

Thermodynamic Stability Explains the Differential Evolutionary Dynamics of Cytochrome b and COX I in Mammals

Juan Carlos Aledo · Héctor Valverde ·
Manuel Ruíz-Camacho

Received: 11 September 2011 / Accepted: 2 February 2012 / Published online: 24 February 2012
© Springer Science+Business Media, LLC 2012

Abstract By using a combination of evolutionary and structural data from 231 species, we have addressed the relationship between evolution and structural features of cytochrome b and COX I, two mtDNA-encoded proteins. The interior of cytochrome b, in contrast to that of COX I, exhibits a remarkable tolerance to changes. The higher evolvability of cytochrome b contrasts with the lower rate of synonymous substitutions of its gene when compared to that of COX I, suggesting that the latter is subjected to a stronger purifying selection. We present evidences that the stability effect of mutations ($\Delta\Delta G$) may be behind these differential behaviour.

Keywords COX I · Cytochrome b · Evolvability · mtDNA · Natural selection · Protein evolution

Introduction

In addition to their central role in the oxidative phosphorylation (OXPHOS), mitochondria are involved in many cellular processes such as growth, apoptosis and ageing (Aledo 2004; Aledo et al. 2011; Navarro and Boveris 2007). Not surprisingly, mitochondrial defects have been

associated with a number of diseases (Scharfe et al. 2009), which may be the result of spontaneous or inherited mutations in the mitochondrial genome (mtDNA) or in nuclear genes (nDNA) that code for mitochondrial components (DiMauro and Schon 2008; Gallardo et al. 2006). In mammals, mtDNA encodes only 13 proteins of the respiratory chain, while the bulk of mitochondrial proteins are encoded by nuclear genes. The evolution of mtDNA contrasts with that of nDNA. Indeed, mitochondrial and nuclear genomes differ in many ways, such as the total length, ploidy level, mode of inheritance, recombination rate, presence of introns, effective population size and repair mechanisms. The fact that mtDNA has been evolving much more rapidly than nDNA in higher animals is currently undisputed, although it came as a surprise (Brown et al. 1979). Given the importance of the mtDNA-encoded proteins in OXPHOS, their more rapid rates of change seemed to challenge the idea that the more important the function of a protein, the more slowly it undergoes evolutionary change in primary structure (Zhang and He 2005).

Identifying factors that determine protein evolution rate has attracted considerable attention. Recent evidence suggests that all the events influencing protein expression, such as transcription initiation, splicing and translation, need to be considered when explaining variation in the rate at which different proteins evolve (Pál et al. 2006; Warnecke et al. 2009). However, when addressing the evolutionary rate of different residues within a single protein, the attention should be focused on functional–structural aspects (Franzosa and Xia 2009). In this sense, the functional properties of a protein, including the interactions with other proteins and post-translational modifications, are all related to its surface properties. However, the ability of this protein to fold correctly and the thermodynamic stability of this fold, which ultimately assures the protein function, are greatly

Electronic supplementary material The online version of this article (doi:10.1007/s00239-012-9489-0) contains supplementary material, which is available to authorized users.

J. C. Aledo (✉) · H. Valverde
Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain
e-mail: caledo@uma.es

M. Ruíz-Camacho
Departamento de Estadística e Investigación Operativa, Facultad de Ciencias, Universidad de Málaga, 29071 Málaga, Spain

influenced by features of the protein interior (Eilers et al. 2000; Mirny and Shakhnovich 2001). Therefore, exposure to solvent is a structural property that has received particular attention as a potential determinant of protein evolution.

Studies carried out on yeast (Conant and Stadler 2009; Franzosa and Xia 2009; Lin et al. 2007) and bacterial proteins (Bustamante et al. 2000) support the view that residues buried in a protein's core are most likely to remain conserved during evolution compared to their solvent exposed counterparts. From these findings, one may be tempted to speculate that proteins with a small proportion of solvent exposed residues should evolve slowly. Although such reasoning has received support from some authors (Lin et al. 2007), others have reported the opposite observation, that is, proteins with fewer exposed residues evolve more rapidly (Bloom et al. 2006a). In a recent work, Franzosa and Xia (2009) suggested that increasing core size has little effect on evolutionary rate among solvent excluded residues while yields a more rapid relaxation of constraint for those residues exposed to the solvent.

Since the effect of structural features on protein evolution has been matter of debate as to its mechanism and significance, we found interesting to address these issues using mtDNA-encoded proteins, which often show a differential evolutionary pattern with respect to nuclear-encoded proteins (Schmidt et al. 2001; Welch et al. 2008). In the current work, we have explored the evolutionary dynamics of the interior and surface of cytochrome b and COX I, two mtDNA-encoded proteins commonly used in phylogenetic studies.

Materials and Methods

Data Sources and Molecular Modelling

A collection of 231 mammalian mitochondrial genomes (Fig. 1) was obtained from the National Center for Biotechnology Information (NCBI) genome database (www.ncbi.nlm.nih.gov). These mammalian species encompass 27 orders: Afrosoricida ($n = 3$), Artiodactyla ($n = 25$), Carnivora ($n = 56$), Cetacea ($n = 24$), Chiroptera ($n = 10$), Cingulata ($n = 1$), Dasyuromorphia ($n = 5$), Dermoptera ($n = 1$), Didelphimorphia ($n = 4$), Diprodontia ($n = 11$), Erinaceomorpha ($n = 4$), Hyracoidea ($n = 2$), Lagomorpha ($n = 5$), Macroscelidea ($n = 2$), Monotremata ($n = 3$), Paramelomorphia ($n = 4$), Paucituberculata ($n = 2$), Perissodactyla ($n = 5$), Pholidota ($n = 1$), Pilosa ($n = 3$), Primates ($n = 32$), Proboscidea ($n = 4$), Rodentia ($n = 13$), Scandentia ($n = 1$), Sirenia ($n = 2$), Soricomorpha ($n = 7$) and Tubulidentata ($n = 1$). Multiple sequence alignments of orthologous proteins were performed using ClustaX 2.0.9. The gap opening and extension penalties were 15 and 6.66,

respectively. The delay divergent sequences option was set to 30%. A value of 0.5 was chosen for the transition weight parameter. Sequence identities were higher than 73 and 90% for cytochrome b and COX I, respectively. Three-dimensional models structures for cytochrome b and COX I were generated by alignment with the experimental crystal structures of corresponding bovine sequences (Protein Data Bank, PDB, 1be3 chain C and 2occ chain A, respectively). Structural calculations were performed on the Swiss-Model workspace (Bordoli et al. 2009). Owing to the difficulties of obtaining reliable structural models for all the 231 mammalian sequences, the analyses described below were done using sets containing 221 and 189 PDB files for cytochrome b and COX I, respectively (Fig. 1).

