Overexpression and Characterization of the Copper A Domain from Cytochrome ba_3 of *Thermus thermophilus*

Thesis by

Claire Ellen Slutter

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1996

(Submitted February 8, 1996)

ACKNOWLEDGMENTS

ABSTRACT

Recently, the genes of cytochrome ba_3 from *Thermus thermophilus* [Keightley, J. A. *et al.* (1995) *J. Biol. Chem.* 270 20345-20358], a homolog of the heme-copper oxidase family, have been cloned. We report here expression of a truncated gene, encoding the copper A (Cu_A) domain of cytochrome ba_3, downstream from the T7 RNA polymerase promoter in *Escherichia coli*. The Cu_A domain is obtained in high yields as a water-soluble, thermostable, purple copper protein. The absorption spectrum of the Cu_A site, free of the heme interference in cytochrome ba_3, is similar to the spectra of other soluble fragments from the aa_3-type oxidase of *Paracoccus denitrificans* [Lappalainen, P. *et al.* (1993) *J. Biol. Chem.* 268 26416-26421] and the caa_3-type oxidase of *Bacillus subtilis* [von Wachenfeldt, C. *et al.* (1994) *FEBS Lett.* 340 109-113]. There are intense bands at 480 nm (3,100 M^{-1} cm^{-1}) and 530 nm (3,200 M^{-1} cm^{-1}), a band in the near-IR centered at 790 nm (1,900 M^{-1} cm^{-1}) and a weaker band at 363 nm (1,300 M^{-1} cm^{-1}). The secondary structure prediction from the far-UV CD spectrum indicates that this domain is predominantly β-sheet, in agreement with the recent X-ray structure reported for the complete *P. denitrificans* cytochrome aa_3 molecule [Iwata, S. *et al.* (1995) *Nature* 376 660-669] and the engineered, purple CyoA protein [Wilmanns *et al.*, *Proc. Natl. Acad. Sci. USA*, in press]. Soluble Cu_A fragments from other terminal oxidases have been expressed; however, the thermostability of the fragment described here (T_m = 80 °C) and the stable binding of copper over a broad pH range (pH 3-9) makes this protein uniquely suitable for detailed physical-chemical study. Copper analysis by chemical assay, mass spectrometry, X-ray fluorescence, and EPR spin quantification all indicate that this protein contains two copper ions bound in a mixed-valence state, consistent with the prediction that the Cu_A site in cytochrome ba_3 is a binuclear center.

Flash photolysis has been used to initiate electron transfer from excited tris(2,2'-bipyridyl)ruthenium(II) to the Cu_A site of the soluble *Thermus* domain. Luminescence

quenching of the excited state of the ruthenium(II) complex was observed at low protein concentrations (20-200 μM Cu_A domain), with second-order kinetics and rate constants of 2.9×10^9 $M^{-1}s^{-1}$ and 1.3×10^9 $M^{-1}s^{-1}$ at low and high ionic strength, respectively, At high protein concentrations (>250 μM Cu_A) and low ionic strength, the quenching rate saturates due to ground-state complex formation; a first-order rate constant of 1.5×10^5 s^{-1} was estimated for ET in the complex.

TABLE OF CONTENTS

Chapter 3

Appendix

LIST OF TABLES AND FIGURES

LIST OF ABBREVIATIONS

1D	one-dimensional
BCA	bicinchoninic acid
BCS	bathocupreine sulfonate
CDBLI	corrected double integral
CD	circular dichroism
CCO	cytochrome *c* oxidase
Cu_A	copper A
CV	cyclic voltammetry
CyoA	cytochrome *o* quinol oxidase
cyt *c*	cytochrome *c*
ε	extinction coefficient
ε_l	extinction coefficient-left (polarized light)
ε_r	extinction coefficient-right (polarized light)
EDTA	ethylenediaminetetraacetic acid
EPR	electron paramagnetic resonance
ESI-MS	electrospray ionization mass spectrophotometry
ET	electron transfer
FPLC	fast protein liquid chromatography
GHz	gigahertz
ICP-MS	induction coupled plasma mass spectrometry
IEF	isoelectric focusing
kb	kilobase
K_d	dissociation constant
kDa	kilodalton

LB	Luria Bertani media
LMTC	ligand to metal charge transfer
min	minutes
ms	milliseconds
μs	microseconds
mV	millivolt
MW	molecular weight
NHE	normal hydrogen electrode
nm	nanometers
NMR	nuclear magnetic resonance
N_2OR	nitrous oxide reductase
OD	optical density
P. aeruginosa	*Pseudomonas aeruginosa*
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
P. denitrificans	*Paracoccus denitrificans*
pI	isoelectric point
rpm	revolutions per minute
RR	Raman resonance
$Ru(bpy)_3^{2+}$	tris (2,2'-bipyridyl)ruthenium(II)
S	spin
SCE	saturated calomel electrode
SDS	sodium dodecylsulfate
TCA	trichloroacetic acid
Tris	tris-hydroxymethyl aminomethane
T_m	melting temperature

T. thermophilus	*Thermus thermophilus*
T. versutus	*Thiobacillus versutus*
TXRF	total-reflection X-ray fluoresence
UV/vis	ultraviolet/visible
WT	wild-type

You should never be afraid of failure when you try new or difficult things. Because no man can be an absolute failure at anything. If nothing else, he can always serve as a poor example. - Roger George Slutter (1931-1989)

Chapter 1

Overview:

The Binuclear Copper A (Cu_A) Site

BACKGROUND AND SIGNIFICANCE

Overview of aerobic respiration

Dioxygen (O_2) currently accounts for ~20% (by volume) of the atmosphere; however, this free O_2 is a relative latecomer. The original atmosphere is believed to have formed from outgassing of compounds trapped during the rock aggregation process that formed the earth ~ 4.5 billion years ago. Outgassing still occurs; volcanoes, for example, still release large quantities of water vapor and some hydrogen, hydrogen chloride, carbon monoxide, carbon dioxide, nitrogen and sulfur compounds. A portion of the O_2 in this original atmosphere may have reformed from the solar radiolysis products of outgassed H_2O. But, it is believed that most of the O_2 was released by the first photosynthesizing prokaryotes. Margulis *et al.* (1976) have proposed that these organisms were ancestors of the modern cyanobacteria. This atmospheric change is dated between 2.4 and 2.8 billion years ago, marked in the geologic record by the appearance of significant amounts of red, oxidized iron in rock strata (Knoll, 1992). Figure 1.1 summarizes some general features of this speculative picture of earth's history, as well as the important interrelationship between geological composition and evolution.

Ironically, O_2 was initially an atmospheric pollutant created by a successful organism adapted to thrive on water, carbon dioxide and light. Moreover, this radical shift from a reducing atmosphere to a more reactive, oxidizing atmosphere was not a benign change for these organisms. The unregulated presence of partially reduced O_2 in a cell, leading to superoxide ($O_2^{\cdot -}$), hydrogen peroxide (H_2O_2) and hydroxyl radical ($OH^{\cdot}$) formation, has well characterized toxic effects (Rodgers & Powers, 1981). Arguably this pollutant was as antagonistic to the unadapted extant life as current pollution problems.

