2007-07-16

Protein Coevolution and Coadaptation in the Vertebrate bc1 Complex

Kimberly Kay Baer
Brigham Young University - Provo

Protein Coevolution and Coadaptation in the Vertebrate Cytochrome bc_1 Complex

by

Kimberly Kay Baer

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Integrative Biology

Brigham Young University

August 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Kimberly Kay Baer

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	David A. McClellan, Chair
Date	Dixon J. Woodbury
Date	David M. Belnap

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Kimberly Kay Baer in its final form and have found that (1) its format, citations and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date	David A. McClellan Chair, Graduate Committee

Accepted for the Department

Date	Keith A. Crandall Department Chair

Accepted for the College

Date	Rodney Brown Dean, College of Life Sciences

ABSTRACT

Protein Coevolution and Coadaptation in the Vertebrate Cytochrome bc_1 Complex

Kimberly Kay Baer

Department of Integrative Biology

Master of Science

The cytochrome bc_1 complex of the mitochondrial electron transport chain accomplishes the enzymatic reaction known as the modified Q-cycle. In the Q-cycle the bc_1 complex transports protons from the matrix to the intermembrane space of the mitochondria, creating the proton gradient used to make ATP. The energy to move these protons is obtained by shuttling electrons from the coenzyme ubiquinol (QH_2) to coenzyme ubiquinone (Q) and the mobile cytochrome *c*. This well studied complex is ideal for examining molecular

adaptation because it consists of ten different subunits, it functions as a dimer, and it includes at least five different active sites.

The program TreeSAAP was used to characterize molecular adaptation in the bc_1 complex and identify specific amino acid sites that experienced positive destabilizing (radical) selection. Using this information and three-dimensional structures of the protein complex, selection was characterized in terms of coevolution and coadaptation. Coevolution is described as reciprocal local biochemical shifts based on phylogenetic location and results in overall maintenance. Coadaptation, on the other hand, is more dynamic and is described as coordinated local biochemical shifts based on phylogenetic location which results in overall adaptation.

In this study both coevolution and coadaptation were identified in various locations on the protein complex near the active sites. Sites in the pore region of cyt c_1 were shown to exhibit coevolution, in other words maintenance, of many biochemical properties, whereas sites on helix H of cyt *b*, which flanks the active sites Q_o and Q_i, were shown to exhibit coadaptation, in other words coordinated shifts in the specific properties equilibrium constant and solvent accessible reduction ratio. Also, different domains of the protein exhibited significant shifts in drastically different amino acid properties: the protein imbedded in the membrane demonstrated shifts in mainly functional properties, while the part of the complex in the intermembrane space demonstrated shifts in conformational, structural, and energetic properties.

ACKNOWLEDGEMENTS

I thank my advisor, David McClellan for his guidance, direction, knowledge and patience throughout this process. He has contributed not only funding, but more importantly his time providing the resources and opportunities necessary to shape me into the scientist he knew I would become. His countless hours spent brainstorming, formatting, editing and revising this thesis are greatly appreciated. I also thank his wife Linda for being so supportive and giving of her husband's time.

I thank my committee members Dixon Woodbury and David Belnap for their suggestions and time spent editing and reviewing this thesis.

I acknowledge the students in my lab for computer support, suggestions and discussions: Rick Smith, Brady Curtis, Wes Beckstead, Mark Ebbert, Srikar Chamala and Jason Primavera. I also acknowledge the graduate students and post-docs who contributed by discussions, suggestions and support: Nicole Lewis-Rogers, Carissa Jones, Dan Hardy, Scott Peat, Martha Yoke, and Stephen Cameron.

I thank the experts in the field I was able to correspond with for sharing their knowledge: Dr. Mark Rowe from the Department of Nutrition, Dietetics & Food Science at BYU and Dr. Edward A. Berry from the Physical Biosciences Division of the Lawrence Berkeley National Laboratory.

I acknowledge the Department of Integrative Biology and Keith Crandall for funding. I also thank those professors: Duane Jeffery, Duke Rogers, and David McClellan who provided the opportunity for me to serve as a teaching assistant. I also acknowledge the department secretaries and staff for their assistance through this process.

I thank my parents Don and Sally for their support, both emotionally and financially, during my graduate education. I also thank my siblings Jenny, Tina and Todd for their support.

I would especially like to thank my significant other Gavin Svenson, who was able to provide the motivation and support I needed to finish this degree while he was also finishing a PhD. Also, the countless hours spent discussing theory, editing papers, and teaching me phylogenetics and illustration programs are greatly appreciated. Thanks so much for your love and understanding.

Table of Contents

Thesis Manuscript: *Formatted for Submission to the Journal of Molecular Biology and Evolution*

Protein Coevolution and Coadaptation in the Vertebrate Cytochrome bc_1 Complex

By Kimberly K. Baer

Protein Coevolution and Coadaptation in the Vertebrate Cytochrome bc_1 Complex

Kimberly Baer

Abstract

The cytochrome bc_1 complex of the mitochondrial electron transport chain transports protons from the matrix to the intermembrane space, creating the proton gradient used to make ATP—at the same time shuttling electrons from ubiquinol to ubiquinone and the mobile cytochrome *c*. This well studied complex is ideal for examining molecular adaptation because it consists of ten different subunits and at least five different active sites. Using the program TreeSAAP molecular adaptation was characterized in this complex and specific amino acid sites were identified that experienced positive destabilizing (radical) selection. Coevolution is characterized in two different ways in this paper, first coevolution is described as reciprocal changes in a microenvironment that result in the maintenance of the overall function of that region, while coadaptation is the more dynamic form of coevolution specifically coordinated changes in a microenvironment that result in a shift of a property in one direction. In this study both types—coevolution and coadaptation were identified in various locations on the protein complex. Most adaptive sites are on the protein complex surface. Some regions of the protein have clusters of radical shifts in a variety of physicochemical properties, including the head of the ISP subunit, the pore of cyt c_1, and also helix H of cyt *b* which flanks the active sites Q_o and Q_i.

Introduction

The electron transport chain (ETC) system in the mitochondria is comprised of five protein complexes, which work to create a proton gradient between the matrix and the intermembrane space to drive ATP synthesis (Fig. 1). ATP is the fuel of all cells, and subtle changes in the proteins of the ETC may have profound effects on metabolism (Chamala et al. 2006), obesity (Chamala et al. 2006; Okura et al. 2003), aging (Trifunovic et al. 2004), insulin resistance (Petersen et al. 2003) and overall health (Wibrand et al. 2001). The ETC functions as a result of a series of electron carriers and proton pumps that work to shuttle protons across the membrane. This proton gradient fuels ATP synthesis via ATP synthase (Fig. 1, complex 5). The third complex, formally called ubiquinol: cytochrome *c* oxidoreductase (E.C. 1.10.2.2) is also known as the cytochrome bc_1 complex (named for cytochrome *b* and cytochrome c_1, two major components of the complex) and helps to maintain the proton gradient. For every Q-cycle the complex oxidizes 2 ubiquinols, reduces cytochrome *c*, and uses the free energy to transport 4 H+ across the membrane from the matrix to the intermembrane space helping to create the proton motive force (Crofts 2004). The bc_1 complex of vertebrates is made up of 11 different subunits and functions as a dimer (Fig. 2). Of these subunits, three have known redox capabilities that function to transfer electrons from electron carriers ubiquinol and cytochrome *c*, and protons from ubiquinol to the intermembrane space. Cytochrome *b* contains two hemes, heme b_H, which is associated with the binding site for ubiquinone (Q) at the Q_i site, and heme b_L which is associated with the binding of ubiquinol (QH_2) at the Q_o site. The Rieske iron-sulfur protein contains an iron sulfur center and undergoes a conformational change to transport electrons from the Q_o site, to

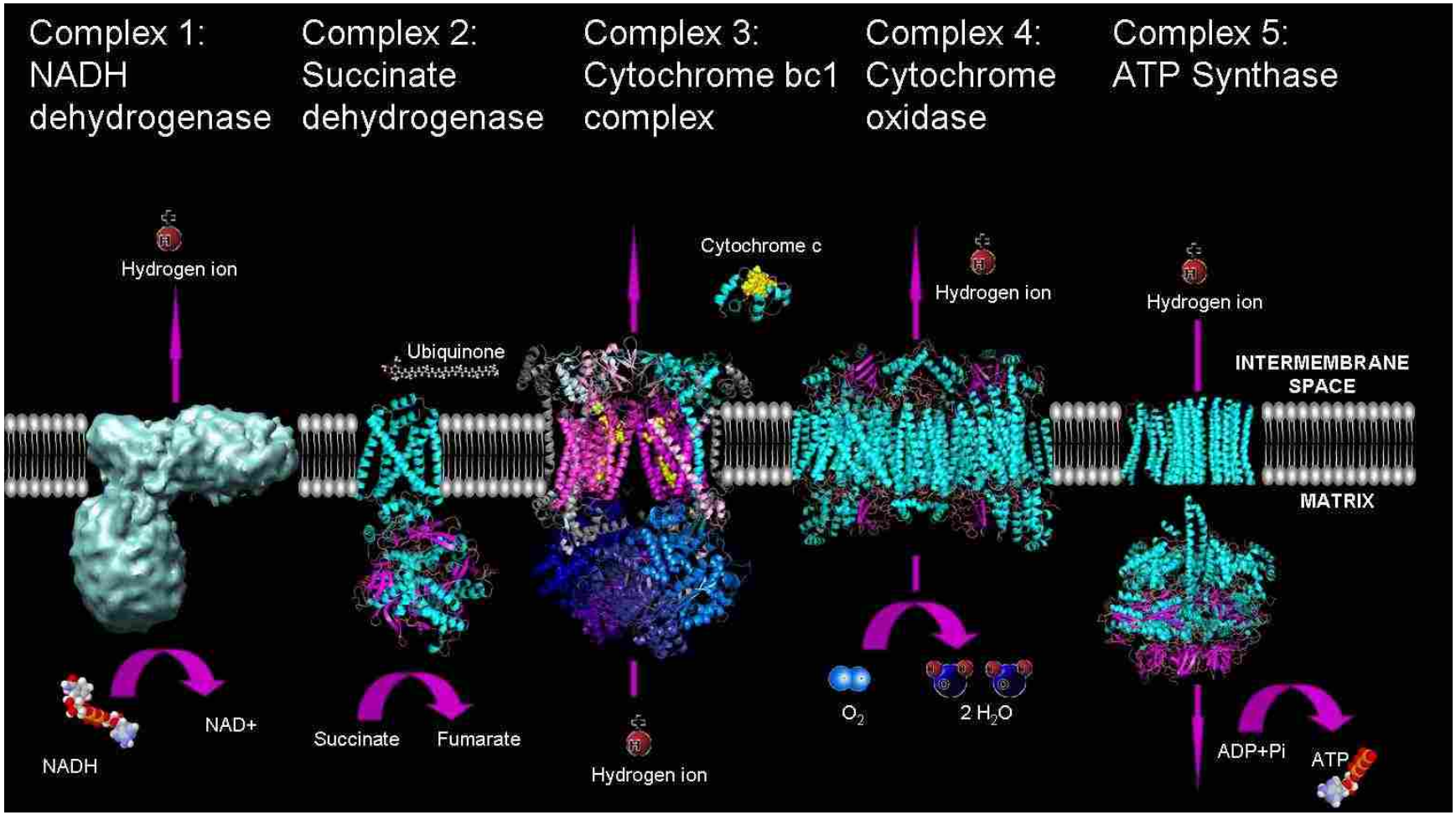


Figure 1: The electron transport chain is shown here. Complex 1 (which has not been crystallized yet) is from Grigorieff (1998). Complex 2, PDB ID: 1YQ3; Complex 3, PDB ID: 2A06; Complex 4, PDB ID: 1OCC; Complex 5, PDB ID: 1BMF and 1C17; and Cytochrome *c*, PDB ID: 1I6E. Each subunit works to increase the positive charge in the intermembrane space so complex 5 can synthesize ATP. Complex 3, the cytochrome bc_1 complex, has its major subunits colored.

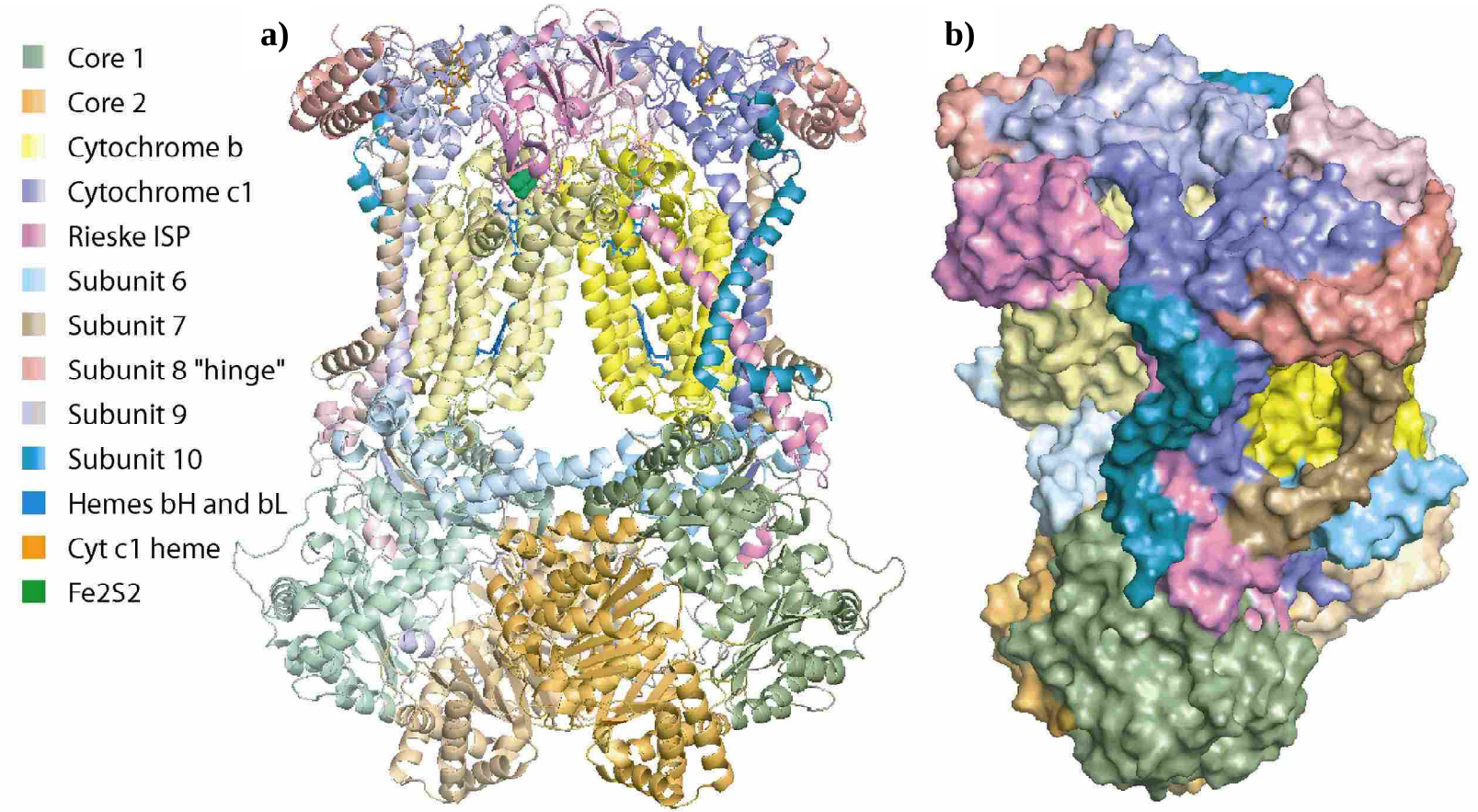


Figure 2: Cartoon and surface view of bc_1 complex. Both monomers are shown to display the functional dimer. (a) Cartoon view of bc_1 complex, angle is looking through the membrane. The key shows colors consistent in all figures from this point forward. Hemes and Fe_2S_2 are also shown here. (b) Surface view of the bc_1 complex. Angle is slightly tilted to show the top of the structure in the intermembrane space where cyt c_1 and ISP come in contact.

the heme of cyt c_1 with its iron sulfur center (Crofts et al. 1999b; Esser et al. 2006; Izrailev et al. 1999; Kim et al. 1998; Zhang et al. 1998). Cytochrome c_1 also contains a heme, which functions to transfer the electrons from ubiquinol to cytochrome *c*. This happens in what is called the modified Q cycle. For a more detailed description of the Q cycle, see Crofts et al. 1999a.

The bc_1 complex is a good model to study protein-protein interactions because its structure and function are well known and well researched. Many researchers have published three-dimensional crystal structures of the bc_1 complex, each showing something slightly different and informative, some with inhibitors bound to the active sites, others with the ubiquinone and ubiquinol bound, and even one with cytochrome *c* bound (Gao et al. 2003; Zhang et al 1998; Iwata et al. 1998; Lange and Hunte 2002; Huang et al. 2005). In addition these researchers have used these structures to propose a mechanism for the Q cycle, and detail the path of each electron and proton involved, and how they are transferred (Izrailev et al. 1999; Crofts et al. 1999; Crofts 2004). The three-dimensional structure is especially important when studying coevolution and molecular evolution of a protein complex because when the protein is folded, regions come into contact that are very distant on the linear protein, or even on different subunits altogether. This gives us a more in depth perspective of where adaptation is occurring in this protein complex.

Cytochrome bc_1 is an appropriate molecular system for this study of coevolution and coadaptation, which proposes a new method to look at protein-protein interactions and molecular adaptation using physicochemical properties. This protein complex is an ideal model because it has multiple subunits, complex active sites, and has moving parts.

Multiple subunits allow an investigation into signs of coevolution, both within regions where subunits come in contact with each other, and also in the multi-subunit active sites of the overall complex. Because conformational changes occur, as the complex carries out its function, there may be a more dynamic type of coevolution and adaptation at multiple points of articulation and within the secondary structures that allow movement.

Additionally, study of radical changes in amino acid properties can help to show what regions of the protein have historically been affected by selection. This information can help us assess how the protein complex is holistically evolving, and in what regions major physicochemical adaptation is taking place. This deep understanding of protein complexes may become useful to the pharmaceutical industry in designing novel drugs.

Knowledge of adaptation in three dimensions is prerequisite to understanding the array of mitochondrial disorders. There are many know human disorders resulting from changes in the proteins of the bc_1 complex. Changes to the subunit cytochrome *b* include Leber hereditary optic neuropathy (Riordan-Eva and Harding 1995), exercise intolerance (Keightley et al. 2000), myoglobinuria (Keightley et al. 2000), myopathy (Andreu et al. 2000), histiocytoid cardiomyopathy (Andreu et al. 2000), colorectal cancer (Polyak et al. 1998), obesity (Okura et al. 2003), and increased hydrogen peroxide production leading to parkinsonism and mitochondrial encephalopathy with lactic acidosis and stroke-like episodes (Rana et al. 2000). There may also be multi-system manifestations (e.g. deafness, mental retardation, retinitis pigmentosa, cataract, growth retardation, epilepsy) (Wibrand et al. 2001). Cytochrome c_1 deficiencies include myopathy, encephalomyopathy, and cardiomyopathy (Hoffman et al 1993).

An example of the complexity of cytochrome bc_1 disorders is warranted. A 4-bp deletion (nucleotides 338-341) was documented in exon 4 of the UQCRB gene (subunit 6), predicting a change in the last 7 amino acids and the addition of a stretch of 14 amino acids at the C-terminal end of the protein. The patient had hypoglycemia and liver dysfunction (Haut et al. 2003). Both parents were healthy and were found to be heterozygous for the deletion.

Not only do mutations in the subunits of the bc_1 complex lead to disorders, but mutations in the chaperone proteins for this complex can confer effects equally as detrimental. For example the bcs1 protein is a constituent of the inner mitochondrial membrane and is required for the expression of functional ubiquinol-cytochrome-*c* reductase (bc_1) complexes. Found within the inner mitochondrial membrane, BCS1L is presumed to facilitate insertion of Rieske Fe/S protein into precursors to complex III during assembly of the respiratory chain (Cruciat et al. 1999). Changes in this gene can lead to neonatal proximal tubulopathy, hepatic involvement, encephalopathy (De Lonlay et al. 2001), GRACILE syndrome (growth retardation, amino aciduria, cholestasis, iron overload, lactic acidosis, and early death) (Visappa et al. 1998) and others.

In addition to problems resulting from mutations, mitochondrial complexes are susceptible to, and aid in cancer growth. It has been shown that alterations to the oxidative phosphorylation function in tumor cell mitochondria play a causative role in cancerous growth, (Warburg, 1956) largely as a result of the role they play in apoptosis and other aspects of tumor biology. The mitochondrial genome is particularly susceptible to mutations because of the high level of reactive oxygen species (ROS) generated in this organelle, coupled with a low level of DNA repair (Polyak et al. 1998).

Methods

Sampling

Sequences for vertebrate model organisms including; *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Mus musculus* (mouse), *Rattus novegicus* (Norway rat), *Bos taurus* (cow), *Canis familiaris* (domestic dog), *Gallus gallus* (chicken), *Xenopus laevis* (frog), and *Danio rerio* (zebrafish) were obtained from GenBank. This sampling represents all species of vertebrate for which the gene sequences for the protein subunits of the cytochrome bc_1 complex were available. Genes sampled include cytochrome *b* (cyt *b*), cytochrome c_1 (cyt c_1), rieske iron sulfur protein (ISP), core 1, core 2, subunit 6, subunit 7, subunit 8, subunit 9, and subunit 10. Subunit 11, present in some vertebrate crystal structures, is known as the first subunit to dissociate from the complex when treated with detergents. It comes in contact with subunit 10 and core 1 forming another helix through the membrane (Iwate et al. 1998; Gao et al. 2003). Because subunit 11 readily dissociates it is not easily crystallized and thus was not included in this study. Appendix 1 shows accession number and reference for each sequence used (Wong et al. 1983; Klein et al. 2002; Anderson et al. 1982; Moore et al. 2005; Gencic et al. 1991; Usui et al. 1990; Desjardins and Morais 1990; Intl. Chicken Genome Sequence Consortium 2004; Hubbard et al. 2005; Caldwell et al. 2005; Hixson & Brown 1986; Chimp Sequence Consortium 2005; Doan et al. 2005; Bjornerfeldt et al. 2006; Lindblad-Toh, K. et al. 2005; Lundrigan et al. 2002; Da Cruz et al. 2003; Brady et al. 1997; Mammalian gene collection program team 2002; Sadakata and Furuichi 2006; Behar et al. 2006; Grosskopf and Feldmann 1981; Venter et al. 2001; Nishikimi et al. 1989; Rat Genome Sequencing Project Consortium 2004; Broughton et al. 2001).

Phylogenetic Reconstruction

The five largest proteins of the bc_1 complex were used to reconstruct the phylogeny for the vertebrate cytochrome bc_1 complex with a total of 5801 bp: cyt c_1 – 981 bp; ISP – 825 bp; cyt *b* – 1148 bp; core 1 – 1446 bp; core 2 – 1401 bp (Appendix 1). Nucleotide gene sequences were aligned by inferred amino acid sequence in MEGA3 (Kumar et al. 1994) using ClustalW (Thompson et al. 1994). The five genes were concatenated using MacClade (Maddison and Maddison, 2000), and then analyzed in PAUP* (Swofford, 2003). A phylogeny was estimated using a parsimony optimality criterion. A heuristic search algorithm was employed, with 1,000 random addition replicates and TBR branch swapping. A strict consensus of all most parsimonious trees was constructed for use in subsequent analyses. Bootstrap values were also calculated using 1000 replicates with 20 random additions for each replicate and all other settings were at the PAUP* default under parsimony criteria. The most parsimonious tree resulted in the phylogeny shown in figure 3. This tree is not an estimation of a species tree, and clearly does not reflect the accepted relationships of mammals. The clade representing artiodactyls and carnivores is placed as sister to primates with rodents as outgroup, when in fact rodents is the accepted sister to primates with artiodactyls and carnivores as outgroups. The reason for this shift in placement in this particular tree is probably a function of convergent evolution of respiratory genes. Those mammals that were grouped together are all large mammals and would have different metabolic needs than very small mammals, or rodents.

Biochemical Adaptation Analysis

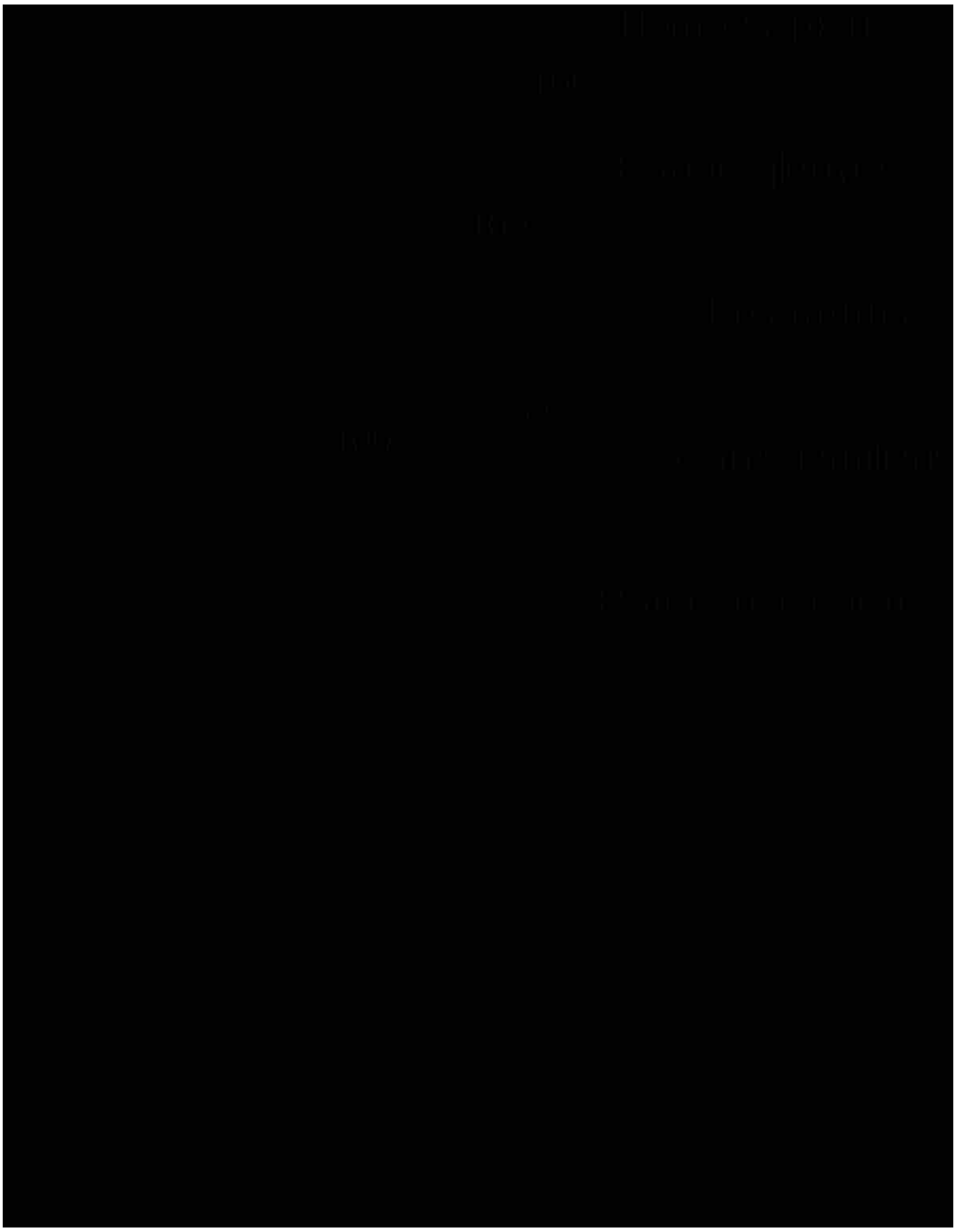


Figure 3: Parsimony tree for the cytochrome bc_1 gene sequences used in this study. Tree is representative of the gene complex relationships, not the actual species relationships.

To test for selection, the software package TreeSAAP (Selection on Amino Acid Properties using phylogenetic trees) (Woolley et al. 2003) was used. TreeSAAP uses an existing phylogenetic topology (in Newick tree format) and a protein coding nucleotide sequence alignment, to test for positive selection on quantitative physicochemical amino acid properties. Descriptions of these properties can be found in Appendix 2. This program identifies the properties historically affected by selection, and locates the exact amino acid sites that experienced significant biochemical or conformational shifts. The program TreeSAAP implements the baseml algorithm (Yang, 1997) to reconstruct ancestral character states at the nodes on a given phylogeny, then evaluates the program evaluates the average influence of selection on amino acid properties for a variable number of magnitude categories. For this study, eight categories of selection were used, and only categories 6, 7, and 8 (positive-destabilizing selection) were considered because they are unambiguously associated with molecular adaptation (McClellan et al. 2005; McClellan et al. 2001). Each of the ten genes were analyzed independently. The inferred pattern of amino acid replacement was further statistically evaluated using a sliding window to test for significant clustering of radical physicochemical shifts. A Bonferroni correction for multiple sampling was employed. All results are presented in the context of human sequences.

Visualization

Pre-existing three-dimensional structures (PDB 2A06 Huang et al. 2005; 1NTZ Gao et al. 2003; 1PPJ Huang et al. 2005) were used with the program Pymol (DeLano 2002) to show amino acid residues under heavy selection. This provides a spatial context

that transcends the linear arrangement of amino acids by presenting a more information-rich view of how these specific amino acids interact with known active sites.

Results

Cytochrome b

Cyt *b* is a membrane spanning protein 379 amino acids long that consists of eight alpha helices (A-H) and their corresponding loop regions (Fig. 4). It contains two heme molecules, heme b_H (high energy) and b_L (low energy) which are held in a four alpha helix bundle formed by A, B, C, and D helices (Iwata et al. 1998). The hemes are anchored by the conserved histidines 83 and 182 and 97 and 196. They are oriented in the membrane between the helices and almost at right angles to one another and they function to stabilize the electrons as they are shuttled from QH_2 either to cyt *c*, or to Q.

