIS CON
FALLA DE ORIGEN

A



TESIS CON
FALLA DE ORIGEN

B

Cr		NALQQA	PRVFGLLGRA	PVALGQSGIL	TGSSGPRNQG	PNGSLQSVEN	-61		
So		-----	-----	-----	-----	-----			
Cr	▼	HVYAQAFSTS	SQEEQAAPSI	QGASGMKLPG	MAGSMLLQKS	RSGLRTGSMV	PFAAQQAMNM	-1	
So		-----	-----	-----	-----	-----	-----		
Cr		-----	▼	STAAQ VQSGATVNSI	LLGIY----	-RFWRSQAPM	DKPHAPVDDR	39	
So		MILNYFLFNY	SSRIIASAAE	AWQSVPLVSL	HLPFMGAVFS	LNIWAVTVAI	AMLVLVVNHS		
		: : * : *	: * * :	* :	: * :	: * :	: * :		
Cr		MLPAIVDASD	-----	-----	-----	-----NRAA	LGT-----	56	
So		LNHSSAQAGN	SALSGFRKSV	SGFWNTQTL	RVHLKWRQLS	ARSAQTDRVS	LSTNRLPSAW		
		: : * * :				: * :	: * * :		
Cr		-----	-----	WATALFC	TILASNLLGL	VPTNEAPTSG	LGFATGLGVS	VWATATLGL	103
So		RALSDIQGKT	TVVWSSVLEF	ILLASNSFGL	VPLCSAITGQ	AGFTLGHMSIR	LLSGITYLGV		
		: * * :	: * * :	: * * * * :	: * * * :	: * * * :	: * * * :	: * * * :	
Cr		YKLGFSPPGH	FIPGGTWPWM	AFIFVPLETI	SYTFRAVSLG	VRLVWNMLAG	HTLLHLITGM	163	
So		KROGMRFVRL	FLPSPGSPWPM	APLFILLESI	SYGFRAISLG	TRLYCNMFAG	HLHLHLFTSL		
		: * * :	: * * * * :	: * * * * :	: * * * * * :	: * * * * * :	: * * * * * :		
Cr		ALALPFSLSF	FAMVPATFAV	ACLLSALVGL	BYLVAVLQSG	VFSILSTVYV	GEFNSVKLAG	223	
So		SLVPVLCTPV	LLGAPFTVIA	ASILMALSAL	ETIVAVLQSG	VFCLLSGFYV	TEVLRKDE-		
		: * :	: * * * * :	: * * * * * :	: * * * * * * * :	: * * * * * * * :	: * * * * * * * :		
Cr		PLAKVVKKAL	233						
So		-LASAL----							
		: * * :							

Fig. 1. Organization of the *atp6* gene of *C. reinhardtii* and sequence comparison of *C. reinhardtii* and *S. obliquus* ATP6 subunits. A, diagram of the organization of the *C. reinhardtii atp6* gene. The coding regions of the *atp6* cDNA and *atp6* gene are shown as boxes, the nontranslated regions are shown as thick black lines, and introns are indicated by a thin black line. The putative mitochondrial targeting sequence is gray and the polyadenylation signal is indicated as a vertical black bar. B, the amino acid sequence alignment of the ATP6 proteins of the *F₁F₀*-ATP synthase from *C. reinhardtii* (Cr) and *S. obliquus* (So) are shown. Black triangles indicate the positions of introns in the corresponding gene sequence of *C. reinhardtii*. The *C. reinhardtii* sequence is numbered, using 1 as the first amino acid of the mature protein. The amino acids of the MTS have negative numbers. The MTS of the *C. reinhardtii* protein is in boldface type. The amino-terminal sequence of mature Cr-ATP6 obtained by Edman degradation is boxed. *, identical amino acids; :, similar amino acids. The sequence of the mature Cr-ATP6 is 39% identical and 46% similar to that of the *S. obliquus* ATP6. UTR, untranslated region.

present in the majority of nuclear genes (70% of the reported genes use this stop codon), and the flanking nucleotides G/TAAAG are identical to the consensus sequence (G/CYTAAG/G/A), characteristic of *C. reinhardtii* nuclear genes (27). The deduced sequence predicts a Cr-ATP6 preprotein of 340 amino acids. A BLAST search of the SwissProt nonredundant protein data base using the Cr-ATP6 sequence produced the highest similarity with the mtDNA-encoded ATP6 subunit from the green alga *S. obliquus* (28, 29), also a chlorophyte alga. The two proteins shared 39% identity and 46% similarity (Fig. 1B). The similarity is highest at the carboxyl-terminal region of the protein (59% similarity over amino acids 114–233).

The *atp6* cDNA contains a 3'-untranslated region of 1299 nt

(Fig. 1A). Two different putative polyadenylation signals were found in the Cr-ATP6 cDNA: TGTAAG, the typical signal of nuclear genes of this alga, located 15 nt before the end of the cDNA sequence, and TGTAG, a variation of the most common signal (27), located 1048 nt upstream from the first polyadenylation site. When total RNA was isolated and hybridized with the *atp6* probe in a Northern blot analysis, a single band of 2.4 kb was observed. This result suggested that only the orthodox TGTAAG polyadenylation site is functional in the *atp6* RNA sequence (Fig. 2C).

Genomic Sequence of *atp6* from *C. reinhardtii*—The complete genomic sequence of the *atp6* gene was obtained following PCR amplification of three overlapping genomic regions using prim-

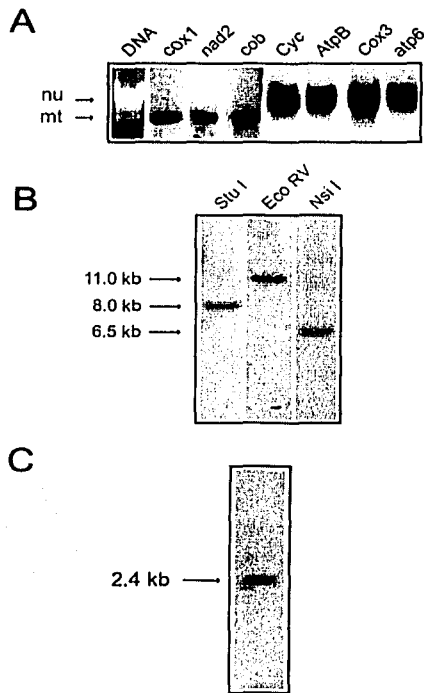


Fig. 2. The *atp6* gene is nucleus-localized, present in a single copy, and expressed in *C. reinhardtii*. *A*, nuclear localization of the *atp6* gene. Thirty micrograms of total DNA from *C. reinhardtii* was electrophoresed through a 0.7% agarose gel. The first lane (DNA) shows the ethidium bromide-stained gel. The subsequent lanes show Southern blot analyses of the same gel hybridized with different mtDNA-encoded (*cox1*, *nad2*, and *cob*) and nucleus-encoded (*Cyc*, *AtpB*, and *Cox3*) gene probes. The *atp6* probe hybridized with the nuclear DNA and not with the mtDNA. *B*, *atp6* is a single copy gene. Total DNA from *C. reinhardtii* was digested with the restriction enzymes *StuI*, *EcoRV*, and *NsiI* and subjected to Southern blot analysis with a labeled probe for the *atp6* gene. Single hybridizing fragments were detected, suggesting the presence of a single *atp6* gene. *C*, Northern blot analysis of total RNA from *C. reinhardtii*. Twenty micrograms of total RNA from *C. reinhardtii* was electrophoresed through 1% agarose, 0.66 M formaldehyde gels and subjected to Northern blot analysis with a labeled probe for the *atp6* gene. A single hybridizing band of 2.4 kb was detected.

crs derived from the cDNA sequence. The 3577-nt gene contained seven introns, ranging from 72 to 263 nt in length. These introns were present in the coding regions for both the mitochondrial targeting sequence and the mature portion of the protein (Fig. 1, A and B).

The *atp6* Gene Is Located in the Nucleus in *C. reinhardtii*—When total DNA isolated from *C. reinhardtii* was electrophoresed on agarose, the mtDNA separated as a discrete band running below the major band representing nuclear and chloroplast DNA (11). Southern blot analysis was carried out to ascertain if the *atp6* gene was present in the nuclear genome. The lower band hybridized with three different mtDNA probes from *C. reinhardtii*, *cox1*, *nad2*, and *cob* (7). In contrast, nuclear DNA hybridized with the *atp6* gene, obtained in this study, and three probes derived from *C. reinhardtii* nuclear genes: *Cyc* encoding cytochrome *c* (30), *AtpB* encoding the β subunit of the ATP synthase (16), and *Cox3* encoding subunit

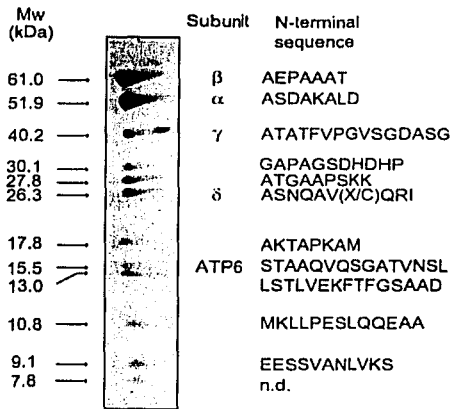


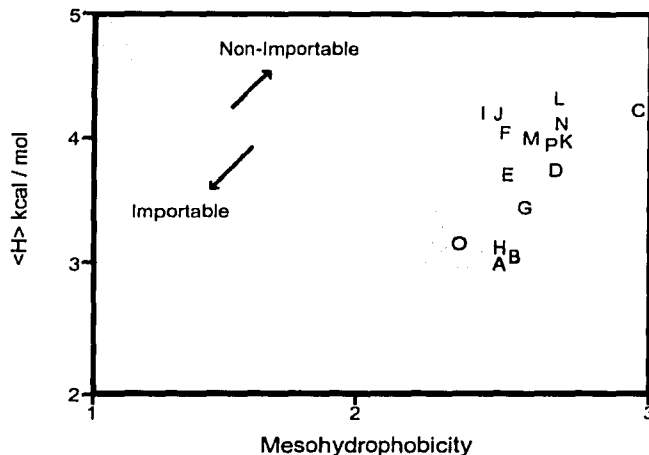
Fig. 3. Subunit composition of the F_1F_0 -ATP synthase complex from *C. reinhardtii*. The mitochondrial respiratory chain complexes were separated by blue native electrophoresis and in the second dimension by Tricine-SDS-PAGE. Shown is the ATP synthase complex after electrophoresis on a Tricine-SDS-polyacrylamide gel. The amino-terminal sequences of the polypeptides were determined, and the identified subunits are indicated. *n.d.*, not determined.

III of cytochrome *c* oxidase (11) (Fig. 2A). A second Southern blot analysis was carried out to determine whether the *atp6* gene was present as a single copy gene in the genome of *C. reinhardtii*. Total DNA was digested with three different restriction enzymes. Southern analysis utilizing an 822-nt PCR product of the coding region of the *atp6* gene as a probe, resolved a single band in each of the restriction digests, suggesting that this gene is present in a single copy in the *C. reinhardtii* genome (Fig. 2B).

The pattern of codon utilization of the *atp6* gene was compared with the pattern of codon usage of other known nuclear, chloroplast, and mitochondrial genes of *C. reinhardtii*. Each codon family of the *atp6* gene favored C or G in the third position and is therefore typically nuclear and different from mitochondrial and chloroplast genes (27) (data not shown).

ATP6 Is a Constituent of F_1F_0 -ATP Synthase of *C. reinhardtii*—If the nuclear localized *atp6* gene of *C. reinhardtii* is expressed, the corresponding Cr-ATP6 subunit should be present in the F_1F_0 -ATP synthase complex. To identify the Cr-ATP6 subunit, isolated mitochondria from *C. reinhardtii* were solubilized and subjected to blue native electrophoresis followed by second dimension analysis by denaturing Tricine-SDS-PAGE. The F_1F_0 -ATP synthase was identified by its characteristic electrophoretic pattern on the second dimension gels. Twelve different polypeptides were present in this complex (Fig. 3). The three major subunits were identified as subunits α , β , and γ , since their apparent molecular masses are similar to those reported previously (31). Their identities were confirmed by N-terminal sequence analysis of their first five amino acids. The additional eight polypeptides were subjected to more extensive N-terminal sequencing (Fig. 3). The 26.3-kDa subunit was identified as subunit δ of F_1F_0 -ATP synthase because of its sequence similarity with other δ subunits. The 15.5-kDa polypeptide was identified as subunit ATP6, the protein product encoded by *atp6*, since its initial 15 amino acids matched exactly amino acids 108–122 of the deduced protein sequence encoded by the *atp6* cDNA. This confirms that the *atp6* gene product is a constituent of the F_1F_0 -ATPase complex of *C. reinhardtii*. The N-terminal sequences of the eight remaining

Fig. 4. Plot of mesohydrophobicity versus maximal local hydrophobicity for ATP6 proteins. The MitoProtII program was used to calculate local hydrophobicity values ($\langle H \rangle$) and mesohydrophobicity values for ATP6 proteins, using the PRIFT scale. Proteins are distributed on the *abscissa* according to their mesohydrophobicity value and on the *ordinate* according to the hydrophobicity of the most hydrophobic 17-residue segment. The hypothetical boundary between importable and nonimportable proteins, indicated by a broad, gray diagonal, was derived from Claros *et al.* (39) and Pérez-Martínez *et al.* (11). The following proteins with their symbols were analyzed: *C. reinhardtii* (A), *S. obliquus* (B), *P. minor* (C), *H. annuus* (D), *P. subcordiformis* (E), *Z. mays* (F), *P. wicherhamii* (G), *O. sativa* (H), *S. tuberosum* (I), *N. tabacum* (J), *B. napus* (K), *T. aestivum* (L), *M. polymorpha* (M), *B. vulgaris* (N), *G. max* (O), *A. thaliana* (P).



polypeptides showed no evident similarity with other ATPase subunits, and their identities remain to be established.

Our data indicate that the mature Cr-ATP6 subunit is a protein of 233 amino acids with an expected molecular mass of 24,577 Da and predicts that Cr-ATP6 has an MTS of 107 amino acids, which is cleaved upon import of this polypeptide into mitochondria. For numbering purposes, the mature protein Cr-ATP6 is considered to start at amino acid number 1, and the amino acids that belong to the putative MTS are numbered accordingly (Fig. 1B).

Hydrophobicity and Importability of the Nucleus-encoded Subunit ATP6 from *C. reinhardtii*—Mitochondrial protein import studies suggest that the highest average hydrophobicity over 60–80 amino acids of a polypeptide chain (mesohydrophobicity), along with the maximum hydrophobicity of the putative transmembrane segments, are useful indicators of the likelihood that a protein could be imported into mitochondria (32). We have previously shown that the transfer of genes from the mtDNA to the nucleus in chlamydomonal algae is accompanied by a decrease of these parameters for the COX IIA, COX IIB, and COX III subunits of cytochrome *c* oxidase (11, 12). The physical characteristics of the predicted Cr-ATP6 polypeptide were examined *in silico* and compared with those of ATP6 subunits encoded by other complete plant and algal mitochondrial *atp6* genes. Fig. 4 shows the plot of mesohydrophobicity versus maximal local hydrophobicity ($\langle H \rangle$) for Cr-ATP6 and for different mtDNA-encoded ATP6 sequences. When compared with several of its mitochondrial counterparts, Cr-ATP6 displays both decreased $\langle H \rangle$ and mesohydrophobicity. However, other mtDNA encoded ATP6 subunits, like those from *S. obliquus*, *O. sativa*, and *G. max*, also exhibit reduced mesohydrophobicity and $\langle H \rangle$.

The reduction of mesohydrophobicity and $\langle H \rangle$ depends on two factors: the length of the hydrophilic loops between transmembrane domains and the mean hydrophobicity of each transmembrane segment. The alignment of all available ATP6 sequences reveals that the distances between transmembrane domains have not changed (results not shown). Hence, the mean hydrophobicity of each transmembrane segment ought to be reduced in Cr-ATP6. To determine in which regions the hydrophobicity was diminished, hydropathy plots of Cr-ATP6 were compared with those of ATP6 sequences from plants and algae. Analysis of Cr-ATP6 predicts that it contains at least six

hydrophobic regions that could be transmembrane regions (data not shown). However, topological considerations argue against the possibility that all hydrophobic segments span the membrane. The most precise positioning of the putative transmembrane helices was obtained by aligning all reported ATP6 sequences and predicting the secondary structure. This analysis suggested a model in which there are five transmembrane segments, named A–E (Fig. 5). This model is in accordance with the topology of the amino and carboxyl termini and the presence of conserved and functionally required amino acids (33–36).

Establishing the boundaries of the five putative transmembrane regions allowed us to quantitate the mean hydrophobicity for each in *C. reinhardtii* and mtDNA-encoded ATP6 subunits from other algae and plants (Fig. 6). While transmembrane helices D and E exhibit similar $\langle H \rangle$ values when compared with the mean $\langle H \rangle$ values in helices D and E of plant and algal mitochondrial sequences, significant differences in $\langle H \rangle$ values were found in the putative transmembrane regions A, B, and C, which are not believed to participate directly in proton translocation and are not thought to interact with the multimeric ring of *c* subunits. This decrease was more than 50% in helices A and C of Cr-ATP6 in comparison with the $\langle H \rangle$ values of helices A and C of other mtDNA-encoded ATP6 subunits. We conclude that the Cr-ATP6 subunit exhibits diminished overall hydrophobicity as compared with the majority of mtDNA-encoded homologs and that the main decrease in hydrophobicity occurred in those transmembrane regions that exhibit poor sequence conservation and that seem not to be critical for proton translocation.

DISCUSSION

The Gene *atp6* of *C. reinhardtii* Exhibits Nucleus-encoded Characteristics—The 15.8-kb linear mitochondrial genome from *C. reinhardtii* lacks several genes that are typically mtDNA-encoded, including three genes that encode subunits of the F_1F_0 -ATP synthase (7). In this work, we show that the *atp6* gene was transferred from the mitochondrial genome to the nucleus in *C. reinhardtii* and demonstrate that the ATP6 protein is present in the mitochondrial F_1F_0 -ATP synthase. The *atp6* gene has been previously found only in mitochondrial genomes. This is the first biochemical and genetic evidence for the nuclear localization of *atp6*. We show that the *atp6* gene is

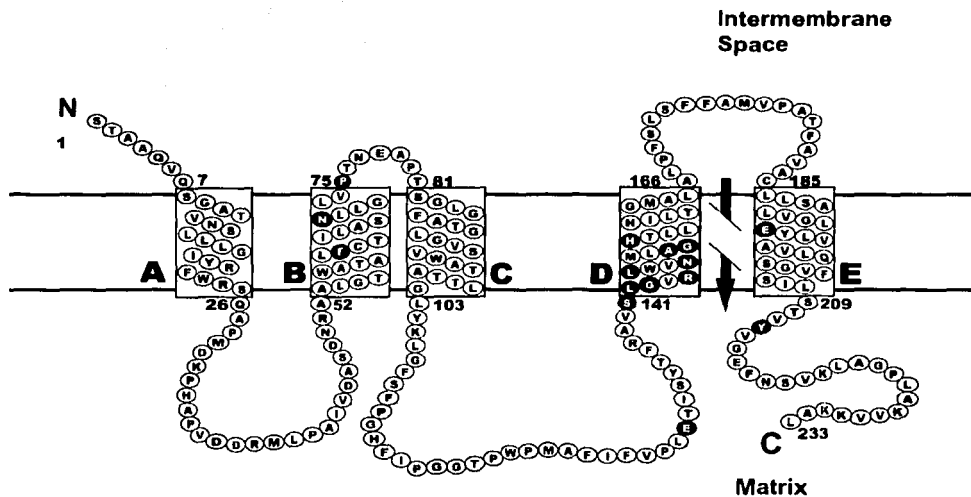


Fig. 5. Model for the transmembrane helices of the ATP6 subunit of the F_1F_0 -ATP synthase from *C. reinhardtii*. Shown is a model for the location of the transmembrane helices of Cr-ATP6. Conserved amino acids are shown as black circles. Transmembrane regions D and E participate in proton translocation. Proton translocation is depicted as a broken arrow to illustrate the participation of two hemichannels in subunit ATP6. For simplicity, the multimeric ring of c subunits is not depicted.

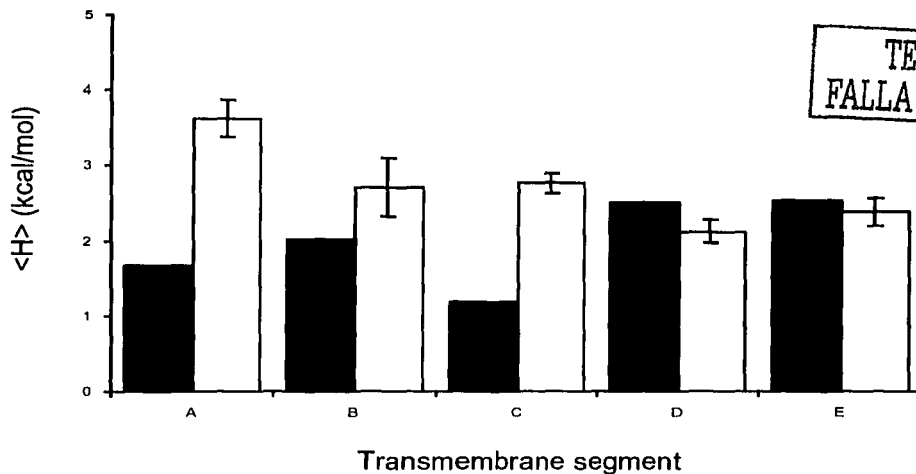


Fig. 6. Analysis of the mean hydrophobicity values of the five transmembrane regions. Hydrophobicity values for the five putative transmembrane helices of the Cr-ATP6 (black bars) were compared with the mean value of hydrophobicity for 21 mtDNA-encoded ATP6 proteins from plants (white bars). In Cr-ATP6, mean hydrophobicity is reduced in transmembrane regions A, B, and C (54, 25, and 57% diminished) as compared with the mean of the mean hydrophobicity values of algal and plant transmembrane regions.

nucleus-localized in *C. reinhardtii* by Southern blot hybridization (Fig. 2A). The *atp6* gene exhibits all of the characteristics of a mitochondrial gene that was functionally transferred to the nucleus (37): a nuclear codon usage, a typical nuclear polyadenylation signal, the presence of introns, acquisition of a DNA sequence encoding an amino-terminal MTS, and diminished hydrophobicity relative to many equivalent mtDNA-encoded proteins.

The deduced amino acid sequence of the ATP6 cDNA and the N-terminal sequence of the mature protein predict the exist-

ence of a 107-amino acid MTS. This MTS is likely to have been acquired after the original mitochondrial gene was transferred to the nuclear genome. This MTS has a low number of acidic residues, and amino acids 2–24 are predicted to form the amphipathic domain necessary for import of the protein into mitochondria. The methionine (Met¹) adjacent to the MTS cleavage site of the full-length Cr-ATP6 may have been the initiating methionine of the ancestral mitochondrial protein before its corresponding gene was transferred to the nucleus. The MTS is unusually long when compared with other *C.*

reinhardtii MTSs, which exhibit a mean length of 30 amino acids (38). However, we previously found that the nuclear genes encoding COX IIA and COX III of *C. reinhardtii* and *Polytomella* sp., proteins that are typically encoded in the mtDNA, are also predicted to have long MTSs, of 98 and 130 amino acids (11, 12). This suggests that the mitochondrial import of the more hydrophobic mitochondrial proteins, such as COX IIA, COX III, and ATP6, may benefit from an extended MTS.

The duplication of the MTS was found to improve the *in vitro* and *in vivo* import of hydrophobic proteins into yeast mitochondria (32, 39). It was suggested that a long MTS can improve the interaction of the precursor with the mitochondrial import machinery. Alternatively, a long MTS could facilitate the folding of the protein to increase its importability and assembly in the membrane.

The vast majority of the plant mitochondrial ATP6 sequences exhibit physical constraints that are expected to hinder their import into mitochondria (Fig. 4), consistent with their presence in the mitochondrial genome. In contrast, Cr-ATP6 has a lower overall hydrophobicity that allows its import into mitochondria, most probably through the TOM-TIM machinery (40). In addition, some mtDNA-encoded ATP6 sequences also exhibit low $\langle H \rangle$ and mesohydrophobicity values (*S. obliquus*, *O. sativa*, and *G. max*). These genes seem to be "prepared," from the hydrophobicity point of view, for transfer to the nucleus. We hypothesize that the ancestral Cr-ATP6 protein had a low overall hydrophobicity that enabled the *atp6* gene to be functionally transferred to the nucleus.

Transfer of *atp6* and Other mtDNA-encoded ATP Synthase Subunits to the Nucleus—Mitochondrial genomes range widely in size and gene content. Only two genes encoding components of the mitochondrial respiratory chain are invariably present in all mitochondrial genomes known to date: *cob* and *cox1*. Their polypeptide products are central components of proton-translocating complexes: cytochrome *b* of the *bc₁* complex and subunit I of cytochrome *c* oxidase. Cr-ATP6 is the first example of a mitochondrial protein with more than two transmembrane helices, directly involved in proton translocation, that is nucleus-encoded.

The highly reduced 6-kb mtDNA of the parasite *Plasmodium falciparum* contains three genes encoding components of the mitochondrial respiratory chain: *cob*, *cox1*, and *cox3* (41). At the other end of the mtDNA size spectrum, the 60-kb mtDNA of the flagellate *Reclinomonas americana* (42) encodes 23 proteins that participate in oxidative phosphorylation. Five of these are constituents of the F_1F_0 -ATP synthase complex: the proteins encoded by the genes *atp1*, *atp3*, *atp6*, *atp8*, and *atp9*. In the mitochondrial genomes of fungi and plants, three genes encoding hydrophobic components of the F_1 sector of F_1F_0 -ATP synthase are usually found: *atp6*, *atp8*, and *atp9*. These three genes are absent in the mtDNA of *C. reinhardtii*.

The relocation of *C. reinhardtii atp6* to the nucleus occurred relatively late in evolution, after the massive transfer of genes from the protomitochondrion to the nucleus (43). In the evolution of the green algae, the transfer of *atp6* to the nucleus occurred after the separation of the *Scenedesmus* and *Prototheca* lineages, where *atp6* is still mtDNA-encoded, from the *Chlamydomonas* and *Chlorogonium* lineages, where the *atp6* gene is no longer encoded in the mitochondrial genome.

The transfer of the *atp6* gene to the nuclear genome is unlikely to be exclusive to *C. reinhardtii*. Other organisms that lack the *atp6* gene in their mitochondrial genomes and that are likely to have transferred it to the nucleus are the closely related algae *C. eugametos* (9) and *C. elongatum* (10), the apicomplexan organisms *P. falciparum* (41) and *Plasmodium*

reichenowi (44), and the ciliates *Paramecium aurelia* (45) and *Tetrahymena pyriformis* (46). It would be interesting to examine the sequences of ATP6 from more distantly related organisms to determine whether they have reduced physical constraints for import and to determine where reductions in hydrophobicity occur in the protein.

Mean Hydrophobicity Has Strongly Decreased in Those Transmembrane Regions of Cr-ATP6 That Are Not Critical for Function—The transfer of the *atp6* gene from the mtDNA to the nucleus was accompanied by, or preceded by, a strong decrease in the overall hydrophobicity of the encoded protein (Figs. 4 and 5) and probably facilitates the import and assembly of Cr-ATP6 into an active F_1F_0 -ATP synthase. Hydrophobicity analyses, in combination with the alignment of all ATP6 sequences available and previous published observations (33–36), allowed us to predict the presence of five well defined transmembrane regions in Cr-ATP6, named A–E (Fig. 5). The decreased hydrophobicity occurs predominantly in transmembrane regions A, B, and C (Fig. 6), which are poorly conserved among other ATP6 proteins and are thought not to participate directly in the proton translocating function of ATP6. In contrast, the well conserved transmembrane regions D and E maintain similar levels of mean hydrophobicity to mtDNA-encoded ATP6 subunits. We previously observed a similar phenomenon in the COX III proteins of *C. reinhardtii* and *Polytomella* sp. COX III is typically mtDNA-encoded but has been transferred to the nucleus in these algae (11). The nucleus-encoded COX III proteins showed greater diminished hydrophobicity in regions of the protein not in contact with the COX I subunit. This suggests that decreases in mean hydrophobicity of mitochondrial proteins whose genes have been relocated to the nucleus are more likely to occur in regions of the protein not involved in subunit-subunit interactions or in protein function.

Cr-ATP6 contains several amino acids that are conserved in all ATP6 sequences. Helix B ends in a highly conserved Pro and contains two additional conserved amino acids, Phe⁶² and Asn⁶⁹. The presence of Asn is not very common in transmembrane segments, so we believe its presence may have a role in transmembrane helix association through interhelical hydrogen bonding (47, 48). Transmembrane helices D and E contain most of the conserved amino acids; helix D is preceded by a highly conserved Glu¹³¹ involved in the proton translocation pathway and contains the invariant amino acids Ser¹⁴¹, Leu¹⁴², Gly¹⁴³, Leu¹⁴⁶, Asn¹⁴⁹, Ala¹⁶², Gly¹⁶³, and His¹⁶⁴. There is also an invariant Arg¹⁴⁵ that has been previously described as a residue involved in the protonation of the *c* subunit oligomer. Helix E contains the conserved Glu¹⁹¹, which is hypothesized to be involved in proton translocation, and Tyr²¹², which may be structurally important, after the end of this helix (Fig. 5).

Implications for Human Mitochondrial Gene Therapy—A considerable number of human diseases have been associated with point mutations or deletions in the mitochondrial genome (49–51), and strategies to develop mitochondrial gene therapies have been suggested (52, 53). One promising approach for overcoming mutations in mtDNA-encoded proteins is to place a wild-type copy of the affected gene in the nucleus and target the expressed protein to the mitochondrion to replace the defective mtDNA-encoded protein. Such allotypic expression of a normally mtDNA-encoded gene has been successfully performed in *S. cerevisiae* to overcome a mutation in the mtDNA-encoded *atp8* gene (54).

Analysis of nuclear forms of proteins that are normally encoded in the mtDNA may provide insights that will facilitate the allotypic expression of genes harboring mutations in the mtDNA-encoded genes. The observed decreases in hydrophobicity of nuclear expressed COX III and ATP6 of *C. reinhardtii*

suggest that human mitochondrial genes could potentially be engineered for allotropic expression for gene therapy purposes by modifications to decrease the mean hydrophobicity of the protein product, especially in transmembrane regions that are not highly conserved or that are known to be noncritical for function. Such changes would be in addition to the changes in codon usage and the addition of an MTS-coding sequence that are a prerequisite for nuclear expression of a mitochondrial protein. The presence of long MTSs in COX III, COX IIA, and ATP6 from *C. reinhardtii* suggests that a long MTS would facilitate import in allotropic expression systems. A detailed knowledge of the topology and function of the nucleus-encoded hydrophobic mitochondrial proteins may enhance our understanding of the problems involved in allotropic expression in addition to providing insights into the evolutionary forces that have led to the nuclear localization of genes that are typically mtDNA-encoded.

Acknowledgments—We thank Dr. J. d'Alayer (Institut Pasteur) for expert help in sequencing peptides, Dr. L.-G. Franzén (Göteborg University) for the kind gift of the cDNA library of *C. reinhardtii*, Drs. M. Futai and H. Omote (Osaka University) for carrying out initial immunoblot analysis with antibodies against the *E. coli* subunit *a*, Dr. S. B. Vik (Southern Methodist University) for helpful discussions while predicting the Cr-ATP6 model, and several colleagues and researchers who kindly provided DNA probes that were used in this work: Drs. H. Handa (Hokkaido National Agricultural Experiment Station, Japan), K. Kadovaki (National Institute of Agrobiological Resources, Ibaraki, Japan), and F. Kempken (Ruhr-Universität Bochum, Germany). We are grateful to Drs. A. Atteia (Instituto de Fisiología Celular, Universidad Nacional Autónoma de México), D. W. Krogmann (Purdue University), M. Tuena (Instituto de Fisiología Celular, Universidad Nacional Autónoma de México), and S. B. Vik (Southern Methodist University) for critical comments to the manuscript.

REFERENCES

- Mitchell, P. (1966) *Biol. Rev.* 41, 445-502
- Pedersen, P. L., and Amzel, L. M. (1993) *J. Biol. Chem.* 268, 9937-9940
- Abrahams, J. P., Leslie, A. G., Lutter, R., and Walker, J. E. (1994) *Nature* 370, 621-628
- Stock, D., Leslie, A. G., and Walker, J. E. (1999) *Science* 286, 1700-1705
- Vik, S. B., Pedersen, A. R., and Antonio, B. J. (1998) *J. Biol. Chem.* 273, 16229-16234
- Rastogi, V. K., and Girvin, M. E. (1999) *Nature* 402, 263-268
- Gray, M. W., and Boer, P. H. (1988) *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 319, 135-147
- Michaelis, G., Vahrenholz, C., and Pratje, E. (1990) *Mol. Gen. Genet.* 223, 211-216
- Denovan-Wright, E. M., Nedelcu, A. M., and Lee, R. W. (1998) *Plant Mol. Biol.* 36, 285-295
- Kroymann, J., and Zatsche, K. (1998) *J. Mol. Evol.* 47, 431-440
- Pérez-Martínez, X., Vázquez-Acevedo, M., Tolkunova, E., Funes, S., Claros, M. G., Davidson, E., King, M. P., and González-Halphen, D. (2000) *J. Biol. Chem.* 275, 30144-30152
- Pérez-Martínez, X., Antaramian, A., Vázquez-Acevedo, M., Funes, S., Tolkunova, E., d'Alayer, J., Claros, M. G., Davidson, E., King, M. P., and González-Halphen, D. (2001) *J. Biol. Chem.* 276, 11302-11309
- Gorman, D. S., and Levine, R. P. (1985) *Proc. Natl. Acad. Sci. U. S. A.* 54, 1665-1669
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Asanizu, E., Nakamura, Y., Sato, S., Fukuzawa, H., and Tabata, S. (1999) *DNA Res.* 6, 369-373
- Franzén, L.-G., and Falk, G. (1992) *Plant Mol. Biol.* 19, 771-780
- Frohman, M. A. (1993) *Methods Enzymol.* 218, 340-356
- Eriksson, M., Gadeström, P., and Samuelsson, G. (1995) *Plant Physiol.* 107, 479-483
- Schägger, H. (1994) in *A Practical Guide to Membrane Protein Purification* (van Jagow, G., and Schägger, H., eds) pp. 81-104, Academic Press, Inc., San Diego, CA
- Schägger, H., and von Jagow, G. (1987) *Anal. Biochem.* 166, 368-379
- Markwell, M. A. K., Hans, S. M., Bieber, L. L., and Tolbert, N. E. (1978) *Anal. Biochem.* 87, 206-210
- Atteia, A., Dreyfus, G., and González-Halphen, D. (1997) *Biochim. Biophys. Acta* 1320, 75-84
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* 24, 4876-4882
- Claros, M. G. (1995) *Comput. Appl. Biosci.* 11, 441-447
- Claros, M. G., and von Heijne, G. (1994) *Comput. Appl. Biosci.* 10, 685-686
- Rost, B. (1996) *Methods Enzymol.* 266, 525-539
- Sillhoo, C. D. (1998) in *Rochevix, J. D., Goldschmidt-Clermont, M., and Morchant, S. (eds) The Molecular Biology of Chlamydia and Mitochondria in Chlamydomonas*, pp. 25-40, Kluwer Academic Publishers, Dordrecht, The Netherlands
- Nedelcu, A. M., Lee, R. W., Lemieux, C., Gray, M. W., and Burger, G. (2000) *Genome Res.* 10, 819-831
- Küek, U., Jekosch, K., and Holzamer, P. (2000) *Gene (Amst.)* 253, 13-18
- Amati, B. B., Goldschmidt-Clermont, M., Wallace, C. J., and Rochaix, J. D. (1988) *J. Mol. Evol.* 28, 151-160
- Nurani, G., and Franzén, L.-G. (1996) *Plant Mol. Biol.* 31, 1105-1116
- Claros, M. G., Perea, J., Shu, Y., Samaty, P. A., Popot, J. L., and Jacq, C. (1995) *Eur. J. Biochem.* 228, 762-771
- Lang, J. C., Wang, S., and Vik, S. B. (1998) *J. Biol. Chem.* 273, 16235-16240
- Valiyaveetil, F. L., and Fillingame, R. H. (1998) *J. Biol. Chem.* 273, 16241-16247
- Jiang, W., and Fillingame, R. H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 6607-6612
- Wada, T., Long, J. C., Zhang, D., and Vik, S. B. (1999) *J. Biol. Chem.* 274, 17353-17357
- Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V., and Schuster, W. (1993) *FEBS Lett.* 325, 140-145
- Franzén, L.-G. (1994) *Biologisches Membran* 11, 304-309
- Galanis, M., Devenish, R. J., and Nagley, P. (1991) *FEBS Lett.* 282, 425-430
- Bauer, M. F., Hoffmann, S., Neupert, W., and Brunner, M. (2000) *Trends Cell Biol.* 10, 5-31
- Festin, J. E., Werner, E., Gardner, M. J., Williamson, D. H., and Wilson, R. J. (1992) *Nucleic Acids Res.* 20, 879-887
- Lang, B. F., Burger, G., O'Kelly, C. J., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., and Gray, M. W. (1997) *Nature* 387, 493-497
- Gray, M. W., Lang, B. F., and Lang, F. (1999) *Science* 283, 1476-1481
- Conway, D. J., Fanella, C., Lloyd, J. M., Al-Joubori, B. M., Baloch, A. H., Somanath, S. D., Roper, C., Oduola, A. M., Mulder, B., Povoln, M. M., Singh, B., and Thomas, A. W. (2000) *Mol. Biochem. Parasitol.* 111, 163-171
- Pritchard, A. E., Seilhammer, J. J., Mahalingam, R., Sable, C. L., Venuti, S. E., and Cummings, D. J. (1990) *Nucleic Acids Res.* 18, 173-180
- Burger, G., Zhu, Y., Littlejohn, T. G., Greenwood, S. J., Schenar, M. N., Lang, B. F., and Gray, M. W. (2000) *J. Mol. Biol.* 297, 385-389
- Choma, C., Gratkowsky, H., Lear, J. D., and De Grado, W. F. (2000) *Nat. Struct. Biol.* 7, 161-166
- Zhou, F.-X., Merianos, H. J., Bronger, A. T., and Engelman, D. M. (2001) *Proc. Natl. Acad. Sci. U. S. A.* 98, 2250-2255
- Davidson, E., and King, M. P. (1997) *Trends Cardiovasc. Med.* 7, 16-24
- DiMauro, S., Bonilla, E., Davidson, M., Hirano, M., and Schon, E. A. (1998) *Biochim. Biophys. Acta* 1366, 199-210
- Wallace, D. C. (1999) *Science* 283, 1482-1488
- Collombet, J.-M., and Coutelle, C. (1998) *Mol. Med. Today* 4, 31-38
- de Grey, A. D. N. J. (2000) *Trends Biotechnol.* 18, 394-399
- Gray, R. E., Law, R. H. P., Devenish, R. J., and Nagley, P. (1996) *Methods Enzymol.* 264, 269-389

TESIS CON
FALLA DE ORIGEN

Apéndice II.

Artículo publicado

Subunit II of cytochrome c oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes.

Subunit II of Cytochrome *c* Oxidase in *Chlamydomonas* Algae Is a Heterodimer Encoded by Two Independent Nuclear Genes^{†*}

Received for publication, November 9, 2000
Published, JBC Papers in Press, November 27, 2000, DOI 10.1074/jbc.M010244200

Xochitl Pérez-Martínez‡, Anaid Antaramian‡, Miriam Vázquez-Acevedo‡, Soledad Funest‡, Elena Tolkunova§, Jacques d'Alayer¶, Manuel G. Claros||, Edgar Davidson§, Michael P. King§, and Diego González-Halphen‡**

From the ‡Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, México 04510, D.F. Mexico, §Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, ¶Laboratoire de Microsynthèse des Protéines, Département des Biotechnologies, Institut Pasteur, 75724 Paris Cedex 15, France, and ||Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, E-29071 Málaga Spain

The mitochondrial genomes of *Chlamydomonas* algae lack the *cox2* gene that encodes the essential subunit COX II of cytochrome *c* oxidase. COX II is normally a single polypeptide encoded by a single mitochondrial gene. In this work we cloned two nuclear genes encoding COX II from both *Chlamydomonas reinhardtii* and *Polytomella* sp. The *cox2a* gene encodes a protein, COX IIA, corresponding to the N-terminal portion of subunit II of cytochrome *c* oxidase, and the *cox2b* gene encodes COX IIB, corresponding to the C-terminal region. The *cox2a* and *cox2b* genes are located in the nucleus and are independently transcribed into mRNAs that are translated into separate polypeptides. These two proteins assemble with other cytochrome *c* oxidase subunits in the inner mitochondrial membrane to form the mature multi-subunit complex. We propose that during the evolution of the Chlorophyte algae, the *cox2* gene was divided into two mitochondrial genes that were subsequently transferred to the nucleus. This event was evolutionarily distinct from the transfer of an intact *cox2* gene to the nucleus in some members the Leguminosae plant family.

phobic polytopic proteins that contain two or more transmembrane stretches (5, 6). The genes for *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, and *nad6* (encoding subunits 1, 2, 3, 4, 4L, 5, and 6 of NADH:ubiquinone oxidoreductase), *cob* (encoding cytochrome *b* of the *bc* complex), *cox1*, *cox2*, and *cox3* (encoding subunits COX I, COX II, and COX III of cytochrome *c* oxidase), and *atp6* and *atp8* (encoding subunits *a* and A6L of the F₀ portion of ATP synthetase) are found in the mitochondrial genomes of most organisms.

The transfer of mitochondrial genes to the nucleus is an ongoing process, as shown by the presence of particular genes encoded in both the mitochondrial and the nuclear genomes in certain species. These are exemplified by COX II in some members of the family leguminosae (7–9) and by subunit 9 of ATP synthetase in *Neurospora crassa* (10).

The algae of the family Chlamydomonadaceae, including *Chlamydomonas reinhardtii* and *Polytomella* sp., lack some of the genes that are typically found in mitochondrial genomes, including *nad3*, *nad4L*, *cox2*, *cox3*, *atp6*, and *atp8* (11–13)¹. We have shown that, in at least two members of this family, *C. reinhardtii* and *Polytomella* sp., the *cox3* gene was transferred to the nucleus, and the corresponding mitochondrial copy has been lost (14).

Several mitochondrial respiratory chain complexes have been isolated and characterized from the colorless alga *Polytomella* sp. (15–17), a close relative of *Chlamydomonas*, including an active, cyanide-sensitive cytochrome *c* oxidase preparation (14). In this work, we show that COX II is present as a heterodimer in this complex. All COX II sequences that have been described to date are single polypeptides encoded by one gene, normally in the mitochondrial genome. In both *Polytomella* sp. and *C. reinhardtii*, COX II is encoded by two nuclear genes that were named *cox2a* and *cox2b*. The *cox2a* gene encodes a protein, COX IIA, corresponding to the N-terminal half of a typical one-polypeptide COX II, and the *cox2b* gene encodes a protein, COX IIB, equivalent to the C-terminal half of the same subunit. We propose that these separate genes give rise to a heterodimeric COX II that results from the noncovalent assembly of the COX IIA and COX IIB polypeptides in cytochrome *c* oxidase of the inner mitochondrial membrane.

EXPERIMENTAL PROCEDURES

Strain and Culture Conditions—*Polytomella* sp. (198.80, E. G. Pringsheim), from the Sammlung von Algenkulturen (Göttingen, Germany), was grown as previously described (15).

¹ S. Funas, A. Antaramian, and D. González-Halphen, unpublished results.

Mitochondria are thought to descend from free-living α -proteobacteria (1), whose closest extants are bacteria of the genus *Rickettsia* (2). After the endosymbiotic event, there was a transfer of genes from the protomitochondrion to the nucleus such that few genes now remain in mitochondrial genomes (3). Those genes that remain encode only a sub-set of the mitochondrial proteins needed for oxidative phosphorylation and a portion of the factors necessary for their expression (4). The mtDNA-encoded respiratory chain subunits are highly hydro-

* This work was supported by National Institutes of Health Grants TW01176 and HL59646, Consejo Nacional de Ciencia y Tecnología, Mexico Grant 27754N, and Dirección General de Apoyo al Personal Académico, Universidad Nacional Autónoma de México Grant IN202598. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF305078 (*Polytomella* sp. *cox2a* cDNA), AF305079 (*Polytomella* sp. *cox2b* cDNA), AF305080 (*C. reinhardtii* *cox2a* cDNA), AF305540 (*C. reinhardtii* *cox2b* cDNA), AF305541 (*Polytomella* sp. genomic *cox2a*), AF305542 (*Polytomella* sp. genomic *cox2b*), and AF305543 (*C. reinhardtii* genomic *cox2b*).

** To whom correspondence should be addressed. Tel.: 011-525-622-5620; Fax: 011-525-622-5611 or 011-525-548-0387; E-mail: dhalphen@fisiol.unam.mx.

TESIS CON
FALLA DE ORIGEN

Protein Analysis—Cytochrome c oxidase from *Polytomella* sp. was obtained as previously described (14). Polyacrylamide gel electrophoresis was performed as in Schagger *et al.* (18) using 16% acrylamide gels. For tryptic digestion analysis, gels were stained with Amido Black, and the polypeptide of interest was excised from the gel. Polypeptides were isolated as previously described (15) for N-terminal sequencing. Amino-terminal Edman degradation was carried out on an Applied Biosystems Sequencer at the Laboratoire de Microséquence des Protéines, Institut Pasteur, Paris, France. An 18.6-kDa polypeptide was isolated from polyacrylamide gels and subjected to tryptic and endolysin-C digestion and separation on DEAE-C14 and DEAE-C18 HPLC² columns. Peaks eluted from the columns were subjected to N-terminal sequence analysis.

Nucleic Acids Preparation—Total DNA and total RNA from *Polytomella* sp. and *C. reinhardtii* were isolated as previously described (14). PCR fragments were cloned into pMOS blue-T (Amersham Pharmacia Biotech) or pGEM-T easy (Promega). cDNA was prepared from 1–2 µg of total RNA with Moloney murine leukemia virus reverse transcriptase (Promega) or Superscript II reverse transcriptase (Life Technologies). All standard molecular biology techniques were as described (19). Sequencing was carried out by the Kimmel Cancer Center DNA Sequencing Facility, Thomas Jefferson University, and at the Unidad de Biología Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México.

Cloning the cox2b Gene from Polytomella sp.—A genomic *Polytomella* sp. *cox2b* fragment was amplified by PCR using two degenerate primers: F1 (5'-CA(A/G)GA(C/T)AG(C/T)GC(C/T)AC(A/T)AG(C/T)CA(G/A)GC(C/T)CA(A/G)G-3') based on internal sequence (QD-SATSQAQA) of COX II, and B1 (5'-TG(GA/AT)(C/T)AA(AV/GC)(A/T)CC(AAT)GG(A/G)AT(AV/G)CC(A/G)TC(CAT-3') based on the internal sequence MDAIPGR(L)N. The resulting 300-nt product was used to isolate genomic *cox2b* clones from a library of *Polytomella* DNA *Pst*I fragments of ~2 kilobases in length in pBluescript.

The *cox2b* cDNA sequence from *Polytomella* sp. was obtained by PCR and 5'-RACE (20) using primers based on the genomic sequence obtained above. A poly(dT) tail was added to the 5' end of the cDNA with a terminal transferase (Roche Molecular Biochemicals). The forward primer was oligo(dT)/adapter primer: 5'-GACTCGAGTCGACATC-GATTTTTTTTTTTTTTTT-3', and the reverse primer (B2) was 5'-AGCTGTTAAGACCATGACTCT-3'.

Cloning the cox2a Gene from Polytomella sp.—A *cox2a* cDNA fragment was amplified using the degenerate primers F2 (5'-GAG(A)GC(T/C)CC(T/C)GT(T/C)GC(T/C)TGG CAG CT(T/G)GG-3'), based on the N-terminal protein sequence EAPVAWQLG, and B3 (5'-CCA(AV)GTA CCA CTG(AV)TG(AV)TCC(AV)AT(AV)GG C-3'), based on the internal conserved sequence KAIGHQWYV. Nested PCR was done with degenerate primers F3 (5'-CAG GA(T/C)TC(T/C)GC(T/C)AC(T/C)TC(T/C)CAG GC(T/C)CAG G-3'), based on the N-terminal sequence of the protein QDSATSQAQA and B4 (5'-GA(A/G)TA(AV)G(AV)G(AV)G(AV)GC(AV)AA(AV)G(AV)G(AV)GG-3'), based on the internal conserved sequence PSFALLYS. For 3'-end cDNA cloning, oligo(dT)/adapter primer and primer F4 (5'-TCCCTTACCACATCGCCACCC-3') were used. For nested PCR, adapter (5'-GACTCGAGTCGACATCGA-3') and primer F5 (5'-ACTACTACTAAGCAAGCTCTCCCTG-3') were used. For 5'-RACE, primers B5 (5'-TCAGGGAGAGCTTGTAGTGTAG-3'), with oligo(dT)/adapter primer, and B6 (5'-TTGGTGGCAGTGTGGTAGAGG-3'), with adapter for nested PCR, were used. Primers F6 (5'-AATGCTCGCCACGCTATC-3') and B7 (5'-AAACCTTCACACCCATAGG-3'), derived from the 5' and 3'-RACE products, were used to amplify the full-length cDNA of the *cox2a* gene. The genomic sequence of the *Polytomella* sp. *cox2a* gene was obtained by PCR amplification of total *Polytomella* sp. DNA with the same primers.

