Evolution of Nuclear- and Mitochondrial-Encoded Subunit Interaction in Cytochrome c Oxidase

Timothy R. Schmidt,* Wei Wu,*† Morris Goodman,*† and Lawrence I. Grossman*

*Center for Molecular Medicine and Genetics; and †Department of Anatomy and Cell Biology, Wayne State University

School of Medicine

Mitochondrial DNA (mtDNA)-encoded proteins function in eukaryotes as subunits of respiratory complexes that also contain nuclear DNA (nDNA)-encoded subunits. The importance of functional interactions between mtDNA- and nDNA-encoded proteins was previously demonstrated by testing the survivability of cybrid cells or individuals containing nDNA and mtDNA from different populations or species. This report focuses on the multisubunit respiratory complex cytochrome c oxidase (COX), made up of both mtDNA-encoded and nDNA-encoded subunits. A combination of evolutionary and crystallographic data is employed to determine whether rates of nonsynonymous substitutions have been higher, the same, or lower for residues in close proximity that are encoded by a different genome (nDNA or mtDNA). This determination is performed by simply taking the ratio, called the interaction ratio \( i \), of the nonsynonymous substitution rate of the close-contact residues to the nonsynonymous substitution rate of the noncontact residues. We assume that the close-contact residues (which are more likely to interact) are functionally important and that, therefore, amino acid replacements among these residues cannot escape the scrutiny of natural selection. \( i = 1 \) indicates that the close-contact residues have been under neither greater purifying selection nor greater positive selection than the noncontact residues as a specific consequence of their being encoded by separate genomes. \( i < 1 \) indicates that the close-contact residues have been under greater purifying selection than less positive selection than have the noncontact residues. Conversely, \( i > 1 \) indicates that the close-contact residues have been under less purifying but greater positive selection than have the noncontact residues. \( i < 1 \) may be referred to as a constraining interaction; i.e., the close-contact residues compared with the noncontact residues appear to be under greater structural-functional constraints. On the other hand, \( i > 1 \) may be referred to as an optimizing interaction; i.e., apparently many different amino acid replacements are required to optimize this subunit’s interaction with the other subunit. A major finding is that the nDNA-encoded residues in close physical proximity to mtDNA-encoded residues evolve more slowly than the other nuclear-encoded residues (and thus display a constraining interaction), whereas the mtDNA-encoded residues in close physical proximity to nDNA-encoded residues evolve more rapidly than the other mitochondrial-encoded residues (and thus display an optimizing interaction). A possible reason for this striking difference between the nuclear- and mitochondrial-encoded COX subunits in how their functional interaction evolves is discussed.

Introduction

The term “coevolution” was originally used to describe how the evolution of one species of organism was linked to that of another, e.g., the linkage between toxin-producing plants and herbivorous insects (Ehrlich and Raven 1964), between a toxic and nontoxic butterfly species (the latter’s coloration mimicking the former’s) (Turner 1977), and between host and parasite (Brooks 1979; Hafner and Nadler 1988). More recently, the term has been applied to the evolution of different members of gene families that encode proteins that interact coadaptively with one another (Hughes 1992; Fryxell 1996; Sitnikova and Su 1998). Protein-protein interactions have been studied most frequently at the functional level provide unique opportunities to study the evolution of protein-protein interactions and the effects of these interactions on the evolution of their respective genomes.
The presumed importance of functional interactions between mtDNA- and nDNA-encoded proteins has been experimentally confirmed. Kenyon and Moraes (1997) determined that cybrid cells constructed with human nDNA and mtDNAs from different primate species varied in ability to survive in media that required oxidative metabolism. In those experiments, mtDNAs that were evolutionarily close to humans (chimps and gorillas) were able to functionally complement the human nDNA gene products. In contrast, cybrids constructed with more evolutionarily distant mtDNAs did not survive. Conceptually similar experiments have been performed by repeatedly backcrossing copepod Tigriopus califonicus populations, thereby placing the maternally inherited mtDNA genome of one population with the paternal nDNA of another (reviewed in Burton, Rawson, and Edmonds 1999). COX activities of the resulting progeny were significantly lower than those of both native populations. One interpretation of these studies is that mtDNA-encoded proteins are less able to function with the nuclear counterparts of other populations or species because they disrupt protein interactions that have co-evolved over time.