The Q_o site is near the intermembrane space, and forms a hydrophobic pocket for the binding of QH_2, and is composed of many different regions of cyt *b*; L121-Y131 of helix C, F140-S151 of the cd1 helix, I268-R282 of the ef helix and EF loop, V291-L299 of helix F (Esser et al. 2006) (Fig. 4 and Fig. 12). Sites surrounding the heme b_L are: I45-T56 of helix A, A62-C70 of the ab helix, Y81-F91 of helix B, F178-P186 of helix D. A conserved site, Glu-272 (Hunte et al. 2000), functions in the Q_o site with His-161 from ISP to transport electrons from QH_2 to heme b_L. Near the Q_o site the ISP head moves and docks via hydrogen bonding (dock sites are primarily on the cd1 helix, GH loop, the EF loop and ef helix) (Esser et al. 2006). This allows the iron sulfur center to shuttle electrons from the QH_2 site and transport them to the heme of cyt c_1 by movement in the neck region of the protein. During this exchange the first proton is released. Also present near the Q_o site is a water bridge that helps transport the second proton from QH_2 to

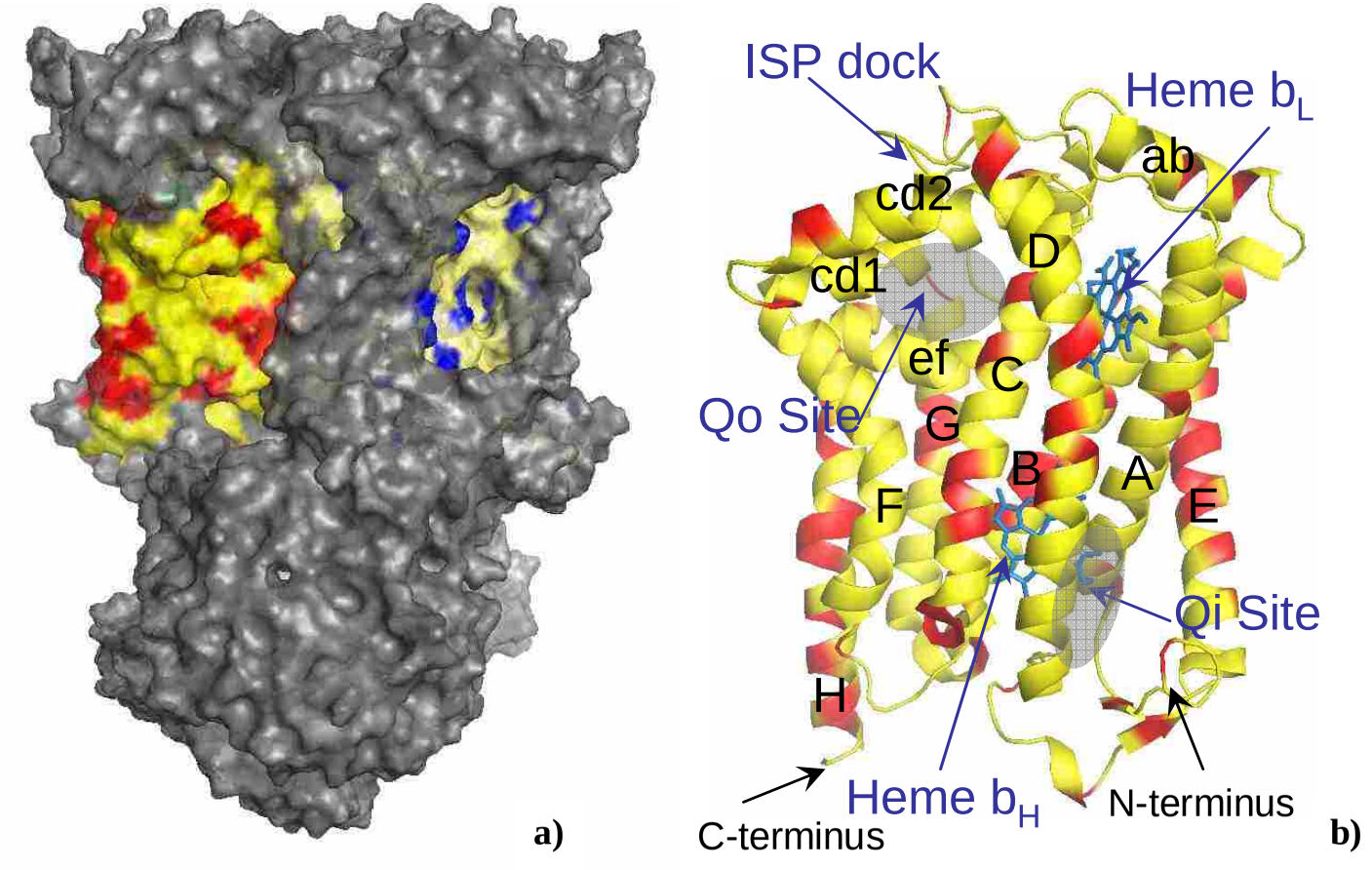


Figure 4: (a) Location of cytochrome *b* in the whole complex. Blue and red sites indicate those found to have experienced selection (different monomers of cyt *b*). (b) One subunit of cytochrome *b* is shown with active sites and hemes indicated by blue arrows and gray shaded regions. Helices are labeled with uppercase indicating transmembrane helices and lowercase indicating helices near the intermembrane space. C and N termini are also labeled.

Heme b_L to the water bridge via Glu-271, and out to the intermembrane space (Crofts et al. 1999a). The proposed exit of the proton via the water bridge is near the interface of cyt *b* and cyt c_1 just beyond the membrane (exit loop region C70-Y75). The water bridge is necessary to transport the polar proton out of the hydrophobic interior of the protein complex.

The Q_i site is less complex and functions to restore one Q to a QH_2 per Q-cycle, and is composed of A17-I42 of the N terminal and helix A, F90-V123 of helices B and C, I189-N206 of helix D, F220-D228 of helix E. It does this by using the electrons from QH_2 that were transferred through heme b_L to heme b_H, coupled with protons recruited from the matrix of the mitochondria. The first 20 amino acids of the protein are thought to be involved in recruiting these protons (McClellan and McCracken 2005).

The program TreeSAAP recovered 399 amino acid replacements in cyt *b* (Fig. 5). Of the 29 properties utilized in TreeSAAP, 18 were found to have at least one site under selection in cyt *b*. In total, 100 sites were found to be experiencing radical selection of some property, and 69 of the 379 amino acids in cyt *b* have experienced selection (some sites had multiple properties under selection). The sliding-window analysis showed multiple properties under heavy selection in four general regions of the protein, and a few properties under selection in other regions. There were six properties under selection in the region from the end of helix B to the beginning of helix C, including the bc loop region (C93-L119). This region is known to be part of the Q_i site of the protein complex. The next region under heavy selection with nine properties was in the loop region between helices C and D, which form small alpha helices, referred to as the cd1 helix and the cd2 helix (M138-G167). This region is known to be both the docking region for the

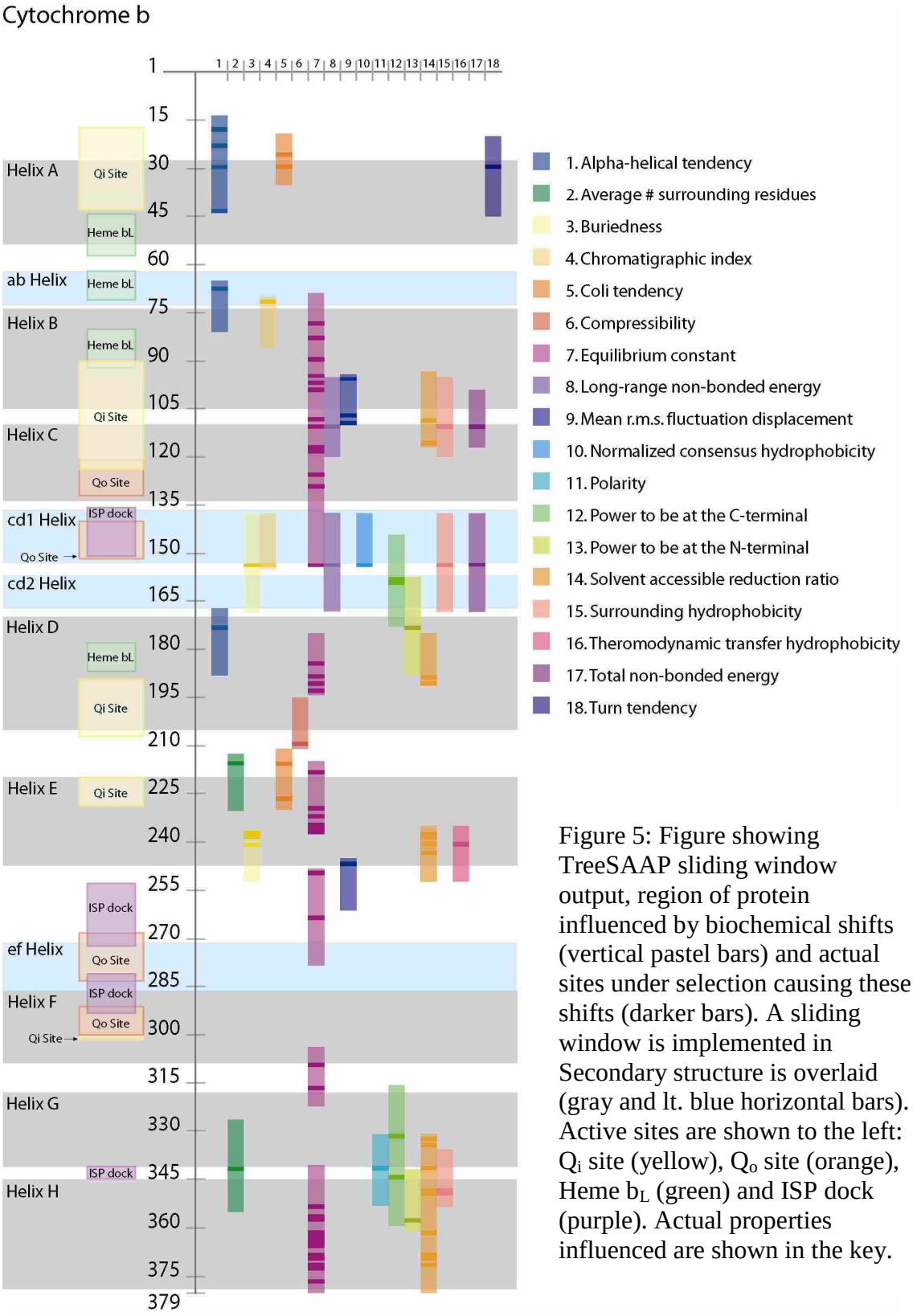


Figure 5: Figure showing TreeSAAP sliding window output, region of protein influenced by biochemical shifts (vertical pastel bars) and actual sites under selection causing these shifts (darker bars). A sliding window is implemented in Secondary structure is overlaid (gray and lt. blue horizontal bars). Active sites are shown to the left: Q_i site (yellow), Q_o site (orange), Heme b_L (green) and ISP dock (purple). Actual properties influenced are shown in the key.

movable ISP protein, and part of the Q_o site. The sliding window analysis also showed the region including the end of helix G and the beginning of helix H, and its corresponding loop region to have seven sites under heavy selection (W326-L360). This region is also part of the ISP dock when folded in three dimensions. The analysis also shows two overlapping regions under selection, one with three properties, the other with five. These regions start in the DE loop, include helix E, and end in the EF loop region (S213-N260). The beginning region has some sites that form part of the Q_i site, and the latter region surrounds the water bridge that carries the proton from the Q_o site to the intermembrane space. To see sites under selection for all amino acid properties see figure 5. To see all sites under selection in a three dimensional view see figure 4.

Cytochrome c_1

Cyt c_1 is made up of one alpha-helix that spans the membrane, and a head region that functions in the intermembrane space (Fig. 6). Protruding from the head region is a loop on the end of two beta sheets. This loop region allows cyt c_1 to come in contact with its dimer, forming what looks like a pore at the apex of the protein complex. This loop structure motif is highly conserved and present in all cytochromes, even the simplest bacteria (Baymann et al. 2004). The function of this pore is not known, but it is postulated that the first proton released may exit cyt *b* through this region (Baer and McClellan, 2006). The protein is 242 amino acids long and contains one heme with ligands His-41 and Met-160, covalent anchors Cys-37 and Cys-40 (Iwata et al. 1998). The heme of cyt c_1, functions to receive the electrons from QH_2 (active site on cyt *b*), via the iron sulfur center (Fe_2S_2) of ISP. During this transfer t releases the first proton into the intermembrane space, and transports the electrons to cytochrome *c*, a protein that

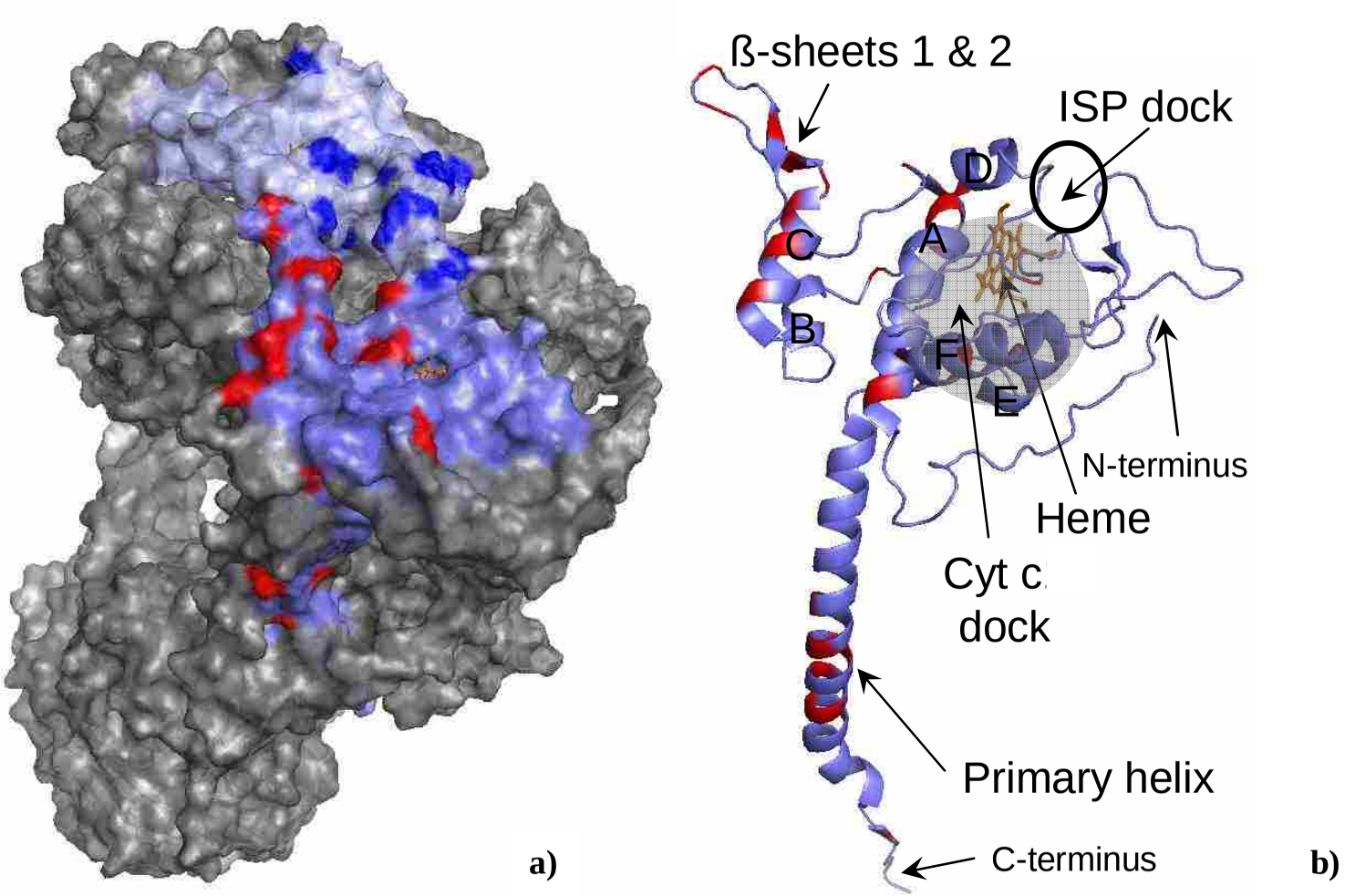


Figure 6: (a) View from the intermembrane space of cytochrome c_1 showing its location in the bc_1 complex. Subunits are colored slightly different colors of blue. Sites under selection are colored red and dark blue. (b) Only one monomer is shown here with the helices and beta sheets labeled. Also the active sites, ISP dock and cyt *c* dock are indicated. The C and N termini are labeled and the heme is shown in orange.

carries the electrons to complex 4 of the electron transport chain. The region that forms a dock for the movable ISP head includes G107-S114, R118-H121, L143-L147, and F153-I158. The heme is further surrounded by I116, E124-L131, and V186-L190. The cyt *c* dock location was crystallized in yeast by Hunte et al. (2000), and homologous regions were used to determine the binding sites on the bovine structure. Those sites are Q35-H41, Y95, A103-N106, E145-Y148, and G159-E170.

TreeSAAP recovered 93 amino acid replacements in cyt c_1 (Fig. 7). Analysis of these changes resulted in detection of selection in 15 of 29 properties. Of those amino acid replacements, 41 total sites were found to be experiencing adaptation relative to an amino acid property. Of those, eight sites experienced adaptation for multiple amino acid properties. On the linear protein, one major region (E67-A103) corresponding to the highly conserved loop and beta sheets (discussed above) of the protein, was found to be under selection with nine properties represented. Another adaptive region (L17-L51), which is part of the cyt *c* dock, had three properties under selection. Yet another region (D172-A193), which forms a region in contact with the heme group, experienced adaptation relative to three amino acid properties. Other regions experiencing adaptation can be seen in figure 7. Figure 6 shows a three dimensional view of the adaptive regions within each dimer.

Rieske ISP

The 179 amino acid long rieske iron sulfur protein (ISP) is made up of a primary alpha helix that extends through the membrane, a moveable neck region, and a head region that sits between dimers of cytochrome c_1 on the intermembrane side of the membrane (Fig. 8). The ISP is essential to the function of the Q cycle and contains an

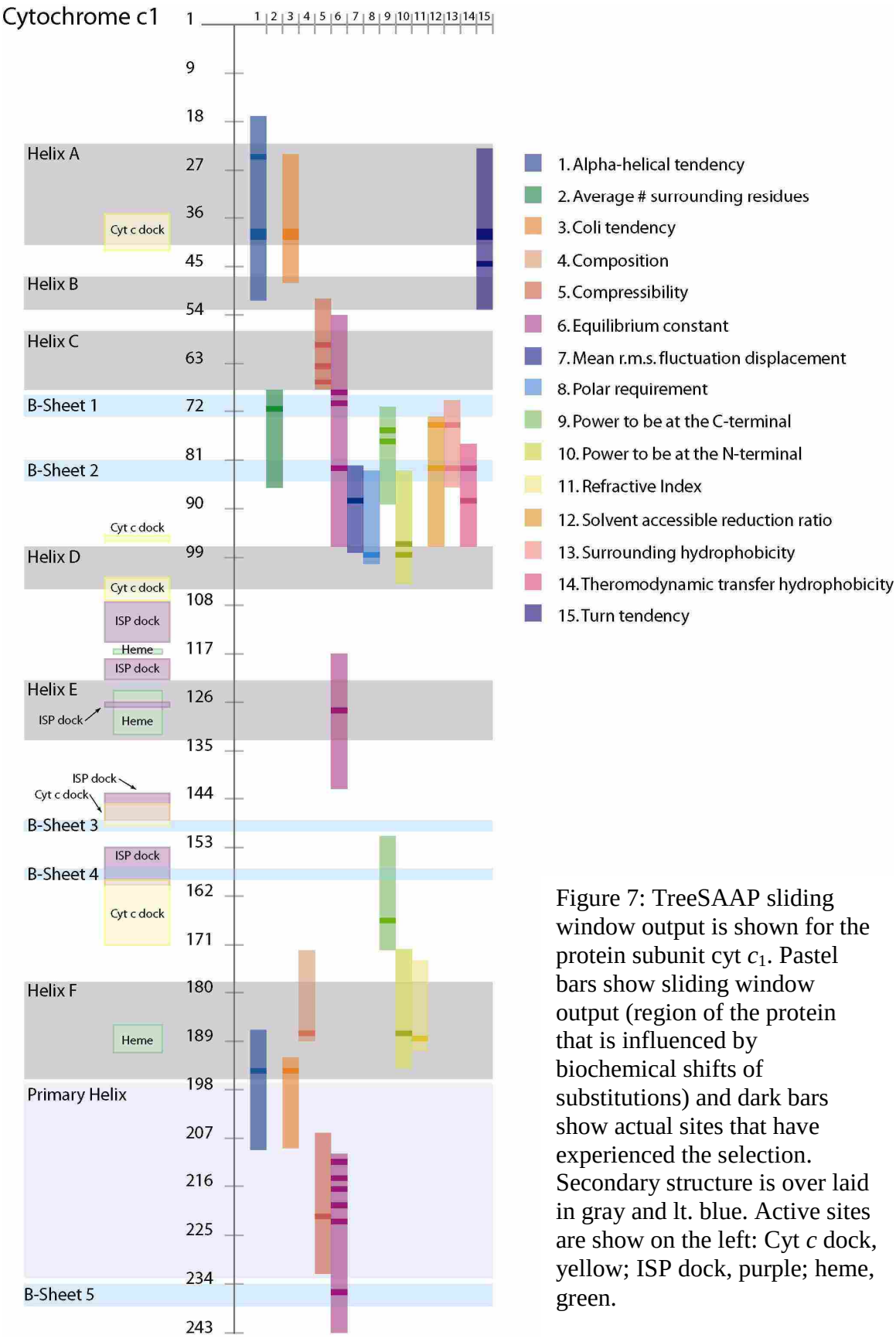


Figure 7: TreeSAAP sliding window output is shown for the protein subunit cyt c_1. Pastel bars show sliding window output (region of the protein that is influenced by biochemical shifts of substitutions) and dark bars show actual sites that have experienced the selection. Secondary structure is over laid in gray and lt. blue. Active sites are show on the left: Cyt *c* dock, yellow; ISP dock, purple; heme, green.

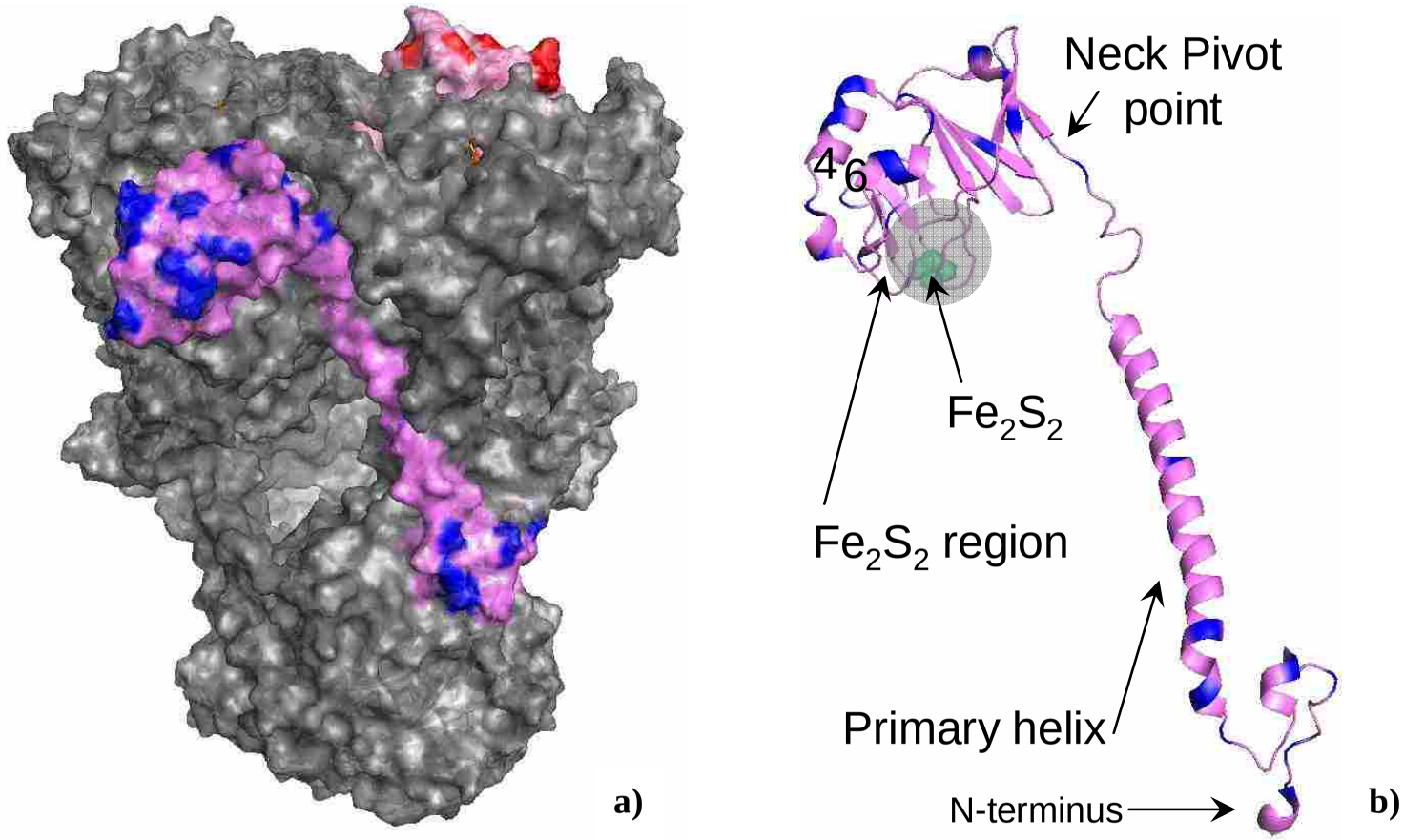


Figure 8: (a) View of both dimers of the rieske iron-sulfur protein and their location in the bc_1 complex. Specific sites under selection are shown in blue and red. (b) View of one subunit if ISP. Fe_2S_2 is shown in green. Active sites, N terminus, and helices that will be discussed later in the paper are indicated.

iron sulfur center (Fe_2S_2) anchored by His-141 and Cys-158 (Gurung et al. 2005). The head region of ISP undergoes a conformational change to allow the bifurcated reactions of the Q cycle (details of the movement of ISP are presented in the discussion). The primary alpha-helix of the ISP is K23-M62. The head region is made up of 10 short beta sheets, of which 5-8 surround the Fe_2S_2 center (V47-A48, G154-C158, S163-D166, and I171-K173, respectively) (Hunte et al. 2000). Loop regions that are in close contact with the Fe_2S_2 center, and thus interacting with both cyt c_1 and cyt *b* include R118, T140-V145, Y157-S163, and K173-L178.

TreeSAAP recovered 139 amino acid replacements in the rieske protein (Fig. 9). Out of the 29 amino acid properties used in TreeSAAP, 21 properties were found to be adapting in the ISP. There were 43 instances of sites under positive-destabilizing selection: of these, eight sites had more than one property adapting. There were four major regions found to have multiple properties under positive-destabilizing selection in this protein subunit. The first region, (S1-Y37), had five adaptive properties. This region of the protein extends through the membrane and into the matrix of the mitochondria. The next adaptive region (A88-Q108) which includes part of beta strand 2 and all of beta strand 3; TreeSAAP identified six adapting properties in this region. The third adaptive region immediately follows (E109-I134) and has nine properties under positive-destabilizing selection. It includes part of beta sheet 4, and is on the exterior surface of the protein head region. The last major region experiencing heavy adaptation (G155-P177) includes beta strands 6, 7, and 8 and functions in the stabilization of the Fe_2S_2 center. Other regions experiencing adaptation are shown in figure 9. Figure 8 presents a three-dimensional view of these adaptive regions.

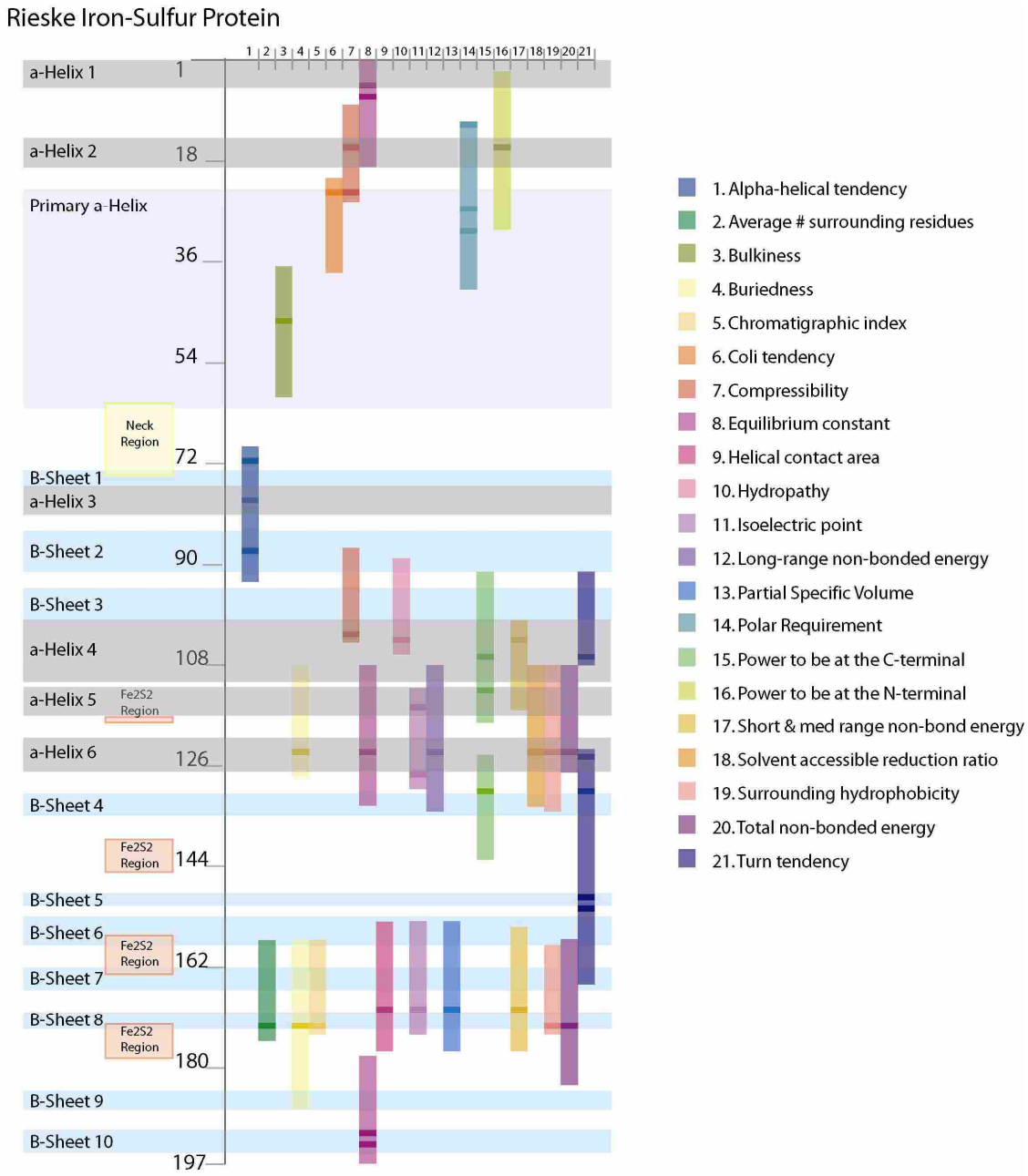


Figure 9: TreeSAAP sliding window output is shown for ISP. Pastel bars show sliding window (region of the protein influenced by biochemical shifts) and dark bars show actual sites experiencing selection. Secondary structure is over laid in gray and lt. blue. Active sites are show on the left: Neck region, yellow; regions surrounding Fe_2S_2, orange).

Other Subunits

Core 1 and core 2 are the largest subunits in the protein complex, one having 443 and the other 400 amino acids. The function of these large subunits is unknown, but studies have shown that they are required for the proper assembly of the complex, and may protect a subset of subunits against proteolytic degradation (Crivellone et al. 1988; Gatti and Tzagoloff 1990; Marx et al. 2003). Sequence homology and subsequent experimentation has shown that core 1 and 2 continue to have the same proteolytic qualities as the mitochondrial processing peptidase (MPP) (Marx et al. 2003), and it was proposed that activity had ceased due to the binding of subunit 9 in the active site at the interface of core 1 and 2 (Deng et al. 2001; Hunte et al. 2000). Others suggest that activity is only ceased because the presequence is actually cleaved in the site (Iwata et al. 1998). Each core protein is made up of two parts, the lower and upper, each consisting of a series of alpha helices and beta sheets. The beta sheets are lined up in rows and in core 1 the 5 beta sheets in the upper portion of the protein are extended by 2 more beta sheets, one from subunit 7 and the other from cyt c_1. The beta sheets of core 2 are extended by two more from subunit 9. Subunit 9 is bound by the loop region F64-N73 in core 1 (Hunte et al. 2000). Table 1 gives a description of the location, structure and function of core 1 and core 2, as well as selection information from TreeSAAP. Figure 10 shows a three dimensional view of core 1, core 2 and subunit 9.

Of the 5 other subunits, two contain membrane-spanning helices and the other three are more compact. Subunit 10 (a hydrophobic subunit), adds 2 additional helices to the complex, and may play a role in the proper assembly of the complex (Berry et al. 2000, Iwata et al. 1998). Subunit 7, also spanning the membrane, consists of a helix that

	Core 1	***Core 2***
Location	On the matrix side, this subunit sits below cyt *b*. Dimers do not come in contact at central axis of complex. Comes in contact with subunit 7, tail of rieske, tail of cyt c_1, and with the very tip of subunit 6.	Located entirely on the matrix side of the protein complex. Its subunits are sandwiched between the two core 1 protein dimers. Core 2 also comes in contact with subunits 8 and 9.
Structure	Core 1 is made up of both ß-sheets and α-helices. The ß-sheets are arranged into two sets: both with 6 ß-sheets, with one closer to cyt *b* and the other closer to the distal surface of the complex. The sets of ß-sheets are surrounded by short α- helices and loop regions.	Two clusters of ß-sheets and α-helices connected by a long loop region. First cluster has six ß-sheets arranged linearly, second has five ß-sheets, also arranged linearly. The first set of ß-sheets lines up with corresponding beta sheets of the alternative dimer in the middle of the complex. Second set of ß-sheets lines up with the ß-sheets of core 1.
Function	Exact function is not known, but they are essential for the assembly and proper association of the redox centers, though they do not contain redox centers themselves. Both proteins protect a subset of subunits against proteolytic degradation, with core 1 having a greater influence than core 2. (Crivellone et al. 1988).	
Areas under Selection	219-240 (9 properties) This region corresponds to a long loop region on the extreme exterior of the subunit that also forms an exterior surface of the complex. This region also comes in contact with the tail end of ISP.	112-166 (9 properties) This region forms 3 α-helices and 2 loops, and articulates with the core 1 protein, at one the regions experiencing selection (aa 267-302). In addition, this core 2 region comes into contact with subunits 6 and 9.
	267-302 (9 properties) This region is toward the midline of the complex, and is on the border of where core 1 comes in contact with core 2 and subunit 9.	211-237 (4 properties) This region is composed of core 2 ß-sheets 6 and 7. It is on the exterior surface of the protein, and in the three dimensional structure is adjacent to region (112-166) experiencing selection and also adjacent to core 1.
	425-443 (11 properties) This is the C-terminus of the protein that comes in contact with the N-terminus of ISP and cyt *b*, cyt c_1, and subunit 7. The N-terminal region of cyt *b* which is known to function in the recruitment of protons for attachment to Q.	345-423 (9 properties) This region is made up of α-helices and loops and is sandwiched between the two sets of core 2 ß-sheets. This region is in direct contact with core 1 on the extreme exterior of the protein, and subunit 9 within the complex.
Length	443	440
TreeSAAP properties	28	17
Selected sites/ NSyn Subs	95/310	65/504
Total aa sites selected	49	46

Table 1: Description of core proteins 1 and 2 including sites experiencing selection.