Aside from these oxygen toxicity problems which necessitated some adaptation, O_2 has enormous advantages over other terminal electron acceptors (Babcock &

Wikström, 1992). Dioxygen has been described as the ideal terminal oxidant since no other stable electron acceptor can provide more free energy to the cell (George, 1965). For this reason, yeast cells--eukaryotes which can utilize both anaerobic an aerobic pathways--consume less food when the more efficient aerobic respiration is activated. Dioxygen is also ideal for use as a reactant in enzymatically catalyzed systems because it is kinetically controllable. The relative sluggishness of single electron reductions is attributed to the large differences in bond energy between O_2 and the superoxide radical $O_2{\cdot}^-$. Also, the ground electronic triplet state of O_2 (two unpaired electrons) restricts reactivity with singlet-state two electron donors, a consequence of spin restrictions. In fact, if this spin restriction is removed (spin-paired electrons), as is the case for singlet oxygen, a dangerously reactive molecule results that is very destructive to the cell.

The complete oxidation of glucose in aerobic respiration can be summarized by the following simple redox equation.

$$C_6H_{12}O_6 + 6\ O_2 \longrightarrow 6\ CO_2 + 6\ H_2O \qquad \Delta G^{\circ\prime} = -2823\ \text{kJ-mol}^{-1}$$

This reaction can be subdivided into two biologically relevant half-reactions. The oxidation of glucose catalyzed by the glycolytic pathway and the citric acid cycle; and, the reduction of O_2 to H_2O.

$$C_6H_{12}O_6 + 6\ H_2O \longrightarrow 6\ CO_2 + 24\ H^+ + 24\ e^-$$

$$6\ O_2 + 24\ H^+ + 24\ e^- \longrightarrow 12\ H_2O$$

The electrons generated in glucose oxidation are not transferred directly to O_2, but are instead used to reduce NAD^+ and FAD, forming 10 NADH and 2 $FADH_2$ molecules per oxidized glucose molecule. NADH and $FADH_2$ are then regenerated after transferring electrons to complex I and complex II, respectively, transmembrane proteins of the electron-transport chain (See Figure 1.2). The oxidation of NADH has a $\Delta G^{\circ\prime} = -218$ kJ mol^{-1}; more than sufficient to drive synthesis of three ATP molecules from ADP and P_i ($\Delta G^{\circ\prime} = 30.5$ kJ mol^{-1}). However, the oxidation of $FADH_2$ yields only two ATP molecules; thus Complex II simply recycles FAD, injecting the released electrons into the

transport chain. The electrons (from NADH and $FADH_2$) are transferred to a series of redox centers with progressively greater electron affinity (more positive reduction potentials) until they reach the terminal electron acceptor in the chain, O_2. Complex IV- more commonly called the terminal oxidase- catalyzes this reduction of O_2 to H_2O.

Because ATP synthase (complex V) is a distinct enzyme from those in the electron transport chain, energy most be stored in a form which is accessible to it. The coupling of electron transfer to proton pumping was originally described by Mitchell in his chemiosmotic hypothesis (Mitchell, 1961). Here the generation of a proton gradient across the membrane, analogous to a biological capacitor, was proposed as the method of energy storage. The transmembrane complexes of the electron transport chain are able to efficiently utilize the liberated free energy of electron transfer to pump protons into the periplasmic space (or mitochondrial space). The free flow of protons back to the cytoplasmic side of the membrane is then used as an energy source for ATP synthetase, permitting the synthesis of 36 ATP's per glucose molecule.

Important structural/function features of cytochrome *c* oxidase

Cytochrome *c* oxidase, one of several terminal oxidase types, catalyzes the oxidation of four consecutive ferrocytochrome *c* molecules and the concurrent reduction of dioxygen to water. These oxidases range in complexity from the two-subunit, *Thermus thermophilus* cytochrome ba_3 (Keightley *et al.*, 1995) to the 13-subunit bovine cytochrome *c* oxidase. However, in all cases, the three redox centers of these enzymes are located in two subunits- the functional core of the oxidases. Subunit I contains the bimetallic active site where O_2 is bound and reduced; and subunit II contains the Cu_A domain, anchored to the periplasmic (or mitochondrial space) side of the membrane with one or, more typically, two transmembrane helices.

Based upon sequence analysis, Castresana *et al.* (1994) have proposed that all known terminal oxidases evolved from one common ancestor called uroxidase. The

phylogenetic tree for subunit I of known oxidases is shown in Figure 1.3. Several functions for uroxidase have been suggested, including the possibility that uroxidase may have limited oxygen toxicity by scavenging electrons to reduce O_2 to H_2O, may have reduced a more abundant, but structurally similar oxidant like nitric or nitrous oxide, or may have functioned in an oxygen-poor atmosphere analogous to the *Bradyrhizobium japonicum* oxidase in root nodules (Preisig *et al.*, 1993).

In eukaryotes, the origin of mitochondria are traced to an endosymbotic event between an ancestor of the Alpha subclass of Proteobacteria (Margulis, 1981; Yang *et al.*, 1985; Cedergren *et al.*, 1988) and an eubacterium capable of oxidative phosphorylation and/or photosynthesis. Most of the genes of this enveloped bacterium were eventually transferred to the nucleus of the eukaryotic host; however, several remain, including the subunit I, II and III genes of cytochrome *c* oxidases. The number of nuclear-encoded subunits associated with the mitochondrial functional core increases as the complexity of the organism increases- 4 in *Dictyostelium discoideum* (Bisson and Schiavo, 1986), 5 in *Neurospora crassa* (Werner, 1977), 6 in yeast (Power *et al.*, 1984), 7 to 9 in birds and fish (Montecucco *et al.*, 1987) and 10 in mammals (Kadenbach *et al.*, 1983, 1987 for review). The function of these subunits is believed to be largely regulatory; in fact, the more complex oxidases have been shown to be catalytically depressed relative to their bacterial relatives (for review see Kadenbach *et al.*, 1991). Specifically, ATP, ADP, phosphate and free fatty acids are known to modulate activity, but these interactions have yet to be localized to specific subunits. Changes in the conformation of the catalytic core upon binding an allosteric effector has been the hypothesized mechanism of regulation; however, obtaining structural details of these changes has not been possible until very recently. One noteworthy exception is the case of ATP and ADP. ATP and ADP are known to increase and decrease, respectively, the K_M of ferrocytochrome *c* binding for the bovine oxidase (Hüther & Kadenbach, 1988). Bacterial oxidases (that lack regulatory subunits) are typically run as controls in these kinetics experiments. In this study, the rate

of ferrocytochrome *c* oxidation and the binding of ferrocytochrome *c* by *P. denitrificans* oxidase was found to be unaffected by the presence of ATP or ADP.

Despite the fact that cytochrome *c* oxidases are the most studied redox-linked proton pumps, relatively little was known about the mechanistic details of proton pumping until the report of two X-ray structures in 1995- the aa_3-type *Paracoccus denitrificans* (Iwata *et al.*, 1995) and bovine cytochrome *c* oxidases (Tsukihara *et al.*, 1995). Previously, the low-spin cytochrome *a* and Cu_A sites have been suggested as the point of redox linkage (Chan & Li, 1990). The quinol oxidase, cytochrome *o*, receives electrons from the membrane soluble ubiquinol and menaquinol carriers; therefore, it lacks the Cu_A site found in most cytochrome *c* utilizing enzymes. Additionally, these oxidases lack the formyl group of the cytochrome *a*. Since the *P. denitrificans* oxidase and the *E. coli* quinol oxidase share considerable sequence homology, it is likely that a similar proton pumping mechanism is used, thus casting doubt on redox-linkage at these two sites.