Determining Exposed and Buried Positions

Solvent accessible surface areas (ASA) were computed using the SurfRace program (Tsodikov et al. 2002). The accessibility of a given amino acid residue in a protein was calculated as the ratio of its ASA in the native protein structure to that it would have in an unfolded and extended polypeptide chain ($\Phi = -120^\circ$, $\Psi = 140^\circ$), with the side-chain conformations corresponding to the one most frequently observed in proteins (Miller et al. 1987). Amino acid residues exhibiting accessibilities below 5% were defined as buried residues. Then, following a multiple sequence alignment of orthologous proteins, the number of instances (species) a given position appeared as buried was computed. This position was considered as a buried position when in most of the species (>50%) the residue found showed an accessibility below the threshold of 5%.

Shannon's Entropy Determinations

For the mitochondrial protein being analyzed, multiple alignments of orthologous sequences were obtained, which allowed us to compute variability at position j of the alignment as the corresponding Shannon's entropy:

$$H(j)_c = - \sum_{i=1}^c p_i(j) \log_c p_i(j) \quad (1)$$

where $p_i(j)$ is the frequency of residue from class i in position j . Since different amino acid classification schemes can be contemplated, c stands for the number of different classes. More concretely, we used six classes of residues, $c = 6$, to reflect physico-chemical properties of amino acids and their natural pattern of substitution (Mirny and Shakhnovich 2001; Thompson and Goldstein 1996). In addition, for each position the Shannon's entropy was computed without grouping amino acids, that is, for $c = 20$.



Fig. 1 Mammalian species used in the study. A collection of 231 mammalian mitochondrial genomes was obtained from the National Center for Biotechnology Information. For each species, the amino acid sequences of cytochrome b and COX I were concatenated and subjected to alignments to construct a tree using the program Promlk from the PHYLIP package. The name of each species is accompanied

by a bidimensional vector. The first coordinate of the vector is 1 if a reliable structural model for cytochrome b could be obtained for that species, and 0 in the opposite case. Similarly, the second coordinate is 1 or 0, depending on whether or not a COX I structural model could be obtained, respectively

A low entropy value for a given position is interpreted as a lowly variable position. The extreme case being an invariant position, which would yield a null entropy. In the opposite extreme, an entropy value of 1 for a given position, is interpreted as a highly variable position.

Determining Unconstrained, Conservative and Radical Positions

The comparison between $H_6(j)$ and $H_{20}(j)$ can provide some insights into the forces behind the evolution of

residues at position j . For instance, positions yielding low values for both entropy measures, H_6 and H_{20} , can be considered as constrained positions. In contrast, those positions exhibiting high values for both entropy functions are expected to be less critical, allowing a relaxation of the purifying selection. These positions are designated as *unconstrained*. On the other hand, a high value for H_{20} along with a low H_6 value inform us that at that position, the properties of the residues rather than the amino acids themselves, are critical and subjected to selection; thus, we labelled this positions as *conservative*. More interesting are those positions exhibiting a low H_{20} value accompanied by high H_6 entropy. In this case, we have few changes at that positions (low H_{20}) but these changes are radical because they are taking place between amino acids belonging to groups with very different properties (high H_6), which may reflect adaptive changes driven by positive selection. These positions are referred to as *radical*. Thus, for each protein we assembled three sets of positions designed as *Unconstrained*, *Conservative* and *Radical*, according to the following criteria:

$$\text{Unconstrained} = (j \in J : H_{20}(j) \geq \text{UQ}_{20} \text{ and } H_6(j) \geq \text{UQ}_6)$$

$$\text{Conservative} = (j \in J : H_{20}(j) \geq \text{UQ}_{20} \text{ and } H_6(j) \leq \text{LQ}_6)$$

$$\text{Radical} = (j \in J : H_{20}(j) \leq \text{LQ}_{20} \text{ and } H_6(j) \geq \text{UQ}_6)$$

where J is the set containing either the positions of the 262 variable sites of cytochrome b or the 230 variable positions of COX I. LQ_c and UQ_c are the lower and upper quartiles of Shannon's entropy distribution, respectively. To calculate these quartiles, invariant positions, $H_{20}(j) = 0$, were previously removed. In this way, for cytochrome b we obtained the following values: $\text{LQ}_{20} = 0.0079$ and $\text{UQ}_{20} = 0.0501$; $\text{LQ}_6 = 0.0049$ and $\text{UQ}_6 = 0.0650$. For COX I, the quartiles took the following values: $\text{LQ}_{20} = 0.008$, $\text{UQ}_{20} = 0.0376$, $\text{LQ}_6 = 0.0040$ and $\text{UQ}_6 = 0.0222$.

mtDNA Substitutions Rates

Whole-tree estimation can be unreliable if nuisance parameters, such as base composition, vary across groups. This problem is particularly relevant to mitochondrial genes, since uncorrected nucleotide bias in mtDNA can mimic the effect of positive selection (Albu et al. 2008). For these reasons, we only calculated substitution rates for closely related pairs of species (see Online Resource 1). To assist in the assemblage of a data set formed by pairs of close species, we generated maximum likelihood phylogenetic subtrees using the PHYLIP package. Each subtree accounted for the species belonging to the same mammalian order. Only those species directly connected to the same internal node were

considered as a suitable pair. A phylogeny reliably reconstructed from data unrelated to the primary structure of cytochrome b and COX I was preferable to avoid any semblance of circularity. Thus, the complete sequences of the two mitochondrial ribosomal RNA genes were concatenated for each species, and used to generate the phylogenetic reconstructions that assisted in the assemblage of an initial data set formed by 54 pairs of close species. We then computed the mean number of nucleotide differences per site by pairwise comparison using the Nei–Gojobori method (Nei and Gojobori 1986) and the Jukes-Cantor correction to account for multiple substitutions at the same site. In this way, for each gene (cytochrome b and COX I) we obtained 54 points of the $d_S \times d_N$ plane, where the d_S variable represents the mean number of synonymous differences per synonymous site while d_N is the mean number of nonsynonymous differences per nonsynonymous site.

Next, each DNA orthologous sequence was divided into two sets on the basis of the accessibility of the amino acid residue being encoded for the considered codon. In other words, mtDNA nucleotide sequences encoding cytochrome b and COX I buried residues were segregated and placed in a separate data set from those encoding for exposed residues. Afterwards, for each data set d_S and d_N were computed as explained above.

Thermodynamic Stability Changes

The thermodynamic stability changes ($\Delta\Delta G$) of mutations were computed using the protein design tool FoldX version 3.0 (Guerois et al. 2002; Schymkowitz et al. 2005; Tokuriki et al. 2007). FoldX uses a full atomic description of the structure of the protein, to provide a quantitative estimation of the importance of the interactions contributing to the stability of this protein. The different energy terms taken into account, which have been described in detail somewhere else (Guerois et al. 2002), have been weighted using empirical data obtained from protein engineering experiments.