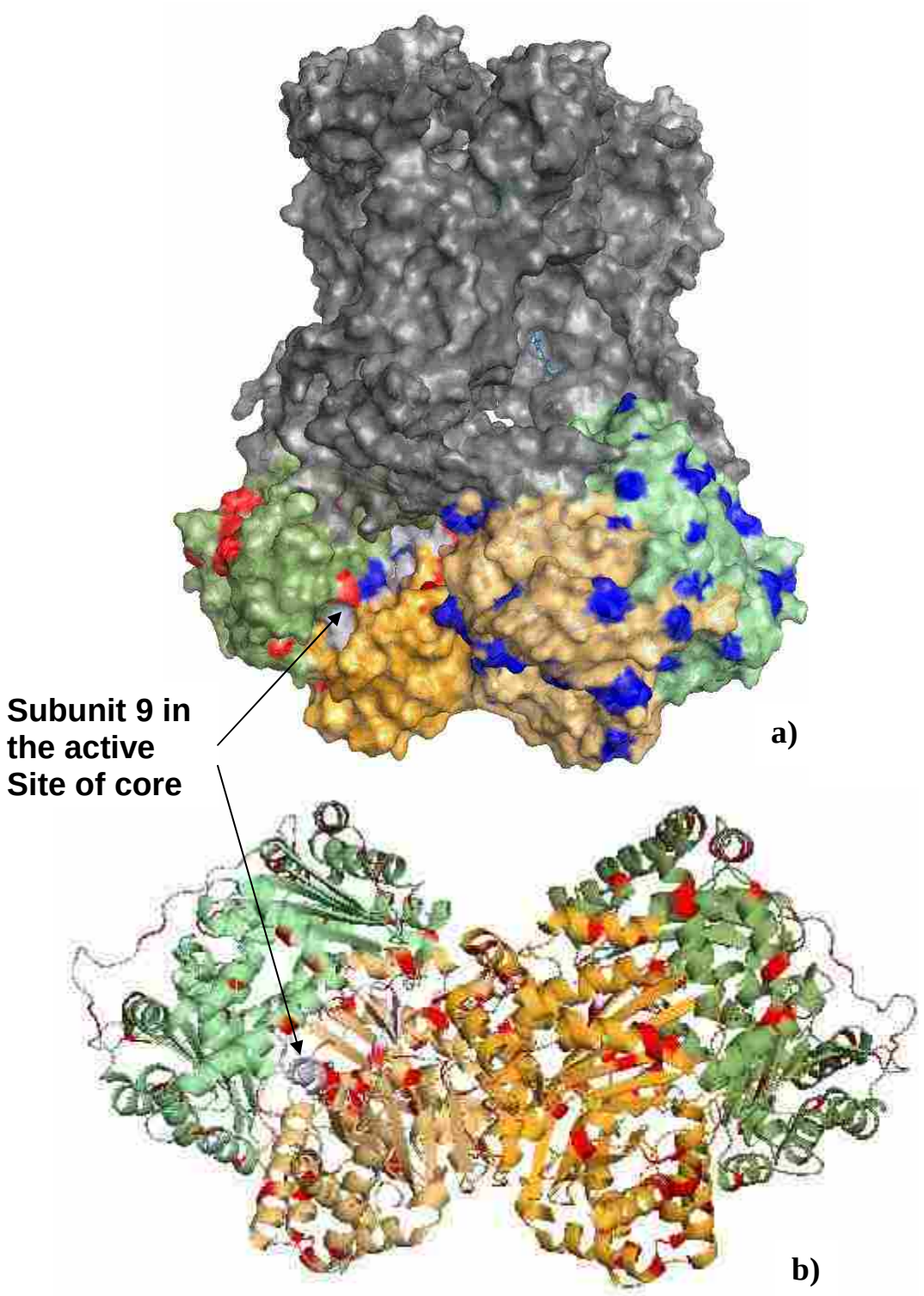


Figure 10: (a) Location of core 1 (green) and core 2 (orange), along with subunit 9 (gray) are shown here. All are found only in the matrix of the mitochondria. Sites that have experienced selection are shown in red and blue. (b) Cartoon model shows the structure of core 1 and core 2, as well as the actual sites that have experienced selection. These sites are further discussed in table 1.

stands alone across cyt *b* while other non-cyt *b* helices group together between the dimers. Subunit 6, a relatively hydrophilic subunit, forms the connection between cyt *b* and core 2 (Berry et al. 2000), while subunit 9 nestles between core 1 and core 2. Subunit 8, also called the "hinge protein" or "acidic protein," is located on the outside of cyt c_1, is the only one of these five proteins with any known catalytic property (Iwata et al. 1998). It has been shown to interact with the binding of cyt *c* and may be required for complex formation between cytochrome c_1 and cytochrome *c* (Stonehuerner et al. 1985). This binding is probably mediated by the N-terminus of the hinge protein which is highly enriched in glutamic and aspartic acid residues (Hunte et al. 2000). Table 2 gives a description of the location, structure and function of these subunits, as well as adaptation information from resulting from the TreeSAAP analysis. Figure 11 shows the specific sites under selection on subunits 6, 7, 8, and 10.

There are some disagreements among crystal structures as to the placement of subunit 9. It consists of the first 78 amino acids of the ISP targeting presequence. This leader sequence is used to transport the protein into the mitochondria (Brandt et al. 1993). Once inside it is cleaved off and becomes subunit 9. The crystal structures PDB 2A06 and 1PPJ (Huang et al. 2005) show codons 32-78, while PDB 1NTZ (Gao et al. 2003) shows 1-56. The two do not seem to be a continuation of the same sequence, rather two different sequences that form the same three-dimensional structure. Both sequences have two beta sheets that function in proximity with the first set of beta sheets in core 2 (Beta sheets add on to beta sheet 4 of core 2, 94-100). Both structures also have a small alpha helix within the loop region located between the second set of beta sheets of core 2 and the closest loop regions of core 1. For ease of discussion the three-dimensional

	Subunit 6	***Subunit 7***	***Subunit 8***	***Subunit 9***	***Subunit 10***
Location	Between cyt *b* and the core 2 and cyt C_1 helix. It also comes in contact with subunit 7, just barely with core 1. Dimers do not come in contact.	Helix spans the membrane on the exterior of cyt *b*. Intermembrane space side: contact with subunit 8 & cyt C_1. Matrix side: contact with helices of cyt C_1 & ISP and sub. 6, core 1 & 2	Nicknamed "Hinge Protein." Entirely in the intermembrane space, and is attached to cyt C_1. This subunit also comes in contact with the end of the helix comprising subunit 7.	Between core 1 and core 2, on the matrix side of the protein complex.	Has 3 helices, 1 in the intermembrane space (comes in contact with cyt C_1) and the other 2 in the membrane. Helices come in contact with the α-helices of cyt C_1 and ISP. Loop comes in contact with core 1.
Structure	4 α-helices, all at different angles, with connecting loop regions. Some loop regions contain small helices 3-4 aa.	One long α-helix with a loop region on the intermembrane space side of the complex and a ß-sheet on the matrix side. ß-sheet lines up with those in core 1, extending the set.	The subunit is essentially two α-helices held together by two sets of cysteine sulfur bridges (aa 24-S-S-68, and 40-S-S-54).	There are 2 ß-sheets and one short α-helix, the rest is loops. The two ß-sheets continue the ß-sheets of core 2.	Made up of 3 α-helices, first on cyt C_1, second crossing from cyt C_1 to helix of ISP, and the third bending to stay entirely on ISP. The loop between helix 2 and 3 comes in contact with core 1.
Function	unknown	unknown	interacts with binding of cytochrome c	unknown	unknown
Areas under Selection	14-38, 7 prop, comes in contact with G and H helices, F-G loop and B-C loop of cyt *b*	36, 38, 49 in a ring near contact with cyt C_1	18 and 65 are facing up, toward where cyt *c* comes in contact with cyt C_1	70 and 76 are on the beta sheets	5, 25, 61 are all on the surface of the protein.
	44-64, 3 prop, comes in contact with N-terminus of cyt C_1, and N-terminus of subunit 7 (regions in contact with each other)	45, 61, 68, 70, 76 all on surface of alpha helix,	23, 29, 32, 39 are all on the outside or underside of the subunit.	All others are on the loop regions.	52 is at the interface with cyt C_1 (still visible on surface)
Length	109	79	78 (13-78 crystallized)	78 (32-78 crystallized) aa 43-48 missing	62
TreeSAAP properties	12	4	5	5	4
Selected sites/ NSyn Subs	20/88	10/97	9/44	9/85 in 32-78 (140 total)	5/51
Total aa sites selected	14	10	7	8	5

Table 2: Description of subunits 6-10 including sites experiencing selection

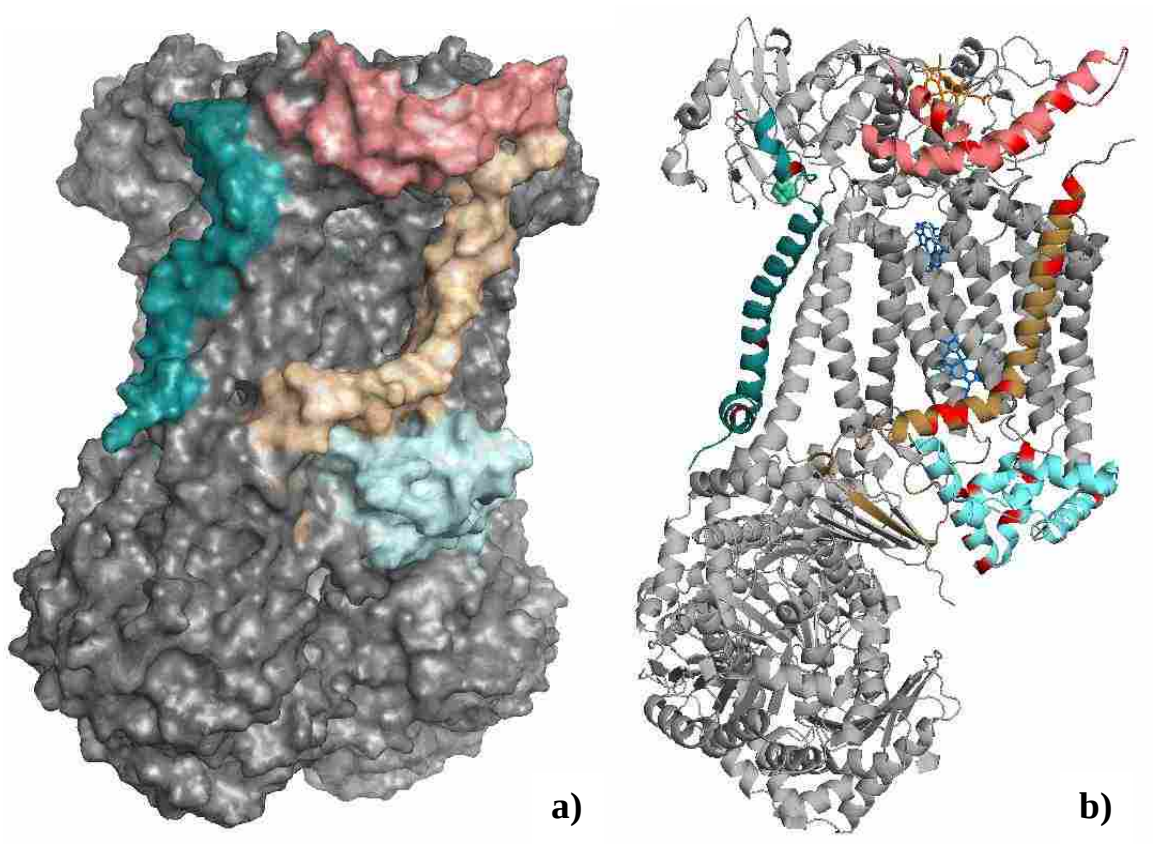


Figure 11: (a) Subunit 6 (cyan), subunit 7 (tan), subunit 8 (coral) and subunit 10 (teal) are shown. All other subunits (cyt *b*, cyt c_1, ISP, core 1, core 2, subunit 9) are colored gray. (b) Only one monomer of the bc_1 complex is shown. Sites under selection are shown in red and are further discussed in table 2.

conformation in PDB 2A06 will be used throughout the rest of the paper (Huang et al. 2005).

Discussion

As discussed above, the bc_1 complex essentially has five active sites, Q_o, Q_i, Cyt *c* dock (Fig. 13), Rieske dock on cyt *b*, and Rieske dock on cyt c_1 (Fig. 12). In addition to these, there are a few other sites of interest: (1) the rieske neck region (allows movement of head region and iron sulfur center in rieske protein), (2) the N-terminus of cyt *b* that functions in proton recruitment for the oxidized ubiquinone (near Q_i and interior of the complex) (3) the proposed exit of the first proton, the pore region formed by the monomer of cyt c_1 (Fig. 14) and (4) the water bridge that stabilizes the second proton from the Q_o site (in cyt *b*) through the membrane to the intermembrane space (Fig. 15). Of these combined active and functional sites, most are comprised of multiple protein subunits that function in concert. Generally, molecular adaptation does not work on the specific amino acids necessary for a catalytic reaction or conformational change. To the contrary, adaptation results from changes in those amino acids that surround active/functional residues to enact a slight change in the way such sites perform their role. Consequently, protein constituents of active/functional sites are forced to co-evolve in such a way that the overall roles of such regions are maintained, while allowing an overall flexibility necessary for the maintenance of organismal fitness in a dynamic environment. Thus protein coevolution can take two forms: (1) the overall maintenance of local microenvironments at protein-protein interfaces despite a mutational dynamic, and (2) the maintenance of complex function, while allowing global adaptation in response to dynamic organismal needs. The former requires compensatory

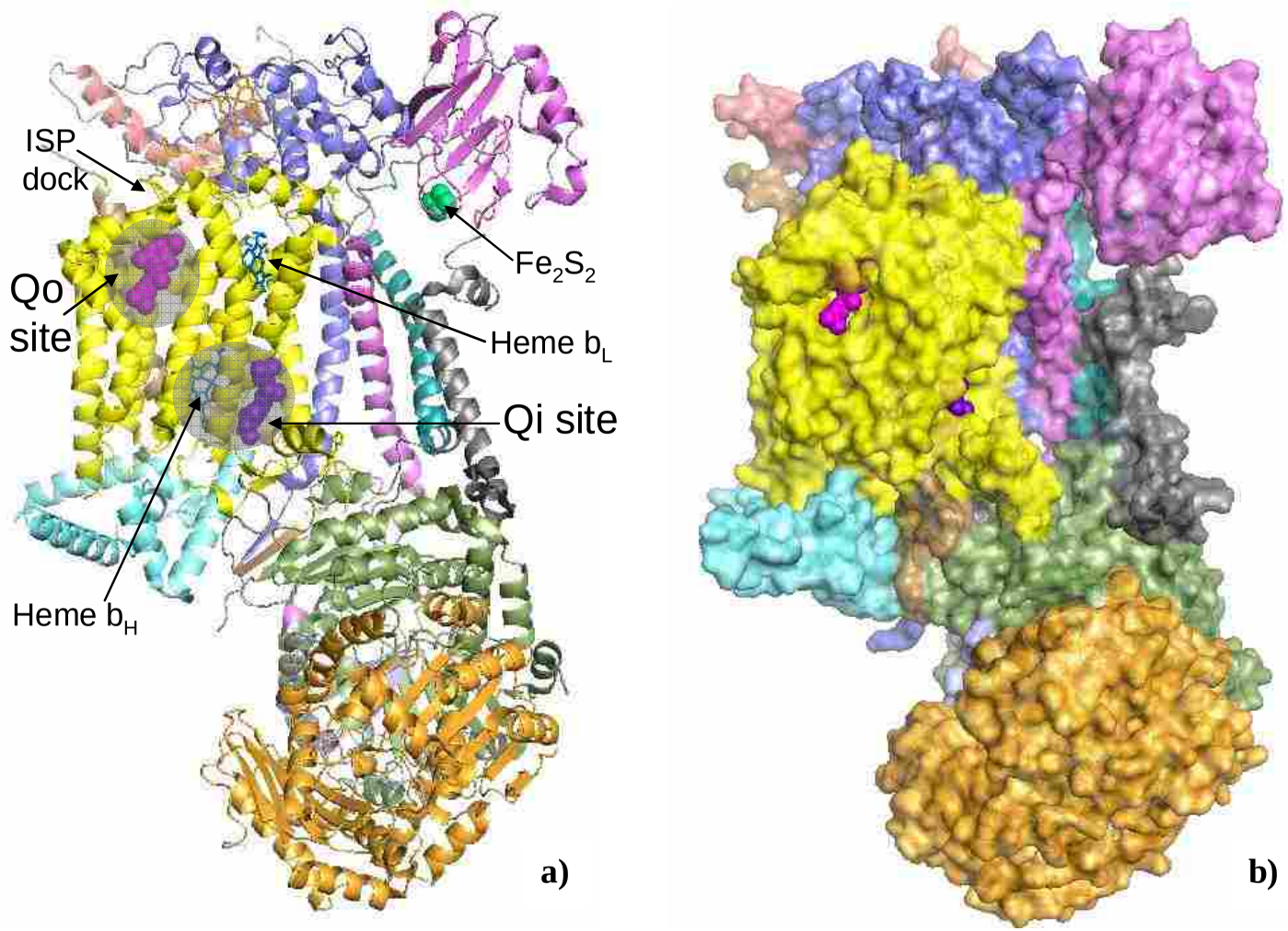


Figure 12: (a) Active sites and functional regions of the *bc*$_1$ complex are shown in only one monomer from PDB 1NTZ (Gao et al. 2003). Locations of QH_2 (magenta) in Q_o site and Q (purple) in Q_i site are shown. (b) Surface view shows the depths of the binding pockets and is slightly transparent so proximity to hemes can be seen. Subunit 11 (gray) is shown, but is not included in this study.

physicochemical changes at the protein-protein interface and can be referred to as a more classical type of coevolution, while the latter requires net physicochemical changes among proteins that are functioning in concert and represents a more dynamic form of coadaptation. Although the remainder of this discussion contains glimpses at coevolution, it will focus on the coadaptation of protein constituents that comprise the active and functional regions of the cytochrome bc_1 complex. Finally, there are many individual sites in cyt *b* that have experienced selection, however only on those sites that seem to form contiguous clusters of residues on cyt *b* near active sites, or interfaces with other subunits will be discussed.

The Q cycle, Q_o site, and the water bridge

The Q cycle begins with QH_2 binding at a hydrophilic pocket, called the Q_o site, which is located on cyt *b*, toward the intermembrane space, slightly to the side of heme b_L, and requires ISP to be functional (Crofts et al. 1999; Hunte et al. 2000) (Fig. 12). From this active site electrons either enter the high or low potential chains of the bifurcated reaction (Berry et al. 1999; Brunga et al. 2000; Crofts et al. 1999b; Esser et al. 2006). The high potential chain consists of ISP, cyt c_1 and cyt *c.* The first electron and proton are transferred from QH_2 (bound at the Q_o site) to ISP (docked on cyt *b*), then to cyt c_1 (via the movement of ISP) and ultimately to cyt *c* (Crofts et al. 1999b). ISP is essential for this process because the distance between QH_2 and the heme of cyt *c* is greater than an electron can readily diffuse. To overcome this distance, the ISP picks up an electron and its proton (His-161 and Fe_2S_2 are the actual proton and electron carriers) from QH_2 bound in the Q_o site and undergoes a conformational change to then transfer that electron to the heme of cyt c_1, at the same time releasing the proton into the

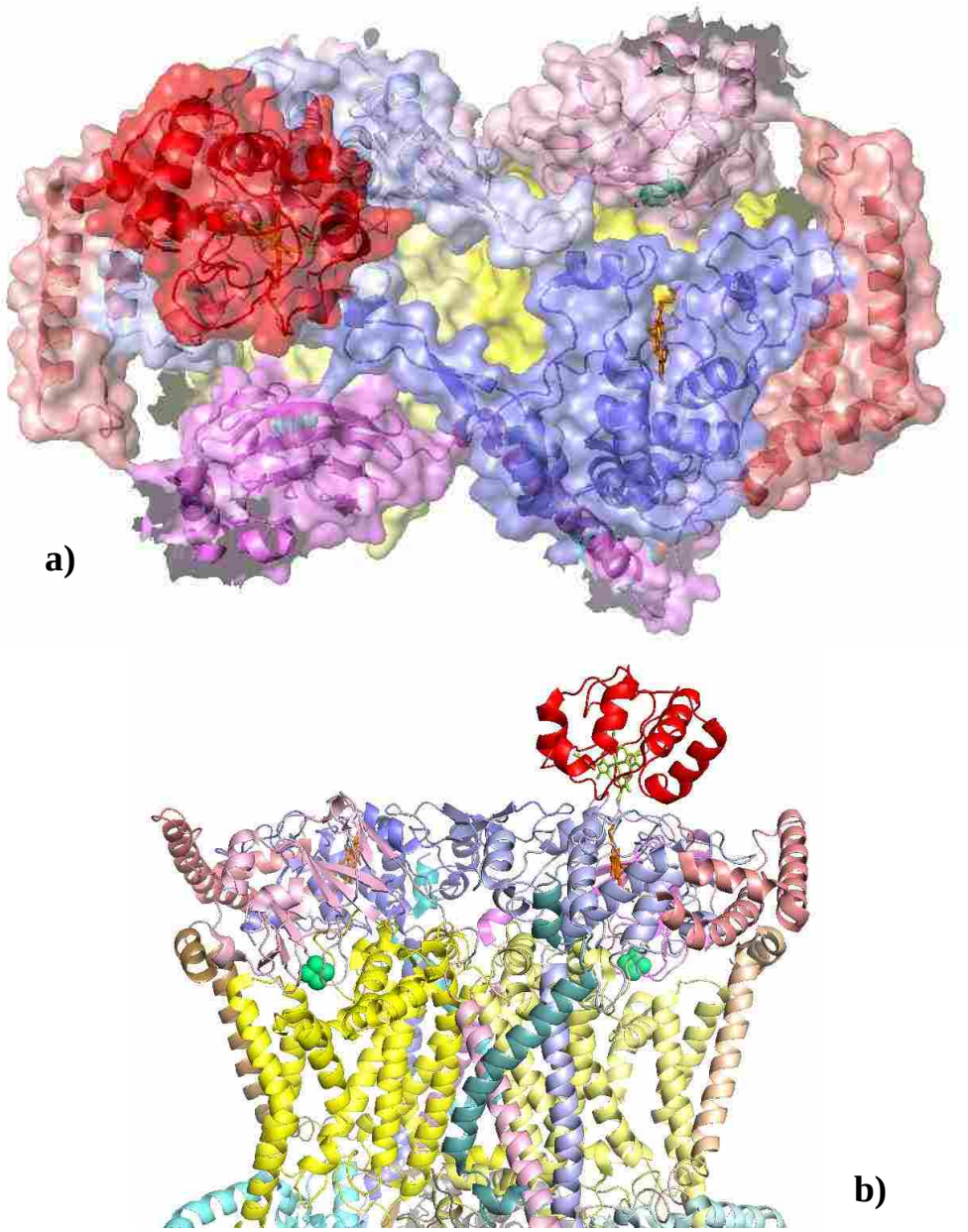


Figure 13: Location of cyt *c* dock on cyt c_1 from PDB 1KYO (Hunte et al. 2000). (a) View from the intermembrane space. Cyt *c* is shown (red), in its dock location on cyt c_1. The surface is transparent to show the location of the heme (yellow) in cyt *c*, and the heme (orange) in cyt c_1. (b) shows a cartoon model looking through the membrane. The proximity of the heme of cyt *c* and the heme of cyt c_1 can be easily seen in this view.

intermembrane space. The low potential chain consists of semiquinone in the Q_o site, the two cyt *b* hemes, and Q in the Q_i site. Electron transfer goes from the semiquinone (formed by the high potential chain), to heme b_L via Glu-271, then to heme b_H and finally to Q (bound at the Q_i site), changing it to another semiquinone. This electron is not transported by the ISP, but rather by a specific amino acid, Glu-271, that undergoes a rotational change to transport the electron to Heme b_L and the proton to the nearby water bridge. Another turn of the Q cycle to reduces this semiquinone to a QH_2. Figure 12 (PDB 1NTZ, Gao et al. 2003) shows QH_2 and Q bound at the Q_o and Q_i sites respectively.

Electron transfer has been well documented in the Q-cycle, but the location and mechanisms involved in proton release and capture are still largely unclear. The protons released from QH_2 are referred to as the "first proton" that enters the high potential chain and the "second proton" that enters the low potential chain. The first proton is released during the exchange of the electron from ISP to cyt c_1 and the second proton is released via the water bridge, which functions to take the second proton from heme b_L inside of the membrane (near the Q_o site) to the intermembrane space (Crofts et al. 1999; Izrailev et al. 1999) (Fig. 14). Not much is known about this water bridge, but it may stabilize the hydrophilic proton while inside of the hydrophobic region of the protein complex. The water chain occupies a channel leading from the heme b_L to the external aqueous phase (Crofts et al. 1999). It is postulated that the protons exit between the helices of cyt c_1 and ISP. Fig 15 shows the location of the exit of the second proton via the water bridge. Also indicated is the only surface site in this region that is adaptive, R71F (Fig. 16).

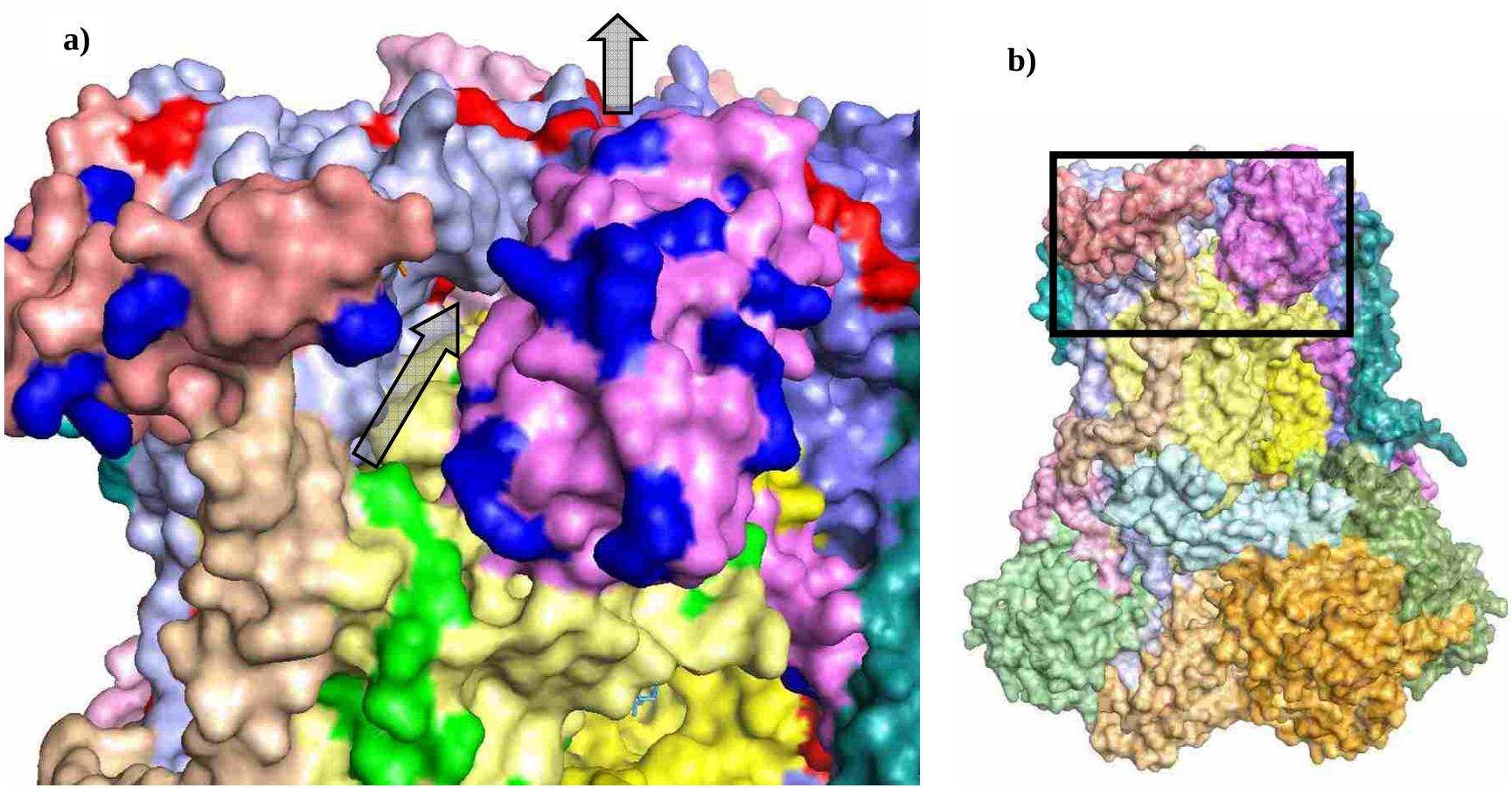


Figure 14: (a) Location of the first proton exit. Proton is released when ISP head moves toward cyt c_1 dock to transfer the first electron. The proton travels in the direction the arrows indicate, and exits through the pore in the intermembrane space created by the two subunits of cyt c_1. (b) Square shows orientation and location of (a) on the bc_1 complex.

There is a short distance between the Q_o and Q_i sites, with only a few residues between them on helices C and D of cyt *b*. This region has many sites under selection and shows a few examples of coevolution and coadaptation (Fig. 16). The basal phylogenetic branch leading to mammals has an example of coevolution on sites 184 and 192. There is a change, L184I, which causes a decrease in equilibrium constant (EC) and a change, I192L, which causes a corresponding increase in EC (Fig. 3). This reciprocal change results in the overall maintenance of the microenvironment of this region. Another example of reciprocal coevolution is on the terminal branch of *Canis familiaris*. There is a change, V117I, causing a decrease in EC and a reciprocal change, I118V, causing a corresponding increase in the property. In this same region there is an example of coadaptation on the branch leading to *Gallus gallus* where both I188L and I115T lead to a radical decrease in solvent accessible reduction ratio (SARR), making this region less accessible to water.