These studies demonstrated the functional importance of mtDNA-nDNA-encoded protein interactions, but they did not explore the evolution of such interactions. In this report, a combination of evolutionary and crystallographic data from the COX holoenzyme was used to study interactions at the DNA level. Residues in close physical proximity to those of a subunit encoded by another genome are clearly functionally important. Thus, nonsynonymous mutations in codons encoding such close-contact residues are not likely to escape the scrutiny of natural selection, either in its positive form when selection for advantageous amino acid replacements spreads the replacements through a species lineage or, alternatively, in its purifying form when selection continues to favor these amino acid replacements after they have spread throughout the lineage. We found not only that the close-contact residues have different nonsynonymous substitution rates than the residues not involved in protein interactions, but also that the type of interaction differs for mtDNA- and nDNA-encoded residues.

Materials and Methods
Interaction Ratio

Interaction Ratio

We tested for statistical evidence, in the form of alteration of nonsynonymous substitution rates, for the functional interaction between nDNA-encoded and mtDNA-encoded amino acid residues. To accomplish this, we defined an interaction ratio, $i$, for when part of one protein (A) is in close physical proximity to part of another (B). Two interactive values can be calculated: one for protein A interacting with protein B ($i_{A,B}$), and one for protein B interacting with A ($i_{B,A}$). Specifically, an estimate of $i_{A,B}$ requires knowledge of the physical distances between the residues of protein A and protein B. The residues of protein A that are in close contact with residues of protein B are then removed from the protein A data set and placed into a new data set, which is the “contact” data set. The remaining residues of protein A form the “noncontact” data set. Using the sequence data that exist from different species for protein A, we calculate two rates of nonsynonymous substitutions, one for the close-contact residues and the other for the noncontact residues. The interaction ratio ($i_{A,B}$) is then simply the ratio of the contact nonsynonymous rate to the noncontact nonsynonymous rate. Similarly, the residues of protein B can be segregated to calculate an interaction ratio for protein B ($i_{B,A}$).

Three categories of interaction ratios are expected. An $i$ indistinguishable from 1 would result if the evolutionary rate of the close-contact residues were the same as the evolutionary rate of the noncontact residues, indicating that the close-contact residues were under neither greater purifying selection nor greater positive selection than the noncontact residues. $i < 1$ would result from a reduced evolutionary rate of the close-contact residues relative to the noncontact residues, indicating that greater purifying selection but less positive selection acted on the close-contact residues than on the noncontact residues. Finally, $i > 1$ would be due to a more rapid nonsynonymous substitution rate for the close-contact residues, indicating that they are under less purifying selection but under more positive selection than the noncontact residues. On the one hand, we interpret $i < 1$ as being due to the protein’s close-contact residues (those involved in protein interaction) being under greater structural-functional constraints than the protein’s noncontact residues (those less likely to be involved in protein interaction). We refer to such a state as a “constraining interaction.” On the other hand, we interpret $i > 1$ as being due to many different amino acid replacements among the close contact residues being required to optimize this protein’s interaction with the other protein. We refer to such a state as an “optimizing interaction.”

Since $i$ is a comparison of nonsynonymous substitutions for independently selected residues of a protein with the evolutionary rate of the remaining residues of the same protein, the residues have a common evolutionary history. Mutation rates, population size, phylogenetic history, and other evolutionary parameters are unlikely to differ within a protein. Consequently, for the purposes of our study, relative rates of evolution rather than absolute rates are used since for any protein under study, the period of time over which the nonsynonymous substitutions occurred is the same for the close-contact residues as for the noncontact residues.

Data Sets

Crystallographic data for COX genes from bovine heart mitochondria (Tsukihara et al. 1995, 1996) were first grouped by encoding genome. Nucleotide sequences encoding COX nDNA residues that were also in close proximity to COX mtDNA-encoded residues (fig. 1) were then segregated and placed in a separate data set from those not in close contact with mtDNA-encoded residues (fig. 2). Similarly, COX residues encoded by
Evolution of Subunit Interaction 565

Fig. 1.—Flow chart for the calculation of the interaction ratio of nDNA-encoded residues ($i_{np}$). In brief, an initial data set is divided into two data sets on the basis of physical distance between amino acids. Pairwise nonsynonymous sequence divergences were calculated for each of the two data sets. The ratio (contact/noncontact) was calculated from the sums of the nonsynonymous divergences and is the interaction ratio. Values of $i$ that are statistically different from 1 reflect that residues in close contact have evolved at a different rate than residues not in close contact. The same procedure was followed for the mtDNA-encoded residues.

mtDNA were segregated on the basis of proximity to nDNA-encoded residues. Physical distances between residues were calculated with a computer program provided by Dr. Philip D. Martin. Residues ≤4 Å apart, the nominal upper limit for weak interactions (Martin et al. 1997), were segregated.

