



Mitochondrial Cytochrome C Oxidase Subunit I of *Manduca sexta* and a Comparison with Other Invertebrate Genes

D. R. Frohlich, B. A. Stevenson, A. M. Peterson and M. A. Wells

DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF ARIZONA, BIOLOGICAL SCIENCES WEST 440, TUCSON, AZ 85721, USA

ABSTRACT. A cDNA encoding mitochondrial cytochrome *c* oxidase subunit I (mt COI) from *Manduca sexta* (Lepidoptera: Sphingidae) was cloned and sequenced. AT (adenine-thymine) content is high and codon usage is biased and likely reflects the role of mt COI in electron transport. The encoded protein is 514 amino acids long, contains seven invariant His residues observed in COIs in all organisms and would be predicted to be composed of 12 transmembrane regions. COMP BIOCHEM PHYSIOL 113B, 785–788, 1996.

KEY WORDS. Cytochrome *c* oxidase I, Lepidoptera, *Manduca sexta*, Sphingidae

INTRODUCTION

The tobacco hornworm moth, *Manduca sexta*, is commonly used to model biochemical and physiological systems in insects. Because of its wide utility, we have initiated an effort to characterize the mitochondrial genome and assess its usefulness in insect phylogenetic and population analysis. The animal mitochondrial genome is unique in that it is small, circular, consists almost entirely of coding sequences (13 protein genes, 22 tRNA genes, 2 rRNAs) and is maternally inherited. Sequences of some mitochondrial protein coding regions (e.g., cytochrome *b*) are believed to be so highly conserved as to be of use in resolving deep phylogenies. Yet, certain other portions of the mitochondrial genome, such as the control region, exhibit enough variability to allow characterization of single populations (1).

Of the proteins encoded by animal mtDNA, three are subunits of cytochrome *c* oxidase that form the membrane-bound catalytic core involved in the final steps of electron transport (2,3). The complex uses heme and protein-bound copper atoms to accept four electrons from cytochrome *c*, transferring them to a single bound O₂ molecule and ultimately producing two molecules of water by a mechanism that is not completely understood. Although it is estimated that the cytochrome oxidase reaction accounts for some 90% of oxygen uptake in cells, no mitochondrial cytochrome *c* oxidase proteins have as yet been purified and crystalized. Thus, no three-dimensional structures are known, and protein function has been deduced only from peptide sequences inferred from nucleic acid sequences (2,3).

Among the insects, six complete mitochondrial cytochrome

c oxidase subunit I (mt COI) sequences have been reported (5 Diptera, 1 Hymenoptera), and over the invertebrates as a whole, at least 11 complete sequences, including the insects, are known. Herein, we report the first complete nucleotide sequence of an mt COI cDNA from a lepidopteran, *M. sexta*, examine codon usage patterns and compare representative complete invertebrate mt COI nucleotide and inferred peptide sequences.

MATERIALS AND METHODS

A partial clone of mt COI cDNA was obtained from a cDNA midgut library and subcloned into a pBluescript SK (Stratagene, La Jolla, CA) vector as described by Peterson *et al.* (4). Because the length of the clone was approximately 1.6 kb, ExoIII (Stratagene, La Jolla, CA) deletion mapping was used to generate a population of smaller subclones for sequencing. Double-stranded sequencing was performed as described by Sambrook *et al.* (5).

Based on similarity to other known mt COI sequences, the length of the original cDNA was judged to be approximately 30 bp short of full length at the putative 5' end. Consequently, we used the polymerase chain reaction (PCR) to amplify and obtain the missing sequence. Two oligonucleotide primers were designed, one based on known *M. sexta* sequence (5'-AAGCTTAAAGAAGTTCCAACCTATTCCTGCT-3') and the other based on *Drosophila yakuba* sequence from mt NADPHase 2 (5'-GGTGGATTACCTCCATTTTTAGGATTTTTACC-3'). The second primer sequence was chosen based on the assumption that gene order was conserved and that NADPHase 2 was adjacent to COI as in *D. yakuba* (6). A single PCR product, 0.5 kb long, was obtained and subcloned into a commercially available T-vector (Invitrogen, San Diego, CA) (7). PCR conditions were as reported by Kocher *et al.* (8).