3D structures for both cytochrome b and COX I were subjected to an optimization procedure using the repair function of FoldX. Then for each protein in each species, an alanine scan was carried out. That is, every single residue was replaced by alanine one by one, and the resulting $\Delta\Delta G$ was computed and recorded as a function of the residue position in the primary protein structure. This procedure provided two matrices of 221×379 and 189×514 , containing $\Delta\Delta G$ values for cytochrome b and COX I, respectively (raw data can be provided under request).

Computation and Statistical Analyses

Random distributions were generated using Perl scripts. Probability calculations were assisted by Wolfram

Mathematica 8.0. All other statistic analyses were done with SPSS 15.0.

Results

Invariant Positions are Accumulated in the Interior of COX I but May be Randomly Distributed Through the Whole Cytochrome b Protein

To address whether or not the surface and interior of mitochondrial proteins evolve differentially, we started sorting out each residue position as buried or surface site according to the criteria exposed in the methodological section. Once this segregation was accomplished, we addressed whether evolutionary rates varied between these different residue sets. As a first approach, we focused our interest on those residues that have remained invariant during the diversification of mammals. More concretely, we tested the following null hypothesis: invariant residues are randomly distributed between the interior and surface of the considered proteins. To this end, we defined the random variable X as the number of invariant residues that are buried in the protein interior. Beside the current value of X (designed by lowercase x), the proportion of invariant residues ($p_x = \text{number of invariant residues}/\text{total number of residues}$) and the number of buried residues, n , were computed for each protein (Table 1). In this way, under the null hypothesis conditions, we can assume that X follows a binomial distribution, $X \sim \text{Bin}(n, p_x)$, which allows us to calculate the probability of finding by chance a number of invariant buried residues equal or higher to that observed for each protein, $P[X \geq x]$. Although such probabilities were relatively low for both proteins cytochrome b and COX I, only in the latter case the null hypothesis could be rejected at a significance level of 1%.

COX I, but Not Cytochrome b, Protein Interior Shows a Low Shannon’s Entropy that Departs from Random Expectations

Hitherto we have analyzed the departure from random distribution of invariant positions, considering a position as

Table 1 Abundances of exposed, buried and invariant residues in cytochrome b and COX I

Protein	P_{exp}	p_x	n	x	$P[X \geq x]$
Cytochrome b	0.8482	0.314	58	22	0.175
COX I	0.6718	0.556	170	110	0.009

P_{exp} is the proportion of exposed residues, p_x is the proportion of invariant residues, n is the number of buried residues, x is the number of invariant buried residues, $P[X \geq x]$ is the probability of finding by chance a number of invariant buried residues equal to or higher than that observed

invariant when the same amino acid is found, without exception, in all the analyzed species. We next used an information theoretic formalism to study the evolutionary conservation of the protein interiors.

After computing H_6 and H_{20} for each position, the averaged entropies for the buried residues were worked out. The mean values for $H_6(\text{buried})$ and $H_{20}(\text{buried})$ were 0.023 and 0.019, respectively, in the case of cytochrome b. These values went down to $H_6(\text{buried}) = 0.006$ and $H_{20}(\text{buried}) = 0.005$, when COX I was considered. In this way, the conservation of the buried protein core can be statistically compared with the conservation of all residues in the protein. To this end, we tested the following null hypothesis: buried residues are not more conserved than the whole protein sequence. To contrast this hypothesis, we compared the above means with the distribution of mean entropy values of the same number of residues randomly chosen from the whole protein, $H_c(\text{random})$. These distributions were obtained by taking 10^5 random sets of 58 residues for cytochrome b and 170 residues for COX I. Then, the fraction of instances with $H_c(\text{random}) < H_c(\text{buried})$ gives the probability of observing by chance a mean entropy value lower than that computed for the buried residues. In other words, it gives us the type I error rate. We were unable to reject the null hypothesis for cytochrome b at a confidence level of $\alpha = 10\%$. In contrast, for COX I the hypothesis was rejected at confidence levels as low as $\alpha = 0.34$ and 0.00%, for H_6 and H_{20} , respectively. Thus, these results are in line with those previously shown on invariant positions.

Since COX I buried positions are preferentially enriched with invariant residues, it might be that the low $H_c(\text{buried})$ we have reported above were due to the contribution of invariant amino acids from the protein core. In other words, we wanted to address whether buried positions were still more constrained than positions at the protein surface, once those invariant positions were excluded from the study. For this purpose, random sets of 36 and 60 residues from cytochrome b and COX I, respectively, were used to generate random distributions as explained above. Figure 2 shows the results of such analyses. As it can be deduced from this figure, the variable buried residues from COX I are much more constrained than the rest of COX I variable positions ($\alpha \leq 5\%$). However, variable buried positions from cytochrome b failed again to exhibit a statistically lower variability with respect to their exposed counterpart ($\alpha > 18\%$).

Unconstrained Positions are Randomly Distributed in Cytochrome b but are Selectively Excluded from the Interior of COX I

For each protein we assembled three sets of positions designed as *Unconstrained*, *Conservative* and *Radical*,

according to the criteria specified in “[Determining Unconstrained, Conservative and Radical Positions](#)” (Fig. 3). Afterwards, we assessed whether or not these position categories were preferentially located in or excluded from the protein interior. To this end, we computed the frequencies of unconstrained ($p_u = 0.068$ and 0.054), conservative ($p_c = 0.018$ and 0.037) and radical position ($p_r = 0.034$ and 0.031) for cytochrome b and COX I, respectively. These frequencies were used as proxies of the probabilities of a given position being an unconstrained, conservative or radical position, respectively. Also, we defined three random variables (U, C, R) as the number of unconstrained buried, conservative buried and radical buried positions. If these three types of positions are randomly distributed between the protein interior and the protein surface, one would expect that each of these random variables would follow a binomial distribution: $U \sim \text{Bi}(n, p_u)$, $C \sim \text{Bi}(n, p_c)$, $R \sim \text{Bi}(n, p_r)$; where n is the number of buried residues of the protein under consideration. In this way, after computing the current values that these random variables take for each protein (represented by lower-case letters: u, c and r , respectively), we were in conditions to calculate the probability of finding by chance, in the protein interior, a number equal or higher to that observed. That is, $P[U \geq u]$, $P[C \geq c]$ and $P[R \geq r]$. Since all these probabilities were much higher than 0.01 , for both cytochrome b and COX I (Table 2), we concluded that the

interior of these mitochondrial proteins is not particularly enriched in any of these type of positions. We next tested the possibility of any of these position categories being selectively excluded from the protein core. For this purpose, we calculated, under the null hypothesis conditions (random distribution), the probability of finding by chance, in the protein interior, a number of unconstrained, conservative or radical positions lower or equal to that observed for each protein. That is, $P[U \leq u]$, $P[C \leq c]$ and $P[R \leq r]$. When these probabilities are below 0.01 we can reject the null hypothesis of random distribution. As it can be observed in Table 2, only COX I gave a $P[U \leq u]$ value below the threshold of 0.01 , which suggests that unconstrained positions are selectively excluded from the COX I core, but not from the cytochrome b interior.