Cloning the cox2a and cox2b Genes of C. reinhardtii—The cDNAs for *cox2a* and *cox2b* were obtained by screening a *C. reinhardtii* cDNA library in Agt10 (21) using the *Polytomella* sp. *cox2a* cDNA or *cox2b* genomic DNA as probes. A total of 17 positive clones were obtained from 5 × 10⁴ plaque-forming units screened. PCR using two primers based on Agt10 sequences was used to identify the positive clones with the largest inserts. Phage DNA from these clones was isolated with the Qigen Lambda mini kit. The 5' end of the *cox2a* cDNA was completed by

RACE using primers based on the cDNA sequence obtained.

A bacterial artificial chromosome clone containing *cox2b* was obtained from a BAC genomic library from *C. reinhardtii* (22) by Genome Systems using the *C. reinhardtii* *cox2b* cDNA obtained above. BAC DNA was sequenced directly using internal primers.

Protein Sequence Analysis—Mitochondrial targeting sequences were analyzed using MitoProt II (23, 24). The same program was used to calculate the segments with high local hydrophobicity (<H>) in a distance comprising 13 to 17 residues. The *mesoH* was determined by scanning each sequence for a maximum average hydrophobicity measured in windows from 60 to 80 residues and averaging the values (24). More hydrophobicity scales were included to reduce the possibility of bias. Protein transmembrane stretches were predicted using the program TopPred II (25). Three-dimensional structure modeling was carried out using SWISS-MODEL (26).

Data Base Accession Numbers—The nucleotide sequences are in the DDBJ/EMBL/GenBankTM nucleotide sequence data base under the accession numbers AF305078 (*Polytomella* sp. *cox2a* cDNA), AF305079 (*Polytomella* sp. *cox2b* cDNA), AF305080 (*C. reinhardtii* *cox2a* cDNA), AF305540 (*C. reinhardtii* *cox2b* cDNA), AF305541 (*Polytomella* sp. genomic *cox2a*), AF305542 (*Polytomella* sp. genomic *cox2b*), and AF305543 (*C. reinhardtii* genomic *cox2b*).

RESULTS

Subunit II Is Present in the Cytochrome c Oxidase Complex from Polytomella sp.

An active cytochrome c oxidase from *Polytomella* sp. was isolated as previously described (14). The four largest polypeptides with apparent molecular masses of 54.6, 29.6, 18.6, and 14.5 kDa were subjected to Edman degradation. The 54.6-kDa polypeptide, not susceptible to Edman degradation, was identified as subunit I of cytochrome c oxidase (COX I) since its mass was similar to that predicted for the mitochondrial *cox I* gene sequence (54,781 Da (27)). The 29.6-kDa polypeptide (N-terminal sequence: SSDAGHLLSPREYRLV) was previously identified as subunit III (14). The N-terminal sequence of the 14.5-kDa polypeptide, DANSELVLEPTRKYKAGLATRELV, did not show similarity with any sequences in GenBankTM.

The N-terminal sequence of the 18.6-kDa polypeptide, EAPVAWQLGFQDSATSQAQA, was similar to COX II from other species, allowing identification of this polypeptide as COX II. A second sequence, MDAIPGR(L)NQLWLTINREG, was obtained during the Edman degradation of the 18.6-kDa polypeptide region at a yield of less than 50% that of the yield of the first sequence. This sequence also was similar to COX II from other species and was identified as an internal fragment of COX II that was obtained after partial cleavage of the protein during Edman degradation.

Cloning of the cox2b Gene from Polytomella sp. and C. reinhardtii—On the basis of the primary amino acid sequences obtained for COX II, two degenerate oligodeoxynucleotide primers were designed. Using these primers, a PCR amplification product of 300 nt was obtained using total DNA from *Polytomella* sp. as a template. This PCR product encoded the C-terminal portion of the *cox2* gene but lacked the region that encodes the N-terminal sequence of COX II. The absence of an N-terminal sequence was attributed to nonspecific annealing of the primer based on the N-terminal amino acid sequence.

The 300-nt *cox2* gene fragment hybridized to a 2-kilobase *Pst*I fragment of *Polytomella* sp. total DNA in Southern analyses. To obtain a full-length gene, a mini-library was constructed from *Pst*I fragments of ~2 kilobases, and a positive clone was isolated and sequenced. This genomic sequence contained a 462-nt open reading frame that encoded a 153-amino acid protein homologous to the C terminus of COX II. The gene was named *cox2b*. There was no open reading frame corresponding to the N-terminal portion of COX II in the 960 nt preceding the 462-nt open reading frame.

Using primers based on the genomic sequence, a portion of a cDNA corresponding to the *cox2b* open reading frame was

²The abbreviations used are: HPLC, high precision liquid chromatography; *mesoH* mesohydrophobicity; mtDNA mitochondrial DNA; nt, nucleotide(s); PCR, polymerase chain reaction; <H>, maximum local hydrophobicity of a sequence segment; µH, hydrophobic moment; RACE, rapid amplification of cDNA ends; BAC, bacterial artificial chromosome.

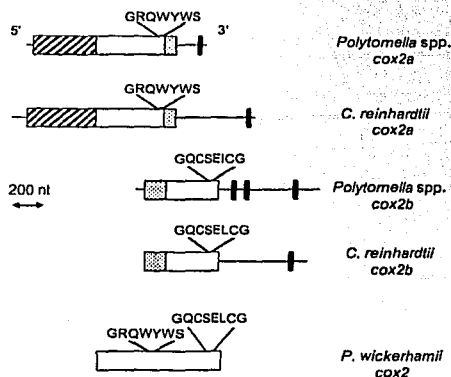


Fig. 1. Organization of the *cox2a* and *cox2b* genes of *Chlamydomonas* algae. The coding regions of *cox2a* and *cox2b* cDNAs from *Polytomella* sp. and *C. reinhardtii* are indicated by a box. The regions that are conserved with other COX II proteins are white, the putative mitochondrial-targeting sequences are striped, and the highly charged sequences unique to the *Chlamydomonas* *cox2* genes are stippled. Non-translated regions are shown as a black line. Putative polyadenylation signals are indicated as vertical black bars. The locations of the highly conserved COX II sequences GRQWYWSY and GQCSE(I/L)CG are indicated above the boxes. The single *cox2* mitochondrial gene of *P. wickerhamii*, representing a canonical *cox2* gene, is included for comparison.

amplified from *Polytomella* sp. total RNA using reverse transcription-PCR. The remainder of the 5'-cDNA sequence was obtained by 5'-RACE. The full-length cDNA obtained contained a 65-nt 5'-untranslated region and a 462-nt open reading frame identical to that of the genomic clone. This showed that this gene encoded only the C-terminal portion of COX II and did not contain introns. The overall organization of the *cox2b* gene is shown in Fig. 1. The predicted protein contained the MDAIPGRLNQIWLINREG internal sequence that was obtained from direct protein sequencing as well as the sequence GQCSEICG known to be the COX II binding site for binuclear copper (28). The major N-terminal sequence of COX II, determined by Edman degradation, was not present in the deduced protein sequence. The N-terminal 43 amino acids lacked homology to any COX II proteins. The remaining sequence was homologous with the C-terminal half of many COX II proteins. The highest similarity was with COX II from the alga *Prototheca wickerhamii* (29). Altogether, these data suggested that the *cox2* gene had been split into two genes in *Polytomella* sp.

The PCR amplification product of the *cox2b* gene of *Polytomella* sp. was also used to isolate a *cox2b* cDNA from a λ gt10 cDNA library of *C. reinhardtii*. A comparison of this *cox2b* cDNA with the genomic sequence (see below) showed that this *cox2b* cDNA sequence lacked the first four codons of the *cox2b* gene and a 5'-untranslated region. It exhibited 85% identity with the *Polytomella* sp. *cox2b* (Fig. 2). The *cox2b* cDNA from *C. reinhardtii* was used to isolate *cox2b* from a *C. reinhardtii* BAC genomic library. The genomic *cox2b* sequence contained the complete coding region and was identical to the cDNA except for the presence of one intron of 187 nt, located 21 nucleotides upstream of the stop codon. There was no open reading frame encoding a known protein in the 1.7 kilobases preceding the start codon for the *C. reinhardtii* *cox2b* gene. The predicted COX IIB protein was 153 amino acids long and was 85% identical and 92% similar to COX IIB of *Polytomella* sp. Both *C. reinhardtii* and *Polytomella* sp. are therefore likely to use two genes to code for COX II.

Cloning the *cox2a* Gene from *Polytomella* sp. and *C. reinhardtii*—To clone the gene that encoded the N-terminal region of COX II, nested PCR was performed with primers derived from the major N-terminal sequence obtained from the protein, and internal sequences (KAIGHQWYW and PSFALLYYS) were conserved among COX II proteins. Using *Polytomella* sp. cDNA as a template, a 250-nt PCR product was obtained that exhibited similarity with other *cox2* genes. The full-length cDNA, obtained using 5'- and 3'-RACE (Fig. 1), contained an open reading frame of 816 nt predicted to encode a protein of 271 amino acids including the sequence EAPVAWQLGF determined for the N terminus of the 18.6-kDa polypeptide. This gene was named *cox2a*, since it encoded the N-terminal portion of the COX II protein (COX IIA).

The N terminus of the mature *Polytomella* sp. COX IIA protein, derived from direct sequencing, corresponded to Glu-131 of the predicted sequence, indicating that this protein contains a mitochondrial targeting sequence of 130 amino acids. The mature protein is predicted to be 141 amino acids long and to contain two putative transmembrane stretches, from Ile-28 to Thr-48 and from Val-69 to Leu-89, and the highly conserved sequence GRQWYWSY present in all sequences of COX II subunits known to date (Fig. 2). The sequence from Glu-131 to Glu-246 was homologous with the N-terminal portion of many COX II proteins. This sequence was most similar to COX II from the alga *P. wickerhamii*. The COX IIA protein contained a C-terminal 20-amino acid region, lacking similarity to conventional COX II proteins, that had a high density of charged amino acids.

Primers corresponding to the 5' and 3' ends of the *cox2a* cDNA sequence of *Polytomella* sp. were used to amplify a portion of the *cox2a* nuclear gene from total DNA. This 1773-nt PCR product was cloned and sequenced. The genomic *cox2a* gene contained 6 introns, ranging in size from 84 to 136 nt.

The *Polytomella* sp. *cox2a* cDNA was used to isolate a partial cDNA clone of *cox2a* from a *C. reinhardtii* cDNA library. The complete sequence of *C. reinhardtii* *cox2a* was obtained by 5'-RACE (Fig. 1). The *C. reinhardtii* *cox2a* cDNA contained a 5'-untranslated region of 30 nt, an open reading frame of 855 nt, encoding a protein of 284 amino acids, and a 3'-untranslated region of 201 nt. The predicted *C. reinhardtii* COX IIA mature polypeptide exhibited 72% identity and 81% similarity with the sequence predicted for the COX IIA protein from *Polytomella* sp. (Fig. 2). The gene sequence predicts an extension of 21 residues at the C-terminal end that lacked homology to COX II proteins but that was highly similar to the extension predicted by the *cox2a* gene from *Polytomella* sp.

The *cox2a* and *cox2b* Genes of *Polytomella* sp. and *C. reinhardtii* Are Nuclear-localized, Single-copy Genes—The nuclear location of both *cox2a* and *cox2b* was confirmed by Southern analysis. After electrophoresis of *Polytomella* sp. total DNA through agarose gels, the mtDNA was detected as a discrete band below the bulk of the nuclear DNA (Fig. 3A, left panel). This was confirmed with a *cox1* gene probe (27) that hybridized with the mtDNA and a β -tubulin gene probe (30) that hybridized with the nuclear DNA (Fig. 3A, left panel). The *cox2a* and *cox2b* genes of *Polytomella* sp. hybridized with the major DNA fraction and not with the mtDNA band, confirming their nuclear localization. A similar analysis was carried out with total DNA from *C. reinhardtii* using a *cox1* gene probe that hybridized with the mtDNA (31) and a cytochrome c gene probe (32) that hybridized with the nuclear DNA (Fig. 3A, left panel). Those results confirmed that the *cox2a* and *cox2b* genes were also nuclear-encoded in *C. reinhardtii*.

To determine whether *cox2a* and *cox2b* were present as single-copy genes in the genomes of *Polytomella* sp. and *C. rein-*

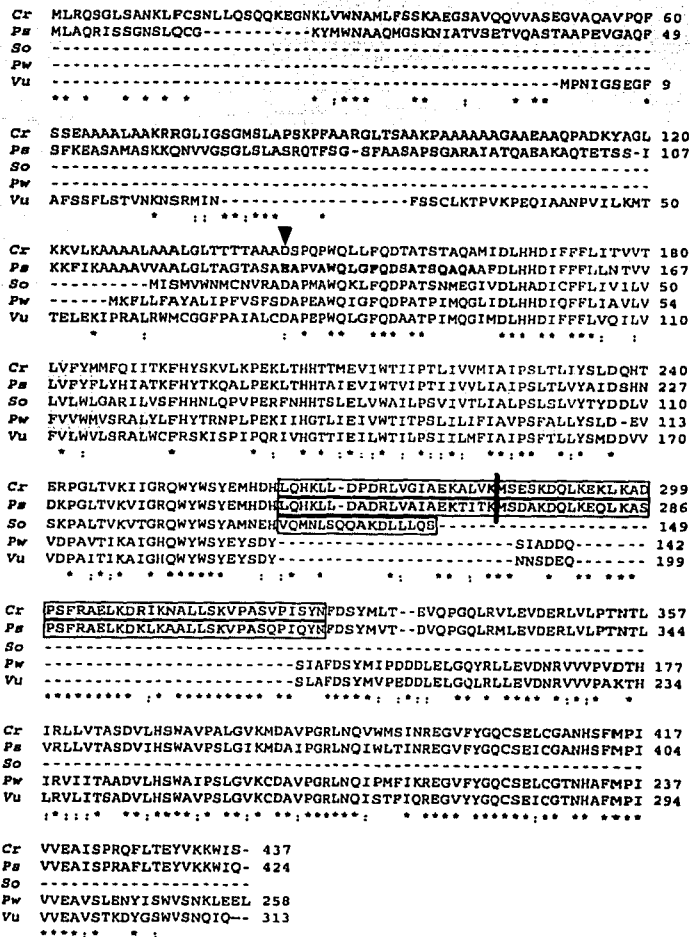


Fig. 2. Sequence comparison of COX II, COX IIA, and COX IIB polypeptides. Shown are the sequences of COX II of *P. wickerhamii* (*Pw*), *Vigna unguiculata* (*Vu*), and *S. obliquus* (*So*). Also shown are COX IIA and COX IIB of *C. reinhardtii* (*Cr*), and *Polytomella* sp. (*P#*). A black bar indicates the end of the COX IIA protein of *C. reinhardtii* and *Polytomella* sp. and the N terminus of the COX IIB protein. An asterisk denotes an identical amino acid, and a colon indicates a similar amino acid. The 20-amino acid extension at their C terminus of COX IIA and the 42-amino acid extension at the N terminus of COX IIB are boxed. These unique extensions of COX IIA and COX IIB are hypothesized to interact for the assembly of the two COX II subunits in the cytochrome *c* oxidase complex. The putative transmembrane stretches are indicated within gray boxes. The black triangle indicates the N terminus of the mature COX IIA that was determined for *Polytomella* sp. The highly conserved COX II regions G(R/H)QWYWSY present within COX IIA and GQCS(E/L)CG present within COX IIB are underlined. Numbers indicate the last amino acid for each protein.

hardtii, additional Southern blot analyses were performed. Single hybridization bands were found for the *cox2a* and *cox2b* genes of *Polytomella* sp. and *C. reinhardtii* with several restriction enzymes (Fig. 3, panel B and data not shown). This suggested that both genes were present in only one copy in their respective genomes.

The expression of *cox2a* and *cox2b* of *Polytomella* sp. and *C. reinhardtii* was examined by Northern analyses. Probes derived from the *cox2a* and *cox2b* genes hybridized to independent transcripts of sizes consistent with the corresponding cDNA sequences for both algae (Fig. 3, panel C). The presence of a larger, mature transcript that could suggest a transcribed

product was not observed in any case. The *cox2b* gene of *Polytomella* sp. exhibited a double band. Since the genomic sequence of this *cox2b* gene contained three putative polyadenylation sites in the 3'-noncoding region, it is possible that these bands correspond to mRNAs that have different sites of polyadenylation.

There is a significant bias in codon usage in these genera of algae (30), and this bias differs between mitochondrial and nuclear genes (14). Analysis of the codon usage for the *cox2a* and *cox2b* genes of *Polytomella* sp. and *C. reinhardtii* indicated that the codon usage was consistent with their nuclear localization (data not shown). In addition, the conserved poly-

TESIS CON FALLA DE ORIGEN

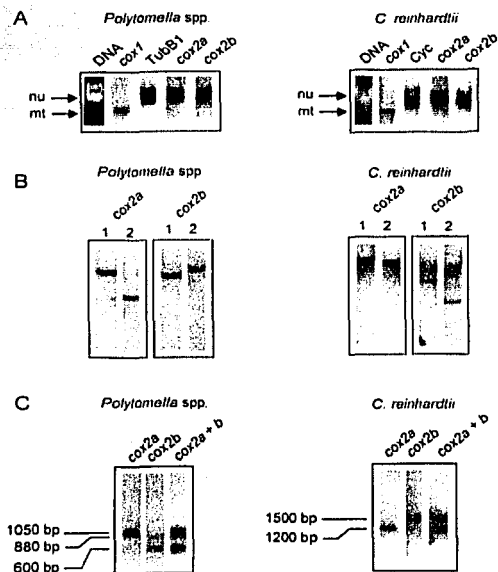


FIG. 3. The *cox2a* and *cox2b* genes are single-copy, nuclear-localized, and functionally expressed in Chlamydomonas algae. **A**, nuclear localization of the *cox2a* and *cox2b* genes. Thirty μ g of total DNA from *Polytomella* sp. (left panel) or *C. reinhardtii* (right panel) was electrophoresed through a 0.7% agarose gel. The first lanes (DNA) of both panels are photographs of the ethidium bromide-stained sample after electrophoresis. The positions of the nuclear DNA (*nu*) and mtDNA (*mt*) are shown by arrows. The remaining lanes show autoradiograms after Southern analysis. The left panel shows *Polytomella* sp. DNA hybridized with probes for mtDNA-encoded *Polytomella* sp. *cox1* (second lane), nuclear DNA-encoded *Polytomella agilis* β -tubulin (*TubB1*, third lane), and *Polytomella* sp. *cox2a* (fourth lane) and *cox2b* (*cox2b*, fifth lane) genes. The right panel shows Southern analyses of the *C. reinhardtii* DNA hybridized with mtDNA-encoded *C. reinhardtii* *cox1* (second lane), nuclear DNA-encoded cytochrome *c* of *C. reinhardtii* (*Cyc*, third lane), and *C. reinhardtii* *cox2a* (fourth lane) and *cox2b* (*cox2b*, fifth lane) genes. **B**, *cox2a* and *cox2b* are single-copy genes. Total DNA from *Polytomella* sp. (left) was digested with *EcoRV* (lane 1) or *PvuII* (lane 2) for *cox2a* and *EcoRV* (lane 1) or *EcoRI* (lane 2) for *cox2b*. Total DNA from *C. reinhardtii* (right) was digested with *KpnI* (lane 1) or *XbaI* (lane 2) for *cox2a* and *KpnI* (lane 1) or *HindIII* (lane 2) for *cox2b*. Shown are autoradiograms of the Southern blot analyses of the restricted DNAs hybridized with probes for *cox2a* or *cox2b*. **C**, Northern blot analysis. Shown are autoradiograms of Northern blots of total RNA isolated from *Polytomella* sp. (left) or *C. reinhardtii* (right) hybridized with probes specific for *cox2a* and *cox2b* genes from *Polytomella* sp. and *C. reinhardtii*. The sizes calculated for the hybridizing RNAs are indicated to the left of the autoradiograms. *bp*, base pairs.

adenylation signals TGTA (33), present in the vast majority of nuclear genes in the Chlamydomonas family, were present at the 3' ends of the cDNA sequences of *cox2a* and *cox2b* for both algae.

Two COX II Polypeptides Are Present in *Polytomella* sp. Cytochrome *c* Oxidase—Since the primary protein sequences derived from the 18.6-kDa region of a polyacrylamide gel corresponded to the predicted amino acid sequences of both COX IIA and COX IIB from *Polytomella* sp., it is likely that this region of the gel contained both subunits. To confirm that *Polytomella* sp. contained two independent COX II polypeptides, the purified cytochrome *c* oxidase complex of this alga was subjected to matrix-assisted laser desorption-time of flight mass spectrometry analysis. The complex contained two major

TABLE I
Mass spectrometry analysis of the tryptic fragments obtained from the COX IIB subunit of *Polytomella* sp.

The experimental values were obtained by matrix-assisted laser desorption-time of flight mass spectrometry analysis of tryptic digestion fragments of COX IIB purified by reverse transcription-HPLC from isolated cytochrome *c* oxidase. These masses are compared with the predicted molecular masses for the tryptic fragments from the *cox2b* gene product.

Sequence of the tryptic peptide	Theoretical molecular mass	Experimental molecular mass
	<i>D_a</i>	<i>D_a</i>
AFLETEYVK	970.12	970.448
LVLPTNTLVR	1125.37	1125.649
LNQIWLTVNR	1270.48	1270.702
LLVTASDV...AVPSLGIK	2106.47	2123.172
VPASQPIQ...DVQPGQLR	2854.18	2854.383
EGVIFYGQC...VVEAISPR	3141.56	3140.432

polypeptides in the expected mass range for COX IIA and COX IIB, one of 15,984 Da (theoretical 16,222 Da for mature COX IIA) and one of 17,169 Da (theoretical 17,219 Da for full-length COX IIB). The differences between the predicted and observed masses are likely due to post-translational modifications of the proteins. The 17,169-Da polypeptide was isolated by reverse phase-HPLC, trypsinized, and analyzed by matrix-assisted laser desorption-time of flight mass spectrometry analysis. The trypsin digestion products exhibited molecular masses that were consistent with the theoretical molecular masses expected for COX IIB peptides (Table I).

We wished to determine whether COX IIB contained a cleavable mitochondrial-targeting sequence or if the novel N-terminal charged domain was present in the mature protein. The observed band of 18.6 kDa (containing COX IIA and COX IIB) was purified from polyacrylamide gels and subjected to digestion by trypsin or endolysin followed by HPLC separation and Edman degradation (Fig. 4). Amino-terminal sequences were obtained for all the endolysin C-derived fragments, except for the 16.0-kDa polypeptide, which exhibited a blocked N terminus. The sequence was obtained from COX IIB starting at Asp-5 of the full-length sequence (tryptic fragment DQLK). That the mass spectrometry analysis of the mature protein gave a mass consistent with an intact N terminus suggests mature COX IIB is the full-length protein. Several overlapping sequences were obtained for the tryptic products and for the endolysin C fragments of the protein. A total of 154 amino acids out of the 294 residues predicted by the *cox2a* and *cox2b* genes were identified in the 18.6-kDa region of the gel by direct amino acid sequence analysis. This analysis, using a different cytochrome *c* oxidase preparation from that first used by Edman degradation, confirmed that Glu-131 was the N terminus of the mature COX IIA. No difference was found between the predicted sequences and the sequences obtained by Edman degradation.

DISCUSSION

COX II Is Encoded by Two Distinct Nuclear Genes in the Algae of the Family Chlamydomonadaceae—COX II is conserved in cytochrome *c* oxidases throughout all phyla. It constitutes the main core of the enzyme along with COX I and COX III (34). Although the *cox2* gene is absent from the mtDNA of *Polytomella* sp., the presence of COX II in cytochrome *c* oxidase from *Polytomella* sp. was confirmed by N-terminal and internal amino acid sequence analysis of polypeptides with an apparent molecular mass of 18.6 kDa from the purified complex. This size is consistent with a 14.5-kDa polypeptide that was identified immunochemically as COX II in cytochrome *c* oxidase isolated from *C. reinhardtii* (35).

Polytomella sp. and *C. reinhardtii* each contain two distinct

TESIS CON
FALLA DE ORIGEN

N-terminal region of the Ps-COX II subunit (encoded by the *cox2a* gene):

EAPVAWQLGFGDSATSQAQAAFDLHHDIFFFLLNTVVL
VVFYFLYHIATK'FHYTK'QALPEK'LTHTTAIEVIWTVIPTII
VVLIAIPSLTLVYAIDSHNDK'PGLTVK'VIGR'QWYWSYE
MHDHLQHK'LLDADR'LVIAIEK'TITK

C-terminal region of the Ps-COX II subunit (encoded by the *cox2b* gene):

MSADK'DQLK'EQLK'ASPSFR'AEK'DK'LK'AALLSK'VPA
SQPIQYNFDSYMTVDVQPGQLR'MLEVDER'LVLPNTL
VR'LLVTSADVHSWAVPSLGIK'**MDAIPGR'LNQIWLTI**N
R'EGVVFYQQCEICGANHSFMPIVVEAISPR'AFLEYV
K'K'WIQ

Fig. 4. Protein sequence analysis of the COX IIA and COX IIB polypeptides of *Polytomella* sp. Shown are the amino acid sequences of COX IIA and COX IIB. The polypeptides present in the region of the polyacrylamide gel (containing COX IIA and COX IIB) were isolated and subjected to digestion by trypsin or endolysin. HPLC separation, and Edman degradation. The N-terminal sequences were obtained for the endolysin C-derived fragments and tryptic digestion products of the proteins. The amino acids directly identified by these amino acid sequence analyses are in *bold italic* characters. The regions identified by N-terminal Edman degradation analysis of the 18.6-kDa region of the polyacrylamide gel are boxed. A total of 154 amino acids out of the 294 amino acids were identified in this analysis. No difference was found between the predicted sequences and the sequences obtained by Edman degradation.

nuclear genes that encode two COX II homologs. *cox2a* encodes a protein, COX IIA, corresponding to the N-terminal half of a canonical COX II. *cox2b* encodes a protein, COX IIB, corresponding to the C-terminal half of a canonical COX II. The *cox2a* and *cox2b* genes from *Polytomella* sp. and *C. reinhardtii* were shown to be nuclear-localized by Southern hybridization analyses. These genes also displayed other features typical of nuclear genes in Chlamydomonas algae: a biased codon usage, introns in the genomic sequences, and a TGTA polyadenylation signal.

COX IIA and IIB were unresolved by SDS-polyacrylamide gel electrophoresis and migrated with an apparent molecular mass of 18.6 kDa. Mass spectrometry and reverse phase HPLC analyses confirmed that distinct COX IIA and IIB proteins existed in the purified cytochrome *c* oxidase of *Polytomella* sp. The COX IIB polypeptide may have a blocked N terminus (MSADKQL), since only the major N-terminal sequence EAPVAWQLGF of COX IIA was detected by direct N-terminal sequence analysis of the 18.6-kDa polypeptide.

COX IIA and IIB Have Been Adapted to Facilitate Import into Mitochondria.—Most nuclear-encoded mitochondrial proteins contain a mitochondrial targeting sequence that directs the import of the protein into mitochondria. Such sequences have been identified in the nuclear-encoded *cox2* genes from cowpea (7) and soybean (8). The predicted full-length COX IIA proteins each has a mitochondrial targeting sequence of 130 amino acids, since the amino acid sequence obtained for the mature *Polytomella* sp. COX IIA commences at Glu-131 of the predicted polypeptide. These sequences are 37% identical (Fig. 2). The COX IIB proteins from these organisms also contain unusually long (>100 amino acids) mitochondrial-targeting sequences (14). It is possible that during import into mitochondria the targeting sequences may act as chaperones for the import of the COX subunits and the assembly of the enzyme complex. Alternatively, after cleavage they may be maintained

as components of the cytochrome *c* oxidase complex, as was observed with the targeting sequence of the Rieske subunit of beef heart mitochondrial cytochrome *bc₁* complex (36).

In contrast to COX IIA, the COX IIB proteins lack a canonical, cleavable, mitochondrial-targeting sequence, since the COX IIB sequence was obtained starting at Asp-5 of the full-length sequence (Fig. 4). Additionally, mass spectrometry analysis of the mature protein gave a mass consistent with an intact N terminus. However, the N-terminal 43 amino acids in both predicted COX IIB proteins show no similarity with other known COX II sequences and contains a high density of charged residues. The numerous negatively charged amino acids are not consistent with a mitochondrial-targeting sequence, normally characterized by a paucity of acidic residues and abundance of basic residues (37). It is possible that the 43-amino acid stretch may function as a noncleavable mitochondrial-targeting sequence. The N-terminal region of *Polytomella* sp. COX IIB possesses a positively charged amphiphilic α -helix from position 19 to 44 with a high hydrophobic moment (μ H ranging from 5.98 to 7.42) and a very hydrophobic face (maximum hydrophobicity ranging from 3.55 to 6.35). Similarly, the N-terminal region of *C. reinhardtii* COX IIB possesses a putative amphiphilic helix from residues 14 to 32, with μ H ranging from 7.93 to 8.92 and a hydrophobic face (maximum hydrophobicity ranging from 4.70 to 6.44). The amphiphilicity is essential for the function of a mitochondrial targeting sequence (38). Alternatively, there may be no conventional N-terminal, cleavable mitochondrial targeting sequence, as in eukaryotic cytochrome *c*. Like COX IIB, cytochrome *c* is imported into the inner membrane space, but a minimum length of N-terminal region rather than a specific sequence is necessary for efficient mitochondrial import (39).

The highest average hydrophobicity over 60–80 amino acids of a polypeptide chain (termed mesohydrophobicity, or *mesoH*) along with the maximum local hydrophobicity (<H>) of likely transmembrane segments are useful indicators of the likelihood that a protein could be imported into the mitochondrion (23). Both *Polytomella* sp. and *C. reinhardtii* mature COX IIA proteins, despite their two transmembrane domains, exhibited <H> and *mesoH* patterns significantly lower than mtDNA-encoded COX II polypeptides from other organisms but still higher than most proteins imported into mitochondria (data not shown). The nuclear-encoded COX II proteins from legumes were also less hydrophobic than their mtDNA-encoded counterparts and more similar to the COX IIA proteins of the Chlamydomonas algae.

The low hydrophobicity parameters exhibited by the COX IIB proteins may also facilitate their import into mitochondria. These proteins contain no transmembrane stretches, exhibit very low <H> and *mesoH* values, and are predicted to be imported readily into mitochondria (data not shown).

Assembly of COX IIA and COX IIB in Cytochrome *c* Oxidase.—The normal COX II subunit exhibits two transmembrane stretches and a hydrophilic region containing the consensus copper binding site. We predict that the two mature Chlamydomonas COX II polypeptides would assemble to give a heterodimeric COX II with an overall structure similar to that determined by x-ray crystallography of the bacterial and mammalian monomeric COX II subunits (28, 40).

Hydropathy profile analysis predicted two transmembrane stretches for the COX IIA polypeptides from both algae (data not shown), similar to conventional COX II proteins. In addition, both contain a C-terminal 20-amino acid region that has a high density of charged amino acids and is not homologous to known COX II proteins. The two COX IIB polypeptides contain a 42-amino acid extension at their N terminus that also exhib-

TESIS CON
 FALLA DE ORIGEN

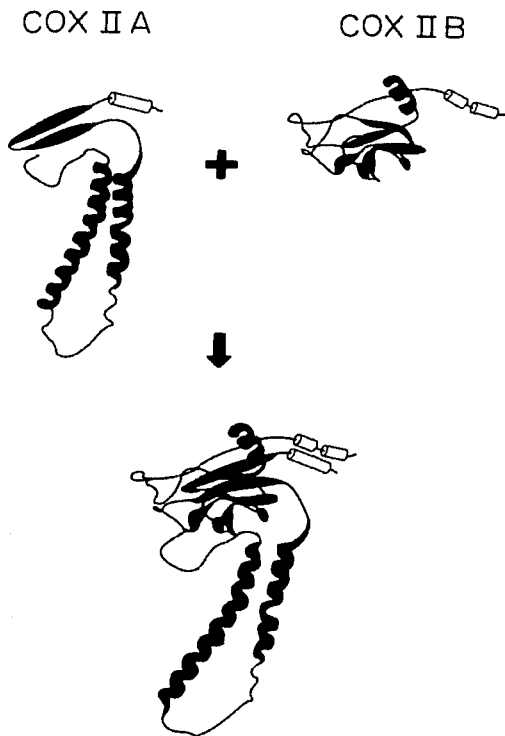


Fig. 5. Model of the interaction of COX IIA with COX IIB to form the heterodimeric COX II of *Chlamydomonas* algae. A three-dimensional structure of *Polytomella* sp. COX IIA and COX IIB was modeled with SWISS-MODEL based upon the three-dimensional structure of COX II from the bovine enzyme (28). This model proposes an interaction between the unique COX IIA C-terminal domain and the highly charged COX IIB N-terminal domain. We hypothesize that the loop formed by the interaction of the N-terminal and C-terminal extensions is topologically distant from the site of interaction of soluble cytochrome *c* with the COX IIB subunit.

its a high density of charged amino acids and that is also not homologous to known COX II proteins. We propose that the C-terminal extension of COX IIA interacts with the N-terminal extension of the COX IIB protein. This interaction may stabilize the two COX II subunits in the cytochrome *c* oxidase complex (Fig. 5). According to this model, the extra loop formed by the interaction of the N-terminal and C-terminal extensions is topologically distant from the site of interaction of soluble cytochrome *c* with the COX IIB subunit.

Evolutionary Considerations—The transfer of *cox2* genes from the mitochondria to the nucleus in the *Chlamydomonas* algae satisfies many of the criteria proposed by Brennicke *et al.* (41) and Claros *et al.* (42) for this event: acquisition of a region encoding a mitochondrial-targeting sequence (at least for *cox2a*), altered codon usage, acquisition of a polyadenylation signal, and diminished $\langle H \rangle$ and *mesoII* of the protein products. In addition, any mitochondrial *cox2* genes that presumably existed have been eliminated, suggesting that this transfer occurred early in evolution.

There are several precedents for mitochondrial genes being

split into two. The mitochondrial *nad1* genes of *Tetrahymena pyriformis* and *Paramecium aurelia* (43) and rapeseed mitochondrial gene encoding a homologue of the bacterial protein Ccl1 (44) have also been divided into two independently transcribed reading frames, although retained on the mtDNA.

It is possible that in some organisms the transfer of *cox2a* and *cox2b* to the nucleus is ongoing, as in some legumes, or has been arrested at an intermediate stage. *cox1*, *cox2*, and *cox3* were present in the mtDNA of the Chlorophycean alga *Scenedesmus obliquus* (45, 46). However, the *cox2* gene was truncated and was predicted to be a pseudogene. The putative *Scenedesmus* COX II protein exhibits high similarity with COX IIA from *Polytomella* sp. and *C. reinhardtii* (Fig. 2), although it lacks a mitochondrial-targeting sequence, as would be expected for a mtDNA-encoded protein. The C-terminal region of the *Scenedesmus* COX II is also similar to the unique C-terminal regions of COX IIA of *Polytomella* sp. and *C. reinhardtii*. We suggest that the mitochondrial *cox2* gene from *S. obliquus* is likely to be an active *cox2a* gene and that the missing fragment of the gene (*i.e.* *cox2b*) has been transferred to the nucleus.

The gene content and fragmentation pattern of the ribosomal RNA genes on the mtDNA of *S. obliquus* suggest that this mitochondrial genome represents an intermediate stage between the *Prototheca*-like green algae and the *Chlamydomonas*-like green algae (45). *S. obliquus* may represent a stage of green algal evolution in which the *cox2b* gene (encoding for a highly hydrophilic polypeptide) has been transferred to the nucleus, whereas the *cox2a* gene (encoding a more hydrophobic polypeptide) is retained in the mitochondrial genome. Since the mitochondrial genetic code of *S. obliquus* has diverged from the standard genetic code (45, 46), it is likely that the *cox2a* gene has been retained on the mtDNA since it could not be functionally transferred to the nucleus in its current form. We hypothesize that two independent *cox2a* and *cox2b* genes existed in the mitochondrial genome of the common ancestor of *Chlamydomonas* algae and that these genes were transferred from the mitochondria to the nucleus before the separation of the colorless genera of *Polytomella* from the main photosynthetic lineage of *Chlamydomonas*.

The functional transfer of intact *cox2* genes to the nucleus has been reported in detail for members of the legume subfamily Papilionoideae (7, 8, 9). Our analysis of the *Chlamydomonas* *cox2a* and *cox2b* genes suggests that the division of the mitochondrial *cox2* gene into two nuclear genes that encode two independent proteins is an evolutionary event unrelated to the functional gene transfer of intact mitochondrial *cox2* genes to the nucleus in some legumes.

A number of organisms, particularly protists, have mitochondrial genomes that lack the *cox2* gene, including the apicomplexan protozoan *Plasmodium falciparum* (47) and the Chlorophyte alga *Pedinomonas minor* (1), whose mtDNA has been sequenced. It is likely that these organisms have transferred their *cox2* genes to the nucleus and that, in the green alga *P. minor*, this has also involved the transfer of *cox2a* and *cox2b* to the nucleus. Further analyses of such organisms will determine whether the process outlined here has an evolutionary lineage outside of the Chlorophyta.

The present work highlights a novel strategy utilized by some Chlorophyte algae that allowed functional transfer of a mitochondrial gene to the nucleus. This strategy has allowed the functional transfer to the nucleus of a mitochondrial gene encoding an essential, hydrophobic subunit of a crucial respiratory complex. It is remarkable that a protein with such specific functional requirements (*e.g.* a cytochrome *c* docking site, binding site for metals involved in catalysis) should have

TESIS CON
FALLA DE ORIGEN

evolved into two components containing added domains that maintain the ability of the protein to perform its roles.

Acknowledgments—We thank Yun-Lu and T. A. Neubert (New York University) for mass spectroscopy analysis, M. W. Gray (Dalhousie University) and J. D. Palmer (Indiana University) for encouragement at the starting point of this project, L.-G. Franzén (Göteborg University) for the kind gift of the cDNA library of *C. reinhardtii*, L. Ongay, G. Codiz, and M. Sosn (Unidad de Biología Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México) for the synthesis of various oligonucleotides, and several colleagues and researchers who kindly provided DNA probes: T. D. Fox (Cornell University), M. Goldschmidt-Clermont (University of Geneva), E. H. Harris (Duke University), J. D. Palmer (Indiana University), and C. D. Sillflow (University of Minnesota). We are also indebted to A. Attein and A. Gómez-Puyou (Instituto de Fisiología Celular, Universidad Nacional Autónoma de México) and D. W. Krogmann (Purdue University) for critical review of the manuscript.

REFERENCES

- Gray, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brassard, N., Delage, F., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., and Burger, G. (1998) *Nucleic Acids Res.* **26**, 865–878
- Andersson, S. G. E., Zomorodipour, A., Andersson, J. O., Sicheritz-Pontén, T., Alsmark, U. C. M., Podowski, R. M., Nslund, A. K., Eriksson, A.-S., Winkler, H. H., and Kurland, C. G. (1998) *Nature* **396**, 133–140
- Gray, M. W. (1992) *Int. Rev. Cytol.* **14**, 233–357
- Attardi, G., and Schatz, G. (1988) *Annu. Rev. Cell Biol.* **4**, 289–333
- von Heijne, G. (1986) *FEBS Lett.* **198**, 1–4
- Popot, J. L., and de Vitry, C. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19**, 369–403
- Nugent, J. M., and Palmer, J. D. (1991) *Cell* **66**, 473–481
- Covello, P. S., and Gray, M. W. (1992) *EMBO J.* **11**, 3815–3820
- Adams, K. L., Song, K., Roessler, P. G., Nugent, J. M., Doyle, J. L., and Palmer, J. D. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 13863–13868
- van den Boogaart, P., Samallo, J., and Agsteribbe, E. (1982) *Nature* **298**, 187–189
- Michaelis, G., Vahrenholz, C., and Prütje, E. (1990) *Mol. Gen. Genet.* **223**, 211–216
- Demvian-Wright, E. M., Nedeleu, A. M., and Lee, R. W. (1998) *Plant Mol. Biol.* **36**, 285–295
- Krogmann, J., and Zetsche, K. (1998) *J. Mol. Biol.* **47**, 431–440
- Pérez-Martínez, X., Vázquez-Acevedo, M., Tolkunova, E., Funes, S., Claros, M. G., Davidson, E., King, M. P., and González-Halphen, D. (2000) *J. Biol. Chem.* **275**, 30144–30152
- Gutiérrez-Cirlos, E. B., Antaramian, A., Vázquez-Acevedo, M., Coria, R., and González-Halphen, D. (1994) *J. Biol. Chem.* **269**, 9147–9154
- Attein, A., Dreyfus, G., and González-Halphen, D. (1997) *Biochim. Biophys. Acta* **1320**, 275–284
- Bramme, S., Krauß, V., Schmitz, U.-K., and Braun, H.-P. (1998) *J. Biol. Chem.* **273**, 13143–13149
- Schöttger, H., Link, T. A., Engel, W. D., and von Jagow, G. (1986) *Methods Enzymol.* **126**, 224–237
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Frohman, M. A. (1993) *Methods Enzymol.* **218**, 340–356
- Franzén, L.-G., and Falk, G. (1992) *Plant Mol. Biol.* **19**, 771–780
- Lefebvre, P. A., and Sillflow, C. D. (1999) *Genetics* **151**, 9–14
- Claros, M. G. (1995) *Comput. Appl. Biosci.* **11**, 441–447
- Claros, M. G., and Vincens, P. (1996) *Eur. J. Biochem.* **241**, 779–786
- Claros, M. G., and von Heijne, G. (1994) *Comput. Appl. Biosci.* **10**, 685–686
- Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714–2723
- Antaramian, A., Coria, R., Ramirez, J., and González-Halphen, D. (1996) *Biochim. Biophys. Acta* **1273**, 198–202
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. (1996) *Science* **272**, 1136–1144
- Wolff, G., Plante, I., Lang, B. F., Kück, U., and Burger, G. (1994) *J. Mol. Biol.* **237**, 75–86
- Conner, T. W., Thompson, M. D., and Sillflow, C. (1989) *Gene* **84**, 345–358
- Gray, M. W., and Buer, P. H. (1988) *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* **319**, 135–147
- Amati, B. B., Goldschmidt-Clermont, M., Wallace, C. J., and Rochaix, J. D. (1988) *J. Mol. Evol.* **28**, 151–160
- Sillflow, C. (1998) *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas* (Rochaix, J.-D., Goldschmidt-Clermont, M., and Merchant, S., eds) pp. 25–40, Kluwer Academic Publishers Group, Dordrecht, Netherlands
- Saraste, M. (1990) *Q. Rev. Biophys.* **23**, 331–366
- Bennoun, P., Attein, A., Pierre, Y., and Delosme, M. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10202–10206
- Iwata, S., Lee, J. W., Okada, K., Lee, J. K., Iwata, M., Rasmussen, B., Link, T. A., Ramaswamy, S., and Jap, B. K. (1998) *Science* **281**, 64–71
- Claros, M. G., Brunnak, S., and von Heijne, G. (1997) *Curr. Opin. Struct. Biol.* **7**, 394–398
- Roise, D., Theiler, F., Horvath, S. J., Tomich, J. M., Richards, J. H., Allison, D. S., and Schatz, G. (1988) *EMBO J.* **7**, 649–653
- Wang, X., Dumont, M. E., and Sherman, F. (1996) *J. Biol. Chem.* **271**, 6584–6594
- Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. (1995) *Nature* **376**, 660–669
- Brennick, A., Grohmann, L., Hiesel, R., Knoop, V., and Schuster, W. (1993) *FEBS Lett.* **325**, 140–145
- Claros, M. G., Perea, J., Shu, Y., Sumatey, F. A., Popot, J. L., and Jacq, C. (1995) *Eur. J. Biochem.* **228**, 762–771
- Edqvist, J., Burger, G., and Gray, M. W. (2000) *J. Mol. Biol.* **297**, 381–393
- Handa, H., Bonnard, G., and Grienberger, J.-M. (1996) *Mol. Gen. Genet.* **252**, 292–302
- Nedeleu, A. M., Lee, R. W., Lemieux, C., Gray, M. W., and Burger, G. (2000) *Genome Res.* **10**, 819–831
- Kück, U., Jekosch, K., and Holzamer, P. (2000) *Gene* **253**, 13–18
- Fenglin, J. E. (1992) *Mol. Biochem. Parasitol.* **62**, 146–148

TESIS CON
FALLA DE ORIGEN

Apéndice III.

Artículo publicado

**Unusual location of a mitochondrial gene: subunit III of
cytochrome c oxidase is encoded in the nucleus of
Chlamydomonad algae.**

Unusual Location of a Mitochondrial Gene

SUBUNIT III OF CYTOCHROME *c* OXIDASE IS ENCODED IN THE NUCLEUS OF CHLAMYDOMONAD ALGAE*

Received for publication, May 9, 2000, and in revised form, July 14, 2000
Published, JBC Papers in Press, July 17, 2000, DOI 10.1074/jbc.M003940200

Xochitl Pérez-Martínez‡, Miriam Vázquez-Acevedo‡, Elena Tolkunova§, Soledad Funes‡,
Manuel G. Claros¶, Edgar Davidson§, Michael P. Kings§, and Diego González-Halphen‡

From the ‡Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510, Distrito Federal México, the §Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, and the ¶Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, E-29071 Málaga, Spain

The algae of the family Chlamydomonadaceae lack the gene *cox3* that encodes subunit III of cytochrome *c* oxidase in their mitochondrial genomes. This observation has raised the question of whether this subunit is present in cytochrome *c* oxidase or whether the corresponding gene is located in the nucleus. Cytochrome *c* oxidase was isolated from the colorless chlamydomonad *Polytomella* spp., and the existence of subunit III was established by immunoblotting analysis with an antibody directed against *Saccharomyces cerevisiae* subunit III. Based partly upon the N-terminal sequence of this subunit, oligodeoxynucleotides were designed and used for polymerase chain reaction amplification, and the resulting product was used to screen a cDNA library of *Chlamydomonas reinhardtii*. The complete sequences of the *cox3* cDNAs from *Polytomella* spp. and *C. reinhardtii* are reported. Evidence is provided that the genes for *cox3* are encoded by nuclear DNA, and the predicted polypeptides exhibit diminished physical constraints for import as compared with mitochondrial-DNA encoded homologs. This indicates that transfer of this gene to the nucleus occurred before *Polytomella* diverged from the photosynthetic *Chlamydomonas* lineage and that this transfer may have occurred in all chlamydomonad algae.

(mtDNA)¹ and are synthesized inside the organelle (1). These subunits are homologous to the three major polypeptides of bacterial cytochrome *c* oxidases. There are strong similarities in the primary, secondary, and tertiary structures of these subunits, as evidenced by the x-ray crystallographic models of cytochrome *c* oxidases from *Paracoccus denitrificans* and bovine mitochondria (2, 3). In addition, a variable set of nuclear-encoded subunits that exhibit either no transmembrane stretch or a single transmembrane stretch are synthesized in the cytoplasm and imported into mitochondria (4).

A striking example of simplicity, in size and composition, is the 15.8-kilobase linear, double-stranded mtDNA from the green alga *Chlamydomonas reinhardtii*. This compact and highly diverged mitochondrial genome has been entirely sequenced (5). Several genes that encode essential components of oxidative phosphorylation that are usually found in mitochondrial genomes are absent in this mtDNA: *nad3* and *nad4L* (encoding subunits 3 and 4L of NADH-ubiquinol oxidoreductase), *cox2* and *cox3* (encoding COX II and COX III), and *atp6* and *atp8* (encoding subunits *a* and *A6L* of the F₀ portion of ATP synthase). These genes are also absent in the mtDNAs of the closely related algae *Chlamydomonas smithii* (6), *Chlamydomonas eugametos* (7), *Chlamydomonas moewussii* (8), and *Chlorogonium elongatum* (9). The absence of this set of genes seems to be a common feature of the algae from the family Chlamydomonadaceae and suggests that the complexes that participate in oxidative phosphorylation lack some of their classical polypeptide constituents or that the corresponding genes were transferred to the nucleus.

The algae of the colorless genus *Polytomella* are members of the family Chlamydomonadaceae (10) that diverged from the *Chlamydomonas* lineage by losing both the cell wall and the photosynthetic apparatus (11). The close relationship between the genera *Polytomella* and *Chlamydomonas* is supported by numerous morphological (10, 12), molecular genetic (13–16), and biochemical (17) studies. This colorless alga has been used to characterize the mitochondrial complexes of the chlamydomonad algae, because there is no interference by thylakoid components or by the cell wall during purification (17, 18). In the experiments described below, we report the isolation of an active, cyanide-sensitive cytochrome *c* oxidase from *Polytomella* spp. that contains COX III. The corresponding cDNAs of

Mitochondrial cytochrome *c* oxidase (EC 1.9.3.1), the terminal component of the respiratory chain, is an oligomeric membrane protein complex of 10–13 subunits that contains four redox components: a binuclear center Cu_A, heme *a*, heme *a*₃, and Cu_B. The transfer of electrons from reduced cytochrome *c* to molecular oxygen is coupled to proton translocation from the matrix to the intermembrane space. In most eukaryotic cells, the three largest subunits of cytochrome *c* oxidase (COX I, COX II, and COX III) are encoded by the mitochondrial DNA

* This work was supported by Grant TW01176-01 from the National Institutes of Health/Fogarty, Grant 2775-4N from Consejo Nacional de Ciencia y Tecnología, and Grant IN202598 from DGAPA, UNAM. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF233514 and AF233515.