ISP movement and docking zones

When ISP is docked on cyt *b*, it is very close to the Q_o site allowing electron transfer (Fig. 17). There are some specific regions of cyt *b* involved in the docking of ISP via hydrogen bonding and van der Walls forces (Esser et al. 2006, Hunte et al. 2000). The cd1 helix and the ef helix have been known to help in docking the Rieske head. ISP can be seen in the docked position on cyt *b* in the structure PDB 2A06 (Huang et al. 2005), which is used for most figures in this paper. Once the electron is picked up by His-161 of ISP, the head region undergoes a conformational change and rotates the Fe_2S_2 toward the heme of cyt c_1. The iron sulfur center is essentially moved from being close to heme b_L to being close to the heme in cyt c_1. In each of the docking positions the head region is held

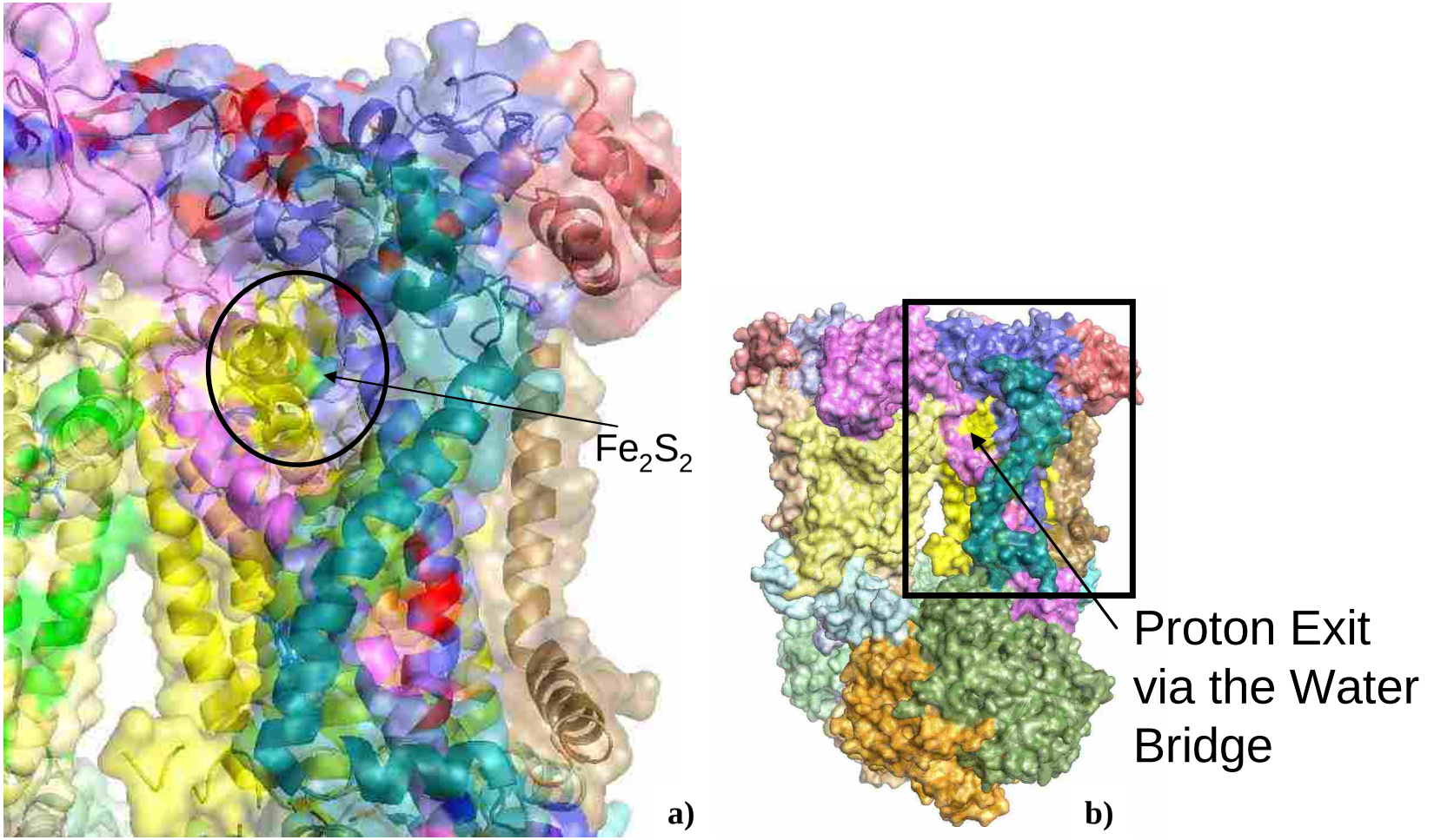


Figure 15: (a) The second proton exits via the water bridge that extends from the Q_o site through cyt *b* to the intermembrane space. The proposed exit of this water bridge is shown here. (b) Square shows orientation and location of (a) on the bc_1 complex.

in place by hydrogen bonds that form and break sequentially to allow movement between both docking sites (Iwata et al. 1998, Izrailev et al. 1999) (Fig. 12).

ISP is anchored by an N-terminus alpha-helix in the membrane (Fig. 8). The "neck region" (directly following the anchor) is the moveable part of the subunit acting as a pivot point for the large head region to move from its dock on cyt *b* to its dock on cyt c_1 (Izrailev et al. 1999). Studies have shown that there may also be a third position of the ISP head, called the intermediate position, where it is neither docked on cyt *b* or c_1, but hydrogen bonded somewhere in between and is inactive (Iwata et al. 1998). The neck region of the protein is from A70-I74 and the major pivot point is Ile-74 (Shown in Fig. 18, Iwata et al. 1998). Without this ISP motion between cyt *b* and cyt c_1, electrons would not be able to readily diffuse between the Q_o site and the heme of cyt c_1 (Fig. 14). As a result, the Q cycle would not be completed and the whole complex would lose efficiency and function (Kim et al. 1998).

One site on ISP in the highly conserved neck region had experienced positive-destabilizing selection for an increase in alpha-helical tendency (Fig. 18). This change (S72A) occurs on the phylogenetic branch separating all vertebrates from humans and chimps (Fig. 3). This change is consistent with the findings of Doan et al. (2005) who concluded that cyt *b* and ISP exhibited accelerated rates of amino acid replacement as a result of positive selection in primates. Since this conformational property occurred in a loop region and not a helix of the protein, it is hard to interpret. But the residue's proximity to the pivot point and to cyt c_1 suggests it may be affecting the efficiency of the pivot.

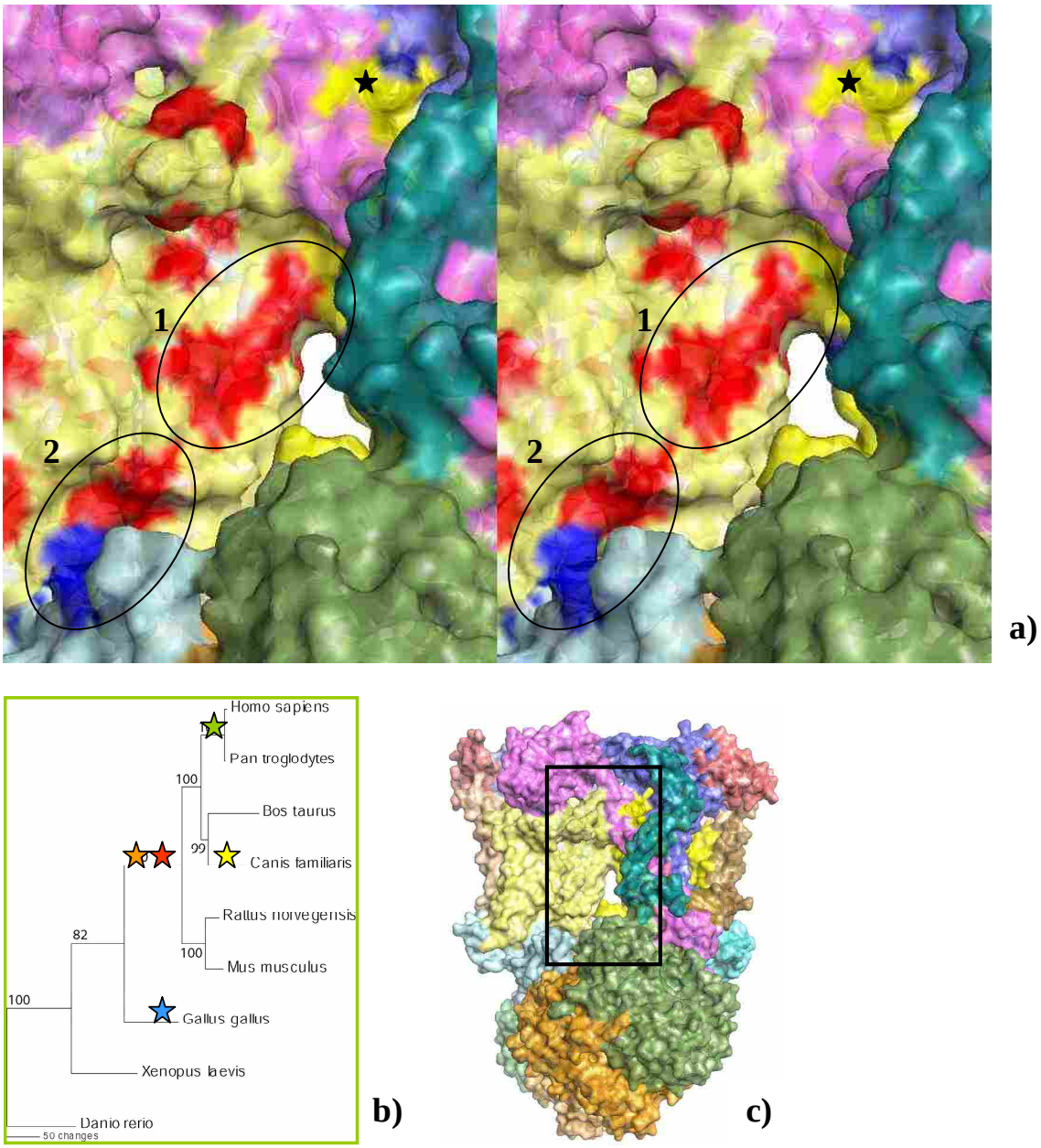


Figure 16: (a) Close up view of the transmembrane region, specifically the region of cyt *b* that is between heme b_H and heme b_L. Selection on cyt *b* is shown in red on one dimer and indigo on the other dimer. Selection on subunit 6 is shown in bright blue. In circle 1, which is directly between the two hemes of cyt *b*, there are 2 examples of coevolution and one of coadaptation. The changes: L184I and I192L (cyt *b*) increase and decrease equilibrium constant respectively on the branch leading to mammals (red star). The changes: V117I and I118V (cyt *b*) increase and decrease equilibrium constant respectively on the terminal branch of *Canis familiaris* (yellow star). The changes: I188L and I115T (cyt *b*) decrease SARR on the terminal branch leading to *Gallus gallus* (blue star). In circle 2, which is directly behind heme b_L, there is one example of coevolution and one example of coadaptation between cyt *b* and subunit 6. Sites L108I, K110I (cyt *b*) decrease equilibrium constant, Y109F (cyt *b*) decreases mean r.m.s. fluctuation displacement & D40N (sub 6) decreases power to be at the C-terminus on the branch leading to all mammals (orange star). Later in the phylogeny, on the branch leading to primates (green star) there are reversals of the sites above, except I110L instead of K. The star shows the one site under selection near the location of the second proton exit. The indigo site indicates R71F (cyt *b*) and is only under selection on the terminal branch to *Xenopus laevis*. (b) Phylogeny showing location of changes discussed in the paper. Red, yellow and blue starts are in circle 1 and orange and green are in circle 2. (c) Entire complex shown, with box indicating the close-up region in (a).

Figure 19 shows a large contiguous region or ‘stripe’ of adapting sites on the surface of cyt *b*, which starts in the ISP dock site, and continues down helix H through the inner mitochondrial membrane. This region is near the interface of cyt *b* with subunit 7 and ISP. There are examples of both coevolution and coadaptation occurring here. On the branch separating *Danio rerio* from the ingroup, there are six sites experiencing radical biochemical shifts, with four of them decreasing equilibrium constant (EC) (L361I, V364I, L372I, A376I). This coordinated negative shift in EC is an example of coadaptation—a concerted directional change. All four of these changes are from a hydrophobic amino acid to an isoleucine, that is slightly more hydrophobic than the original residues (L, V, and A).

In this same ‘stripe’ on cyt *b* coevolution also is seen on the branch leading to *Xenopus laevis* (frog) where there are 13 sites experiencing biochemical shifts (Fig. 3). Of these, six sites experienced an increase in EC (I353L, I356V, I368L, I372V, and I376L of cyt *b* and I58V of subunit 7) while three sites experienced a decrease (L357I, L363I, L365I of cyt *b*) the EC of the site. These changes resulted in an overall maintenance of the microenvironment in *Xenopus*; they were exchanges between nearly equally hydrophobic amino acids, which only caused a shift in one other property: EC. Additionally, one site experienced an increase in SARR (T369I of cyt *b*), while two sites decreased (I348T and I332T of cyt *b*). One of these sites also experienced a decrease in surrounding hydrophobicity (I348T) (Fig. 19).

Other parts of the phylogeny also have exhibited adaptation within this region of cyt *b*. The phylogenetic branch leading to *Gallus gallus* (chicken) shows both coadaptation and coevolution—three sites experienced coordinated increases in EC

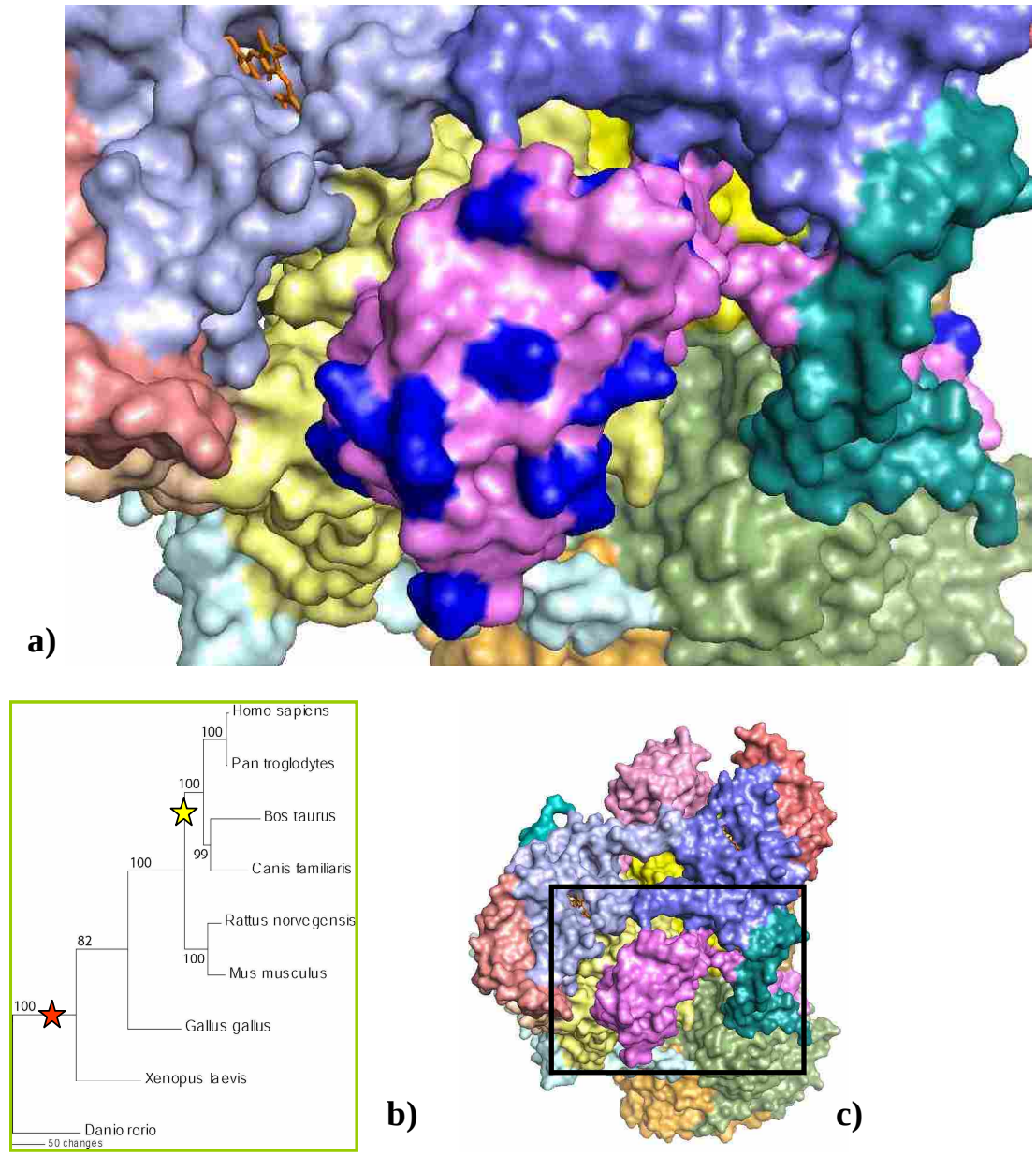


Figure 17: (a) Selection on the ISP head region is shown in bright blue. The ISP head region shows many sites experiencing physicochemical shifts. There is one example of coadaptation on the head region of ISP: E128K & S170R both increase isoelectric point and one example of coevolution: D131E & E107D reciprocal increase and decrease in turn tendency on the phylogenetic branch separating *Danio rerio* as the outgroup (red star). There is also one example of coevolution on the branch grouping humans, chimps, cow, and dog: I192M & V194I show reciprocal increase and decrease in the equilibrium constant (yellow star). The other changes, that do not show coevolution or coadaptation include: T88A (alpha helical tendency), S72A (alpha helical tendency), S79A (alpha helical tendency), E103A (compressibility), K104A (hydropathy, short and med range non-bonded energy), D107N (power to be at the C-terminus), N113D (power to be at the C-terminus), E116K (isoelectric point), I124K (buriedness, equilibrium constant, long-range non-bonded energy, SARR, surrounding hydrophobicity, total non-bonded energy), D125E (turn tendency), D131S power to be at the C-terminal), A150S (turn tendency), E152D (turn tendency), and S170R (short and medium range non-bonded energy, partial specific volume, helical contact area). (b) Phylogenetic tree and locations of the evolutionary changes discussed above. (c) Square shows the region of the bc_1 complex that is enlarged in (a).

(I372L, I356L of cyt *b*, and I58L of subunit 7), while SARR was maintained by reciprocal changes (I371T and I368T on cyt *b* experienced decreases and T369I experienced an increase in this property). Again fluctuations are seen in SARR and EC, both properties associated with protein-protein interactions, EC dealing with the ability to release a proton, and SARR with water interactions. Also, the branch leading to *Bos taurus* (cow) presents evidence of coadaptation with four substitutions causing a coordinated increase in EC (I361L, I356V, I364V, I376L of cyt *b*), and two causing a coordinated decrease in SARR (I371T, I349T). An increase in EC and a decrease in SARR, seem to be highly correlated and cause an overall change in this region. The branch leading to *Canis familiaris* (dog) shows coadaptation as well, with all three substitutions increasing EC (I376L, I371V, I362L of cyt *b*) (Fig. 19).

There are other substitutions not mentioned, three changes on the branch leading to *Rattus norvegensis* (rat) (I376M and I372V EC increase, L357S power to be at the N-terminus), five changes on the branch grouping *Bos* and *Canis* (I353L of cyt *b* and I58V of subunit 7 increasing EC, I268T and I344T of cyt *b* decreasing SARR, and E344S of cyt *b* decreasing power to be at the C-terminus), four changes on the branch grouping *Pan troglodytes* and *Homo sapiens* (I371L and I356V increasing EC, I361T decreasing SARR, and K173L of ISP increasing total non-bonded energy), and five changes on the branch to *Homo sapiens* (A369I and T309I decreasing and I353V increasing EC, I368T and I348T decreasing SARR, and I348T also decreasing surrounding hydrophobicity). These changes show more examples of coadaptation by coordinated directional selection. In these membrane-bound helices there is a high instance of substitutions between similarly hydrophobic amino acids that increase or decrease EC, SARR, and occasionally

surrounding hydrophobicity. All three of these properties essentially cause shifts in an amino acids ability to interact with water and protons just in slightly different ways. This allows the proteins to dynamically adapt to changing energy requirements without losing overall protein-complex function (Fig. 19).

The entire head of the ISP (not just the docking region) has experienced adaptive change at 18 sites, and nearly 95% of them are on the surface facing out toward the intermembrane space (Fig. 17). There are two clear examples of coevolution and one of coadaptation on the ISP head region. The first is an example of coevolution on the phylogenetic branch separating *Danio rerio* from the ingroup. The change is between the closely associated D131E and E107D which decrease and increase turn tendency, respectively. On this same node coadaptation occurs between E128K, which increases isoelectric point and S170R which also increases isoelectric point (in addition to increasing helical contact area and partial specific volume and decreasing short and medium range non-bonded energy). The other coevolutionary changes, I192M and V194I, occur on the branch separating *Homo sapiens*, *Pan troglodytes*, *Canis familiaris* and *Bos taurus* from the rest of the phylogeny and result in an increase and decrease of EC, respectively (Fig. 17). The suite of properties affected in this region is drastically different than that of the membrane-spanning portion of the bc_1 complex. Most of these changes involve either a D, E, A, or K. Amino acids D, E and K all have relatively large and reactive R-groups which are often involved in the function of active sites, hydrogen bonding and protein-protein interactions. These properties include two conformational properties (alpha helical tendency and turn tendency), two that measure proton release (isoelectric point and EC), and three that measure types of energetic interactions (short

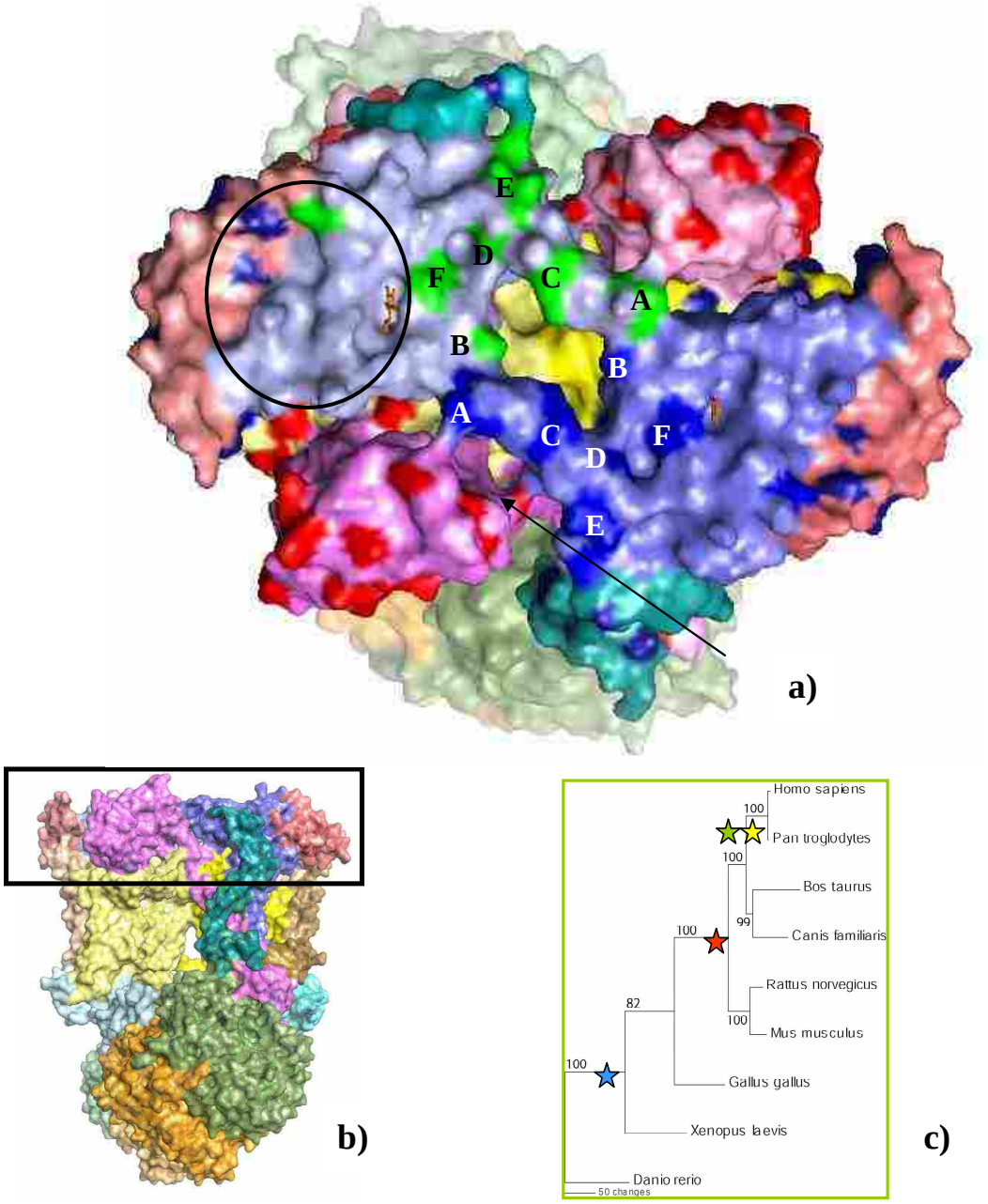


Figure 18: (a) The pore region of the protein complex, in the intermembrane space, is formed by the two monomers of cyt c_1. Selection on ISP is shown in red and selection on subunit 8 is shown in indigo. Selection on cyt c_1 is shown in bright green and bright blue. A-F indicate areas (clusters) of selection on both dimers. A: N77D & D75N ± power to be at the C-terminal, while A96P region B shows a decrease in power to be at the N-terminal (red star). C and D: T82I & I68V ± equilibrium constant which shows coevolution (blue star). E: G66A & A63G ± compressibility which shows coevolution (yellow star). F: A39S decreases alpha-helical tendency while increasing coil and turn tendencies and E44D increases turn tendency, shows coadaptation (red star). At the interface between cyt *b* and cyt c_1, just below the pore region there is an example of coevolution T67A and A173P ± alpha helical tendency (green star). The circle indicates the docking region of cyt *c*. Most changes here do not indicate coevolution or coadaptation, rather suggest a conservation of this site, and occur on numerous branches throughout the tree. The arrow indicates site S72A, alpha helical tendency, in neck region of ISP. (b) View of the bc_1 complex indicating the region in part (a). (c) Phylogenetic tree showing the locations of the changes discussed above.

and medium non-bonded energy, total non-bonded energy, and power to be at the C-terminus). This suite of properties could influence conformational changes that affect ion exchange between secondary structures of the moveable ISP.

Cyt c binding site and Cyt c_1 pore

Once an electron is transferred from His 161 and Fe_2S_2 of ISP to cyt c_1, it must then be transferred to the mobile carrier cyt *c* to complete the high-energy chain of the Q cycle (Figs. 13 and 14). Cyt *c* binds near the heme of cyt c_1 on the intermembrane side of the complex. The binding location is shown in figure 13 using a crystal structure of yeast (Lange and Hunte 2002) that has the mobile cyt *c* bound to cyt c_1. Although yeast and bovine amino acid sequences are not identical, the overall function and structure of the protein complex is similar. Using the yeast structure as a model, the binding site on bovine cyt c_1 was inferred (shown in three dimensions on Fig. 18 and labeled on the linear protein in Fig. 7).

Between the binding sites of cyt *c* on each dimer is the pore region of the complex formed by beta strands of cyt c_1 (Fig. 18). This region has been postulated to be the location where the first proton is released during the exchange of the electron between ISP and cyt c_1 (Baer and McClellan 2006). Most of the adaptation seen here seems to be the result of coevolution, or the biochemical maintenance of the region. In each of the small clusters the same properties are being conserved by reciprocal changes in adjoining sites. Much of the adaptation here is occurring on the phylogenetic branch between mammals and other vertebrates (Fig. 3). For example, the cluster on the loop region between the beta sheets 1 and 2 exhibit the reciprocal changes D75N and N77D, one decreasing and the other increasing power to be at the C-terminal (A in Fig. 18). And

example of coadaptation is seen in A39S of cyt c_1 (F in Fig. 18) decreases alpha-helical tendency while increasing coil and turn tendencies while E44D of cyt c_1, also increase turn tendency. These coordinated directional shifts in the structure of sites at the interface of cyt c_1 and cyt *b* is a clear case of protein-protein coadaptation.

Within the mammal clade, on the branch leading to primates (*Homo sapiens* and *Pan troglodytes*), there are two examples of coevolution where reciprocal changes preserve the biochemical microenvironment of the pore region shown in figure 18. The cyt c_1 changes, G66A and A63G, increase and decrease compressibility, respectively (E in Fig. 18). Also on this branch, changes in cyt *b* located near the protein interface of the pore region of cyt c_1 (T67A and A173P) increase and decrease alpha-helical tendency, respectively. These are changes of both a conformational (alpha-helical tendency) and a structural (compressibility) property, which makes sense because the change in conformation would necessitate the change in structure of a particular region in the complex.

Another example of coevolution is on the phylogenetic branch separating *Danio rerio* from the ingroup. There are reciprocal changes, I68V which increases equilibrium constant (EC) (D in Fig. 18), and T82I which decreases EC, as well as increases surrounding hydrophobicity, SARR, and thermodynamic transfer hydrophobicity (C in Fig. 18). These two regions are adjacent in three-dimensions and both face the pore opening.

The only other branch on the phylogeny with more than one change is the terminal branch leading to *Xenopus laevis*, where there are 5 changes. The reciprocal changes, V70I and I82L, decrease and increase EC respectively (C in Fig. 18). The other

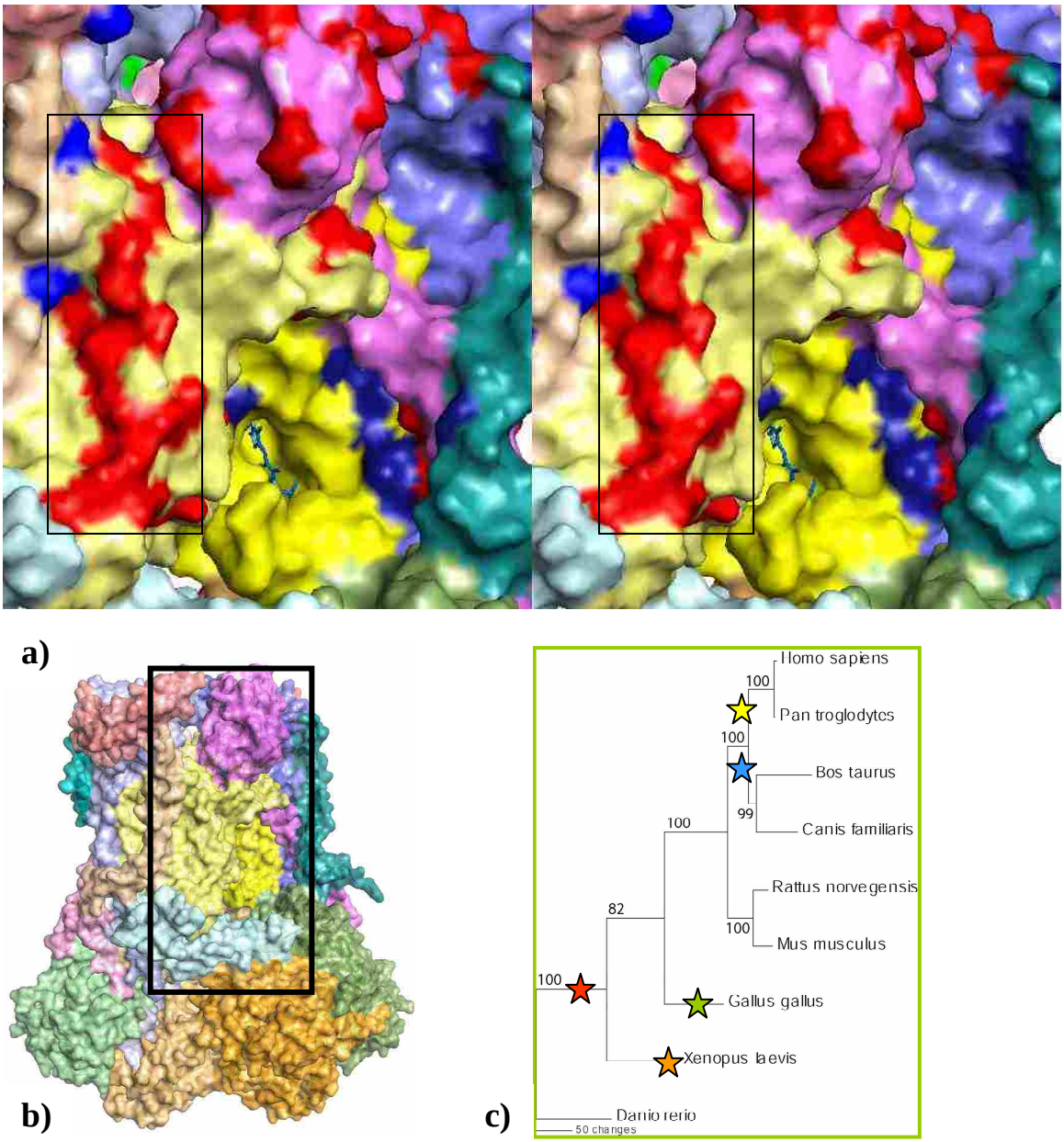


Figure 19: The box indicates those sites that are under selection on and around helix H of cyt *b*. Selection on cyt *b* is shown in red, as well as selection on ISP. Subunit 7 selection is shown in bright blue. This region is undergoing heavy selection with many examples of coadaptation occurring here. The changes L361I, V364I, L372I, A376I (cyt *b*) all decrease equilibrium constant (red star). The changes I356V& I371L (cyt *b*) increase equilibrium constant while I361T (cyt *b*) decreases SARR (yellow star). The changes I353L (cyt b) & I58V (sub 7) increase equilibrium constant while I368T& I334T (cyt *b*) decrease SARR (blue star). The terminal branches to *Xenopus laevis* and *Gallus gallus* show more examples of coevolution and coadaptation. *Xenopus* shows six sites that increase and three that decrease equilibrium constant, while two sites decrease and one site increases SARR in this 'stripe' region (orange star). *Gallus* shows three sites that increase equilibrium constant, while two sites decrease and one site increases SARR (green star). (b) Shows the region of the protein viewed in (a). (c) Shows phylogenetic tree with locations of changes discussed above.

changes on this branch radically shift different properties: G59A increases compressibility (E in Fig. 18), S38A increases alpha-helical tendency and decreases coil and turn tendencies (F in Fig. 18), and P98E increases power to be at the N-terminal and polar requirement (B in Fig. 18). Many of these changes in the pore region occurred on the backbone branches of the phylogeny. This placement is of interest because it may mean that radical changes in this area of the protein are likely to result in significant shifts in overall metabolism strategy, which would be expected in the change from aquatic to terrestrial, and then again from cold-blooded to warm-blooded.