Correspondence to: Department of Biology, University of St. Thomas, 3800 Montrose Blvd., Houston TX, 77006, USA. Tel 713-525-3165; Fax 713-525-2125.

Received 21 August 1995; revised and accepted 20 October 1995.

DNA and amino acid sequences were aligned and compared with the aid of the Genetics Computer Group (Madison, WI) (GCG) package based on the method of Needleman and Wunsch (9). Distance and parsimony analyses were performed on nucleic acid sequences, inferred translated sequences and second codon positions with the GCG package (10), the branch and bound and bootstrapping options of PAUP (Phylogenetic Analysis Using Parsimony) (11) and the 3.04 release of MacClade (12).

The invertebrate species, with known complete COI sequences, used for comparison and their GenBank accession numbers were Platyhelminthes: *Fasciola hepatica*, M93388; Nematoda: *Caenorhabditis elegans*, X54252, *Ascaris suum*, X54253; Echinodermata: *Pisaster ochraceus*, X55514, *Paracentrotus lividus*, J04815; Arthropoda: Insecta: Diptera, Culicidae: *Anopheles quadrimaculatus*, L04272, Diptera, Calliphoridae: *Lucilia illustris*, L14945, *Phormia regina*, L14946, *Phaenicia sericata*, L14947, Diptera, Drosophilidae: *Drosophila yakuba*, X03240, Hymenoptera, Apidae: *Apis mellifera*, M23409.

The *M. sexta* mt COI discussed herein is GenBank Accession No. U09843.

RESULTS AND DISCUSSION

M. sexta mt COI

The initiation codon of *M. sexta* mt COI (ATT) begins at nucleotide 74 (sequence not shown) and aligns with the putative initiation tetranucleotide (ATAA) of *D. yakuba* mt COI (6). Codons ATT, ATG and ATA are believed to act as initiation codons in an array of animal mitochondrial genomes (6), and ATT has been shown to be the initiation codon for mt COII in two beetles (13). ATT encodes Ile internally but may code for Met in the initiation position in animal mitochondrial genomes (14).

Like other mitochondrial genomes, no specific termination codon is present. In addition to TAA, TAG, AGA, AGG

and GTG, a single T immediately adjacent to the 5' terminal nucleotide of the sense strand of a tRNA gene is believed to signal termination in a number of animal mitochondrial genomes (6). A single T occurs at nucleotide 1616 and signals the end of the open reading frame. If *M. sexta* mitochondrial gene order is conserved, with respect to *Drosophila*, the 5' end of a tRNA gene for leucine would be expected to immediately follow the 3' portion of *M. sexta* COI.

A search of GenBank with the 5' portion of the *M. sexta* sequence (nucleotides 1–73) indicates greatest similarity to regions that encode an mt tRNA for tyrosine, on the complementary strand, in three other invertebrates. In fact, the sequence can be folded into the standard cloverleaf secondary structure and displays the correct anticodon for tyrosine, GTA (residues 50–48).

Codon usage appears to be the same as in other insect mitochondrial genomes, that is, TGA encodes Trp and is not a termination codon, ATA encodes Met and not Ile and AGA encodes Ser not Arg (6). Like other known mitochondrial sequences, AT content is relatively high at 72.5% overall. Only 7% of all codons used end in G or C; and four doublets end only in A or T: AAX, CTX, GAX and GTX. Codon usage bias is quite strong. Table 1 shows the codon usage frequency for *M. sexta* mt COI. Note that for two acidic amino acids (Asp and Glu), two basic amino acids (Lys and Arg) and two polar amino acids (Asn and Gln), only a single codon is used. For His, an important amino acid in binding the two oxidation-reduction centers involved in electron transport, an alternate codon is used only once.

The inferred amino acid sequence of *M. sexta* COI is shown in Fig. 1. If the inferred protein sequence is subjected to Kyte–Doolittle (15) analysis, the hydrophobicity plot in Fig. 2 is produced. From the plot it can be seen that *M. sexta* mt COI would be predicted to be a membrane spanning protein composed of 12 hydrophobic regions that is identical to profiles produced by prokaryotic counterparts as well (2,3).