A long-range conformational coupling mechanism has also been discussed in the literature. But, this conformational pumping mechanism is not strongly supported by sequence homology information either (for reviews see Morgan *et al.*, 1994; Babcock & Wikström, 1992). There are three major types of terminal oxidases: the cytochrome *c* oxidases of the aa_3-type, the quinol oxidases of the cytochrome bo_3-type and the more divergent cytochrome cbb_3-type oxidases that accept electrons from cytochrome *c* but lack a Cu_A site. All of these oxidases have been shown to pump protons, yet only the six histidines ligands to the two hemes and Cu_B centers in SUI are conserved (van der Oost *et al.*, 1992; Hosler *et al.*, 1993). If there were a large number of residues participating in conformational changes, more conserved amino acids would be expected.

The *P. denitrificans* structure clearly shows the presence of two proton channels near the bimetallic active site of subunit I (SUI)- strong evidence that the binuclear active site participates in proton pumping. The presence of two separate proton channels has been suspected since the discovery that the SUI D124—›N mutation blocks proton

pumping but not water formation (Thomas *et al.*, 1993; Garcia-Horseman *et al.*, 1995). Additionally, proposed direct redox linkage mechanisms with the active site required that the pumped (or vectorial) protons are positioned in the channel before the scalar protons reach the oxygen intermediate in the active site, thus suggesting two separate channels.

Iwata *et al.* (1995) have described a proton pumping mechanism which is based upon a histidine cycle mechanism proposed by Morgan *et al.* (1994). Here protonation/deprotonation of a conserved Cu_B histidine ligand, SUI-H325 links pumping to the oxygen chemistry occurring in the active site. Like the previous histidine cycle proposal, the SUI-H325 is suggested to cycle through the imidazolate, imidazole and imidazolium states (Figure 1.4). In the oxidized or O-state, SUI-H325, is in the imidazolate state and bound to Cu_B (Figure 1.4A). Here the negatively charged imidazolate is stabilized by the positively charged Fe and Cu ions in the bimetallic site and hydrogen bonding from SUI-T344. The first electron transferred to the Cu_B ion converts the imidazolate to the neutral imidazole form (Figure 1.4B). Now the imidazole ring is a hydrogen donor to SUI-T344. The second electron transferred to Cu_B results in simultaneous ligand exchange of the SUI-H325 and uptake of another proton, resulting in an imidazolium state of the histidine which is bound to a second site (Figure 1.4C). The imidazolium ring also needs to rotate during this step to orient the two protons toward the periplasmic side of the membrane. Dioxygen then binds between the Fe and Cu. A third electron converts the active site into the peroxy-state (Figure 1.4D). In order to retain an electroneutral site, another proton needs to bind to another residue in the vicinity. SUI-E278 has been suggested as the most likely proton acceptor. The peroxy-state is then double protonated, requiring that the two imidazolium protons must be released to the periplasmic side of the membrane (Figure 1.4E). This release of water yields the oxoferryl state (Figure 1.4F). After release of the two protons, the SUI-H325 accepts the proton from SUI-E278 and binds to Cu_B and SUI-T344 (Figure 1.4G). Transfer of another electron results in the formation of water and the fully protonated pump again releases two

protons similar to Figure 1.4C.

The general features of this mechanism are supported by available electron transfer data in the cytochrome *c* oxidases (see Babcock & Wikstrom, 1992 for a review). Two fast electron transfers to the active site are essential in trapping the dioxygen in the peroxy form. It has been demonstrated that O_2 does not bind to the ferrous heme until both of these electrons have been transferred to the active site (Lindsay *et al.*, 1975; Malatesta *et al.*, 1990). Once the peroxy state forms, the oxygen affinity of the site increases to $K_M \sim 10^{-7}$ M. This trapping chemistry and increased oxygen affinity is important since subsequent electron transfer steps are accompanied by proton pumping activity (i.e., conformational changes) and require more time.

The third and fourth electron transfers to the active site are responsible for initiating two cycles of proton release to the periplasm, supporting pumping mechanisms which are asymmetric- translocating four vectorial protons during the final two electron transfers. Additionally, it has been demonstrated by reverse electron flow experiments that the translocation of two protons occurs during the peroxy to ferryl transition and, again, during the ferryl to oxidized transition (Wikström, 1989). The overall kinetics of the oxidases studied thus far appear to be controlled by these rate limiting third and fourth electron transfers. In these peroxy and ferryl states, the slowest ET rate is 2×10^2 s^{-1} and appears to be rate limiting; this is very close to the k_{cat} for the enzyme (300 s^{-1}; however, this rate is a function of experimental conditions; van Kuilenburg *et al.*, 1992).

THE PURPLE COPPER (Cu_A) CENTER

Important structural/function features of the Cu_A site

All of the available structural and electron transfer data suggests that the Cu_A site functions as the initial electron transfer acceptor in the cytochrome *c* oxidases and is not involved directly in proton translocation. Rapid electron injection into the oxidized

cytochrome *c* oxidase always occurs through the Cu_A site (Pan *et al.*, 1993) and is followed by rapid equilibration with cytochrome *a* regardless of the electron donor used in the experiment (Nilsson, 1992; Kobayashi *et al.*, 1989; Morgan *et al.*, 1989; Pan *et al.*, 1993; Brzezinski *et al.*, 1995 and references therein). However, the injection rate to the Cu_A site, while detectable, is too fast to measure. The reduction state of the bimetallic active site is known to effect the redox potential of cytochrome *a*, making it a stronger oxidant in the oxidized enzyme (Ellis *et al.*, 1986; Blair *et al.*, 1986). The Cu_A—>cytochrome *a* rate is two-fold greater in the a_3-Cu_B oxidized enzyme (2.0×10^4 vs. 1.0×10^4 s^{-1} for the half reduced oxidase; for a review see Winkler *et al.*, 1995).

In addition to the oxidase structures, a higher resolution (2.5 Å) structure of a soluble fragment containing the purple center engineered into subunit II of the *E. coli* quinol oxidase (CyoA), is also available (Wilmanns *et al.*, 1995). The Cu_A site is a binuclear Cu site with two bridging thiolate ligands, two terminal histidine ligands and a weak ligand to each Cu- a methionine sulfur and a main chain carbonyl from a glutamic acid residue. (See Figure 1.5.) This ligand arrangement and cupredoxin fold is comparable to that found in the type 1 (blue copper) proteins. Figure 1.6 shows a ribbon structure comparison of *Populus nigra* plastocyanin and the CyoA fragment. Previously, sequence homology studies of the purple copper domain of the oxidases and blue (type 1) copper proteins have indicated that there is a close evolutionary relationship between these proteins (van der Oost *et al.*, 1992). And in addition to the CyoA fragment (Kelly *et al.*, 1993), several workers have demonstrated that a purple center can be engineered into azurin (Hay *et al.*, 1995) and amicyanin (Dennison, *et al.*, 1995) by addition of a cysteine ligand.