Q_i Site and proton recruitment

The Q_i site is imbedded in cyt *b* close to the matrix side of the membrane (Fig. 12). The site is associated with the heme b_H, and electron transport to Q is direct—not facilitated directly by other amino acids like at the Q_o site. The oxidized ubiquinone (Q) binds in this pocket and after two turns of the Q cycle (2 QH_2 are oxidized to Q at the Q_o site) ubiquinone in the Q_i site is reduced to ubiquinol (QH_2) and can be used in the cycle again. This takes two protons and two electrons, or simply two hydrogens to change the ketones of Q to the alcohols of QH_2. The electrons used come from the low-energy chain of the Q cycle, but the protons once associated with these electrons are shuttled across the membrane to the intermembrane space to create the gradient needed for ATP synthesis. Therefore protons must be recruited from the matrix to reduce Q. The first 20 amino acids of cyt *b* have been shown to function in this proton recruitment (Degli Esposti et al. 1993). This region contains two small alpha helices located near the matrix side of the complex (and can be seen in Fig. 12 directly below where Q is bound). Also, the specific cyt *b* sites His-201 and Lys-227, both part of the Q_i site, have been shown to undergo

conformational changes coupled to ubiquinone reduction, and it has been suggested that they are involved in recruiting protons from the matrix side (Gurung et al. 2005). Additionally between the dimer subunits of cyt *b* there is a hole filled with ordered water molecules, this may contribute to the protons used at the Q_i site (Gurung et al. 2005). Other subunits in this region may also be involved in this recruitment of protons, especially core 1 and subunit 6, subunit 7, and the helices of subunit 10, ISP and cyt c_1.

Many of the sites in the region involved in proton recruitment are changing on the terminal branches of the phylogeny only, so for clarity, mainly those changes on the internal branches will be discussed. On the phylogenetic branch separating *Danio rerio* from the ingroup there are five changes: V13I (sub. 6) decreases equilibrium constant (EC) while I216L and I222M (cyt c_1) increase EC. The other two changes are A17S from cyt *b* that decreases alpha-helical tendency and G47T from ISP that increases bulkiness. The changes in EC may be coevolutionary in maintaining the biochemical property of the region. There are also three changes on the branch separating *Xenopus* and *Danio* from the ingroup: Y221A of cyt c_1 increases compressibility, while L229I and N226I of cyt *b* decrease EC and coil tendency respectively. Although these changes are not reciprocal or coordinated, they do suggest that radical changes to this region of the complex may be evolutionarily important. Interestingly 9 of the changes in core 1 were on the terminal branch leading to *Pan troglodytes*, and 3 were on the terminal branch leading to *Gallus gallus*.

There are many sites experiencing selection in and around the Q_i site that may or may not be involved in water recruitment, but are certainly influencing the function of the active site due to proximity. One such cluster of sites experiencing selection involves

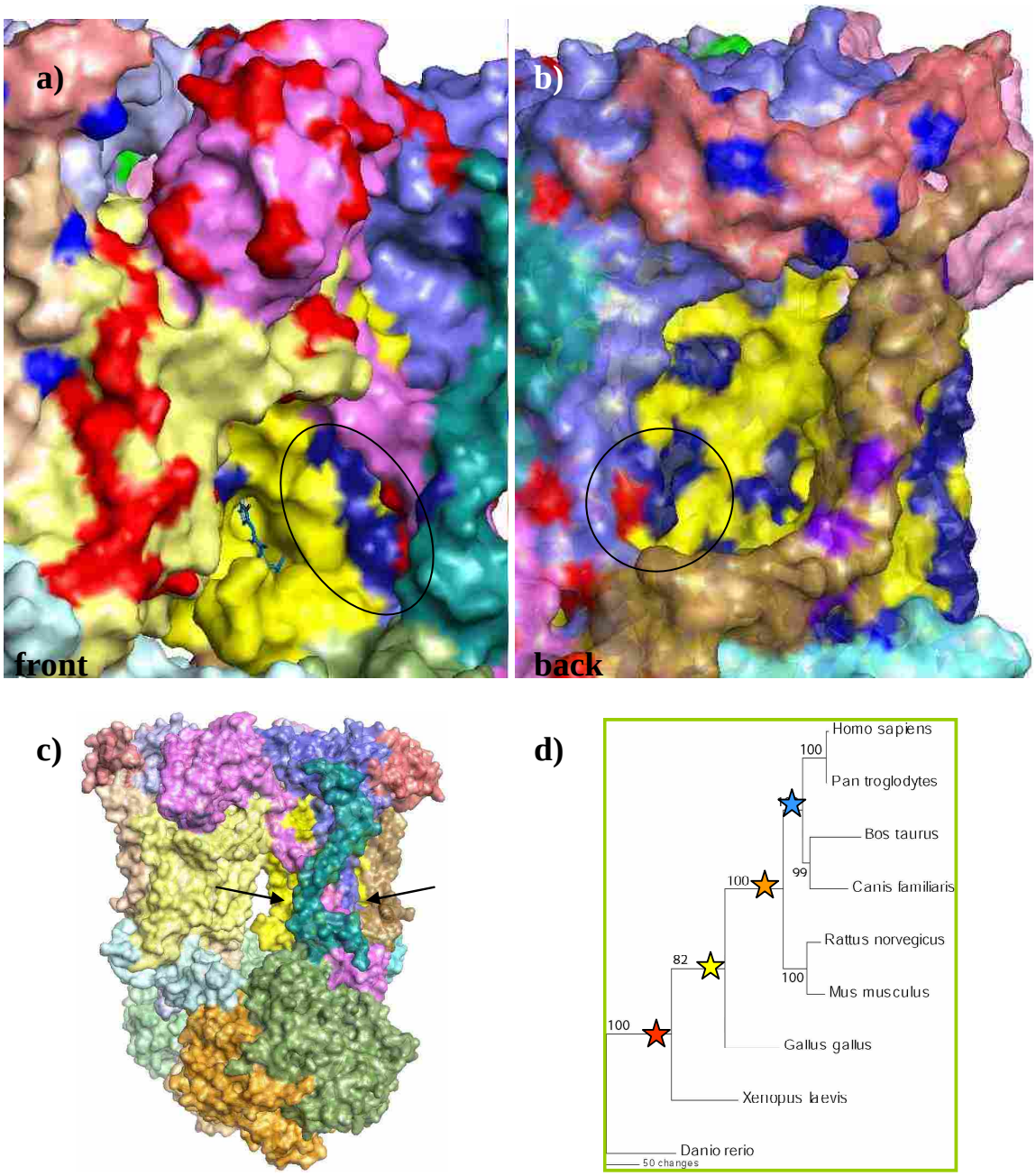


Figure 20: The selection on cyt *b* is shown in indigo, while the selection on ISP and cyt c_1 is shown in red. Selection on subunit 10 is shown in dark blue, while selection on subunit 7 is shown in purple. This view is of selection on helix E of cyt *b*, which is located adjacent to the Q_i site (shown with heme) and directly behind the transmembrane helices of cyt c_1, ISP and subunit 10. (a) Shows the front face of helix E, while (b) shows the back face of helix E. Selection on the front shows one example of coadaptation, I82L (cyt *b*) & I122M (cyt c_1) both increase equilibrium constant (red star). Also on the front side of this helix, changes from different amino acids of cyt *b* all to isoleucine cause shifts in different properties: 226I decrease in coil tendency, L229I decrease in equilibrium constant, T240I increase in buriedness, SARR and thermodynamic transfer hydrophobicity (yellow star). While the back side of this helix shows reversals of these changes from isoleucine to leucine: I235L and I234L increase in equilibrium constant (cyt *b*) (orange star), and I216L increase in equilibrium constant (cyt c_1) (blue star). (c) Indicates the location of helix E via the arrows. (d) Phylogenetic tree showing locations of changes discussed above.

changes behind the Q_i site, surrounding the heme. This site is near the matrix side, just below the QH_2 binding site (Fig. 16). Sites 107-110 of cyt *b* come in direct contact with subunit 6 and show an interesting pattern of coadaptation. On the phylogenetic branch separating *Gallus gallus* from mammals, there are three changes in cyt *b*: L108I, which decreases EC; K110I, which also decreases EC while increasing long range non-bonded energy, surrounding hydrophobicity, and total non-bonded energy; and Y109F, which decreases mean r.m.s. fluctuation displacement. There is also one change in subunit 6: D40N, which decreases power to be at the C-terminal. The coordinated shift in EC is an example of coadaptation. In this same region, but chronologically later in the phylogeny, on the branch leading to primates, there are several reversals: I108L, which increases EC; I110L, which also increases EC; F109Y, which increases mean r.m.s. fluctuation displacement; and N40D of subunit 6, which increases power to be at the C-terminal. These changes are an example of coadaptation because this major overall shift occurred in a punctuated manner on the branch between birds and mammals, and then reciprocal changes reversing the first changes occurred in an evolutionary burst on the primate lineage.

Another region on the Q_i site that shows clusters of coadaptation is where helix E of cyt *b* comes in contact with the helices of ISP and cyt c_1 (denoted the “front side” in Fig. 20). On the branch that separates the outgroup *Danio rerio* (zebrafish) from the in-group (amphibians, birds, and mammals) there are four sites showing radical shifts and two of them result in coadaptation. I82L on cyt *b* and I222M on cyt c_1 both increases EC while T43A on cyt *b* increases alpha-helical tendency and G47T on ISP increases bulkiness. On the next node that excludes *Xenopus laevis* (frog) from the in-group there

are four more adaptive changes, three of which change an amino acid to an isoleucine. In cyt *b*, N226I decreases coil tendency, L229I decreases EC, T240I increases buriedness, SARR, and thermodynamic transfer hydrophobicity, and on cyt c_1 Y221A increases compressibility. On the other side of cyt *b* helix E there is adaptive selection on the residues on the opposing surface of cyt *b*, just beyond the helices of subunit 10, ISP and cyt c_1 (Fig. 20). There are some changes in these amino acids that reverse these isoleucine changes, but not until much later in the phylogeny—on the branch separating cow, dog, human and chimp. In cyt *b*, I235L and I234L both increase EC. I216L on an adjoining region of cyt c_1 also increased EC just one branch earlier and included all mammals. Perhaps one side is decreasing EC by increasing isoleucine, and the other side is increasing EC by decreasing isoleucines.

Conclusions

Although there are many regions of the cytochrome bc_1 complex experiencing physicochemical shifts, there does seem to be a pattern to not only where on the protein these shifts are taking place, but what properties are being shifted. As predicted, those functional amino acid residues in the active sites are not undergoing positive destabilizing selection, but are highly conserved or undergoing negative selection. Those regions surrounding these residues in and around active sites are rapidly adapting. There are examples of both coevolution—changes in different amino acids of the same region to maintain the physicochemical environment with compensatory reciprocal shifts, and coadaptation—changes in a specific region to enact a coordinated increase or decrease in a particular physicochemical property. These changes are herein demonstrated to be immediately adjacent to both the Q_o and Q_i sites (Figs. 16 and 20), the ISP head region

(Fig. 17), and the pore region of cyt c_1 (Fig. 18). Although these adjacent regions are not directly influencing the binding of substrates necessary for electron transfer, they do influence the microenvironment of those active sites, and therefore the efficiency of the complex as a whole. The long 'stripe' of selected residues starting at the ISP dock and continuing through the inner mitochondrial membrane, may ultimately be shown to have some function in either the assembly of the structure or the interaction of helix H with the lipid membrane (Fig. 19).

Because this study is focused mainly on cyt *b*, cyt c_1, and ISP, those changes that were found in the membrane-spanning helices, and the intermembrane space were discussed in more detail than others. Most of the sites in these helices under selection were for the property equilibrium constant (EC). This supports the findings by McClellan et al. (2005) in their 2004 paper on cetacean cytochrome *b* adaptation. They found an abundance of changes in EC in the transmembrane helices of cyt *b*. These changes in EC are caused by substitutions between isoleucine, leucine and valine—all hydrophobic amino acids. An increase in this property means that the residue is more likely to ionize –COOH (Gromiha and Ponnuswamy 1993). Because they have largely the same properties (hydrocarbon chains) they can be interchanged keeping the alpha helices in tact and in the correct orientation in the membrane. This switching of Ile, Leu, and Val is only abundant in the helices imbedded in the membrane, and not the extra-membranous helices (like those of the core proteins and the heads of cyt c_1 and ISP). Additionally, when Ile, Leu, or Val changes to a threonine (Thr), EC is not changed, but the property solvent accessible reduction ratio (SARR) is. The property SARR is a ratio of surface area of one residue to the solvent (water) over the surface area of that residue in an extended

tripeptide (Ala-X-Ala) (Gromiha et al. 1999). A decrease in this property would therefore make the specific residue more accessible to water and an increase would make it less accessible to water. These definitions of the properties explain that when there was an increase in EC, there would also be a decrease in SARR—since the relationship between accessibility to water and proton release are closely related.

The majority of sites found to be experiencing physicochemical shifts were on the outer surface of the complex, not buried within. This may be because the external-facing residues are more subject to selection by interactions with molecules in the membrane, matrix and intermembrane space. The amino acid substitutions discussed in this paper were only those that were statistically found to be experiencing radical or positive-destabilizing selection. There were many other nonsynonymous substitutions that did not radically change any physicochemical property. Figure 21 shows those nonsynonymous changes that were not found to cause radical shifts in a property along with those nonsynonymous changes that did cause radical shifts in a particular property. From this view of the intermembrane side of the bc_1 complex it is evident that most of the nonsynonymous substitutions occurred in sites on the perimeter of subunit 8 and cytochrome c_1. While those sites that were radical are shown surrounding the pore of c_1 and the head regions of the ISP. The binding site of cyt *c* is largely conserved with only minimal substituted sites. This is an interesting method of identifying where on the protein only radical, verses conservative changes are taking place. The pore of c_1 and the head of ISP are radically adapting in comparison to those other regions on the intermembrane side of the bc_1 complex.

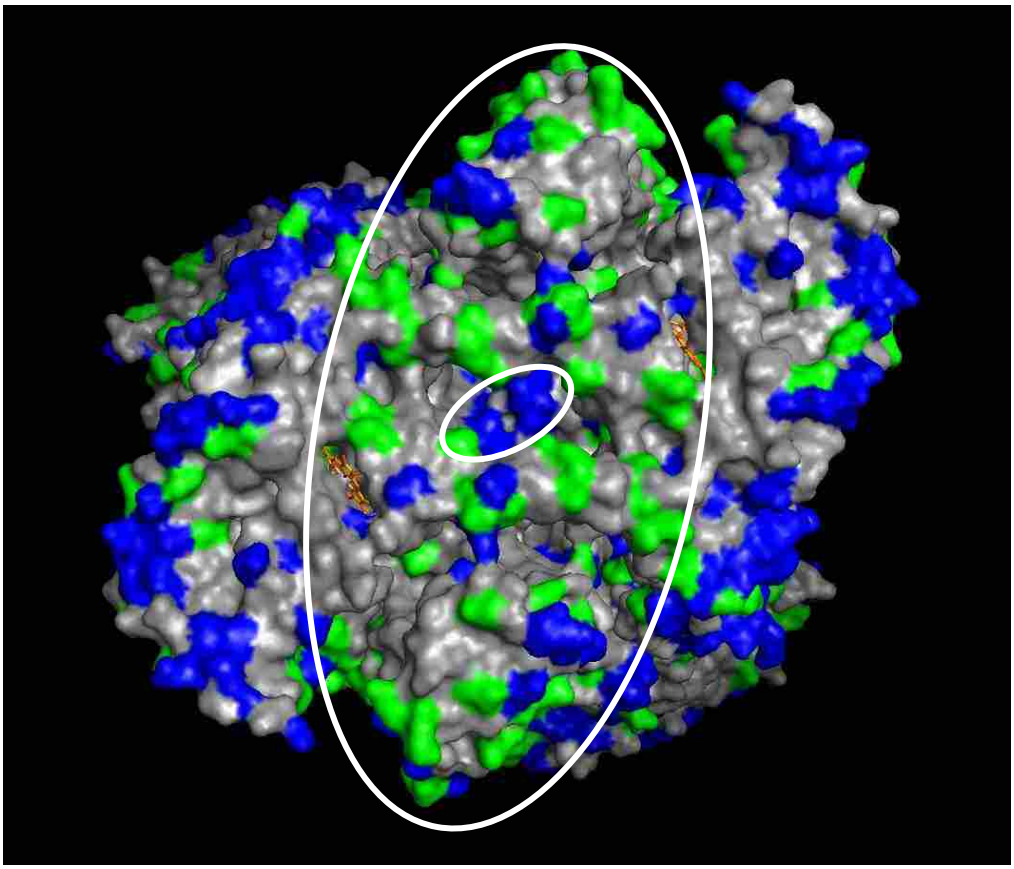


Figure 21: View from the intermembrane space looking down on the membrane. Cyt c_1, ISP, and subunit 8 are most readily visible. All subunits are colored gray to show the non-synonymous substitutions more clearly. Radical substitutions (positive-destabilizing selection) are shown in bright green, while conservative non-synonymous substitutions are shown in blue. You can see that the pore of c_1 and the head of ISP show mostly green (radical substitutions), while the peripheral edge of cyt c_1 and subunit 8 show mostly blue, or conservative substitutions. The cyt *c* docking region surrounding the heme (orange) is primarily gray, showing that this region is highly conserved relative to the rest of the protein in the intermembrane space.

This novel approach to characterizing molecular coadaptation and coevolution has demonstrated its usefulness in studying these complex protein complexes. Those regions of the protein that have been historically affected by selection can be studied not only by whether a change was synonymous or non-synonymous, but by the actual biochemical shift that was a result of a specific change. This cause and effect of amino acid substitution resulting in a biochemical shift has not been looked at in great detail until this study. Not only does the location of the changes on the folded protein and the phylogeny tell a story of a protein's evolution, but the different types of physicochemical properties influenced also shed light on the specific function and domain of the protein complex.

Acknowledgments

I thank my advisor, David A. McClellan for his time, energy, guidance, and expertise. I also thank my committee members Dixon Woodbury and David Belnap for suggestions and editing. I thank my labmates Rick Smith, Wes Beckstead, Srikar Chamala and Mark Ebbert for their comments, discussions, and computer expertise. I also thank Gavin Svenson, Nicole Lewis-Rogers, Carissa Jones, Dan Hardy, Martha Yoke, Scott Peat, and Stephen Cameron for their comments and feedback. I acknowledge the experts in the field Dr. Mark Rowe, Department of Nutrition, Dietetics and Food Science, BYU and Dr. Edward A. Berry, Lawrence Labs, Berkeley CA for sharing their knowledge. I also acknowledge the Department of Integrative Biology at BYU and Keith Crandall for funding and this opportunity.

References

1. Alff-Steinberger C. 1969. The genetic code and error transmission. PNAS. USA 64:584-591.
2. Anderson S, de Bruijn MH, Coulson AR, Eperon IC, Sanger F, Young, IG. 1982. Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome. J. Mol. Biol. 156 (4): 683-717.
3. Andreu AL, Checcarelli N, Iwata S, Shanske S, DiMauro S. 2000. A missense mutation in the mitochondrial cytochrome *b* gene in a revisited case with histiocytoid cardiomyopathy. Pediat. Res. 48: 311-314.
4. Baer KK and McClellan DA. 2006. Molecular coevolution of the vertebrate cytochrome c_1 and Rieske Iron Sulfur Protein in the cytochrome bc_1 complex. Proceedings of the Biotechnology and Bioinformatics Symposium. 3: 10-17.
5. Baymann F, Lebrun E, Nitschke W. 2004. Mitochondrial cytochrome c_1 is a collapsed di-heme cytochrome. PNAS. 101 (51): 17737-17740.
6. Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Herrnstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R, Skorecki K. 2006. The matrilineal ancestry of ashkenazi jewry: portrait of a recent founder event. Am. J. Hum. Genet. 78 (3): 487-497.
7. Berry EA, Huang LS, Zhang Z, Kim SH. 1999. Structure of the avian mitochondrial cytochrome bc_1 complex. J. Bioenergetics and Biomembranes. 31 (3): 177-190.
8. Berry EA, Guergova-Kuras M, Huang L-S, Crofts AR. 2000. Structure and function of cytochrome *bc* complexes. Annu. Rev. Biochem. 69: 1005-1075.
9. Bjornerfeldt S, Webster MT, Vila C. 2006. Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res., 16 (8): 990–994.
10. Brady KP, Rowe LB, Her H, Stevens TJ, Eppig J, Sussman DJ, Sikela J, Beier DR. 1997. Genetic mapping of 262 loci derived from expressed sequences in a murine interspecific cross using single-strand conformational polymorphism analysis. Genome Res. 7 (11): 1085-1093.
11. Brandt U, Yu L, Yu C-A, Trumpower BL. 1993. The mitochondrial targeting presequence of the Rieske iron-sulfur protein is processed in a single step after insertion into the cytochrome bc_1 complex in mammals and retained as a subunit in the complex. J. Biol. Chem. 268 (12): 8387-8390.
12. Broughton RE, Milam JE, Roe BA. 2001. The Complete Sequence of the Zebrafish (*Danio rerio*) Mitochondrial Genome and Evolutionary Patterns in Vertebrate Mitochondrial DNA. Genome Res. 11 (11): 1958–1967.
13. Brunga M, Rodgers S, Schricker A, Montoya G, Kazmeier M, Nitschke W, Sinning I. 2000. A spectroscopic method for observing the domain movement of the Rieske iron-sulfur protein. PNAS. 97 (5): 2069-2074.
14. Caldwell RB, Kierzek AM, Arakawa H, Bezzubov Y, Zaim J, Fiedler P, Kutter S, Blagodatski A, Kostovska D, Koter M, Plachy J, Carninci P, Hayashizaki Y, Buerstedde JM. 2005. Full-length cDNAs from chicken bursal lymphocytes to facilitate gene function analysis. Genome Biol. 6 (1): R6.
15. Chamala S, Beckstead WA, Rowe MJ, McClellan DA. 2006. Evolutionary selective pressure on three mitochondrial SNPs is consistent with their influence on metabolic efficiency in Pima Indians. Proceedings of the Biotechnology and Bioinformatics Symposium. 3: 35-44.
16. Charton M, Charton B. 1983. The dependence of the Chou-Fasman parameters on amino acid side chain structure. J. Theor. Biol. 111: 447-450.
17. Chimpanzee Sequencing and Analysis Consortium. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437 (7055): 69-87.
18. Chou PY, Fasman GD. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. J. Adv. Enzymol. 47: 45-148.
19. Crivellone MD, Wu M, Tzagoloff A. 1988. Assembly of the mitochondrial membrane system. Analysis of structural mutants of the yeast coenzyme QH_2-cytochrome *c* reductase complex. J. Biol. Chem. 263: 14323-14333.
20. Crofts AR. 2004. The cytochrome bc_1 complex: function in the context of structure. Annu. Rev. Physiol.. 66: 689–733.
21. Crofts AR, Hong S, Ugulava N, Barquera B, Gennis R, Guergova-Kuras M, Berry EA. 1999. Pathways for proton release during ubihydroquinone oxidation by the bc_1 complex. PNAS 96: 10021-10026.
22. Crofts AR, Hong S, Zhang Z, Berry EA. 1999. Physicochemical aspects of the movement of the Rieske iron-sulfur protein during quinol oxidation by the bc_1 complex from mitochondria and photosynthetic bacteria. Biochemistry. 38: 15827-15839.

23. Cruciat CM, Hell K, Folsch H, Neupert W, Stuart RA. 1999. Bsc1p, an AAA-family member, is a chaperone for the assembly of the cytochrome bc_1 complex. EMBO J. 18: 5226-5233.

24. Da Cruz S, Xenarios I, Langridge J, Vilbois F, Parone PA, Martinou JC. 2003. Proteomic analysis of the mouse liver mitochondrial inner membrane. J. Biol. Chem. 278 (42): 41566-41571.

25. Degli Esposti M, De Vries S, Crimi M, Ghelli A, Paternello T, Meyer A. 1993. Mitochondrial cytochrome *b*: evolution and structure of the protein. Biochem. Biophys. Acta. 1143: 243-271.

26. DeLano W. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, California.

27. DeLonlay P, Valnot I, Barrientos A, Gorbatyuk M, Tzagoloff A, Taanman J-W, Benayoun E, Chretien D, Kadhom N, Lombes A, Ogier de Baulny H, Niaudet P, Munnich A, Rustin P, Rotig A. 2001. A mutant mitochondrial respiratory chain assembly protein causes complex III deficiency in patients with tubulopathy, encephalopathy and liver failure. Nature Genet. 29: 57-60.

28. Deng K, Shenoy SK, Tso S-C, Yu L, Yu C-A. 2001. Reconstitution of mitochondrial processing peptidase from core proteins (subunits I and II) of bovine heart mitochondrial cytochrome bc_1 complex. J. Biol. Chem. 276 (9): 6499-6505.

29. Desjardins P, Morais R. 1990. Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. J. Mol. Biol. 212 (4): 599-634.

30. Doan JW, Schmidt TR, Wildman DE, Goodman M, Weiss ML, Grossman LI. 2005. Rapid nonsynonymous evolution of the iron-sulfur protein in anthropoid primates. J Bioenerg Biomembr. 37: (1): 35–41.

31. Esser L, Gong X, Yang S, Yu L, Yu C-A, Xia D. 2006. Surface-modulated motion switch: Capture and release of iron-sulfur protein in the cytochrome bc_1 complex. PNAS. 103 (35): 13045-13050.

32. Gao X, Wen X, Esser L, Quinn B, Yu L, Yu C-A, Xia D. 2003. Structural basis for the quinone reduction in the bc_1 complex: a comparative analysis of crystal structures of mitochondrial cytochrome bc_1 with bound substrate and inhibitors at the Q_i site. Biochemistry. 42: 9067-9080.

33. Gatti DL, Tzagologg A. 1990. Structure and function of the mitochondrial bc_1 complex. J. Biol. Chem. 265 (35): 21468-21475.

34. Gencic S, Schagger H, von Jagow G. 1991. Core I protein of bovine ubiquinol-cytochrome-*c* reductase; an additional member of the mitochondrial-protein-processing family. Cloning of bovine core I and core II cDNAs and primary structure of the proteins. Eur. J. Biochem. 199 (1): 123-131.

35. Grantham R. 1974. Amino acid difference formula to help explain protein evolution. Science 185:862-864.

36. Grigorieff N. 1998. Three-dimensional structure of bovine NADH:ubiquinone oxidoreductase (complex I) at 22 Å in ice. J. Mol. Biol. 277 (5): 1033-1046.

37. Gromiha MM, Ponnuswamy PK. 1993. Relationship between amino acid properties and protein compressibility. J. Theor. Biol. 165: 87-100.

38. Gromiha MM, Oobatake M, Kono H, Uedaira H, Sarai A. 1999. Role of structural and sequence information in the prediction of protein stability changes: comparison between buried and partially buried mutations. Protein Engineering. 12(7): 549-555.

39. Grosskopf R, Feldmann H. 1981. Analysis of a DNA segment from rat liver mitochondria containing the genes for the cytochrome oxidase subunits i, ii, iii, atpase subunit 6, and several tRNA genes. Curr. Genet. 4: 151-158.

40. Gurung B, Yu L, Xia D, Yu C-A. 2005. The iron-sulfur cluster of the rieske iron-sulfur protein functions as a proton-exiting gate in the cytochrome bc_1 complex. J. Biol. Chem. 280 (26): 24895-24902.

41. Haut S, Brivet M, Touati G, Rustin P, Lebon S, Garcia-Cazorla A, Saudubray JM, Boutron A, Legrand A, Slama A. 2003. A deletion in the human QP-C gene causes a complex III deficiency resulting in hypoglycaemia and lactic acidosis. Hum. Genet. 113: 118-122.

42. Hixson JE, Brown WM. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications Mol. Biol. Evol. 3 (1): 1-18.

43. Hoffman GG, Lee S, Christiano AM, Chung-Honet LC, Cheng W, Katchman S, Uitto J, Greenspan DS. 1993. Complete coding sequence, intron/exon organization, and chromosomal location of the gene for the core I protein of human ubiquinol-cytochrome *c* reductase. J. Biol. Chem. 268: 21113-21119.

44. Huang LS, Cobessi D, Tung EY, Berry EA. 2005. Binding of the Respiratory Chain Inhibitor Antimycin to the Mitochondrial bc(1) Complex: A New Crystal Structure Reveals an Altered Intramolecular Hydrogen-bonding Pattern. J. Mol. Biol. 351: 573-597.

45. Hubbard SJ, Grafham DV, Beattie KJ, Overton IM, McLaren SR, Croning MDR, Boardman PE, Bonfield JK, Burnside J, Davies RM, Farrell ER, Francis MD, Griffiths-Jones S, Humphray SJ, Hyland C, Scott CE, Tang H, Taylor RG, Tickle C, Brown WRA, Birney E, Rogers J, and Wilson SA. 2005. Transcriptome analysis for

the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags. Genome Res. 15: 174–183.

46. Hunte C, Koepke J, Lange C, Roßmanith T, Michel H. 2000. Structure at 2.3 Åresolution of the cytochrome bc_1 complex from the yeast Saccharomyces cerevisiae co-crystallized with an antibody Fv fragment. Structure 2000 (8): 669-684.
47. International chicken genome sequencing consortium. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution, Nature, 432 (7018) 695–716.
48. Iwata S, Lee JW, Okada K, Lee JK, Iwata M, Rasmussen B, Link TA, Ramaswamy S, Jap BK. 1998. Complete structure of the 11-subunit bovine mitochondrial cytochrome bc_1 complex. Science 281:64-71.
49. Izrailev S, Crofts AR, Berry EA, Schulten K. 1999. Steered Molecular Dynamics Simulation of the Rieske Subunit Motion in the Cytochrome bc_1 Complex. Biophys. J. 77: 1753-1768.
50. Keightley JA, Anitori R, Burton MD, Quan F, Buist NRM, Kennaway NG. 2000. Mitochondrial encephalomyopathy and complex III deficiency associated with a stop-codon mutation in the cytochrome *b* gene. Am. J. Hum. Genet. 67: 1400-1410.
51. Kim H, Xia D, Yu C-A, Xia J-Z, Kachurin AM, Zhang L, Yu L, Deisenhofer J. 1998. Inhibitor binding changes domain mobility in the iron-sulfur protein of the mitochondrial bc_1 complex from bovine heart. Biophysics. 95: 8026-8033.
52. Klein SL, Strausberg RL, Wagner L, Pontius J, Clifton SW, Richardson P. 2002. Genetic and genomic tools for *Xenopus* research: The NIH *Xenopus* initiative. Dev. Dyn. 225 (4): 384-391.
53. Kumar S, Tamura K, Nei M. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput. Appl. Biosci. 10: 189-191.
54. Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. J. Molecular Biology. 157:105-132.
55. Lange C, Hunte C. 2002. Crystal structure of the yeast cytochrome bc_1 complex with its bound substrate cytochrome *c*. PNAS. 99: 2800-2805.
56. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DM, Kamal M, Clamp M, Chang JL, Kulbokas III EJ, Zody MC, Mauceli E, Xie X, Breen M, Wayne RK, Ostrander A, Ponting CP, Gailbert F, Smith DR. deJong PJ, Kirkness E, Alvarez P, Biagi T, Brockman W, Butler J, Chin C, Cook A, Cuff J, Daly MJ, DeCapriol D, Gnerre S, Grabherr M, Kellis M, Kleber M, Bardeleben C, Goodstadt L, Heger A, Hitte C, Kim L, Koepfli K, Parker HG, Pollinger JP, Searle SMJ, Sutter NB, Thomas R, Webber C, Lander ES. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438 (7069): 803-819.
57. Lundrigan BL, Jansa SA, Tucker PK. 2002. Phylogenetic relationships in the genus *mus*, based on paternally, maternally, and biparentally inherited characters. Syst. Biol. 51 (3) 410-431.
58. Maddison DR, Maddison WP. 2000. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Massachusetts. 492 pp book (as PDF file) + computer program.
59. Mammalian gene collection program team. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. PNAS. 99 (26): 16899–16903.
60. Marx S, Baumgartner M, Kunnan S, Braun H-P, Land BF, Burger G. 2003. Structure of the bc_1 complex from *Seculamonas ecuadoriensis*, a Jakobid flagellate with an ancestral mitochondrial genome. Mol. Biol. Evol. 20 (1): 145-153.
61. McClellan DA, and McCracken KG. 2001. Estimating the influence of selection on the variable amino acid sites of the cytochrome *b* protein functional domains. Mol. Biol. Evol. 18: 917–925.
62. McClellan DA, Moss JL, Palfreyman EJ, Smith MJ, Christensen RG, Sailsbery JK. 2005. Physicochemical evolution and molecular adaptation of the cetacean cytochrome *b* protein. Mol. Biol. Evol. 22: 437-455.
63. Moore S, Alexander L, Brownstein M, Guan L, Lobo S, Meng Y, Tanaguchi M, Wang Z, Yu J, Prange C, Schreiber K, Shenmen C, Wagner L, Bala M, Barbazuk S, Barber S, Babakaiff R, Beland J, Chun E, Del Rio L, Gibson S, Hanson R, Kirkpatrick R, Liu J, Matsuo C, Mayo M, Santos RR, Stott J, Tsai M, Wong D, Siddiqui A, Holt R, Jones SJ, Marra MA. 2005. Genbank: Direct Submission. BC Cancer Agency, Canada's Michael Smith Genome Sciences Centre, Suite 100, 570 West 7th Avenue, Vancouver, British Columbia V5Z 4S6, Canada.
64. Nishikimi M, Hosokawa Y, Toda H, Suzuki H, Ozawa T. 1989. Cloning and sequence analysis of a cDNA encoding the Rieske iron-sulfur protein of rat mitochondrial cytochrome bc_1 complex. Biochem. Biophys. Res. Commun. 159 (1): 19-25.