‡ To whom correspondence should be addressed: Dept. de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, 04510, D.F. México. Tel.: 525-622-5620; Fax: 525-622-5611; E-mail: dhalphen@ifisiol.unam.mx.

cox3 were cloned and sequenced from both *Polytomella* spp. and *C. reinhardtii*. Evidence is provided that in these algae, this gene is not localized in the mitochondrial genome but in the nuclear genome. This contrasts with the location in the majority of eukaryotes. To our knowledge, this is the first example of a *cox3* gene that is found in the nuclear genome. Therefore, our data indicated that the transfer of the *cox3* gene occurred before the genus *Polytomella* diverged from the *Chlamydomonas* lineage and that such transfer is a common feature of the Chlamydomonadales family. The results also show that the nuclear-localized *cox3* gene encodes a polypeptide that exhibits diminished values for <H> and *meso*H (19), when compared with their mitochondrial counterparts in other organisms.

EXPERIMENTAL PROCEDURES

Strain and Culture Conditions—*Polytomella* spp. (198.80, E.G. Pringsheim) from the Sammlung von Algenkulturen (Göttingen, Germany), was grown as described previously (18).

Isolation and Solubilization of Mitochondria—Mitochondria from *Polytomella* spp. were obtained and solubilized in the presence of lauryl maltoside as described previously (18).

Isolation of Cytochrome c Oxidase from *Polytomella* spp.—Solubilized mitochondria were dialyzed against 50 mM Tris-HCl (pH 8.0), 1 mM MgSO₄, 1 mM phenylmethylsulfonyl fluoride, and 50 µg/ml of N^ε-*p*-tosyl-L-lysine chloromethyl ketone (TM buffer) containing 100 mM NaCl. The mixture was taken to 40% saturation of ammonium sulfate and 1.6% sodium cholate and then centrifuged at 10,000 × g for 15 min. The resultant green pellet was resuspended in TM buffer containing 100 mM NaCl and lauryl maltoside (1.2%). The mixture was centrifuged at 80,000 × g, and the supernatant was dialyzed against 10 volumes of TM buffer. The sample was applied to a DEAE-Bioelg A column equilibrated with TM buffer containing 0.1% lauryl maltoside. The column was washed with three bed volumes of the same buffer and a linear gradient from 0 to 100 mM NaCl was applied. A cytochrome c oxidase-enriched fraction identified by spectroscopy was eluted with the same equilibration buffer that contained 200 mM NaCl. The samples were concentrated by ultrafiltration on an Amicon YM100 filter and stored at -70 °C until use.

Purification of Cytochrome c Oxidase from Bovine Heart Mitochondria—Bovine heart cytochrome c oxidase was purified according to the method of Capaldi and Hayashi (20) and stored in small aliquots at -70 °C until use.

Spectroscopic and Activity Measurements—Cytochrome c oxidase activity was measured spectrophotometrically in a final volume of 3 ml that had 50 mM Tris-HCl (pH 8.0), 1 mM MgSO₄, 0.1 mg/ml dodecyl maltoside, 20 µM antimycin, and 30 µM reduced cytochrome c. The reaction was started by addition of the enzyme, and the absorbance change at 550 nm was followed. Cytochrome c concentration was calculated using the extinction coefficient 16.5 nm⁻¹ cm⁻¹ (21) and cytochrome c determination was done as described previously (22). Visible spectra were recorded at room temperature with a DW-2a UV/Vis SLM-Aminco spectrophotometer modified with the OLIS DW2 Conversion and OLIS software (On-line Instrument System Inc.).

Polyacrylamide Gel Electrophoresis, Immunoblots, and Protein Determination—Polyacrylamide gel electrophoresis was performed as described by Schägger *et al.* (23), using 1.2-mm-thick slab gels (16% acrylamide). Gels were fixed and stained as described in the same work. Apparent molecular masses were calculated based on the reported molecular masses of bovine cytochrome c oxidase (24). Immunoblotting was carried out in González-Halphen *et al.* (25). Antibodies against yeast Cox III were obtained from Molecular Probes (Eugene, Oregon). Protein concentrations were determined according to Markwell *et al.* (26).

Sequencing of Subunit III by Edman Degradation—The isolation of polypeptides for N-terminal sequencing was carried out as described previously (18). N-terminal sequencing was carried out by Dr. J. D'Alayer on an Applied Biosystems Sequencer at the Laboratoire de Microséquence des Protéines (Institut Pasteur, Paris, France).

Nucleic Acid Preparation—Two-liter cultures of *Polytomella* spp. grown for 72 h were collected and resuspended in 10 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM EDTA, 2% Triton X-100, and 1% SDS. Total DNA was extracted from broken cells two times with phenol/chloroform 1:1 and once with chloroform. The aqueous phase was precipitated with 3 M sodium acetate (pH 5.3) in the presence of ethanol, and the pellet was resuspended in water free of nucleases. The mixture was incubated in the presence of 2.5 µg of RNase DNase-free (Roche Molecular Bio-

chemicals) for 3 h, and DNA was extracted and precipitated as above. *C. reinhardtii* cells were collected and washed with TE buffer and resuspended in 100 mM sodium citrate (pH 7.0). The cells were frozen in liquid nitrogen and incubated at 60 °C for 15 min in the presence of 1 volume of 2% SDS. Total DNA was extracted and precipitated from broken cells as above. Total RNA from *Polytomella* spp. was obtained using the kit RNeasy Mini Kit (Qiagen).

Cloning and Sequencing of the Gene *cox3* from *Polytomella* spp.—A genomic *Polytomella* spp. *cox3* fragment was amplified by PCR using two degenerate oligodeoxynucleotides. The first one was based on the N-terminal sequence of the protein SDAGHLSL: 5'-TC(C/T) GA(C/T) GC(C/T) GG(C/T) CA(C/T) CA(C/T) CT(C/T) TC(C/T) CC-3'. The second one was based on an internal highly conserved sequence of the protein WH(F/M)VDVWV: 5'-AG CCA (G/A)AC (G/A)AC (G/A)TC (G/A)AC (C/G)A(T/A) (G/A)TG CC-3'. For PCR amplification using Vent polymerase (New England Biolabs), samples were denatured for 5 min at 94 °C and subjected to 30 cycles of 1 min denaturation at 94 °C, 1 min annealing at 40–55 °C, and 2 min extension at 72 °C. A 780-nucleotide product containing a fragment of the *cox3* gene was obtained and cloned with the pGEM-T Easy Vector System from Promega.

Cloning and Sequencing of cDNA of *cox3* from *Polytomella* spp.—cDNA sequence from *cox3* cDNA from *Polytomella* spp. was obtained with 5'- or 3'-RACE-PCR (27) using primers based on the genomic sequence obtained above. First strand cDNA templates were prepared from 1–2 µg of total RNA with Moloney murine leukemia virus reverse transcriptase from Promega or Display Thermo RT from Display Systems Biotech and using oligo dT/adaptor as first strand cDNA primer (oligo dT/adaptor, 5'-GAC TCG AGT CGA CAT CGA TTT TTT TTT TTT TTT TT-3'). For 3' end cDNA cloning, oligo dT/adaptor and B1 primer 5'-GCA TTA CCT CGT CCA CAC TGC-3' were used. A 1.1-kilobase product was amplified. Nested PCR was done with primers B2 (5'-GAT GGG CAT GCA TAC CGA TG-3') and B3 (5'-CAT GAA GAA GGT GGT ACC GTA GG-3') to confirm *cox3* identity. For 5' end cloning, a poly(A) tail was attached to 5' end with terminal transferase from Roche Molecular Biochemicals. For PCR amplification primers B4 (5'-CAT CCG TAT GCA TGC CCA TC-3') with oligo dT/adaptor and B5 (5'-CAA CGG ATC CGA ACA ACA AGG-3') with adaptor for nested PCR were used. A 600-nucleotide PCR product was obtained. Both RACE products were cloned with the pGEM-T easy vector system from Promega. The cDNA sequence was confirmed using primers B6 (5'-GAG GTC TCA GCT TCT TAA GGC TC-3') and B7 (5'-CGC ATA ACG CGA AGT CAC TA-3'). For PCR amplification, samples were denatured for 5 min at 94 °C and subjected to 30 cycles of 45 s denaturation at 94 °C, 1 min annealing at 48 °C, and 2.5 min extension at 72 °C. Primers B8 (5'-ATG AGG TCT CAG CTT CTT AAG GCT C-3') and B9 (5'-CGG ATA ACG CGA AGT CAC TAC-3') were used to amplify the complete *cox3* gene.

Cloning and Sequencing of cDNA of *cox3* from *C. reinhardtii*—A *C. reinhardtii* cDNA library in λgt10 (28) was screened using the *Polytomella* spp. 780-base pair PCR product corresponding to a portion of the genomic *cox3* gene. Eight positive clones were obtained from 5 × 10⁴ plaque-forming units screened. Two deoxyoligonucleotides based on λgt10 sequences were used to identify the longest positive clones (forward, 5'-AGC AAG TTC AGC CTG GTT AAG T-3', and reverse, 5'-CTT ATG AGT ATT TCT TCC AGG GTA-3'). Phage DNA from the clone containing the largest cDNA was isolated with the Qiagen Lambda Mini Kit. The *cox3* gene was subcloned into pBluescript. The 5' end of cDNA was completed by RACE PCR (the primers used were: forward, oligo dT/adaptor, and reverse, 5'-TGC TCC ATG TAG AAC TCC TTG C-3'). The sequences for nested PCR were: forward, oligo adapter, and reverse, 5'-GTT GGG GAC CTG AGG CTG C-3').

Sequence Analysis *In Silico*—Sequences were compared using the GCG Sequence Analysis Software Package (Genetics Computer Group, Madison, Wisconsin) (29). Alignments and construction of the cladogram were carried out with the Clustal X program (30) using sequences in the Swissprot data bank. Mitochondrial targeting sequences were analyzed and predicted using MitoProt II (31), including calculations of hydrophobic moment (μ H), high local hydrophobicity (<H>), and *meso*H. Protein transmembrane stretches were predicted using the program TodPred II (32). Three dimensional structure modelling was carried out using SWISS-MODEL (33).

Data Base Accession Numbers—The nucleotide sequences discussed in this paper will appear in the DDBJ/EMBL/GenBank™ nucleotide sequence data base under the accession numbers AF233514 (*cox3* cDNA sequence from *Polytomella* spp.) and AF233515 (*cox3* cDNA sequence from *C. reinhardtii*).

TESIS CON
 PALA DE ORIGEN

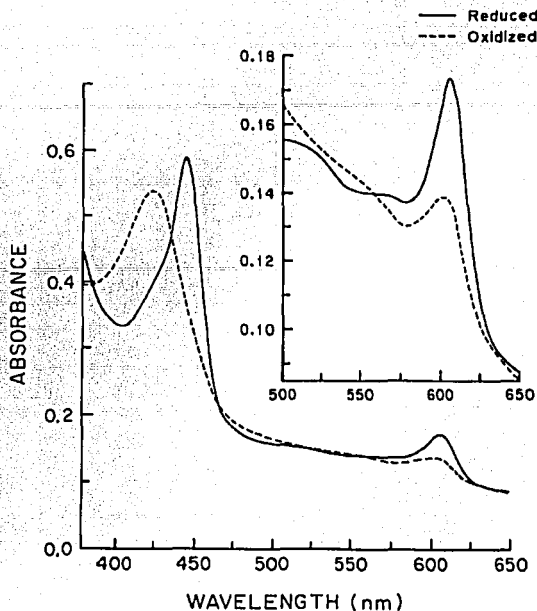


Fig. 1. Visible spectra of cytochrome *c* oxidase from *Polytomella* spp. The cytochrome *c* oxidase was diluted in 50 mM Tris-HCl (pH 8.0) containing 1 mM MgSO₄ and 0.1 mg/ml of lauryl maltoside. Broken lines, oxidized sample, as obtained. Continuous line, cytochrome *c* oxidase fully reduced in the presence of a small amount of dithionite. Inset, enlargement of the α absorption bands of the oxidized and reduced samples.

RESULTS

Isolation and Characterization of the Cytochrome *c* Oxidase Complex from *Polytomella* spp.—Cytochrome *c* oxidase was purified from the colorless alga *Polytomella* spp. The complex catalyzed electron transfer from horse heart cytochrome *c* to oxygen with a specific activity of 2.8 μ mol O₂/mg of protein/min, an activity that was completely abolished by cyanide or azide (data not shown). Absorption spectra of the cytochromes of the complex are shown in Fig. 1. The oxidized complex displayed a major absorbance peak in the Soret region at 425 nm; after reduction with dithionite, there was an increase in intensity and a shift of its maximal absorbance to 445 nm. The α -absorption peak exhibited a maximum at 605 nm in its reduced form, shifted 4–5 nm toward the red when compared with the absorption spectrum of cytochrome *c* oxidase type *aa₃* from other species. A red-shifted α -absorption peak at 606 nm was also described for reduced cytochrome *c* oxidase of *C. reinhardtii* (34). From the difference spectra (reduced with dithionite minus air-oxidized), a heme content of 3.03 nmol of heme *a*/mg of protein for the cytochrome *c* oxidase of *Polytomella* spp. was calculated.

The cytochrome *c* oxidase of *Polytomella* spp. exhibited seven polypeptides with molecular masses of 54.6, 29.6, 18.6, 14.5, 13.4, 10.8, and 9.6 kDa (Fig. 2A). The 29.6-kDa band was identified as subunit III of cytochrome *c* oxidase (see below). Two additional bands were present in this preparation, with apparent molecular masses of 80.0 and 41.8 kDa. These bands were considered contaminants and were not further explored.

In immunoblots, the 29.6-kDa polypeptide of *Polytomella*

spp. cytochrome *c* oxidase exhibited cross-reactivity with an antibody raised against COX III of *Saccharomyces cerevisiae* (Fig. 2B, lane 3). This band had a molecular mass similar to that of the corresponding subunit III of cytochrome *c* oxidase from yeast (Fig. 2B, lane 1). The 29.6-kDa polypeptide of *Polytomella* spp. was excised from the gel and extracted. The purified polypeptide still showed cross-reactivity with the anti-yeast antibody (Fig. 2B, lane 2). Accordingly, it was subjected to N-terminal sequencing. The N-terminal sequence obtained (SSDAGHHLSRPRERYLV) showed no similarity with any other COX III in the NCBI sequence data banks.

Characterization of the *cox3* Gene from *Polytomella* spp. and the *cox3* cDNA from *C. reinhardtii*—Two degenerate deoxyoligonucleotides were designed based on the N-terminal sequence of COX III from *Polytomella* spp. (Ps-COX III) and on a highly conserved internal sequence of COX III present in different organisms. With these oligonucleotide primers, a PCR amplification product of 780 nucleotides was obtained using total DNA from *Polytomella* spp. as a template. The DNA sequence obtained from the amplified product was predicted to encode a COX III protein. This sequence was used to design primers for use in 5'- and 3'-RACE using cDNA made from *Polytomella* spp. total RNA (27). The overlapping cDNA clones thus obtained were sequenced and a full-length cDNA sequence (DDBJ/EMBL/GenBank™ accession number AF233514) was obtained. The amplified genomic fragment was also used as a probe to screen a λ gt10 cDNA library from *C. reinhardtii*, and eight positive plaques were identified, isolated, and sequenced. The clone containing the longest cDNA was identified by PCR and sequenced, confirming the cDNA as encoding COX III. The overlapping regions of the genomic and cDNA sequences were identical. The sequence of the *cox3* cDNA from *C. reinhardtii* is not shown but is available in DDBJ/EMBL/GenBank™ with accession number AF233515.

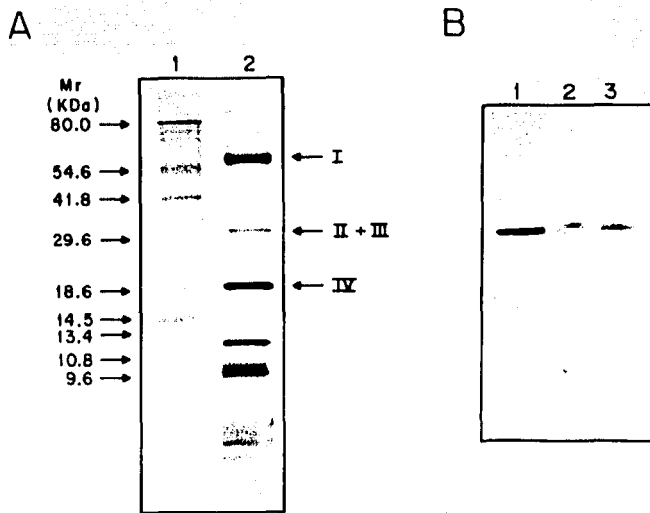
Translation of the cDNA sequences predicts a mature protein of 272 residues with a molecular mass of 29,978 Da for Ps-COX III, and a mature polypeptide of 272 residues (29,967 Da) for its close relative *C. reinhardtii* (Cr-COX III). Comparison of Ps-COX III with Cr-COX III indicated that the first 17 residues of the mature COX III sequences are highly conserved between *Polytomella* spp. and *C. reinhardtii* but are not present in the COX III sequence of the chlorophyte alga *Prototheca wickerhamii*. The alignment of the overall amino acid sequences of Ps-COX III and Cr-COX III (Fig. 3A) revealed an identity of 66.5% and a similarity of 73.9%. The similarity between the two subunits III of the cytochrome *c* oxidase is very high and extends over the complete protein sequences.

The *Polytomella* spp. *cox3* cDNA contains an open reading frame of 1113 base pairs, our identification of the N terminus of the mature protein as amino acid 99, allows us to predict a 98-amino acid mitochondrial targeting sequence (MTS). In *C. reinhardtii* the *cox3* cDNA contains an open reading frame of 1146 base pairs. Assuming that the N-terminal sequence of the mature protein corresponds to that of *Polytomella* spp., three different ATG codons could correspond to the initiation of the MTS. The upstream methionine predicts a presequence of 109 amino acids, the second methionine predicts a 51 residues MTS, and the downstream one predicts a presequence of 40 amino acids. It is known that the sequence surrounding start codon affects the efficiency of translation; in *C. reinhardtii* there is a consensus of (A/C)A(A/C)(A/C)ATG(G/C)C(C/G) for the start codon (35). According to these data, the upstream methionine that predicts a MTS of 110 amino acids is appropriate for translation initiation site.

The alignment of the presequences of Ps-COX III and Cr-COX III revealed 45.4% identity and a 50.5% similarity. This

TESIS CON
FALLA DE ORIGEN

FIG. 2. Subunit composition and immunoblot analysis of the cytochrome *c* oxidase complex from *Polytomella* spp. **A**, the cytochrome *c* oxidase preparation was analyzed on a 16% acrylamide gel stained with Coomassie Brilliant Blue (23) and compared with the bovine enzyme. **Lane 1**, cytochrome *c* oxidase from *Polytomella* spp. (20 μ g of protein). **Lane 2**, cytochrome *c* oxidase from beef-heart mitochondria (30 μ g of protein); its four major subunits are indicated. The apparent molecular masses are shown in kDa. **B**, blot immunostained with antibodies raised against COX III from *S. cerevisiae*. **Lane 1**, yeast cytochrome *c* oxidase (20 μ g of protein per lane). **Lane 2**, purified COX III from *Polytomella* spp. **Lane 3**, isolated cytochrome *c* oxidase from *Polytomella* spp. (20 μ g of protein/lane).



values are much higher when the first 17 residues MRSQLL(K/R)ALTRAPAGFS are compared or when the 8-residues region ALAALPPR just before the mature protein is compared. These sequences must play an important role in the processing of the MTS or in the import of COX III in these algae. Immediately upstream of the N terminus of the mature protein (as determined by protein sequencing of *Polytomella* spp. COX III) in both algae there is a methionine that could have been retained from the ancestral mitochondrial copy.

A cladogram was generated with COX III sequences from different organisms (Fig. 3B). The result obtained showed that Ps-COX III clearly affiliates with Cr-COX III, but surprisingly, these chlamydomonad COX III sequences appear close to yeast COX III sequences and relatively far away from the mitochondrial COX III sequences from other algae and from plants.

The pattern of codon utilization for the *cox3* gene of *Polytomella* spp. was compared with the pattern of codon usage of known nuclear and mitochondrial genes of this alga (Table I). As in other nuclear-localized genes, there is a significant bias in each codon family; this is because triplets that end in A are rare in the nuclear genome of this alga (14). The codon usage of the *cox3* gene of *Polytomella* spp. is typically nuclear and different from mitochondrial codon usage. A similar analysis was carried out for the *cox3* cDNA from *C. reinhardtii*. The codon usage pattern was similar to nuclear codon usage and differed from codon usage in the mitochondrial genome. In addition, the polyadenylation signals TGTAA (35) were found at the end of the cDNA sequences.

DNA blot analysis was carried out to ascertain that the *cox3* genes were encoded by the nuclear genome. Total DNA isolated from *Polytomella* spp. was electrophoresed through agarose. The mtDNA separated as a discrete band running below the major band representing nuclear DNA. The DNA from these gels was transferred to nylon membranes and subjected to hybridization analysis with a battery of probes from mitochondrial and nuclear origins. The smaller band hybridized with three different mitochondrial probes (Fig. 4A) *cob1*, encoding cytochrome *b* from *Polytomella* spp. (36); *nad1*, encoding subunit 4 of NADH-ubiquinone oxidoreductase from *Polytomella*

spp.²; and *cox1*, encoding subunit I of cytochrome *c* oxidase from *Polytomella* spp. (15). In contrast, nuclear DNA hybridized with the following nuclear probes: *Cytc1*, a partial sequence of the gene encoding cytochrome *c*₁ (18), and *TubB1*, the gene encoding β -tubulin from *Polytomella agilis* (now renamed *Polytomella parva*) (14). The *cox3* gene hybridized with the major DNA fraction and not with the mtDNA band, confirming its nuclear localization. A similar analysis was carried out with total DNA from *C. reinhardtii* (Fig. 4B). The smaller band hybridized with three different mitochondrial probes from *C. reinhardtii*, *cob1*, *nad2*, and *cox1* (5). In contrast, nuclear DNA hybridized with the following nuclear probes from *C. reinhardtii*: the gene *Cyc* encoding cytochrome *c* (37), the gene *Fes1* encoding the Rieske iron-sulfur protein (38), the gene *AtpB* encoding the β subunit of the ATP synthase (28), and the *cox3* gene obtained in this study.

Hydrophobicity and Importability of the Nuclear-encoded Subunit III of Cytochrome *c* Oxidase from *Chlamydomonas* Algae—Import studies suggest that the highest average hydrophobicity over 60–80 amino acids of a polypeptide chain (termed mesohydrophobicity), along with the maximum hydrophobicity of likely transmembrane segments, are useful indicators of the likelihood that a protein could be imported into the mitochondrion (19). The predicted Ps-COX III and Cr-COX III subunits were tested for their physical characteristics *in silico*. The computational analyses suggested that both proteins contain a bipartite MTS. The first 25 amino acids are predicted to be a MTS, whereas the segment up to residue 50 is predicted to be a mitochondrial inner membrane signal that should direct the peptide to its final location (Fig. 5). Both COX III polypeptides were compared with those encoded by other complete *cox3* genes in the data base; all are located at mitochondrial genomes. Fig. 6A shows a *mesoH versus* maximal local hydrophobicity (<H>) plot for different COX III sequences. In comparison with all their mitochondrial counterparts, Ps-COX III and Cr-COX III display both decreased local hydrophobicity and

² S. Funes, A. Antaramian, and D. González-Halphen, unpublished results.

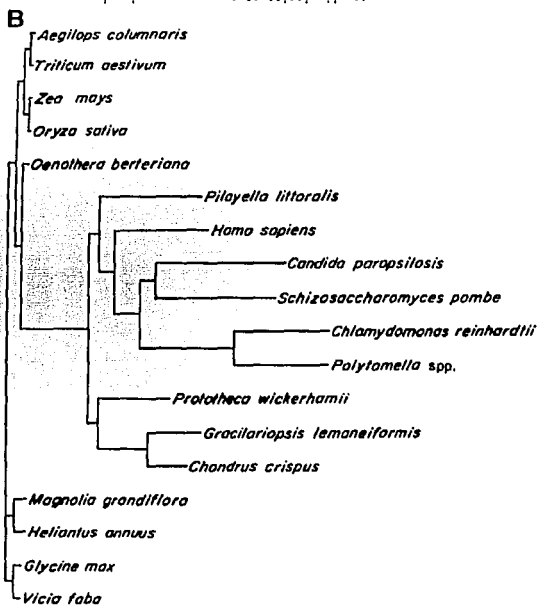
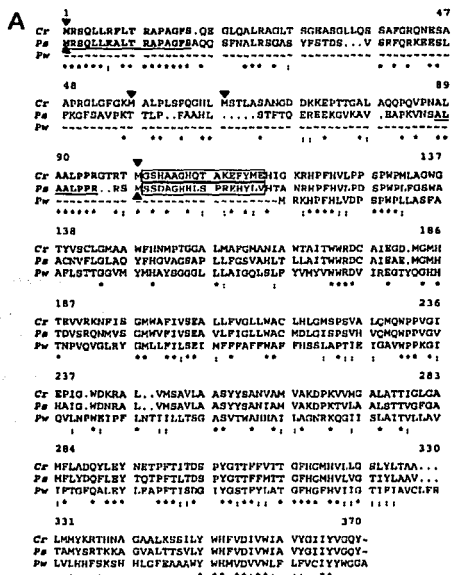


Fig. 3. Sequence alignment and phylogenetic analysis of cytochrome *c* oxidase subunit III sequences. **A**, sequence alignment of COX III from *Polytomella* spp. (Ps), *C. reinhardtii* (Cr), and *P. wickerhamii* (Pw) (62). Numbering is referred to the *Polytomella* spp. sequence. Black triangles indicate methionine residues present in the putative MTS sequences. The boxed regions indicate the N-terminal sequence of COX III from *Polytomella* spp. determined by Edman degradation and its homologous region in *C. reinhardtii*. The underlined sequences shown in bold are highly conserved amino acids present before the N-terminal sequence. Asterisks denote identical residues.

TABLE I

Codon usage of nuclear and mitochondrial genes of *Polytomella* spp.

Values are shown as percentages. Conspicuous differences in the codon usage are indicated in bold characters and gray boxes. Nuclear gene sequences used to construct this table were the TabB1 gene encoding β -tubulin from *P. agilis* (now *P. parva*) (14), partial sequence of the gene *Cytc1* encoding cytochrome *c*₁ (18), and partial sequence of *Atpa*, encoding subunit α of ATP synthase from *Polytomella* spp. (Xiao, Antaramian, and González-Halphen, unpublished results). Mitochondrial gene sequences from *Polytomella* spp. were *cox1* (15), *cob1* (36), and *nad4* (Funes, Antaramian, and González-Halphen, unpublished results).

Res	Codon	Mito	Nuc	cox3	Res	Codon	Mito	Nuc	cox3
Gly	GGG	8	0	0	Trp	TGG	100	100	100
Gly	GGA	34	2	5	End	TGA	0	0	0
Gly	GGT	46	75	65	Cys	TGT	43	0	0
Gly	GGC	12	23	30	Cys	TCC	57	100	100
Glu	GAG	47	95	100	End	TAG	33	0	0
Glu	AAA	53	5	0	End	TAA	67	100	100
Asp	GAT	64	64	30	Tyr	TAT	45	17	0
Asp	GAC	36	36	70	Tyr	TAC	55	83	100
Val	GTG	19	4	9	Leu	TYG	24	7	12
Val	ATA	34	0	0	Leu	TTA	28	1	0
Val	GTT	35	37	35	Phe	TTT	50	8	10
Val	GTC	12	59	56	Phe	TTC	50	92	90
Ala	GCG	5	0	0	Ser	TCG	4	0	0
Ala	GCA	34	1	2	Ser	TCA	6	2	4
Ala	GCT	66	36	51	Ser	TCT	21	32	43
Ala	GCC	15	63	47	Ser	TCC	8	49	39
Arg	AGG	4	9	18	Arg	CGG	4	0	0
Arg	AGA	3	0	0	Arg	CGA	21	0	0
Ser	AGT	29	1	0	Arg	CGT	43	62	46
Ser	AGC	32	15	14	Arg	CGC	25	29	36
Lys	AAG	32	98	90	Gln	CAG	41	96	100
Lys	AAA	68	2	10	Gln	CAA	59	4	0
Asn	AAT	57	11	0	His	CAT	64	24	27
Asn	AAC	43	89	100	His	CAC	36	76	73
Met	ATG	100	100	100	Leu	CTG	6	1	6
Ile	ATN	25	2	0	Leu	CFA	12	0	0
Ile	ATT	58	37	50	Leu	CTT	25	44	44
Ile	ATC	17	61	50	Leu	CTC	5	47	38
Thr	ACG	3	0	0	Pro	CCA	2	0	0
Thr	ACA	33	0	0	Pro	CCA	59	0	6
Thr	ACT	50	36	44	Pro	CCT	17	38	33
Thr	ACC	14	64	56	Pro	CCC	22	62	61

mesohydrophobicity. This strengthens the observation that mitochondrial imported proteins have diminished physical constraints ($\langle H \rangle$ and $\langle \text{mesoH} \rangle$) when compared with polypeptides encoded by mitochondrial genes. The figure presents the results using the scale PRIFT (32), but similar results were obtained with the scale GES and with other scales based on physicochemical amino acid properties OMH or KD (39) (results not shown). It is noteworthy that all COX III proteins that are encoded in the mitochondrial genome have higher hydrophobicity values and are grouped in the upper right corner of the graph.

Hydropathy profile analysis, carried out with different scales (32), predicted seven transmembrane stretches for COX III polypeptides from both algae (Fig. 6B). This suggests a structure of these polypeptides similar to the ones determined by

Two dots indicate similar residues. Sequence comparisons made without considering the putative presequences showed 66.5% identity and 73.9% similarity between *Polytomella* spp. and *C. reinhardtii*; 33.6% identity and 42.7% similarity between *Polytomella* spp. and *P. wickerhamii*; and 35.6% identity and 46.6% similarity between *C. reinhardtii* and *P. wickerhamii*. **B**, phylogenetic analysis of cytochrome *c* oxidase subunit III sequence. To construct the cladogram, the amino acid sequences of cytochrome *c* oxidase subunit III were compared among different organisms and the sequences obtained in this study, corresponding only to the mature proteins.

TESIS CON
FALLA DE ORIGEN

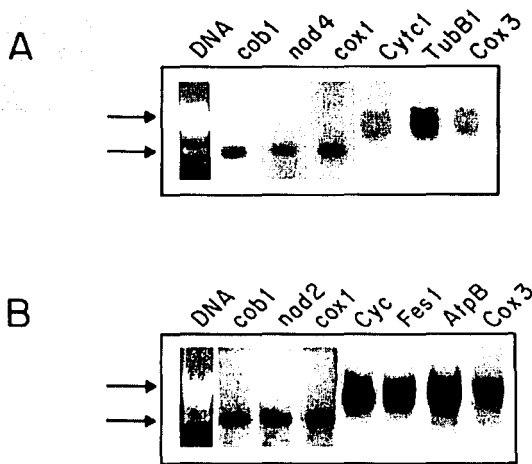


FIG. 4. The gene *cox3* is nuclear-localized in *Polytomella* spp. and in *C. reinhardtii*. A, 30 μ g of total DNA from *Polytomella* spp. was run in a 0.7% agarose gel. The gel transferred to a nylon membrane and hybridized with different nuclear and mitochondrial probes described in the text. Arrows indicate the positions of nuclear DNA and mtDNA. B, 30 μ g of total DNA from *C. reinhardtii* were run in a 0.7% agarose gel. Hybridization analysis was carried out with different nuclear and mitochondrial probes as indicated (see text).

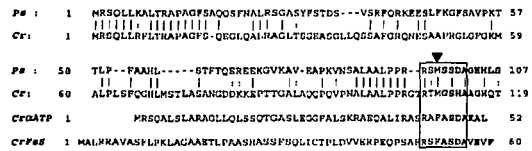


FIG. 5. Comparison of presequences in nuclear encoded mitochondrial proteins in *Polytomella* spp. and *C. reinhardtii*. Identical and similar residues are denoted with straight vertical lines and double dots, respectively. Only the sequences from *Polytomella* spp. (Ps) and *C. reinhardtii* (Cr) are compared. These sequences exhibited 45.4% identity and 50.5% similarity. The arrow indicates the site where presequences are cleaved by the mitochondrial processing protease. The N-terminal sequences of the mature proteins are shown in bold characters. The data shown for the α subunit of ATP synthase (CrFATP) and the Rieske iron sulfur protein (CrFes) of *C. reinhardtii* were taken from Nurani and Franzén (63) and from Atteia and Franzén (38), respectively. The box indicates conserved residues before and after the cleavage site according to the consensus sequence R(A/S/T)(M/F)↓(A/S/G/S)(D/H/A) (45).

x-ray crystallography for the bacterial and mammalian COX III subunits (2, 3). As an initial approach to gain insights on its topological arrangement, Ps-COX III was modelled over the three-dimensional structure of the bovine COX III (3). The predicted structure of Ps-COX III shows an overall topology similar to the bovine COX III but exhibiting shorter or incomplete transmembrane stretches as compared with the bovine counterpart (results not shown). Altogether, these observations suggest that the cytoplasmic-synthesized COX III polypeptides from chlamydomonad algae are imported into mitochondria and assembled in the inner mitochondrial membrane, with a topology similar but not identical to that of its mitochondrial-synthesized counterparts in other organisms.

DISCUSSION

Subunit III Is a Bona Fide Constituent of Cytochrome *c* Oxidase from *Polytomella* spp.—Cytochrome *c* oxidase from *C.*

reinhardtii has been purified and partially characterized (34, 40). In those works, Ps-COX III was not ascertained. In our hands, Ps-COX III was present in the intact cytochrome *c* oxidase of *Polytomella* spp. and was shown to be a bona fide constituent of this complex by immunochemical analysis. Therefore, we suggest that this subunit must exist in the mitochondrial complexes of algae of the family Chlamydomonadaceae. Moreover, because the corresponding gene is absent in the mitochondrial genomes of these algae (Refs. 5–9 and this work), it is likely that it was transferred to the nucleus early in evolution but previous to speciation.

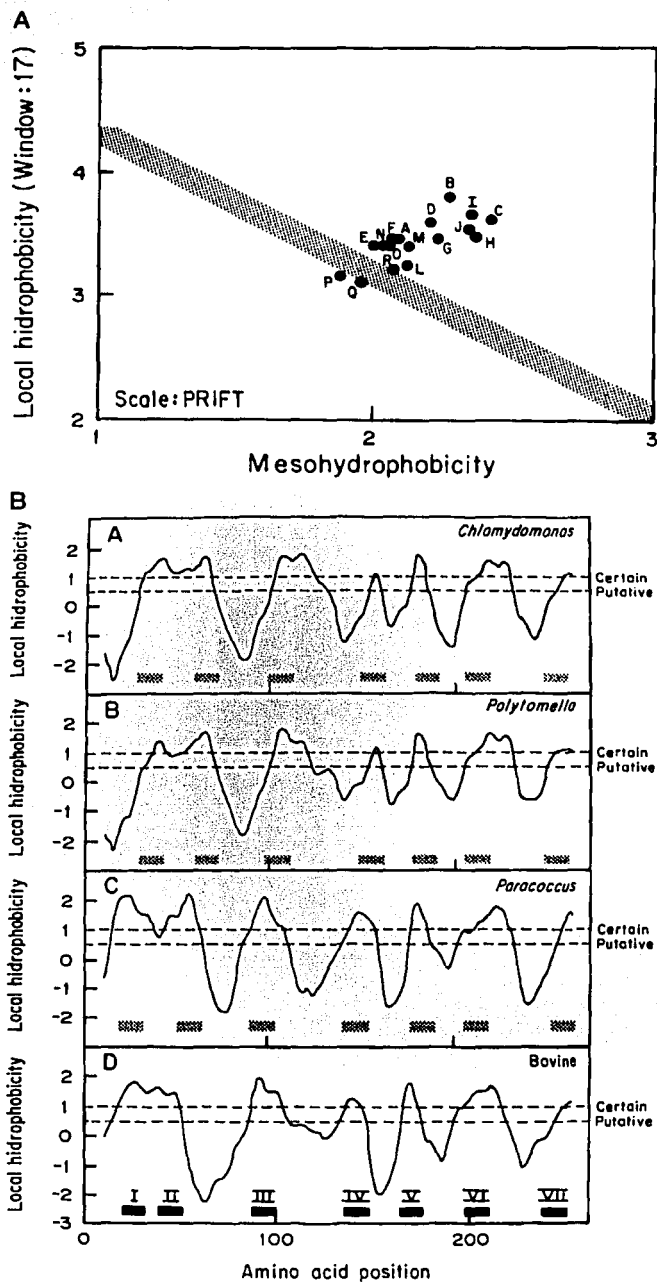
COX III Is Nuclear-encoded in the Algae of the Family Chlamydomonadaceae—This work also describes the cloning and complete sequencing of two new members of the *cox3* gene family from two chlamydomonad algae. Up until now the genes that encode COX III have been found only in mitochondrial genomes. The gene *cox3* is found even in the most reduced mitochondrial genome known to date, that of *Plasmodium falciparum* (41). The existence of a nuclear-encoded *cox3* gene was proposed for the lycopod *Selaginella*, because it was not present in the mitochondrial genome (42). However there is no evidence for its presence in the nuclear genome. Here we show that the *cox3* gene is nuclear-localized in the algae of the family Chlamydomonadaceae, as shown by Southern blot hybridization (Fig. 4), the presence of a biased codon usage typical of nuclear-localized genes in chlamydomonad algae (Table I), the presence of a polyadenylation signal TGTA usually found in the nuclear-localized genes of these algae, the existence of a sequence encoding a putative bipartite MTS (Fig. 5), a diminished <H> and mesoH of the predicted protein product (Fig. 6A), and the presence of introns in the corresponding *cox3* genomic sequences.³ The *cox3* gene is expressed as demonstrated by Northern blot hybridization (data not shown). In addition, the corresponding subunit is present in the mature and isolated cytochrome *c* oxidase complex from *Polytomella* spp., as shown by N-terminal sequencing and immunochemical analysis. To our knowledge, this is the first report of a nuclear-localized and active *cox3* gene. Some portions of the *cox3* gene of *C. reinhardtii* described in this work are similar to three cDNA sequences (AV386752, AV391757, and AV393074) recently deposited in the expressed sequence tags data base (43). Other organisms that lack the *cox3* gene in their mitochondrial genomes are the chlorophyte alga *Pedinomonas minor* and the ciliates *Paramecium aurelia* and *Tetrahymena pyriformis* (44). It is possible that these organisms may have also transferred their *cox3* genes to the nucleus control.

Nucleotide sequences encoding putative MTS were identified in the *cox3* genes of *Polytomella* spp. and *C. reinhardtii*. The MTS of Ps-COX III and Cr-COX III show some similarities with the mitochondrial targeting sequence of the Rieske iron-sulfur protein from *C. reinhardtii* (Fig. 5). The MTS sequences from chlamydomonad algae are rich in alanines, prolines, and charged amino acids. These sequences predict an amphiphilic α -helix structure in the N-terminal region. In addition, they share a similar site for cleavage for the mitochondrial processing peptidase, which seems to recognize the consensus sequence R(A/S/T)(M/F)↓(A/S/G/S)(D/H/A) (45).

Characteristics of Gene Transfer from the Mitochondria to the Nucleus in the Algae of the Family Chlamydomonadaceae—The theory of the origin of mitochondria proposes that there was a gradual transfer of genes from the original bacterial endosymbiont to the nucleus (46). This transfer is an ongoing process, as exemplified by the presence of genes encoded in

³ X. Pérez-Martínez, S. Funes, E. Davidson, M. P. King, and D. González-Halphen, unpublished observations.

FIG. 6. Mesohydrophobicity and hydrophobicity plots of cytochrome c oxidase subunit III from different organisms. A, mesohydrophobicity versus maximal local hydrophobicity plot for cytochrome c oxidase subunit III from different organisms using the PRIFT scale. Proteins are distributed on the abscissa according to their maximum hydrophobicity value and on the ordinate according to the hydrophobicity of the most hydrophobic segment. The boundary was calculated as in Claros *et al.* (19). The GenBank™ accession numbers of the nuclear COX III sequences used to construct this graph were: A, *Aegilops columnaris* (U46765); B, *Candida parapsilosis* (X75679); C, *Chondrus crispus* (P48872); D, *Helianthus annuus* (X57669); E, *Magnolia grandiflora* (Z68127); F, *Zea mays* (X53055); G, *Oenothera berteriana* (X04764); H, *Pilayella littoralis* (Z37967); I, *P. wickerhamii* (Q37620); J, *Gracilaria lemaneiformis* (AF118119); K, *Oryza sativa* (X17040); L, *Schizosaccharomyces pombe* (X16868); M, *Glycine max*, soybean (X15131); N, *Vicia faba* (X51690); O, *Triticum aestivum* (X15944); P, *C. reinhardtii* (AF233515, this work); Q, *Polytomella* spp. (AF233514, this work); and R, *Homo sapiens*, human (P00414). B, hydrophobicity plots comparing the deduced COX III sequences of *Polytomella* spp. (B) and *C. reinhardtii* (A) with the ones of the bovine enzyme (D) and *P. denitrificans* (C) are shown. Black boxes with roman numerals indicate the positions of certain transmembrane stretches based on the crystallographic structure of bovine COX III. Gray boxes indicate calculated transmembrane stretches.



TESIS CON
FALLA DE ORIGEN

both the mitochondrial and the nuclear genomes, *i.e.* ATP synthase subunit 9 of *Neurospora crassa* (47), and COX II of some leguminosae (48). In several species, the process of moving mitochondrial genes to the nucleus may have a selective advantage, because nuclear genes exhibit a lower mutation rate and the nucleus seems to have a more sophisticated DNA repair system than mitochondria (49). Gene transfer from organelles to the nucleus is also thought to increase its rate of recombination and reduce accumulation of deleterious mutations (50).

The *cox3* gene transferred from the mitochondria to the nucleus in the chlamydomonad algae satisfies many of the criteria necessary for a gene that has been translocated from the mitochondrial to the nuclear genome, as proposed by Brennicke *et al.* (42) and by Claros *et al.* (19). It acquired a presequence for targeting into mitochondria, changed its codon usage, acquired a polyadenylation signal, and diminished the $\langle H \rangle$ and *mesoH* of its protein product. In addition, the corresponding mitochondrial copy that presumably existed has completely disappeared, suggesting that this transfer occurred early in evolution.

The high sequence similarity found between a region of the putative MTS encoded by the *cox3* genes from *C. reinhardtii* and *Polytomella* spp. (Fig. 3A) suggests that the transfer of this gene from the mitochondria to the nucleus and the corresponding acquisition of the presequence, occurred before the *Polytomella* colorless genus diverged from the main *Chlamydomonas* photosynthetic lineage. Otherwise no conservation of the presequences would be expected. The drastic change in codon usage, which is more remarkable in these algae because of its highly biased nuclear codon usage (14), also suggests that the transfer of the *cox3* gene in these organisms occurred early in evolution, when there was a massive transfer of genes from the protomitochondrion to the nucleus (46). This process might have occurred before the Post Cretaceous era (65 or more million years ago), when the nonphotosynthetic algae are thought to have derived from the green lineages (11). The phylogenetic analysis carried out with the predicted COX III sequences (Fig. 3B) shows similar results to those obtained with classical mitochondrial proteins like COX I (15), or cytochrome *b* (36). Sequences from the algae of the genera *Polytomella* and *Chlamydomonas* tend to strongly affiliate in these phylogenetic analyses.

Importability of Nuclear-encoded Subunits into Mitochondria: Subunit III Exhibits Diminished Local $\langle H \rangle$ and Diminished *mesoH*—Why have some genes remained in the mitochondrial DNA? One explanation has been the variation of the genetic code in mitochondria, where the triplet UGA encodes tryptophan instead of a polypeptide chain termination signal (51). Another explanation suggests that organelle genomes have persisted by encoding structural proteins that maintain redox balance within the bioenergetic membranes (52). Alternatively, it has been proposed that the genes that remained localized in the mitochondrial genome are those that encode highly hydrophobic polytopic proteins, containing two or more helices that span the membrane (53). This is because the presence of a larger number of hydrophobic segments in a polypeptide could impair its import into mitochondria (54) or cause mistargeting to the endoplasmic reticulum (53). Moreover, the synthesis of hydrophobic polypeptides inside the mitochondria may ensure their proper insertion in the inner membrane and the correct topological arrangement required for vectorial proton translocation. Two classic examples are the cytochrome *b* gene (*cob1*), which encodes an 8-transmembrane-stretch polypeptide (55), and the cytochrome *c* oxidase subunit I (*cox1*), which encodes a protein with 12 membrane-associated helices

(2, 3). Both genes are present in all mitochondrial genomes characterized to date. Other genes that encode highly hydrophobic polypeptides are also present in the majority of mtDNAs (56), *i.e.* *atp6* (encoding 5 transmembrane helices), *atp8* (encoding 2 transmembrane helices), *nad1* (encoding 8–9 transmembrane helices), *nad2* (encoding 13–14 transmembrane helices), *nad3* (encoding 3 transmembrane helices), *nad4* (encoding 13–14 transmembrane helices), *nad4L* (encoding 3 transmembrane helices), *nad5* (encoding 15–16 transmembrane helices), *nad6* (encoding 5 transmembrane helices), *cox2* (encoding 2 transmembrane helices), and *cox3* (encoding 7 transmembrane helices).

In yeast, *in vivo* studies with cytoplasmically synthesized constructs of variable lengths of apocytochrome *b*, showed that in mitochondria, the import of polypeptides with more than three or four transmembrane helices is strongly hindered (19). Analysis of sequences from nuclear-encoded and mitochondrial-encoded mitochondrial proteins suggested that low values of *mesoH* and $\langle H \rangle$ are more useful indicators than the number of transmembrane regions in determining whether a protein could be imported into the mitochondrion. It is known that mitochondria readily import proteins with several transmembrane stretches, for example the adenine nucleotide translocator, if they possess low $\langle H \rangle$ and low *mesoH* (19). However, the import pathway of the translocator differs greatly from the "conservative intramitochondrial sorting pathway," in which polypeptides are transferred to the mitochondrial matrix space and then sorted to its final membrane destination (57). We hypothesize that the latter may be the mechanism for the biogenesis of the COX III proteins described in this work.

Transfer of genes from organelles to the nucleus involves several steps (42): (i) the export of the nucleic acid molecule as DNA or RNA (48), (ii) integration into the nucleus by nonhomologous recombination (58) or by a common end-joining mechanism (59), (iii) acquisition of a presequence by duplication of existing targeting signals (60), (iv) acquisition of a promoter, a ribosome binding site, and a polyadenylation signal (49), (v) change in codon usage (49), (vi) modification of the nucleotide sequence to encode for a polypeptide with diminished local hydrophobicity ($\langle H \rangle$) and diminished *mesoH*, which may allow the import of the protein products into mitochondria (19), (vii) inactivation of the mitochondrial gene copy, and (viii) stepwise loss of the mitochondrial gene (61).

Our data support the hypothesis that the genes that encode proteins with high $\langle H \rangle$ and high *mesoH* have remained in the mitochondrial genome, whereas those genes that encode proteins with low values of $\langle H \rangle$ and *mesoH* have been exported to the nucleus, and their protein products imported back into mitochondria (19). The strategy used by the algae of the family Chlamydomonadaceae seems to involve the acquisition of a large and possibly bipartite MTS and a lowering of $\langle H \rangle$ and *mesoH* in the COX III polypeptides, which are requirements for the proper insertion of the protein into the mitochondrial inner membrane. We hypothesize that the limiting step in gene transfer from organelles to the nucleus has not been the differences in genetic code but hindrances to the import into the mitochondrial inner membrane polytopic proteins whose membrane topology is a critical requirement for its catalytic activity (vectorial proton pumping).

Hydropathy analysis of Cr-COX III and Ps-COX III showed the presence of seven putative transmembrane stretches. The hydrophobicity of these seven helices seems to be lower in the chlamydomonad algae when compared with the *P. denitrificans* or the bovine subunits (Fig. 6B). This is more evident in the three-dimensional model for Ps-COX III built upon the crystallographic coordinates of its bovine counterpart (3). In our model

TESIS CON
FALLA DE ORIGEN

(data not shown), shorter transmembrane stretches are observed as well as interruptions in the middle section of the membrane helices. In addition, helix VII of the Ps-COX III protein is only half the size of the corresponding helix in the bovine polypeptide and may not span the membrane bilayer. The helices that are in contact with COX I (helices I and III) do not exhibit structural modifications, suggesting that the diminished hydrophobicity of COX III is stronger in those regions of the protein that seem not to be involved in subunit-subunit interactions.