TABLE 1. Codon usage by *Manduca sexta* mitochondrial cytochrome c oxidase I (first column) and all invertebrate taxa combined (second column)

Arg	AGG	0	11	Gln	CAG	0	7	Gly	GGG	9	76	Stop	TAG	0	2
Ser	AGA	6	69		CAA	9	94		GGA	27	306		TAA	0	8
	AGT	3	83	His	CAT	15	147		GGT	8	154	Tyr	TAT	17	167
	AGC	1	4		CAC	1	56		GGC	0	14		TAC	4	64
Lys	AAG	0	36	Leu	CTG	0	12	Glu	GAG	0	15	Leu	TTG	3	117
	AAA	6	66		CTA	5	99		GAA	9	86		TTA	51	470
Asn	AAT	25	182		CTT	8	105	Asp	GAT	15	136	Phe	TTT	35	405
	AAC	0	28		CTC	0	11		GAC	0	46		TTC	4	118
Met	ATG	1	112	Pro	CCG	0	13	Val	GTG	0	26	Ser	TCG	0	7
	ATA	28	239		CCA	10	97		GTA	14	169		TCA	16	135
Ile	ATT	57	478		CCT	13	159		GTT	8	216		TCT	16	183
	ATC	2	24		CCC	3	31		GTC	0	16		TCC	1	26
Thr	ACG	0	10	Trp	TGG	0	35	Ala	GCG	0	8	Arg	CGG	0	6
	ACA	16	141		TGA	14	145		GCA	16	112		CGA	9	73
	ACT	10	200	Cys	TGT	1	19		GCT	11	220		CGT	0	25
	ACC	2	27		TGC	0	1		GCC	3	45		CGC	0	6

```

      10           30           50
ILFLRKLWLYSTNHKDIGTLYFIFGIVWAGMVGTSLSLLIRAEELGNPGLIGDDQTYNTIVT
  ↓           70           90           110
AHAFIMIFMFPMPIMIGGFGNWLVLPLMLGAPDMAFPRMNMNSFWLLPSSLMLLISSSIVE
  ↓           130          150          170
NGAGTGWIVYPPLSSNIAHSGSSVDLAIFSLHLGAISSILGAINFITTIINMRINNSFD
  ↓           190          210          230
QMPFLVWAVGITAFLLLLILFVLGAIITMLLTDRLNLTSPFDPAAGGDPILYQHLFWFFG
  ↓           250          270          290
HPEVYILILPGFCMISHIISQESTKKEITFCGLMIYAMMAIGLLGFIVWAHMFETIGMDI
  ↓           310          330          350
DTRAYFTSATMIIAVPTGKIFSWLATLHGTOINYNPSILWSLGFVFLFTVGGGLTGVLIA
  ↓           370          390          410
NSSIDITLHDTYVVAHFHYVLSMGAVFAIMGGFIHWYPLFLGLNLNPNYLLKIQFFIMFL
  ↓           430          450          470
GVNLTFFPQHFGLAGMPRRYSYDPSYISWNLISSLSGSYISLLAVMMILIIWESMTYQ
  ↓           490          510
RITLFLPLMSSSIEWYQNLPPAEHSYNELPILSN

```

FIG. 1. Amino acid sequence of *Manduca sexta* mitochondrial cytochrome c oxidase I. Histidines that are invariant in all known COI sequences are indicated by arrows.

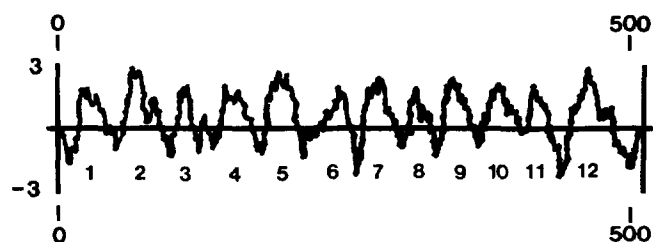


FIG. 2. Hydrophobicity plot showing twelve transmembrane regions of *Manduca sexta* mitochondrial cytochrome c oxidase I. Hydrophobic regions project above x-axis.