The COX I Gene is Subjected to Both a Higher Mutation Pressure and a Stronger Purifying Selection than the Cytochrome b Gene

Under the assumption that changes at silent sites are mainly neutral (Kimura 1977), the comparison between the rates of synonymous substitutions per site within the cytochrome b with those for COX I, suggested a surprisingly higher mutational pressure for COX I with respect to cytochrome b (Fig. 4a). In contrast, when the rates of nonsynonymous substitutions per site were the subject of comparisons, COX I

Fig. 2 Shannon’s entropy of cytochrome b and COX I buried positions. After removing invariant positions, the mean H_6 and H_{20} entropies for buried residues were calculated and are shown in the upper-left corner of each plot. These mean values were compared with the distribution of mean entropy values of the same number of residues randomly chosen from the protein (surface + interior)

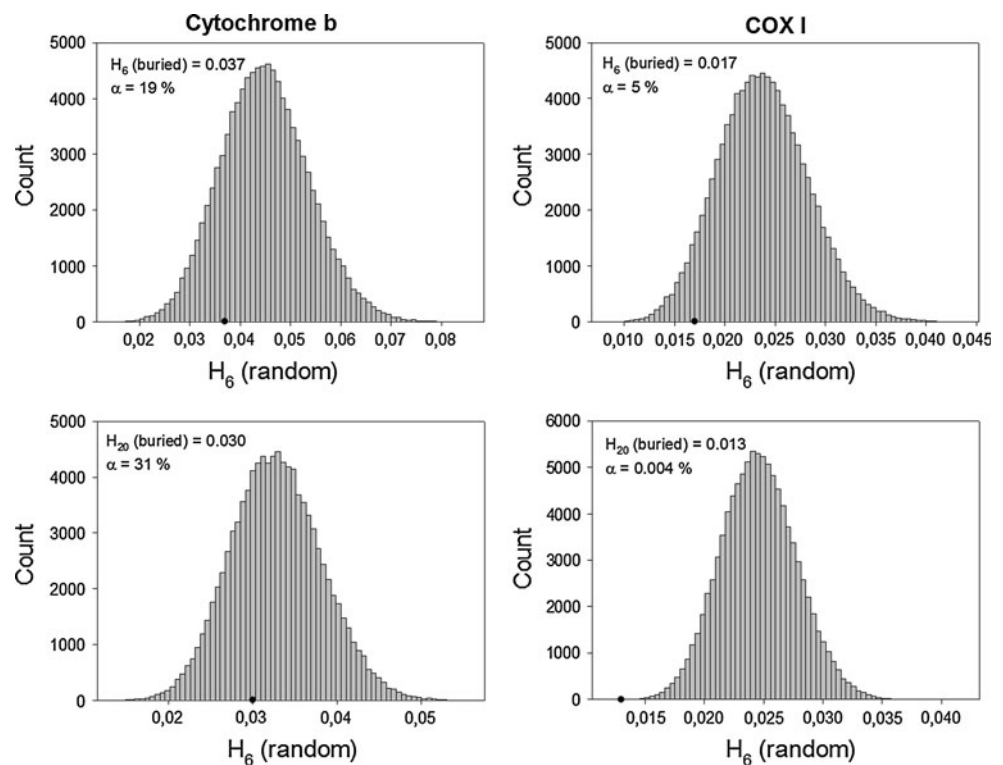
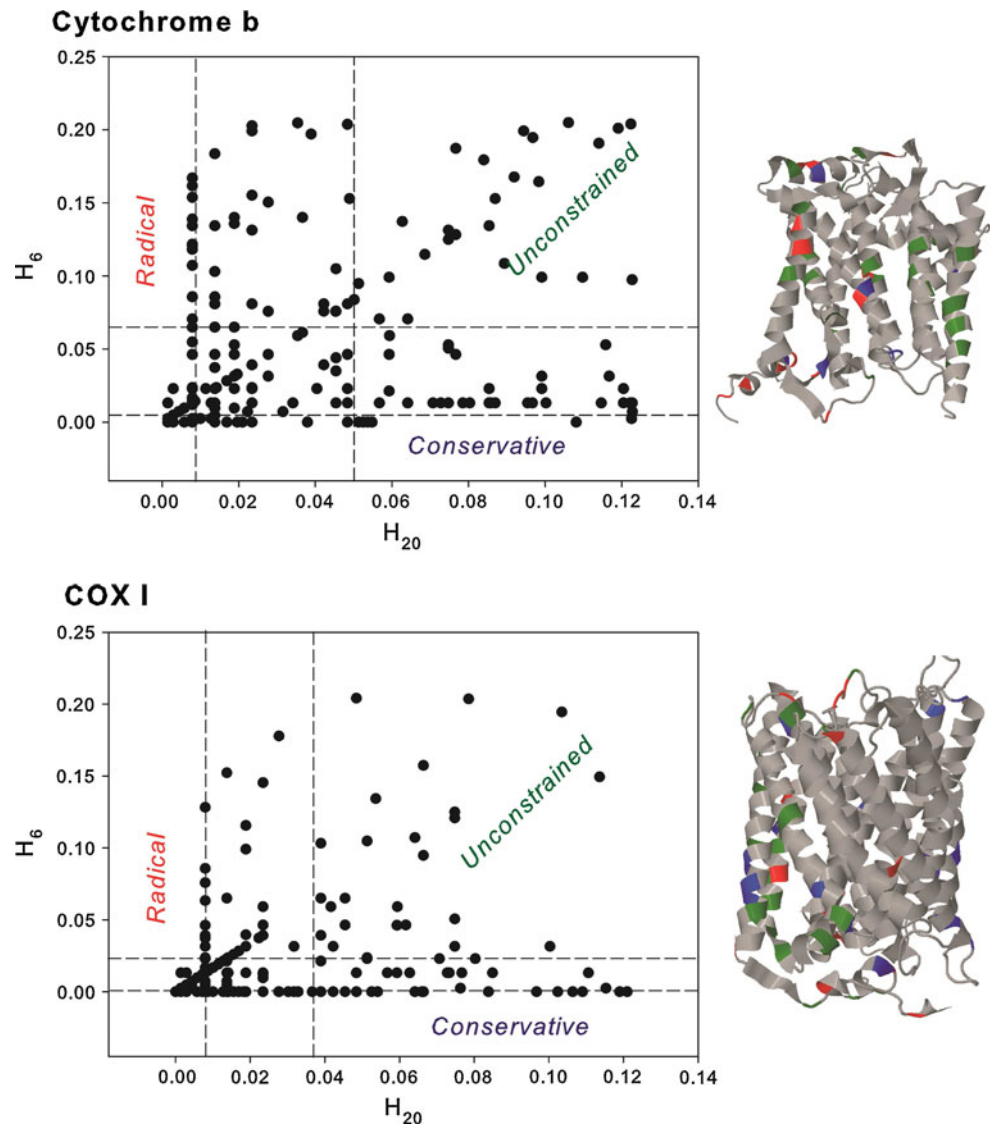


Fig. 3 Identification and location of unconstrained, conservative and radical sites. H_6 was plotted against H_{20} and horizontal dashed lines were drawn to indicate the lower and upper quartiles of the H_6 distribution. Similarly, vertical dashed lines indicate the lower and upper quartiles of the H_{20} distribution. Residues belonging to each of the three categories (see the text for details) were identified on the folded structure (either cytochrome b or COX I) according to the following colour code: unconstrained (green), conservative (blue) and radical (red)



showed d_N values significantly lower than those calculated for cytochrome b, indicating that the COX I protein is much more constrained by purifying selection (Fig. 4b).