Acknowledgments—We thank Dr. J. d'Alayer (Institut Pasteur) for expert help in sequencing peptides. Drs. A. Antaramian and R. Coria (Instituto de Fisiología Celular (IFC), Universidad Nacional Autónoma de México (UNAM México) for participation in the initial stages of this project. Dr. L.-G. Franzén (Göteborg University) for the kind gift of the cDNA library of *C. reinhardtii*, Drs. L. Ongay, M. Sosa, and G. Codiz (Unidad de Biología Molecular, IFC, UNAM) for the synthesis of various oligonucleotides, Dr. L. J. Prochaska (Wright State University) for carrying out immunoblot analysis with the battery of anti-COX III antibodies, J. Ramirez for helping us in the construction of the Ps-COX III crystallographic model, Dr. Y. Z. Zhang (Molecular Probes) for calling our attention to the monoclonal anti-COX III yeast antibodies, and several colleagues and researchers who kindly provided us with DNA probes that were used in this work: Drs. T. D. Fox (Cornell University), L.-G. Franzén (Göteborg University), A. García-Horsman (University of Illinois), R. B. Gennis (University of Illinois), M. Goldschmidt-Clermont (University of Geneva), E. H. Harris (Duke University), G. Aurrignat (Instituto de Biocología, UNAM), L. J. Prochaska (Wright State University), B. Schäfer (Rheinisch-Westfälische Technische Hochschule, Aachen), and C. D. Sillflow (University of Minnesota). We are also indebted to Drs. A. Atteia, F. Bastarrachea, M. Gavilanes, and A. Gómez-Puyou (UNAM) and D. W. Krogmann (Purdue University) for helpful discussions and critical review of the manuscript.

REFERENCES

- Capaldi, R. A. (1990) *Annu. Rev. Biochem.* **59**, 569-596
- Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. (1995) *Nature* **376**, 660-669
- Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. (1996) *Science* **272**, 1136-1144
- Grossman, L. I., and Lomax, M. I. (1997) *Biochim. Biophys. Acta* **1352**, 174-192
- Michaelis, G., Vahrenholz, C., and Pratz, E. (1990) *Mol. Gen. Genet.* **223**, 211-216
- Collenx, L., Michel-Wulvertz, M.-R., Matagne, R. F., and Dujon, B. (1990) *Mol. Gen. Genet.* **223**, 288-296
- Denovan-Wright, E. M., Nedelcu, A. M., and Lee, R. W. (1998) *Plant Mol. Biol.* **36**, 285-295
- Turmel, M., Mercier, J.-P., and Côté, M.-J. (1993) *Nucleic Acids Res.* **21**, 5242-5250
- Kroymann, J., and Zetsche, K. (1998) *J. Mol. Evol.* **47**, 431-440
- Melkonian, N. (1990) in *Handbook of Protocista* (Margulis, L., Corliss, J. O., Melkonian, M., and Chapman, D. J., eds) pp. 608-616, Johns and Bartlett, Boston, MA
- Round, F. E. (1980) *Bioessays* **12**, 61-69
- Mattox, K. R., and Stewart, K. D. (1984) in *Systematics of the Green Algae* (Irvine, D. E. G., and John, D. M., eds) pp. 29-72, Academic Press, London
- Melkonian, M., and Surek, B. (1995) *Bull. Soc. Zool. Fr.* **120**, 191-208
- Conner T. W., Thompson M. D., and Sillflow C. (1989) *Gene (Amst.)* **84**, 345-358
- Antaramian, A., Coria, R., Ramirez, J., and González-Halphen, D. (1996) *Biochim. Biophys. Acta* **1273**, 198-202
- Nedelcu, A., Spence, D. F., Denovan-Wright, E. M., and Lee, R. W. (1996) *J. Phycol.* **32**, 103-111
- Atteia, A., Dreyfus, G., and González-Halphen, D. (1997) *Biochim. Biophys. Acta* **1320**, 275-284
- Gutiérrez-Cirlos, E. B., Antaramian, A., Vázquez-Acevedo, M., Coria, R., and González-Halphen, D. (1994) *J. Biol. Chem.* **269**, 9147-9154
- Claros, M. C., Peren, J., Shu, Y., Samatay, F. A., Popot, J. L., and Jacq, C. (1995) *Eur. J. Biochem.* **228**, 762-771
- Capaldi, R. A., and Hayashi, H. (1972) *FEBS Lett.* **20**, 261-263
- Horie, S., and Morrison, M. (1963) *J. Biol. Chem.* **238**, 1855-1860
- González-Halphen, D., Vázquez-Acevedo, M., and García-Ponce, B. (1991) *J. Biol. Chem.* **266**, 3870-3876
- Schagger, H., Link, T. A., Engel, W. D., and von Jagow, G. (1986) *Methods Enzymol.* **120**, 224-237
- Capaldi, R. A., González-Halphen, D., Zhang, Y. Z., and Yanamura, W. (1988) *J. Bioenerg. Biomembr.* **20**, 291-311
- González-Halphen, D., Lindorfer, M. A., and Capaldi, R. A. (1988) *Biochemistry* **27**, 7021-7031
- Markwell, M. A. K., Haas, S. M., Bieber, L. L., and Tolbert, N. E. (1978) *Anal. Biochem.* **87**, 206-210
- Fruhman, M. A. (1993) *Methods Enzymol.* **218**, 340-356
- Franzén, L.-G., and Falk, G. (1992) *Plant Mol. Biol.* **19**, 771-780
- Desvergne, J., Haeblerli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res.* **24**, 4876-4882
- Claros, M. G., Brunak, S., and von Heijne, G. (1997) *Curr. Opin. Struct. Biol.* **7**, 394-398
- Claros, M. G., and von Heijne, G. (1994) *Comput. Appl. Biosci.* **10**, 685-686
- Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714-2723
- Atteia, A. (1994) *Contribution à l'Étude des Protéines de la Chaîne Respiratoire Mitochondriale Chez l'Algue Verte Chlamydomonas reinhardtii* Ph.D. thesis, University of Paris VII, Paris, France
- Sillflow, C. (1998) *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas* (Rocheix, J.-D., Goldschmidt-Clermont, M., and Merchant, S., eds) pp. 25-40, Kluwer Academic Publishers, Dordrecht, The Netherlands
- Antaramian A., Funes S., Vázquez-Acevedo M., Atteia A., Coria R., and González-Halphen, D. (1998) *Arch. Biochem. Biophys.* **354**, 206-214
- Amati, B. B., Goldschmidt-Clermont, M., Wallace, C. J., and Rocheix, J. D. (1988) *J. Mol. Evol.* **28**, 151-160
- Atteia, A., and Franzén, L.-G. (1996) *Eur. J. Biochem.* **237**, 792-799
- Crimi, M., and Degli Esposti, M. (1991) *Trends Biochem. Sci.* **16**, 119
- Bennoun, P., Atteia, A., Pierre, Y., and Delosme, M. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 10202-10208
- Pequin, J. E., Werner, E., Gardner, M. J., Williamson, D. H., and Wilson, R. J. M. (1992) *Nucleic Acids Res.* **20**, 879-887
- Bronnacke, A., Grohmann, L., Hiesel, R., Knoop, V., and Schuster, W. (1993) *FEBS Lett.* **325**, 140-145
- Asamizu, E., Nakamura, Y., Sato, S., Fukuzawa, H., and Tabata, S. (1999) *DNA Res.* **6**, 369-373
- Gruy, M. W., Lang, B. F., Cedergren, R., Golding, G. B., Lemieux, C., Sankoff, D., Turmel, M., Brussaard, N., Delage, E., Littlejohn, T. G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., and Burger, G. (1998) *Nucleic Acids Res.* **26**, 865-878
- Nurani, G. (1997) *Studies on the Structure and Biogenesis of Membrane Proteins from the Green Alga Chlamydomonas reinhardtii*. Ph.D. thesis, University of Stockholm, Stockholm, Sweden
- Gruy, M. W., Burger, G., and Lang, B. F. (1999) *Science* **283**, 1476-1481
- van den Boogaart, P., Samallo, J., and Agsteribbe, E. (1982) *Nature* **298**, 187-189
- Adams, K. L., Song, K., Roessler, P. G., Nugent, J. M., Doyle, J. L., and Palmer, J. D. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96**, 13863-13868
- Neckelmann, N., Li, K., Wade, R. P., Shuster, R., and Wallace, D. C. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7580-7584
- Martin, W., Steebe, B., Gorenzkin, V., Hansmann, S., Hasegawa, M., and Kowalik, K. V. (1998) *Nature* **393**, 162-165
- Fox, T. D. (1987) *Annu. Rev. Genet.* **21**, 67-91
- Roca, H. L., Herrmann, R. G., and Martin, W. (1999) *Trends Genet.* **15**, 364-370
- von Heijne, G. (1986) *FEBS Lett.* **198**, 1-4
- Popot, J. L., de Vitry, C., and Atteia, A. (1994) in *Membrane Protein Structure: Experimental Approaches*, pp. 41-96, Oxford University Press, London
- Xia, D., Yu, C. A., Kim, H., Xia, J. Z., Kachurin, A. M., Zhang, L., Yu, L., and Deisenhofer, J. (1997) *Science* **277**, 60-66
- Burger, G., Plante, I., Lonergan, K. M., and Gray, M. W. (1995) *J. Mol. Biol.* **246**, 522-537
- Voos, W., Martin, H., Krimmer, T., and Pfanner, N. (1999) *Biochim. Biophys. Acta* **1429**, 205-254
- Sun, C.-W., and Callia, J. (1993) *Plant Cell* **5**, 97-107
- Blanchard, J. L., and Schmidt, G. W. (1996) *Mol. Biol. Evol.* **13**, 537-548
- Kadowaki, K., Kubo, N., Ozawa, K., and Hirai, A. (1996) *EMBO J.* **15**, 6652-6661
- Grohmann, L., Brennick, A., and Schuster, W. (1992) *Nucleic Acids Res.* **20**, 5641-5646
- Wolff, G., Plante, I., Lang, B. F., Kück, U., and Burger, G. (1994) *J. Mol. Biol.* **237**, 75-86
- Nurani, G., and Franzén, L.-G. (1996) *Plant Mol. Biol.* **31**, 1105-1116

Apéndice IV.

Artículo publicado:

Structure of nuclear-localized *cox3* genes in *Chlamydomonas reinhardtii* and in its colorless close relative *Polytomella* sp.

Xochitl Pérez-Martínez · Soledad Funes
Elena Tolkunova · Edgar Davidson · Michael P. King
Diego González-Halphen

Structure of nuclear-localized *cox3* genes in *Chlamydomonas reinhardtii* and in its colorless close relative *Polytomella* sp.

Received: 17 November 2001 / Accepted: 21 December 2001 / Published online: 13 February 2002
© Springer-Verlag 2002

Abstract Several chlorophyte algae do not have the *cox3* gene, encoding subunit III of cytochrome *c* oxidase, in their mitochondrial genomes. The *cox3* gene is nuclear-encoded in the photosynthetic alga *Chlamydomonas reinhardtii* and in the colorless alga *Polytomella* sp. In this work, the genomic sequences of the *cox3* genes of these two closely related algae are reported. The *cox3* genes of both *C. reinhardtii* and *Polytomella* sp. contain four introns in the region encoding the putative mitochondrial-targeting sequences. These four introns show low sequence identities, but their locations are conserved between these species. The *cox3* gene of *C. reinhardtii* has five additional introns in the region encoding the mature subunit III of cytochrome *c* oxidase. Sequence analysis of intron 6 of the *cox3* gene of *C. reinhardtii* revealed similarity with two sequence elements present in introns of several other nuclear genes from this green alga. In the majority of the genes, these conserved sequences are located either near the 3' end or near the 5' end of the introns. Based on these data, we propose that the colorless genus *Polytomella* separated from *C. reinhardtii* after the *cox3* gene was transferred to the nucleus. The data also support the evolutionary hypothesis of a recent acquisition of introns in *C. reinhardtii*.

Keywords *Chlamydomonas reinhardtii* · Cytochrome *c* oxidase · Mitochondrial-targeting signals · *Polytomella* sp.

Communicated by B.B. Sears

X. Pérez-Martínez · S. Funes · D. González-Halphen (✉)
Departamento de Genética Molecular,
Instituto de Fisiología Celular,
Universidad Nacional Autónoma de México,
Apartado Postal 70-243, México 04510, D.F., Mexico
E-mail: dhalphen@ifisiol.unam.mx

E. Tolkunova · E. Davidson · M.P. King
Department of Biochemistry and Molecular Pharmacology,
Thomas Jefferson University, 233 South 10th Street,
BLSB 308, Philadelphia, PA 19107, USA

Introduction

Members of the genus *Chlamydomonas* form a monophyletic group, but are distributed in at least six distinct lineages (Buchheim et al. 1996; Friedl 1997). The green alga *C. reinhardtii* – a model of choice in the molecular studies of photosynthesis and flagellar structure – has been grouped with members of the genus *Volvox* in several phylogenetic analyses (Buchheim et al. 1996; Nakayama et al. 1996; Rumpf et al. 1996; Friedl 1997). The colorless chlorophytes of the genus *Polytomella* are also close relatives of *C. reinhardtii*; complete 18S rDNA sequence analysis suggests that *C. reinhardtii*, *Polytomella parva*, *Polytomella* sp. and *Volvox carterii* are members of a single monophyletic clade, the so-called *Volvox* clade (Melkonian and Surek 1995; Nakayama et al. 1996). The genus *Polytomella* is believed to have evolved from a *Chlamydomonas*-like ancestor by loss of its cell wall and functional chloroplasts (Round 1980; Melkonian and Surek 1995). The close phylogenetic relationship between members of the genus *Polytomella* and *C. reinhardtii* is also supported by morphological and structural similarities (Mattox and Stewart 1984), the similarity of its nuclear encoded β -tubulin genes (Conner et al. 1989), its mitochondrial *cox1* genes (Antaramian et al. 1996), and the biochemical similarities in the α and β subunits of their mitochondrial F_1F_0 -ATPases (Atteia et al. 1997).

The mitochondrial genomes of several chlorophyte algae lack some of the genes that are commonly found in the mitochondrial genome of the majority of eukaryotes. The missing genes include *cox2*, *cox3*, *atp6*, *atp8*, *atp9*, *nad3* and *nad4L* in *C. reinhardtii* (Michaelis et al. 1990), *C. eugametos* (Denovan-Wright et al. 1998), and *Chlorogonium elongatum* (Kroymann and Zetsche 1998). In the green alga *Chlamydomonas reinhardtii* and in the colorless alga *Polytomella* sp., the *cox2* and *cox3* genes encoding subunits II and III of cytochrome *c* oxidase have been transferred from the mitochondrial genome to the nucleus (Pérez-Martínez et al. 2000, 2001). The

TESIS CON
FALLA DE ORIGEN

nuclear-localized *cox2* and *cox3* genes reflect the characteristics of genes transferred from the mitochondrion to the nucleus proposed by Brennicke et al. (1993) and Claros et al. (1995): (1) they acquired a nucleotide sequence encoding a putative mitochondrial-targeting sequence (MTS), (ii) they exhibit a nuclear codon-usage pattern, (3) they acquired polyadenylation signals typical of chlamydomonad nuclear genes that differ from the consensus in other eukaryotes, and (4) their deduced protein products exhibit an overall diminished mesohydrophobicity, when compared with their counterparts encoded in the mitochondrial genome.

In this work, we describe the cloning and characterization of the genomic *cox3* genes from *C. reinhardtii* and *Polytomella* sp. This is the first description of *cox3* genes that are localized in the cell nucleus. The data suggest that transfer of the *cox3* gene to the nucleus occurred before the divergence of *Polytomella* sp. and *C. reinhardtii*. They also support the hypothesis of a late acquisition of introns in *C. reinhardtii*.

Materials and methods

Strains and culture conditions

Polytomella sp. (198.80, E.G. Pringsheim), from the Sammlung von Algenkulturen (Göttingen, Germany), was grown as described by Gutiérrez-Cirlos et al. (1994). *C. reinhardtii* (wild type, CC125 mt⁺) was obtained from the *Chlamydomonas* Genetic Center (Duke University) and was grown as described by Rochaix et al. (1988).

Cloning and sequencing of genomic *cox3* from *Polytomella* sp. and *C. reinhardtii*

Total DNA was obtained from *Polytomella* sp. and *C. reinhardtii* as described by Pérez-Martínez et al. (2000). All other molecular biology techniques were standard (Sambrook et al. 1989). Nucleotide sequencing was performed by the Kimmel Cancer Center DNA Sequencing Facility, Thomas Jefferson University.

The genomic sequence of the *cox3* gene of *Polytomella* sp. was obtained by PCR amplification with Hot Start Taq polymerase (Qiagen), using primers designed to anneal to the 5' and 3' non-coding regions of the *cox3* cDNA. The forward primer was 5'-CGT TTT TGG TCA AGT TGA AA-3' and the reverse primer was 5'-CGC ATA ACG CGA AGT CAC TAC -3'. Total DNA was denatured for 5 min at 94 °C and subjected to 30 cycles of: 45 s denaturation at 94 °C, 1 min annealing at 50 °C, and 2.5 min extension at 72 °C. PCR products were cloned into the pGEM-T easy vector system (Promega) and sequenced.

C. reinhardtii *cox3* genomic sequence was obtained by amplification of three overlapping PCR products, using Hot Start Taq polymerase (Qiagen). The sets of primers used were: forward 1: 5'-AGC GCG ACC GGT GAA ACC AG-3', reverse 1: 5'-TGG AAG GGG TGG CGC TTG CCG-3', forward 2: 5'-CCA AGG AGT TCT ACA TGG AGC AC-3', reverse 2: 5'-CCT TGG CCA CCA TGG CCA CGT TGG-3', forward 3: 5'-GTG GCG CTG CAG ATG CAG TGG C-3', reverse 3: 5'-CTG CCA CAC ACA CCC GTC ATA CG-3'. Total DNA was denatured for 5 min at 94 °C, and subjected to 30 cycles of: 1 min denaturation at 94 °C, 1 min annealing at 60 °C, and 2.5 min extension at 72 °C. PCR products were cloned into the pGEM-T easy vector system from Promega.

The genomic sequences of the *cox3* genes of *Polytomella* sp. and *C. reinhardtii* are not shown, but are available in the DDBJ/EMBL/

GenBank data banks under accession numbers AF286057 and AF286058, respectively.

Results and discussion

The *cox3* genomic sequences of chlamydomonad algae contain several introns

Based on the sequences of the *cox3* cDNA clones from *Polytomella* sp. and *C. reinhardtii* (Pérez-Martínez et al. 2000), deoxyoligonucleotide primers were designed and used for PCR amplification of the corresponding genomic coding regions. When total genomic DNA from *Polytomella* sp. was used as template for amplification, a 2.0-kb product was obtained. In contrast, when cDNA from the same alga was used as template, the predicted 1.2-kb product was obtained. When genomic DNA from *C. reinhardtii* was used as template to amplify the complete coding region with three different primer pairs, the genomic sequence was estimated to be 3.2 kb. The mRNA gave rise to three amplicons, reflecting a total size of 1.2 kb. The PCR fragments obtained from amplification using genomic DNA from the two algae were cloned and sequenced. A comparison of the sequences with the *cox3* cDNA sequences previously obtained (Pérez-Martínez et al. 2000) confirmed that they represented *cox3* genes and allowed identification of introns in the genomic sequence. Figure 1 shows the overall genomic organization of the two *cox3* genes.

The two chlamydomonad *cox3* genes have conserved intron positions in the regions encoding the putative MTS

The boundaries, phases and sizes for the introns of both *Polytomella* sp. and *C. reinhardtii* *cox3* genes are shown in Table 1. The *cox3* gene of *Polytomella* sp. has four introns (numbered 1-4) that range in size over 71-444 nucleotides. All are located in the region encoding the putative MTS that is thought to direct and insert subunit COX III into the mitochondrial inner membrane. This MTS is not present in the mature COX III protein of *Polytomella* sp. (Pérez-Martínez et al. 2000). In contrast, the genomic sequence of *cox3* of *C. reinhardtii* is interrupted by nine introns (numbered 1-9) ranging in size over 133-323 nucleotides. The first four introns, like those of *Polytomella* sp., are located in the region encoding the putative MTS of the *C. reinhardtii* COX III polypeptide. Introns 5-9 are distributed along the nucleotide sequence that encodes the mature COX III subunit.

An alignment of the MTSs encoded by both *cox3* genes indicates that the location and the phases of introns 1, 3, and 4, have been conserved between the two species (Fig. 2). Intron 2 also has a conserved phase, but its position is displaced by three nucleotides. This suggests that all four introns were acquired after the *cox3*

TESTIS CON
 FALLA DE ORIGEN

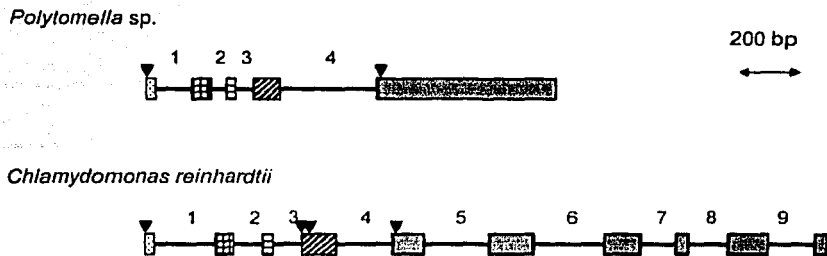


Fig. 1. Overall organization of exons and introns in *Polytomella* sp. and *Chlamydomonas reinhardtii* *cox3* genes. Rectangles show exons and lines represent introns. Gray rectangles show the region encoding the mature protein. Rectangles marked with dots, squares, horizontal lines and dashed lines represent homologous exons encoding the mitochondrial-targeting sequence (MTS) found in both algae. Black triangles indicate the relative positions of methionines in the encoded MTS and the first methionine in the encoded COX III subunit

gene was transferred to the nucleus in the common ancestor, but before *Polytomella* diverged from *C. reinhardtii*. An alternative scenario would imply intron insertion in preferential sites of *cox3* sequences. When other intron-containing mitochondrial *cox3* genes were explored, it was found they all exhibit introns that are localized in different positions. The bryophytes *Pellia epiphylla* (Malek et al. 1996) and *Marchantia polymorpha* (Oda et al. 1992) contain one and two introns, respectively, in the mitochondrial *cox3* gene. These introns are present in different positions from those found in the *Chlamydomonas* and *Polytomella* nuclear *cox3* genes. The plant *Lycopodium squarrosum* (Hiesel et al. 1994) and the yeast *Yarrowia lipolytica* (Matsuoka et al. 1994) *cox3* genes each contain a single, non-conserved intron. The position of these introns is also different from that found in the *Chlamydomonas* and *Polytomella* nuclear *cox3* genes. These comparisons indicate the absence of a single set of preferential intron insertion sites in *cox3* sequences.

Although the regions of the *cox3* genes encoding the MTSs of both chlamydomonad algae share a conserved location of introns, the nucleotide sequences of these

introns are poorly conserved (less than 38% identity). This is likely due to the rapid rate of nucleotide substitutions in introns of volvocine algae, which is ten-fold higher than in exons (Liss et al. 1997). Nevertheless, the presence of conserved intron locations in the highly conserved regions encoding the MTS in the *cox3* genes of *Polytomella* sp. and *C. reinhardtii* indicates that these algae acquired the region encoding the MTS and their respective introns before speciation. The presence of five additional introns in the *C. reinhardtii* *cox3* gene also suggests that these introns originated during the period that occurred after the divergence of these two species. Alternatively, but less likely, they were present in the common ancestor and then lost from the *cox3* gene of *Polytomella* sp. This last scenario would imply a selective loss of introns in regions encoding the *Polytomella* mature COX III subunit, without loss of introns in the region encoding the MTS.

Splicing junctions and internal conserved sequences in the *cox3* gene introns

All the introns in chlamydomonad *cox3* genes show orthodox splice sites, exhibiting GT at the 5' end and AG at the 3' end (Table 1). In *Chlamydomonas*, consensus-conserved sequences surrounding the splicing sites have been identified: (C/A)(A/C)G↓GTG(A/C)G for the 5' splice site and (G/A)CAG↓(G/A) for the 3' splice site (Silflow 1998). The splice junctions of the introns of the *cox3* gene of *C. reinhardtii* also conform to these consensus-conserved sequences. The sequences

Table 1. Phases, sizes, and flanking sequences of 5' and 3' splice sites for *Chlamydomonas reinhardtii* (Cr) and *Polytomella* sp. (Ps) *cox3* introns

Intron	Phase		5' splice site		Size (bp)		3' splice site	
	Ps	Cr	Ps	Cr	Ps	Cr	Ps	Cr
1	1	1	CTG/GTAAGA	CAG/GTGTGT	167	267	TGTAG/GCT	TGCAG/GCT
2	2	2	GAG/GTAAGA	CGG/GTGAGC	71	133	TTTAG/GAA	CGCAG/CCG
3	0	0	AAG/GTAAAT	AAG/GTAAGG	85	139	TATAG/ACT	CGCAG/ATG
4	2	2	CAG/GTAAGA	CCG/GTGAGG	444	256	TATAG/GTC	TGCAG/CAC
5	-	1	-	TGG/GTGAGC	-	290	-	CGCAG/GCA
6	-	2	-	GCT/GTGAGT	-	323	-	CGCAG/CTG
7	-	0	-	AAG/GTGGGG	-	159	-	CGCAG/GTG
8	-	2	-	CCA/GTGAGT	-	180	-	CGCAG/GTA
9	-	2	-	CTG/GTGGGT	-	204	-	CACAG/GCA

TESIS CON
FALLA DE ORIGEN

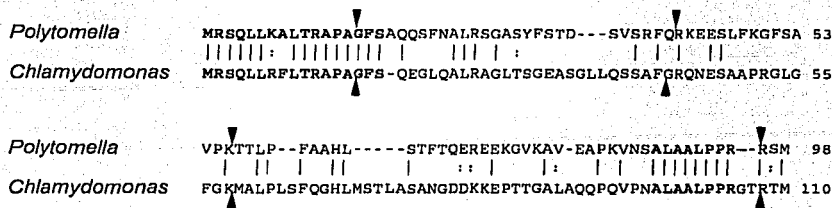


Fig. 2. Amino acid sequence alignment of MTS of *Polytomella* sp. and *C. reinhardtii* *cox3* and comparison of intron positions. Arrows indicate the positions of introns 1-4 in the corresponding gene of each alga. An arrow over an amino acid means that the intron is not in phase with the open reading frame (orf). Only intron 3 is in phase with the orf

surrounding splicing sites in the *cox3* gene of *Polytomella* sp. show some differences from the *C. reinhardtii* consensus splicing sites; and they conform more to the sites (A/T)G↓GTAA for the 5' splice site and TAG↓(G/A) for the 3' splice site. Additional genomic sequences of *Polytomella* sp. genes are needed to establish canonical splicing sites in this organism.

A general feature of *Chlamydomonas* genes is that introns are often small, ranging from 57 bp to 1,318 bp, with an average size of 219 bp (Silflow 1998). The *Chlamydomonas* introns in the *cox3* gene range from 133 bp to 323 bp. *C. reinhardtii* contains a high intron density in its nuclear genome, estimated at 3.9-4.2 introns/kb of genomic sequence (Silflow 1998; Funke et al. 1999). For the *C. reinhardtii* *cox3* gene, a frequency of 2.9 introns/kb is observed. In contrast, the mitochondrial genomes of chlamydomonad algae exhibit a lower density of introns, as compared with the nuclear genome. Introns are present in the *cob* gene of *C. smithii* (Colleaux et al. 1990), the *cox1*, *nad5*, and *cob* genes of *C. elongatum* (Kroymann and Zetsche 1998), and in the *cox1*, *nad1*, *nad5*, *cob*, and ribosomal RNA coding regions of *C. eugametos* (Denovan-Wright et al. 1998). A degenerate group II intron is also present in the intronless mitochondrial genome of *C. reinhardtii* (Nedelcu and Lee 1998).

Introns 5-9 of the *cox3* gene of *C. reinhardtii* are distributed along the sequence encoding the mature COX III polypeptide (Fig. 1). Intron 6, the largest intron in the *C. reinhardtii* *cox3* gene, shows two internal conserved sequences (ICS) found in several introns of *C. reinhardtii* and in members of the genus *Volvox* (Table 2). In the majority of these *Chlamydomonas* and *Volvox* genes, these ICS are located near the 5' splicing junctions (ICS 1), or near the 3' splicing junctions (ICS 2). The ICSs reported in this work differ from most other *Chlamydomonas* ICS in that they lack the consensus sequence NCTAG located 15-51 bp upstream of the 3' splice site (Lee et al. 1997). The ICS described here also differ from ICS found in yeast, where the great majority of the introns found in its entire genome (214 out of 231) contain a 7-nucleotide

ICS, with the consensus sequence TACTAAC (Spingola et al. 1999).

The chlamydomonad ICS seem to be present in the larger introns of a gene or in introns not smaller than 212 bp. This observation is consistent with the fact that two ICS are present only in intron 6 of the *Chlamydomonas* *cox3* gene, which is the largest intron (323 bp). It is important to note that the relative position of ICS 1 within the intron is usually conserved, being near the 5' end. In some examples in Table 2, the sequence of ICS 1 is inverted with respect to the *cox3* intron 6 sequence. In these cases, the relative position of ICS 1 within the intron is also inverted and is found near the 3' end of the intron. The relative position of ICS 2 within the intron is also conserved and is usually found near the 3' end of the intron. When the sequence of ICS 2 is inverted, the position within the intron is also inverted and it is present near the 5' end of the intron. These ICS are not present in all *Chlamydomonas* introns, but they may play an important role in the processing of particular introns in this alga. ICS 2 was previously described by Funke et al. (1999), when comparing an intron from the gene encoding acetolactate synthase (*ALS*) with an intron of the gene encoding carbonic anhydrase (*CAH3*). Those authors proposed that this nucleotide stretch might represent a sequence encompassing a branch site of these introns. Or, it might represent a signal of yet unknown importance in the RNA-splicing process or in the expression of genes in *C. reinhardtii* (Lumbreras et al. 1998).

This is the first description of the sequences of *cox3* genes localized in the nucleus. Based on these sequences, it is proposed that *cox3* genes were transferred to the nucleus in a common ancestor of *Polytomella* sp. and *C. reinhardtii*. This process was accompanied by the loss of the mitochondrial *cox3* gene copy. The inactivation of the original mitochondrial gene and its subsequent disappearance likely occurred once the nuclear *cox3* gene copy became active. It is also proposed that the acquisition of an additional nucleotide sequence encoding a MTS and its corresponding introns occurred before the divergence of these two species. Nucleotide insertions, deletions and substitutions have occurred in introns 1-4, which changed their sequence and size, while retaining their original positions. Although it is difficult to distinguish between evolutionary scenarios, intron-loss against intron-gain for chlamydomonad *cox3* genes, the simplest interpretation is that intron acquisition in *C. reinhardtii* occurred after its separation from the

TESIS CON
 FALLA DE ORIGEN

Table 2. Internal conserved sequences (ICS 1, ICS 2) of *C. reinhardtii* introns. *C. reinhardtii* *cox3* intron 6 contains an ICS 1 and an ICS 2 that are present in other *Chlamydomonas* and *Volvox*

introns. *Superscript a* denotes ICS 1 or ICS 2 sequence that is inverted with respect to *cox3* intron 6. GenBank accession numbers are given in parentheses

Intron	Size (bp)	Gene name	Location in the intron
ICS 1: GGGAGGGGGGAGAGG			
6	323	<i>C. reinhardtii cox3</i>	31 bp from 5'
8	443	<i>C. reinhardtii</i> acetolactate synthase gene, <i>ALS</i> (AF047459)	6 bp from 5'
14	288	<i>C. reinhardtii</i> variable flagellar number protein gene, <i>VFL1</i> (AF154916)	16 bp from 5'
2	253	<i>C. reinhardtii</i> <i>EYE2</i> gene (AF233430)	22 bp from 5'
2 ^a	228	<i>C. reinhardtii</i> actin-related protein gene, <i>ACT2</i> (U68060)	21 bp from 3'
5 ^a	454	<i>Volvox carteri</i> argininosuccinate lyase gene, <i>VASL</i> (AF233374)	5 bp from 3'
1	748	<i>V. spermatozophera</i> UTEX 2273 GTPase Ypt4p gene (U55929)	2 bp from 5'
ICS 2: ATCATGAATGTAACCC			
6	323	<i>C. reinhardtii cox3</i>	25 bp from 3'
3	493	<i>C. reinhardtii</i> class II DNA photolyase gene, <i>PHR2</i> (AF129458)	177 bp from 3'
8	694	<i>C. reinhardtii</i> class II DNA photolyase gene, <i>PHR2</i>	17 bp from 3'
2	480	<i>C. reinhardtii</i> phosphatidylinositol 3-kinase gene (U97663)	21 bp from 3'
5 ^a	500	<i>C. reinhardtii</i> acetolactate synthase gene, <i>ALS</i>	2 bp from 5'
9	582	<i>C. reinhardtii</i> acetolactate synthase gene, <i>ALS</i>	34 bp from 3'
4	287	<i>C. reinhardtii</i> carbonic anhydrase α -type gene, <i>CAH3</i> (U73856)	26 bp from 3'
8 ^a	490	<i>C. reinhardtii</i> argininosuccinate lyase gene, <i>ARG7</i> (X16619)	27 bp from 5'
8	212	<i>C. reinhardtii</i> nitrate reductase gene, <i>NIT1</i> (AF203033)	20 bp from 3'
2	281	<i>C. reinhardtii</i> <i>zys1B</i> gene (AB001486)	32 bp from 3'

colorless genus *Polytomella*. This proposal is in agreement with the suggestion of a late occurrence of multiple intron-insertion events in the evolution of eukaryotic genes (Palmer and Logsdon 1991) and also with the proposal of a late acquisition of introns in the genes of *C. reinhardtii* and its multicellular relative, *V. carteri* (Funke et al. 1999). A test of the late-acquisition hypothesis could be accomplished through a comparative study of intron distribution among a sampling of chlamydomonad taxa that includes exemplars from all of the known lineages (Buchheim et al. 1996; Friedl 1997). In particular, it will be important that future investigations include those taxa currently regarded as sister to *C. reinhardtii* (e.g., *C. globosa*, *C. incerta*, *C. zebra*, *V. carteri*).

Acknowledgements The authors are indebted to Dr. E. Harris (Duke University) for providing the *C. reinhardtii* wild-type strain, to M. Vázquez-Acevedo for technical assistance, and to Drs. A. Atteia (IFC, UNAM), D.W. Krogmann (Purdue University), A. Lazcano-Araujo (Facultad de Ciencias, UNAM), and F. Recillas-Targa (IFC, UNAM) for reviewing the manuscript. We are indebted to M.A. Buchheim (University of Tulsa) for helpful suggestions and critical review of the manuscript. This work was supported by grants TW01176 from Fogarty International Center at NIH and HL59646 from NHLBI, NIH, USA, grant 27754 N from the Consejo Nacional de Ciencia y Tecnología, Mexico, and grant IN202598 from the Dirección General de Asuntos del Personal Académico, Universidad Nacional Autónoma de México, Mexico.

References

Antaramian A, Coria R, Ramírez J, González-Halphen D (1996) The deduced primary structure of subunit I from the cytochrome c oxidase suggests that the genus *Polytomella* shares a common mitochondrial origin with *Chlamydomonas*. *Biochim Biophys Acta* 1273:198–202

Atteia A, Dreyfus G, González-Halphen D (1997) Characterization of the α and β subunits of the F_0F_1 -ATPase from the alga *Polytomella* sp., a close relative of *Chlamydomonas reinhardtii*. *Biochim Biophys Acta* 1320:275–284

Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett* 325:140–145

Buchheim MA, Lemieux C, Otis C, Gutell RR, Chapman RL, Turmel M (1996) Phylogeny of the Chlamydomonadales (Chlorophyceae): a comparison of ribosomal RNA gene sequences from the nucleus and the chloroplast. *Mol Phylogenet Evol* 5:391–402

Claros MG, Perca J, Shu Y, Samatey FA, Popot JL, Jacq C (1995) Limitations to in vivo import of hydrophobic proteins into yeast mitochondria. The case of a cytoplasmically synthesized apocytochrome *b*. *Eur J Biochem* 228:762–771

Colleaux L, Michel-Woltertz M-R, Matagne RF, Dujon B (1990) The apocytochrome *b* gene of *Chlamydomonas smithii* contains a mobile intron related to both *Saccharomyces* and *Neurospora* introns. *Mol Gen Genet* 223:288–296

Conner TW, Thompson MD, Silflow CD (1989) Structure of the three β -tubulin-encoding genes of the unicellular alga, *Polytomella agilis*. *Gene* 84:345–358

Denovan-Wright EM, Nedelcu AM, Lee RW (1998) Complete sequence of the mitochondrial DNA of *Chlamydomonas eugametos*. *Plant Mol Biol* 36:285–295

Friedl T (1997) The evolution of green algae. *Plant Syst Evol [Suppl]* 11:87–101

Funke RP, Kovar JL, Logsdon JM Jr, Corrette-Bennet JC, Straus DR, Weeks DP (1999) Nucleus-encoded, plastid-targeted acetolactate synthase genes in two closely related chlorophytes, *Chlamydomonas reinhardtii* and *Volvox carteri*: phylogenetic origins and recent insertion of introns. *Mol Gen Genet* 262:12–21

Gutiérrez-Cirlos EB, Antaramian A, Vázquez-Acevedo M, Coria R, González-Halphen D (1994) A highly active ubiquinol-cytochrome c reductase (*bc₁* complex) from the colorless alga *Polytomella* sp., a close relative of *Chlamydomonas*. *J Biol Chem* 269:9147–9154

Hiesel R, Haeseler A von, Brennicke A (1994) Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc Natl Acad Sci USA* 91:634–638

TESIS CON
FALLA DE ORIGEN

- Kroymann J, Zetsche K (1998) The mitochondrial genome of *Chlorogonium elongatum* inferred from the complete sequence. *J Mol Biol* 47:431-440
- Lee VD, Stapleton M, Huang B (1997) Genomic structure of *Chlamydomonas caltractin*. Evidence for intron insertion suggests a probable genealogy for the EF-hand superfamily of proteins. *J Mol Biol* 22:175-191
- Liss M, Kirk DL, Beyser K, Fabry S (1997) Intron sequences provide a tool for high-resolution phylogenetic analysis of volvocine algae. *Curr Genet* 31:214-227
- Lumbreras V, Stevens D, Purton S (1998) Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. *Plant J* 14:441-447
- Malek O, Lattig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J* 15:1403-1411
- Matsuoka M, Matsubara M, Kakehi M, Imanaka T (1994) Homologous maturase-like proteins are encoded within the group I introns in different mitochondrial genes specifying *Yarrowia lipolytica* cytochrome *c* oxidase subunit 3 and *Saccharomyces cerevisiae* apocytochrome *b*. *Curr Genet* 26:377-381
- Mattox KR, Stewart KD (1984) Classification of the green algae. In: Irvine DEG, John DM (eds) Systematics of the green algae. Academic Press, London, pp 29-72
- Melkonian M, Surek B (1995) Phylogeny of the Chlorophyta: congruence between ultrastructural and molecular evidence. *Bull Soc Zool Fr* 120:191-208
- Michaelis G, Vahrenholz C, Prätze E (1990) Mitochondrial DNA of *Chlamydomonas reinhardtii*: the gene for apocytochrome *b* and the complete functional map of the 15.8 kb DNA. *Mol Gen Genet* 223:211-216
- Nakayama T, Watanabe S, Mitsui K, Uchida H, Inouye I (1996) The phylogenetic relationship between the Chlamydomonadales and Chlorococcales inferred from 18S rRNA sequence data. *Phycol Res* 44:47-55
- Nedelcu A, Lee RW (1998) A degenerate group II intron in the intronless mitochondrial genome of *Chlamydomonas reinhardtii*: evolutionary implications. *Mol Biol Evol* 15:918-922
- Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Kohchi T, Ogura Y, Kanehara T, Akashi K, Ohshima K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J Mol Biol* 223:1-7
- Palmer JD, Logsdon JM (1991) The recent origins of introns. *Curr Opin Genet Dev* 1: 470-477
- Pérez-Martínez X, Vázquez-Acevedo M, Tolkunova E, Funes S, Claros MG, Davidson E, King MP, González-Halphen D (2000) Unusual location of a mitochondrial gene: subunit III of cytochrome *c* oxidase is encoded in the nucleus of chlamydomonad algae. *J Biol Chem* 275:30144-30152
- Pérez-Martínez X, Antaramian A, Vázquez-Acevedo M, Funes S, Tolkunova E, d'Alayer J, Claros MG, Davidson E, King MP, González-Halphen D (2001) Subunit II of cytochrome *c* oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J Biol Chem* 276:11302-11309
- Rochaix JD, Mayfield S, Goldschmidt-Clermont M, Erickson J (1988) Plant molecular biology. In: Shaw CH (ed) Molecular biology of *Chlamydomonas*. IRL Press, Oxford, pp 253-275
- Round FE (1980) The evolution of pigmented and unpigmented unicells - a reconsideration of the Protista. *Biosystems* 12:61-69
- Rumpf R, Vernon D, Schreiber D, Birky CW Jr (1996) Evolutionary consequences of the loss of photosynthesis in Chlamydomonadaceae: phylogenetic analysis of *Rrm18* (18S rRNA) in 13 *Polytoma* strains (Chlorophyta). *J Phycol* 32:119-126
- Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- Sillflow C (1998) Organization of the nuclear genome. In: Rochaix JD, Goldschmidt-Clermont M, Merchant S (eds) The molecular biology of chloroplasts and mitochondria in *Chlamydomonas*. Kluwer, Rotterdam, pp25-40
- Spingola M, Grate L, Haussler D, Ares M Jr (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 5:221-234

Apendice V.

Artículo publicado:

Two unusual amino acid substitutions in cytochrome *b* of the colorless alga *Polytomella* spp.: correlation with the atypical spectroscopical properties of the *b_H* heme.

Two Unusual Amino Acid Substitutions in Cytochrome b of the Colorless Alga *Polytomella* spp.: Correlation with the Atypical Spectral Properties of the b_H Heme¹

Anaid Antaramian, Soledad Funes, Miriam Vázquez-Acevedo, Ariane Atteia, Roberto Coria, and Diego González-Halphen²

Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510 Mexico City, D.F., Mexico

Received January 7, 1998

The dithionite-reduced spectra of the purified bc_1 complexes from the colorless alga *Polytomella* spp. and the closely related green alga *Chlamydomonas reinhardtii* were compared. The spectrum of the bc_1 complex from *C. reinhardtii* showed a profile similar to those of the bc_1 complexes from other species. In contrast, the bc_1 complex from *Polytomella* spp. exhibits a double-peak spectrum in the α -band region, where the absorption bands of cytochrome c_1 and cytochrome b are completely resolved. To further understand the molecular basis of these spectroscopic differences, the mitochondrial gene encoding cytochrome b of *Polytomella* spp. was cloned, sequenced, and compared with that of *C. reinhardtii*. The *Polytomella* spp. cytochrome b gene is 1113 bp long and does not contain introns. The deduced protein sequence exhibits 56% identity and 68% similarity with the cytochrome b of *C. reinhardtii*, and in a phylogenetic analysis it clearly affiliated with the b-type cytochromes of *C. reinhardtii* and *C. smithii*. A comparison of the primary sequences of the *Polytomella* spp. cytochrome b with other b-type cytochromes, and its analysis based on the structure featuring eight transmembrane stretches, allowed the identification of a tyrosine in position 114, which substitutes for a tryptophan present in all mitochondrial b-type cytochromes sequenced to date. In addition, the primary sequence of the cytochrome b from *Polytomella* spp.

has a serine at position 36, instead of a nonpolar residue (alanine or leucine) found in all other species. In the proposed model for cytochrome b, both residues Tyr₁₁₄ and Ser₃₆ are in close proximity to the high-potential b_H heme. The above data suggest that the polar residues Y₁₁₄ and S₃₆, each one by itself or in combination, may interact with heme b_H of *Polytomella* spp. and, thus, may be responsible for the unique spectroscopic characteristics of cytochrome b.

© 1998 Academic Press

Key Words: *Polytomella*; *Chlamydomonas*; mitochondria; bc_1 complex; cytochrome b; Chlorophyceae.

The algae of the genus *Polytomella* (1, 2) are colorless members of the family Chlamydomonadaceae (3, 4). The genus *Polytomella* is thought to have arisen by secondary evolution from the *Chlamydomonas* lineage, losing its cell wall and its functional chloroplasts (5, 6). Structural, functional, biochemical, and molecular evidence supporting the close relationship between the genera *Polytomella* and *Chlamydomonas* have been provided by different authors (4, 7–11).

Ubiquinol-cytochrome c oxidoreductase (bc_1 complex) is a multisubunit mitochondrial membrane enzyme that binds three redox prosthetic groups: a c-type heme in cytochrome c_1 , an Fe₂S₂ center in the Rieske protein, and two b-type hemes in cytochrome b (12). Embedded in the inner mitochondrial membrane, cytochrome b constitutes the hydrophobic core of the bc_1 complex (13, 14). The available data indicate that the tertiary structure of cytochrome b is dominated by eight transmembrane stretches (15–18), two of which contain invariant histidine residues that ligate the heme irons: H₈₂ and H₁₈₃ (yeast numbering) coordinate

¹ The nucleotide sequence data reported in this paper have been submitted to GenBank under Accession No. U87396.

² To whom correspondence should be addressed at Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, 04510 México, D.F., México. Fax: (525) 622-5611 or (525) 548-0387. E-mail: dhalphen@ifcsun1.ifisiol.unam.mx.

heme b_L on the P-side of the membrane, and H₉₈ and H₁₉₇ ligate heme b_H on the N-side of the membrane (19). The crystal structure of the bc₁ complex from bovine heart mitochondria (20) has confirmed the predicted topology of cytochrome b. The two heme groups b_L and b_H are associated with two redox catalytic units in the bc₁ complex, the Q_o center and the Q_i center, respectively (21–25).

A highly active bc₁ complex from *Polytomella* spp. has been purified and characterized (26). It exhibits a clearly defined double-peak spectrum in the α -region with absorption maxima at 553 nm for cytochrome c₁ and at 567 nm for cytochrome b. The spectroscopic characteristics of *Polytomella* spp. cytochrome b are thus distinct from those in mammals, plants, yeasts, and bacteria. In order to interpret on a molecular basis the spectroscopic differences of the cytochrome b, we have compared the fully reduced spectra of the bc₁ complexes from *Polytomella* spp. and *Chlamydomonas reinhardtii*. We also describe the nucleotide sequence of the *cytb* gene and the deduced primary structure of the mitochondrial apocytochrome b of *Polytomella* spp. and compare it to its homologue in *C. reinhardtii*. Two striking substitutions were found in the amino acid sequence of cytochrome b from *Polytomella* spp.: residues S₃₆ and Y₁₁₄ were unique to this sequence, replacing residues A₃₆ and W₁₁₄ present in the cytochrome b of *C. reinhardtii* and of several other species. In particular, W₁₁₄ is one of the invariant amino acids in all apocytochrome b sequences reported to date, and its replacement by a tyrosine could account for the peculiar spectroscopic characteristics of the cytochrome b of *Polytomella* spp.

MATERIALS AND METHODS

Strains and plasmids. *Polytomella* spp. (198.80, E. G. Pringsheim) was grown in the medium described by Wise (3) supplemented with vitamins B₁ (0.06 mg/liter) and B₁₂ (0.08 mg/liter) (27). *Escherichia coli* DH5 α was used to amplify recombinant plasmids. Bluescript phagemid (Stratagene) was used for subcloning and sequencing.

Mitochondrial DNA isolation and construction of the genomic mitochondrial DNA library. Mitochondria from *Polytomella* spp. were prepared as described by Gutiérrez-Cirlos *et al.* (26). The mitochondrial *Hind*III DNA library was constructed as described by Antaramian *et al.* (10).

Amplification and partial sequencing of the *cytb* gene. Mitochondrial DNA (1 μ g) from *Polytomella* spp. was used as a template for amplification by the polymerase chain reaction (PCR) in the presence of 8 mM MgCl₂ and 200 ng of each of the following degenerate oligodeoxynucleotides, which were synthesized on an Applied Biosystems Model 381A DNA synthesizer at the Molecular Biology Facilities of the Instituto de Fisiología Celular, UNAM (Mexico). The forward primer, 5'-agt(c) ttg ttc(t) ttc(t) att gta(c) gta(c) tac(t) t(c)tg(a) cac(t)-3', corresponded to the conserved region from amino acid 87 to 95 of *C. reinhardtii* cytochrome b (SLFFIVVYLH). The backward primer, 5'-ta a(t)ac cca a(a)gaa a(g)ta cca ctc t(a)gg (a)ac a(g)at atg-3', corresponded to the complementary region of *C. reinhardtii* cytochrome b from amino acid 267 to 278 (HIVPEWYFL-

WVY). The oligodeoxynucleotides were designed based on the mitochondrial codon usage of *C. reinhardtii* (28).

For PCR amplification, samples were denatured for 12 min at 94°C and subjected to 50 cycles of 30-s denaturation at 94°C, 45-s annealing at 55°C, and 2-min extension at 72°C. A final extension for 12 min at 72°C was carried out. The PCR product was gel purified and ligated into the pCRII plasmid (Invitrogen). Transformed *E. coli* DH5 α cells were plated on ampicillin LB plates and recombinant plasmids were selected by loss of β -galactosidase activity. Positive clones were sequenced using the kit Seq Ver 2 from USB, based on the dideoxy chain termination method of Sanger *et al.* (29). Standard recombinant DNA techniques were carried out essentially as described by Sambrook *et al.* (30).