Invertebrate mt COI Comparisons

If the representative invertebrate mt COI inferred protein sequences available in GenBank are aligned (data not shown), a high degree of conservation is apparent and only the extreme termini are highly variable. Four amino acid regions are invariant over all taxa and surround six of seven invariant His residues: amino acids 233–245, 287–293, 368–371, and 377–383. Seven His residues that have been found to be invariant in other organisms (3) are His₆₂, His₂₃₄, His₂₄₁, His₂₉₁, His₂₉₂, His₃₆₉, and His₃₇₇ (Fig. 1). If a hydrophobicity plot is constructed for each inferred invertebrate protein according to the method of Kyte and Doolittle (15) and is overlaid onto the same plot for *M. sexta* in Fig. 2, the profiles are identical and sequences composed of 12 hydrophobic or transmembrane regions (data not shown) are predicted.

Tables 2 and 3 show distance matrices (PAUP) for the invertebrate mt COI sequences based both on nucleic acid and amino acid sequence. Not surprisingly, mean distances are rather high. At the amino acid level, mean distances vary from 0.01 to 0.56, whereas at the nucleic acid level, mean distances vary from 0.06 to 0.50. Interestingly, these differences are more conservative than for similar comparisons among the insects for COII (13). Nucleotide divergences between within-order taxa, for example, are as high as 45% (beetles) with respect to COII, whereas within-order differences with respect to COI here are as low as 6–20% among the five dipteran sequences.

Because of its role in electron transport and general conservation, we also attempted to assess the usefulness of mt COI nucleotide (unweighted and weighted second positions of codons) and amino acid sequences for phylogenetic analysis of the higher invertebrate taxa. Using parsimony methods (branch and bound, bootstrapping, PAUP, MacClade), we consistently found only one most parsimonious reconstruction

TABLE 2. Distance matrix for invertebrate mitochondrial COI nucleic acid sequences (PAUP)

	1	2	3	4	5	6	7	8	9	10	11	12
1 <i>Caenorhabditis elegans</i>	—	0.21	0.41	0.40	0.35	0.35	0.36	0.33	0.37	0.36	0.37	0.45
2 <i>Ascaris suum</i>	326	—	0.42	0.43	0.37	0.37	0.38	0.36	0.39	0.36	0.39	0.38
3 <i>Pisaster ochraceus</i>	633	646	—	0.21	0.32	0.32	0.32	0.31	0.32	0.33	0.36	0.49
4 <i>Paracentrotus lividus</i>	625	671	328	—	0.32	0.31	0.32	0.32	0.32	0.33	0.36	0.50
5 <i>Lucilia illustris</i>	540	571	493	491	—	0.06	0.09	0.13	0.17	0.19	0.26	0.46
6 <i>Phaenicia sericata</i>	540	567	493	474	85	—	0.09	0.12	0.16	0.19	0.26	0.45
7 <i>Phormia regina</i>	556	576	495	487	139	138	—	0.13	0.18	0.20	0.28	0.46
8 <i>Drosophila yakuba</i>	511	552	480	483	197	184	199	—	0.17	0.20	0.27	0.44
9 <i>Anopheles quadrimaculatus</i>	563	595	491	488	261	252	283	262	—	0.19	0.27	0.46
10 <i>Manduca sexta</i>	531	556	514	513	293	293	304	299	296	—	0.26	0.45
11 <i>Apis mellifera</i>	576	608	554	549	399	406	424	409	416	393	—	0.46
12 <i>Fasciola hepatica</i>	698	584	740	762	697	687	700	675	708	680	699	—

Below diagonal = absolute distances. Above and below the dashes = mean distances.

TABLE 3. Distance matrix for invertebrate mitochondrial COI peptic sequences (PAUP)

	1	2	3	4	5	6	7	8	9	10	11	12
1 <i>Caenorhabditis elegans</i>	—	0.12	0.42	0.40	0.38	0.37	0.37	0.38	0.40	0.39	0.42	0.56
2 <i>Ascaris suum</i>	62	—	0.43	0.41	0.37	0.37	0.37	0.38	0.39	0.39	0.43	0.55
3 <i>Pisaster ochraceus</i>	217	221	—	0.14	0.30	0.30	0.30	0.29	0.30	0.32	0.39	0.55
4 <i>Paracentrotus lividus</i>	207	211	73	—	0.27	0.27	0.27	0.26	0.27	0.28	0.36	0.55
5 <i>Lucilia illustris</i>	192	191	155	139	—	0.01	0.02	0.04	0.12	0.16	0.31	0.53
6 <i>Phaenicia sericata</i>	191	191	155	138	3	—	0.01	0.04	0.12	0.16	0.31	0.53
7 <i>Phormia regina</i>	191	191	154	136	9	7	—	0.04	0.12	0.15	0.31	0.53
8 <i>Drosophila yakuba</i>	187	191	150	133	20	19	19	—	0.11	0.16	0.31	0.53
9 <i>Anopheles quadrimaculatus</i>	204	201	153	140	62	62	59	54	—	0.16	0.32	0.53
10 <i>Manduca sexta</i>	201	200	162	144	81	81	78	80	81	—	0.30	0.54
11 <i>Apis mellifera</i>	217	222	201	187	158	157	157	156	164	151	—	0.54
12 <i>Fasciola hepatica</i>	283	280	277	278	269	269	270	267	268	270	273	—