We next addressed the question of whether the rates of mutation and selection varied between those codons coding for buried amino acids and those coding for exposed residues. As shown in Fig. 4c, d, we failed to observe significant differences in d_S between the interiors and surfaces. In contrast, the rates of nonsynonymous substitutions were remarkably lower among nucleotides encoding for buried residues, regardless the gene being considered (Fig. 4e, f). Nevertheless, d_N for buried COX I was still much lower than d_N for buried cytochrome b (Fig. 5). Of note is that the rate of nonsynonymous substitutions affecting cytochrome b buried residues was much higher than that for solvent-exposed COX I residues (Fig. 5c).

Thermodynamic Stability Can Explain the Differential Behaviour of COX I and Cytochrome b

Often a single amino acid substitution can dramatically alter the stability of a protein. Not surprisingly, protein stability has been pointed out as a determinant of evolvability (Bloom et al. 2006b; Tokuriki and Tawfik 2009). According to this view, the stability effects of mutations may underlie the differential evolutionary dynamics of cytochrome b and COX I described above. In other words, we hypothesized that mutations affecting buried residues in COX I are more destabilizing than mutations taking place in the interior of cytochrome b. To address this working hypothesis, we assessed the thermodynamic stability effect ($\Delta\Delta G$) of substituting each single buried residue into alanine. To this end, the thermodynamic stability changes of

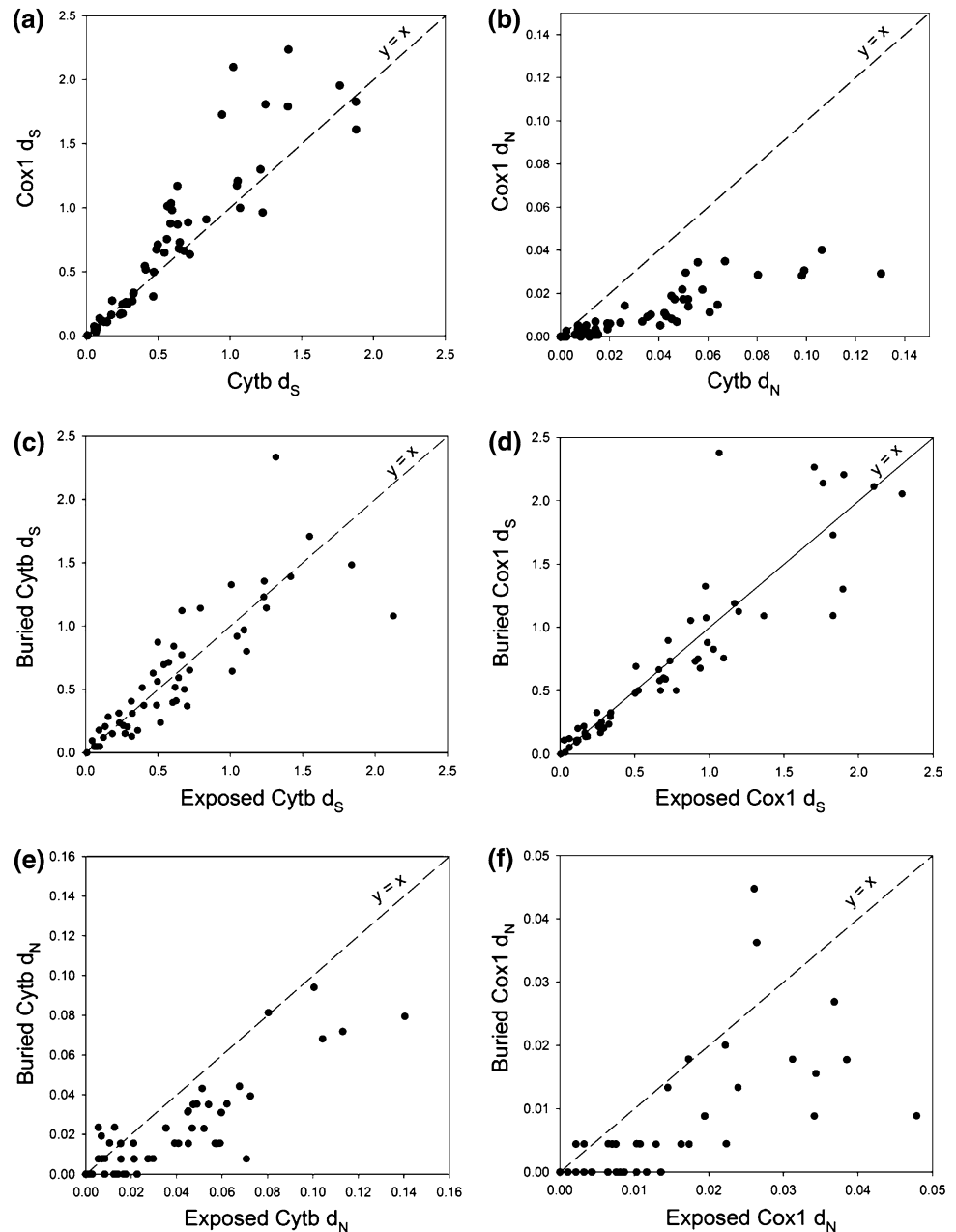
Table 2 Probabilities, according to the null hypothesis, of finding in the protein interior a number of unconstrained (u), conservative (c), or radical (r) positions lower or higher to those observed for each protein

Protein	n	$P[U \leq u]$	$P[C \leq c]$	$P[R \leq r]$	$P[U \geq u]$	$P[C \geq c]$	$P[R \geq r]$
Cyt b	58	0.438	0.913	0.135	0.764	0.281	1.000
COX 1	170	0.008**	0.05	0.101	0.991	0.983	0.969

n is the number of buried residues of the protein under consideration

The null hypothesis assumes that the numbers of buried unconstrained, conservative and radical positions follow binomial distributions. See the text for details. ** Confidence level lower than 0.01