The unique spectral features of Cu_A, which have puzzled investigators for more than three decades (see Antholine *et al.*, 1992; Gurbiel *et al.*, 1993; Farrar *et al.*, 1995 for references to early work) can now begin to be understood in terms of the cluster-like structure described above. Unlike a structurally similar $(Cys)_2[2Fe\text{-}2S](Cys)_2$ center,

which shows strong valence *localization* (Sands & Dunham, 1975), the Cu_A center unexpectedly behaves as a spin-delocalized, mixed-valence structure in which the single unpaired electron is found with equal probability at either metal. This would correspond to a Class III complex in the scheme of Robin and Day (1967). (Interestingly, there is no direct experimental evidence for the formal Cu(II)/Cu(I) valence states, although this is a reasonable starting assumption.) This property was first demonstrated in N_2OR (Antholine *et al.*, 1992) and was confirmed for cytochrome ba_3-Cu_A by Fee *et al.*, 1995. Thus, spin (or valence) delocalization appears to be a fundamental property of the Cu_A center that distinguishes it from the Fe/S clusters of ferredoxins.

Studies of electron tunneling in ruthenium-modified azurins (Langen *et al.*, 1995; Regan *et al.*, 1995), taken together with the discovery that the Cu_A domain has a cupredoxin fold (van der Oost *et al.*, 1992; Wittung *et al.*, 1994), lead to the tempting speculation that good electronic coupling over these distances would be mediated along the β-strands. Therefore, the fact that the Cu_A domain is not oriented properly to take advantage of this possibility is somewhat surprising. Ramirez *et al.* (1995) have proposed an electron transfer (ET) pathway in the *P. denitrificans* structure. There is a direct ET pathway from Cu_A to the cytochrome *a* center composed of 14 covalent bonds and 2 hydrogen bonds. ET from cytochrome *c* to the a_3-Cu_B site is proposed to occur through the SUII-H181 ligand of the Cu_A site (where cytochrome *c* binds), through the Cu_A center to the SUII-H224 ligand, through a hydrogen bond to loop XI-XII of SUI and to the heme *a* propionate by a hydrogen bond. ET to the bimetallic site then proceeds across helix X as shown in Figure 1.7. The calculated Cu_A—›cytochrome *a* ET rate for this pathway is between 4×10^4 and 8×10^5 s^{-1}. These rates limit the ET reorganization energy (λ) to 0.15 and 0.5 eV, respectively. By comparison to other published λ values for protein reactions, 0.7 to 1.3 eV (Bjerrum *et al.*, 1995; Winkler *et al.*, 1992), this predicted λ is quite low. But, this may be reasonable for an intramolecular, solvent inaccessible pathway coupled through a uniquely delocalized Cu_A site. Also, Larsson and co-workers (1995) have used a

theoretical analysis based on electronic spectroscopic properties to predict that λ should decrease from about 0.4 eV in a mononuclear site to 0.2 eV in a purple center.

A regulatory role for Cu_A?

There are two evolutionary questions which pertain to a discussion of the functional role of the Cu_A site. The first--What kind of biological diversity exists in the terminal oxidases?--has been instrumental in eliminating mechanisms with direct Cu_A or cytochrome *a* participation in proton pumping. However, there is an equally important evolutionary question which has yet to be fully considered. Namely, which type of terminal oxidase is the "best"? Clearly, the efficient and highly regulated aa_3-type oxidases exclusively found in the most complex organisms are the "best" terminal oxidases. Since these proteins contain Cu_A sites, it is reasonable to suspect that this particular initial electron acceptor has an advantage over others. It is probable, although highly speculative at this point, that this regulatory role is important for improving the efficiency of the aerobic respiration process.

In the membrane there are many oxidases operating independently, creating a distribution of enzymes at different stages of the dioxygen reduction cycle. Efficiency mandates to discrimination between an oxidase which requires rapid electron transfer for oxygen trapping (the first two electrons or high-gear) and an oxidase which requires slower electron injection (the last two electrons or low-gear) as it pumps protons. The same electron injection rate, determined by unregulated binding of ferrocytochrome *c*, would necessitate storing any queued electrons on the cytochrome *a* center. The problem with this approach is that there may be other oxidases that have not reached the peroxy state because the needed electrons are unavailable. Clearly, the optimal process would contain the smallest possible number of terminal oxidases working constantly. In fact, there is good evidence that oxidases are not this wasteful. In the peroxy and ferryl states, where the driving force is very high, the oxidation rate of cytochrome *a* ($2 \times 10^2\ s^{-1}$) is

slower than the electron back flow rate (2×10^5 s^{-1}; Oliveberg & Malmström, 1991).

Therefore, the efficiency of the electron-transport chain may be greatly enhanced by communication strategies which signal whether the enzyme is in high-gear or low-gear. This could be accomplished by a change in cytochrome *c* binding affinity (known to occur in mammalian oxidases with ATP and ADP addition) or a slower electron transfer rate to the cytochrome *a* site or both. The latter effect can easily be accomplished by increasing the λ of the Cu_A site. By binding each Cu in a three-coordinate environment with a fourth axial (weak) ligand, the protein environment can disfavor valence trapping (Ramirez *et al.*, 1995). Intriguingly, by moving between a fully delocalized and valence-trapped conformation, the Cu_A site could modulate its λ, altering electron transfer to the cytochrome *a*.

Recently, Zickermann *et al.*, 1995, have shown that mutation of the Cu_A methionine ligand (SUII-M227) in the *P. denitrificans* oxidase to isoleucine generates a valence-trapped site with a slow Cu_A—›cytochrome *a* into the oxidized enzyme. This mutation illustrates the dramatic affect of perturbing a weak ligand around the Cu_A site. (A more detailed discussion of this topic can be found in Chapter 3.) These affects could also be caused by conformational changes of SUI or regulatory subunits, switching from a delocalized (fast injection) to a valence trapped (slow injection) site.

Another less drastic modification would be the use of these weak ligands in changing the redox potential. Guckert *et al.*, 1995, have suggested that the weak ligand distances may be important in fine-tuning the redox potentials of the blue copper proteins. Stellacyanin, for example, has the lowest redox potential (180 mV *vs.* NHE) of the small blue copper proteins, because it has the strongest axial bond. Here the methionine ligand, conserved in most blue copper proteins, is replaced by a glutamine, producing a stronger Cu(II)-ligand bond which stabilizes the oxidized state and lowers the redox potential. This ease of altering the ET properties of the Cu_A site suggests the very interesting possibility that the cupredoxin fold of the Cu_A domain could be functioning as a rather sophisticated

injection valve. Moreover, the allosteric control potential of this initial electron acceptor with additional nuclear-encoded subunits could explain its prevalence in the complex mammalian oxidases.