Below diagonal = absolute distances. Above diagonal = mean distances.

(data not shown), regardless of the sequence analyzed. However, that tree was consistently incorrect.

Kumazawa and Nishida (17) recently completed a detailed study comparing mt COI, mt cytochrome *b* and all 22 mt tRNAs of six vertebrate taxa and a single echinoderm. They concluded that those particular mitochondrial protein genes are of limited usefulness in resolving relationships among animals whose divergence times are greater than 300 Myr with statistical reliability (i.e., bootstrap values >95% probability). Compared with the stem portions of mt t-RNA genes, mt protein genes display multiple sequence changes at the same site, resulting in a saturation of sequence differences among distantly related taxa and are relatively more susceptible to natural selection (17). It appears that the same could be said to be true for deep splits among invertebrates; the divergence of such groups is probably too ancient to be resolved by using mitochondrial protein or nucleic acid sequences. However, as greater numbers of lepidopteran sequences become available, mt COI information may prove to be useful at the genus-species population level.

References

- Moritz, C.; Dowling, T.E.; Brown, W.M. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annu. Rev. Ecol. Syst.* 18:269–292;1987.
- Capaldi, R.A. Structure and function of cytochrome *c* oxidase. *Annu. Rev. Biochem.* 59:569–596;1990.
- Saraste, M. Structural features of cytochrome oxidase. *Q. Rev. Biophys.* 23:331–366;1990.
- Peterson, A.M.; Barrillas-Muray, C.; Wells, M.A. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. *Insect Biochem. Mol. Biol.* 24:463–472;1994.
- Sambrook, J.; Fritsch, E.F.; Maniatis, T. *Molecular cloning: a laboratory manual*. New York: Cold Spring Harbor Laboratory Press; 1989.
- Clary, D.O.; Wolstenholme, D.R. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* 22:252–271;1985.
- Holton, T.A.; Graham, M.W. A simple and efficient method for direct cloning of PCR product using ddT-tailed vectors. *Nucleic Acids Res.* 19:1156;1991.
- Kocher, T.D.; Thomas, W.K.; Meyer, A.; Edwards, S.V.; Paabo, S.; Villablanca, F.X.; Wilson, A.C. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA* 86:6196–6200;1989.
- Needleman, S.B., 1975. *Protein sequence determination: A sourcebook of methods and techniques*. New York: Springer-Verlag, p. 393.
- Feng, D.-F.; Doolittle, R.F. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evol.* 25:351–360;1985.
- Swofford, D.L. *PAUP: phylogenetic analysis using parsimony, Version 3.0s*. Champaign, IL: Illinois Natural History Survey; 1991.
- Maddison, W.P.; Maddison, D.R. *MacClade, Version 3.0*. Sunderland, MA: Sinauer Press; 1992.
- Liu, H.; Beckenbach, A.T. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylo. Evol.* 1:41–52;1992.
- Wolstenholme, D.R. Genetic novelties in mitochondrial genomes of multicellular animals. *Curr. Opin. Genet. Dev.* 2:918–925;1992.
- Kyte, J.; Doolittle, R.F. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157:105–132; 1982.
- Holm, L.; Saraste, M.; Wikstrom, M. Structural models for redox centres in cytochrome oxidase. *EMBO J.* 6:2819–2823; 1987.
- Kumazawa, Y.; Nishida, M. Sequence evolution of mitochondrial tRNA genes and deep-branch animal phylogenetics. *J. Mol. Evol.* 37:380–398;1993.