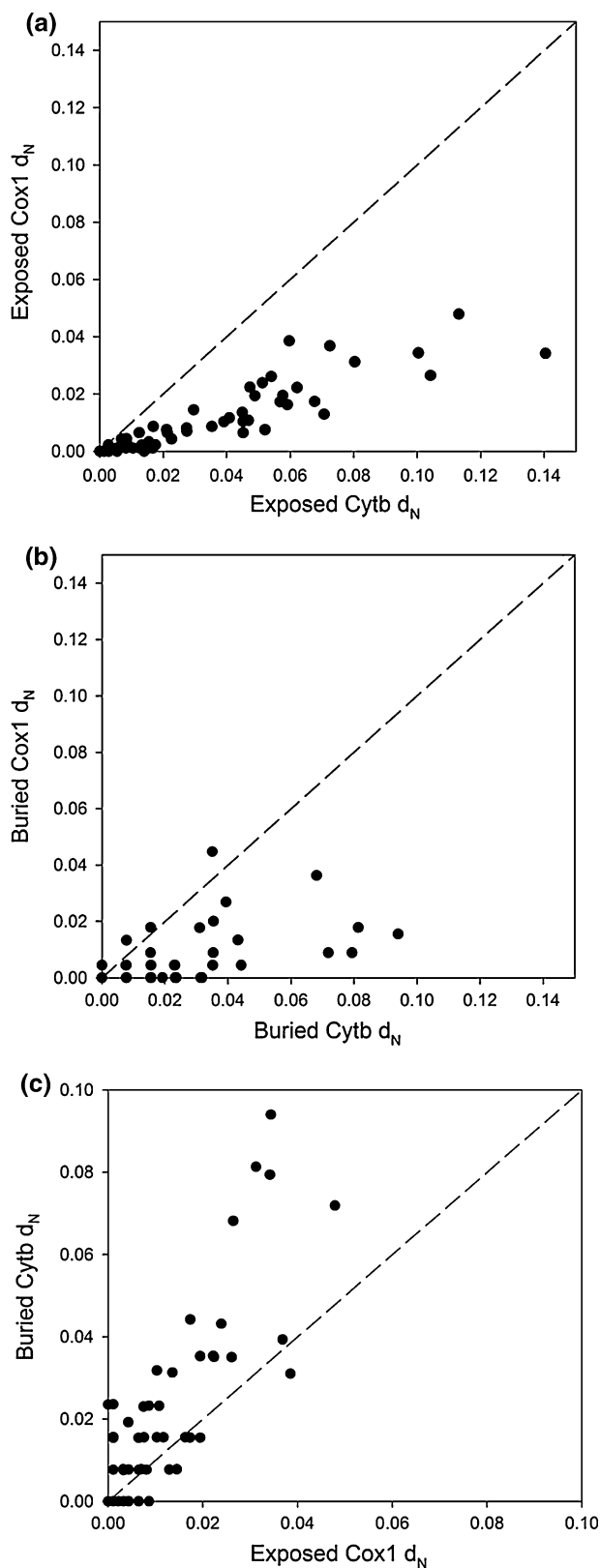
Fig. 4 Intergenic and intragenic substitution rate comparisons. For each pair of proximal species, the substitution rate per site was computed using different sets of residues from cytochrome b and COX I (see text for details), and plotted for comparative purposes



mutations were computed using the force-field FoldX (Guerois et al. 2002; Schymkowitz et al. 2005).

First, we formulated the null hypothesis that single point mutations at buried positions are no more destabilizing

than mutations at randomly chosen positions from the whole protein. To test this hypothesis we computed the mean free energy change upon mutation of buried residues to alanine, and compared it with the distribution of means



of the same number of residues randomly chosen in the same protein. This distribution was obtained by choosing 10^6 random sets of either 58 (for cytochrome b) or 170 (for

◀ **Fig. 5** Comparison of nonsynonymous substitution rates among different residues sets. The rates of nonsynonymous substitutions per site affecting the exposed regions of cytochrome b and COX I were compared (a). Similarly, the substitutions rates within the interior of these proteins were directly compared (b). The interior of cytochrome b was also compared to the surface of COX I (c)

COX I) residues and computing the mean $\Delta\Delta G$ upon single point mutation of each residue from the set. As expected, mutations affecting buried residues from COX I were strongly destabilizing. The null hypothesis could be rejected at a significance level as low as $\alpha < 10^{-6}$. Changes affecting the cytochrome b interior were also significantly ($\alpha = 0.0002$) more destabilizing than changes on the surface. However, when the $\Delta\Delta G$ distributions for cytochrome b and COX I were compared (Fig. 6), a clear-cut conclusion emerged: cytochrome b seemed to be much more robust to mutations from the thermodynamic stability point of view. Thus, nonsynonymous substitutions affecting buried residues may be easily tolerated in the case of cytochrome b, but much more unlikely for COX I, as also suggested in Fig. 5.

Both, cytochrome b and COX I are single polypeptides that form part of large multisubunit complexes. Therefore, many of the residues being classified as exposed in the single polypeptide chain, may be involved in functional and structural interactions with other subunits. Since these interactions would generate evolutionary constraints and change the stability status of new mutations, it seemed relevant to examine this issue. To this end, and using the bovine quaternary structures of complexes III and IV, we determined the $\Delta\Delta G$ after carrying out alanine scans of cytochrome b and COX I while being part of their respective complexes. Figure 6 summarizes the results of such analyses. While the $\Delta\Delta G$ distribution for cytochrome b as part of the cytochrome bc1 complex was similar to that of the single chain (with respect to mean and variance) the distribution of COX I in the complex IV showed a remarkable increased mean and reduced variance (compare Fig. 6a, b). More importantly, the comparison of $\Delta\Delta G$ for cytochrome b and COX I as part of their respective complexes, strongly suggest that, from the thermodynamic point of view, cytochrome b is much more robust to mutations than COX I.

Discussion

Understanding variability in substitution rates between different proteins and different regions of proteins is of considerable interest to molecular evolutionists, as well as to biotechnologists engaged in protein engineering. Studies carried out in bacteria (Bustamante et al. 2000) and more recently in yeast (Bloom et al. 2006a; Conant and Stadler