65. Okura T, Koda M, Ando F, Niino N, Tanaka M, Shimokata H. 2003. Association of the mitochondrial DNA 15497G/A polymorphism with obesity in a middle-aged and elderly Japanese population. Hum. Genet. 113: 432-436.

66. Oobatake M, Ooi T. 1977. An analysis of non-bonded energy of proteins. J. Theoretical Biology 67:567-584.

67. Petersen KF, Befroy D, Dufour S, Dziura J, Ariyan C, Rothman DL, DiPietro L, Cline GW, Shulman GI. 2003. Mitochondrial dysfunction in the elderly: possible role in insulin resistance. Science. 300 (5622): 1140-1142.

68. Polyak K, Li Y, Zhu H, Lengauer C, Willson JKV, Markowitz SD, Trush MA, Kinzler KW, Vogelstein B. 1998. Somatic mutations of the mitochondrial genome in human colorectal tumours. Nature Genet. 20: 291-293.

69. Ponnuswamy PK, Prabhakaran M, Manavalan P. 1980. Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins. J. Biochim. Biophys. Acta. 623:301-316.

70. Rana M, de Coo I, Diaz F, Smeets H, Moraes CT. 2000. An out-of-frame cytochrome *b* gene deletion from a patient with parkinsonism is associated with impaired complex III assembly and an increase in free radical production. Ann. Neurol. 48: 774-781.

71. Rat Genome Sequencing Project Consortium. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428 (6982): 493-521.

72. Riordan-Eva P, Harding AE. 1995. Leber's hereditary optic neuropathy: the clinical relevance of different mitochondrial DNA mutations. J. Med. Genet. 32: 81-87.

73. Sadakata T, Furuichi T. 2006. Identification and mRNA expression of Ogdh, QP-C, and two predicted genes in the postnatal mouse brain. Neurosci. Lett. 405 (3): 217-222.

74. Stonehuerner J, O'Brien P, Geren L, Millett F, Steidl J, Yu L, and Yu C-A. 1985. Identification of the Binding Site on Cytochrome c_1 for Cytochrome *c*. Journal of Biological Chemistry. 260 (9): 5392-5398.

75. Swofford DL. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

76. Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through weighting, position-specific gap penalties and weight matrix choice. Nuc. Acids Res. 22: 4673-4680.

77. Trifunovic A, Wredenberg A, Falkenberg M, Spelbrink JN, Rovio AT, Bruder CE, Bohlooly-Y M, Gidlof S, Oldfors A, Tornell J, Jacobs HT, Larsson NG. 2004. Premature ageing in mice expressing defective mitochrondrial DNA polymerase. Nature. 429 (6990): 357-359.

78. Usui S, Yu L, Yu C-A. 1990. Cloning and sequencing of a cDNA encoding the Rieske iron-sulfur protein of bovine heart mitochondrial ubiquinol-cytochrome *c* reductase. Biochem. Biophys. Res. Commun. 167 (2): 575-579.

79. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA. et al. (274 authors). 2001. Celera genomics. The sequence of the human genome. Science 291: 1304-1351.

80. Visappa I, Fellman V, Varilo T, Palotie A, Raivio KO, Peltonen L. 1998. Assignment of the locus for a new lethal neonatal metabolic syndrome to 2q33-37. Am. J. Hum. Genet. 63: 1396-1403.

81. Wibrand F, Ravn K, Schwartz M, Rosenberg T, Horn N, Vissing J. 2001. Multisystem disorder associated with a missense mutation in the mitochondrial cytochrome *b* gene. Ann. Neurol. 50: 540-543.

82. Warburg O. 1956. On the origin of cancer cells. Science 123: 309-314.

83. Woese CR, Dugre DH, Dugre SA, Kondo K, and Saxinger WC. 1966. On the fundamental nature and evolution of the genetic code. Cold Spring Harb. Symp. Quant. Biol. 403:304-308.

84. Wong JF, Ma DP, Wilson RK, Roe BA. 1983. DNA sequence of the *Xenopus laevis* mitochondrial heavy and light. Nucleic Acids Res. 11 (14): 4977-4995.

85. Woolley S, Johnson J, Smith MJ, Crandall KA, and McClellan DA. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19: 671–672.

86. Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13: 555-556.

87. Zhang Z, Huang L, Shulmeister VM, Chi YI, Kim KK, Hung LW, Crofts AR, Berry EA, Kim SH. 1998. Electron transfer by domain movement in cytochrome bc_1. Nature 392: 677-84.

Appendix 1: Sequence accession numbers and references for all sequences used in this study.

Common Name	Genus species	Gene	GenBank #	Reference
Frog	*Xenopus laevis*	cytochrome *b*	M10217	Wong et al. 1983
		cytochrome c_1	BC045127	Klein et al. 2002
		core protein I	BC098177	Klein et al. 2002
		core protein II	BC106252	Klein et al. 2002
		ISP, subunit 9	BC041528	Klein et al. 2002
		subunit 6	BC078487	Klein et al. 2002
		subunit 7	BC110732	Klein et al. 2002
Bovine	*Bos taurus*	cytochrome *b*	V00654	Anderson et al. 1982
		cytochrome c_1	BC109917	Moore et al. 2005
		core protein I	NM_174629	Gencic et al. 1991
		core protein II	BC102337	Moore et al. 2005
		ISP, subunit 9	NM_174813	Usui et al. 1990
		subunit 6	BC103057	Moore et al. 2005
		subunit 7	BC103110	Moore et al. 2005
		subunit 8	BC102729	Moore et al. 2005
		subunit 10	BC102720	Moore et al. 2005
Chicken	*Gallus gallus*	cytochrome *b*	X52392	Desjardins and Morais, 1990
		core protein I	XM_414356	Intl. Chicken Genome Seq. Consortium 2004
		core protein II	BX950415	Hubbard et al. 2005
		ISP, subunit 9	AJ719669	Caldwell et al. 2005
		subunit 6	BX934091	Hubbard et al. 2005
		subunit 7	CR405815	Hubbard et al. 2005
		subunit 8	CR385442	Hubbard et al. 2005
		subunit 10	BX935769	Hubbard et al. 2005
Chimp	*Pan troglodytes*	cytochrome *b*	D38113	Hixson & Brown, 1986
		cytochrome c_1	XM_528263	Chimp Seq. Consortium, 2005
		core protein I	XM_516440	Chimp Seq. Consortium, 2005
		core protein II	XM_523485	Chimp Seq. Consortium, 2005
		ISP, subunit 9	AY387497S2	Doan et al. 2005
		subunit 7	XM_527009	Chimp Seq. Consortium, 2005
		subunit 8	XM_525881	Chimp Seq. Consortium, 2005
		subunit 10	XM_525558	Chimp Seq. Consortium, 2005
Dog	*Canis familiaris*	cytochrome *b*	DQ480496	Bjornerfeldt et al. 2006
		cytochrome c_1	XM_532351	Lindblad-Toh, K. et al. 2005
		core protein I	XM_846116	Lindblad-Toh, K. et al. 2005
		core protein II	XM_536942	Lindblad-Toh, K. et al. 2005

		ISP, subunit 9	XM_533711	Lindblad-Toh, K. et al. 2005
		subunit 6	XM_535132	Lindblad-Toh, K. et al. 2005
		subunit 7	XM_847071	Lindblad-Toh, K. et al. 2005
		subunit 8	XM_845161	Lindblad-Toh, K. et al. 2005
		subunit 10	XM_847309	Lindblad-Toh, K. et al. 2005
Mouse	*Mus musculus*	cytochrome *b*	AY057804	Lundrigan et al. 2002
		cytochrome c_1	NM_025567	Da Cruz et al. 2003
		core protein I	NM_025407	Brady et al. 1997
		core protein II	BC003423	Mammalian gene collection program team 2002
		ISP, subunit 9	NM_025710	Da Cruz et al. 2003
		subunit 6	NM_026219	Sadakata & Furuichi, 2006
		subunit 7	NM_025352	Sadakata & Furuichi, 2006
		subunit 8	BC011388	Mammalian gene collection program team 2002
		subunit 10	BC024518	Mammalian gene collection program team 2002
Human	*Homo sapiens*	cytochrome *b*	DQ301818	Behar et al. 2006
		cytochrome c_1	BC001006	Mammalian gene collection program team 2002
		core protein I	BC009586	Mammalian gene collection program team 2002
		core protein II	BC003136	Mammalian gene collection program team 2002
		ISP, subunit 9	BC021057	Mammalian gene collection program team 2002
		subunit 6	BC005230	Mammalian gene collection program team 2002
		subunit 7	BC090048	Mammalian gene collection program team 2002
		subunit 8	BC001426	Mammalian gene collection program team 2002
		subunit 10	BC015971	Mammalian gene collection program team 2002
Rat	*Rattus norvegensis*	cytochrome *b*	X14848	Grosskopf and Feldmann, 1981
		cytochrome c_1	XM_001072221	Venter et al. 2001
		core protein I	BC078923	Mammalian gene collection program team 2002
		core protein II	BC083610	Mammalian gene collection program team 2002
		ISP, subunit 9	NM_001008888	Nishikimi et al. 1989
		subunit 6	XM_343225	Rat Genome Sequencing Project Consortium 2004
		subunit 7	AY323237	Rat Genome Sequencing Project Consortium 2004
		subunit 8	BC086954	Mammalian gene collection program team 2002
		subunit 10	XM_001067698	Rat Genome Sequencing Project Consortium 2004
Zebrafish	*Danio rerio*	cytochrome *b*	AC024175	Broughton et al. 2001
		cytochrome c_1	BC080245	Mammalian gene collection program team 2002
		core protein II	BC097011	Mammalian gene collection program team 2002
		ISP, subunit 9	BC059475	Mammalian gene collection program team 2002
		subunit 6	BC095035	Mammalian gene collection program team 2002
		subunit 7	BC076310	Mammalian gene collection program team 2002
		subunit 8	NM_001037108	Mammalian gene collection program team 2002

Appendix 2: List of physicochemical properties that come with TreeSAAP with descriptions and citations.

Property	Category	Approximate Description	Citation
alpha-helical tendency	conformational	the tendency of a residue to be part of an alpha helical region of a protein	Gromiha & Ponnuswamy 1993
bulkiness	structural	a measure of the amount of three-dimensional space a residue resides within	Zimmerman et al. 1968
buriedness	structural	the tendency of a residue to reside within the interior of a protein	Gromiha & Ponnuswamy 1993
coil tendency	conformational	the tendency of a residue to be part of a coiled region of a protein	Charton & Charton 1983
compressibility	structural	the contribution of a residue to the local density of protein secondary structures	Gromiha & Ponnuswamy 1993
equilibrium constant (ionization of COOH)	functional	the ability of the carboxyl group to give up an additional proton	Gromiha & Ponnuswamy 1993
isoelectric point	energetic	a measure of the pH of the residue at the point that the residues do not move in an electric field	Alff-Steinberger 1969
long-range non-bonded energy	energetic	the tendency of a residue to be part of interactions involving van der Waals forces	Oobatake & Ooi 1977
mean r.m.s. fluctuation displacement	structural	the ability of a residue to change position in three-dimensional space	Gromiha & Ponnuswamy 1993
polar requirement	hydrophobicity	a measure derived from the paper chromatographic mobility of a residue in a pyridine-water mixture	Woese et al. 1966
power to be at the C-terminal	energetic	the energy potential of the C-terminus of an alpha helix to interact with other residues	Gromiha & Ponnuswamy 1993
power to be at the N-terminal	energetic	the energy potential of the N-terminus of an alpha helix to interact with other residues	Gromiha & Ponnuswamy 1993
short-range & medium-range nonbonded energy	energetic	the tendency of a residue to be part of interactions involving repulsion and electrostatic forces	Oobatake & Ooi 1977
solvent accessible reduction ratio	functional	the reduction ratio of cross peak intensity when residues are/are not irradiated with aliphatic protons	Ponnuswamy et al. 1980
surrounding hydrophobicity	hydrophobicity	a measure of the reactivity of the reside to the surrounding solvent	Gromiha & Ponnuswamy 1993
total nonbonded energy	energetic	the tendency of a residue to be part of any non-bonded interaction	Oobatake & Ooi 1977
turn tendency	conformational	the tendency of a residue to be part of a protein turn	Chou & Fasman 1978

In Press: *International Journal of Bioinformatics Research and Applications*
Submitted: November 2006
In Print: August 2007, vol. 3, no. 3

Molecular Coevolution of the Vertebrate Cytochrome c_1 and Rieske Iron Sulfur Protein in the Cytochrome bc_1 Complex

By Kimberly K. Baer and David A. McClellan

Molecular coevolution of the vertebrate cytochrome c_1 and Rieske Iron Sulfur Protein in the cytochrome bc_1 complex

Kimberly K. Baer and David A. McClellan*

Department of Integrative Biology,
Brigham Young University, Provo, UT 84602, USA
E-mail: kimberly.baer@gmail.com
E-mail: david_mcclellan@byu.edu
*Corresponding author

Abstract: Cytochrome c_1 (cyt-c_1) and the Rieske Iron Sulfur Protein (ISP) are subunits of the cytochrome bc_1 – complex 3 in the electron transport chain in mitochondria functioning both as a proton pump and an electron transporter. Vertebrate model organism phylogenies were used in conjunction with existing 3D protein structures to evaluate the biochemical evolution of cyt-c_1 and ISP in terms of selection on amino acid properties. We found selection acting on the exterior surfaces of both proteins and specifically the core region of cyt-c_1. There is evidence supporting coevolution of these proteins relative to alpha helical tendencies, compressibility and equilibrium constant.

Keywords: molecular coevolution; cytochrome c_1; Rieske iron sulfur protein; cytochrome bc_1 complex; biochemical adaptation.

Reference to this paper should be made as follows: Baer, K.K. and McClellan, D.A. (xxxx) 'Molecular coevolution of the vertebrate cytochrome c_1 and Rieske Iron Sulfur Protein in the cytochrome bc_1 complex', *Int. J. Bioinformatics Research and Applications*, Vol. x, No. x, pp.xxx–xxx.

Bibliographical notes: Kimberly K. Baer is a Master of Science student in the Integrative Biology Department at Brigham Young University. She earned her Bachelor of Science Degree in Biology at Brigham Young University. Her research interests include protein 3D structure mapping, biochemical adaptation and protein physiology.

David A. McClellan is an Assistant Professor in the Department of Integrative Biology, Brigham Young University. His research focuses on methods for detecting molecular adaptation and identifying the physicochemical amino acid properties affected by natural selection. He and his students developed the TreeSAAP software that automates such analyses, as well as several other software tools. He also is interested in diagnosis of single nucleotide polymorphisms, evolution of genetic code structure, patterns of nucleotide variation and phylogenetic methods.

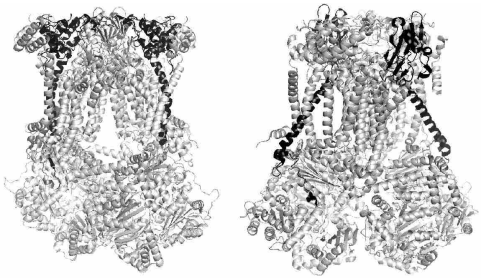
1 Introduction

The cytochrome bc_1 complex (also known as complex III or ubiquinol: cytochrome c oxidoreductase) is made up of 11 different protein subunits. These include the

well-studied cytochrome *b* (cyt-*b*), cytochrome c_1 (cyt-c_1), the Rieske ISP and eight others. These proteins span the inner mitochondrial membrane and have functional domains that work in the matrix, membrane and intermembrane space of the mitochondria.

This study represents a newly developed method for characterising protein-protein interactions. As an example of the method, we evaluate historical evolutionary biochemical interactions between cyt-c_1 and ISP within the bc_1 complex (Figure 1).

Figure 1 Location of cytochrome c_1 (black, left) and Rieske ISP (black, right) dimers in cytochrome bc_1 complex (grey)



The exploration of new methods by which historical and potentially important protein interactions can be described is vital to the study of genetic disease, infection and cancer. With this study we are proposing a new method for accomplishing this goal.

2 Methods

Sampling

All vertebrate cyt-c_1 and ISP gene sequences were obtained from GenBank for the *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Bos taurus* (cow), *Canis familiaris* (dog), *Rattus norvegensis* (Norway rat), *Mus musculus* (house mouse), *Gallus gallus* (chicken), *Xenopus laevis* (African clawed frog) and *Danio rerio* (zebrafish). Five protein-coding gene sequences were used to estimate the cytochrome bc_1 phylogeny for a total of 5801 bp: cyt-c_1 – 981 bp; ISP – 825 bp; cyt-*b* – 1148 bp; core 1 – 1446 bp; core 2 – 1401 bp (Wong et al., 1983; Klein et al., 2002; Anderson et al., 1982; Moore et al., Gencic et al., 1991; Usui et al., 1990; Desjardins and Morais, 1990; International Chicken Genome Sequencing Consortium, 2004; Hubbard et al., 2005; Caldwell et al., 2005; Hixson and Brown, 1986; Chimpanzee Sequencing and Analysis Consortium, 2005; Doan et al., 2003; Bjornerfeldt et al., 2006; Lindblad-Toh et al., 2005; Lundrigan et al., 2002; Da Cruz et al., 2003; Brady et al., 1997; Mammalian gene collection program team, 2002; Behar et al., 2006; Grosskopf and Feldmann, 1981; Venter et al., 2001; Nishikimi et al., 1989, Broughton et al., 2001). See Appendix A.

Phylogeny reconstruction

Protein-coding DNA sequences were translated into amino acid data in MEGA3 (Kumar et al., 1994) and aligned using ClustalW (Thompson et al., 1994). The resulting protein alignments were then used as templates for aligning the protein-coding DNA data. After alignment, the sequences were concatenated using MacClade (Maddison and

Maddison, 2000). PAUP* (Swofford, 2003) was used to estimate a phylogeny for the cytochrome bc_1 complex using a parsimony optimality criterion. A heuristic search algorithm was employed, with 1,000 random addition replicates and TBR branch swapping. A strict consensus of all most parsimonious trees was constructed for use in subsequent analyses.

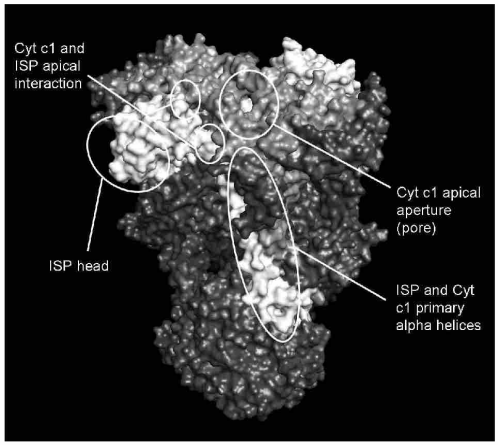
Biochemical adaptation analysis

The software package TreeSAAP (Woolley et al., 2003) evaluates protein-coding nucleotide sequences using a given phylogenetic topology to test for positive selection on quantitative physicochemical amino acid properties. This program identifies the properties historically affected by selection and locates the exact amino acid sites that experienced significant biochemical or conformational shifts. The baseml algorithm (Yang, 1997) is used to reconstruct ancestral character states at the nodes on a given phylogeny and then the program evaluates the average influence of selection on amino acid properties for a variable number of magnitude categories. For this study eight categories of selection were used and only categories 6–8 were considered because they are unambiguously associated with molecular adaptation (McClellan and McCracken, 2001; McClellan et al., 2005). Cyt-c_1 and ISP were analysed in TreeSAAP independently. The results were further analysed using a sliding window to statistically analyse the data for significant clustering of radical physicochemical shifts resulting from naturally occurring amino acid replacements. A Bonferroni correction for multiple sampling was employed. All results are relative to human sequences. Software without citations can be found at: http://inbio.byu.edu/faculty/dam83/cdm/.

Visualisation

Regions under heavy selection were correlated to pre-existing three-dimensional structures of the bc_1 complex (PDB A206) using the program Pymol (DeLano, 2002). In addition, Pymol was used to annotate the specific amino acids that were found to have undergone radical biochemical shifts. This technique allows us to visualise the influences of selection by differentiating those regions historically influenced by selection on specific amino acid properties on functional protein domains (Figure 2).

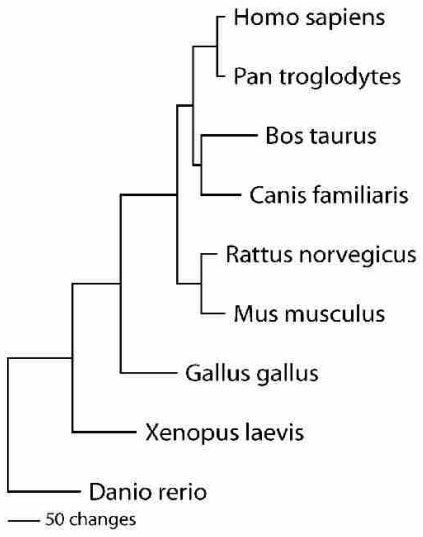
Figure 2 Functional domains of cyt-c_1 (grey) and ISP (white) in the broader context of the 3-dimensional structure of the cytochrome bc_1 protein complex



3 Results

The most parsimonious tree resulted in the phylogeny illustrated in Figure 3.

Figure 3 Parsimony tree for the cytochrome bc_1 gene sequences used in this study



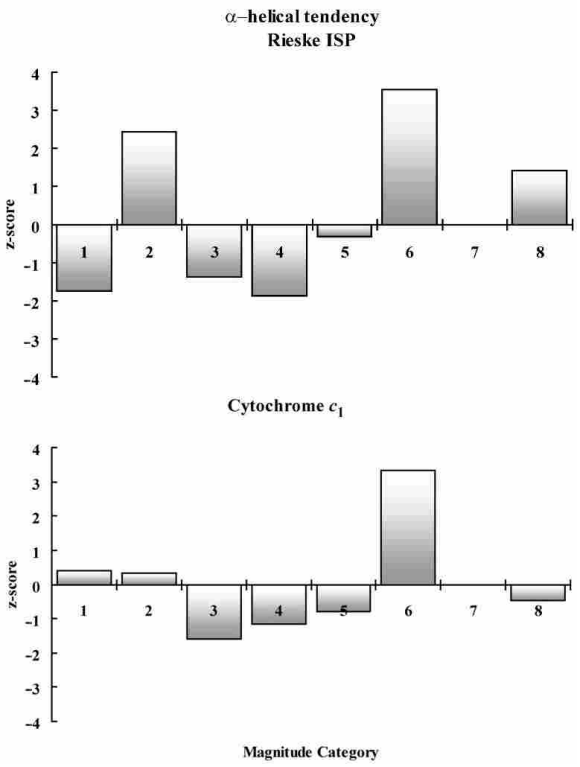
We recovered 187 historical amino acid replacements from the cyt-c_1 data and 279 amino acid replacements from the ISP data. TreeSAAP analysis of these changes resulted in the detection of the influence of selection on several physicochemical amino acid properties, which clustered in three regions of both proteins, with amino acid replacements in these regions being affected by multiple amino acid properties.

Detected regions correspond to residues on the outer surface of each protein, including surfaces on these proteins that are in direct contact. Table 1 shows the amino acid properties under positive-destabilising selection and domain locations of each specific property. For example, α-helical tendency was influenced by selection at four sites in cyt-c_1 and three sites in ISP. Decreases in α-helical tendency broadens the outer surface area of a helix and allows a more flexible surface area contact with neighbouring protein constituents, while increases have the opposite effect (Chou and Fasman, 1987). Both cyt-c_1 and ISP exhibit significant positive-destabilising selection for moderately radical changes (category 6) in this property (Figure 4). In ISP, there are radical shifts in α-helical tendency at the apical interaction between cyt-c_1 and ISP as a result of mutations that changed both a threonine and a serine to alanine (Figure 5). Cyt-c_1 shows moderately radical shifts in α-helical tendency resulting from mutations that changed an alanine to glycine, a threonine to serine and a serine to alanine. These replacements resulted in reciprocal α-helical structural changes (increases in the property in ISP and corresponding decreases in cyt-c_1) in the region surrounding the apical aperture of the complex, likely resulting in a coevolutionary modification of overall pore conformation.

Table 1 Amino acid properties and sites identified as being under positive-destabilising selection in cyt-c_1 and ISP protesins

Property	*cyt-c_1*	*Domain*	*ISP*	*Domain*	*Property*	*cyt-c_1*	*Domain*	*ISP*	*Domain*
α-helical tendency	24	helix	72	apical	isoelectric point	–	–	116	head
	38	pore	79	apical		–	–	128	head
	39	pore	88	apical		–	–	170	head
	190	interior	–	–	lng-range nb energy	–	–	124	head
average # surround res	71	pore	173	head	mean rms fluc displ	88	pore	–	–
bulkiness	–	–	47	helix	partial specific vol	–	–	170	head
buriedness	–	–	124	head	polar requirement	98	pore	12	helix
	–	–	173	head		–	–	27	helix
chromatograph index	–	–	173	head		–	–	31	helix
coil tendency	38	pore	24	helix	pwr to be C-terminal	75	apical	107	head
	39	pore	–	–		77	apical	113	head
	194	interior	–	–		166	exterior	131	head
composition	187	interior	–	–	pwr to be N-terminal	96	pore	16	helix
compressibility	59	pore	16	helix		98	pore	–	–
	63	pore	24	helix		187	interior	–	–
	66	pore	103	head	refractive index	187	interior	–	–
	221	helix	–	–	sh & med range nb energy	–	–	104	head
equilibrium constant	68	pore	5	helix		–	–	170	head
	70	pore	7	helix	solvent accessible red	71	pore	124	head
	82	pore	124	head		82	pore	–	–
	127	pore	192	apical	surrounding hydrophobicity	71	pore	124	head
	211	helix	194	apical		82	pore	173	head
	214	helix	–	–	therm transfer hydroph	82	pore	–	–
	216	helix	–	–		88	pore	–	–
	219	helix	–	–	total nb energy	–	–	124	head
	222	helix	–	–		–	–	173	head
	235	helix	–	–	turn tendency	38	pore	107	head
helical contact area	–	–	170	head		39	pore	125	head
hydropathy	–	–	104	head		44	pore	131	head
						–	–	150	head
						–	–	152	head

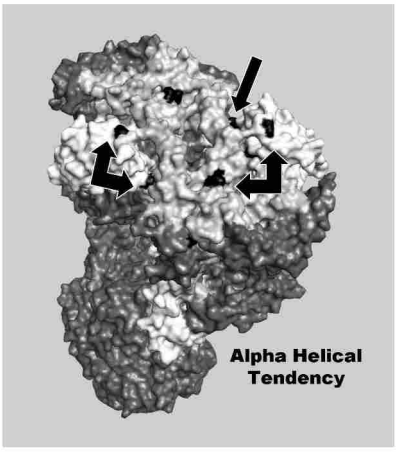
Figure 4 Moderately radical changes (category 6) in α-helical tendency in ISP and cyt-c_1 proteins are affected by positive-destabilising selection



$p < 0.001$.

z-score critical value = 3.09.

Figure 5 Three-dimensional representation of specific amino acids undergoing destabilising structural shifts in α-helical tendency. ISP is white, cyt-c_1 is shaded light grey and adaptive sites are black. All other subunits of the bc_1 complex are shaded dark grey. Arrows indicate positively selected amino acid sites buried in the tertiary structures



Another amino acid property that appears to be important to the adaptation of ISP and cyt-c_1 is equilibrium constant. Equilibruim constant, with reference to the ionisation property of the COOH group (pK') is a biochemical amino acid property that is a measure of the net change in the overall charge resulting from the ionisation of carboxyl groups (Gromiha and Ponnuswamy, 1993). When an acidic amino acid is exchanged for a pH neutral or basic amino acid (or vice versa), a radical shift in pK' will be produced. Radical shifts will result from certain exchanges between some similarly hydrophobic amino acids, such as between isoleucine and valine, leucine or methionine.

Analysis of pK' indicates that it has made an important contribution to the adaptation of ISP and cyt-c_1 (Figure 6). Ten sites in the cyt-c_1 pore and helix regions and five sites in the ISP head, helix and apical aperture regions have experienced radical shifts in pK' more often than expected by chance (Figure 7).

Figure 6 Extremely radical changes (category 8) in *pK'* in ISP ($p < 0.01$) and cyt-c_1 ($p << 0.001$) are affected by positive-destabilising selection. Positive-stabilising selection also is at work in ISP (category 1), promoting only the most conservative changes ($p < 0.05$)

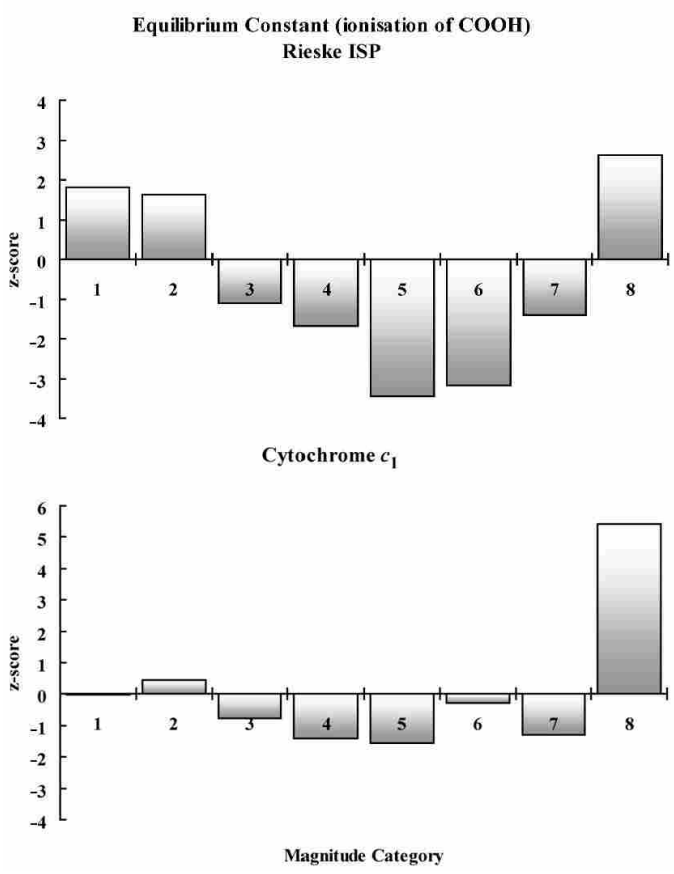
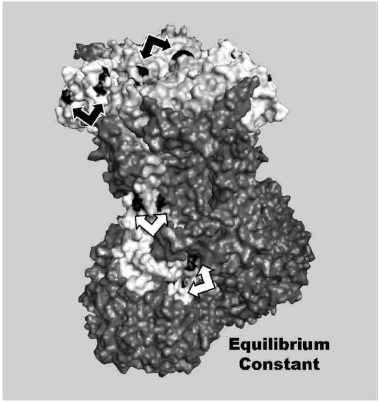


Figure 7 Amino acids undergoing radical biochemical shifts in *pK'*. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Figure 5. Arrows indicate the locations of amino acid sites discussed in the text

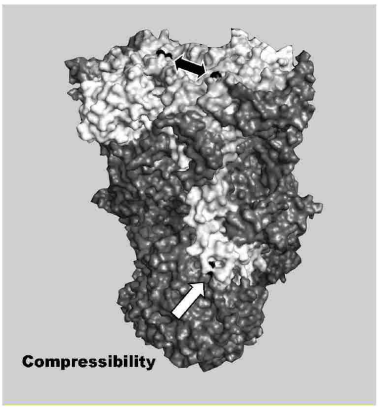


The changes that took place in the helical regions of the two proteins are largely reciprocal in nature resulting in a relatively constant micro-environment. There were also several biochemical shifts in both proteins that surrounded the central pore of the complex that contributed to the adaptation of the apical aperture. Finally, there was one radical pK' shift that was adaptive on the outer surface of the ISP head that likely is involved in the interaction of the bc_1 complex and the electron transporter ubiquinone.