Southern analysis. The mitochondrial library clones were transferred to a nylon membrane and hybridized with the ³²P radioactive probe of the *cytb* gene PCR-amplified product from *Polytomella* spp. The hybridization was carried out at 55°C, and the nylon membrane was washed twice with 5 \times SSC³ containing 0.5% SDS and once with 0.2 \times SSC containing 1% SDS at the same temperature and was exposed overnight with X-OMAT AR Kodak film. With this procedure, clone 9H from the *Hind*III mtDNA library was found to hybridize with the radioactive probe and to contain a partial sequence of the *cytb* gene.

PCR amplification of the upstream region of the *cytb* gene. Since clone 9H from the *Hind*III library did not contain the complete *cytb* gene, an additional PCR amplification step was carried out in order to obtain the sequence of the corresponding 3' end region. The PCR was done with mtDNA from *Polytomella* spp. as a template, with the specific forward primer 5'-tta acc aag gca aag ctg aac c-3' (nucleotides 1101 to 1122 from the known *cytb* gene sequence from clone 9H) and five different randomly synthesized decamers as backward primers (synthesized at the Instituto de Fisiología Celular, UNAM). The mixture for PCR contained 1 mM MgCl₂, 0.5% dimethyl sulfoxide, and 0.5% bovine serum albumin. Samples were denatured for 10 min at 94°C and subjected to 50 cycles of 45-s denaturation at 94°C, 45-s annealing at 50°C, a 1-min extension at 72°C, and an additional extension for 10 min at 72°C. The PCR products thus obtained were subjected to Southern blot analysis using the same 9H clone as a probe. Two different positive 1-kb products were identified, cloned in pMOS (Amersham), and sequenced; both were found to contain the 3' end of the *cytb* gene.

Purification of bc₁ complexes. The bc₁ complex from *Polytomella* spp. was purified as described by Gutiérrez-Cirlos *et al.* (26). The bc₁ complex from *C. reinhardtii* was purified following the methods of Attala (31) and Attala and Franzén (32), using the multiple-photosynthetic mutant strain BF4F54F14 (33). Both bc₁ complexes were kept at -70°C until used.

Difference spectra of the bc₁ complex. Difference spectra were recorded using bc₁ complexes at concentrations of 2.0–6.0 μ M cytochrome b, using a few grains of dithionite as reductant.

Purification of cytochrome b from *Polytomella* spp. and determination of its N-terminal sequence. Subunit III (cytochrome b) from the bc₁ complex from *Polytomella* spp. was purified by SDS-polyacrylamide gel electrophoresis. For this purpose, the purified bc₁ complex (200 μ g of protein per lane, 10 lanes) was run in a Coomassie blue SDS-PAGE as described by Schagger *et al.* (34) with the modifications described by Vázquez-Acevedo *et al.* (35). The stained band of interest was excised from the gel and ground in a mortar in the presence of sand (1:1 w/w). When a fine paste was formed, 2 vol of 20 mM Tris-HCl (pH 7.0) was added, and the mixture centrifuged at 30,000g for 15 min. The supernatant, which contained the extracted and purified polypeptide, was lyophilized and subjected to SDS-

³ Abbreviations used: CAPS, 3-cyclohexylamino-1-propanesulfonic acid; mtDNA, mitochondrial DNA; SSC, standard saline citrate.

TESIS CON
FALLA DE ORIGEN

PAGE as described by Schagger *et al.* (36). Subunit III (cytochrome b) was transferred onto a ProBlott membrane (Applied Biosystems) at 400 mA for 2.5 h (4°C) in the presence of 10 mM Caps (pH 11.0) and 10% methanol following the procedure of Matsudaira (37). The membrane was stained with Coomassie brilliant blue, destained, and air-dried as described by the same author. Amino-terminal sequence analysis was carried out by Dr. J. D'Alayer on an Applied Biosystems sequencer at the Laboratoire de Microséquençage des Protéines, Institut Pasteur, Paris (France).

Sequence analysis. Sequences were analyzed using the GCG Sequence Analysis Software Package (Genetics Computer Group, Madison, WI). Alignments were carried out with the Pileup program (38) using sequences in the SwissProt data bank. Distances in the cladogram were corrected according to Kimura (39). The terms similarity and homology are used as suggested by Reeck *et al.* (40). Accepted amino acid substitutions for considering similarity were K=R, D=E, S=T, I=L=V=M, and F=Y.

RESULTS

Spectral characterization of the *bc₁* complexes. The dithionite-reduced spectra of the *bc₁* complexes purified from *Polytomella* spp. and from its photosynthetic counterpart *C. reinhardtii* were recorded at room temperature. The fully reduced *bc₁* complex of *Polytomella* spp. exhibited absorption maxima at 553 nm for cytochrome *c₁* and at 567 nm for cytochrome *b* in the α -region of the spectrum (Fig. 1A). This gives rise to a double-peak spectrum of the *bc₁* complex, in which the absorption maximum of cytochrome *c₁* is clearly resolved from that of cytochrome *b* at room temperature. In contrast, the absorption spectrum of the dithionite-reduced *bc₁* complex from the closely related alga *C. reinhardtii* (Fig. 1B) showed a maximum at 561.5, corresponding to heme *b*, and a shoulder at 554.7 nm, characteristic of cytochrome *c₁*. This profile is similar to those of *bc₁* complexes from several other organisms, including mammalian, plant, yeast, and bacterial *bc₁* complexes, in which the absorption bands of cytochromes *c₁* and *b* are overlapped. Thus, the *b*-type heme from *Polytomella* spp. exhibits a redshifted α -band, which seems to be unique to the *bc₁* complex from this colorless alga. To understand the molecular basis of these spectroscopic differences, the mitochondrial gene encoding cytochrome *b* (*cytb*) of *Polytomella* spp. was cloned and sequenced.

Characterization of the *cytb* Gene from *Polytomella* spp. Two degenerate oligodeoxynucleotides were designed based on highly conserved regions of the *cytb* gene of *C. reinhardtii*. With these primers, a PCR-amplification product of 0.8 kb was obtained, using mtDNA from *Polytomella* spp. as a template. Nucleotide sequence analysis showed that it was an amplified product of the *cytb* gene that exhibited 68% similarity with the *cytb* gene from *C. reinhardtii*. The 0.8-kb PCR product was used as a probe to screen the *Hind*III mtDNA library of *Polytomella* spp. A single clone from this library (clone 9H) gave a positive signal in Southern blot analysis. Nucleotide sequencing of the 9H clone revealed the

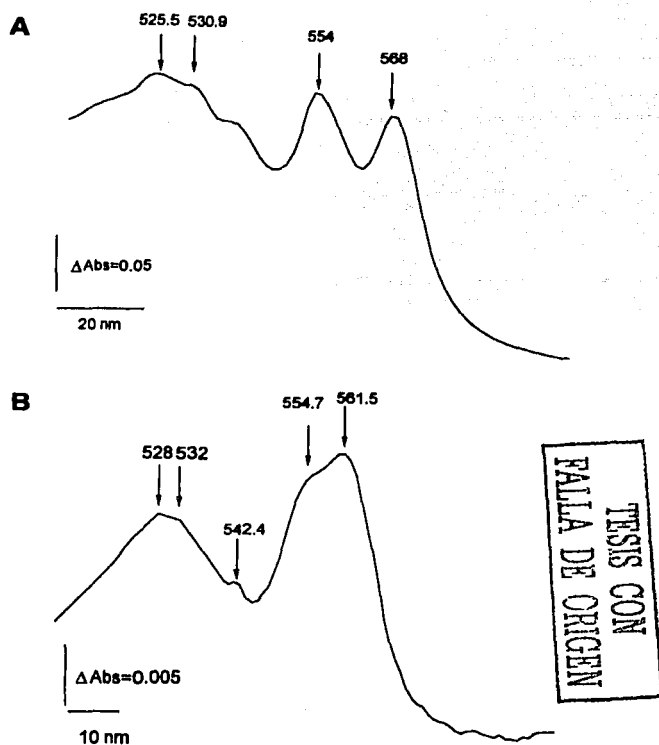


FIG. 1. Spectra of the dithionite-reduced *bc₁* complexes from *Polytomella* spp. and *C. reinhardtii*. (A) The *bc₁* complex from *Polytomella* spp. was diluted in 25 mM potassium phosphate buffer (pH 7.2) containing 2 mM EDTA and 0.1 mg/ml lauryl maltoside and was fully reduced with a few grains of dithionite. (B) The *bc₁* complex from *C. reinhardtii* was diluted in 20 mM Tris-acetate buffer (pH 7.0) containing 0.05% Triton X-100 and fully reduced with a few grains of dithionite.

presence of an open reading frame of 1029 bp encoding the 343 residues of the sequence of cytochrome *b*, including its N-terminus. Since clone 9H from the *Hind*III library did not contain the complete sequence of the *cytb* gene, an additional PCR amplification step with randomly synthesized decameric primers was carried out in order to obtain the sequence that encoded the carboxy-terminal region of cytochrome *b* (see Materials and Methods). The PCR products were transferred to a nylon membrane and hybridized with the *cytb* gene from clone 9H as a probe. Two different positive 1-kb products that contained the 3' end of the *cytb* gene were identified and sequenced.

The complete nucleotide sequence of the *cytb* gene from *Polytomella* spp. and the predicted primary struc-

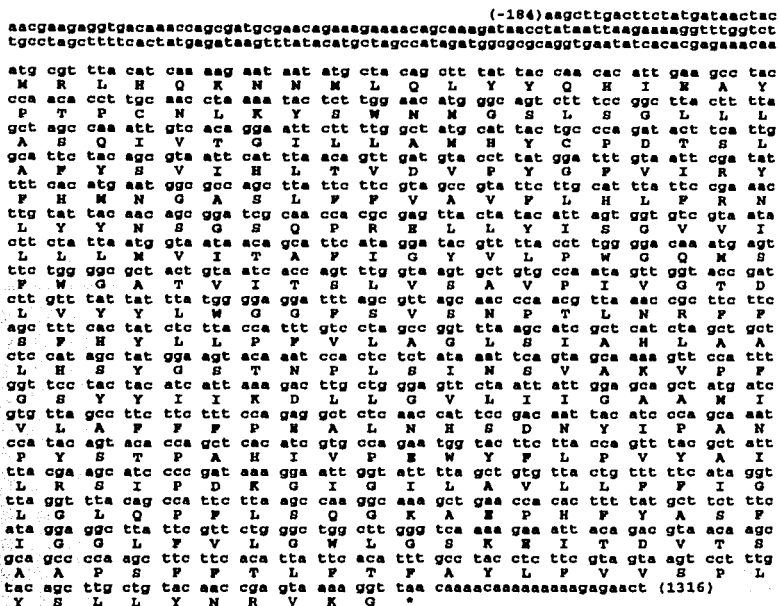

 FESIS CON
 FALLA DE ORIGEN

FIG. 2. Nucleotide sequence of the mitochondrial *cytb* gene from *Polytomella* spp. The derived primary sequence of apocytochrome b is also shown.

ture of the apocytochrome are illustrated in Fig. 2. The nucleotide sequence showed an open reading frame of 1113 bp that was not interrupted by intervening sequences. Translation of the DNA sequence predicts a protein of 370 residues with a molecular mass of 41,226 Da. To corroborate that the deduced primary sequence was also present in the cytochrome b of the mitochondrial bc₁ complex from *Polytomella* spp., subunit III (35.5 kDa) was purified and subjected to 15 cycles of N-terminal sequencing. The identified residues are shown in Fig. 3. These first 15 amino acid residues were identical to those deduced from the nucleotide sequence (Fig. 2), indicating that the sequence of the mature and assembled cytochrome b is encoded by the *cytb* gene. The predicted apoprotein of cytochrome b from *Polytomella* spp. shows a hydropathy profile similar to those of cytochrome b from other organisms (data not shown) and fits the eight-transmembrane-helices structure (20). It contains the four invariant histidine residues (H₈₂, H₉₆, H₁₈₃, and H₁₉₇) that ligate heme b_L and b_H. In addition, it contains the other highly conserved residues found in the cytochrome b sequences and considered to be invariant (41): G₃₃, G₄₇, C₇₅, G₁₁₇, G₁₃₁, S₁₄₀, T₁₇₅, H₂₀₂, D₂₂₉, P₂₇₁, E₂₇₂, W₂₇₃, L₂₈₂, and G₂₉₁ (yeast numbering). Notably, the completely conserved residue W₁₁₄ was substituted by a

tyrosine in the *Polytomella* spp. cytochrome b sequence. The derived sequence of apocytochrome b of *Polytomella* spp. was also compared with the one from *C. reinhardtii* (28, 42), revealing an identity of 56% and an overall similarity of 68% (Fig. 4). This similarity is less evident in the carboxy-terminal regions of the two

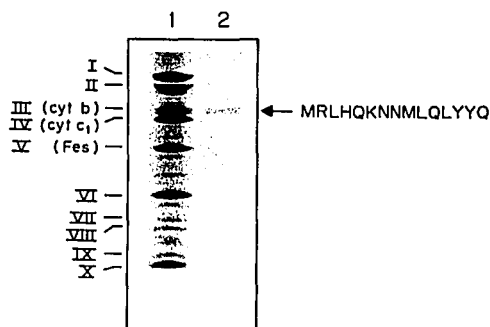


FIG. 3. N-terminal sequence of cytochrome b (subunit III) of the bc₁ complex from *Polytomella* spp. Lane 1, purified bc₁ complex from *Polytomella* spp. The 10 subunits which constitute the complex are indicated (26). Lane 2, the N-terminal sequence of the purified 35.5-kDa polypeptide (subunit III) is shown.

		2nd					
		U	C	A	G		
1st	U	Phe 10 5	Ser 3 7	Tyr 9 5	Cys 0 1	U	
		Phe 20 20	Ser 3 1	Tyr 17 17	Cys 2 2	C	
		Leu 24 6	Ser 3 0	Stop 1 1	Stop 0 0	A	
		Leu 8 27	Ser 1 0	Stop 0 0	Trp 6 11	G	
	C	Leu 7 5	Pro 4 1	His 7 2	Arg 1 5	U	
		Leu 5 1	Pro 1 0	His 6 12	Arg 2 2	C	
		Leu 7 19	Pro 17 18	Gln 6 9	Arg 4 0	A	
		Leu 4 1	Pro 0 0	Gln 2 2	Arg 0 0	G	
	A	Ile 13 16	Thr 2 14	Asn 7 0	Ser 8 7	U	
		Ile 7 11	Thr 2 7	Asn 8 15	Ser 14 10	C	
		Ile 7 0	Thr 10 0	Lys 7 5	Arg 0 0	A	
		Met 8 15	Thr 1 0	Lys 1 0	Arg 0 0	G	
G	Val 9 4	Ala 14 21	Asp 4 3	Gly 8 29	U		
	Val 3 7	Ala 7 8	Asp 3 4	Gly 7 5	C		
	Val 14 19	Ala 6 0	Glu 4 0	Gly 12 1	A		
	Val 4 1	Ala 0 0	Glu 2 5	Gly 1 0	G		

FIG. 6. Comparison between the codon usage of the mtDNA sequences of *cytb* from *Polytomella* spp. and from *C. reinhardtii*. Boxes indicate the main differences in codon usage. Codon usages were derived from the sequences of *cytb* from *Polytomella* spp. (bold characters, this study) and from *C. reinhardtii* (25).

pombe (44), *Aspergillus nidulans* (45), *Neurospora crassa* (46), or the related alga *C. smithii* (47).

The deduced primary structure of cytochrome b from *Polytomella* spp. exhibited similarity with its green counterparts *C. reinhardtii* and *C. smithii*, clearly affiliating with these two species in a phylogenetic analysis and thus providing additional evidence for the close relationship between the genus *Polytomella* and the genus *Chlamydomonas*. The similarity between the cytochromes b from *Polytomella* spp. and *C. reinhardtii* is significantly higher in the N-terminal regions (67% identity), including the first four transmembrane helices, than in the C-terminal part (49% identity), consistent with all other sequences of apocytochrome b (41, 48, 49).

Although the apocytochromes b from *Polytomella* spp. and *C. reinhardtii* share highly similar amino acid sequences (56% identity), the cytochromes exhibit different spectroscopic properties when assembled in the mitochondrial bc₁ complex (Fig. 1). In *Polytomella* spp., the maximum absorption peak of cytochrome b is

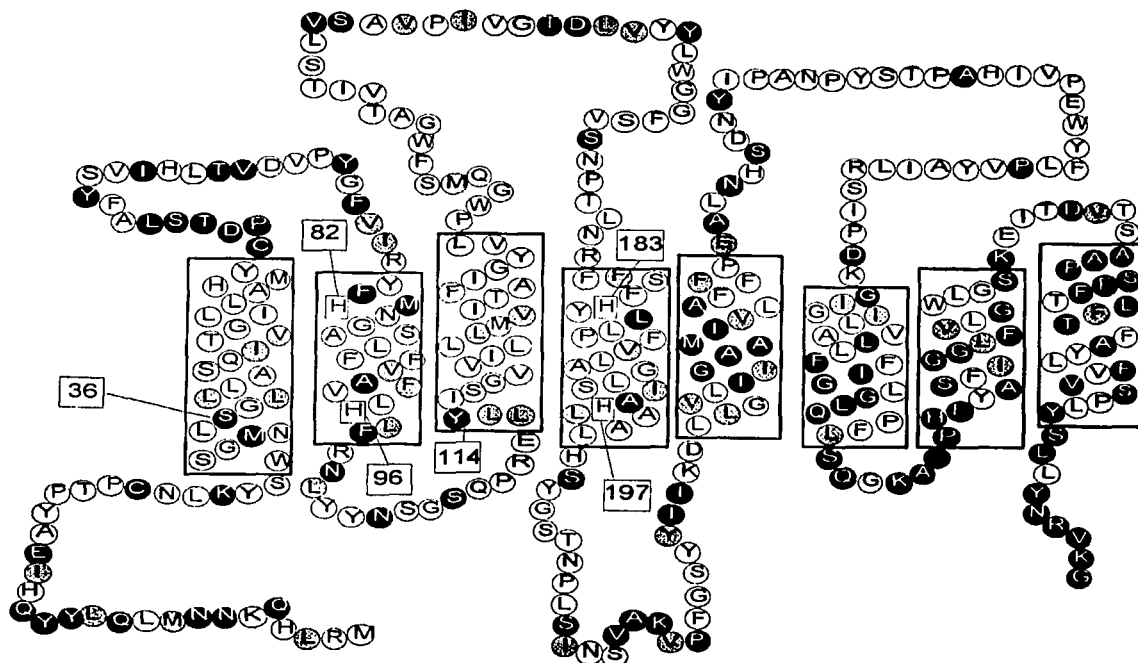


FIG. 7. Proposed transmembrane structure of cytochrome b from *Polytomella* spp. A model based on hydropathy plots that proposes eight transmembrane segments (A to H), with amino and carboxy termini exposed to the mitochondrial matrix surface. The model also shows a comparison of amino acid residues between *Polytomella* spp. and *C. reinhardtii*: Identical residues are shown in white circles, conserved residues in gray, and variant residues in black. The four invariant histidines that bind the hemes are indicated, as well as residues S₃₆ and Y₁₁₄.

TESIS CON
FALLA DE ORIGEN

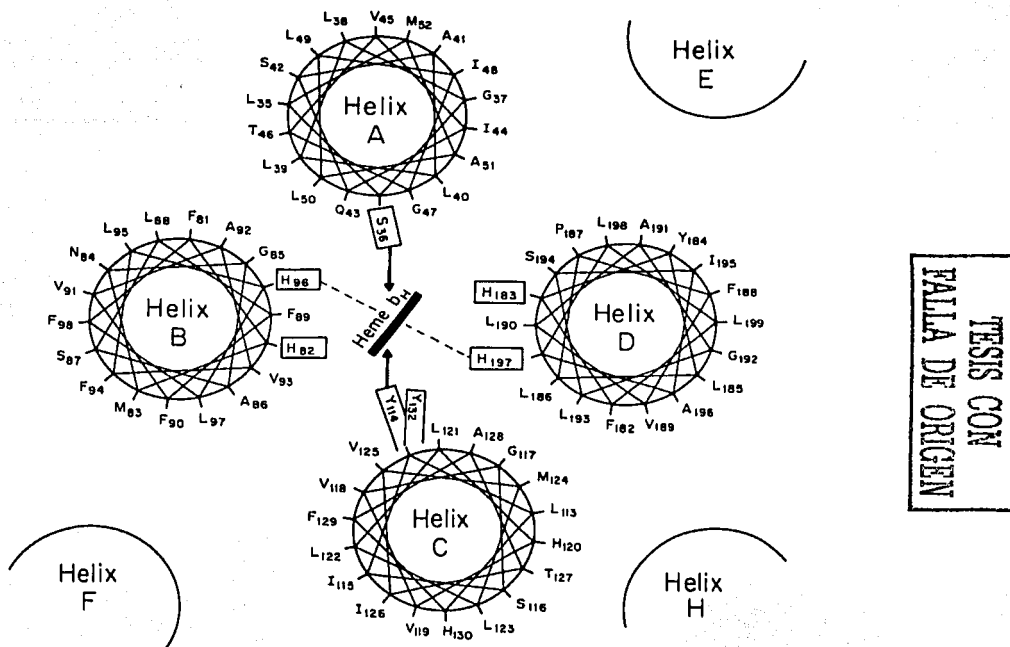


FIG. 8. Helical wheel representation of helices A to D of *Polytomella* spp. cytochrome b as seen from the mitochondrial intermembrane space. The pairs of histidines that bind the heme groups are H82, 96, 183, and 197. The two putative changes in the amino acid residues that may interact with the b_L heme group are S_{36} in helix A, and Y_{114} in helix C. The position of residue Y_{132} , exactly 18 amino acids away from Y_{114} , is also depicted (see text).

shifted 3–4 nm to the red, compared with its counterpart in *C. reinhardtii*.

Residue Y_{114} and possibly S_{36} may confer the atypical spectroscopic characteristics on the b_H heme of *Polytomella* spp. A comparison of the primary structures of the b-type cytochromes from *Polytomella* spp. and all other species, including the closely related alga *C. reinhardtii*, shows the unique substitution of W_{114} for Y (yeast numbering). Based on the eight-transmembrane-helices structure, we propose that S_{36} and Y_{114} are likely to be in close proximity to the b_H heme in *Polytomella* spp. (Figs. 7 and 8). Tryptophan 114 is one of the invariant amino acids present in all cytochromes b sequenced to date (41). Although the exchange Trp to Tyr is relatively common in members of the same protein family, the fact that this occurs only in *Polytomella* spp. at position 114 could be related to the unusual spectroscopic properties of cytochrome b. The aromatic residues surrounding the heme groups seem to maintain a tightly packed hydrophobic environment around these prosthetic groups and must contribute to their spectroscopic properties. As discussed by Mathews

(50), cytochrome b_{562} from *E. coli* and the cytochrome c' monomer from *Rhodospirillum molischianum* have highly similar three-dimensional structures, and the heme group is located in very similar structural environments in the two proteins. The large differences in the spectroscopic properties of the two hemes are the result of differences in the amino acids that surround these prosthetic groups, particularly tryptophans, tyrosines, and phenylalanines.

Using computer simulations, Link *et al.* (51) proposed a model of the amino acid residues that constitute the environments of hemes b_L and b_H . In this model heme b_H is surrounded by three tryptophan residues, W_{29} , W_{30} , and W_{114} (yeast numbering). It is relevant that W_{114} was predicted to have an angle of 70° at a distance of 3.2 Å from heme b_H . X-ray crystallography of the bovine bc_1 complex (20) has confirmed a closest distance of 4 Å between this residue and heme b_H (Chang-An Yu, personal communication). The replacement of W_{114} by Y_{114} in *Polytomella* spp. cytochrome b maintains the aromatic character of the residue, but introduces differences regarding the size and H-bonding capacity of the amino

acid group. Tyrosine 114 could produce a different electronic environment around the b_H heme, delocalizing electrons of the tetrapyrrole ring in a manner different from that of tryptophan.

Only a few inhibitor-resistant mutations of cytochrome b have been characterized in *C. reinhardtii*. Bennoun *et al.* (52) described a $Y_{132}C$ mutation that confers resistance to the inhibitors myxothiazol and mucidin. Data obtained with this mutant suggest a close proximity of Y_{132} to the b_L heme, an essential component of the Q_0 center. This residue is also present at position 132 of the deduced *Polytomella* spp. cytochrome b sequence. When the α helix modeling is carried out on the C transmembrane stretch, Y_{132} is found to lie 18 residues away from Y_{114} , which as suggested here would interact with the b_H heme (Fig. 8). This face of the C transmembrane α helix would be exposed to the space in which both heme moieties are present, and thus Y_{114} must be located near the b_H heme and Y_{132} near the b_L heme. These data are in accord with the proposal made here of a close proximity of Y_{114} with the b_H heme.

The precise nature of the protein factors that determine the optical properties of b-type cytochromes is not known. However, it is most likely that the electronic density and relative polarity of the heme pocket influences the magnitude and polarization of the orthogonally polarized transitions of the heme group ($Q_{X,Y}$) that determine the α -band (53). In this view, the proximity and orientation of electron-dense aromatic groups close to the tetrapyrrole ring should have a dominant effect on the properties of the electronic transitions, thus influencing their optical absorption band. In the case of *Polytomella* spp. cytochrome b, it is feasible that Y_{114} (alone or in combination with S_{36}) introduces an electronic environment distinct from that in other species that have W_{114} , so that the energy levels of the Q_X and Q_Y transitions are much more separated. This induces a splitting of the α -band of the b_H heme, when normally they are so close that they fuse into a single band of larger width (25). To our knowledge, this is the first documented case of a natural variation of a single amino acid that may confer detectable differences in the spectroscopic characteristics of a heme moiety in a b-type cytochrome. The present results also predict that site-directed mutagenesis $Y_{114}W$ on the *Polytomella* spp. *cytb* gene or, alternatively, the $W_{114}Y$ mutation on the *C. reinhardtii* gene should alter the spectroscopic properties of their respective b_H hemes.

ACKNOWLEDGMENTS

The authors acknowledge the technical expertise of Dr. Laura Ongay (Unidad de Biología Molecular, IFC, UNAM) in the synthesis of *cytb* oligodeoxynucleotides. We are grateful to Drs. M. Degli Esposti, C. Gómez-Lojero, A. Gómez-Puyou, E. B. Gutiérrez-Cirlos, and

D. W. Krogmann, for helpful discussions and critical review of the manuscript. This work was supported by Grants IN-204595 from DGAPA, UNAM, and 2260P-N from CONACyT, Mexico.

REFERENCES

1. Aragao, H. (1910) *Mem. Inst. Oswaldo Cruz* **2**, 42-57.
2. Pringsheim, E. G. (1955) *J. Protozool.* **2**, 137-145.
3. Wise, D. L. (1959) *J. Protozool.* **6**, 19-23.
4. Melkonian, M. (1990) in *Handbook of Protoctista* (Margulis, L., Corliss, J. O., Melkonian, M., and Chapman, D. J., Eds.), pp. 608-616, Jones & Bartlett, Boston.
5. Preisig, H. R., and Melkonian, M. (1984) *Plant Syst. Evol.* **146**, 57-74.
6. Round, F. E. (1980) *Biosystems* **12**, 61-69.
7. Mattox, K. R., and Stewart, K. D. (1984) in *Systematics of the Green Algae* (Irvine, D. E. G., and John, D. M., Eds.), pp. 29-72, Academic Press, London.
8. Conner, T. W., Thompson, M. D., and Silflow, C. (1989) *Gene* **84**, 345-358.
9. Melkonian, M., and Surek, B. (1995) *Bull. Soc. Zool. Fr.* **120**, 191-208.
10. Antaramian, A., Corla, R., Ramírez, J., and González-Halphen, D. (1996) *Biochim. Biophys. Acta* **1273**, 198-202.
11. Atteia, A., Dreyfus, G., and González-Halphen, D. (1997) *Biochim. Biophys. Acta* **1320**, 275-284.
12. Trumppower, B. L., and Gennis, R. B. (1994) *Annu. Rev. Biochem.* **63**, 675-716.
13. González-Halphen, D., Lindorfer, M. A., and Capaldi, R. A. (1988) *Biochemistry* **27**, 7021-7031.
14. Akiba, T., Toyoshima, C., Matsunaga, T., Kawamoto, M., Kubota, T., Fukuyama, K., Namba, K., and Matsubara, H. (1996) *Nature Struct. Biol.* **3**, 553-561.
15. Saraste, M. (1984) *FEBS Lett.* **166**, 367-372.
16. Widger, W. R., Cramer, W. A., Herrman, R. G., and Trebst, A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 674-678.
17. Crofts, A. R., Robinson, H., Andrews, K., Van Doren, S., and Berry, E. (1987) in *Cytochrome Systems, Molecular Biology and Bioenergetics* (Papa, S., Chance, B., and Ernster, L., Eds.), pp. 617-624, Plenum, New York.
18. Yun, C. H., Van Doren, S. R., Crofts, A. R., and Gennis, R. B. (1991) *J. Biol. Chem.* **266**, 10967-10973.
19. Yun, C. H., Crofts, A. R., and Gennis, R. B. (1991) *Biochemistry* **30**, 6747-6754.
20. Xia, D., Yu, C. A., Kim, H., Xia, J. Z., Kachurin, A., Zhang, L., Yu, L., and Delsenhofer, J. (1997) *Science* **277**, 60-66.
21. Mitchell, P. (1976) *J. Theor. Biol.* **62**, 327-367.
22. Trumppower, B. L. (1981) *J. Bioenerg. Biomembr.* **13**, 1-24.
23. de Vries, S., Albracht, S. P. J., Berden, J. A., and Slater, E. C. (1981) *J. Biol. Chem.* **256**, 11996-11998.
24. Robertson, D. E., and Dutton, P. L. (1988) *Biochim. Biophys. Acta* **935**, 273-291.
25. Howell, N., and Robertson, D. E. (1993) *Biochemistry* **32**, 11162-11172.
26. Gutiérrez-Cirlos, E. B., Antaramian, A., Vázquez-Acevedo, M., Corla, R., and González-Halphen, D. (1994) *J. Biol. Chem.* **269**, 9147-9154.
27. Cantor, M. H., and Burton, M. D. (1975) *J. Protozool.* **22**, 135-139.
28. Michaelis, C., Vahrenholz, C., and Pratje, E. (1990) *Mol. Gen. Genet.* **223**, 211-216.

TESIS CON
FALLA DE ORIGEN

29. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
30. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31. Attela, A. (1994) *C. R. Acad. Sci. Ser. III* **317**, 11-19.
32. Attela, A., and Franzén, L.-G. (1996) *Eur. J. Biochem.* **237**, 792-799.
33. Chua, N. H., Matlin, K., and Bennoun, P. (1975) *J. Cell Biol.* **67**, 361-377.
34. Schägger, H., Aquila, H., and von Jagow, G. (1988) *Anal. Biochem.* **173**, 201-205.
35. Vázquez-Acevedo, M., Antaramian, A., Corona, N., and González-Halphen, D. (1993) *J. Bioenerg. Biomembr.* **25**, 401-410.
36. Schägger, H., Link, T. A., Engel, W. D., and von Jagow, G. (1986) *Methods Enzymol.* **126**, 224-237.
37. Matsudaira, P. (1987) *J. Biol. Chem.* **262**, 10035-10038.
38. Devereux, J., Haerberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395.
39. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge Univ. Press, Cambridge, UK.
40. Reeck, G. R., de Haen, C., Teller, D. C., Doolittle, R. F., Fitch, W. M., Dickerson, R. E., Chambon, P., McLachlan, A. D., Margoliash, E., Jukes, T. H., and Zuckerkand, E. (1987) *Cell* **50**, 667.
41. Degli Esposti, M., de Vries, S., Crimi, M., Ghelli, A., Partanello, T., and Meyer, A. (1993) *Biochim. Biophys. Acta* **1143**, 243-271.
42. Ma, D. P., Yang, Y. W., King, T. Y., and Hasnain, S. E. (1990) *Plant Mol. Biol.* **15**, 357-359.
43. Nobrega, F. G., and Tzagoloff, A. (1980) *J. Biol. Chem.* **255**, 9828-9837.
44. Zimmer, M., Wesler, F., Oraler, G., and Wolf, K. (1987) *Curr. Genet.* **12**, 329-336.
45. Waring, R., Davies, R. W., Lee, S., Crisi, E., McPhail Berks, M., and Scazzocchio, C. (1981) *Cell* **27**, 4-11.
46. Burke, J. M., Breitenberger, C., Heckman, J. E., Dujon, B., and RajBhandary, U. L. (1984) *J. Biol. Chem.* **259**, 504-511.
47. Colleaux, L., Michel-Wolwertz, M. R., Matagne, R. F., and Dujon, B. (1990) *Mol. Gen. Genet.* **223**, 288-296.
48. Howell, N. (1989) *J. Mol. Evol.* **29**, 157-169.
49. Irwin, D. M., Kocher, T. D., and Wilson, A. C. (1991) *J. Mol. Evol.* **32**, 128-144.
50. Mathews, F. S. (1985) *Prog. Biophys. Mol. Biol.* **45**, 1-56.
51. Link, T. A., Wallmeier, H., and von Jagow, G. (1994) *Biochem. Soc. Trans.* **22**, 197-203.
52. Bennoun, P., Delosme, M., Godehart, I., and Kück, U. (1992) *Mol. Gen. Genet.* **234**, 147-154.
53. Gouterman, M. (1961) *J. Mol. Spectrosc.* **6**, 138-163.

TESIS CON
FALLA DE ORIGEN

Apéndice VI.

Artículo en prensa

**Transfer of mitochondrial genes to the nucleus in
chlamydomonad algae: perspectives for the allotopic
expression of OX-PHOS proteins and future human therapies.**

Transfer of mitochondrial genes to the nucleus in chlamydomonad algae: perspectives for the allotopic expression of OX-PHOS proteins and future human therapies.

Soledad Funes¹, Xochitl Pérez-Martínez¹, Anaid Antaramian¹, Miriam Vázquez-Acevedo¹, Robert van Lis¹, A. Reyes-Prieto¹, J.L. Santillán-Torres¹, M.Gonzalo Claros², Edgar Davidson³, Michael P. King³ and Diego González-Halphen^{1*}

1. Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, México D.F. (Mexico)
2. Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, E-29071 (Spain).
3. Department of Biochemistry and Molecular Pharmacology, Thomas Jefferson University, Philadelphia, PA. (USA).

(*) To whom correspondence should be addressed: Departamento de Genética Molecular, Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, Apartado Postal 70-243, 04510 México D.F. Tel. (52-55) 5622-5620; Fax (52-55) 5622-5611. E-mail dhalphen@ifisiol.unam.mx

Abbreviations

<H>	local hydrophobicity
<i>mesoH</i>	mesohydrophobicity
mtDNA	mitochondrial DNA
MTS	mitochondrial targeting sequence MTS
nt	nucleotides
TIM-TOM	translocases of the mitochondrial inner and outer
<i>atp6</i>	gene encoding subunit <i>a</i> (ATP6) of the F ₁ F ₀ -ATP synthase.
<i>cox2</i>	gene encoding subunit II (COXII) of cytochrome <i>c</i> oxidase.
<i>cox3</i>	gene encoding subunit III (COXIII) of cytochrome <i>c</i> oxidase.

ABSTRACT

Mitochondrial genomes encode a limited set of polypeptides that are components of the membrane-embedded oligomeric complexes that participate in oxidative phosphorylation (OX-PHOS). It has been proposed that the genes that remained localized in the mitochondrial genome are those that encode highly-hydrophobic polytopic proteins, containing two or more helices that span the membrane. The 15.8 kb linear mitochondrial genome (mtDNA) of the green alga *Chlamydomonas reinhardtii* lacks several of the genes, encoding essential components of OX-PHOS, that are typically found on mtDNA, including *cox2*, *cox3* and *atp6*. These genes were transferred to the nucleus of this alga where they were modified to permit their successful expression in the nucleus of *C. reinhardtii*. In an extreme example, the *cox2* gene, encoding the COXII subunit, has been split into two genes whose products, COX IIA and COX IIB are predicted to form a heterodimeric subunit. The predicted COX IIA, COX III, and ATP6 polypeptides have long putative mitochondrial targeting sequences (MTS) and exhibit diminished physicochemical constraints for import into mitochondria. The mean hydrophobicity of the nucleus-encoded proteins is diminished, particularly in those transmembrane stretches that are thought not to participate directly in proton translocation or in inter-subunit contacts, as shown for the ATP6 and COXIII subunits. Based on these results, we suggest some conclusions on the allotopic expression of human mitochondrial genes and on the design of mitochondrial gene therapy strategies.

1. INTRODUCTION

In some members of the chlorophyte algae from the family Chlamydomonadaceae, several genes that are normally found in the mitochondrial genomes, like *cox2*, *cox3*, and *atp6*, naturally reside in the nucleus. The study of these genes, and the structural changes that allowed their successful relocation to the nucleus, are the subject of this review. To illustrate this transfer process, we first briefly address the large diversity of mitochondrial genomes in nature, and in particular, those of Chlorophycean algae. We then summarize the evidence for the prevalent and on-going transfer of mitochondrial genes to the nucleus, review this process in chlamydomonad algae, and address why some genes have remained in mitochondrial genomes. We suggest that nuclear genes encode mitochondrial membrane proteins whose overall hydrophobicity has decreased as compared to its mitochondria-encoded counterparts. We also show that the hydrophobicity has decreased preferentially in those transmembrane regions of the protein that seem not to be critical for function or for inter-subunit interactions. Finally, we suggest some conclusions for consideration when designing mitochondrial gene therapy strategies involving allotopic expression of mitochondrial genes.

2. THE LARGE DIVERSITY OF MITOCHONDRIAL GENOMES IN NATURE.

Mitochondria are believed to have evolved from endosymbionts [1,2] that derived originally from free-living α -proteobacteria [3,4,5,6], related to the contemporary members of the genus *Rickettsia* [7]. The nature of the host with whom the endosymbiosis was established remains a matter of debate, and it is thought to be either an archeon or an amitochondrial eukaryote [8]. It is believed that the endosymbiotic event that gave rise to mitochondria occurred only once in evolution [9], and that it was followed by a massive transfer of genes to the nucleus, followed by a more gradual transfer [10] that seems to be an ongoing process [11]. This long-

term migration process gave rise to the present highly reduced mtDNAs.

Mitochondrial genomes from different species vary in size as well as in organization. In general, they encode ribosomal RNAs, tRNAs, and a limited set of polypeptide subunits of OX-PHOS proton translocating complexes, plus a variable set of proteins involved in mitochondrial protein synthesis [5]. The majority of mitochondrial proteins are nucleus-encoded, translated in the cytoplasm, and translocated through a specialized import-machinery known as the TOM-TIM complex [12]. In yeast, more than 400 mitochondrial proteins (97% of all proteins required for mitochondrial function) are encoded in the nucleus [13]. Many proteins imported into mitochondria contain a mitochondrial targeting sequence (MTS), generally a small, cleavable presequence of 20 to 40 amino acids, capable of forming an amphiphilic α -helix, that is recognized by the mitochondrial TOM-TIM machinery.

A wide spectrum of gene content of mitochondrial genomes is found in nature. The 60 kb mtDNA of the flagellate *Reclinomonas americana* [14], "the mitochondrion that time forgot" [15], resembles a highly reduced bacterial genome. This mtDNA encodes 24 proteins that participate in OX-PHOS, plus a set of 38 additional proteins involved in translation, transcription, protein import, and maturation. The *R. americana* mitochondrial genome is thought to be the extant mtDNA that most closely resembles the proto-mitochondrial genome. At the other end of the mtDNA complexity-spectrum, there is the highly reduced 6 kb mtDNA of the apicomplexan parasite *Plasmodium falciparum*, that contains only 3 genes encoding components of the mitochondrial respiratory chain - *cob*, *cox1*, and *cox3* [16,17]. *cob* and *cox1* are present in all mitochondrial genomes known to date. Their gene products are highly hydrophobic polypeptide proteins that function as central components of proton translocating complexes: cytochrome *b* of the *bc₁* complex, and subunit I of cytochrome *c* oxidase. In the small circular mitochondrial genomes of metazoans, including humans, 13 protein-coding genes, all encoding components of oxidative phosphorylation are present: *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *cob*, *cox1*, *cox2*, *cox3*, *atp6*, *atp8* [18].

3. THE MITOCHONDRIAL GENOMES OF CHLOROPHYCEAN ALGAE.

The class Chlorophyceae, members of the phylum Chlorophyta (green algae), contains more than 355 genera and 2650 species. Most live in freshwater but there are also marine and terrestrial species [19]. The mitochondrial genomes of 7 species of chlorophycean algae have been sequenced, *Chlamydomonas eugametos* [20], *Scenedesmus obliquus* [21,22], *Chlorogonium elongatum* [23], *Chlamydomonas reinhardtii* [24,25], *Nephroselmis olivacea* [26], *Pedinomonas minor* [26], and *Prototheca wickerhamii* [27]. Table 1 summarizes the gene content of the mitochondrial genomes of these green algae. The 15.8 kb linear mtDNA from *C. reinhardtii* lacks several genes encoding essential components of OX-PHOS: *nad3*, *nad4L*, *cox2*, *cox3*, *atp6*, and *atp8*, that until recently had only been characterized as mtDNA-encoded genes. These six genes are also absent from the circular genomes of the related green algae *C. eugametos* and *C. elongatum*. It was hypothesized that these genes had been transferred to the nucleus in these species. The products of two of these genes, subunit III of cytochrome *c* oxidase (COX III) and subunit 6 of F₁F₀-ATP synthase (ATP6, equivalent to subunit *a* in *Escherichia coli*), are usually highly-hydrophobic polypeptides, with 7 [28] and 5 [29,30] transmembrane stretches respectively. As outlined below, we have demonstrated that functional *atp6*, *cox2*, and *cox3* genes are nuclear localized in *C. reinhardtii* [31,32,33]. The modifications that enabled these once mitochondrial genes to become nuclear localized and fully functional offers insights into the evolutionary processes involved, and suggest possible strategies for the allotopic expression of proteins as potential genetic therapies for human mitochondrial disorders.

Table 1. Gene content in the mitochondrial genomes of Chlorophycean algae.

Genes	TR	<i>P. wi</i>	<i>N. ol</i>	<i>P. mi</i>	<i>S. ob</i>	<i>C. re</i>	<i>C. eu</i>	<i>C. el</i>
COMPLEX I								
<i>nad1</i>	8-9	+	+	+	+	+	+	+
<i>nad2</i>	13-14	+	+	+	+	+	+	+
<i>nad3</i>	3	+	+	+	+	-	-	-
<i>nad4</i>	13-14	+	+	+	+	+	+	+
<i>nad4L</i>	3	+	+	+	+	-	-	-
<i>nad5</i>	15-16	+	+	+	+	+	+	+
<i>nad6</i>	5	+	+	+	+	+	+	+
<i>nad7</i>	1-2	+	+	-	-	-	-	-
<i>nad8</i>	1	-	+	-	-	-	-	-
<i>nad9</i>	1	+	+	-	-	-	-	-
COMPLEX III								
<i>cob</i>	9	+	+	+	+	+	+	+
COMPLEX IV								
<i>cox1</i>	12	+	+	+	+	+	+	+
<i>cox2</i>	2	+	+	-	+(<i>cox2a</i>)	-	-	-
<i>cox3</i>	7	+	+	-	+	-	-	-
COMPLEX V								
<i>atp1</i>	0	+	+	-	-	-	-	-
<i>atp6</i>	5	+	+	+	+	-	-	-
<i>atp8</i>	2	+	+	+	-	-	-	-
<i>atp9</i>	2	+	+	-	+	-	-	-
rib RNAs								
<i>rnl</i>		+	+	+(2)	+(4)	+(8)	+(6)	+(6)
<i>rns</i>		+	+	+	+(2)	+(4)	+(3)	+(3)
5S		+	+	-	-	-	-	-
tRNAs		26	26	8	27	3	3	3

TR= transmembrane regions of the protein encoded, *P. wi*= *Prototheca wickerhamii*, *N. ol*=*Nephroselmis olivacea*, *P. mi*= *Pedinomonas minor*, *S. Ob*= *Scenedesmus obliquus*, *C. re*= *Chlamydomonas reinhardtii*, *C. eu*= *Chlamydomonas eugametos*, *C. el*= *Chlorogonium elongatum*. Numbers in parenthesis indicate the number of fragments that constitute the ribosomal RNAs

4. TRANSFER OF MITOCHONDRIAL GENES TO THE NUCLEUS.

The theory of mitochondrial origin proposes that there was a gradual transfer of genes from the endosymbiont to the nucleus [10]. This transfer of genetic material may have happened in the form of DNA or of cDNA [34]. In *S. cerevisiae*, the current rate of transfer of DNA from the mitochondria to the nucleus is 10^5 times more frequent than the rate of transfer in the opposite direction [35]. This suggests a favored unidirectional flux of genetic material from organelles to nuclear control. Transfers of mtDNA to the nucleus can involve fragments ranging from 31 nt [36] up to 620 kb, as in the case of *Arabidopsis thaliana*, where a complete mtDNA copy was found in the nucleus [37]. The human genome project has also revealed transfer of large mitochondrial fragments to the nucleus, ranging from 106 to 14,654 bp [38].

The continuous transfer of organellar DNA to the nucleus predicts that all coding sequences should eventually be displaced from the mitochondrion [39]. However, this transfer is not always successful, since on many occasions genes do not establish themselves functionally in the nucleus. Numerous copies of mitochondrial genes reside in the nucleus as pseudogenes in over 64 animal species [40,41]. These pseudogenes are considered relics of anciently transferred mtDNA that remain as molecular fossils in the nucleus [42]. Mitochondrial DNA has been transferred to the nucleus on numerous independent occasions, and the same gene can be sent to the nucleus at different times in closely related organisms [9]. These multiple transfers sometimes resulted in a successful integration of the genes and their establishment as functional genes. In other occasions, after integration, the genes were not activated, and remained as pseudogenes.

Transfer of mitochondrial genes to the nucleus seems to be an ongoing process [43], exemplified by the presence of similar genes for the same protein encoded in both the mitochondrial and the nuclear genomes of an organism, i.e. F_1F_0 -ATP synthase subunit 9 (ATP9) of *Neurospora crassa* [44,45] and cytochrome *c* oxidase subunit II (COX II) of some higher plants [11,46]. The functional relocation of mitochondrial genes to the nucleus required several steps [39,47]: *i*) The transfer of a gene to the nucleus, while an active copy is still retained in the mitochondria. *ii*) The activation of the nuclear copy of the gene by acquisition of sequences encoding nuclear promoters, ribosome binding sites, mitochondrial targeting sequences (MTS), and polyadenylation signals. These structural transformations may also be accompanied by changes in codon usage and the acquisition of introns. The relocalization of mitochondrial genes to the nucleus implies the coexistence of active nuclear and mitochondrial genes during a variable period of time, as described for the *cox2* gene in some legumes [48]. *iii*) Inactivation of the mitochondrial gene. Such is the case of the mitochondrial ribosomal protein S14 in rice, that has been transferred to the nucleus, while the original mitochondrial copy has been interrupted by stop codons [49]. Ultimately, the original mitochondrial gene will be eliminated. Alternatively, the nuclear copy may be the one to inactivate, resulting in the original gene being retained in the mitochondrion and the appearance of mtDNA pseudogenes in the nucleus [50].

The transfer of mitochondrial genes to the nucleus may confer a selective advantage, since nuclear genes in some organisms exhibit a lower mutation rate [51]. This is not necessarily true of fungi, where mutation rates of the nuclear and mitochondrial genes are almost equivalent, or in plants, where nuclear genes mutate at a higher rate than mitochondrial genes [52,53,54]. Gene transfer from organelles to the nucleus is also thought to reduce the accumulation of deleterious mutations, since the genes migrate from a predominantly asexual to a predominantly sexual genome [55,56].

5. TRANSFER OF MITOCHONDRIAL GENES TO THE NUCLEUS IN CHLAMYDOMONAD ALGAE.

Since the genes *cox2*, *cox3*, and *atp6* were absent in the mtDNA of *C. reinhardtii*, we investigated whether these genes were expressed in the nucleus. We used the standard techniques of molecular biology to address the presence of these genes in *C. reinhardtii* combined with a biochemical approach to show the presence of the corresponding proteins in the mitochondrion. For this second approach, we also used the colorless alga *Polytomella* sp. as it lacked a cell wall and functional chloroplasts [57,58]. These characteristics allowed the isolation of algal mitochondria free of thylakoid contaminants which facilitated the purification of several OX-PHOS complexes [31,59,60,61]. The colorless algae of the genus *Polytomella* are closely related to *C. reinhardtii*, as demonstrated by the similarities between their nucleus-encoded beta-tubulin [62] and 18S rRNA [58,63], and mtDNA-encoded *cox1* genes [64].