Sequence Divergences and Interaction Ratios

Nonsynonymous sequence divergences for nDNA-encoded residues were calculated from pairwise comparisons between primates (humans), bovines (cows), and rodents (mice and rats) with the contact and noncontact data sets. Distances were computed both with the method of Jukes and Cantor (1969) using Molecular Evolutionary Genetics Analysis (MEGA, version 1.01; Kumar, Tamura, and Nei 1993) and with the maximum-likelihood method using Phylogenetic Analysis by Maximum Likelihood (PAML; Yang 2000). The mtDNA-encoded data sets (26 species) were analyzed with methods described below for sequences of human, cow, mouse, and rat. There were not sufficient evolutionary data to include COX5A and COX6C. COX8H was excluded from the analyses because it is absent in humans (Van Kuilenburg et al. 1988; Rizzuto et al. 1989). Nucleotide sequences of 26 mammalian taxa were analyzed for the mtDNA-encoded COX genes (COX1, COX2, and COX3). These mammalian taxa were Bos taurus, Balaenoptera physalus, Balaenoptera musculus, Equus caballus, Equus asinus, Rhinoceros unicornis, Ceratotherium simum, Homo sapiens, Pan troglodytes, Pan paniscus, Gorilla gorilla, Pongo pygmaeus, Hylobates lar, Felis catus, Halticurus grrypus, Phoca vitulina, Mus musculus, Rattus norvegicus, Erinaceus europaeus, Didelphis virginiana, Macroopus robustus, Ornithorhynchus anatinus, Oryctolagus cuniculus, Myoxus glis, Cavia porcellus, and Armadillo officinalis.

Results

Interaction Ratios (Pairwise Comparisons)

Results from two methods (the maximum-likelihood and Jukes-Cantor methods) of calculating nonsynonymous sequence divergence differed little, suggesting that either model is adequate for calculation of interaction ratios. Unless otherwise noted, results from the maximum-likelihood analysis are used below. The interaction ratio of nDNA-encoded COX residues in contact with mtDNA-encoded COX residues ($i_{nm}$) was 0.82 ± 0.18, which is not significantly less than 1 (table 1). However, the Jukes-Cantor method of calculating nonsynonymous sequence divergence gave a smaller standard error, which resulted in a ratio that was significantly

were influenced by the number of taxa analyzed, analyses were also performed with the mtDNA-encoded data sets on the same taxa utilized for nuclear-encoded data sets. A Z-test was performed to assess the significance of differences between the contact and noncontact nonsynonymous sequence divergences (and, therefore, the difference between the interaction ratio and 1).

A seven-taxon data set was constructed from the mtDNA-encoded data in order to calculate $i$ on a lineage-by-lineage basis using the codeml program of the PAML package. Tree topologies were constrained to be consistent with relationships derived from analyses of complete mitochondrial genomes (Arnason, Gullberg, and Janke 1997, 1998; Janke, Xu, and Arnason 1997) and morphological and other molecular data (Liu and Miyamoto 1999). In all cases, seven-taxon data sets were used to limit computation time.