OVERVIEW OF THE THESIS PROJECT

Chapter 2

The study of soluble fragments containing only the purple copper center- separated from the spectral interference of the hemes in oxidases- has contributed substantially to an understanding of this site. When this project began, there were two soluble native Cu_A domains available from *Paracoccus denitrificans* (Lappalainen *et al.*, 1993) and *Bacillus subtilis* (von Wachenfeldt *et al.*, 1994). Additionally, the engineered purple site in the soluble subunit II domain of the *E. coli* quinol oxidase, the CyoA fragment (van der Oost, *et al.*, 1992; Kelley *et al.*, 1993) was available. Characterization of these fragments revealed for the first time the very interesting UV/Vis absorption spectrum of the Cu_A site. In the oxidases, only the far infrared (IR) band centered at 790 nm was visible. These soluble fragments revealed the presence of the unique double peak feature near 500 nm. The bands of nearly equal intensity have been attributed to exciton splitting in the binuclear site of the thiolate-Cu charge-transfer band (Larsson *et al.*, 1995; Farrar *et al.*, 1991) found in the type 1 copper proteins near 600 nm. This UV/Vis spectrum, unique to the purple centers, provided tantalizing evidence that the Cu_A site was *not* another mononuclear redox center. However, more definitive experiments which could, for example, clarify the mononuclear *vs.* binuclear debate were hampered by stability problems. Simple but critical data, like copper quantitation and redox potential measurements, were not straightforward due to pesky stability problems and copper loss upon reduction. For this reason, we decided to construct a soluble, thermostable fragment from *Thermus thermophilus*. Two engineered purple centers in the small blue copper proteins azurin (Hay *et al.*, 1995) and amicyanin (Dennison *et al.*, 1995) were

constructed. However, the *T. thermophilus* Cu_A fragment which is described in Chapter 2 remains the most stable of this group and the best candidate for detailed biophysical analysis.

Chapter 2 presents clear evidence that this fragment contains 2 Cu's per protein molecule. This technical advantage has permitted the combination of Cu quantitation and other characterization techniques. Chapter 2 presents extinction coefficients for the transitions in the absorption spectrum; the first visible CD spectrum for this site; the first assignment of the hyperfine EPR structure for the Cu_A site; the first report (in a native not engineered purple center) of paramagnetically shifted protons near the active site and the unusual temperature dependence of the chemical shifts of these protons. Importantly, these ^{1}H-NMR experiments also demonstrate that this fragment is not plagued by the aggregation problems that have hampered NMR experiments with other fragments.

Since this work began, structural information (discussed earlier) has become available, showing a binuclear site; but the interest in Cu_A is far from over. The unique electronic properties of this site and their functional implications are not fully understood. In addition, the electron transfer properties of the Cu_A site have not been closely examined in the literature. Here too, data from a stable soluble fragment has much to contribute.

Chapter 3

Chapter 3 contains a description of the first electron transfer experiments performed on this fragment. Here, excited $Ru(bpy)_3^{2+}$ has been used to reduce the oxidized Cu_A fragments. This demonstration that this fragment can be reversibly oxidized and reduced without loss of copper is important for more spohisticated ET experiments. Work is continuing using ruthenated-Cu_A (Winkler & Gray, 1992) in second-generation experiments (DiBilio, unpublished results).

Preliminary binding experiments have also been included. Since this work was completed, structural information (Iwata *et al.*, 1995) has been published supporting the suggestion that a 10 acidic residue patch on subunit III could form a portion of the

cytochrome *c* binding site in the *P.denitrificans* oxidase (Reider & Bosshard, 1980). Thus, the horse heart cytochrome *c* used in the binding experiments may not have a native docking site which is quite different from the *T. thermophilus* cytochrome *c*. The natural cytochrome ba_3 substrate binds only to subunit II (this oxidase has only two subunits), making these unnatural partners less than ideal. Moreover, kinetics studies of the cytochrome ba_3 enzyme with the *T. thermophilus* cytochrome *c* and horse heart cytochrome *c* substrates give substantially faster rates when the natural partners are used (Fee, J. A., personal communication). A discussion of the original binding experiments is included in Chapter 3 since the spectral perturbation of the heme is interesting, if not entirely explainable, (Spectral changes are not typical of cytochrome *c* binding interactions.) Consequently, the binding results should be examined with some caution since they may not represent specific binding.

Another important advantage of the *T. thermophilus* fragment, namely that electrochemical techniques can be used to directly measure the redox potential of this fragment, is discussed. Accurate redox potential measurements are important for future ET experiments designed to measure the reorganization energy of the site. Additionally, theoretical (Guckert *et al.*, 1995) work indicating that both the λ and the redox potential can be modulated in this site are discussed.

REFERENCES

Antholine, W. E., Kastrau, D. H. W., Steffens, G. C. M., Buse, G., Zumft, W. G. & Kroneck, P. H. M. (1992) *Eur. J. Biochem. 209*, 875-888.

Babcock, G. T. & Wikström, M. (1992) *Nature 356*, 301-307.

Bisson, R. & Schiavo, G. (1986) *J. Biol. Chem. 261*, 4373-4376.

Bjerrum, M. J., Casimro, D. R., Chang, I.-J., DiBilio, A. J., Gray, H. B., Hill, M. G., Langen, R., Mines, G. A., Skov, L. K., Winkler, J. R. & Wuttke, D. S. (1995) *J Bioenerg. Biomembr. 27*, 295-302.

Blair, D. F., Gelles, J. & Chan, S. I. (1986) *Biophys. J. 50*, 713 - 733.

Brzezinski, P., Sundahl, M., Ädelroth, P., Wilson, M. T., El-Agez, B., Wittung, P. & Malmström, B. G. (1995) *Biophys. Chem.* 54, 191-197.

Castresana, J., Lübben, M., Saraste, M. & Higgins, D. G. (1994) *EMBO J. 13 (11)* 2516-2525.

Cedergren, R., Gray, M. W., Abel, Y. & Sankoff, D. (1988) *J. Mol. Evol. 28*, 98-112.

Chan, S. I. & Li, P. M. (1990) *Biochemistry 29*, 1-12.

Dennison, C., Vijenboom, E., de Vries, S., van der Oost, J. & Canters, G. (1995) *FEBS Lett. 365*, 92-94.

Ellis, W. R., Wang, H., Blair, D. F., Gray, H. B. & Chan, S. I. (1986) *Biochemistry 25*, 161-167.

Farrar, J. A., Lappalainen, P., Zumft, W. G., Saraste, M. & Thomson, A. J. (1995) *Eur. J. Biochem. 232*, 294-303.

Fee, J. A., Sanders, D., Slutter, C. E., Doan, P. E., Aasa, R., Karpefors, M. & Vänngård, T. (1995) *Biochem. Biophys. Res. Commun. 212*, 77-83.

Garcia-Horsman, J. A., Puustinen, A., Gennis, R. B. & Wikström, M. (1995) *Biochemistry 34*, 4428-4433.

George, P. (1965) in Oxidases and Related Redox Systems (eds. King, T. E., Mason, H.

S. and Morrison, M.) p. 3-36, Wiley, New York.

Guckert, J. A., Lowery, M. D. & Solomon, E. I. (1995) *J. Am. Chem. Soc. 117*, 2817-2844.

Gurbiel, R. J., Fann, Y.-C., Surerus, K. K., Werst, M. M., Musser, S. M., Doan, P. E., Fee, J. A. & Hoffman, B. M. (1993) *J. Am. Chem. Soc. 115*, 10888-10894.

Hay, M., Yi, L. & Richards, J. H. (1995) *Proc. Natl. Acad. Sci. USA 93*, 461-464.