Compressibility is a structural property that correlates with the local density of protein secondary structures (Gromiha and Ponnuswamy, 1993). Whereas ISP has experienced positive-destabilising selection for this property, cyt-c_1 has not, but appears to have been influenced by stabilising selection.

Compressibility was influenced by selection at three sites in ISP helix and head regions (only helix region is illustrated in Figure 8). The change occurring in the helix is situated at the protein-protein interface with core-1 protein. Future study that includes the core-1 protein will determine if this change resulted in coevolution.

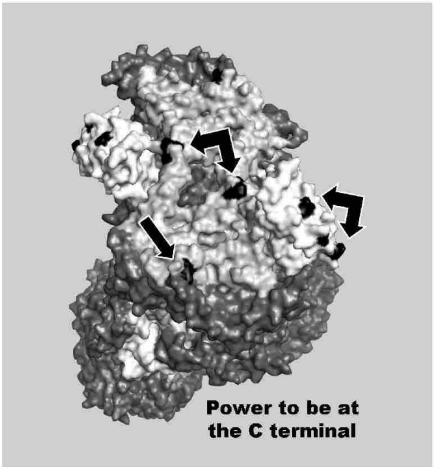
Figure 8 Amino acids undergoing radical biochemical shifts in compressibility. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Figure 5. Arrows indicate the locations of amino acid sites discussed in the text



Destabilising amino acid replacements for the compressibility did not occur more often that expected by chance in cyt-c_1, but four such replacements did occur in the pore and helix regions of the protein. The mutation occurring near the pore is noteworthy because the site at which it is situated is immediately distal (in the context of the tertiary structure) to the radical shifts in α-helical tendency described above (Figure 8) and may have contributed to the overall conformational changes in the apical aperture.

Power to be at the *C*-terminal is an amino acid property characterised by an increase in the energy potential of an amino acid to interact with surrounding residues or subunits (Gromiha and Ponnuswamy, 1993). Analysis of biochemical shifts in this property indicates only local regions being affected by selection, but no significant influence overall. Local regions being influenced by selection include three residues in the ISP head and two residues in the cyt-c_1 near the apical aperture (Figure 9). Finally, analyses of the properties coil tendency and turn tendency also indicated a significant influence of selection for radical structural shifts. Coil tendency refers to the ability of an amino acid to contribute to or initiate a coil in the protein, which are generally composed of more bulky amino acids (Charton and Charton, 1983). Turn tendency refers to the ability of a specific amino acid to contribute to or initiate a turn, a tight association usually involving the bent amino acid praline, in the protein backbone (Gromiha and Ponnuswamy, 1993). The analysis of coil tendency implicated three adaptive sites in cyt-c_1 and a single site in ISP for coil tendency (Table 1), while the analysis of turn tendency implicated three sites in cyt-c_1 and five sites in ISP (Table 1).

Figure 9 Amino acids undergoing radical biochemical shifts in power to be at the *C*-terminal. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Figure 5. Arrows indicate the locations of amino acid sites discussed in the text



Structural analysis indicates that the regions implicated in the physicochemical evolution sliding window analysis also are correlated to the larger three-dimensional structures of the cytochrome bc_1 complex. An example shows that two separate regions of selection when viewed in two dimensions actually form a continuous surface when viewed in three dimensions (Figure 10). The regions circled in black in Figure 11 indicate that the influences of selection for structural and biochemical shifts have affected residues that completely surround the apical aperture of the complex as seen in Figure 10 and include

sites on both ISP and cyt-c_1. When viewed in the context of the individual residues discussed above, these properties form clusters of sites on the apical intermembrane surface of the complex, primarily on the head of ISP and around the pore of cyt-c_1 and along the length of the transmembrane helix of each protein (Figure 2).

Figure 10 TreeSAAP sliding window output from areas in Figure 11 circled in black and here shaded black that complexly surround the apical aperture of the complex when view in three dimensions

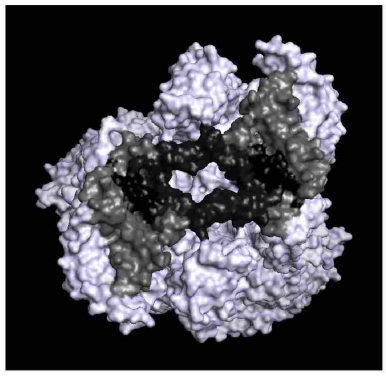
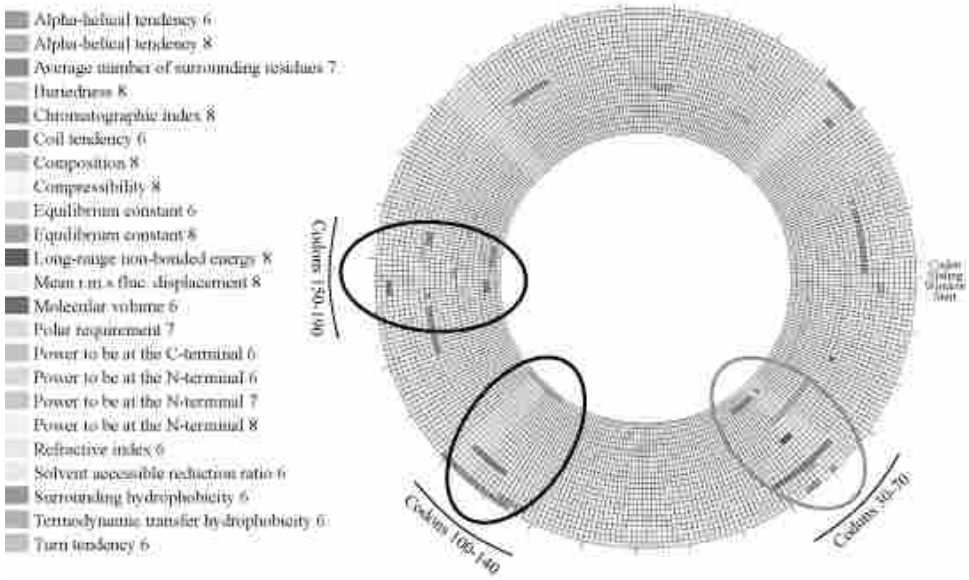


Figure 11 Three main regions implicated in the TreeSAAP sliding window analysis. The regions circled in black form a continuous surface surrounding the apical aperture of the cytochrome bc_1 protein complex. The region circled in grey is part of the leader sequence of this protein and is cleaved to become subunit 9 in the same complex (Doan et al., 2005). It was not considered in this study



4 Discussion

Cross-referenced analytical results indicate that several regions under heavy selection, when considered in the context of the 3-dimensional protein structure, are interacting and coevolving on the cyt-c_1 and ISP proteins. Analysis of the property α-helical tendency (Figure 5) implicates several amino acid residues in the protein coadaptation of the apical interaction between cyt-c_1 and ISP. Furthermore, the compressibility shows evidence of coadaptation along the helix interface of the proteins. These examples of structural coevolution demonstrate how these proteins work together over an evolutionary time scale and how each protein evolves in response to changes in the other at the protein-protein interface in order to maintain overall function. Furthermore, equilibrium constant has affected both of these regions, while power to be at the *C*-terminal has affected primarily the apical region. Both of these properties are biochemical in nature and likely adjust the ability of these proteins to maintain the function of the proton pump and to interact with each other and those other proteins in the complex that surround them.

The cyt-c_1 pore in the intermembrane space, through which protons are pumped into the intermembrane space of the mitochondria, also is under heavy positive selection (Figure 11). The MAdS Visualiser shows regions of cyt-c_1 that are under heavy selection. Those two regions shown in the 'onion' graph correspond to the pore region of cyt-c_1. This is where protons are actively pumped out of the complex, from the matrix to the intermembrane space. This creates a proton gradient and ultimately fuels ATPase in making ATP. Selection in this region of the bc_1 complex may have an effect on the efficiency of the protein and therefore the efficiency of the organism as a whole. Over 60% of all selection detected by TreeSAAP was in the pore region of cyt-c_1.

The MAdS Visulaiser gave similar results for the head of ISP (not shown). It is well known that ISP undergoes a conformational change in the course of transferring electrons and pumping protons (Zhang et al., 1998; Doan et al., 2005). This movement may have to do with the location of the iron sulfur centre and holding, releasing, or stabilising the protons coming from the heme groups in cytochrome *b*. Selection in this region of ISP may be result in adjustments in the ability or efficiency of this movement. Over 65% of all the selection detected by TreeSAAP for ISP was located on the head of ISP.

Finally, the vast majority of sites experiencing positive selection are located on the exterior surface of each protein. Selection may be occurring primarily on the exterior of the proteins because these are the areas that come into contact with cytochrome *c* and ubiquinol (coenzyme *Q*). These constitute the active sites of ISP and cyt-c_1, and selection promoting change in these regions may adjust binding affinity or optimise the efficiency of the complex as a whole.

Acknowledgements

We would like to thank Brady Curtis, Rick Smith, Wes Beckstead, Srikar Chamala and Gavin Svenson for their comments and insight. We would also like to thank the Department of Integrative Biology at Brigham Young University for funding.

References

Anderson, S., de Bruijn, M.H., Coulson, A.R., Eperon, I.C., Sanger, F. and Young, I.G. (1982) 'Complete sequence of bovine mitochondrial DNA. Conserved features of the mammalian mitochondrial genome', *J. Mol. Biol.*, Vol. 156, No. 4, pp.683–717.

Behar, D.M., Metspalu, E., Kivisild, T., Achilli, A., Hadid, Y., Tzur, S., Pereira, L., Amorim, A., Quintana-Murci, L., Majamaa, K., Herrnstadt, C., Howell, N., Balanovsky, O., Kutuev, I., Pshenichnov, A., Gurwitz, D., Bonne-Tamir, B., Torroni, A., Villems, R. and Skorecki, K. (2006) 'The matrilineal ancestry of ashkenazi jewry: portrait of a recent founder event', *Am. J. Hum. Genet.*, Vol. 78, No. 3, pp.487–497.

Bjornerfeldt, S., Webster, M.T. and Vila, C. (2006) 'Relaxation of selective constraint on dog mitochondrial DNA following domestication', *Genome Res.*, Vol. 16, No. 8, pp.990–994.

Brady, K.P., Rowe, L.B., Her, H., Stevens, T.J., Eppig, J., Sussman, D.J., Sikela, J. and Beier, D.R. (1997) 'Genetic mapping of 262 loci derived from expressed sequences in a murine interspecific cross using single-strand conformational polymorphism analysis', *Genome Res.*, Vol. 7, No. 11, pp.1085–1093.

Broughton, R.E., Milam, J.E. and Roe, B.A. (2001) 'The complete sequence of the zebrafish (Danio rerio) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA', *Genome Res.*, Vol., 11, No. 11, pp.1958–1967.

Caldwell, R.B., Kierzek, A.M., Arakawa, H., Bezzubov, Y., Zaim, J., Fiedler, P., Kutter, S., Blagodatski, A., Kostovska, D., Koter, M., Plachy, J., Carninci, P., Hayashizaki, Y. and Buerstedde, J.M. (2005) 'Full-length cDNAs from chicken bursal lymphocytes to facilitate gene function analysis', *Genome Biol.*, Vol. 6, No. 1, p.R6.

Charton, M. and Charton, B. (1983) 'The dependence of the Chou-Fasman parameters on amino acid side chain structure', *J. Theor. Biol.*, Vol. 111, pp.447–450.

Chimpanzee Sequencing and Analysis Consortium (2005) 'Initial sequence of the chimpanzee genome and comparison with the human genome', *Nature*, Vol. 437, No. 7055, pp.69–87.

Chou, P.Y. and Fasman, G.D. (1987) 'Prediction of the secondary structure of proteins from their amino acid sequence', *J. Adv. Enzymol.*, Vol. 47, pp.45–148.

Da Cruz, S., Xenarios, I., Langridge, J., Vilbois, F., Parone, P.A. and Martinou, J.C. (2003) 'Proteomic analysis of the mouse liver mitochondrial inner membrane', *J. Biol. Chem.*, Vol. 278, No. 42, pp.41566–41571.

DeLano, W. (2002) 'The PyMOL molecular graphics system', *DeLano Scientific*, San Carlos, California.

Desjardins, P. and Morais, R. (1990) 'Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates', *J. Mol. Biol.*, Vol. 212, No. 4, pp.599–634.

Doan, J.W., Schmidt, T.R., Wildman, D.E., Goodman, M., Weiss, M.L. and Grossman, L.I. (2005) 'Rapid nonsynonymous evolution of the iron-sulfur protein in anthropoid primates', *J. Bioenerg Biomembr*, Vol. 37, No. 1, pp.35–41.

Gencic, S., Schagger, H. and von Jagow, G. (1991) 'Core I protein of bovine ubiquinol-cytochrome-c reductase; an additional member of the mitochondrial-protein-processing family. Cloning of bovine core I and core II cDNAs and primary structure of the proteins', *Eur. J. Biochem.*, Vol. 199, No. 1, pp.123–131.

Gromiha, M.M. and Ponnuswamy, P.K. (1993) 'Relationship between amino acid properties and protein compressibility', *J. Theor. Biol.*, Vol. 165, pp.87–100.

Grosskopf, R. and Feldmann, H. (1981) 'Analysis of a DNA segment from rat liver mitochondria containing the genes for the cytochrome oxidase subunits i, ii, iii, atpase subunit 6, and several tRNA genes', *Curr. Genet.*, Vol. 4, pp.151–158.

Hixson, J.E. and Brown, W.M. (1986) 'A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications', *Mol. Biol. Evol.*, Vol. 3, No. 1, pp.1–18.

International chicken genome sequencing consortium (2004) 'Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution', *Nature*, Vol. 432, No. 7018, pp.695–716.

Hubbard, S.J., Grafham, D.V., Beattie, K.J., Overton, I.M., McLaren, S.R., Croning, M.D.R., Boardman, P.E., Bonfield, J.K., Burnside, J., Davies, R.M., Farrell, E.R., Francis, M.D., Griffiths-Jones, S., Humphray, S.J., Hyland, C., Scott, C.E., Tang, H., Taylor, R.G., Tickle, C., Brown, W.R.A., Birney, E., Rogers, J. and Wilson, S.A. (2005) 'Transcriptome analysis for the chicken based on 19,626 finished cDNA sequences and 485,337 expressed sequence tags', *Genome Res.*, Vol. 15, pp.174–183.

Klein, S.L., Strausberg, R.L., Wagner, L., Pontius, J., Clifton, S.W. and Richardson, P. (2002) 'Genetic and genomic tools for Xenopus research: the NIH Xenopus initiative', *Dev. Dyn.*, Vol. 225, No. 4, pp.384–391.

Kumar, S., Tamura, K. and Nei, M. (1994) 'MEGA: molecular evolutionary genetics analysis software for microcomputers', *Comput. Appl. Biosci.*, Vol. 10, pp.189–191.

Lindblad-Toh, K. *et al.* (2005) 'Broad sequencing platform. Genome sequence, comparative analysis and haplotype structure of the domestic dog', *Nature*, Vol. 438, No. 7069, pp.803–819.

Lundrigan, B.L., Jansa, S.A. and Tucker, P.K. (2002) 'Phylogenetic relationships in the genus mus, based on paternally, maternally, and biparentally inherited characters', *Syst. Biol.*, Vol. 51, No. 3, pp.410–431.

Maddison, D.R. and Maddison, W.P. (2000) *MacClade 4: Analysis of Phylogeny and Character Evolution*, Sinauer Associates, Sunderland, Massachusetts, p.492, book (as PDF file) + computer program.

Mammalian gene collection program team (2002) 'Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences', *Proc. Natl. Acad. Sci.*, USA, Vol. 99, No. 26, pp.16899–16903.

McClellan, D.A. and McCracken, K.G. (2001) 'Estimating the influence of selection on the variable amino acid sites of the cytochrome *b* protein functional domains', *Mol. Biol. Evol.*, Vol. 18, pp.917–925.

McClellan, D.A., Moss, J.L., Palfreyman, E.J., Smith, M.J., Christensen, R.G. and Sailsbery, J.K. (2005) 'Physicochemical evolution and molecular adaptation of the cetacean cytochrome *b* protein', *Mol. Biol. Evol.*, Vol. 22, pp.437–455.

Moore, S., Alexander, L., Brownstein, M., Guan, L., Lobo, S., Meng, Y., Tanaguchi, M., Wang, Z., Yu, J., Prange, C., Schreiber, K., Shenmen, C., Wagner, L., Bala, M., Barbazuk, S., Barber, S., Babakaiff, R., Beland, J., Chun, E., Del Rio, L., Gibson, S., Hanson, R., Kirkpatrick, R., Liu, J., Matsuo, C., Mayo, M., Santos, R.R., Stott, J., Tsai, M., Wong, D., Siddiqui, A., Holt, R., Jones, S.J. and Marra, M.A. (2005) 'Direct Submission', *BC Cancer Agency, Canada's Michael Smith Genome Sciences Centre*, Suite 100, 570 West 7th Avenue, Vancouver, British Columbia V5Z 4S6, Canada.

Nishikimi, M., Hosokawa, Y., Toda, H., Suzuki, H. and Ozawa, T. (1989) 'Cloning and sequence analysis of a cDNA encoding the Rieske iron-sulfur protein of rat mitochondrial cytochrome bc_1 complex', *Biochem. Biophys. Res. Commun.*, Vol. 159, No. 1, pp.19–25.

Swofford, D.L. (2003) *PAUP (and Other Methods). Phylogenetic Analysis Using Parsimony*, Version 4, Sinauer Associates, Sunderland, Massachusetts.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through weighting, position-specific gap penalties and weight matrix choice', *Nuc. Acids Res.*, Vol. 22, pp.4673–4680.

Usui, S., Yu, L. and Yu, C.A. (1990) 'Cloning and sequencing of a cDNA encoding the Rieske iron-sulfur protein of bovine heart mitochondrial ubiquinol-cytochrome c reductase', *Biochem. Biophys. Res. Commun.*, Vol. 167, No. 2, pp.575–579.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A. *et al.* (2001) 'Celera genomics. The sequence of the human genome', *Science*, Vol. 291, pp.1304–1351.

Wong, J.F., Ma, D.P., Wilson, R.K. and Roe, B.A. (1983) 'DNA sequence of the Xenopus laevis mitochondrial heavy and light', *Nucleic Acids Res.*, Vol. 11, No. 14, pp.4977–4995.

Woolley, S., Johnson, J., Smith, M.J., Crandall, K.A. and McClellan, D.A. (2003) 'TreeSAAP: selection on amino acid properties using phylogenetic trees', *Bioinformatics*, Vol. 19, pp.671, 672.

Yang, Z. (1997) 'PAML: a program package for phylogenetic analysis by maximum likelihood', *CABIOS*, Vol. 13, pp.555, 556.

Zhang, Z., Huang, L., Shulmeister, V.M., Chi, Y.I., Kim, K.K., Hung, L.W., Crofts, A.R., Berry, E.A. and Kim, S.H. (1998) 'Electron transfer by domain movement in cytochrome bc_1', *Nature*, Vol. 392, pp.677–684.

Appendix A

Accession numbers for protein-coding DNA from the cytochrome bc_1 complex that was used to reconstruct the phylogeny in Figure 1 (Wong et al., 1983; Klein et al., 2002; Anderson et al., 1982; Moore et al., Gencic et al., 1991; Usui et al., 1990; Desjardins and Morais, 1990; International Chicken Genome Sequencing Consortium, 2004; Hubbard et al., 2005; Caldwell et al., 2005; Hixson and Brown, 1986; Chimpanzee Sequencing and Analysis Consortium, 2005; Doan et al., 2003 Bjornerfeldt et al., 2006; Lindblad-Toh et al., 2005; Lundrigan et al., 2002; Da Cruz et al., 2003; Brady et al., 1997; Mammalian gene collection program team, 2002; Behar et al., 2006; Grosskopf and Feldmann, 1981; Venter et al., 2001; Nishikimi et al., 1989, Broughton et al., 2001).

Common name	*Genus*	*Species*	*Gene*	*GenBank number*	*Reference*
Frog	*Xenopus*	*Laevis*	Cytochrome *b*	M10217	Wong et al. (1983)
			Cytochrome c_1	BC045127	Klein et al. (2002)
			Core protein I	BC098177	Klein et al. (2002)
			Core protein II	BC106252	Klein et al. (2002)
			Rieske ISP	BC041528	Klein et al. (2002)
Bovine	*Bos*	*Taurus*	Cytochrome *b*	V00654	Anderson et al. (1982)
			Cytochrome c_1	BC109917	Moore et al. (2005)
			Core protein I	NM_174629	Gencic et al. (1991)
			Core protein II	BC102337	Moore et al. (2005)
			Rieske ISP	NM_174813	Usui et al. (1990)
Chicken	*Gallus*	*Gallus*	Cytochrome *b*	X52392	Desjardins and Morais (1990)
			Core protein I	XM_414356	International Chicken Genome Sequencing Consortium (2004)
			Core protein II	BX950415	Hubbard et al. (2005)
			Rieske ISP	AJ719669	Caldwell et al. (2005)

Common name	*Genus*	*Species*	*Gene*	*GenBank number*	*Reference*
Chimp	*Pan*	*Troglodytes*	Cytochrome *b*	D38113	Hixson and Brown (1986)
			Cytochrome c_1	XM_528263	Chimpanzee Sequencing and Analysis Consortium (2005)
			Core protein I	XM_516440	Chimpanzee Sequencing and Analysis Consortium (2005)
			Core protein II	XM_523485	Chimpanzee Sequencing and Analysis Consortium (2005)
			Rieske ISP	DQ480496	Bjornerfeldt et al. (2006)
Dog	*Canis*	*Familiaris*	Cytochrome *b*	AY729880	Zhu et al. (2004)
			Cytochrome c_1	XM_532351	Lindblad-Toh et al. (2005)
			Core protein I	XM_846116	Lindblad-Toh et al. (2005)
			Core protein II	XM_536942	Lindblad-Toh et al. (2005)
			Rieske ISP	XM_533711	Lindblad-Toh et al. (2005)
Mouse	*Mus*	*Musculus*	Cytochrome *b*	AY057804	Lundrigan et al. (2002)
			Cytochrome c_1	NM_025567	Da Cruz et al. (2003)
			Core protein I	NM_025407	Brady et al. (1997)
			Core protein II	BC003423	Mammalian gene collection program team (2002)
			Rieske ISP	NM_025710	Da Cruz et al. (2003)
Human	*Homo*	*Sapiens*	Cytochrome *b*	DQ301818	Behar et al. (2006)
			Cytochrome c_1	BC001006	Mammalian gene collection program team (2002)
			Core protein I	BC009586	Mammalian gene collection program team (2002)
			Core protein II	BC003136	Mammalian gene collection program team (2002)
			Rieske ISP	BC021057	Mammalian gene collection program team (2002)
Rat	*Rattus*	*Norvegicus*	Cytochrome *b*	X14848	Grosskopf and Feldmann (1981)
			Cytochrome c_1	XM_001072221	Venter et al. (2001)
			Core protein I	BC078923	Mammalian gene collection program team (2002)
			Core protein II	BC083610	Mammalian gene collection program team (2002)
			Rieske ISP	NM_001008888	Nishikimi et al. (1989)
Zebrafish	*Danio*	*Rerio*	Cytochrome *b*	AC024175	Broughton et al. (2001)
			Cytochrome c_1	BC080245	Mammalian gene collection program team (2002)
			Core protein II	BC097011	Mammalian gene collection program team (2002)
			Rieske ISP	BC059475	Mammalian gene collection program team (2002)

In Print: *Proceedings of the Bioinformatics and Biotechnology Symposium*
Submitted: April 2006
Printed: October 2006, vol. 3, pp. 10-17

Molecular Coevolution of the Vertebrate Cytochrome c_1 and Rieske Iron Sulfur Protein in the Cytochrome bc1 Complex

By Kimberly K. Baer and David A. McClellan

Molecular Coevolution of the Vertebrate Cytochrome c_1 and Rieske Iron Sulfur Protein in the Cytochrome bc_1 Complex

Kimberly Baer*, David McClellan†
Department of Integrative Biology, Brigham Young University, Provo, UT 84602
* *kimberly.baer@byu.net,* † *david_mcclellan@byu.edu*

Key words: Molecular coevolution, cytochrome c1, Rieske iron sulfur protein, cytochrome bc1 complex.

SUMMARY
The cytochrome bc_1 complex is the third protein complex in the electron transport chain in mitochondria. It functions both as a proton pump and an electron transporter and consists of 11 subunits including cytochrome c_1 (cyt-c_1) and the Rieske iron sulfur protein (ISP). Vertebrate model organism phylogenies were used in conjunction with existing 3D protein structures to evaluate the biochemical evolution of cyt-c_1 and ISP in terms of selection on amino acid properties. We found acting on the exterior surfaces of both proteins, and specifically on the surface of the cyt-c_1 core region. There is evidence to support the coevolution of these proteins relative to alpha helical tendencies, compressibility, and equilibrium constant.

The cytochrome bc_1 complex (also known as complex III or ubiquinol: cytochrome c oxidoreductase) is made up of 11 different protein subunits. These include the well-studied cytochrome *b* (cyt-*b*), cytochrome c_1 (cyt-c_1), the Rieske iron sulfur protein (ISP), and eight others. These proteins span the inner mitochondrial membrane, and have functional domains that work in the matrix, membrane, and intermembrane space of the mitochondria.

This study represents a newly developed method for characterizing protein-protein interactions. As an example of the method, we evaluate historical evolutionary biochemical interactions between cyt-c_1 and ISP within the bc_1 complex (Fig 1).

FIGURE 1: Location of cytochrome c_1 (black, left) and Rieske ISP (black, right) dimers in cytochrome bc_1 complex (gray).

The exploration of new methods by which historical and potentially important protein interactions can be described is vital to the study of genetic disease, infection, and cancer. With this study we are proposing a new method for accomplishing this goal.

Methods

Sampling: All vertebrate cyt-c_1 and ISP gene sequences were obtained from GenBank for the *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Bos taurus* (cow), *Canis familiaris* (dog), *Rattus norvegensis* (Norway rat), *Mus musculus* (house mouse), *Gallus gallus* (chicken), *Xenopus laevis* (African clawed frog), and *Danio rerio* (zebrafish). Five protein-coding gene sequences were used to estimate the cytochrome bc_1 phylogeny for a total of 5801 bp: cyt-c_1 – 981 bp; ISP – 825 bp; cyt-*b* – 1148 bp; core 1 – 1446 bp; core 2 – 1401 bp (Appendix A).

Phylogeny Reconstruction: Protein-coding DNA sequences were translated into amino acid data in MEGA3 (Kumar et al., 1994) and aligned using ClustalW (Thompson et al. 1994). The resulting protein alignments were then used as templates for aligning the protein-coding DNA data. After alignment, the sequences were concatenated using MacClade (Maddison and Maddison, 2000). PAUP* (Swofford, 2003) was used to estimate a phylogeny for the cytochrome bc_1 complex using a parsimony optimality criterion. A heuristic search algorithm was employed, with 1,000 random addition replicates and TBR branch swapping. A strict consensus of all most parsimonious trees was constructed for use in subsequent analyses.

Biochemical Adaptation Analysis: The software package TreeSAAP (Woolley et al. 2003) evaluates protein-coding nucleotide sequences using a given phylogenetic topology to test for positive selection

on quantitative physicochemical amino acid properties. This program identifies the properties historically affected by selection, and locates the exact amino acid sites that experienced significant biochemical or conformational shifts. The baseml algorithm (Yang 1997) is used to reconstruct ancestral character states at the nodes on a given phylogeny, and then the program evaluates the average influence of selection on amino acid properties for a variable number of magnitude categories. For this study eight categories of selection were used, and only categories 6, 7, and 8 were considered because they are unambiguously associated with molecular adaptation (McClellan et al. 2005). Cyt-c_1 and ISP were analyzed in TreeSAAP independently. The results were further analyzed using a sliding window to statistically analyze the data for significant clustering of radical physicochemical shifts resulting from naturally occurring amino acid replacements. A Bonferroni correction for multiple sampling was employed. All results are relative to *Bos taurus* sequences.

Visualization: Regions under heavy selection were correlated to pre-existing three-dimensional structures of the bc_1 complex (PDB A206) using the program Pymol (DeLano 2002). In addition, Pymol was used to annotate the specific amino acids that were found to have undergone radical biochemical shifts. This technique allows us to visualize the influences of selection by differentiating those regions historically influenced by selection on specific amino acid properties on functional protein domains (Fig 2).

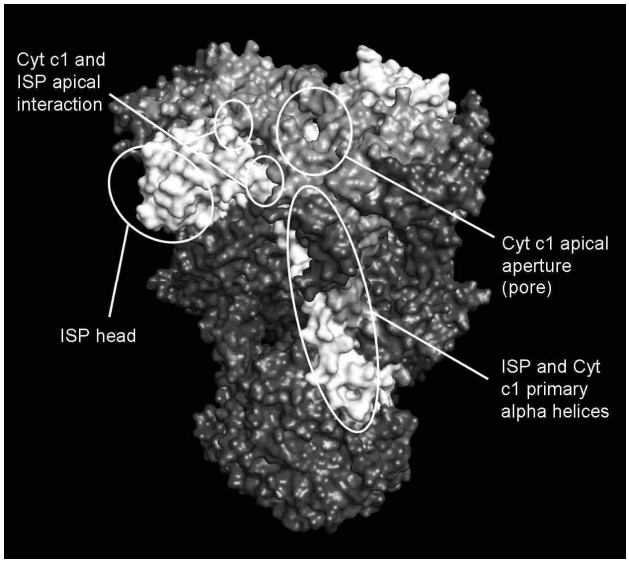


FIGURE 2: Functional domains of cyt-c_1 (gray) and ISP (white) in the broader context of the three-dimensional structure of the cytochrome bc_1 protein complex.

Results

The most parsimonious tree resulted in the following phylogeny (Fig 3).

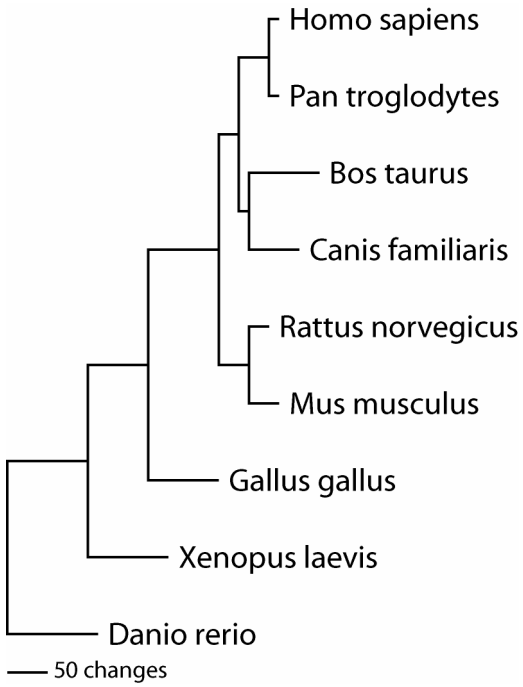


FIGURE 3: Parsimony tree for the cytochrome bc_1 gene sequences used in this study.

We recovered 187 historical amino acid replacements from the cyt-c_1 data, and 279 amino acid replacements from the ISP data. TreeSAAP analysis of these changes resulted in the detection of the influence of selection on several physicochemical amino acid properties, which clustered in three regions of both proteins, with amino acid replacements in these regions being affected by multiple amino acid properties.