DNA Sequences

Most nucleotide sequences of COX genes were acquired from GenBank and aligned with CLUSTAL W (alignments can be obtained from the website of L.I.G., http://cmmg.biosci.wayne.edu/lgross/lgross-home.html). COX crystallography data of Tsukihara et al. (1996) were acquired from the Brookhaven Protein Data Bank. Seven nDNA-encoded COX genes (COX4, COX5B, COX6AH, COX6B, COX7AH, COX7B, and COX7C) were analyzed with methods described below for sequences of human, cow, mouse, and rat. There were not sufficient evolutionary data to include COX5A and COX6C. COX8H was excluded from the analyses because it is absent in humans (Van Kuilenburg et al. 1988; Rizzuto et al. 1989). Nucleotide sequences of 26 mammalian taxa were analyzed for the mtDNA-encoded COX genes (COX1, COX2, and COX3). These mammalian taxa were Bos taurus, Balaenoptera physalus, Balaenoptera musculus, Equus caballus, Equus asinus, Rhinoceros unicornis, Ceratotherium simum, Homo sapiens, Pan troglodytes, Pan paniscus, Gorilla gorilla, Pongo pygmaeus, Hylobates lar, Felis catus, Halticurus grrypus, Phoca vitulina, Mus musculus, Rattus norvegicus, Erinaceus europaeus, Didelphis virginiana, Macroopus robustus, Ornithorhynchus anatinus, Oryctolagus cuniculus, Myoxus glis, Cavia porcellus, and Armadillo officinalis.
Fig. 2.—Stereo view of the mtDNA-encoded COX II (green) and nDNA-encoded COX IV (blue) subunits. Residues of COX II that are in close contact with COX IV are red; those of COX IV in close contact with COX II are yellow. Distances between atoms of residues were calculated with a domain interface program. Close contact between residues was defined as $<$4 Å. Although only COX II and COX IV are depicted here, identification of residues in close contact was performed for each of the nDNA- and mtDNA-encoded COX subunits included in the analyses.

less than 1. Therefore, the statistical evidence suggests that the interaction of nDNA-encoded residues with mtDNA-encoded residues reduces the nonsynonymous substitution rate of the nDNA (a constraining interaction), although the evidence for this is modest. In contrast with results from the nDNA-encoded subunits, analysis of the 26-taxon mtDNA-encoded COX subunits in contact with nDNA-encoded counterparts yielded an $i_{\text{mt}}$ of $2.11 \pm 0.14$, significantly greater than 1. Therefore, the interaction of mtDNA-encoded residues with nDNA-encoded residues seems to increase the nonsynonymous substitution rate, an observation suggesting an optimizing interaction. Analysis of the four-taxon mtDNA-encoded data (using the same taxa as in the nDNA analysis) gave an $i_{\text{mt}}$ value ($2.02 \pm 0.14$) similar to that given by analysis of the entire data set. This suggests that the results of this type of analysis are not overly sensitive to the number of taxa included.

Functional regions of proteins are typically constrained and typically have lower nonsynonymous substitution rates than those of other regions (Li, Wu, and Luo 1985; Nei and Gojobori 1986). Thus, if there were a higher proportion of functionally important residues in the noncontact region of the protein than in the contact region, this would cause $i$ to be greater than 1, independent of protein interactions. This possibility was tested by further dividing the mtDNA contact and noncontact residues into residues that had important known functions other than the contact function (e.g., heme binding, proton channels) and those that did not have such other known functions (Tsukihara et al. 1995, 1996; Yoshikawa et al. 1998) and calculating $i$ for both ($i_{\text{mt,n,F}}$ and...
Since the subdivision of the residues was based on incomplete information about functionality, we initially tested residues that were not in close contact with nDNA-encoded residues to see if the subdivision was reliable. The ratio of nonsynonymous substitutions for this division of functional/nonfunctional residues was 0.32 ± 0.11, suggesting that purifying selection had been acting more strongly on residues of known function and, consequently, that the grouping of residues was not arbitrary. The interaction ratio for residues of known function \( i_{\text{mtDNA}} \) (5.28 ± 0.37) was significantly larger than that for unknown function \( i_{\text{mtDNA}} \) (1.78 ± 0.15), showing that the optimizing interaction was more pronounced for residues of known function (table 1).