Holster, J. P. *et al.* (1993) *J. Bioenerg. Biomembr. 25*, 121-136.

Hüther, F.-J. & Kadenbach, B. (1988) *Biochem. Biophys. Res. Commun. 153*, 525-534.

Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995) *Nature 376*, 660-669.

Kadenbach, B., Stroh, A., Hüther, F.-J., Reimann, A. & Steverding, D. (1991) *J. Bioenerg. Biomembr. 23 (2)*, 321-334.

Kadenbach, B., Jarausch, J., Hartmann, R. & Merle, P. (1983) *Anal. Biochem. 129*, 517-521.

Kadenbach, B., Kuhn-Nentwig, L. & Büge, U. (1987) *Curr. Top. Bioenerg. 15*, 113-161.

Keightley, J. A., Zimmermann, B. H., Mather, M. W., Springer, P., Pastuszyn, A., Lawrence, D. M. & Fee, J. A. (1995) *J. Biol. Chem.,270*, 20345-20358.

Kelly, M., Lappalainen, P., Talbo, G., Haltia, T., van der Oost, J. & Saraste, M. (1993) *J. Biol. Chem. 268*, 16781-16787.

Knoll, A. H. (1992) *Science 256*, 622-627.

Kobayashi, K., Une H. & Hayashi, K. (1989) *J. Biol. Chem.* 264, 7976-7980.

Langen, R., Chang, I.-J., Germanas, J. P., Richards, J. H., Winkler, J. H. & Gray, H. B. (1995) *Science 268*, 1733-1735.

Lappalainen, P., Aasa, R., Malmström, B. G. & Saraste, M. (1993) *J. Biol. Chem. 268*, 26416-26421.

Larsson, S., Källebring, B., Wittung, P. & Malmström, B. G. (1995) *Proc. Natl. Acad. Sci. USA 92*, 7167-7171.

Lindsay, J. G., Owen, C. S. & Wilson, D. F. (1975) *Archs. Biochem. Biophys. 169*, 492-505.

Malatesta, F., Sarti, P., Antonini, G., Vallone, B. & Brunori, M. (1990) *Proc. Natl. Acad. Sci. USA 87*, 7410-7413.

Malmström, B. G. (1994) *Eur. J. Biochem. 223*, 711-718.

Marcus, R. A. & Sutin, N. (1985) *Biochim. Biophys. Acta 811*, 265-322.

Margulis, L., Walker, J. C. G. & Rambler, M. (1976) *Nature 264*, 620-624.

Margulis, L. (1981) *Symbiosis in Cell Evolution*. W. H. Freeman and Company, San Francisco.

Mitchell, P. (1961) *Nature 191*, 144-148.

Montecucco, C., Schiavo, G., Bacci, B. & Bisson, R. (1987) *Comp. Biochem. Physiol. 87B*, 851-856.

Morgan, J. E., Li, P. M., Jang, D.-E., El-Sayed, M. A. & Chan, S. I. (1989) *Biochemistry 28,* 6975-6983.

Morgan, J. E., Verkhovsky, M. I. & Wikström, M. (1994) *J. of Bioeng. and Biomem. 26 (6)*, 599-608.

Nilsson, T. (1992) *Proc. Natl. Acad. Sci. USA 89*, 6497-6501.

Oliveberg, M. & Malmström, B. G. (1991) *Biochemistry 30*, 7053-7057.

Pan, L. P., Hibdon, S., Liu, R.-Q., Durham, B. & Millett, F. (1993) *Biochemistry 32*, 8492-8498.

Power, S. D., Lochrie, M. A., Sevarino, K. A., Patterson, T. E. & Poyton, R. O. (1984) *J. Biol. Chem. 259*, 6564-6570.

Preisig, O., Anthamatten, D. & Hennecke, H. (1993) *Proc. Natl. Acad. Sci. USA 90*, 3309-3313.

Ramirez, B. E., Malmström, B. G., Winkler, J. R. & Gray, H. B. (1995) *Proc. Natl. Acad. Sci. USA 92*, 11949-11951.

Regan, J. J., DiBilio, A.J., Langen, R., Skov, L. K., Winkler, J. R., Gray, H. B. &

Onuchic, J. N. (1995) *Chem. Biol. 2*, 489-496.

Rieder, R. & Bosshard, H. R. (1980) *J. Biol. Chem. 255*, 4732-4739.

Robin, M. & Day, P. (1967) *Advan. Inorg. Radiochem. 10*, 247-422.

Rodgers, M. A. J. & Powers, E. L., eds., (1981) *Oxygen and Oxy-radicals in Chemistry and Biology*, Academic, New York.

Sands, R. H. & Dunham, W. R. (1975) *Q. Rev. Biophys. 7*, 443-504.

Thomas, J. W. *et al.* (1993) *Biochemistry 32*, 10923-10928.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawaitoh, K., Nakashima, R., Yaono, R. &Yoshikawa S. (1995) *Science* 269 1069-1074.

van der Oost, J., Lappalainen, P., Musacchio, A., Warne, A., Lemieux, L., Rumbley, J., Gennis, R. B., Aasa, R., Pascher, T., Malmström, B. G. & Saraste, M. (1992) *EMBO J. 11*, 3209-3217.

van Kuilenburg, A. B. P., Gorren, A. C. F., Dekker, H. L., Nieboer, P., van Gelder, B. F. & Muijsers, A. O. (1992) *Eur. J. Biochem. 205*, 1145-1154.

von Wachenfeldt, C., de Vries, S. & van der Oost, J. (1994) *FEBS Lett. 340*, 109-113.

Werner, S. (1977) *Eur. J. Biochem. 79*, 103-110.

Wikström, M. (1989) *Nature 338*, 776-778.

Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E. & Saraste, M. (1995) *Proc. Natl. Acad. Sci. USA 92*, 11955-11959.

Winkler, J. R. & Gray, H. B. (1992) *Chem. Rev. 92*, 369-379.

Winkler, J. R., Gray, H. B. & Malmström, B. G. (1995) *Biophys. Chem. 54*, 199-209.

Wittung, P., Källebring, B. & Malmström, B. G. (1994) *FEBS Lett. 349*, 286-288.

Yang, D., Oyaizu, H., Olsen, G. J. & Woese, C. R. (1985) *Proc. Natl. Acad. Sci. USA 82*, 4443-4447.

Zickermann, V., Verkhovsky, M., Morgan, J., Wikström, M., Anemüller, S., Bill, E., Steffens, G. C. M. & Ludwig, B. (1995) *Eur. J. Biochem. 234*, 686-693.

Figure 1.1

A speculative history of the earth.