Detected regions correspond to residues on the outer surface of each protein, including surfaces on these proteins that are in direct contact. Table 1 shows the amino acid properties under positive-destabilizing selection, and domain locations of each specific property. For example, α-helical tendency was influenced by selection at four sites in cyt-c_1 and three sites in ISP. Decreases in α-helical tendency broadens the outer surface area of a helix, and allows a more flexible surface area contact with neighboring protein constituents, while increases have the opposite effect (Chou and Fasman 1987). Both cyt-c_1 and ISP exhibit significant positive-destabilizing selection for moderately radical changes (cat 6) in this property (Fig 4). In ISP, there are radical shifts in α-helical tendency at the apical

Table 1. Amino acid properties and sites identified as being under positive-destabilizing selection in cyt-c_1 and ISP proteins.

Property	cyt-c_1	Domain	ISP	Domain	Property	cyt-c_1	Domain	ISP	Domain
α-helical tendency	24	helix	72	apical	isoelectric point			116	head
	38	pore	79	apical				128	head
	39	pore	88	apical				170	head
	190	interior			lng-range nb energy			124	head
ave # surround res	71	pore	173	head	mean rms fluc displ	88	pore		
bulkiness			47	helix	partial specific vol			170	head
buriedness			124	head	polar requirement	98	pore	12	helix
			173	head				27	helix
chromatograph index			173	head				31	helix
coil tendency	38	pore	24	helix	pwr to be C-terminal	75	apical	107	head
	39	pore				77	apical	113	head
	194	interior				166	exterior	131	head
composition	187	interior			pwr to be N-terminal	96	pore	16	helix
compressibility	59	pore	16	helix		98	pore		
	63	pore	24	helix		187	interior		
	66	pore	103	head	refractive index	187	interior		
	221	helix			sh & med range nb en			104	head
equilibrium constant	68	pore	5	helix				170	head
	70	pore	7	helix	solvent accessible red	71	pore	124	head
	82	pore	124	head		82	pore		
	127	pore	192	apical	surr hydrophobicity	71	pore	124	head
	211	helix	194	apical		82	pore	173	head
	214	helix			therm transfer hydroph	82	pore		
	216	helix				88	pore		
	219	helix			total nb energy			124	head
	222	helix						173	head
	235	helix			turn tendency	38	pore	107	head
helical contact area			170	head		39	pore	125	head
hydropathy			104	head		44	pore	131	head
								150	head
								152	head

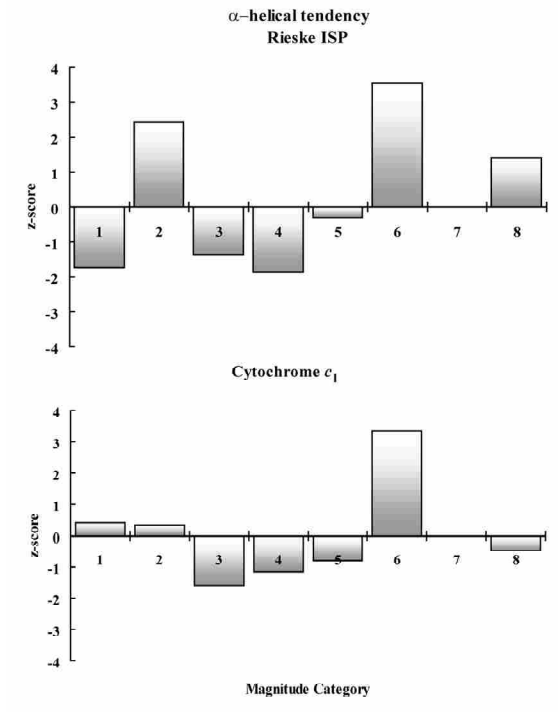


FIGURE 4: Moderately radical changes (category 6) in α-helical tendency in ISP and cyt-c_1 proteins are affected by positive-destabilizing selection ($p < 0.001$; z-score critical value = 3.09).

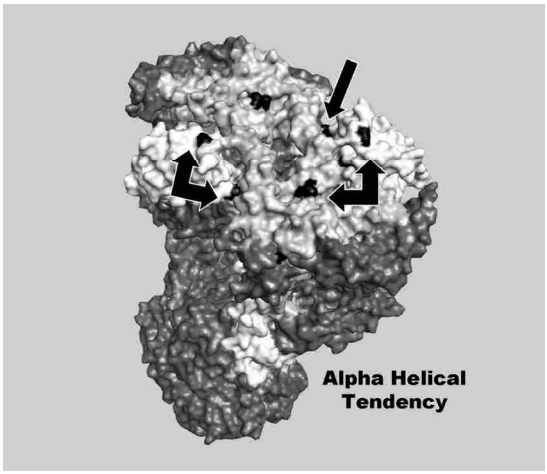


FIGURE 5: Three-dimensional representation of specific amino acids undergoing destabilizing structural shifts in α-helical tendency. ISP is white, cyt-c_1 is shaded light gray, and adaptive sites are black. All other subunits of the bc_1 complex are shaded dark gray. Arrows indicate positively selected amino acid sites buried in the tertiary strutures.

interaction between cyt-c_1 and ISP as a result of mutations that changed both a threonine and a serine to alanine (Fig 5). Cyt-c_1 shows moderately radical shifts in α-helical tendency resulting from mutations that changed an alanine to glycine, a threonine to serine and a serine to alanine. These replacements

resulted in reciprocal α-helical structural changes (increases in the property in ISP and corresponding decreases in cyt-c_1) in the region surrounding the apical aperture of the complex, likely resulting in a coevolutionary modification of overall pore conformation.

Another amino acid property that appears to be important to the adaptation of ISP and cyt-c_1 is compressibility, a structural property that correlates with the local density of protein secondary structures (Gromiha and Ponnuswamy 1993). Whereas ISP has experienced positive-destabilizing selection (category 8) for this property, cyt-c_1 has not (Fig 6), but appears to have been influenced by stabilizing selection (category 1).

Compressibility was influenced by selection at three sites in ISP helix and head regions (only helix region is illustrated in Fig 7). The change occurring in the helix is situated at the protein-protein interface with core-1 protein. Future study that includes the core-1 protein will determine if this change resulted in coevolution.

Destabilizing amino acid replacements for the compressibility did not occur more often that expected by chance in cyt-c_1, but four such

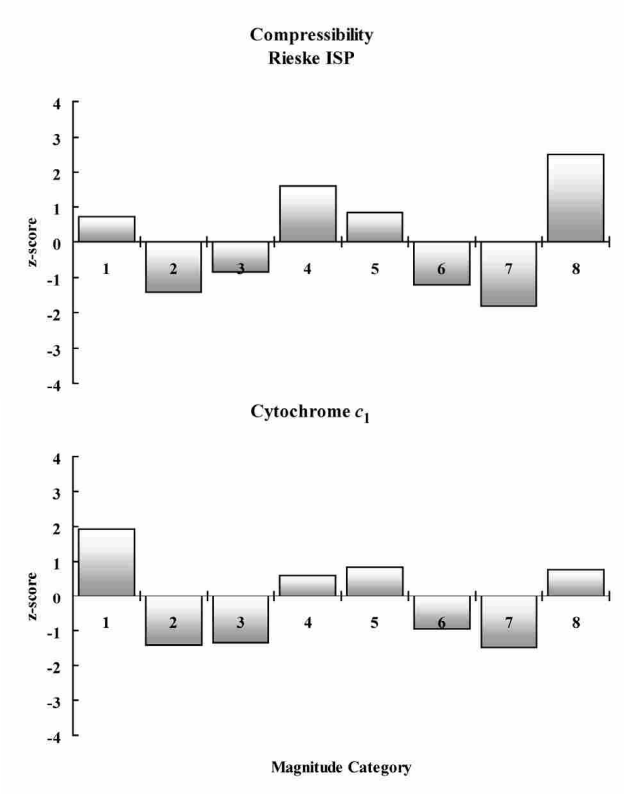


Figure 6: Extremely radical changes (category 8) in compressibility in ISP are affected by positive-destabilizing selection ($p < 0.01$; z-score critical value = 2.32). Positive-stabilizing selection is at work in cyt-c_1, promoting only the most conservative changes (category 1) ($p < 0.05$; z-score critical value = 1.645).

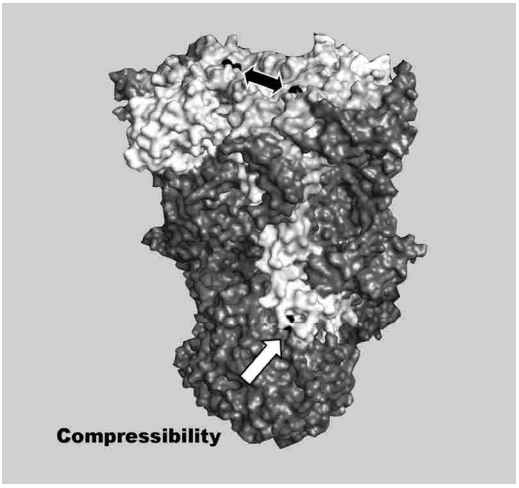


Figure 7: Amino acids undergoing radical biochemical shifts in compressibility. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Fig 5. Arrows indicate the locations of amino acid sites discussed in the text.

replacements did occur in the pore and helix regions of the protein. The mutation occurring near the pore is noteworthy because the site at which is situated is immediately distal (in the context of the tertiary structure) to the radical shifts in α-helical tendency described above (Fig 7), and may have contributed to the overall conformational changes in the apical aperture.

The equilibrium constant with reference to the ionization property of the COOH group (*pK'*) is a biochemical amino acid property that is a measure of the net change in the overall charge resulting from the ionization of carboxyl groups (Gromiha and Ponnuswamy 1993). When an acidic amino acid is exchanged for a pH neutral or basic amino acid (or vice versa), a radical shift in *pK'* will be produced. Radical shifts will result from certain exchanges between some similarly hydrophobic amino acids, such as between isoleucine and valine, leucine or methionine,

Analysis of *pK'* indicates that it has made an important contribution to the adaptation of ISP and cyt-c_1 (Fig 8). Ten sites in the cyt-c_1 pore and helix regions and five sites in the ISP head, helix, and apical aperture regions have experienced radical shifts in *pK'* more often than expected by chance (Fig 9).

The changes that took place in the helical regions of the two proteins are largely reciprocal in nature resulting in a relatively constant micro-environment. There were also several biochemical shifts in both proteins that surrounded the central pore of the complex that contributed to the adaptation of the apical aperture. Finally, there was one radical *pK'* shift that was adaptive on the outer

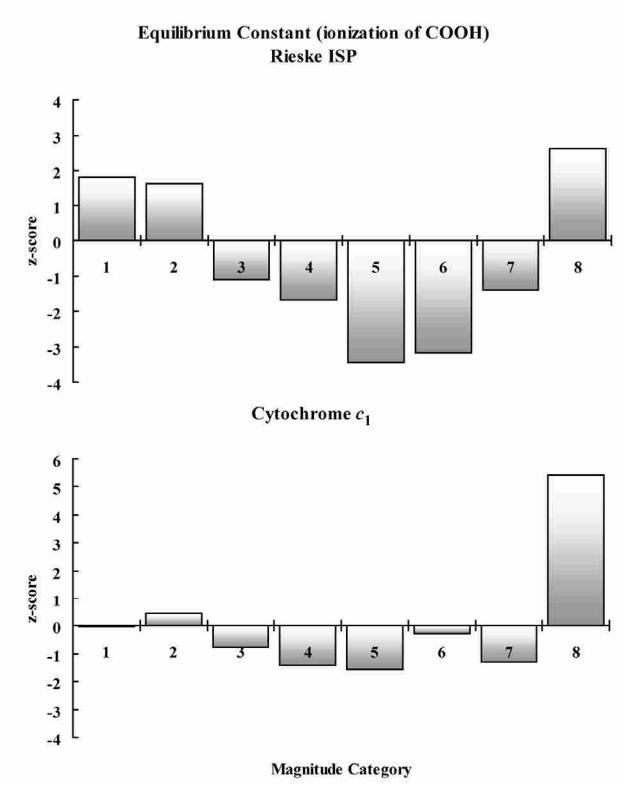


Figure 8: Extremely radical changes (category 8) in *pK'* in ISP ($p < 0.01$) and cyt-c_1 ($p << 0.001$) are affected by positive-destabilizing selection. Positive-stabilizing selection also is at work in ISP (category 1), promoting only the most conservative changes ($p < 0.05$).

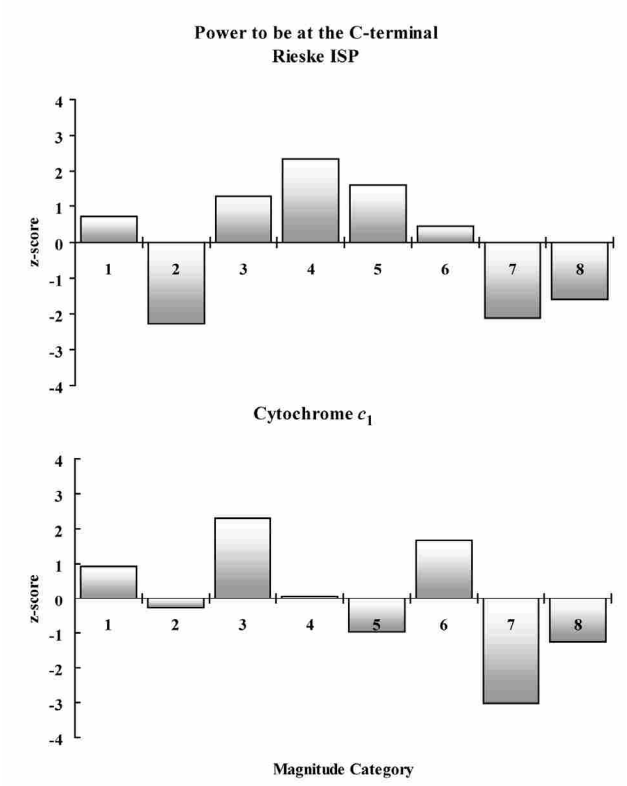


Figure 10: The average influence of power to be at the C-terminal is selectively neutral even though three sites in both proteins experienced more radical shifts than expected by chance.

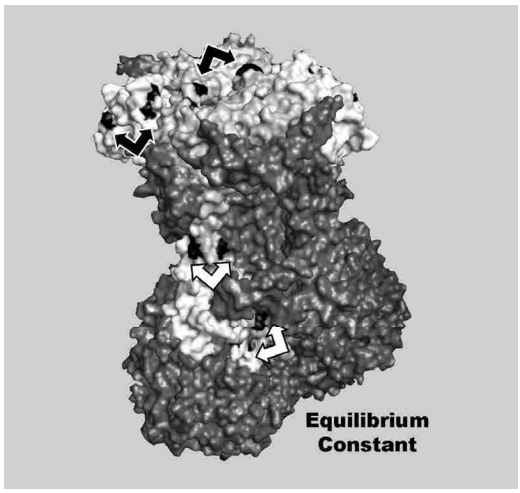


Figure 9: Amino acids undergoing radical biochemical shifts in *pK'*. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Fig 5. Arrows indicate the locations of amino acid sites discussed in the text.

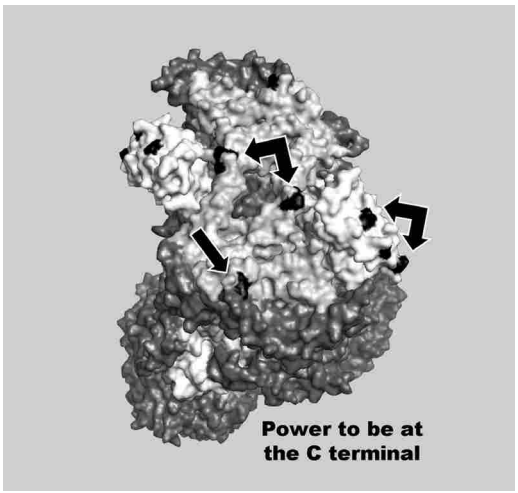


Figure 11: Amino acids undergoing radical biochemical shifts in power to be at the C-terminal. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Fig 5. Arrows indicate the locations of amino acid sites discussed in the text.

surface of the ISP head that likely is involved in the interaction of the bc_1 complex and the electron transporter ubiquinone.

Power to be at the C-terminal is an amino acid property characterized by an increase in the energy potential of an amino acid to interact with surrounding residues or subunits (Gromiha and Ponnuswamy 1993). Analysis of biochemical shifts in this property indicates only local regions being affected by selection, but no significant influence overall (Fig 10). Local regions being influenced by selection include three residues in the ISP head and two residues in the cyt-c_1 near the apical aperture (Fig 11).

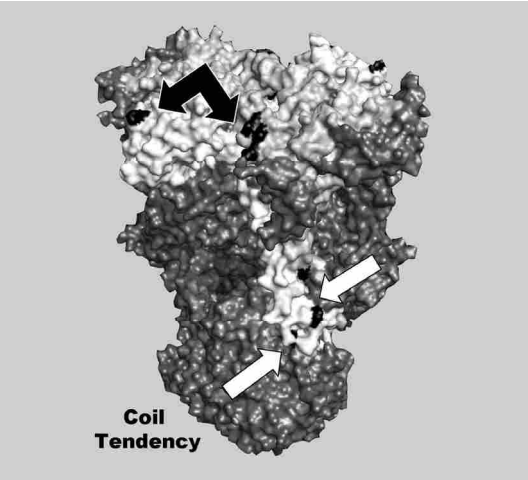


Figure 12: Amino acids undergoing radical biochemical shifts in coil tendency. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Fig 5. Arrows indicate the locations of amino acid sites discussed in the text.

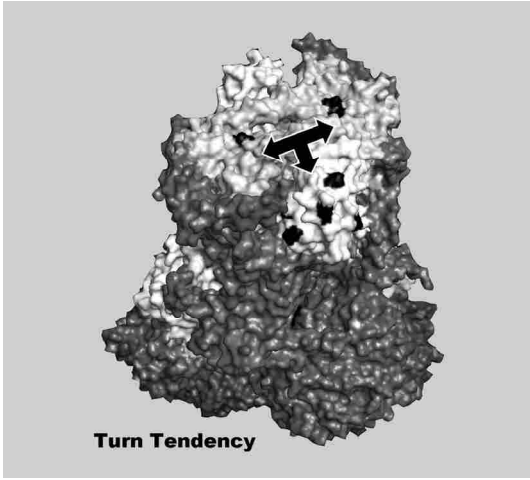


Figure 13: Amino acids undergoing radical biochemical shifts in power to be at the C-terminal. Three-dimensional representations of ISP and cyt-c_1 are shaded as in Fig 5. Arrows indicate the locations of amino acid sites discussed in the text.

Finally, analyses of the properties coil tendency and turn tendency also indicated a significant influence of selection for radical structural shifts. Coil tendency refers to the ability of an amino acid to contribute to or initiate a coil in the protein, which are generally composed of more bulky amino acids (Charton and Charton 1983). Turn tendency refers to the ability of a specific amino acid to contribute to or initiate a turn, a tight association usually involving the bent amino acid praline, in the protein backbone (Gromiha and Ponnuswamy 1993). The analysis of coil tendency implicated three adaptive sites in cyt-c_1 and a single site in ISP for coil tendency (Fig 12), while the analysis of turn tendency implicated three sites in cyt-c_1 and five sites in ISP (Fig 13).

Structural analysis indicates that the regions implicated in the physicochemical evolution sliding window analysis also are correlated to the larger three-dimensional structures of the cytochrome bc_1 complex. An example shows that two separate regions of selection when viewed in two dimensions (Fig 14), actually form a continuous surface when viewed in three dimensions (Fig 15). The regions circled in black in Fig 14 indicate that the influences of selection for structural and biochemical shifts have affected residues that completely surround the apical aperture of the complex as seen in Fig. 15. When viewed in the context of the individual residues discussed above, these properties form clusters of sites on the apical intermembrane surface of the complex, primarily on the head of ISP and around the pore of cyt-c_1, and along the length of the the transmembrane helix of each protein (Fig 16).

Discussion

Cross-referenced analytical results indicate that several regions under heavy selection, when considered in the context of the three-dimensional protein structure, are interacting and coevolving on the cyt-c_1 and ISP proteins. Analysis of the property α-helical tendency (Fig. 5) implicates several amino acid residues in the protein coadaptation of the apical interaction between cyt-c_1 and ISP. Furthermore, the compressibility shows evidence of coadaptation along the helix interface of the proteins. These examples of structural coevolution demonstrate how these proteins work together over an evolutionary time scale and how each protein evolves in response to changes in the other at the protein-protein interface in order to maintain overall function. Furthermore, equilibrium constant has affected both of these regions, while power to be at the C-terminal has affected primarily the apical region. Both of these properties are biochemical in nature and likely adjust the ability of these proteins to maintain the function of the proton pump and to interact with each other and those other proteins in the complex that surround them.

The cyt-c_1 pore in the intermembrane space, through which protons are pumped into the intermembrane space of the mitochondria, also is under heavy positive selection (Fig. 14). The MAdS Visualizer shows regions of cyt-c_1 that are under heavy selection. Those two regions shown in the "onion" graph correspond to the pore region of cyt-c_1. This is where protons are actively pumped out of the complex, from the matrix to the intermembrane space. This creates a proton gradient and ultimately

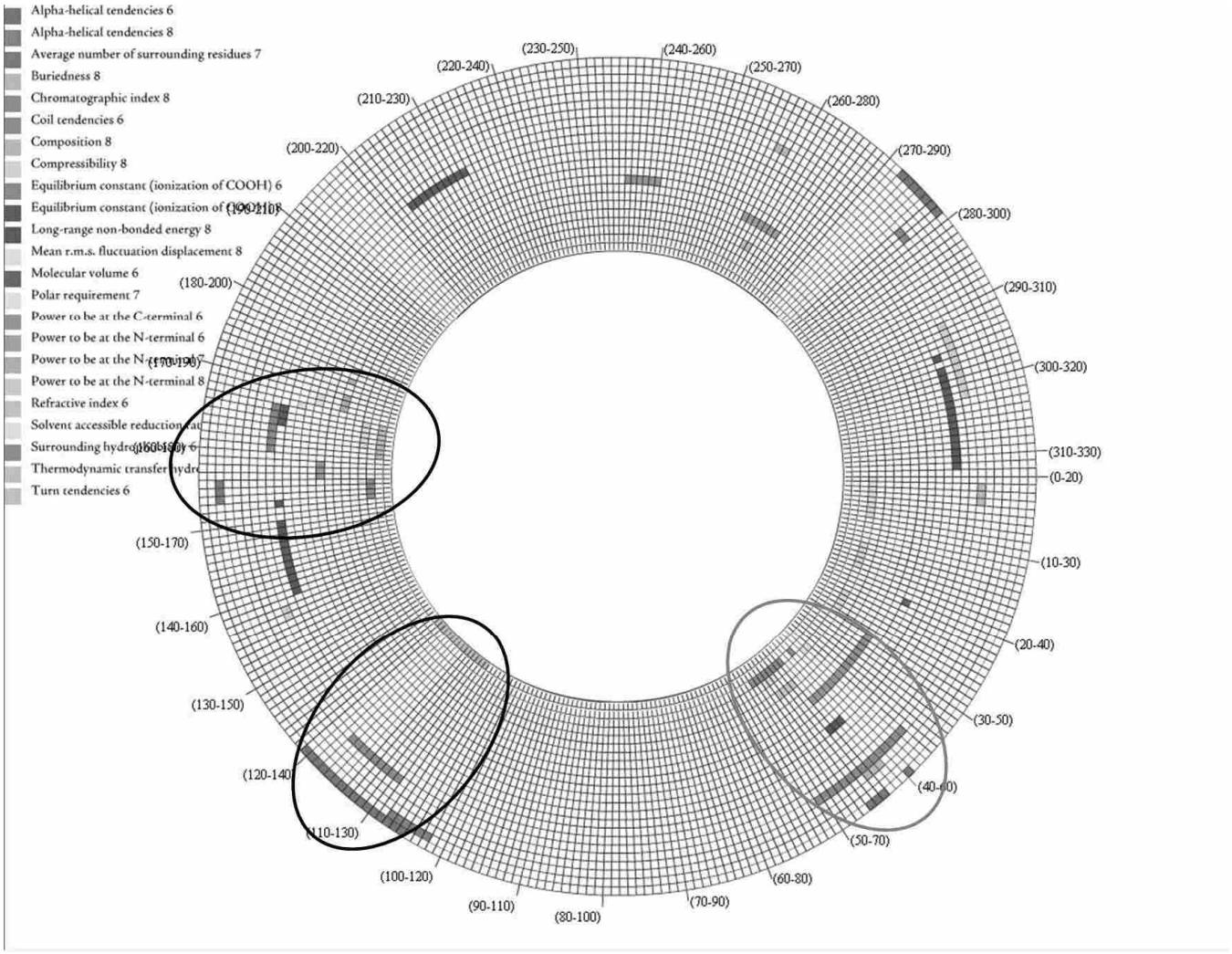


Figure 14: Three main regions implicated in the TreeSAAP sliding window analysis. The regions circled in black form a continuous surface surrounding the apical aperture of the cytochrome bc_1 protein complex. The region circled in gray is part of the leader sequence of this protein and is cleaved to become subunit 9 in the same complex (Doan, et. al. 2005). It was not considered in this study.

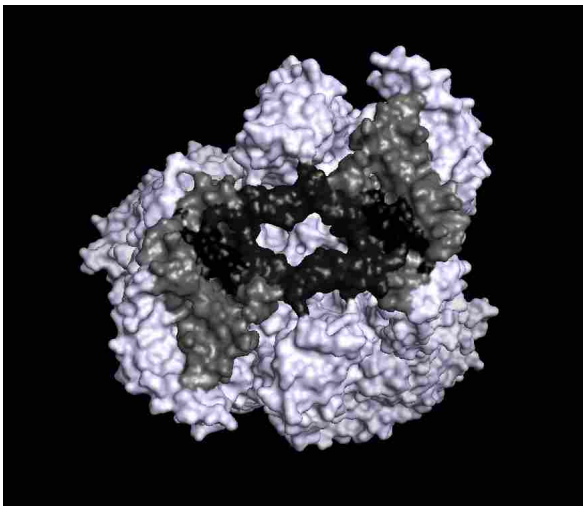


Figure 15: TreeSAAP sliding window output from areas in Fig 14 circled in black and here shaded black that complexly surround the apical aperture of the complex when view in three dimensions.

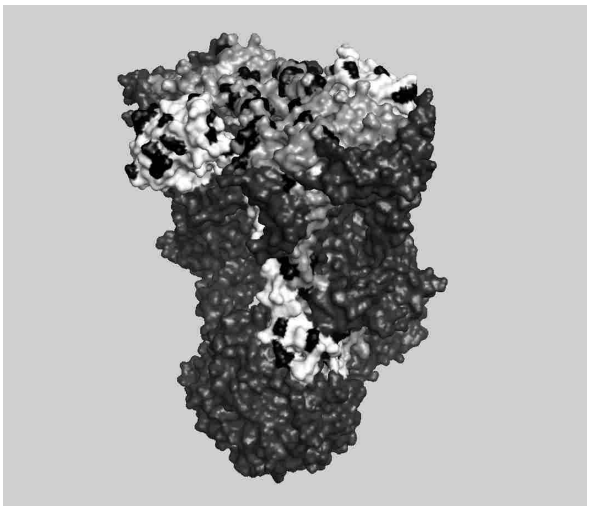


Figure 16: Residues affected by positive destabilizing selection for radical shifts in the properties considered in this study. All biochemical properties under selection are shown in black.

fuels ATPase in making ATP. Selection in this region of the bc1 complex may have an effect on the efficiency of the protein, and therefore the efficiency of the organism as a whole. Over 60% all selection detected by TreeSAAP was in the pore region of cyt-c_1.

The MAdS Visulaizer gave similar results for the head of ISP (no shown). It is well known that ISP undergoes a conformational change in the course of transferring electrons and pumping protons (Zhang et al. 1998). This movement may have to do with the location of the iron sulfur center, and holding, releasing, or stabilizing the protons from coming from the heme groups in cytochrome *b*. Selection in this region of ISP may be result in adjustments in the ability or efficiency of this movement. Over 65% of all the selection detected by TreeSAAP for ISP was located on the head of ISP.

Finally, the vast majority of sites experiencing positive selection are located on the exterior surface of each protein. Selection may be occurring primarily on the exterior of the proteins because these are the areas that come into contact with cytochrome *c* and ubiquinol (coenzyme Q). These constitute the active sites of ISP and cyt-c_1, and selection promoting change in these regions may adjust binding affinity or optimize the efficiency of the complex as a whole.

References

Charton, M. and B. Charton. 1983. The dependence of the Chou-Fasman parameters on amino acid side chain structure. J. Theor. Biol. 111:447-450.

Chou, P.Y. and G.D. Fasman. 1987. Prediction of the secondary structure of proteins from their amino acid sequence. J. Adv. Enzymol. 47:45-148.

DeLano, W. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, California.

Doan, J., T. Schmidt, D. Wildman, M. Goodman, M. Weiss, L. Grossman. 2005. Rapid Nonsynonymous Evolution of the Iron-Sulfur Protein in Anthropoid Primates. J. Bioenerg Biomembr., 37:35-41.

Kumar, S., K. Tamura, and M. Nei. 1994. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput. Appl. Biosci. 10:189-191.

Gromiha, M.M. and P.K. Ponnuswamy. 1993. Relationship Between Amino Acid Properties and Protien Compressibility. J. Theor. Biol. 165:87-100.

Maddison, D.R., and W.P. Maddison. 2000. MacClade 4: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, Massachusetts. 492 pp book (as PDF file) + computer program.

McClellan, D. A., and K. G. McCracken. 2001. Estimating the influence of selection on the variable amino acid sites of the cytochrome b protein functional domains. Mol. Biol. Evol. 18:917–925.

McClellan, D. A., J. L. Moss, E. J. Palfreyman, M. J. Smith, R. G. Christensen, and J. K. Sailsbery. 2005. Physicochemical evolution and molecular adaptation of the cetacean cytochrome *b* protein. Mol. Biol. Evol. 22:437-455.

Swofford, D. L. 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through weighting, position-specific gap penalties and weight matrix choice. Nuc. Acids Res. 22:4673-4680.

Woolley, S., J. Johnson, M. J. Smith, K. A. Crandall, and D. A. McClellan. 2003. TreeSAAP: selection on amino acid properties using phylogenetic trees. Bioinformatics 19:671–672.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555-556.

Zhang, Z, L Huang, V.M. Shulmeister, Y.I. Chi, K.K. Kim, LW Hung, AR Crofts, EA Berry, and SH Kim. 1998. Electron transfer by domain movement in cytochrome bc1. Nature 392: 677-84.

Software without citations can be found at: http://inbio.byu.edu/faculty/dam83/cdm/

Appendix A: Accession numbers for protein-coding DNA from the cytochrome bc_1 complex that was used to reconstruct the phylogeny in Fig 1.

Genus	Species	Common Name	Gene	GenBank Accession
Bos	taurus	Bovine	cyt-*b*	V00654
			cyt-c_1	BC109917
			core 1	NM_174629
			core 2	BC102337
			ISP	NM_174813
Canis	familiaris	Dog	cyt-*b*	AY729880
			cyt-c_1	XM_532351
			core 1	XM_846116
			core 2	XM_536942
			ISP	XM_533711
Danio	rerio	Zebrafish	cyt-*b*	AC024175
			cyt-c_1	BC080245
			core 1	--
			core 2	BC097011
			ISP	BC059475
Gallus	gallus	Chicken	cyt-*b*	X52392
			cyt-c_1	--
			core 1	XM_414356
			core 2	BX950415
			ISP	AJ19669
Homo	sapiens	Human	cyt-*b*	DQ301818
			cyt-c_1	BC001006
			core 1	BC009586
			core 2	BC003136
			ISP	BC021057
Mus	musculus	House Mouse	cyt-*b*	AY057804
			cyt-c_1	NM_025567
			core 1	NM_025407
			core 2	BC003423
			ISP	NM_025710
Pan	troglodytes	Chimpanzee	cyt-*b*	D38113
			cyt-c_1	XM_528263
			core 1	XM_516440
			core 2	XM_523485
			ISP	AY387497S2
Rattus	norvegicus	Norway Rat	cyt-*b*	X14848
			cyt-c_1	XM_001072221
			core 1	BC078923
			core 2	BC083610
			ISP	NM_001008888
Xenopus	laevis	African clawed frog	cyt-*b*	M10217
			cyt-c_1	BC045127
			core 1	BC098177
			core 2	BC106252
			ISP	BC041528