### Interaction Ratios (Phylogenetic)

An increased nonsynonymous substitution rate (or amino acid replacement rate) in primates has been demonstrated for a number of COX subunits and CYC (Baba et al. 1981; Evans and Scarpulla 1988; Adkins and Honeycutt 1994; Wu et al. 1997, 2000; Schmidt, Goodman, and Grossman 1999). To examine whether those results were related to the observation reported here, we determined whether the optimizing interaction between the mtDNA- and nDNA-encoded subunits was emphasized in the primate lineage. Phylogenetic trees were used to apportion the nonsynonymous substitutions for the contact and noncontact mtDNA data sets, and \( i_{\text{mtDNA}} \) was calculated for each lineage (fig. 3). Most lineages showed \( i > 1 \); \( i_{\text{mtDNA}} \) values of <1 were found only for short internal branches. Results of analyses using alternative tree topologies and species did not differ substantially in having an \( i_{\text{mtDNA}} \) of >1, although with some topologies the primate \( i_{\text{mtDNA}} \) was <1. We conclude that primates do not show a markedly different \( i_{\text{mtDNA}} \) than other lineages of mammals. Instead, optimizing interactions have a broad phylogenetic distribution and are present in most mammalian lineages examined. These results suggest that (1) the co-occurring rate accelerations observed for primate COX subunits are not concentrated in the residues that have a direct interaction, and (2) the optimizing interaction is not due to the increased nonsynonymous substitution rate in primates previously observed for several subunits of COX.

### Compensatory Changes

Comparison of amino acid sequences of contact residues among mammalian species identified potential examples of compensating changes. In humans, an E→Q mutation was observed at position 2 of the mature COX VIIa-H protein that eliminated a negative charge, whereas a K→M mutation occurred for an interacting residue (position 224 of COX III) that removed a positive charge. The balance of the overall charge in these mutations suggests that the changes are compensatory. Multiple amino acid replacements can also be found for several subunits of COX.
the same interacting region. For example, a K→Q change at position 61 of subunit Vb was found in the human. Another mutation (D→G) occurred for the interacting residue at position 115 of subunit II, and an additional mutation (E→G) was observed at position 114. It is reasonable to propose that if an amino acid replacement has functional consequences, then compensating change(s) are highly likely, and probably essential, in maintaining a sound interaction.

Discussion

We take it as self-evident that the interaction between nDNA- and mtDNA-encoded proteins in the respiratory complexes is functionally important. Since the two genomes can evolve independently, natural selection seems necessary to maintain their functional interaction. Recent studies have shown that, in fact, coevolution in subunit interaction has occurred. Mixing nuclei and mitochondria from organisms that have evolutionarily diverged by time (Kenyon and Moraes 1997) or isolation (Burton, Rawson, and Edmonds 1999) results in reduced function.

The prevailing (although untested) view, which is based on the generally conservative nature of protein evolution, is that constraining interactions are typical (e.g., see Griffiths 1998; Lockless and Ramanganth 1999). The calculated interaction ratios of <1 (a constraining interaction) for COX subunits encoded by nDNA are consistent with this view. In contrast, an optimizing interaction was detected for mtDNA-encoded residues of COX. These results suggest that the faster mtDNA mutation rate, which allows sampling of more residues in the interacting region, makes mtDNA the predominant partner in accommodating mutations important for subunit interaction.

It would be interesting to examine the evolution of protein interaction in other complexes of the respiratory chain that have a mixed genetic origin for a similar pattern. Moreover, the combined use of crystallographic information and evolutionary data is not limited to studies of mtDNA interactions with nDNA. For example, the methods employed in this study can also be utilized in analysis of the evolution of protein interaction within complexes from plant cells that are encoded by chloroplast DNA and nDNA. Basic requirements for an accurate calculation of z include (1) nucleotide sequences of taxa that are sufficiently divergent to allow for accurate calculation of nonsynonymous substitution rates or amino acid replacements, (2) a reasonable number of residues for each data set, and (3) crystallographic data of molecules that potentially interact. Probably the most important methodological limitation is that a large number of interacting residues are required to develop a statistically relevant sample. However, in the absence of a large sample of residues, important information can still be derived from the combined use of crystallographic and phylogenetic data for identifying residues that covary.

Acknowledgments

We thank Dr. P. D. Martin, Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, for providing a computer program (domain_interface) used in the analyses, and Dr. Derek E. Wildman for help with several analyses. We also thank Drs. O. J. Miller and E. Golenberg for their comments on earlier versions of this manuscript. This work was supported by grants from NSF (MCB-9816923), NIH (GM 48517), and the Center for Molecular Medicine and Genetics of Wayne State University. The analyzed data sets, along with a list of functional residues and other ancillary information, are available at http://cmmg.biosci.wayne.edu/lgross/lgross-home.

T.R.S. and W.W. both contributed equally to this work.

LITERATURE CITED