[Atkins, P. W. (1987) **Molecules**, p. 15, W. H. Freeman and Company, New York]

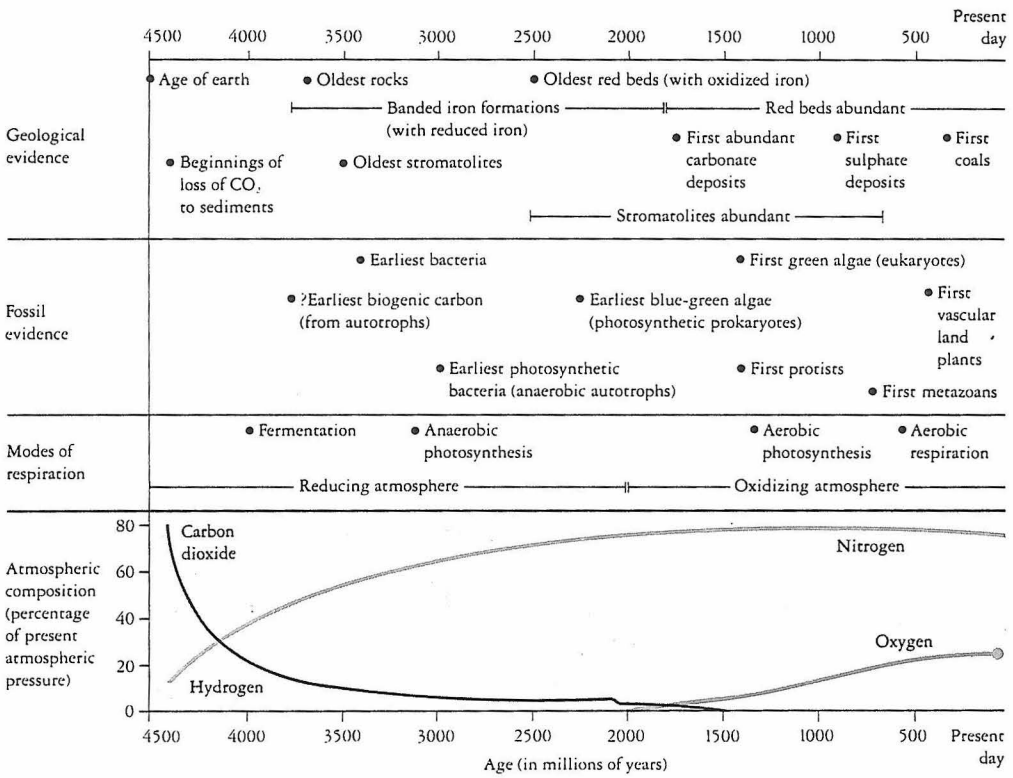

4500
4000
3500
3000
2500
2000
1500
1000
500
Present day
Geological evidence
Age of earth
Oldest rocks
Oldest red beds (with oxidized iron)
Banded iron formations (with reduced iron)
Red beds abundant
Beginnings of loss of CO_2 to sediments
Oldest stromatolites
First abundant carbonate deposits
First sulphate deposits
First coals
Stromatolites abundant
Fossil evidence
Earliest bacteria
First green algae (eukaryotes)
?Earliest biogenic carbon (from autotrophs)
Earliest blue-green algae (photosynthetic prokaryotes)
First vascular land plants
Earliest photosynthetic bacteria (anaerobic autotrophs)
First protists
First metazoans
Modes of respiration
Fermentation
Anaerobic photosynthesis
Aerobic photosynthesis
Aerobic respiration
Reducing atmosphere
Oxidizing atmosphere
Atmospheric composition (percentage of present atmospheric pressure)
80
60
40
20
0
Carbon dioxide
Nitrogen
Oxygen
Hydrogen
Age (in millions of years)

Figure 1.2

The ET events in the mitochondrial respiratory chain with complexes I-IV shown with their common names. Complex I, NADH-Q reductase; Complex II, succinate reductase; Complex III, cytochrome *c* reductase; Complex IV, cytochrome *c* oxidase; NADH, reduced nicotinamide adenine dinucleotide; $FADH_2$, reduced flavin adenine dinucleotide; FMN, flavin mononucleotide; Fe-S, iron sulfur proteins; Q, ubiquione.

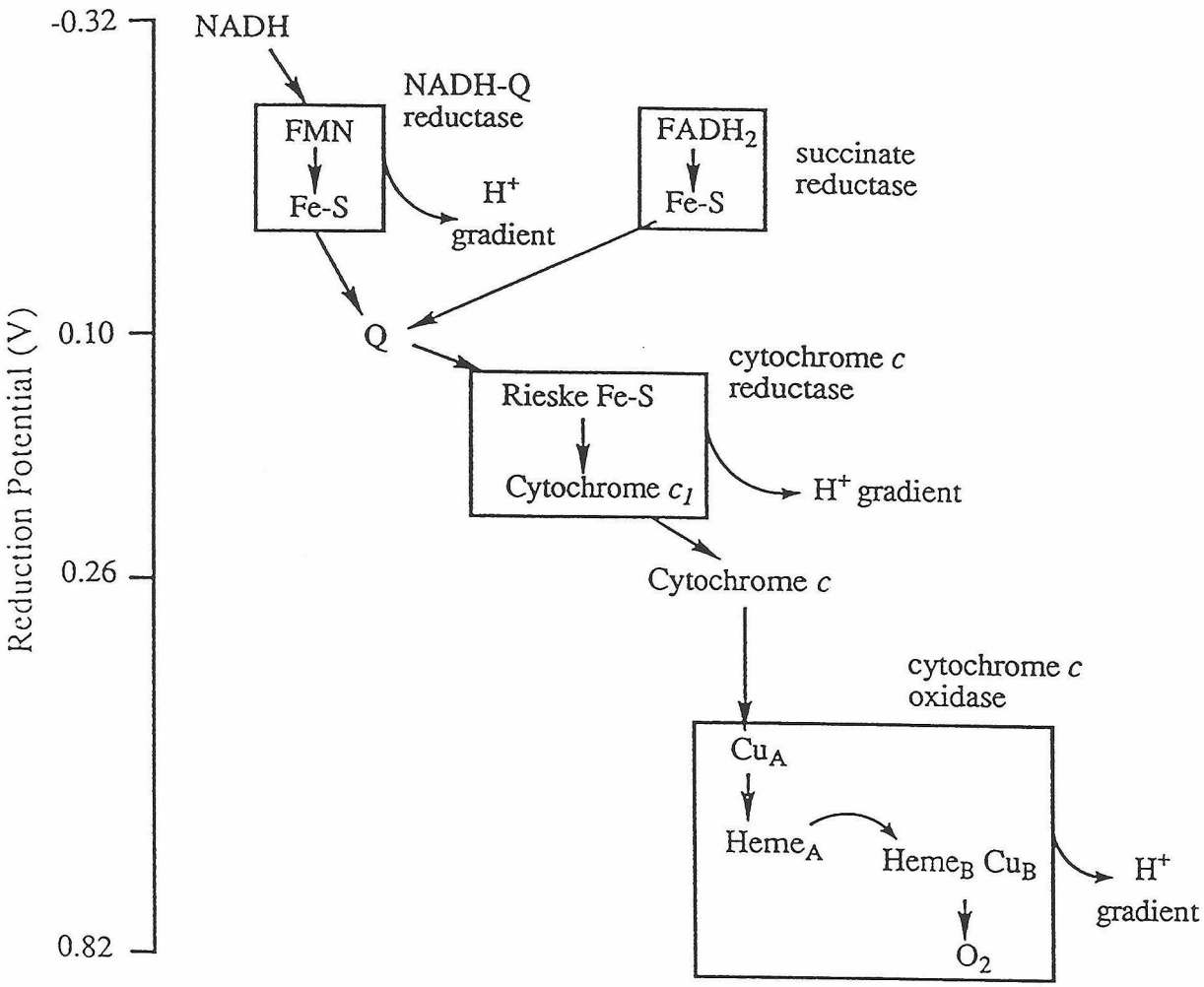

-0.32
0.10
0.26
0.82
Reduction Potential (V)
NADH
NADH-Q reductase
FMN
Fe-S
H+ gradient
FADH2
Fe-S
succinate reductase
Q
cytochrome c reductase
Rieske Fe-S
Cytochrome c1
H+ gradient
Cytochrome c
cytochrome c oxidase
CuA
HemeA
HemeB CuB
O2
H+ gradient

Figure 1.3

Phylogenetic tree of subunit I of oxidases inferred by the neighbor-joining method with Kimura's distance correction. Q, eubacteria quinol oxidases; cQ, archaebacterial quinol oxidases; Cyano, *Cyanobacteria*; Deinoc, *Deinococcaceae*; T, *Thermus*; Gram +, Gram-positive bacteria; Proteob, *Proteobacteria.* The bar represents a distance of 20%.