The combined studies with *C. reinhardtii* and *Polytomella* sp. demonstrated that the gene *cox3*, encoding subunit III of cytochrome *c* oxidase (COX III), had been transferred to the nucleus in both organisms [31]. We also found that in both *Polytomella* sp. and *C. reinhardtii* subunit II of cytochrome *c* oxidase (COX II) is encoded by two separate nuclear genes that were named *cox2a* and *cox2b*. The *cox2a* gene encodes a protein, COX IIA, corresponding to the amino terminal half of a typical single-polypeptide COX II, that includes the two transmembrane stretches. The *cox2b* gene encodes a protein, COX IIB, equivalent to the soluble C-terminal domain of an orthodox COX II subunit. We proposed that the separate *cox2a* and *cox2b* genes gave rise to a heterodimeric COX II that resulted from the non-covalent assembly of the COX IIA and COX IIB polypeptides in the mature cytochrome *c* oxidase complex [32]. This contrasts with the COX II proteins of other eukaryotes, that are single polypeptides encoded by single genes normally localized in the mitochondrial genome, with the exception of full-length nuclear *cox2* genes in the nuclei of some legumes. In addition, the gene *atp6* encoding subunit ATP6, an essential component of the proton translocating F_0 sector of the F_1F_0 -ATP synthase, has also been transferred to the nucleus in *C. reinhardtii* [33]. Therefore, the four genes *cox2a*, *cox2b*, *cox3*, and *atp6* were shown to reside in the nucleus in at least some members of the family Chlamydomonadaceae, in contrast to the mitochondrial location of the genes *cox2*, *cox3* and *atp6* in the vast majority of eukaryotes. The transfer of these genes probably occurred late in evolution, after the massive transfer of genes from the protomitochondrion to the nucleus [5], since nowadays many green algae still retain the *cox2*, *cox3*, and *atp6* genes in their mitochondrial genomes.

The discovery of these genes in *C. reinhardtii* and *Polytomella* sp. allows a consideration of the specific features that accompanied their transfer from the mitochondrion to the nucleus:

- a) *The export of the nucleic acid molecule from the mitochondria in the form of DNA or RNA.* Some genetic material has been transferred to the nucleus via RNA intermediates, shown by the presence of edited versions of plant mitochondrial genes in the nucleus, where the mRNA has a change of C→U when compared to the genomic sequence. [46,48,65,66]. This transfer as RNA must have been followed by a reverse transcription step, before its integration into the nuclear genome. It remains to be ascertained if the genes *atp6*, *cox2a*, *cox2b*, and *cox3* of chlamydomonads were transferred as DNA or RNA intermediates. Mitochondrial editing has not been found in algae, and no evidence for reverse transcription activity has been found, despite the presence of a gene encoding a putative reverse transcriptase-like protein in the mitochondrial genome of *C. reinhardtii* [24]. The absence of editing suggests that the integration of an RNA-derived DNA

fragment was not obligatory, as would be the case in an organism that required RNA editing for functional gene expression.

- b) *Integration into the nucleus.* Integration into the nuclear genome most probably occurred in a non-coding region through nonhomologous recombination [67], or by an end-joining mechanism [36].
- c) *Acquisition of an MTS.* Random genomic DNA sequences preceding mitochondrial genes that have been relocalized in the nucleus could provide an MTS. Two to five % of randomly sheared DNA sequences attached upstream of protein genes *in vitro* functioned as effective MTS [68], showing the potential ease of gaining of an MTS. In some cases, mitochondrial genes have inserted into nuclear genes, acquiring the pre-existing MTS [48,69]. MTS acquisition may also occur by duplication of existing targeting signals [70] or by nuclear exon shuffling [71]. The deduced amino acid sequence located before the N-terminal sequence of the mature COX IIA, COX III and ATP6 proteins of *C. reinhardtii* predicts the existence of unusually large MTSs, of 143, 119 and 107 amino acids respectively, rich in alanines, prolines, and charged amino acids. In yeast, it has been observed that the duplication of MTSs improves the *in vitro* and *in vivo* import of hydrophobic proteins into mitochondria [72]. It has also been suggested that long MTSs can improve the interaction of the precursor with the mitochondrial import machinery [73]. Alternatively, a long MTS could affect the folding of the protein to increase its importability [74]. It is possible that, during import into mitochondria, the targeting sequences are cleaved and maintained as components of the cytochrome *c* oxidase complex, as was observed with the MTSs of the Rieske subunit of yeast and beef heart mitochondrial cytochrome *bc*₁ complexes [75,76]. The function of the targeting sequence retained in the *bc*₁ complex is not known, we speculate that it may act as a chaperone for components of the enzyme complex during assembly. A conventional N-terminal MTS is not absolutely required for a nucleus-encoded protein to be imported into the mitochondrion. Several mitochondrial genes that were transferred to the nucleus have become activated, and their protein products may be imported into mitochondria in the absence of an MTS [50]. This is also the case for the *cox2b* genes of *C. reinhardtii* and *Polytomella* sp. which do not exhibit a region encoding a putative MTS.
- d) *Acquisition of introns, promoters, and ribosome binding sites.* Most nuclear genes in *C. reinhardtii* have introns, with an average number of 3.9 introns per kb of coding sequence. These introns are often small, ranging from 57 bp to 1318 bp, with an average size of 219 bp [77]. We found introns in the genomic sequences of the chlamydomonad mitochondrial genes that were transferred to the nucleus: 6 in *cox2a* of *Polytomella* sp., 7 in *cox2a* of *C. reinhardtii*, 1 in *cox2b* of *C. reinhardtii*, 4 in *cox3* of *Polytomella* sp., 9 in *cox3* of *C. reinhardtii*, and 7 in *atp6* of *C. reinhardtii*. These introns show orthodox splicing sites exhibiting the typical GT sequence at the 5' end and an AG sequence at the 3' end. An exception was *Polytomella* sp. *cox2b* that did not contain introns. The sequences flanking the proposed open reading frame initiating methionine codons corresponded to the consensus translation initiation site (A/C) A (A/C) (A/C) ATG (G/C) reported for *C. reinhardtii* [77]. We have yet to identify promoter regions in the genomic regions upstream from these genes.
- e) *Acquisition of polyadenylation signals.* In *C. reinhardtii* the most common polyadenylation signal in the nuclear genes is TGTA located 10-20 bp upstream of the actual polyadenylation site [77]. This signal was present in the *cox3*, *cox2a*, *cox2b*, and *atp6* genes that were transferred to the nucleus in chlamydomonad algae. These signals are distinctive features of nuclear genes, since they are not present in mitochondrial

genes.

- f) *Change in codon usage.* The *C. reinhardtii* mitochondrial genetic code is the same as the standard genetic code utilized in the nucleus, which may facilitate a continued transfer of mitochondrial genes to the nucleus. However, the chlamydomonad algae exhibit a highly biased codon usage in their nuclear genes. The nuclear genome of *C. reinhardtii* has a high GC content, and this feature is reflected in the pronounced codon bias, which favors triplets with C or G in the third position. In contrast, there is a different bias in the codon usage in the mitochondrial genome. Relocalization of mitochondrial genes to the nucleus in *C. reinhardtii* was followed by changes in codon usage to frequencies typically found for nuclear genes.
- g) *Inactivation of the mitochondrial gene copy, and loss of the original mitochondrial gene.* Successful transfer of genes from organelles to the nucleus is usually followed by inactivation of the mitochondrial copy, its conversion into a pseudogene, and its eventual loss from the mitochondrial genome [78]. There is no evidence for the presence of *cox2a*, *cox2b*, *cox3* or *atp6* genes, gene fragments, or pseudogenes in the mitochondrial genomes of several chlamydomonads, including *Polytomella* sp., *C. reinhardtii*, *C. eugametos*, *C. elongatum* and *C. moewusii*.
- h) *The splitting of the cox2 gene into cox2a and cox2b.* The gene encoding the mitochondrial COX II protein is absent from the mtDNA of a number of chlamydomonad algae. We found that the gene had been split into two parts, *cox2a* and *cox2b*, both expressed from the nucleus, in *Polytomella* sp. and *C. reinhardtii* [32]. We also observed that a *cox2* gene in the mtDNA of *Scenedesmus obliquus* shows strong similarities to the chlamydomonad *cox2a* gene, suggesting that the splitting of *cox2* occurred prior to the transfer of the gene to the nucleus, and that the chlamydomonad transfer of *cox2b* to the nucleus is likely to have preceded that of *cox2a*. It is not known if the *S. obliquus* mtDNA-encoded *cox2a* is functional, or if there is also a nuclear copy of the same gene. It is possible that the *S. obliquus cox2a* was prevented from relocating to the nucleus by the divergence of the mitochondrial genetic code from the standard code in this organism [21,22]. In *Polytomella* sp. and *C. reinhardtii* the COX IIA protein contains a C-terminal 20-amino acid region, lacking similarity to conventional COX II proteins, that had a high density of charged amino acids. The predicted COX IIB polypeptide contains 42-amino acids at the N-terminus with a high density of charged amino acids that are not homologous to known COX II proteins and are not a cleavable MTS. We propose that the C-terminal extension of COX IIA interacts with the N-terminal extension of the COX IIB protein and that these acquired amino acid sequences stabilize the two COX II subunits in the cytochrome *c* oxidase complexes.

6. WHY HAVE SOME GENES REMAINED IN THE mtDNA?

Several ideas have been put forward to explain why mitochondrial genomes still contain a limited set of genes that has not been transferred to the nucleus. One explanation has been the evolution of a different mitochondrial genetic code in some organisms that would inhibit the functional expression of mitochondrial genes transferred to the nucleus. Similarly, some mitochondrial genes have accumulated complex processing patterns like mRNA-editing, which would render the transferred gene inviable when relocated to the nucleus [79]. Another explanation may be that the presence of some organellar proteins in the cytoplasm could have detrimental effects [80], including misrouting of certain highly hydrophobic mitochondrial proteins synthesized in the cytosol to other cell structures, such as the endoplasmic reticulum

[81]. An additional explanation suggests that some genes have remained in the mitochondrial genome to be rapidly regulated by the organelle redox state [82]. Finally, it has been proposed that highly hydrophobic proteins, containing four or more helices that span the membrane, can not be readily imported into mitochondria. Therefore, these polytopic membrane proteins must be synthesized *in situ* to be properly inserted and assembled into the inner mitochondrial membrane [74,83]. The synthesis of these hydrophobic polypeptides inside the mitochondria, may ensure their proper insertion in the inner membrane, giving rise to the correct topological arrangement required for vectorial proton translocation.

There are two universal examples of genes retained in the mitochondrial genomes: the cytochrome *b* gene (*cob*), encoding a protein with 8 transmembrane regions [84], and the cytochrome *c* oxidase subunit I (*cox1*), which encodes a protein with 12 transmembrane helices [28,76]. These subunits also bind prosthetic groups, including heme groups and metal ions. Both genes are present in all mitochondrial genomes so far characterized. Other genes that encode highly hydrophobic polypeptides (*cox2*, *cox3*, *atp6*, *atp8*, *atp9*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*) are also present in the majority of mtDNAs [85]. In some yeasts, the absence of *nad* genes in the mitochondrial genome is related to the lack of respiratory complex I. Those organisms that do contain complex I retain a set of six to seven *nad* genes. The corresponding proteins of the genes retained in mtDNAs exhibit at least two transmembrane helices, and some of them up to 17 putative transmembrane stretches. Therefore, it may be that physicochemical properties (mainly hydrophobicity) are the ultimate limiting step for the transfer of mitochondrial genes, encoding polytopic membrane proteins, to the nucleus.

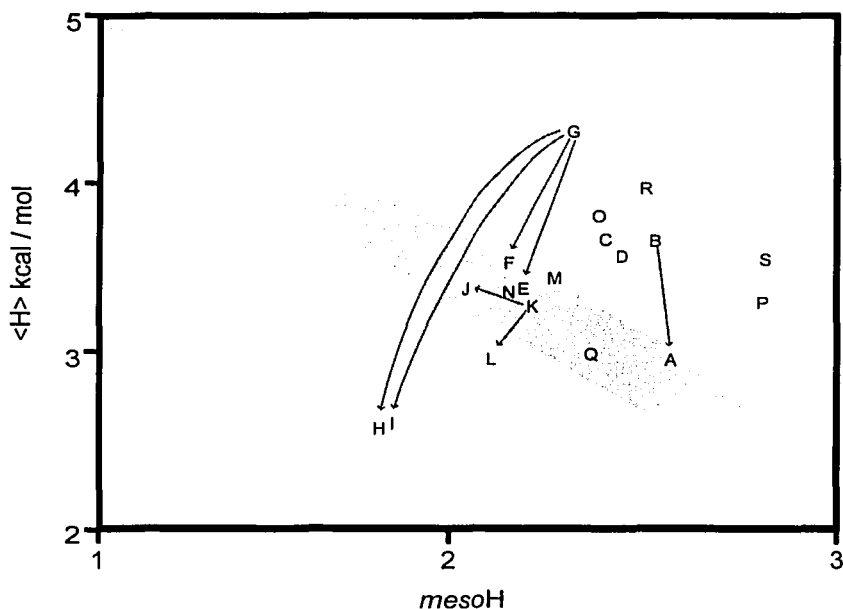
7. NUCLEAR GENES ENCODING MITOCHONDRIAL MEMBRANE PROTEINS DECREASE THEIR HYDROPHOBICITY WHEN COMPARED TO THEIR MITOCHONDRIAL COUNTERPARTS.

In yeast, *in vivo* studies with cytoplasmic synthesized constructs of variable lengths of apocytochrome *b*, showed that the importability of polypeptides into mitochondria is not strictly related to the number of transmembrane domains [74]. These studies suggested that the highest average hydrophobicity over 60 to 80 amino acids of a polypeptide chain (*mesoH*), along with the maximum hydrophobicity of the putative transmembrane segments, are useful indicators of the likelihood that a protein could be imported into mitochondria. Accordingly, mitochondrial *cox2*, *cox3* and *atp6* genes from many eucaryotes encode proteins that exhibit physicochemical characteristics that would block the import of such proteins if they were nucleus-encoded. However, the corresponding nucleus-encoded proteins of chlamydomonad algae COX IIA, COX IIB, COX III and ATP6, all exhibit reduced overall hydrophobicity that allow them to be imported into mitochondria, most probably through the TOM-TIM machinery. Figure 1 shows a mesohydrophobicity (*mesoH*) versus maximal local hydrophobicity (<H>) plot for the *C. reinhardtii* and *Polytomella* sp. COX IIA, COX IIB, COX III and ATP6 sequences as compared to the protein products predicted from the human mitochondrial genes *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *cob*, *cox1*, *cox2*, *cox3*, and *atp6*. The *atp8* gene was not included since its very small size, 68 amino acids, distorts the *mesoH* and <H> values. In comparison with their human mtDNA-encoded counterparts, the nucleus-encoded chlamydomonad sequences display both decreased *mesoH* and <H> (Figure 1). We propose that one of the important modifications required for nuclear genes to encode functional mitochondrial OX-PHOS proteins is a diminished overall mesohydrophobicity of its protein product. The changes in hydrophobicity may have occurred either prior or subsequently to the transfer process and may allow for the successful import and assembly

of these proteins into the mitochondrial inner membrane.

Mitochondria readily import hydrophobic carrier proteins with multiple transmembrane stretches, like the adenine nucleotide translocator. Nevertheless, carrier proteins also seem to follow the low $\langle H \rangle$ and low *mesoH* rule [74]. However, the import pathway of the translocators differs greatly from the "conservative intramitochondrial sorting pathway", in which polypeptides are transferred to the mitochondrial matrix space and then sorted to its final membrane destination. The insertion of carrier proteins into the mitochondrial inner membrane is mediated through different TIM complexes, mainly the so-called tiny-TIMS and the TIM22 complex [12]. This import pathway is distinct from the one followed by orthodox MTS-containing polypeptides.

TESIS CON
FALLA DE ORIGEN



Scale: PRIFT

Scanning Window: 13

Increasing windows from 60 to 80

Figure 1. Plot of mesohydrophobicity (*mesoH*) versus maximal local hydrophobicity ($\langle H \rangle$) of the OX-PHOS mitochondrial proteins encoded in the human mtDNA compared to the nucleus-encoded homologs COX II, COX III and ATP6 of *Chlamydomonas* and *Polytomella*. Arrows indicate the mitochondrial and nuclear counterparts. Letters indicate the following sequences: A. *C.reinhardtii* ATP6, B. human ATP6, C. human COB (cytochrome *b*), D. human COX I, E. *C.reinhardtii* COX IIA, F. *Polytomella* sp. COX IIA, G. human COX II, H. *C.reinhardtii* COX IIB, I. *Polytomella* sp. COX IIB, J. *C.reinhardtii* COX III, K. human COX III, L. *Polytomella* sp. COX III., M. human NAD1, N. human NAD2, O. human NAD3, P. human NAD4L, Q. human NAD4, R. human NAD5, S. human NAD6.

8. MEAN HYDROPHOBICITY HAS DECREASED IN THOSE TRANSMEMBRANE REGIONS THAT ARE NOT CRITICAL FOR FUNCTION.

Hydropathy analysis of the COX III polypeptides of *Polytomella* sp. and *C. reinhardtii* showed the presence of seven putative transmembrane stretches, numbered I to VII. The hydrophobicity of these seven helices seems to be lower in the chlamydomonad algae when compared with the *Paracoccus denitrificans* or the bovine subunits. This was more evident when the three dimensional structure of COX III from *Polytomella* sp. was modelled using the crystallographic coordinates of the bovine subunit [28]. In this model (Figure 2), shorter transmembrane stretches are observed as well as interruptions in the middle-sections of the membrane helices. The helices that are in contact with COX I (helices I and III), do not exhibit significant modifications in hydrophobicity. However, transmembrane domains II, V and VI have diminished their hydrophobicity by 16%, 10% and 12%, respectively. It should be noted that domains IV and VII, that are the least hydrophobic in the mtDNA-encoded COX III proteins ($\Delta G \sim 1.1$ kcal/mol) offer distinct behaviour: IV remains the same and VII has increased its hydrophobicity by 22%, although the final hydrophobicity is not predicted to be a problem for import. This suggests that the diminished hydrophobicity of COX III is stronger in those regions of the protein which seem not to be involved in subunit-subunit interactions (Figure 3).

The *in silico* analysis of the predicted ATP6 sequence of *C. reinhardtii* revealed similar characteristics [33]. Based on multiple sequence comparisons, secondary structure predictions, and available biochemical data [29,30, and see also 118], five hydrophobic regions could be predicted as transmembrane stretches and were named A, B, C, D and E. Hydrophobicity analysis showed that transmembrane helices A, B and C exhibit a highly reduced mean hydrophobicity. In fact, helix A could even be considered not to be membrane-imbedded. A similar situation was observed for the transfer to the nucleus of the mitochondrial *sdh3* gene – encoding subunit 3 of succinate dehydrogenase - in angiosperms: its protein product is predicted to lack one of three transmembrane domains [86]. In contrast ATP6 transmembrane helices D and E, which are believed to interact with the multimeric ring of *c*-subunits (ATP9), and which contain most of the conserved amino acids in the protein, exhibit similar $\langle H \rangle$ values when compared with the helices of other mitochondria-encoded ATP6 subunits.

Transmembrane domains II in COX III and helix A in ATP6 are the most hydrophobic, and have decreased their $\langle H \rangle$ to the greatest extent. Therefore, we conclude that the overall decrease in hydrophobicity in the chlamydomonad subunits COX IIA, COX III, and ATP6 accompanied the functional transfer of their respective genes to the nucleus, and facilitates the import and assembly of these proteins into active cytochrome *c* oxidase and F_1F_0 -ATP synthase complexes, respectively. In addition, we propose that the selective decrease in the hydrophobicity of proteins encoded by nuclear genomes is stronger in those transmembrane regions that seem not to be critical for function, assembly, or participation in inter-subunit interactions with other constituents of the complexes. This conclusion has important implications for the allotopic expression of mitochondrial genes and for future application of allotopic expression to human mitochondrial gene therapy.

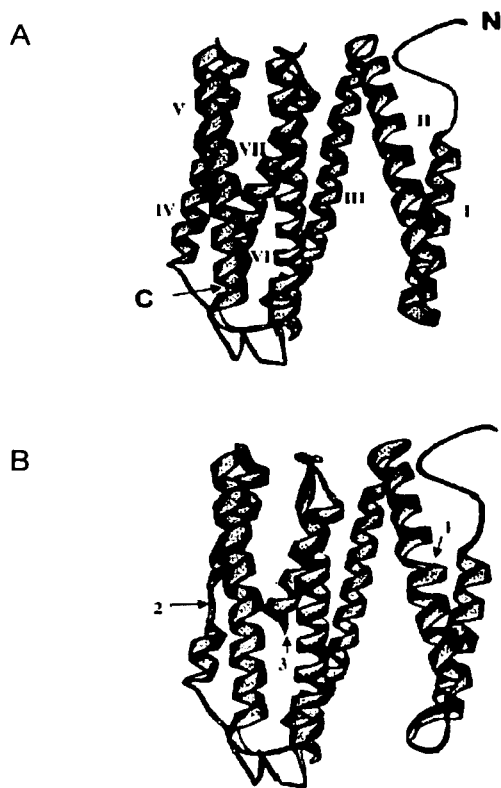


Figure 2. Model for the structural arrangement of COX III from *Polytomella* sp. **Panel A**, three dimensional structure of bovine COX III [28]. The seven transmembrane helices are indicated by roman numerals. The amino and carboxy termini are indicated by the letters N and C, respectively. **Panel B**, three dimensional model for Ps-COX III based on the bovine structure. Arrows indicate distinct features of this polypeptide: 1) the presence of a kink in transmembrane helix II, 2) a distortion in the middle-section of transmembrane helix IV, and 3) a shorter transmembrane helix VII.

TESIS CON
FALLA DE ORIGEN

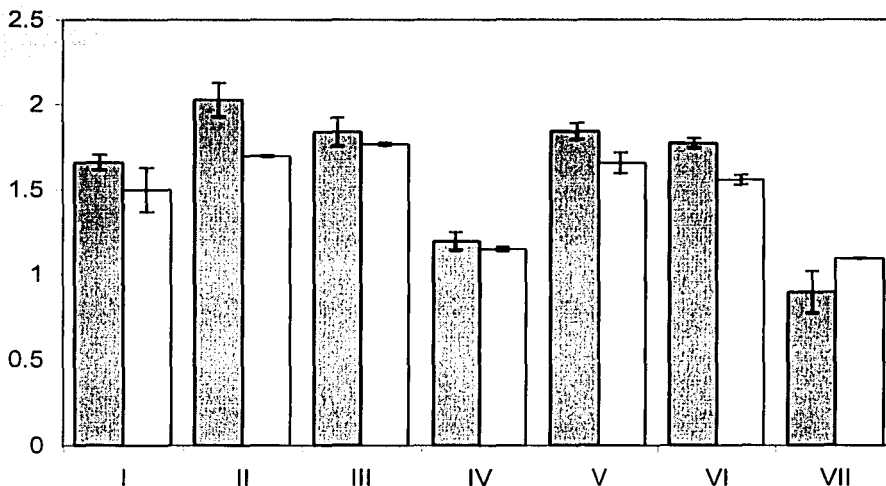


Figure 3. Analysis of the mean hydrophobicity values of the seven transmembrane regions of COX III. Hydrophobicity values for the seven transmembrane helices of COX III from *C. reinhardtii* and *Polytomella* sp. (white bars) were compared with the mean value of hydrophobicity of COX III from different organisms (grey bars). In Cr-COX III and Ps-COX III, mean hydrophobicity is reduced in transmembrane helix I (diminished by 10%, although not statistically significant), helix II (diminished by 16%), helix V (diminished by 10%) and helix VI (diminished by 12%) as compared to the mean values of $\langle H \rangle$ of plant and human transmembrane regions. Domains III and IV, although exhibiting diminished hydrophobicity by 4%, has no significant variation. The same is applicable to domain VII, whose hydrophobicity is increased by 22%, but the $\Delta G = 1.09$ kcal/mol for this transmembrane segment do not impede the import of the protein.

9. THE ALLOTOPIC EXPRESSION OF MITOCHONDRIAL GENES AND THE QUEST FOR HUMAN MITOCHONDRIAL GENE THERAPY.

The first pathogenic mutations of the human mtDNA were described in the pioneering works of Holt *et al.*, [87] and Wallace *et al.*, [88]. Since then, more than 100 unique, pathogenic mutations of mtDNA have been reported [89]. Genetically, the mtDNA mutations fall into three categories 1) mutations in protein encoding genes; 2) mutations in structural RNA genes such as tRNAs and rRNAs; and 3) large-scale rearrangements (deletions or duplications) of the mtDNA. These alterations in human mtDNA result in a broad range of clinical outcomes.

Pathogenic mutations have been described for nearly all protein encoding mitochondrial genes [90]. Perhaps the most common are those associated with Leber's hereditary optic neuropathy (LHON). LHON presents as acute or subacute bilateral visual loss caused by severe bilateral optic nerve atrophy and is associated primarily with mutations in genes for subunits of

NADH dehydrogenase [91]. Mutations in the ATP6 gene of the mtDNA are often associated with two syndromes: NARP (neurogenic muscle weakness, ataxia, and retinitis pigmentosa) [92] and maternally-inherited Leigh's syndrome (MILS; subacute necrotizing encephalomyelopathy) [93,94]. Other protein coding mutations, in the mtDNA-encoded COX subunits, or cytochrome *b* of the *bc₁* complex, present with a wide variety of clinical features, including progressive exercise intolerance, myopathy, encephalopathies, and multi-system disorders [95-102].

Several strategies to develop gene-based mitochondrial therapies for mitochondrial diseases have been reviewed [103-106]. Direct manipulation of the mtDNA is presently not possible. However DNA-protein conjugates can enter mitochondria via the protein import pathway, and chimeras of DNA attached to polypeptide presequences may be imported independently of its DNA length [107-109]. These techniques may eventually lead to the ability to correct mtDNA gene mutations. Another interesting approach for overcoming mutations in mtDNA-encoded proteins is to place a wild-type copy of the affected gene in the nucleus, and target the expressed protein to the mitochondrion to replace the defective mitochondrial protein. This approach requires the nuclear expression of genes that are normally localized in the mtDNA, their synthesis in the cytoplasm, and their successful import into the mitochondrial inner membrane.

Allotopic expression is defined as the functional activation of a gene in a cellular compartment different from its original location. Such "allotopic" expression of mitochondrial genes has been successfully performed in *S. cerevisiae* to overcome defects in mitochondrial *b14*, the RNA maturase of the yeast mitochondrial matrix [110,111]; to study the functions of VAR1, a subunit of the mitochondrial ribosome [112]; and with ATP8 (subunit A6L), a small (48 aa) hydrophobic polypeptide of the F_0 component of the mitochondrial F_1F_0 -ATP synthase [113]. The sequence of the *atp8* gene was genetically engineered for nuclear expression, and a sequence encoding a yeast MTS was artificially attached. This *atp8* gene was expressed, and its corresponding protein product was synthesized in the cytoplasm, imported into the inner mitochondrial membrane, and assembled into a functional F_0 sector of F_1F_0 -ATP synthase.

The fact that several typically mtDNA-encoded genes are nuclear-localized in some organisms, such as chlamydomonad algae, may facilitate the eventual allotopic expression of genes in human cells exhibiting mitochondrial diseases, especially those associated with defects in mitochondrial *atp6* genes [114], MILS and NARP [115], or those related with alterations in *cox* genes [95,99,100].

The allotopic expression of mitochondrial genes in the nucleus for gene therapy purposes, must be accompanied by the appropriate changes in codon usage of the transferred gene, and by the addition of an appropriate MTS. These two requirements may be necessary but not sufficient for successful allotopic expression. Human mitochondrial genes may also require additional alterations for allotopic expression. A decrease in the mean hydrophobicity of the protein product, particularly in those transmembrane stretches that are not highly conserved, and that are known to be non-critical for function may facilitate import of these highly hydrophobic proteins. Where appropriate the hydrophobicity of a protein could be diminished in the same regions where hydrophobicity has been reduced naturally in chlamydomonad algae. These approaches require a more detailed knowledge of the topology and function of the proteins of interest, more extensive site-directed mutagenesis of the allotopically expressed mitochondrial gene than previously considered, and a deep knowledge of the mechanisms of action of the coexpressed proteins.

For genes whose simple allotopic expression cannot produce an active mitochondria-localized protein, it may be plausible to co-express a gene that increases the import of such a protein. In yeast, some non-importable polypeptides can be expressed in the nucleus and

imported efficiently into mitochondria when another gene is co-expressed [73]. The nature of these genes is being elucidated [116,117]. Overexpression of yeast karyopherin Pse1p/Kap121p stimulated the mitochondrial import of hydrophobic proteins [117]. In addition, hydrophobic mitochondrial proteins seem to be imported into mitochondria by means of ribosomes attached to the mitochondrial outer membrane. This scenario is analogous to the export of proteins to the endoplasmic reticulum [M. Corral-Debrinski, personal communication].

A number of genes encoding highly hydrophobic polypeptides have been retained with remarkable consistency in the mitochondrial genomes of a large number of eukaryotes. However, certain members of the chlamydomonad algae have relocated some of those genes to the nucleus. As we have described above, several structural transformations have accompanied chlamydomonad *cox2*, *cox3* and *atp6* genes on their way to the nucleus. Two important changes are the use of relatively long MTS and the reduction of hydrophobicity of their encoded protein products. This reduction of hydrophobicity is more notable in those transmembrane regions that do not participate in proton translocation or in subunit-subunit interactions. Human therapies using allotopic expression of mtDNA-encoded genes, would typically require modifications to the genetic code, and the addition of the necessary transcriptional control signals and a region encoding an appropriate MTS. Further, one should consider reductions in the hydrophobicity of membrane-embedded helices that are not critical for function and assembly to facilitate the functional import of these proteins into mitochondria.

ACKNOWLEDGMENTS

The authors wish to acknowledge Dr. Aubrey de Grey (University of Cambridge) for the criticism and comments about this manuscript, and Dr. M. Corral-Debrinski (Ecole Normale Supérieure, Paris) for her personal communication. Work in our group is supported by grants TW01176 from Fogarty International Center, National Institutes of Health, USA; 27754N from CONACyT, Mexico; and IN202598 from DGAPA, UNAM, Mexico. SF, XPM, AA, RvL, ARP and JLST received fellowships from CONACyT and/or DGEP (UNAM) and graduate student grants from PAEP (UNAM).

REFERENCES

- [1] Margulis, L. 1970, Origin of Eukaryotic Cells, Yale University Press, New Haven, CT.
- [2] Gray, M.W. 1992, Int. Rev. Cytol. 141, 233.
- [3] Yang, D., Oyaizu, Y., Oyaizu, H., Olsen, G.J., and Woese, C.R. 1985, Proc. Natl. Acad. Sci. USA, 82, 4443.
- [4] Martin, W., and Müller, M. 1998, Nature, 392, 37.
- [5] Gray, M.W., Burger, G., and Lang, B.F. 1999, Science, 283, 1476.
- [6] Lang, B.F., Gray, M.W., and Burger, G. 1999, Annu Rev Genet., 33, 351.
- [7] Andersson, S.G.E., Zomorodipour, A., Andersson, J.O., Sicheritz-Pontén, T., Alsmark, U.C.M., Podowski, R.M., Nslund, A.K., Eriksson, A.-S., Winkler, H.H., and Kurland, C.G. 1998, Nature, 396, 133.
- [8] Kurland, C.G., and Andersson, S.G.E. 2000, Microbiol. Mol. Biol. Rev., 64, 786.
- [9] Gray, M.W. 1999, Curr. Opin. Genet. Dev., 9, 678.
- [10] Gray, M.W. 1989, Annu. Rev. Cell Biol., 5, 25.
- [11] Covello, P.S. and Gray, M.W. 1992, EMBO J., 11, 3815.

- [12] Pfanner N., Geissler A. 2001, *Nat. Rev. Mol. Cell Biol.*, 2, 339.
- [13] Hodges, P.E., McKee, A.H.Z., Davis, B.P., Pasyne, W.E.B., and Garrels, J.I. 1999, *Nucleic Acids Res.*, 27, 69.
- [14] Lang, B.F., Burger, G., O'Kelly, C.J., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M. and Gray, M.W. 1997, *Nature*, 387, 493.
- [15] Palmer, J.D. 1997, *Nature*, 387, 454.
- [16] Feagin, J.E., Werner, E., Gardner, M.J., Williamson, D.H. and Wilson, R.J. 1992, *Nucleic Acids Res.*, 20, 879.
- [17] Conway, D.J., Fanello, C., Lloyd, J.M., Al-Joubori, B.M., Baloch, A.H., Somanath, S.D., Roper, C., Oduola, A.M., Mulder, B., Povoia, M.M., Singh, B. and Thomas, A.W. 2000, *Mol. Biochem. Parasitol.*, 111, 163.
- [18] Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J., Staden, R., and Young, I.G. 1981, *Nature*, 290, 457.
- [19] van den Hoek, C., Mann, D.G., and Jahns, H.M. 1995, *Algae. An introduction to phycology*, Cambridge University Press, Cambridge, UK, 627.
- [20] Denovan-Wright, E.M., Nedelcu, A.M. and Lee, R.W. 1998, *Plant Mol. Biol.*, 36, 285.
- [21] Nedelcu, A.M., Lee, R.W., Lemieux, C., Gray, M.W. and Burger G. 2000, *Genome Res.*, 10, 819.
- [22] Kück U., Jekosch K. and Holzamer P. 2000, *Gene.*, 253, 13.
- [23] Kroymann, J. and Zetsche, K. 1998, *J. Mol. Evol.*, 47, 431.
- [24] Gray, M.W. and Boer, P.H. 1988, *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 319, 135.
- [25] Michaelis, G., Vahrenholz, C. and Pratje, E. 1990, *Mol. Gen. Genet.*, 223, 211.
- [26] Turmel, M., Lemieux, C., Burger, G., Lang, B.F., Otis, C., Plante, I., and Gray, M.W. 1999, *The Plant Cell*, 11, 1717.
- [27] Wolff, G., Plante, I., Lang, B.F., Kück, U., and Burger, G. 1994, *Gene content and genome organization. J. Mol. Biol.*, 237, 75.
- [28] Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R., and Yoshikawa, S. 1996, *Science*, 272, 1136.
- [29] Valiyaveetil, F.L. and Fillingame, R.H. 1998, *J. Biol. Chem.*, 273, 16241.
- [30] Wada, T., Long, J.C., Zhang, D. and Vik, S.B. 1999, *J. Biol. Chem.*, 274, 17353.
- [31] Pérez-Martínez, X., Vázquez-Acevedo, M., Tolkunova, E., Funes, S., Claros, M.G., Davidson, E., King, M.P. and González-Halphen D. 2000, *J. Biol. Chem.*, 275, 30144.
- [32] Pérez-Martínez, X., Antaramian, A., Vázquez-Acevedo, M., Funes, S., Tolkunova, E., d'Alayer, J., Claros, M.G., Davidson, E., King, M.P. and González-Halphen, D. 2001, *J. Biol. Chem.*, 276, 11302.
- [33] Funes, S., Davidson, E., Claros, M.G., van Lis, R., Pérez-Martínez, X., Vázquez-Acevedo, M., King, M.P. and González-Halphen, D. 2002, *J. Biol. Chem.*, 277, 6051.
- [34] Henze, K., and Martin, W. 2001, *Trends in Genet.*, 17, 383.
- [35] Thorsness, P.E., and Fox, T.D. 1990, *Nature*, 346, 376.
- [36] Blanchard, J. L., and Schmidt, W. G. 1996, *Mol. Biol. Evol.*, 13, 537.
- [37] Stupar, R.M., Lilly, J.W., Town, C.D., Cheng, Z., Kaul, S., Buell, C.R., and Jiang, J. 2001, *Proc. Natl. Acad. Sci. U S A.*, 98, 5099.
- [38] Mourier, T., Hansen, A.J., Willerslev, E., and Arctander, P. 2001, *Mol. Biol. Evol.*, 18, 1833.
- [39] Berg, O.G., and Kurland, C.G. 2000, *Mol. Biol. Evol.*, 17, 951.
- [40] Wallace, D.C., Stugard, C., Murdock, D., Schurr, T., Brown, M.D. 1997, *Proc. Natl. Acad. Sci. USA.*, 94, 14900.

- [41] Bensasson, D., Zhang, D., Hartl, D.L., and Hewitt, G.M. 2001, *Trends Ecol. Evol.*, 16, 314.
- [42] Perna, N.T., and Kocher, T.D. 1996, *Curr Biol.*, 6, 128.
- [43] Palmer, J.D. 1991, *The Molecular Biology of Plastids*, L. Bogorad and I.K. Vasil (Eds.), Academic Press, San Diego, 5.
- [44] van den Boogaart, P., Samallo, J. and Agsteribbe, E. 1982, *Nature*, 298, 187.
- [45] Sebald, W. and Hoppe, J. 1981, *Curr. Top. Bioenerg.*, 12, 1.
- [46] Nugent, J.M. and Palmer, J.D. 1991, *Cell*, 66, 473.
- [47] Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. and Schuster, W. 1993, *FEBS Lett.* 325, 140.
- [48] Adams, K.L., Song, K., Roessier, P.G., Nugent, J.M., Doyle, J.L., and Palmer, J.D. 1999, *Proc. Natl. Acad. Sci. USA*, 96, 13863.
- [49] Kubo, N., Harada, K., Hirai, A., and Kadowaki, H.-I. 1999, *Proc. Natl. Acad. Sci. USA*, 96, 9207.
- [50] Adams, K.L., Daley, D.O., Qiu, Y.-L., Whelan, J., and Palmer, J.D. 2000, *Nature*, 408, 354.
- [51] Neckelmann, N., Li, K., Wade, R.P., Shuster, R. and Wallace, D.C. 1987, *Proc. Natl. Acad. Sci. USA*, 84, 7580.
- [52] Wolfe, K.H., Li, W.-H., and Sharp, P.M. 1987, *Proc. Natl. Acad. Sci. USA*, 84, 9054.
- [53] Palmer, J.D., and Herbon, L.A. 1988, *J. Mol. Evol.*, 28, 87.
- [54] Lynch, M. 1997, *Mol. Biol. Evol.*, 14, 914.
- [55] Martin, W., and Herrmann, R.G. 1998, *Plant Physiol.*, 118, 9.
- [56] Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., and Kowallik, K.V. 1998, *Nature*, 393, 162.
- [57] Round, F.E. 1980, *Biosystems*, 12, 61.
- [58] Melkonian, M. and Surek, B. 1995, *Bull. Soc. zool. Fr.*, 120, 191.
- [59] Gutiérrez-Cirlos, E.B., Antaramian, A., Vázquez-Acevedo, M., Coria, R., and González-Halphen, D. 1994, *J. Biol. Chem.*, 269, 9147.
- [60] Atteia, A., Dreyfus, G. and González-Halphen, D. 1997, *Biochim. Biophys. Acta.* 1320, 275.
- [61] Atteia, A., van Lis, Robert, Ramírez, J., and González-Halphen, D. 2000. *Eur. J. Biochem.*, 267, 2850.
- [62] Conner, T.W., Thompson, M.D., and Silflow, C.D. 1989, *Gene*, 84, 345.
- [63] Nakayama, T., Watanabe, S., Mitsui, K., Uchida, H., and Inouye, I. 1996, *Phycol. Res.*, 44, 47.
- [64] Antaramian, A., Coria, R., Ramírez, J., and González-Halphen, D. 1996, *Biochim. Biophys. Acta.*, 1273, 198.
- [65] Blanchard, J.L., and Schmidt, G.W. 1995, *J. Mol. Evol.*, 41, 397.
- [66] Steinhäuser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. 1999, *J. Mol. Evol.*, 48, 303.
- [67] Sun, C.-W., and Callis, J. 1993, *Plant Cell*, 5, 97.
- [68] Baker, A., and Schatz, G. 1987, *Proc. Natl. Acad. Sci. U S A.*, 84, 3117.
- [69] Figueroa, P., Gómez, I., Holuigue, L., Araya, A., Jordana, X. 1999, *Plant J.*, 18, 601.
- [70] Kadowaki, K., Kubo, N., Ozawa, K., and Hirai, A. 1996, *EMBO J.*, 15, 6652.
- [71] Long, M., de Souza, S.J., Rosenberg, C., and Gilbert W. 1996, *Proc. Natl. Acad. Sci. U S A.*, 93, 7727.
- [72] Galanis, M., Devenish, R.J., and Nagley, P. 1991, *FEBS Lett.*, 282, 425.

- [73] Claros, M.G., Perea, J., and Jacq, C. 1996, *Methods Enzymol.*, 264, 389.
- [74] Claros, M.G., Perea, J., Shu, Y., Samatey, F.A., Popot, J.L. and Jacq, C. 1995, *Eur. J. Biochem.*, 228, 762.
- [75] Brandt, U., Yu, L., Yu, C.A., and Trumpower, B.L. 1993, *J. Biol. Chem.*, 268, 8387.
- [76] Iwata, S., Ostermeier, C., Ludwig, B., and Michel, H. 1995, *Nature*, 376, 660.
- [77] Silflow, C. D. 1998, *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas*, J.D.Rochaix, M. Goldschmit-Clermont, and S. Merchant (Eds.) Kluwer Academic, The Netherlands, 25.
- [78] Grohmann, L., Brennicke, A., and Schuster, W. 1992, *Nucleic Acids Res.*, 20, 5641.
- [79] Doolittle, F.W. 1998, *Trends in Genet.*, 14, 307.
- [80] Martin, W., and Schnarrenberger, C. 1997, *Curr. Genet.*, 32, 1.
- [81] von Heijne, G. 1987, *FEBS Lett.*, 198, 1.
- [82] Allen, J. F. 1993, *J. Theor. Biol.*, 165, 609.
- [83] Popot, J.-L., and de Vitry, C. 1990, *Annu. Res. Biophys. Chem.*, 19: 369.
- [84] Xia, D., Yu, C.A., Kim, H., Xia, J.Z., Kachurin, A.M., Zhang, L., Yu, L., and Deisenhofer, J. 1997, *Science*, 277, 60.
- [85] Burger, G., Plante, I., Lonergan, K.M., and Gray, M.W. 1995, *J. Mol. Biol.*, 245, 522.
- [86] Adams, K.L., Rosenblueth, M., Qiu, Y.-L., and Palmer J.D. 2001, *Genetics*, 158, 1289.
- [87] Holt, I.J., Harding, A.E., and Morgan Hughes, J.A. 1988, *Nature*, 331, 717.
- [88] Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A., Elsas, L.J., and Nikoskelainen, E.K. 1988, *Science*, 242, 1427.
- [89] DiMauro, S., Andreu, A.L. 2000, *Brain Pathol.*, 10, 431.
- [90] Kogelnik, A.M., Lott, M.T., Brown, M.D., Navathe, S.B., and Wallace, D.C. 1998, *Nucleic Acids Res.*, 26, 112.
- [91] Newman, N.J. 1993, *Arch Neurol.*, 50, 540.
- [92] Holt, I.J., Harding, A.E., Petty, R.K., and Morgan-Hughes, J.A. 1990, *Am. J. Hum. Genet.*, 46, 428.
- [93] Tatch, Y., Christodoulou, J., Feigenbaum, A., Clarke, J.T., Wherret, J., Smith, C., Rudd, N., Petrova-Benedict, R., and Robinson, B.H. 1992, *Am. J. Hum. Genet.*, 50, 852.
- [94] Carrozzo, R., Tessa, A., Vázquez-Memije, M. E., Piemonte, F., Patrono, C., Malandrini, A., Dionisi-Vici, C., Vilarinho, L., Villanova, M., Schägger, H., Federico, A., Bertini, E. and Santorelli, F. M. 2001, *Neurology*, 56, 687.
- [95] Keightley, J.A., Hoffbuhr, K.C., Burton, M.D., Salas, V.M., Johnston, W.S.W., Penn, A.M.W., Buist, N.R.M., and Kennaway, N.G. 1996, *Nature Genet.*, 12, 410.
- [96] Keightley, J. A., Anitori, R., Burton, M. D., Quan, F., Buist, N. R. and Kennaway, N. G. 2000, *Am. J. Hum. Genet.*, 67, 1400.
- [97] Andreu, A. L., Hanna, M. G., Reichmann, H., Bruno, C., Penn, A. S., Tanji, K., Pallotti, F., Iwata, S., Bonilla, E., Lach, B., Morgan-Hughes, J. and DiMauro, S. 1999, *N. Engl. J. Med.*, 341, 1037.
- [98] Bruno, C., Martinuzzi, A., Tang, Y., Andreu, A. L., Pallotti, F., Bonilla, E., Shanske, S., Fu, J., Sue, C. M., Angelini, C., DiMauro, S. and Manfredi, G. 1999, *Am. J. Hum. Genet.*, 65, 611.
- [99] Rahman, S., Taanman, J. W., Cooper, J. M., Nelson, I., Hargreaves, I., Meunier, B., Hanna, M. G., García, J. J., Capaldi, R. A., Lake, B. D., Leonard, J. V. and Schapira, A. H. 1999, *Am. J. Hum. Genet.*, 65, 1030.
- [100] Tiranti, V., Corona, P., Greco, M., Taanman J.-W., Carrara, F., Lamantea, E., Nijtmans, L., Uziel, G., Zeviani, M. 2000, *Hum. Mol. Genet.*, 9, 2733.
- [101] Karadimas, C. L., Greenstein, P., Sue, C. M., Joseph, J. T., Tanji, K., Haller, R. G.,

- Taivassalo, T., Davidson, M. M., Shanske, S., Bonilla, E. and DiMauro, S. 2000, *Neurology*, 55, 644.
- [102] Campos, Y., García-Redondo, A., Fernández-Moreno, M. A., Martínez-Pardo, M., Goda, G., Rubio, J. C., Martín, M. A., del Hoyo, P., Cabello, A., Bornstein, B., Garesse, R. and Arenas, J. 2001, *Ann. Neurol.*, 50, 409.
- [103] Collombet, J.-M. and Coutelle, C. 1998, *Mol. Med. Today*, 4, 31.
- [104] de Grey, A.D.N.J. 2000, *Trends Biotechnol.*, 18, 394.
- [105] DiMauro S, Hirano M, Schon EA. 2000, *Neurol Sci.*, 2000, S901.
- [106] Taylor RW, Wardell TM, Lightowlers RN, Turnbull DM. 2000, *Neurol Sci.*, 21, S909.
- [107] Rings, E.H., Büller, H., Neele, A.M. and Dekker, J. 1994, *Eur. J. Cell. Biol.*, 63, 161.
- [108] Seibel, P., Trappe, J., Villani, G., Klopstock, T., Papa, S. and Reichmann, H. 1995, *Nucl. Acids Res.*, 23, 10.
- [109] Vestweber, D. and Schatz, G. 1989, *Nature*, 338, 170.
- [110] Banroques, J., Delahodde, A., and Jacq, C. 1986, *Cell*, 46, 837.
- [111] Banroques, J., Perea, J., and Jacq, C. 1987, *EMBO J.*, 6, 1085.
- [112] Sanchirico, M., Tzellas, A., Fox, T.D., Conrad-Webb, H., Periman, P.S., Mason, T.L. 1995, *Biochem Cell Biol.*, 73, 987.
- [113] Gray, R.E., Law, R.H.P., Devenish, R.J. and Nagley, P. 1996, *Methods Enzymol.*, 264, 269.
- [114] Kempken, F., Howard, W. and Pring, D.R. 1998, *FEBS Lett.*, 441, 159.
- [115] Schon, E.A. 2000, *Trends Biochem. Sci.*, 25, 555.
- [116] Torchet, C., Jacq, C., and Hermann-Le Denmat, S. 1998, *RNA*, 4, 1636.
- [117] Corral-Debrinski, M., Belgareh, N., Blugeon, C., Claros, M.G., Doye, V., and Jacq, C. 1999, *Mol. Microbiol.*, 31, 1499.
- [118] Vázquez-Memije, M.E. and García, J.J. 2002, *This Book*.

F a l t a

P á g i n a

179 |

Apéndice VII.
Artículo en preparación:
Apicomplexan split *cox2* genes indicate
a green algal apicoplast ancestor

Apicomplexan split *cox2* genes indicate a green algal apicoplast ancestor

Soledad Funes¹, Edgar Davidson², Adrián Reyes-Prieto¹, Susana Magallón³, Pascal Herion⁴,
Michael P. King², and Diego González-Halphen^{1*}

¹Instituto de Fisiología Celular, Universidad Nacional Autónoma de México, 04510 D.F., Mexico.

²Biochemistry and Molecular Pharmacology, Thomas Jefferson University, PA 19107, USA.

³Instituto de Biología, Universidad Nacional Autónoma de México, 04510 D.F., Mexico.

⁴Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, 04510 D.F., Mexico.

*To whom correspondence should be addressed. E-mail: dhalphen@ifisiol.unam.mx

Members of the phylum Apicomplexa, that includes *Plasmodium* and *Toxoplasma* species of human pathogens, contain a vestigial plastid, the apicoplast (1), whose evolutionary history is controversial (1,2,3,4). Here we propose that distinctive chlorophyte (green algal) nuclear genes, encoding mitochondrial respiratory chain proteins, were transferred laterally to the ancestral apicomplexan, consistent with the apicoplast arising from secondary endosymbiosis of a green alga.

In most eukaryotes, the COXII subunit of the mitochondrial respiratory enzyme cytochrome *c* oxidase, is encoded by a single mitochondrial gene, *cox2*, but members of at least three genetic lineages lack *cox2* in the mitochondrial DNA (mtDNA). A number of leguminous plants have transferred a full-length *cox2* to the nucleus (5), where it acquired a region encoding a mitochondrial targeting sequence (MTS). Several chlorophyte algae lack *cox2* in the mtDNA (6). In *Chlamydomonas reinhardtii* and *Polytomella* sp., two algae of the class Chlorophyceae, *cox2* is present in the nucleus as two distinct genes: *cox2a* encodes a protein (COXIIA) corresponding to the N-terminal, membrane domain of orthodox COXII, and *cox2b* encodes COXIIB corresponding to the C-terminal domain that is exposed to the mitochondrial intermembrane space. Both COXIIA and COXIIB are imported into the mitochondrion (7). A truncated *cox2* in the mtDNA of the chlorophyte alga *Scenedesmus obliquus* (8) can now be recognized as a *cox2a* homolog. We predicted that this alga would have a nuclear *cox2b* (7), and have confirmed this (9)(web Fig. S1), suggesting that *cox2a* and *cox2b* may be widespread in the class Chlorophyceae. The mitochondrial *cox2a* and nuclear *cox2b* of *S. obliquus*

indicates that an ancestral chlorophyte *cox2* was divided in the mtDNA, prior to the independent transfer of *cox2a* and *cox2b* to the nucleus. *S. obliquus* represents an intermediate stage in this transfer with *cox2a* retained on the mtDNA, possibly due to the divergence of the *S. obliquus* mitochondrial genetic code from the standard genetic code (8) occurring after the transfer of *cox2b* to the nucleus.