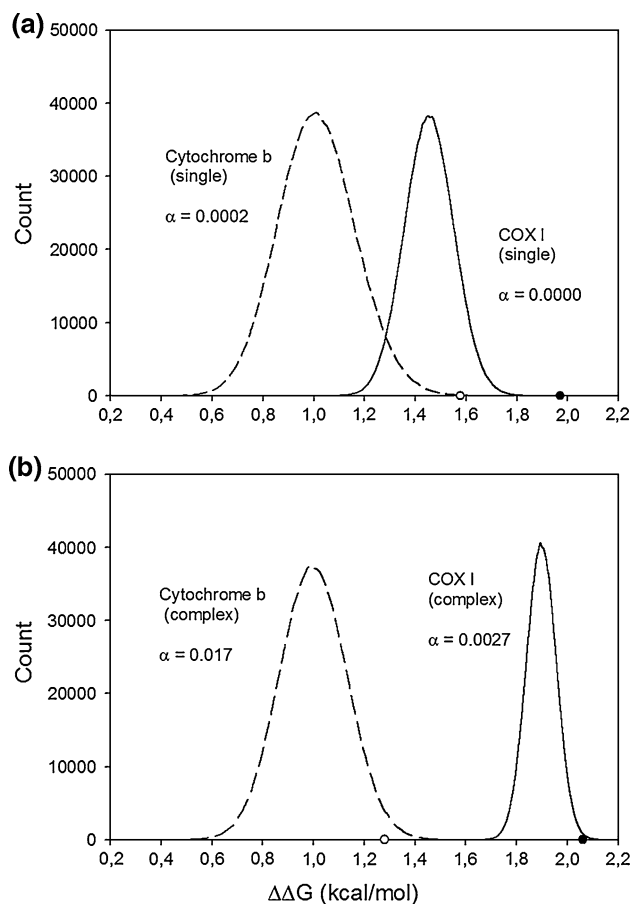


Fig. 6 Destabilizing effect of random mutations in cytochrome b and COX I. The mean free energy changes upon mutation of buried residues to alanine ($\Delta\Delta G$) were computed and indicated as open (cytochrome b) and filled circles (COX I) on the abscises axes. These mean values were compared with the distribution of mean $\Delta\Delta G$ values of the same number of residues randomly chosen from the whole protein (see text for details). Direct comparisons of the distributions of mean $\Delta\Delta G$ in cytochrome b and COX I as single polypeptides (a) or as part of their respective complexes (b), are shown

2009; Franzosa and Xia 2009; Lin et al. 2007), have pointed to residue exposure as an important structural determinant of molecular evolution. In spite of a growing literature supporting the view that solvent exposed residues evolve faster than buried residues, this principle could not be taken for granted when mtDNA-encoded proteins are under consideration, because mitochondrial proteins often evolve under different selective constraints to those of nDNA-encoded proteins. This point is well illustrated by the observation that, while nDNA-encoded residues in the interface of protein complexes are more conserved, their mtDNA-encoded counterparts evolve even faster than other mtDNA-encoded residues (Schmidt et al. 2001). Furthermore, mtDNA encodes for membrane-spanning proteins, which seem to evolve differently than other proteins do (Conant et al. 2007; Popot and Engelman 2000), raising doubts about the applicability of the evolutionary rules that

govern soluble proteins (Conant and Stadler 2009). However, our results show that, for mtDNA-encoded proteins, the surfaces evolve faster than their corresponding interiors (Fig. 4e, f), providing evidence of the generality of this evolutionary rule across genomes and protein types.

Although these two regions also evolve differently in mitochondrial proteins, these differences are not so similar across proteins as it seems to be the case for yeast globular proteins (Conant 2009; Conant and Stadler 2009). In this respect, the interior of cytochrome b, in sharp contrast to that of COX I, shows a remarkable tolerance to changes as indicated by several lines of evidence. First, we failed to observe departures from random expectations in the distribution of invariant residues between the surface and the interior of cytochrome b (Table 1). Second, the mean Shannon's entropy for cytochrome b buried residues is not significantly lower than the mean value for a random set of residues (Fig. 2). Third, while unconstrained residues seem to be depleted from the COX I interior, they are randomly distributed between interior and surface of cytochrome b (Table 2). Fourth, although amino acid changes taking place in the interior of cytochrome b are indeed more constrained than those happening on its surface, the fact to be emphasized here is that the nonsynonymous substitution rates for the interior of cytochrome b were significantly higher than those computed for the surface of COX I (Fig. 5c). Finally, it should be noted that the thermodynamic stability changes of mutations affecting buried residues in cytochrome b are comparable to, or even lower than, those of exposed residues in COX I (Fig. 6).

While there is a consensus in the literature that exposed sites evolve faster than buried sites, the debate on whether the fraction of buried sites and d_N should correlate positively remains open. In other words, there is a wide agreement that solvent accessibility has a strong effect on the conservation of individual residues, but whether this behaviour scales to the level of whole proteins, is openly discussed. Since buried sites are more conserved, we may expect that proteins with a larger fraction of buried sites should evolve slower. On the other hand, it has been argued that although buried residues are generally more conserved, increasing the fraction of buried residues leads to an overall increase in the evolutionary rate of all residues in the protein (Bloom et al. 2006a). According to these authors, this is mainly because the higher number of buried residues yields a higher contact density that in turn contributes to an increase stability of the whole protein. Eventually, this additional stability allows a strong relaxation of constraint on the solved exposed sites. In few words, the reduction in the fraction of exposed residues is more than compensated for by the increased variability of exposed residues in proteins with high contact density. Although it may be the case for certain types of proteins, our results do not support

the generality of such arguments. Clearly, cytochrome b exhibits a much higher proportion of exposed residues than COX I, which is paralleled by a much higher rate of nonsynonymous substitutions. Therefore, the current analysis of mtDNA-encoded proteins lends support to the view of a positive correlation between the proportion of exposed residues and the rate of nonsynonymous substitutions (Lin et al. 2007).

It is widely acknowledged that mtDNA mutation rate can vary from gene to gene (Bielawski and Gold 1996). In fact, it has been suggested that the length of time genes remain in the single-stranded state during mtDNA replication (D_{ssH}), may be an important factor affecting the rate of mutation (Reyes et al. 1998). Since COX I and cytochrome b exhibit the shortest and longest D_{ssH} values, respectively, the superior rate of synonymous substitutions observed for COX I is a somewhat unexpected finding (Reyes et al. 1998). Nevertheless, this result could be explained by other mechanisms such as codon bias or time spent in the single-strand state during transcription (Faith and Pollock 2003).

Residues in the interior of the folded mitochondrial proteins and residues exposed on their surfaces, must experience different selection pressures as suggested by the observation that while the rates of synonymous substitutions are the same (Fig. 4c, d), there are great differences in rates of nonsynonymous changes (Fig. 4e, f). Nevertheless, the relevant question is, what are the biological factors behind this differential pattern of selection? In this sense, a main contribution of the current study is that we are providing compelling and quantitative evidence of the role played by thermodynamic stability. This evidence strongly suggests that the thermodynamic stability effect of mutations may be a key factor driving the evolutionary dynamics of proteins. Thus, we have observed that $\Delta\Delta G$ for mutations affecting buried residues are, in general, more destabilizing than those affecting exposed sites, which is in line with the lower d_N values computed for the interior with respect to the surface, regardless the protein being considered. It is worth noting that, despite the fact that residues from the interior of cytochrome b evolve under stronger constraints than residues exposed on its surface, buried residues from cytochrome b and exposed residues from COX I seem to evolve under comparable selective constraints (Fig. 5c). Interestingly, this observation fits well with the finding reported herein that $\Delta\Delta G$ for mutations affecting cytochrome b buried residues are comparable in magnitude to those for mutations affecting residues from the COX I surface. In few words, even the more destabilizing changes in cytochrome b interior are comparable in importance to the less destabilizing changes in COX I, which may help to explain why the core of cytochrome b seems to be much more tolerant to changes than the interior of COX I.