The third group of organisms lacking mtDNA-encoded *cox2* is the Apicomplexa, consisting of obligate endoparasites. We identified transcripts and cloned nuclear-encoded *cox2a* and *cox2b* homologs from *Toxoplasma gondii* (9)(web Fig. S1). Fragmented *cox2a* and *cox2b* genes are also present in the genomes of *Eimeria tenella*, *Plasmodium yoelii*, and *Theileria parva*. The ancestral algal and apicomplexan *cox2* genes appear to have been split in identical locations (9)(web Fig. S1). In phylogenetic analyses of the available COXIIA and COXIIB sequences and representative conventional COXII sequences, apicomplexan COXIIA and COXIIB grouped with the chlorophyte sequences (Fig. 1). In contrast, apicomplexan mtDNA-encoded cytochrome *b* grouped independently from green algal cytochrome *b* sequences. This strongly argues that apicomplexan *cox2a* and *cox2b* derived from green algal genes and did not result from an independent splitting of *cox2*. All COXIIA proteins have a 125-170 amino acid N-terminal extension that is a cleavable MTS in algal COXIIA (7). The COXIIB proteins have a shorter N-terminal extension (43-60 residues) not cleaved in algae, with no equivalent in conventional COXII (7). The conserved PxxxPxxY motif in this otherwise poorly conserved novel region of COXIIB also indicates a common origin for apicomplexan and chlorophyte *cox2* genes (9)(web Fig. S1).

The *T. gondii cox2a* contains an intron that is conserved in location and phase with introns in *cox2a* of both *C. reinhardtii* and *Polytomella* (9) (web Figs. S1 and S2). This suggests a lateral transfer of *cox2a* and *cox2b* to the apicomplexan nuclear genome from the nucleus, not the mtDNA, of a green alga.

Analysis of the 35-kb circular DNA of apicoplasts, suggested that apicomplexan parasites evolved from a secondary endosymbiotic event (1,2). A green algal ancestor was proposed, based on similarity of *tufA* genes (2). In contrast, cladistic analysis of other genes (1,4) suggested a red algal origin. The mitochondrial genomes of a wide variety of red algae contain a single conventional *cox2* gene (6). Therefore, the nuclear *cox2a* and *cox2b* genes of the Apicomplexa indicate that the endosymbiont was a green alga (2), and specifically from the class Chlorophyceae.

That two genes encoding apicomplexan mitochondrial proteins are derived from a green algal precursor of the apicoplast, is a remarkable example of lateral gene transfer. A single *Toxoplasma* mitochondrial gene product, a component of a multisubunit enzyme, was replaced with two proteins

derived from green algae. The use of functional fragmented genes as atypical molecular markers has helped clarify the complex evolution of the causative agents of malaria and toxoplasmosis.

References and Notes

1. R.J. Wilson *et al.*, *J Mol Biol.* **261**, 155 (1996)
2. S. Köhler *et al.*, *Science* **275**, 1485 (1997)
3. G.I. McFadden, R.F. Waller, *Bioessays* **19**, 1033 (1997)
4. N.M Fast, J.C. Kissinger, D.S. Roos, P.J. Keeling, *Mol. Biol. Evol.* **18**, 418 (2001)
5. K.L. Adams *et al.*, *Proc. Natl. Acad. Sci. U S A* **96**, 13863 (1999)
6. M.W. Gray, *Curr. Opin. Genet. Dev.* **9**, 678 (1999)
7. X. Pérez-Martínez *et al.*, *J. Biol. Chem.* **276**, 11302 (2001)
8. A.M. Nedelcu, R.W. Lee, C. Lemieux, M.W. Gray, G. Burger, *Genome Res.* **10**, 819 (2000)
9. See supplemental information available in Science Online.
10. Sequencing of *T. parva* was funded by the International Livestock Research Institute. The *P. yoelii* sequencing program is carried on by The Institute for Genomic Research in collaboration with the Naval Medical Research Center and is supported by the U.S. Department of Defense. We thank Drs. J.M. Quinn and S. Merchant (UCLA) for the *S. obliquus* genomic library, and A. Gómez-Puyou, D.W. Krogmann, and D.A. Maslov for critical comments. Work supported by TW01176 (NIH), HL59646 (NIH), and IN207201 (UNAM).

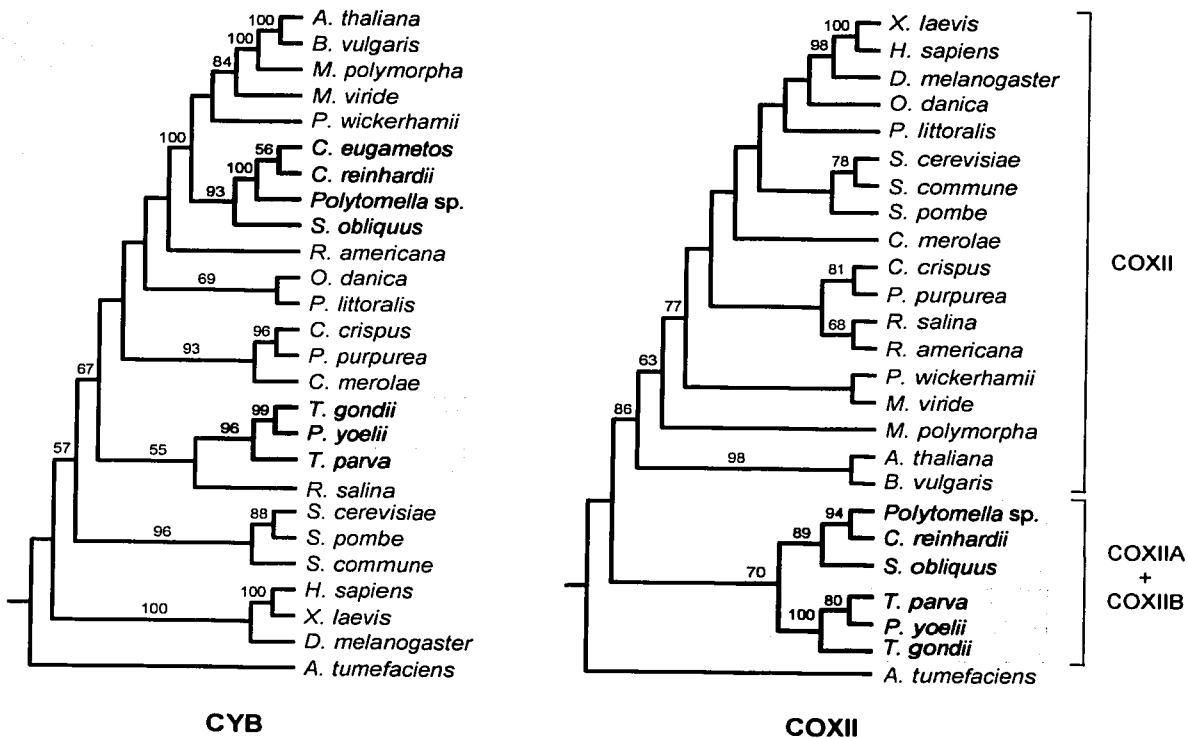


Figure 1

TESIS CON
FALLA DE ORIGEN

FIGURE LEGEND

Fig.1. *Maximum likelihood (ML) analyses showing the position of apicomplexan parasites based on cytochrome *b* and COXIIA + COXIIIB sequences.* Sequences were obtained from GenBank. *S. obliquus cox2b* (AF499003) and *T. gondii cox2a* and *cox2b* (AF499004 and AF503914) are from this work. Preliminary sequence data of *P. yoelii* and *T. parva* were obtained from The Institute for Genomic Research. For phylogenetic analyses, see methods in supplementary material (9). *A. tumefaciens* was specified as the single outgroup. Bootstrap support >50% is indicated above branches. *Left panel.* ML analysis for mitochondrial cytochrome *b* (CYB). *T. parva*, *P. yoelii* and *T. gondii* form a strongly supported clade (96%). *C. reinhardtii*, *C. eugametos*, *Polytomella* sp., and *S. obliquus* form an independent, strongly supported clade (93%) which is sister to a clade that includes other green algae and land plants. The resulting chlorobiont clade is solidly supported (100%). *Right panel.* ML analysis for COXIIA and COXIIIB, and homologous mitochondrial COXII sequences. COXIIA and COXIIIB (excluding their extensions) were fused *in silico* as a single polypeptide, and aligned with orthodox mitochondrial COXII sequences. A clade that includes *T. parva*, *P. yoelii*, and *T. gondii* and the green algae *S. obliquus*, *C. reinhardtii*, and *Polytomella* sp. is formed (70%). Within it, the apicomplexans are grouped in a strongly supported monophyletic group (100%), and the green algae form its sister clade (89%).

SUPPORTING ONLINE MATERIAL

MATERIALS AND METHODS

Total DNA was obtained from tachyzoites of *T.gondii* (Wiktor strain) grown *in vitro* in the presence of Vero cells. The *cox2a* and *cox2b* sequences were obtained by PCR amplification.

The *C. reinhardtii cox2b* gene was used to screen a genomic library of *S. obliquus* in Lambda-ZAP II (1). A positive clone was obtained from 5×10^4 plaque-forming units screened.

For phylogenetic analysis, nucleotide sequences were aligned using ClustalX (2), with subsequent manual refinements. Maximum likelihood (ML) analyses (3) were based on a HKY85 substitution model with gamma-distributed site to site rate variation, tree bisection-reconnection (TBR) branch swapping, with a molecular clock not enforced. For mitochondrial cytochrome *b* sequences (CYB), the transition/transversion ratio (ti/tv=0.9735) and shape of the gamma distribution ($\alpha=0.7095$) of the single most parsimonious tree resulting from the same data (not shown) were obtained via ML. These parameters were used as input criteria in several heuristic ML searches with different starting conditions. All ML searches converged on the same tree, with score $-\ln L=25883.5709$. Support for branches was obtained through 100 replicate bootstrap searches (HKY + Γ) in which ti/tv and α were specified. For COXII sequences, a heuristic ML search was conducted to estimate a topology and its associated parameters. The resulting tree had a score of $-\ln L=21054.4690$, ti/tv=0.871018, and $\alpha=0.810477$. Bootstrap analysis was conducted as described above. The likelihood of an alternative topology where apicomplexans and green algae are not sister taxa was tested by conducting 5000 RELL replicates of the Shimodaira-Hasegawa test (4,5). The likelihood score for the ML topology for CYB, estimated using COXIIA, COXIIB and homologous sequence data via ML (HKY + Γ , with ti/tv and α specified), is $-\ln L=21165.7706$, significantly less likely ($P<0.05^*$) than the ML topology obtained using the same data and parameters.

REFERENCES

1. J.M.Quinn, S.Merchant, *J. Phycol.* **35**, 1253 (1999)
2. J.D. Thompson *et al. Nucl. Acids Res.* **24**, 4876 (1997)
3. J. Felsenstein, *J. Mol. Evol.* **17**, 368 (1981)
4. H. Shimodaira, M. Hasegawa, *Mol. Biol. Evol.* **16**, 1114 (1999)
5. E.M Goldman, J.P. Anderson, A.G. Rodrigo, *Syst. Biol.* **49**, 652 (2000)

LEGENDS TO FIGURES

Fig. S1. *Amino acid sequence alignment of COXIIA and COXIIB subunits and comparison of intron positions.* The sequences of COXIIA and COXIIB of *C. reinhardtii*, *P. yoelii*, *Polytomella* sp., *S. obliquus*, *T. parva*, and *T. gondii* were compared with the orthodox, mitochondrial COXII sequences of *A. thaliana* and *M. viride* (shown in bold characters). Asterisks denote identical amino acids, and colons similar amino acids. Putative mitochondrial targeting sequences of COXIIA subunits are shown on a green background, and the PxxxPxxY motif is indicated in red. The C-terminal amino acid extensions of COXIIA are on yellow background, and the N-terminal amino acid extensions of COXIIB are shown on blue background. These unique extensions are hypothesized to interact for the assembly of COXIIA and COXIIB in the cytochrome *c* oxidase complex. The orange box indicates the experimentally determined N-terminal sequence of the mature COXIIA of *Polytomella* sp. The vertical black line separates COXIIA from COXIIB sequences. Inverted black triangles denote positions of introns in the corresponding genes. If the triangle is over an amino acid, it indicates that the corresponding intron is not in phase with the ORF.

Fig. S2. *Phases and flanking sequences of 5' and 3' splice sites for introns found in the cox2a and cox2b genes of chlorophyte algae and apicomplexan parasites.* For *cox2a* genes, introns 1, 2, 3, 4, and 5 of *Polytomella* sp. share positions and phases with introns 1, 2, 3, 4 and 5 of *C. reinhardtii*. Intron 6 of *Polytomella* sp., intron 7 of *C. reinhardtii*, and intron 1 of *T. gondii* also share positions and phases. In contrast, introns 6 and 8 of *C. reinhardtii* and intron 2 of *T. gondii* exhibit unique positions. For *cox2b* genes, intron 1 of *C. reinhardtii* shares position, but not phase, with intron 2 of *S. obliquus*. Finally, intron 1 of *S. obliquus* exhibits a unique position.

Polytomella sp. **MLAQRIS**-----**GNSLQCG**-----**KYMNAAQMGSKNIATVSETVQASTAAETVGAQFS**-----**PKEASAMASKKQNVVGG**-----**SGLSLAS**
C. reinhardtii **MLRQSGLS**-----**ANKLFCSNLILQSQQKQGNKLVVNMAMLFSSKRAEGSAVQGVVASEGVAQAVPQFS**-----**SEAAALAAKRRLGLG**-----**SGMSLAP**
S. obliquus -----
T. parva **MLTYSGI**-----**TRRLPQSS**-----**LLKNTKQFQTPQSHFGILYPNHVSQRHIFTSRLLL**-----**FTTKKPEPANNSTSDSESSVNM**
P. yoelii **MFNLN**-----**SKLYFC**-----**LRNINQLTSNHKIYYANSKRIVNFP**-----**FKYNEFTTKKDDNN**-----**KVKDI**
T. gondii **MLAYFAASRELAVVFCVGTLPENIFLFLSRRLLAIIPASSPCSLPFGSRLYPFR**-----**SCCGRPAGFPKXIVYVFFLSAHTFPKXMP LSSSLFRYPGHFLHQFPRRFFSSAQR**-----**GDYSL**
M. viride -----
A. thaliana -----

Polytomella sp. **RQTFSGS**-----**FAASAPSGAR**-----**AATAQAEAKAQTETSSIKKFKAAAAVVAALGLTACTASAEAPVAWQLGFQDSATSQAQAA**-----**FDLHDDIFFLLNTVVLVFYFLYIATKEH**
C. reinhardtii **SKPFAARGLISAAPKAAA**-----**AAGAEEAQAQADKQYAGLKKVVLKAAAALAAALGLTTTTAAADSPQWQLLFQDTATSTAQIMIDLHDDIFFLLITVVVLVYVMMFQITKEH**
S. obliquus -----**MISWVNMCMNVRADAPMAWQKLFQDPATSNMEGIVDLHADICFFLIVLVLVWLGARILVFSFH**
T. parva **KNQKQFTHVSETTFGTGKVRGTGLPKPVQPEDLPSIEFGKPT**-----**GMYNFVRHQHGDPRDHLREDGRFKKYATDGFHWYDAYTDPVKQRRTIVNGEAMVLGVETRPMEEELFGVEQTNVPFY**
P. yoelii **HDKTYAEKN**-----**NHGHEH**-----**TKGLYHHVDHGHGPHDHLTEDGTRRPEYDFDNYHDDYWANVVKQNIIVNGQMKIKGEETKPMEYLFNVSQNIPEW**
T. gondii **MLPRHFSSTASSVTKPPQKTDAL**-----**AKHDEHDSHHGTNNFYH**-----**MPSHHSPSRHHLNPDGTMRDILT**-----**TAAETFHWEHAEETPAQQIVSNGRMKVMGVETRDLEVELFVHVKNIPIFW**
M. viride -----**MDAPRPWAMGFPQDPTMIMQGI**-----**IDLHDDIFFLLTAVTFVTVMLARILHFHN**
A. thaliana -----**MIVLKWLFLTISPCDAAEFWLGSQDAATPMIQGI**-----**IDLHDDIFFLLILVFLVWLVLRALWHFH**

Polytomella sp. YTKQALP-EKLTHTHTAIEVI--WTVIPTIIIVVLIAPSLTLVYAIIDSHNDKPGELTVKVIQRQWYSYEMHDHLQHKLLDADRIVAIAEKTITR-----**MSDAKDQLKEQLKAS**-----**PSFR**-----**A**
C. reinhardtii YSKVLKP-EKLTHTHTMEVI--WTIIPTLIVVMAIAPSLTLIYSLDQHTERPLTVKIGRQWYSYEMHDHLQHKLLDADRIVGIAEKALVK-----**MSESKDQLKEKIKAD**-----**PSFR**-----**A**
S. obliquus HNLQVPV-ERFNHHTSLELV--WAILPSVIVTLIALPSLSLVYTYDDLVSFKPALTVKVTGRQWYSYAMNEHVQMNLSQAQAKDLLQS-----
T. parva HRRRLNLWGDHKNALRAEFCFFWI--PTFIIFSLAIPCYTMLYMLDESVTYTM-TVKVIGHQWYVWYEVESPEV-----**MLRGITLESILGNVYAEQ**-----**PLLD**-----**SD**
P. yoelii SRTRLNWGNHNLVVKVEFLFEWI--PTLIIFSIAIPCFMTMLYMLDEIVHTM-TVKVIGRQWYIYEVESPEDEE-----**MEVIRSI FGFNNKI TFNENKPVLLKQVK**
T. gondii PRMRMNVGNHDLKMAEFLFEWT--PTFITWSLAIPMTFLYMLDEAVYARM-TVKVIGRQWYIYEVESPVDEE-----**MDVFRSFWNRRGAATGGAATSSGRPAG**
M. viride **EKKNPINYYLVHGSTLEIV--WTLIPSFILMIALPSFALLYSTDEIVDPAV-TIKVIGHQWYSYEVESDY**
A. thaliana **YKKAIPAIRQIVHGTTIILR--TIFPISIIMFIAIPSFALLYSMDVEVVDPATITKAIKIGHQVITVEYSYD**

Polytomella sp. **ELK**-----**DKLKAALLSKVPASQPIQIN**-----**FDSYMT--DVQPGQLRMLEVDERLVLPNTNLRLLVLTASDVIHSWAVPSLGIKMDAIPGRLNQIWLITINREGVYGCSEICGAN**
C. reinhardtii **ELK**-----**DRINKALLSKVPASVPISTIN**-----**FDSYMLT--EVQPGQLRVLEVDERLVLPNTNLRLLVLTASDVLHSWAVPALGVKMDAVPGRLNQVWMSINREGVYGCSEICGAN**
S. obliquus -----**VEPGQIRLLDCDERLVLPNTNSLRLLVLTSSDVIHSWAVPSLGVKIDAI PGRLNQVWITIQRPVYGCSEICGAN**
T. parva **L-ERSR**-----**ERYPIPKKYLDNFWLIPKYYIS**-----**FQSNLVTDDELQPGMLRQLEVDKRLTLPTRTHIRFLITATDVLHWSVPSLGIKVDVAVPGRTRINTFILREGVYGCSEICGAN**
P. yoelii **DLENIAKAKNIR**-----**AEDYPTPKYIENPDQIPKYIIS**-----**FQSNMVTDEDLQPGMLRQLEVDKRLTLPTRTHIRFLITATDVIHWSVPSLGIKADAI PGRLHKVTFILREGVYGCSEICGAN**
T. gondii **CREENSVLLQSLKANSQQQKYLDDQDKIPTYYV**-----**FQSNMVTDEDLQPGMLRNLLEVDKRLTLPTRTHIRFLITATDVIHSWAVPALGIKADAI PGRLQRINTFQREGVYGCSEICGAN**
M. viride -----**LSDEQDSNLFDSYMPVEDELELGDRLVDRVVPKTHIRVFIITAADVLHWSVPSLGVKCDAI PGRLNQISLTVQREGVYGCSEICGAN**
A. thaliana -----**NSSDEQS-LTDSYMIPEEDELQSRLLLEVDRVVPKTHIRIIVTSADVPHSWAVPSSGVKCDAVPGRLNQISLTVQREGVYGCSEICGAN**

Polytomella sp. **HSEMPIVEAISPRAFLETVVKKWIQ**
C. reinhardtii **HSEMPIVEAISPRQFLTEYVKKWIS**
S. obliquus **HAFMPIVEAVTPRQFVQNYLKKWME**
T. parva **HGFMPIVEAVSPEKYAEHAKKYQD**
P. yoelii **HGFMPIVEAVSPEYAAHAKKYIRE**
T. gondii **HGFMPIVEAVSPEYAAHAKKYIKD**
M. viride **HAFMPIVEAVSVEDYISWVSQRLFKINKR**
A. thaliana **HAFTSIVEAVSPKDYGSRSVNLIPQTEA**

TESIS CON
FALLA DE ORIGEN

Figure S1

cox2a

	Intron n°	Phase	5' splice site	3' splice site
<i>T. gondii</i>	1	0	CAA/GTACGTC	TTTTCAG/AAA
	2	1	CAA/GTAAGAG	TGCGCAG/TTC
<i>Polytomella sp.</i>	1	0	CAG/GTTAATA	TTTATAG/TGC
	2	0	CAG/GTAATTG	CTCACAG/GCC
	3	0	GAG/GTGCGGG	TGTAAAG/GTT
	4	1	TCG/GTAAGTA	CACATAG/GCT
	5	0	AAG/GTAACGC	CCAATAG/GCC
	6	0	CAC/GTAAGTA	TTCGTAG/ATC
<i>C. reinhardtii</i>	1	0	CAG/GTGCAC	CTTGCAG/AGC
	2	0	CAG/GTACAGG	ATTGCAG/CAA
	3	0	CAG/GTGCAG	ATTTAAG/GCT
	4	1	GCA/GTAAGTC	CGTGCAG/GCG
	5	0	AAG/GTGGGGA	CGTGCAG/GCC
	6	0	CAG/GTGCACC	TCGACAG/GCC
	7	0	CAG/GTGGGGC	ACCGCAG/ATT
	8	1	CAG/GTATGAT	GCTGTAG/GCC

cox2b

	Intron n°	Phase	5' splice site	3' splice site
<i>T. gondii</i>	1	0	CAG/TCAGCAC	GCCCCAG/AAA
<i>C. reinhardtii</i>	1	2	CGA/GTGAGTT	ATTGCAG/GTA
<i>S. obliquus</i>	1	0	CAG/GTGAGGG	GCGTCAG/GTG
	2	0	CAG/GTAGGCG	GCTGCAG/AAC

Figure S2

TESIS CON
FALLA DE ORIGEN

Apéndice VIII.

**Comentario publicados sobre uno de los trabajos
realizados en esta tesis.**

Polymorphisms, positive selection and population expansion

The human gene encoding the dopamine receptor D4 (DRD4) is highly variable. Most of this diversity results from a 48 bp tandem repeat (VNTR) in the third exon of the gene, which varies in length and contains single nucleotide polymorphisms. In a worldwide population sample, the most commonly occurring allele (65.1%) contains four repeats (4R) and rarer forms contain anything from two to 11 repeats (2R–11R). Careful analysis of the pattern of haplotype diversity is consistent with a 'two-allele' system, with most of the other polymorphisms generated by recombination between the two most common haplotypes, 4R and 7R.

Ding *et al.* [1] demonstrated that the origin of the 2R–6R alleles can be explained by single one-step recombination/mutation events involving the 4R polymorphism, which they assume is the progenitor allele owing to its frequency in the population and insights from the limited sequence data

from primates. However, the formation of the second most common allele, 7R (19.2% occurrence), cannot be explained by such a simple one-step mutational mechanism. Ding *et al.* postulate that it would require at least one recombination and six mutation events to arise from the 4R allele, although they stress it is impossible to determine exactly how this allele arose. Additionally, the 7R allele is thought to be between fivefold and tenfold younger than the 4R allele. This has led the authors to suggest that the 7R allele is subject to positive selection, increasing its frequency in the general population.

What is fascinating, however, is that the 7R allele has been shown to be associated with attention deficit/hyperactivity disorder (ADHD). ADHD is estimated to affect up to 3% of elementary school children in the USA, with symptoms of inattentiveness, impulsivity and hyperactivity. Why then, if indeed there is a link between 7R and ADHD,

is the allele so frequent in the population? Even more puzzling is why the allele is subject to positive selection as the authors suggest? One idea proposed by Ding *et al.* is that the benefit incurred by possessing the 7R polymorphism related to what was happening in human evolutionary history when the allele arose 40 000–50 000 years ago. Then, the 7R polymorphism would show itself at the phenotype level not as ADHD, but as risk taking, drive and impulsiveness. It is possible that personality traits such as these would have been favourable during, and indeed could have forged, the expansion and changes in human society that occurred at this time.

1 Ding, Y.C. *et al.* (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 309–314

Anthony Isles
anthony.isles@bbsrc.ac.uk

From mitochondrion to nucleus and back again

The extant mitochondrial genomes of virtually all eukaryotes encode only a small number of predominantly hydrophobic polypeptides, almost all of which participate in the respiratory chain or oxidative phosphorylation complexes of the inner mitochondrial membrane. The majority of the protein-coding genes from the eubacterial symbiont that became the mitochondria are presumed to have been transferred to the nucleus, or else lost altogether. Exactly how or why this came about and the reasons, conversely, why a specific subset of genes has been left stranded in the mitochondrial (mtDNA), constitute one of the most intriguing puzzles of mitochondrial evolutionary genetics.

One gene that is encoded in the mtDNA of almost all eukaryotic taxa is subunit 6 of ATP synthase (A6, equivalent to bacterial subunit 6). Prominent exceptions are the malarial parasite *Plasmodium*, the ciliates *Tetrahymena* and *Paramecium*, and the green alga *Chlamydomonas*. In all cases, the completely sequenced mtDNA lacks a gene for A6, and there is not even any extended, unidentified mtDNA sequence that could

constitute an A6 cryptogene. The implication is that A6 is encoded in the nuclear DNA of these species. An understanding of what it takes to express A6 in the nucleus would be valuable, not least because some human diseases are caused by mutations in A6, and the ability to express it allotopically in the nucleus, targeting the protein product back to mitochondria, could offer a potential route to therapy. Thus far, attempts to achieve long-term allotopic expression of human A6 have not succeeded, and the hydrophobicity of the protein, which has no fewer than five membrane-spanning segments, has been suggested to be perhaps an insurmountable barrier.

A recent paper from the groups of Diego Gonzalez-Halphen and Michael King has now identified the nuclear A6 gene from *Chlamydomonas* [1]. The problem of relocation to the nucleus appears to have been solved during the course of evolution by a series of changes to the structure of the protein that significantly reduce its hydrophobicity to levels that the mitochondrial protein import machinery can apparently cope with. Three relatively unconserved transmembrane

segments that are not involved directly in proton translocation are markedly less hydrophobic than their counterparts in mitochondrially encoded A6 in other species. The protein has also acquired an unusually long mitochondrial targeting peptide of 107 amino acids, which might facilitate its solubilization in the hydrophilic environment of the cytosol. These features are shared with the nuclear-encoded COXII and COXIII genes of *Chlamydomonas*.

The transfer of A6 appears to have occurred relatively late in the evolution of this taxon, as the gene is mitochondrial in some other green algae. The implication is that the problem has been 'solved' several times in eukaryote evolution, raising the hope that engineering of mammalian A6 to permit efficient allotopic expression might not be as intractable as once imagined.

1 Funes, S. *et al.* The typically mtDNA-encoded ATP6 subunit of the mitochondrial F1F0-ATPase is encoded by a nuclear gene in *Chlamydomonas reinhardtii*. *J. Biol. Chem.* (in press)

Howy Jacobs
howy.jacobs@uta.fi

TESIS CON FALLA DE ORIGEN

Apéndice IX.
Materiales y Métodos

1. Cepas y Medios de Cultivo

Chlamydomonas reinhardtii

Se utilizó la cepa cw-15 de *C. reinhardtii* obtenida de la colección del "Chlamydomonas Genetics Center" de la universidad de Duke (Carolina del Norte, USA).

Esta cepa se agrupa dentro de un conjunto de mutantes que carecen de pared celular o que producen paredes celulares menos resistentes si se comparan con las cepas silvestres. Concretamente, las células de la cepa cw-15 son incapaces de ensamblar las capas centrales de la pared celular pero si presentan fibras semejantes a la capa externa de las células silvestres. También se ha reportado que proteínas que normalmente se encuentran asociadas a la pared celular, están ausentes en esta cepa (Harris, 1988).

C. reinhardtii se creció en medio TAP (Gorman y Levine, 1965) al cual se le añadió sorbitol (Sigma) hasta una concentración final de 1.0 %. Los medios de cultivo se incubaron a temperatura ambiente, en condiciones de luz continua con una agitación de 100 rpm. Las células se cosecharon en la fase exponencial de crecimiento por centrifugación durante 10 min a 100xg en una centrifuga clínica.

Para mantener las células, se resembraron periódicamente en cajas de Petri con medio TAP añadiendo agar al 1.5%.

Medio TAP:

Solución de Beijerinck 20 X	50 ml
Ácido acético glacial	1 ml
Tris (Base)	2.42 g
Solución de fosfatos 1M pH 7.0	1 ml
Solución de elementos traza	1 ml
H ₂ O	hasta 1000 ml

Solución de Beikerinck 20 X:

NH ₄ Cl	8 g
CaCl ₂ ·2H ₂ O	1 g
MgSO ₄ ·7H ₂ O	2 g
H ₂ O	hasta 1000 ml

Solución de fosfatos 1M pH 7.0:

K ₂ HPO ₄ 1M	250 ml
KH ₂ PO ₄ 1M	~170 ml
(agregar hasta pH 7.0)	

Solución de elementos traza:

EDTA-Na ₂	50.00 g
ZnSO ₄ ·7H ₂ O	22.00 g
H ₃ BO ₃	11.30 g

FeSO ₄ ·7H ₂ O	4.99 g
MnCl ₂ ·4H ₂ O	5.06 g
CoCl ₂ ·6H ₂ O	1.61 g
CuSO ₄ ·5H ₂ O	1.57 g
(NH ₄) ₆ Mo ₇ O ₂₄ ·4H ₂ O	1.10 g

Los compuestos se disolvieron por separado en los siguientes volúmenes de agua:

H ₃ BO ₃	en 200 ml
ZnSO ₄ ·7H ₂ O	en 100 ml
EDTA	en 250 ml

Cada uno de los otros elementos en 50 ml

Se mezclaron todas las soluciones excepto la de EDTA. La solución resultante tomó un tono violeta.

Se hirvió la solución hasta que el color cambió a verde translúcido.

Se añadió la solución de EDTA.

Se enfrió ligeramente (sin que la temperatura fuera menor a 70°C y se ajustó el pH con KOH al 20% hasta un valor entre 6.5 y 6.8.

La solución se diluyó con agua hasta alcanzar un volumen final de 1000 ml y se incubó a temperatura ambiente en un matríz cubierto con un tapón de algodón durante aproximadamente dos semanas hasta que la solución cambió a un color violeta. Se filtró la solución a través de un filtro Millipore de 0.22 µm para eliminar el precipitado rojo formado y se almacenó en un frasco ámbar a temperatura ambiente.

1.2 *Scenedesmus obliquus*

Scenedesmus obliquus (Turp.) Kütz se obtuvo de la colección de algas de la Universidad de Texas (UTEX-78). Las células se crecieron en medio basal de Bold (Bischoff y Bold, 1963) utilizando la solución de elementos traza descrita para el medio TAP:

NaNO ₃	(5.0 g/200 ml)	10 ml
MgSO ₄ ·7H ₂ O	(1.5 g/200 ml)	10 ml
NaCl	(0.5 g/200 ml)	10 ml
K ₂ HPO ₄	(1.5 g/200 ml)	10 ml
KH ₂ PO ₄	(3.5 g/200 ml)	10 ml
CaCl ₂ ·2H ₂ O	(0.5 g/200 ml)	10 ml

H ₃ BO ₃	(1.14 g/100 ml)	1 ml
Solución de Elementos Traza		1 ml
Solución de EDTA		1 ml
Solución de Fierro		1 ml
H ₂ O		hasta alcanzar un volumen final de 1 l

Solución de EDTA (100 ml):

EDTA-Na ₂	5.0 g
KOH	3.1 g

Solución de Fierro (1000 ml):

FeSO ₄ ·7H ₂ O	4.98 g
H ₂ SO ₄ (concentrado)	1.0 ml

Para mantener las células, se utilizaron cajas de Petri con el mismo medio añadiendo agar al 1.5%.

1.3 *Toxoplasma gondii*

Las muestras de DNA de *Toxoplasma gondii* se obtuvieron del laboratorio del Dr. Pascal Herion (Instituto de Investigaciones Biomédicas, UNAM). El DNA se obtuvo a partir de taquizoitos de la cepa WIKTOR crecidos *in vitro* sobre células VERO.

2. Purificación de DNA total de *C. reinhardtii*

El DNA de *C. reinhardtii* se obtuvo por tres métodos, según el tipo de análisis al que se fuera a someter posteriormente.

Para los experimentos tipo Southern, se realizaron dos procedimientos alternativos que varían en el procedimiento mediante el cuál se rompen las células:

En ambos casos se cosecharon 500 a 1000 ml de cultivo por centrifugación durante 10 min a 100xg en una centrifuga clínica.

A continuación se rompen las células por cualquiera de los siguientes mecanismos:

- Las células se lavaron con 100 ml de Tris-HCl 10 mM pH 8.0, EDTA-Na 1 mM (amortiguador TE). Posteriormente las células se resuspendieron en 5 ml de citrato de sodio 100mM y se congelaron en nitrógeno líquido. Se agregó un volumen de SDS 2% y se incubaron a 60°C durante 15 minutos.
- Las células se resuspenden en 25 ml de amortiguador QTP (Tris-HCl 10mM, NaCl 100mM, EDTA-Na 10mM y SDS 1%) y se agitan vigorosamente durante 10 seg.

Las células rotas se extrajeron dos veces con un volumen de fenol-cloroformo (1:1) y una vez con un volumen de cloroformo. La fase acuosa se precipitó incubando durante 15 min a -70°C con 0.1 volúmenes de acetato de sodio 3M pH 5.2 y 3 volúmenes de etanol 100%. Posteriormente se

centrifugó a 10,000xg en un rotor SS34 durante 10 min y el DNA precipitado se lavó con etanol 70%. Se resuspendió en 5 ml de agua y se incubó con 2.5 µg de RNasa (Roche) durante 8 h a 37°C. Después de esta incubación, se llevaron a cabo dos extracciones con un volumen de fenol-cloroformo (1:1) y finalmente con un volumen de cloroformo. El DNA se precipitó y lavó como se describió previamente; finalmente se resuspendió en 1 ó 2 ml de agua y se guardó a 4°C toda la noche para completar su dilución y posteriormente se almacenaron a -20°C para su almacenamiento.

Para los experimentos de PCR, el DNA se obtuvo utilizando el kit DNeasy Plant mini kit (Qiagen) siguiendo las instrucciones del proveedor con algunas modificaciones:

Células de 60 ml de cultivo (aproximadamente 100 mg de peso húmedo) se congelaron en nitrógeno líquido sumergiendo el tubo que las contenía (sin que las células entraran en contacto directo con el nitrógeno líquido) durante 1 min hasta que el botón celular se congeló por completo. De inmediato se calentó a 70°C hasta que el botón se fundió totalmente.

La preparación de células rotas, se limpió de restos celulares y precipitados salinos en un solo paso de centrifugación cargando la muestra en una columna QIAshredder (incluida en el kit).

El lisado limpio se mezcló con un amortiguador de unión que contiene etanol y se transfirió a una nueva columna que tiene una membrana de sílica-gel. Mediante un paso de centrifugación, el DNA se unió a la membrana, mientras que otras moléculas contaminantes como proteínas y polisacáridos pasaron a través de la membrana en dos pasos de lavado. El DNA puro se eluyó en amortiguador de elución precalentado a 65°C.

3. Purificación de RNA total de *C. reinhardtii*.

El RNA total de *C. reinhardtii* utilizado para experimentos tipo Northern se obtuvo a partir de 50 ml de cultivo utilizando el kit RNeasy mini kit (Qiagen) de acuerdo con las instrucciones del proveedor para el tratamiento por centrifugación de células animales. Las células se homogeneizaron utilizando las columnas QIAshredder en presencia de un amortiguador que contiene isotiocianato de guanidina (para inactivar las RNasas que podrían haber degradado la preparación). Una vez homogeneizada la muestra, se agregó etanol para producir las condiciones necesarias para que el RNA se uniera de manera específica a la membrana de sílica-gel que se encuentra en las columnas del kit. La muestra se introduce a las columnas RNeasy y se centrifuga para que al pasar la muestra a través de la membrana, el RNA se una a la misma. Los contaminantes se eliminan en tres pasos de lavado y el RNA limpio se eluye con agua libre de RNasas.

El RNA total de *C. reinhardtii* para realizar la síntesis de la primera hebra de cDNA para experimentos de RT-PCR, 5' y 3' RACE PCR se obtuvo a partir de 50 ml de cultivo utilizando el kit

RNeasy mini kit (Quiagen) utilizando la DNasa libre de RNasas para purificación de RNA en columnas (Qiagen). El procedimiento utilizado en este caso, es muy similar al anterior, sin embargo, el primer paso de lavado se divide en dos: se hace un primer lavado utilizando sólo la mitad del volumen usual, seguido de una incubación durante 15 min a temperatura ambiente con la DNasa y al finalizar esta incubación se realiza un segundo lavado utilizando la otra mitad del volumen usual de amortiguador lavado.

La concentración de RNA se determinó midiendo la absorbancia a 260 nm en un espectrofotómetro de luz UV. La pureza y calidad de la preparación de RNA se determinó a partir de la relación $A_{260\text{nm}}/A_{280\text{nm}}$ y de la forma del espectro de absorción de la muestra en la región entre 260nm y 280nm.

4. Geles para experimentos tipo Southern.

Para analizar muestras de DNA total, se corrieron 30 μg por carril en geles de agarosa al 0.8 o 1.0 %. Los primeros se utilizaron para conocer el tipo de genoma en el cual se localiza cierto gen cargando DNA total sin ningún tratamiento enzimático, y los segundos para hacer análisis de DNA total digerido con diferentes enzimas de restricción.

La transferencia del DNA a membranas de nylon Hybond-N (Amersham ahora Pharmacia) se hizo siguiendo las recomendaciones de Sambrook y col. (1989) toda la noche; o a membranas Zeta-probe GT (Bio-Rad) por capilaridad siguiendo un protocolo de transferencia alcalina incubando el gel 15 min con HCl 0.25 M y transfiriendo durante 45 min en un transferidor de vacío Pharmacia utilizando una solución de NaOH 0.4 N. El DNA se fijó a las membranas Hybond-N utilizando un entrecruzador Stratagene de luz UV y a las membranas Zeta-Probe calentando en un horno de vacío a 80°C durante 30 min.

5. Geles para experimentos tipo Northern.

Para analizar las muestras de RNA total, se corrieron 20 μg por carril en geles de agarosa 1% con formamida 0.66 M. El tratamiento del material de vidrio y de plástico, la preparación de soluciones y la preparación de las muestras se hicieron de acuerdo al método para análisis tipo Northern descrito por Sambrook y col. (1989).

La transferencia del RNA a membranas de nylon Hybond-N (Amersham ahora Pharmacia) siguiendo las recomendaciones de Sambrook y col. (1989). El RNA se fijó a las membranas utilizando un entrecruzador Stratagene de luz UV.

6. Purificación de DNA de geles de agarosa.

La purificación de segmentos específicos de DNA, se realizó colocando muestras de plásmidos dígidos o reacciones de PCR en geles de agarosa al 1 ó 1.2% y sometidos a electroforesis horizontal. De esta manera se separaron, de acuerdo a su masa molecular, los fragmentos de interés de otras porciones de DNA. Una vez que las muestras corrieron lo suficiente como para distinguir la región de interés del resto del DNA, se cortó el trozo de gel que contenía el DNA con una hoja de bisturí limpia y se colocó en un tubo Eppendorf con capacidad para 1.5 ml y se pesó. El DNA se purificó del gel por tres métodos diferentes: i) con el kit para extracción de DNA de geles de agarosa de Boehringer (ahora Roche); ii) con el kit QIAEX II de Qiagen; y iii) con el kit Qiaquick de Qiagen.

Estos tres kits parten de la solubilización por temperatura del trozo de agarosa y aíslan el DNA del resto de las moléculas presentes en la solución mediante su unión a una matriz de sílica (en forma de perlas en los primeros dos kits y en forma de membrana colocada en una columna en el tercero) en presencia de altas concentraciones de sales. La matriz a la cual está unido el DNA se lavó perfectamente eliminando todas las moléculas contaminantes con un amortiguador que contenía etanol. Posteriormente, el DNA se separó de la matriz con agua.

Para cada uno de los sistemas, se siguieron las instrucciones recomendadas por cada uno de los proveedores.

7. Uso de fragmentos purificados de DNA como sondas.

Para la detección radioactiva de genes o de regiones específicas de DNA en experimentos tipo Southern, Northern y búsquedas exhaustivas en bibliotecas (genómicas y de cDNA), se marcaron fragmentos de DNA conocidos con ^{32}P (incorporado a dATP o dCTP 10mCi/ml) utilizando la técnica de cebadores al azar con sistemas de GIBCO (ahora Invitrogen) o de Boehringer (ahora Pharmacia).

Se siguieron las recomendaciones para marcar de cada uno de los sistemas utilizados y posteriormente se limpiaron las sondas de nucleótidos no incorporados por centrifugación utilizando columnas con matrices de Bio-Gel P-30 de poliacrilamida (Micro Bio-Spin chromatography columns de Bio-Rad) o de Sephadex G-50 (Quick Spin Columns for radiolabeled DNA purification de Roche).

El porcentaje de incorporación se cuantificó utilizando un contador de centelleo tomando muestras de la sonda antes y después del paso de purificación.

8. Hibridación de las membranas de Nylon.

Cuando se realizaron experimentos de hibridación en membranas de nylon Hybond-N (durante los experimentos tipo Southern o Northern) y Hybond-NX (para los experimentos de búsqueda exhaustiva en bibliotecas) se utilizaron dos soluciones de hibridación diferentes:

- SSC 6X
Reactivo de Denhardts 5X
Tris-HCl 20 mM pH 8.0
EDTA- Na_2 2mM pH 8.0
SDS 2%
DNA de esperma de salmón 0.1 mg/ml*

* El DNA de esperma de salmón se incubó durante 10 min en agua en ebullición y se enfrió inmediatamente en hielo 10 min antes de agregarse a la solución de hibridación.

- SSC 5X
Reactivo de Denhardts 5X
SDS 0.5%

Cuando se realizaron experimentos tipo Southern con membranas Zeta-Probe GT (Bio-Rad) la solución de hibridación usada fue:

- Fosfato de sodio 0.25 M pH 7.2
SDS 7%

Las membranas se humedecieron en la solución de hibridación correspondiente antes de colocarlas en el recipiente donde se efectuaría la reacción. El volumen de solución de hibridación por cm^2 de membrana se calculó de acuerdo a las recomendaciones de cada uno de los proveedores.

Las membranas se incubaron toda la noche a 65°C después de añadir la sonda correspondiente marcada radioactivamente. A la mañana siguiente se lavaron las membranas con soluciones de concentraciones variables de SDS y SSC de acuerdo a la similitud de la sonda utilizada con el DNA examinado.

El análisis de las señales obtenidas se realizó exponiendo las membranas (sin dejar que se deshidrataran) a películas de autorradiografía o a una placa de detección "Phosphor screen" (Molecular Dynamics) para su análisis en un aparato tipo Storm o Typhon (Molecular Dynamics).

9. Búsquedas exhaustivas en bibliotecas genómicas y de cDNA

9.1 Búsqueda exhaustiva en una biblioteca de cDNA preparada en el fago λ gt10

Para conocer la secuencia de los genes *atp6* de *C. reinhardtii* se realizó una búsqueda exhaustiva en una biblioteca de cDNA preparada en el fago λ gt10 elaborada por Franzén y Falk (1992). La búsqueda y las soluciones empleadas durante el proceso, se realizaron de acuerdo a Sambrook y col. (1989) con las variaciones y detalles descritos por Pérez-Martínez (2000). La cepa de *E. coli* empleada para recibir el fago fue *C600hfz*.

9.2 Búsqueda exhaustiva en una biblioteca genómica preparada en el fago lambda ZAP II (Stratagene)

Para conocer la secuencia del gen *cox2b* de *S. obliquus* se realizó una búsqueda exhaustiva en una biblioteca genómica preparada por la compañía Stratagene y cedida por la Dra. Sabeeha Merchant (Universidad de California en Los Angeles, USA) (Queen y Merchant, 1999).

9.2.1 Titulación de la biblioteca

La titulación de la biblioteca se realizó siguiendo el protocolo descrito por Pérez-Martínez (2000) con las siguientes modificaciones:

Se utilizó la cepa XL1-Blue MRF' (Stratagene) para recibir el fago.

El medio de cultivo empleado fue NZY, en versión líquida para crecer a la cepa receptora, en versión sólida (agar 1.5%) para plaquear las células una vez infectadas y en versión Top-Agar (agar 0.7%) para verter las células infectadas en los platos de medio sólido.

Se añadieron 15 μ l de IPTG 0.5 M y 50 μ l de 250 mg/ml de X-Gal al medio NZY Top-Agar antes de verter en los platos para realizar un análisis de fondo: el vector Lambda ZAP contiene el plásmido pBluescript que es capaz de producir la proteína β -galactosidasa y por lo tanto es posible hacer un análisis de α -complementación de manera que sólo las colonias bacterianas que presenten un color blanco son las que deben considerarse como portadoras de un fragmento de DNA del organismo a estudiar.

9.2.2. Amplificación de la biblioteca

A partir de la biblioteca primaria, se realizó una ronda de amplificación siguiendo las recomendaciones del proveedor:

Se crecieron durante toda la noche a 37°C las células de la cepa XL1-Blue MRF⁺ en medio de cultivo NZY líquido complementado con Maltosa (Fisher) 0.2% y con MgSO₄ 10 mM (Fisher).

Las células se cosecharon por centrifugación 1000xg en una centrifuga clínica y se resuspendieron en MgSO₄ 10 mM hasta que la dilución alcanzó una densidad óptica (OD₆₀₀) de 0.5.

Se tomaron 600 µl de las células receptoras y se combinaron con la solución de fagos que tuviera aproximadamente 5x10⁴ pfu. Se hicieron 20 mezclas iguales en tubos Falcon 2059 de polipropileno, para amplificar alrededor de 1x10⁵ placas líticas. Se incubaron las mezclas durante 30 min a 37°C para que los bacteriófagos infectaran a las células. Al concluir esta incubación, se agregó a cada una de las mezclas entre 6.5 y 7 ml de medio NZY Top agar (previamente fundido y enfriado a 48°C) y se vació completamente en los platos de NZY agar de 150 mm de diámetro. Los platos se dejaron reposar alrededor de 10 min a temperatura ambiente hasta que el NZY top agar solidificó y posteriormente se incubaron aproximadamente 8 h hasta que las placas líticas formadas tuvieron alrededor de 2 mm de diámetro.

Se agregó a cada uno de los platos entre 8 y 10 ml de amortiguador SM y se dejaron incubar toda la noche a 4°C para que los bacteriófagos se difundieran hacia el amortiguador. Al día siguiente, se recuperó el mayor volumen de amortiguador posible. Se realizó un último lavado de los platos con 2 ml adicionales de amortiguador SM y se mezcló con el resto. Se añadió cloroformo hasta que se alcanzó una concentración final de 5% y se dejó incubar durante 15 min a temperatura ambiente. Estas mezclas, se centrifugaron durante 10 min a 500xg en una centrifuga clínica para eliminar los restos celulares y recuperar la suspensión de bacteriófagos lo más limpia posible. A la suspensión final, se le agregó cloroformo hasta una concentración final de 0.3%. El volumen total de la suspensión se dividió en alícuotas en presencia de 7% de DMSO y se almacenaron a -80°C.

La biblioteca amplificada se tituló de la misma manera que la biblioteca primaria.

9.2.3 Tamizaje de la biblioteca

Se plaquearon 6 cajas de NZY agar 1.5% de la misma manera que para titular la biblioteca, añadiendo 50,000 pfu recombinantes por caja. En este caso, no se agregaron ni IPTG ni X-Gal y una vez que las placas líticas alcanzaron el tamaño adecuado, las cajas se incubaron durante 2 h a 4°C para evitar que durante la transferencia de los bacteriófagos a las membranas de nitrocelulosa, fragmentos del medio NZY top agar se pegara a las mismas.