It has been suggested that the stability effect of mutations may show a universal energy distribution (Tokuriki et al. 2007). Therefore, the differential evolvability exhibited by proteins is believed to reside in the absolute thermodynamic stability of the native structure (Bloom et al. 2006b). In other words, random mutations indiscriminately decrease the stability of proteins. However, more stable proteins can tolerate better this decrease in stability, which in turn allows them to evolve faster. Somehow, our results challenge this simple view. We have shown that the magnitude of the stability effect of mutations may strongly depend on the tertiary structure of the protein under consideration. Furthermore, although we do not have data on the absolute stability of cytochrome b and COX I, if we accept the contact density as a proxy for absolute thermodynamic stability, we would expect COX I to be more stable than cytochrome b. However, the last evolves faster than the former. Therefore, we suggest that the stability effect of mutations strongly depends on the native structure and may be a key determinant of protein evolvability.

Conclusions

Although there exists a clear relationship between solvent exposure and the destabilizing effect of mutations (for a given protein, changes in the interior are more perturbing than changes at the surface), the absolute magnitude of the stability effect of mutations strongly depends on the native structure being considered. We suggest that $\Delta\Delta G$ rather than solvent accessibility may be the key determinant of the differential evolutionary behaviour of cytochrome b and COX I in mammals.

Acknowledgments We thank Alicia Esteban del Valle and Miguel Ángel Medina for their helpful comments on the manuscript. The authors are also grateful to two anonymous referees who have helped to improve the original manuscript. We also gratefully acknowledge the support of Grant CGL2010-18124 from the Ministerio de Ciencia e Innovación, Spain.

References

- Albu M, Min XJ, Hickey D, Golding B (2008) Uncorrected nucleotide bias in mtDNA can mimic the effects of positive Darwinian selection. *Mol Biol Evol* 25:2521–2524
- Aledo JC (2004) Glutamine breakdown in rapidly dividing cells: waste or investment? *Bioessays* 26:778–785
- Aledo JC, Li Y, de Magalhães JP, Ruíz-Camacho M, Pérez-Claros JA (2011) Mitochondrially encoded methionine is inversely related to longevity in mammals. *Aging Cell* 10:198–207
- Bielawski JP, Gold JR (1996) Unequal synonymous substitution rates within and between two protein-coding mitochondrial genes. *Mol Biol Evol* 13:889–892

- Bloom JD, Drummond DA, Arnold FH, Wilke CO (2006a) Structural determinants of the rate of protein evolution in yeast. *Mol Biol Evol* 23:1751–1761
- Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006b) Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103:5869–5874
- Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 4:1–13
- Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 76:1967–1971
- Bustamante CD, Townsend JP, Hartl DL (2000) Solvent accessibility and purifying selection within proteins of *Escherichia coli* and *Salmonella enterica*. *Mol Biol Evol* 17:301–308
- Conant GC (2009) Neutral evolution on mammalian protein surfaces. *Trends Genet* 25:377–381
- Conant GC, Stadler PF (2009) Solvent exposure imparts similar selective pressures across a range of yeast proteins. *Mol Biol Evol* 26:1155–1161
- Conant GC, Wagner GP, Stadler PF (2007) Modeling amino acid substitution patterns in orthologous and paralogous genes. *Mol Phylogenet Evol* 42:298–307
- DiMauro S, Schon EA (2008) Mitochondrial disorders in the nervous system. *Annu Rev Neurosci* 31:91–123
- Eilers M, Shekar SC, Shieh T, Smith SO, Fleming PJ (2000) Internal packing of helical membrane proteins. *Proc Natl Acad Sci USA* 97:5796–5801
- Faith JJ, Pollock DD (2003) Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. *Genetics* 165:735–745
- Franzosa EA, Xia Y (2009) Structural determinants of protein evolution are context-sensitive at the residue level. *Mol Biol Evol* 26:2387–2395
- Gallardo ME, Moreno-Loshuertos R, López C, Casqueiro M, Silva J, Bonilla F et al (2006) m.6267G>A: a recurrent mutation in the human mitochondrial DNA that reduces cytochrome c oxidase activity and is associated with tumors. *Hum Mutat* 27:575–582
- Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320:369–387
- Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* 267:275–276
- Lin Y, Hsu W, Hwang J, Li W (2007) Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol Biol Evol* 24:1005–1011
- Miller S, Janin J, Lesk AM, Chothia C (1987) Interior and surface of monomeric proteins. *J Mol Biol* 196:641–656
- Mirny L, Shakhnovich E (2001) Evolutionary conservation of the folding nucleus. *J Mol Biol* 308:123–129
- Navarro A, Boveris A (2007) The mitochondrial energy transduction system and the aging process. *Am J Physiol* 292:C670–C686
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426
- Pál C, Papp B, Lercher MJ (2006) An integrated view of protein evolution. *Nat Rev Genet* 7:337–348
- Popot JL, Engelman D (2000) Helical membrane protein folding, stability, and evolution. *Annu Rev Biochem* 69:881–922
- Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol Biol Evol* 15:957–966
- Scharfe C, Lu HH, Neuenburg JK, Allen EA, Li G, Klopstock T et al (2009) Mapping gene associations in human mitochondria using clinical disease phenotypes. *PLoS Comput Biol* 5(4):e1000374
- Schmidt TR, Wu W, Goodman M, Grossman LI (2001) Evolution of nuclear- and mitochondrial-encoded subunit interaction in cytochrome c oxidase. *Mol Biol Evol* 18:563–569
- Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L (2005) The FoldX web server: an online force field. *Nucleic Acids Res* 33:W382–W388
- Thompson MJ, Goldstein RA (1996) Constructing amino acid residue substitution classes maximally indicative of local protein structure. *Proteins* 25:28–37
- Tokuriki N, Tawfik DS (2009) Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 19:596–604
- Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. *J Mol Biol* 369:1318–1332
- Tsodikov OV, Record MT, Sergeev YV (2002) Novel computer program for fast exact calculation of accessible and molecular surface areas and average surface curvature. *J Comput Chem* 23:600–609
- Warnecke T, Weber CC, Hurst LD (2009) Why there is more to protein evolution than protein function: splicing, nucleosomes and dual-coding sequence. *Biochem Soc Trans* 37:756–761
- Welch JJ, Bininda-Emonds ORP, Bromham L (2008) Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol Biol* 8:53
- Zhang J, He X (2005) Significant impact of protein dispensability on the instantaneous rate of protein evolution. *Mol Biol Evol* 22:1147–1155