Las placas líticas se transfirieron por duplicado a membranas de nylon Hybond-NX circulares, colocando sobre la capa de NZY top agar una membrana durante 5 min y la segunda membrana durante 7 min. Para que después de la hibridación fuera posible orientar y localizar las placas líticas que mostraran una señal positiva, se realizaron tres cortes de forma triangular en posiciones asimétricas en el borde de cada una de las membranas. La silueta de estos cortes, se

reprodujo con un plumón indeleble en la caja correspondiente y las membranas y las cajas se numeraron de la misma manera.

Las membranas se desnaturalizaron durante 5 min en NaOH 1M y NaCl 1.5M y se neutralizaron durante 5 min en Tris 1M pH 7.5 y NaCl 1.5 M. Finalmente se lavaron durante 5 min en una solución de SSC 2x para eliminar los residuos celulares y se dejaron secar al aire. Estas tres incubaciones, se realizaron colocando las membranas sobre papel Whatman saturado con cada una de las soluciones, orientándolas de manera que las placas líticas no quedaran en contacto directo con el papel. El DNA se fijó a las membranas incubándolas durante 2h a 80°C en un horno de vacío.

9.2.4 Tamizaje secundario

Una vez que se realizó la hibridación de las membranas con la sonda correspondiente, se localizaron las placas líticas positivas siguiendo las señales que se observaron en las dos réplicas hechas para cada una de las cajas, de acuerdo a la orientación de los cortes de las membranas y la posición de las marcas en las cajas de cultivo.

Ya que la población de placas líticas en las cajas del primer tamizaje es muy alto, hay una gran cantidad de placas no positivas que se localizan muy cerca de la placa positiva, por lo que resulta casi imposible aislar por completo la placa de interés. Por esto, es necesario hacer un tamizaje secundario en cuyas cajas se localice un número menor de placas y por lo tanto facilitar la purificación de una sola placa lítica.

El tamizaje secundario, se realizó siguiendo el protocolo reportado por Pérez-Martínez (2000) con la diferencia de que las membranas utilizadas en este caso fueron Hybond-NX redondas para cajas con un diámetro de 100 mm.

9.2.5 Recuperación del plásmido pBluescript

Una vez que se logró aislar una placa positiva del resto, se utilizó el fago ayudante Ex-Assist resistente a interferencias (Stratagene), que permite una escisión eficiente del plásmido pBluescript a partir del vector Lambda Zap II. Este fago ayudante, contiene una mutación ambar que evita la replicación de su genoma en algunas cepas de *E. coli* no supresoras como la cepa SOLR. La recuperación del plásmido, se realizó siguiendo las recomendaciones del proveedor para la escisión de una clona aislada:

Se colectó la placa de interés de la caja de NZY agarosa utilizando la punta de una micropipeta cortada y se colocó el fragmento en un tubo para microcentrífuga de 1.5 ml que contuviera 500 μ l de amortiguador SM y 20 μ l de cloroformo. Se agitó vigorosamente la mezcla en un aparato vortex para liberar los bacteriófagos en el amortiguador. Esta mezcla se incubó toda la noche a 4°C.

De manera simultánea, se crecieron independientemente durante toda la noche a 37°C, las cepas XL1-Blue MRF' (en medio NZY complementado con maltosa 0.2% y MgSO₄ 10 mM) y SOLR (en medio LB). Se cosecharon por centrifugación y se resuspendieron en MgSO₄ 10 mM hasta alcanzar una OD₆₀₀ de 1.0.

Se mezclaron 200 µl de células XL1-Blue MRF' con 250 µl de la suspensión de fagos y 1 µl del fago ayudante y se incubaron durante 15 min a 37°C.

Se añadieron 3 ml de medio LB y se incubaron en un tubo Falcon 2059 de polipropileno durante 3 h a 37°C con una agitación de 250 rpm.

Al término de la incubación, se calentaron las muestras durante 20 min a 65°C y se centrifugaron durante 15 min a 1000xg en un rotor SS34. Se recuperó el sobrenadante que contiene la solución madre del fagémido escindido pBluescript empacado como fagos filamentosos.

Finalmente se mezclaron 100 µl y 10 µl de la solución madre de fagémidos con 200 µl de las células de la cepa SOLR en tubos para microcentrifuga independientes. Se incubaron durante 15 min a 37°C y se plaquearon en cajas de LB agar 1.5% complementado con ampicilina 50 µg/ml. Las células se dejaron crecer durante toda la noche a 37°C. Las colonias que crecieron durante la noche, contienen el plásmido de doble cadena con el inserto de interés.

9.2.6 Subclonación de clonas positivas

Los insertos presentes en una biblioteca genómica suelen ser de gran tamaño, por lo que obtener la secuencia completa puede resultar un proceso relativamente largo. Para obtener la secuencia del gen de interés, se realizaron experimentos tipo Southern digiriendo los plásmidos pBluescript purificados con varias enzimas de restricción y transfiriendo el DNA a membranas de nylon para realizar una nueva hibridación y localizar fragmentos relativamente pequeños.

Los fragmentos identificados se subclonaron en el plásmido pBluescript digerido con la misma enzima con la que se digirió el plásmido original. La ligación se realizó con la enzima T4 ligasa (New England Biolabs) incubando toda la noche a 16°C. Se tomó la mitad de la reacción de ligación y se transformaron células competentes *E. coli* de la cepa JM109 (Promega) en cajas de medio LB complementado con ampicilina 50 mg/ml y añadiendo IPTG y X-Gal para poder seleccionar aquellas colonias que presentan el inserto. Para ver los detalles de estas reacciones consultar Pérez-Martínez (2000).

9.2.7 Medios de cultivo y amortiguadores utilizados*NZY*

NaCl	0.5%
MgSO ₄ ·7H ₂ O	0.2%
Extracto de levadura	0.5%
NZ amina (hidrolisado de caseína)	1.0%
Ajustar el pH hasta 7.5 con NaOH	

Cajas de NZY Agar o Agarosa

Medio NZY con agar o agarosa hasta una concentración final de 1.5%

Top NZY Agar o Agarosa

Medio NZY con agar o agarosa hasta una concentración final de 0.7%

LB

NaCl	1.0%
Bactotripton	1.0%
Extracto de levadura	0.5%

Cajas de LB-Agar

Medio LB con agar hasta una concentración final de 1.5%

Si es necesario, añadir ampicilina hasta una concentración final de 50 µg/ml.

Todos los medios de cultivo se esterilizan por autoclave durante 20 min a 15 psi.

Amortiguador SM (cantidades para 1 l)

NaCl	5.8 g
MgSO ₄ ·7H ₂ O	2.0 g
Tris-HCl 1M (pH 7.5)	50 ml
Solución de gelatina 2%*	5 ml
H ₂ O	hasta 1 l

Esterilizar el amortiguador por autoclave durante 20 min a 15 psi.

*La solución de gelatina al 2% se prepara añadiendo 2 g de gelatina a 100 ml de agua y esterilizando la solución durante 20 min a 15 psi.

La cepa XL1-Blue MRF' se conserva en medio LB (líquido o sólido) complementado con tetraciclina hasta una concentración final de 50 µg/ml.

La cepa SOLR se conserva en medio LB (líquido o sólido) complementado con kanamicina hasta una concentración final de 50 µg/ml.

10. Reacción en Cadena de la Polimerasa (PCR).

La mayor parte de las reacciones de PCR realizadas a lo largo de este trabajo, se hicieron utilizando la enzima (*Taq* DNA Polimerasa o la HotStarTaq) y los otros reactivos de la marca Qiagen. Sin embargo, para obtener los productos amplificados de PCR en el caso del gen *atp6* de *C. reinhardtii* se utilizó la *Pfu* turbo DNA Polimerasa (Stratagene) utilizando el amortiguador de la enzima recomendado por el proveedor y los otros reactivos (incluyendo la solución Q) de Qiagen.

De manera general, las reacciones de PCR se realizaron en un volumen final de 50 µl en condiciones de MgCl₂ 2mM, 2.5 unidades de polimerasa, 1x de amortiguador de la enzima, 1x de solución Q, 200 µM de desoxinucleótidos trifosfatados (dNTPs), 20 pmol de oligodesoxinucleótidos, entre 200 ng y 1 µg de DNA o cDNA templado y añadiendo agua hasta completar el volumen de reacción.

Los ciclos de PCR utilizados cuando se esperaban productos amplificados cortos (de menos de 2 kb) fueron:

	5 min a 94°C*	(desnaturalización inicial)	
35 ciclos	{	1 min a 94°C	(desnaturalización)
		1 min a 55°C	(alineamiento)
		2 min a 72°C	(extensión)
		7 min a 72°C	(extensión final)

* Cuando se utilizó como DNA templado una alcuota de alguna de las bibliotecas clonadas en fagos, esta incubación se realizó durante 15 min.

Los ciclos de PCR utilizados cuando se esperaban productos amplificados largos (de más de 2 kb) fueron:

	5 min a 94°C	(desnaturalización inicial)
10 ciclos	10 seg a 94°C	(desnaturalización)
	1 min a 55°C	(alineamiento)
	4 min* a 68°C	(extensión)
25 ciclos	10 seg a 94°C	(desnaturalización)
	1 min a 55°C	(alineamiento)
	4 min* (mas 10 seg/ciclo) a 68°C	(extensión)
	7 min a 68°C	(extensión final)

*se calculó aproximadamente, una extensión de 1 min por cada kb esperada para el producto.

Para analizar los productos de PCR obtenidos, se corrieron 10 ml de cada uno en geles de agarosa al 1% ó 1.2% (dependiendo del tamaño esperado).

11. Preparación de la primera hebra de cDNA

La síntesis de la primera hebra de cDNA se realizó utilizando cuatro sistemas diferentes: Thermo RT™ (Display Systems Biotech), M-MuLV RT (Promega), Superscript™II (Gibco BRL ahora Invitrogen) y Omniscript (Qiagen). En todos los casos se tomaron como templado 1 ó 2 µg de RNA total y 25 unidades de inhibidor de RNasas (Boehringer ahora Roche).

Como en general el DNA nuclear de *C. reinhardtii* tiene un alto contenido de G-C y produce estructuras secundarias que pueden interferir con la síntesis de cDNA, se incubó el RNA templado a 65°C durante 5 minutos y se enfrió rápidamente en hielo, además se agregó amortiguador Q (Qiagen) a una concentración final de 1X independientemente del sistema utilizado.

12. Adición de colas de poli A al extremo 5' del cDNA

Para poder amplificar por PCR el extremo 5' del cDNA se le añadió a la primera hebra de cDNA sintetizada una cola de poliA. Esta reacción se llevó a cabo con la enzima transferasa terminal (Boehringer ahora Roche) siguiendo las recomendaciones del proveedor:

Se mezclaron los siguientes elementos:

- 50 U de transferasa terminal (2 µl)
- dATP 6.25 µM
- Amortiguador de reacción 1X
- CoCl₂ 1.5 mM

TESIS CON
FALLA DE ORIGEN

cDNA 1 μ g

Agua hasta un volumen final de 20 μ l

La mezcla se incubó durante 15 min a 37°C y posteriormente durante 5 min a 65°C para desactivar a la enzima.

13. Identificación de los extremos de las secuencias de cDNA por experimentos de 5' y 3' RACE-PCR (Frohman, 1993)

Para poder amplificar los extremos de los cDNAs, se aprovechó la extensión de poliAs que éstos tienen naturalmente en el extremo 3' y artificialmente en el extremo 5'. En ambas reacciones se tomó 1 μ g de cDNA como templado (sin extensión artificial de poli A para realizar el experimento de 3'-RACE y con extensión artificial de poli A en el extremo 5' para el 5'-RACE).

Estos PCRs se llevaron a cabo en dos reacciones independientes, ambas se realizaron utilizando los ciclos de temperatura descritos para amplificar productos cortos pero cambiando la combinación de oligodesoxinucleótidos y la temperatura de alineamiento: la primera con el oligodesoxinucleótido dT/adaptador y un oligodesoxinucleótido específico para el gen en cuestión se realizó a una temperatura de alineamiento de 48°C mientras que en la segunda reacción se utilizó el oligo adaptador y un oligodesoxinucleótido específico anidado (con respecto al oligo de la primera reacción) y una temperatura de alineamiento de 55°C.

14. Diseño de oligodesoxinucleótidos

14.1 Oligodesoxinucleótidos específicos

Los oligodesoxinucleótidos específicos fueron diseñados en regiones de una secuencia conocida de acuerdo con los siguientes principios generales:

- Que el contenido de G-C fuera alrededor de 50%.
- Que el extremo 3' fuera una C o G (para estabilizar el extremo 3' del oligodesoxinucleótido, que es donde va a empezar a trabajar la polimerasa en la reacción de PCR).
- Que no hubiera más de dos Gs o Cs en el extremo 3' (para evitar que alineamientos inespecíficos se estabilizaran).
- Que la secuencia del oligodesoxinucleótido no formara estructuras diméricas por complementariedad de la secuencia con él mismo.
- Que la secuencia no contenga regiones complementarias que puedan formar estructuras secundarias (por ejemplo asas).
- Que tuvieran una longitud de entre 18 y 30 pb. Teóricamente, una secuencia de 18 pb representa una secuencia única entre 7×10^{10} nt y por lo tanto debería ser

complementario una sólo vez en un genoma eucarionte, que usualmente contienen entre 10^8 y 10^{10} pb.

- Que la T_m teórica fuera alrededor de 55°C .

14.2 Oligodesoxinucleótidos degenerados

Los oligodesoxinucleótidos degenerados se utilizan cuando no se conoce la secuencia de nucleótidos de un gen específico y se diseñan en base a regiones conservadas de residuos de aminoácidos de la proteína codificada por el gen que se está buscando.

Además de las características tomadas en cuenta para el diseño de oligodesoxinucleótidos específicos, se buscó que la secuencia tuviera las siguientes características generales:

- Buscar regiones de la secuencia proteica cuyos residuos de aminoácidos no se codifiquen por muchos codones (evitar en lo posible Arg, Leu, Ser, Pro, Ala, Gly, Thr, Val). Es muy importante considerar el uso de codones preferencial de cada organismo, ya que en algunas ocasiones es tan sesgado que, para estos propósitos, se pueden considerar menos degeneraciones.
- Buscar regiones altamente conservadas sobretodo en el equivalente al final del oligodesoxinucleótido.
- Que las últimas posiciones de la secuencia del oligodesoxinucleótido no contuvieran degeneraciones.

Para analizar la secuencia de los oligodesoxinucleótidos para evitar complementaridades y por lo tanto para evitar la formación de dímeros o de estructuras secundarias, así como para calcular las T_m se utilizaron los programas Amplify 1.2 e HyperPCR.

Todos los oligodesoxinucleótidos utilizados en este trabajo se sintetizaron en la "Nucleic Acid Facility" del Kimmel Cancer Institute de la Universidad Thomas Jefferson (Philadelphia, PA, USA).

La secuencia de los oligodesoxinucleótidos empleados en este trabajo se encuentran en el Apéndice XI.

15. Clonación de los productos de PCR

Una vez que los productos amplificados se purificaron de un gel de agarosa siguiendo los protocolos descritos anteriormente, se clonaron en vectores especiales para productos de PCR:

- Todos los productos obtenidos con una enzima *Taq* o HotStar*Taq* polimerasa se clonaron en vectores como pGEM-T Easy Vector (Promega) o pCR2.1 (Invitrogen), ya que contienen extremos 3' –T que permiten una fácil ligación con los extremos 3' desoxiadenosina que dejan estas polimerasas.

En algunos casos, se construyó un vector con extremos 3'-T para clonar productos de PCR sin modificar siguiendo el protocolo descrito por Marchuk y col. (1991): el plásmido pBluescript (Stratagene) se digirió con la enzima de restricción *EcoRV* y se incubó con la enzima *Taq* polimerasa (Qiagen) en presencia del amortiguador de la enzima diluido hasta una concentración de 1x y dTTP 2 mM durante 2 h a 70°C. El vector se limpió posteriormente con el kit Qiaquick (Qiagen) siguiendo el protocolo recomendado para eliminar nucleótidos y se utilizó en reacciones de ligación.

- A los productos obtenidos con enzimas como la *Pfu* polimerasa, se les añadió una desoxiadenosina en el extremo 3' para poder clonarlos en el mismo vector con extremos 3'-T:
 - En 50 µl de reacción, se mezclaron 4µl de dATP 10 mM, 5 µl de amortiguador de PCR 10x, el fragmento de interés, 2.5 U de *Taq* polimerasa y agua hasta completar el volumen. Esta mezcla se incubó durante 30 min a 72°C y posteriormente se limpió del exceso de dATP y de enzima utilizando las columnas de Qiaquick siguiendo el protocolo descrito por el proveedor para eliminar nucleótidos. Una vez listo el fragmento, se clonó de la misma manera que los productos obtenidos utilizando una *Taq* polimerasa.

En todos los casos las reacciones de ligación, se llevaron a cabo incubando con la enzima T4 ligasa durante toda la noche a 16°C. En el caso de los vectores comerciales de ligación de productos de PCR se utilizó la enzima que incluye el proveedor para este fin y en el caso del vector construido en el laboratorio, se utilizó la enzima de Gibco (ahora Invitrogen).

A continuación, se tomó la mitad de la reacción de ligación y se utilizó para transformar por temperatura a la cepa de *E. coli* JM109 ó DH5α.

16. Secuenciación de DNA

Las secuencias de reportadas en este trabajo se realizaron en secuenciadores automáticos del Kimmel Cancer Center DNA Sequencing Facility de la Universidad Thomas Jefferson (Philadelphia, PA, USA) o en la Unidad de Biología Molecular del Instituto de Fisiología Celular de la UNAM (Ciudad de México, México).

Las secuencias de los genes obtenidas en este trabajo se encuentran en el Apéndice X.

17. Purificación y solubilización de mitocondrias de *C. reinhardtii*

La purificación y solubilización de mitocondrias enteras de *C. reinhardtii*, se realizó siguiendo el protocolo reportado por Eriksson (1995) con algunas modificaciones:

Las células de la cepa cw-15 de *C. reinhardtii* se lavaron dos veces en amortiguador de lavado y se resuspendieron en amortiguador de ruptura hasta alcanzar una concentración final de 0.5 g células/ml. La suspensión de células se mezcló con el equivalente a dos volúmenes de perlas de vidrio de 0.5 mm de diámetro y se agitó en un aparato tipo Vortex a baja velocidad durante 1 min (hasta que aproximadamente el 95% de las células se hubiera roto). La solución obtenida de esta manera, se colectó y separó de las perlas de vidrio. Éstas fueron lavadas tres veces con 10 ml de amortiguador de ruptura recuperando el sobrenadante después de una centrifugación a 2000xg durante 5 min. Todas las soluciones se incorporaron y se mezclaron con un homogeneizador de teflón. El homogenado resultante se centrifugó 10 min a 1500xg en un rotor SS34. El sobrenadante se centrifugó durante 10 min a 12000xg en el mismo rotor y se colectó el botón obtenido que contiene la fracción mitocondrial. El extracto crudo de mitocondrias se resuspendió en 5 ml aproximadamente de amortiguador de ensayo y se mezcló con amortiguador de Percoll hasta alcanzar un volumen final de 35 ml. La mezcla se homogeneizó dos veces con un homogeneizador de teflón. El homogenado se centrifugó durante 40 min a 20000xg en un rotor SS34. Se colectó el botón y se diluyó en amortiguador de dilución de Percoll hasta un volumen final de 40 ml. El homogenado mitocondrial final fue centrifugado durante 10 min a 12000xg en un rotor SS34. El botón que contenía a las mitocondrias, se colectó y lavó nuevamente con el amortiguador de dilución de Percoll. Esta preparación se centrifugó durante 5 min a 10000 en una microcentrífuga Eppendorff. El sobrenadante se descartó. El botón mitocondrial final se resuspendió en 35 ml de amortiguador de dilución de Percoll complementado con inhibidores de proteasas.

17.1 Amortiguadores utilizados

Amortiguador de lavado

HEPES-KOH	20 mM
pH 7.2	

Amortiguador de ruptura

HEPES-KOH pH 7.2	50 mM
EDTA	5 mM
Sorbitol	0.25 M
Cisteína	4 mM
PVP-40	0.5%
BSA	0.1%

Amortiguador de ensayo

Sorbitol	0.25 M
KCl	10 mM

Amortiguador de fosfatos pH 7.2	10 mM
BSA	0.1%
MgCl ₂	5 mM

Amortiguador de Percoll

Sorbitol	0.25 mM
MOPS-KOH pH 7.2	10 mM
PVP-40	0.5%
Percoll	20%
BSA	0.1%
EDTA	1 mM

Amortiguador de dilución de Percoll

Amortiguador de fosfato de potasio	10 mM
Sorbitol	0.25 M
EDTA	10 mM
BSA	0.1%
pH 7.2	

Amortiguador de dilución de Percoll combinado con proteasas

Amortiguador de dilución de Percoll complementado con los siguientes inhibidores de proteasas:
 fenil-metil-sulfonil-fluoruro 0.5 mM
 benzamidina 1 mM
 ácido ϵ -aminocaproico 5 mM

18. Determinación del valor de proteína total

La concentración de proteína en las muestras se calculó mediante el método de Lowry y col. (1951) con la modificación de Markwell y col. (1978).

19. Electroforesis nativa azul

Las mitocondrias de *C. reinhardtii* se solubilizaron en lauril-maltósido 1.0% y se sometieron a una electroforesis nativa azul (Schägger, 1994) usando geles con un gradiente de 5.0 a 12.0% de poliacrilamida.

20. Electroforesis desnaturalizante

Los complejos de la cadena respiratoria mitocondrial de *C. reinhardtii* se separaron en geles de placa de segunda dimensión de 1.2 mm de espesor al 12 % de acrilamida en una cámara BRL (Vertical Gel Electrophoresis System) en presencia de SDS y Tricina (Schägger y von Jagow, 1987).

21. Secuenciación del extremo amino terminal de proteínas

Las proteínas del complejo F_0F_1 ATP sintetasa de *C. reinhardtii* se electrotransferieron a una membrana PVDF ProBlott (Applied Biosystems) a 40 V durante 4.5 h a 4°C en un amortiguador de ácido bórico 50 mM y Tris 50 mM (sin ajustar el pH).

Las membranas se tñeron posteriormente con azul de Coomassie R, se destñeron para eliminar el exceso de colorante y se dejaron secar al aire.

La determinación de la secuencia amino terminal de los polipéptidos se hizo en el Laboratoire de Microséquençage des Protéines del Institut Pasteur (París, Francia) en un secuenciador 473A de Applied Biosystems (Atteia y col., 1997).

22. Análisis *in silico* de las secuencias obtenidas

Las comparaciones entre secuencias, tanto de nucleótidos como de aminoácidos, se realizaron con el programa ClustalX utilizando los parámetros por omisión considerados por el programa.

La búsqueda de secuencias y comparaciones de las bases de datos se realizaron utilizando las distintas versiones del programa Blast (<http://www.ncbi.nlm.nih.gov/BLAST/>).

El análisis de las presecuencias mitocondriales así como el cálculo de los segmentos de Cr-ATP6 entre 13 y 17 aminoácidos con hidrofobicidad local (<H>) se realizaron con el programa MitoProt II (Claros, 1995). La mesohidrofobicidad se estimó al explorar cada secuencia polipeptídica para localizar segmentos con un niveles máximos de hidrofobicidad en ventanas de entre 60 y 80 residuos y promediando los valores obtenidos.

Los cruces transmembranales de Cr-ATP6 se predijeron utilizando el programa TodPred II (Claros y von Heijne, 1994), en el servidor de Internet PredictProtein (www.es.emblnet.org/Services/MolBio/PredictProtein/) (Rost, 1996) y en el ExpASy Molecular Biology Server (www.expasy.ch).

Los datos usados para analizar el uso de codones se obtuvieron de la base de datos de uso de codones (Codon Usage Database) del Kazusa DNA Research Institute (Chiba, Japón) (www.kazusa.or.jp/codon).

Apéndice X.
Secuencias reportadas

atp6 de *C. reinhardtii*

agacgaagaatatagattggattaacc atg gcg
 M A
 ctg cag cag gct gcg ccg agg gtg ttt ggc ctc ttg ggg cgt gcc cct gtc gct ctg ggc
 L Q Q A A P R V F G L L G R A P V 115 L G
 cag agc ggc att ctg act ggt agc agc ggc ttc aag aac cag gga ttc aat ggc tcc ctg
 Q S G I L T G S S G 72 F K N Q G F N G S L
 cag tct gtg gag aac cat gtg tat gcc cag gcc ttc agc acc tct tct cag gag gag cag
 Q S V E N H V Y A Q A F S T S S Q E E Q
 gct gct ccc agc atc cag ggt gct tcg ggc atg aag ctg cct ggt atg gct ggt agc atg
 A A P S I Q G A S G M K L P G M A G S M
 ctg ctg ggc aag tcg cgc agc ggc ctg cgc acc ggc tcg atg gtg cca ttc gct gcc cag
 L L G K S R S G L R T G S M V P F A A Q
 cag gcg atg aac atg tcg acg gcc gcc cag gtg cag tcg ggc gcc acc gtg aac tcc ctg
 Q A M N M S T A A Q V Q S G A T V N S L
 ctg ctc ggc atc tac cgc ttc tgg cgc agc cag gcg ccc atg gat aag ccg cac gcg ccc
 L L G I Y R F W R S Q A P M D K P H A P
 gtg gat gac cgc atg ctc ccc gcc att gtg gat gct agc gac aac cgc gcg gcc ctc ggc
 V 263 D D R M L P A I V D A S D N R A A L G
 acc tgg gcc acc gcc ctg ttc tgc acc atc ctg gcc tcc aac ttg ctg ggt ctg gtg ccc
 T W A T A L F C T I L A S N L L G L V P
 acc aac gag gcg ccc acc tcc ggc ctg ggc ttc gcc acc ggc ctg ggt gtg agc gtg tgg
 T N E A P T S G L G F A T G L G V S V W
 gcc acc gcc acc acc ctg ggt ctg tac aag ctg ggc ttc tcc ttc ccc ggg cac ttc atc
 A 186 A T T L G L Y K L G F S F P G H F I
 cct ggc ggc acc ccc tgg ccc atg gcc ttc atc ttc gtg ccc ctg gag acc att tcg tac
 P G G T P W P M A F I F V P L E T I S Y
 acc ttc cgt gct gtt tcc ctg ggt gtc cgt ctg tgg gtc aac atg ctt gcc gga cac acg
 T F R A V S I G V R L W V N M L A G H T
 ctg ctg cac atc ctg acg ggc atg gcg ctg gcg ctg ccc ttc tcg ctg tcg ttc ttc gcc
 L L H I L T G M A L A L P F S L S F F A
 atg gtg ccc gcc acc ttc gcc gtc gcc tgc ctg ctg tcg gcg ctg gtg ggc ctg gag tac
 M V P A T F A V A C L L S A L V G L E Y

Intrón 1

gtgaggcgggtggccgatttatgtgctgcgaacagctcttccgcgccaccatcacgagctgctgcttccgga
cctgcctcacttgctcggctcccgataatcctcaatcccacag

Intrón 2

gtgagctgcggaatgctggctcgggtgcgagattgctcctcgctgatgttgggaactgactcctgctcgcag

Intrón 3

gtaggagccgtgctagcgcagaagggccgcccacaaaacttgctcttccggtatcggagtagagctcgaccggg
cgtgctcctggccgcag

Intrón 4

gtgagcgcgaggtgcagaacataaaggctacggagctgcattacgggctttagcgtgaaaggattgtgggga
gctctgggttcgggagtggtcctgagggcggggaaactctctggagggctcggcgacctgggttgggttgctg
gtaagctgctgccctcactcagctgctgccccactcctcgcgccacgcgaatttgcggtctcactccacg
cgcccataatgctaccatcgccctccttgcttgtgaacctgcag

Intrón 5

gtaagcagctgaggcaatcctgagggcatggctcggggcgagcagggaggaagtctgagctgcggacaagccta
atgccggggtacggctgctgctggagaggggtgctggattgaggcaattgtgctggttacgggggcatcacct
gctgacaatcatgagcgaggaataagcaggatggctgctgcatgggacgagaatctatgctgggatccttg
acgcggtctcacctcctgctcctgcgcctattccccccccgcag

Intrón 6

gtaggaacgggctgggtgcgggagcccaacgcattgaagggttgaggagcagcgggtttgtgtttgcacaggtcag
caccgcgtgcggggtgtgccgggtatgcacggcatgtgctgggtgcgtgcacgccaacaccgaatggcaaaa
ttgtgctgactagttggttggtgctccttgccgaccgcag

Intrón 7

gtgagtcgctgccgttcgcacagctgcaactgtaaaactacagattaggtgcgctctctgggaccctgcgacgcc
acaacctacaccgctgcttccagaaccccactggctgccaccccgctccccacgcgctctctcacaaccgg
tactcttcccttgag

nd4L de *C. reinhardtii*

(-45) tccaagctgaattccggatgtcttgcctgtaatagagtcacaaga

atg tcg cgg acg cag ctg ctc agg cgt ctg ctt ccc ggc gtg ccc ggc acg cct ggc ctc
M S R T Q L L R R L L P G V P G T P G L

ttc tcg cag tct tgc acc gct gtg cca aat ggc ctg cag cag tgc ggc ttt ttc agc agc
F S Q S C T A V P N G L Q Q C G F F S S

cac gag gag gag cag aag cag caa agc agc ctg cag ccg gct tcc tcg tcg ctc gtg cag
H E E E Q K Q Q S S L Q P A S S S L V Q

ttc gca aac att gtc aac agg ccc atg ccc gtg ccg ctt gcc acc gct gcg ggc ttc gct
F A N I V N R P M P V P L A T A A G F A

gcc tcc cct ctc atg gca atg cca gct cgg cga ggc gcg ggc gtg atg gga gtc agg cgt
A S P L M A M P A R R G A G V M G V R R

cca gct ctg cca atg ctg ccc ggc atg ggt ccc aca acc acg ccc tct gca gcc ctg gcc
P A L P M L P G M G P T T T P S A A L A

cgc tct tac tac aca gag ggc gag att ttc ggc acc acg ttc atg tac acc aca aac atg
R S Y Y T E G E I F G T T F M Y T T N M

atg ttc tgg gcc ggt atc gtg ggc gcg gtg gca ttc agg cgc aac ctc atc atc ctg ctg
M F W A G I V G A V A F R R N L I I L L

ctc agt gcg gag acg gtg atg ctt gcc tgc aac atg aac ttc ctg ttc aca gcc gct tac
L S A E T V M L A C N M N F L F T A A Y

ctc aac gac atc acc ggc gct att atg tca att acc atc acc acc atc gcc gcg tgc gag
L N D I T G A I M S I T I T T I A A C E

acg gcc atc ggc ctg gcg ctg tgc gtg gcg tac ttc cac atg cgc tcg gcg acg gac gtg
T A I G L A L C V A Y F H M R S A T D V

gag gcg ctg aac atg ctg aag tag atc ggc ggg agc tag cat ggg cag caa cag cag ccg
E A L N M L K

gagcagcagcgacaggcgtggcgggggggccgaaacgcgtgcgggcacaaacacgcgtgcagcgacgttccagcaacggcgcg
ggccgcagtgggcgccgctgtgagctagcgcagctgaaagcagatgggggtctggaacggagccagctggcgcgtggcatggg
cactgctgtagccgtgatggagctatgagcgcctctcacaacggtgcggcgggagtgagctttaaagcgatgacgagcag
tatactggcgggcctgggggtggcaagggaaggggcgagcctgccagctgccttgggggcttggggctacatgagggcagg
tgcttgcatgcaaaagcaacttgttgcgagggctggaggcatggcgagagactccgtgaaatagccgagagcgggctgcggtgt
cggagtggcaagatggcgcgctggcgcgacacaaggctgcggtgtgcccgtgcaaggcgtccggccatgttggcttgcgt
gcacgcagaggaatgtgacagggcttatgcattagggcttaatggcatttgacttgagacagtagcgggtgggtgatgag
cgaggggtgaaaatagtagccagaagttgtagattcttacagcagcatgtggagagcgcatgctgcagacagtgctcaggcat
tgtagaataaagtggcagtggaatgcaccaaggtggcagtggaatgcaccgacgatgcaaggaacatgggtagggggtt
tgggtgatgctgtacgtcggggcgaggagccgggattggagcctgggggagacgtggggaagcactgccactgctcgtc
cagtgagcgggctggatgctcaatgggtgtaaacccgcagatagccggaattcttttgcctttt (1647)

atgtccggttctatcgagtcgccaatcacgcgtgccatgagagaagtgacaaaaggataagtcggatccgactgtgtggact
accggacttgaacaaacattctgtgttattcaggaaaggc

Intrón 1

gtacgtccgtcatgagggattcgcgggggtctgtaggccacagaaaaccgacactcttgagccaccgcaaggtcaatg
gaaacgcttcttgttcccgtaatggcaaacgaggtgtctacgtggcatgacccaacaaattctgtgcatttacaagacttt
ttccgttttatttaagatccgtgaaaagtattgacctaatcgtggcaaaaggttggtatgtagggtaggcgacaaagggtg
cgaatcttgcattgatgttagtcggttgattgtgggttttttcag

Intrón 2

gtaagagacgaaagagtgttttgcattcaacaggcctgcgacgctgtagatgactgtgaaaggctcgaactgtgacactt
caccgaaatggcacttgagtgacgaagttccagtgccctgtgacagatgaatttgcctcgcattatgggtattgggtat
tagaggcgaaacgtgaataaaaatgcgcaattttatgttttcgtgagacgcacttgaacgagcagggcggcctgctattat
gttctttatgtcggaagagaggtgacaaagaacgtattcaacgggacgaggttccctgtggcagaaaaatgccgtgtgcatt
ggaggggctagcagttattggacagatgctgagtgggcagttctgtggtgtccgtttaaaaaacaatgtcactgtctcaag
agcatgtgcataaagacaaaggcatgtccctagaggctattgtcgaaggcgaatgttgcactgcaccaagagtacat
gtgtgcgagcactgaatacccatgggcttttctggggacctagcagcattctgtctacgaatggttgcggtgttctgcg
cag

cox2b de T.gondii

(-60) cacgagcttggttgtctttctcgcgacctctgttaccatttctagcctcggettcgcc

atg gat gtt ttc cgg tcc ttc tgg aac cgg cgg ggt gcg gcc ccg ggg ggc gcc gcc acc
M D V F R S F W N R R G A A P G G A A T

agc tca ggt cgt ccc gcg gga tgc cgc gag gag aac tcg tcg gtt ttg ctg cag tct ctg
S S G R P A G C R E E N S S V L L Q S L

aag gct aac tct cag ⁵⁶cag aaa tac ctc gat gac ccg gac aaa att ccc acc tac tac gtg
K A N S Q Q K Y L D D P D K I P T Y Y V

ttc cag tct aac atg gtg aca gat gag gat ctg ttg ccg gga atg ctc cgc aat ttg gaa
F Q S N M V T D E D L L P G M L R N L E

gtc gac aaa cgc ctg act ttg ccg acg cgt acg cac att cgc ttt ctt att acc gcc aca
V D K R L T L P T R T H I R F L I T A T

gac gtc att cac tcc tgg gct gtg cct gcg cta ggc atc aag gct gac gcc atc ccg ggt
D V I H S W A V P A L G I K A D A I P G

agg ctg cag cga atc aac act ttc atc cag aga gaa ggc gtt ttc tat gga cag tgc tca
R L Q R I N T F I Q R E G V F Y G Q C S

gag ctc tgt gga gcg ctc cac ggg ttc atg ccc atc gtc atc gaa gcc gtg agc ccc gaa
E L C G A L H G F M P I V I E A V S P E

acc tac gcg gcg cat gca aag aag tgg tac aaa gac tga gaa ctg gta ttc aag tgg cag
T Y A A H A K K W Y K D *

ctcggacactcgtgaacgaccaggccagaaagtgatcatcgggtcccacattccgcacaattgtgtcctgtcgtctcttggtta
gagcacaactcttatgtggagtgaagaaatgcgcgaggaaggaggcaagcccgcgaggacgtgggtgcacacgaccaggc
aactcttgttgaaaaacagcatcaaaagaaaaatcctccgatttcgcggtgaccattcgttctgtggtaaaagtggcaagatt
gagacggtgtgtgtctcatggagaagccctcggggagggttccaatgtctccagtcctgtagcaagctatgatgggatgac
cgcgacagctagtcacgaggggttcaacctcctcagtgagctctgaaacaaagagagatcggaaaagatcagcaggaacaac
aaatagtgcacgcaacgttgggatgaacgcttgcgaaaactgctacgacgttccatagacgcgcttttccgggctctg
gaagtaacatgttctgcaaaagcagctgcaggattttcttcgcgtgtctcctttgctgttgaacgttctcgttcggtccg
ttcattgtttcgtgcgctgttatccacgtttctccttagtacacagatatcgcgcatctgggaaatcctttgttgaaga
cactcaattacttctgccaatgtgatggcaagccct

Intrón 1

aaaaagggtgaaacgcctctacagaatgtggccgctcggactcgtgacgccccag

cox2a de T.gondii

(-122) aaggagcgtttgaacgcaggcttttgggtggagcaggtta
tgcgggaggacgcactgtactttgcaaggccctcttgcaagtctgtccctcaaaaacgcactgcgaggcggtattccgtggtg

atg cta gca tat ttt gct gca agc agg gaa ctc gct gtt gtt ttc tgt gtc ggc aca acc
M L A Y F A A S R E L A V V F C V G T T

ctt ccg gag aac att ttt ctc ttt ctt tct cga cgg ctt gcc att atc cca gcg tct tcc
L P E N I F L F L S R R L A I I P A S S

ccg tgt tcc ctt ttc ttc ggt tct cga ctg tat ccg atc cgt agc tgt tgc gga cgc ccg
P C S L F F G S R L Y P I R S C C G R P

gct gga ttc cct gcg aaa att gtc tac ccg ttt ttg tct gcc cac acc ttt ccg aag atg
A G F P A K I V Y P F L S A H T F P K M

atg ccg ctg tcc tgg tct ctt ttc cga ccg tat ggt cac ctt ttc cac cag ttc ccc cgg
M P L S S S L F R P Y G H L F H Q F P R

cgc ttc ttt tcc tct gca cag cgg ggt gac tat tct ttg atg ctc ccg cgc cat ttc tcc
R F F S S A Q R G D Y S L M L P R H F S

tcc acc gcg tgg tct gtc acg aaa ccg ccg cag aag acg gat gcc ccc gca aaa cat gat
S T A S S V T K P P Q K T D A P A K H D

gag cac gac tct cac cac ggc aca aac aac ttc tac cac atg cct tcc cac cac tct ccg
E H D S H H G T N N F Y H M P S H H S P

tcc cgt cac cat ctg aat cct gat gga aca atg cgc gac ctc acg acg gca gag act ttc
S R H H L N P D G T M R D L T T A E T F

cac tgg gag cac gct gaa gca gaa acg cct gcc cag cag att gtt tcc gtt aac gga cgc
H W E H A E A E T P A Q Q I V S V N G R

aag atg gtc aag ggc gtt gag acc aga gat ctc gtc gag ctc ttc ctt gtt cac caa²⁸⁵ aaa
K M V K G V E T R D L V E L F L V H Q K

aac att cct ttc tgg cca cga atg cgt atg aac gtc tgg ggc aac cac gac ttg ctc atg
N I P F W P R M R M N V W G N H D L L M⁵⁶³

aag gct gag ttc ctt ttc ttc tgg act ccc aca ttc atc acc tgg tgg ctt gca att ccg
K A E F L F F W T P T F I T W S L A I P

atg ttc acg ctc ttg tac atg ctc gac gaa gcg gtg tat gcg gcc atg acc gtc aaa gtt
M F T L L Y M L D E A V Y A A M T V K V

atc ggc aga cag tgg tac tgg atc tac gaa gtc gag tgg cct gtt gac gac gag gag tag
I G R Q W Y W I Y E V E S P V D D E E *

agagcacaggcgcttctctatgcgtttctctaaaggacggctgtccgagcgttgaatgggcagttgatggtctgcacccgaa
gcaaatggaacaactggatgcccagctttgccagcattctttgagaggttttggaaatcacagtcggttttctcgagttcag
agcggctctgggacagcagcctgggttttgagaagcgcgggaacggcactgctgtaaacgacgcctcatgctgcgctg
gatgagcttgtgaacccgagatgaaatcgacgtttctgtgacaattgagtggtggatggaataacactttgtgaaatc
aaatgtgagtaccgtgaaatacagaccagcgggcccggagatggttctctggacacggcactgcagaagtggtgctt
aaccaggctgtgctcacttgggaagcagacacccgctagcagtggtgaacataaataatgccaattgaaagaaagcggctc

cox2a de T.gondii

(-122) aaggagcgtttgaacgcaggcttttgtgtggagcaggtta
tgcgggaggacgcactgtacttgcaaggccctcttgacgtctgtccctcaaaaacgcactgcgagggcggat tccgtggttg

atg cta gca tat ttt gct gca agc agg gaa ctc gct gtt gtt ttc tgt gtc ggc aca acc
M L A Y F A A S R E L A V V F C V G T T

ctt ccg gag aac att ttt ctc ttt ctt tct cga cgg ctt gcc att atc cca gcg tct tcc
L P E N I F L F L S R R L A I I P A S S

ccg tgt tcc ctt ttc ttc ggt tct cga ctg tat ccg atc cgt agc tgt tgc gga cgc ccg
P C S L F F G S R L Y P I R S C C G R P

gct gga ttc cct gcg aaa att gtc tac ccg ttt ttg tct gcc cac acc ttt ccg aag atg
A G F P A K I V Y P F L S A H T F P K M

atg ccg ctg tcc tcy tct ctt ttc cga ccg tat ggt cac ctt ttc cac cag ttc ccc ccg
M P L S S S L F R P Y G H L F H Q F P R

ccg ttc ttt tcc tct gca cag ccg ggt gac tat tct ttg atg ctc ccg cgc cat ttc tcc
R F F S S A Q R G D Y S L M L P R H F S

tcc acc gcg tcy tct gtc acg aaa ccg ccg cag aag acg gat gcc ccc gca aaa cat gat
S T A S S V T K P P Q K T D A P A K H D

gag cac gac tct cac cac ggc aca aac aac ttc tac cac atg cct tcc cac cac tct ccg
E H D S H H G T N N F Y H M P S H H S P

tcc cgt cac cat ctg aat cct gat gga aca atg ccg gac ctc acg ccg gca gag act ttc
S R H H L N P D G T M R D L T T A E T F

cac tgg gag cac gct gaa gca gaa acg cct gcc cag cag att gtt tcc gtt aac gga cgc
H W E H A E A E T P A Q Q I V S V N G R

aag atg gtc aag ggc gtt gag acc aga gat ctc gtc gag ctc ttc ctt gtt cac caa²⁸⁵aaa
K M V K G V E T R D L V E L F L V H Q K

aac att cct ttc tgg cca cga atg cgt atg aac gtc tgg ggc aac cac gac ttg ctc atg
N I P F W P R M R M N V W G N H D L L M

aag gct gag ttc ctt ttc ttc tgg act ccc aca ttc atc acc tgg tcy ctt gca att ccg
K A E F L F F W T P T F I T W S L A I P

atg ttc acg ctc ttg tac atg ctc gac gaa gcg gtg tat gcg gcc atg acc gtc aaa gtt
M F T L L Y M L D E A V Y A A M T V K V

atc ggc aga cag tgg tac tgg atc tac gaa gtc gag tcy cct gtt gac gac gag gag tag
I G R Q W Y W I Y E V E S P V D D E E *

agagcacaggcggcttccctatgctgttccctaaaggacggctgtccgagcgttgaatgggcagttgatggctctgcaccagaa
gcaaatggaacaactggatgcccgcagtttgccagcatctttgagaggttttggaatcacagtcgtttttctcgagttcagc
agcggctctgggacagcagcctcggttttgagaagcgcgggaacggcactgctgtaaacgacgctcatgtgctgcgctcgc
gatgagcttgtgaaccgagatgaatcgacgttttcgtgtacaattgagtggtggatggaaaaatacactttgttgaatac
aaatgtgagtaccgtgaaaaatacgaaccagcggcggcggagatatgttccctcggacacggcactgcagaagtggtgcggtt
aaccaggcgtgctgctcacttgaagagcagacaccgctagcagtggtgaacataaataatgccaatgaaagaagcggctc

Apéndice XI.
Oligodesoxinucleótidos utilizados

OLIGODESOXINUCLEOTIDOS UTILIZADOS

A continuación se presenta una lista de los oligodesoxinucleótidos que se utilizaron a lo largo de este trabajo. Se incluye el nombre designado para cada uno, la secuencia de nucleótidos en dirección 5'-3', la longitud y la temperatura de apareamiento (tm) en °C.

atp6 de C. reinhardtii

<i>cra61f</i>	aga cga aga ata tag att gga tta acc	27mer	tm 52.2
<i>cra623f</i>	gag ggt gtt cgg cct ctt gg	20mer	tm 58.1
<i>cra6293f</i>	cat tcg ctg ccc agc agg c	19mer	tm 61.0
<i>cra6520f</i>	cct cca act tgc tgg gtc tgg	21mer	tm 58.2
<i>cra61010f</i>	cta cgt tgg cga gtt caa cag c	22mer	tm 57.0
<i>cra61060f</i>	gtg gtc aag aag gcg ctg taa gc	23mer	tm 59.0
<i>cra61027f</i>	aag gtg gtc aag aag gcg ctg taa gc	26mer	tm 63.6
<i>cra6340r</i>	ggc agc gaa tgg cac cat cg	20mer	tm 58.0
<i>cra6429r</i>	cgc cag aag cgg tag atg cc	20mer	tm 58.0
<i>cra6646r</i>	gag aag ccc agc ttg tac aga cc	23mer	tm 59.0
<i>cra6810r</i>	cca tgc ccg tca gaa tgt gc	20mer	tm 58.4
<i>cra6843r</i>	cga aga acg aca gcg aga aag gc	23mer	tm 61.1
<i>cra61378r</i>	aca cag caa cat cta agc gac tc	23mer	tm 52.9
<i>cra62349r</i>	gag acg aag ctt aca gtc tcc tcg	24mer	tm 58.8

nd4L de C. reinhardtii

<i>crnd41r0</i>	tat gcg cgg ttt aca ccc att g	22 mer	tm 59.3
<i>crnd41f1</i>	taa tag agt cac aag aat gtc gcg g	25 mer	tm 56.5
<i>crnd41r2</i>	ccg cga cat tct tgt gac tc	20 mer	tm 53.0
<i>crnd41f2</i>	gta cac cac aaa cat gat gtt ctg g	25 mer	tm 54.9
<i>crnd41f3</i>	gct att atg tca att acc atc acc acc	27 mer	tm 56.5
<i>crnd41r1</i>	ggt ggt gat ggt aat tga cat aat agc	27 mer	tm 56.5

cox2b de S. obliquus

<i>soco2b3utr</i>	tat cct gct gct gca gaa cta cc	23 mer	tm 55.3
<i>soco2b 5'</i>	tgc cca gga gtg gat gac gtc ac	23 mer	tm 62.8
<i>soco2b 5'n</i>	gca cca ggc gct cgt cac agt c	22 mer	tm 63.7
<i>soco2b 3'</i>	acg gcc agt gct ccg agc tgt g	22 mer	tm 65.2
<i>soco2b 3'n</i>	aac cac gcc ttc atg ccc atc gtg	24 mer	tm 67.2

cox2a de T.gondii

tgco2af0	aag gag cgt ttg aac gca gg	20 mer	tm 54.0
tgco2af1	ctc gac tgt atc cga tcc gta gc	23 mer	tm 56.7
tgco2af2	gaa gat gat gcc gct gtc ctc	21 mer	tm 56.0
tgco2af3	gat gtt cac gct ctt gta cat gct c	25 mer	tm 56.8
tgco2af4	aag aag tga caa aag gga tta gtc g	25 mer	tm 54.2
tgco2af5	act tga aca aac att ctg tgt tat tca g	28 mer	tm 54.3
tgco2ar1	gcc ttt cct gaa taa cac aga atg	21 mer	tm 53.9
tgco2ar2	cac aca atc gga tcc gac taa tc	20 mer	tm 55.0
tgco2ar3	gag cat gta caa gag cgt gaa cat c	25 mer	tm 56.8
tgco2ar4	tca gat ggt gac ggg acg gag	21 mer	tm 59.4
tgco2ar5	agc tac gga tcg gat aca gtc g	22 mer	tm 54.9

cox2b de T.gondii

tgco2bf2	cg acc tct gtt acc att tcc	21 mer	tm 52.8
tgco2bf1	cac gag ctt tgt ttg tct ttt ctg c	25 mer	tm 58.3
tgco2br2	gca gaa gta att gag tgt ctt caa caa agg	30 mer	tm 59.9
tgco2br1	agg gct tgc cat cac att gg	20 mer	tm 57.0

Otros oligodesoxinucleótidos

para amplificar productos del vector λ gt10

Lgt10F	agc aag ttc agc ctg gtt aag	21 mer	tm 52.0
Lgt10R	ctt atg agt att tct tcc agg gta	24 mer	tm 52.0

para amplificar productos del vector pBluescript

bst7	gcg cgt aat acg act cac tat agg g	25 mer	tm 57.3
bst3	gcg caa tta acc ctc act aaa gg	23 mer	tm 56.1