[Castresana, J., Lübben, M., Saraste, M., Higgins, D. G. (1994) **Evolution of cytochrome oxidase, an enzyme older than atmospheric oxygen**, *EMBO J. 13 (11)* 2516-2525]

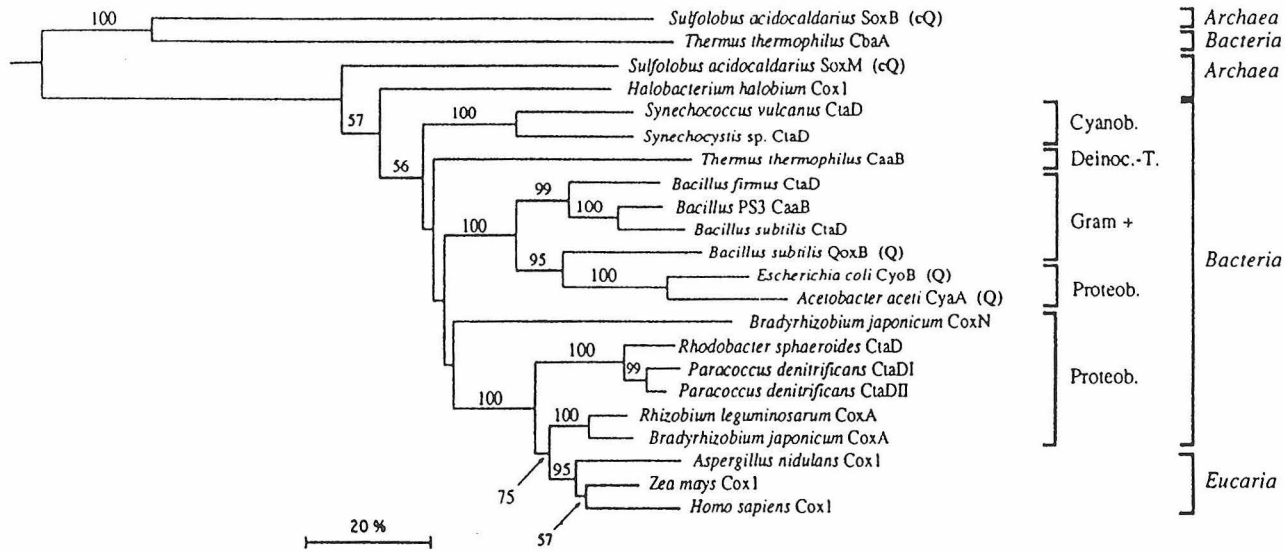
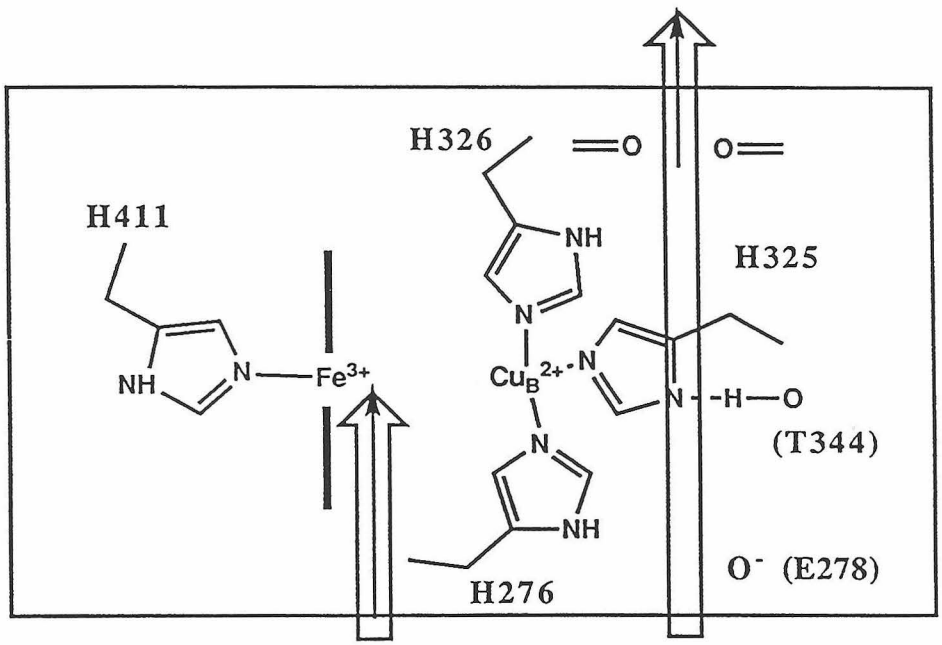

Sulfolobus acidocaldarius SoxB (cQ)
Thermus thermophilus CbaA
Sulfolobus acidocaldarius SoxM (cQ)
Halobacterium halobium Cox1
Synechococcus vulcanus CtaD
Synechocystis sp. CtaD
Thermus thermophilus CaaB
Bacillus firmus CtaD
Bacillus PS3 CaaB
Bacillus subtilis CtaD
Bacillus subtilis QoxB (Q)
Escherichia coli CyoB (Q)
Acetobacter aceti CyaA (Q)
Bradyrhizobium japonicum CoxN
Rhodobacter sphaeroides CtaD
Paracoccus denitrificans CtaDI
Paracoccus denitrificans CtaDII
Rhizobium leguminosarum CoxA
Bradyrhizobium japonicum CoxA
Aspergillus nidulans Cox1
Zea mays Cox1
Homo sapiens Cox1
Cyanob.
Deinoc.-T.
Gram +
Proteob.
Proteob.
Archaea
Bacteria
Archaea
Bacteria
Eucaria
20 %

Figure 1.4

The histidine cycle proton pumping mechanism. The details of this mechanism have been adapted from Iwata, S., Ostermeier, C., Ludwig, B. and Michel, H. (1995) **Structure at 2.8 Å resolution of cytochrome *c* oxidase from *Paracoccus denitrificans*** *Nature 376*, 660-669. Additional structural detail is shown for A for comparison with the more minimalist cartoons (A-F) that present the full cycle. Four consecutive electron transfers to the bimetallic active site pump four vectorial protons and consume four scalar protons. (See text for a discussion of this cycle.)

A: Oxidized Form



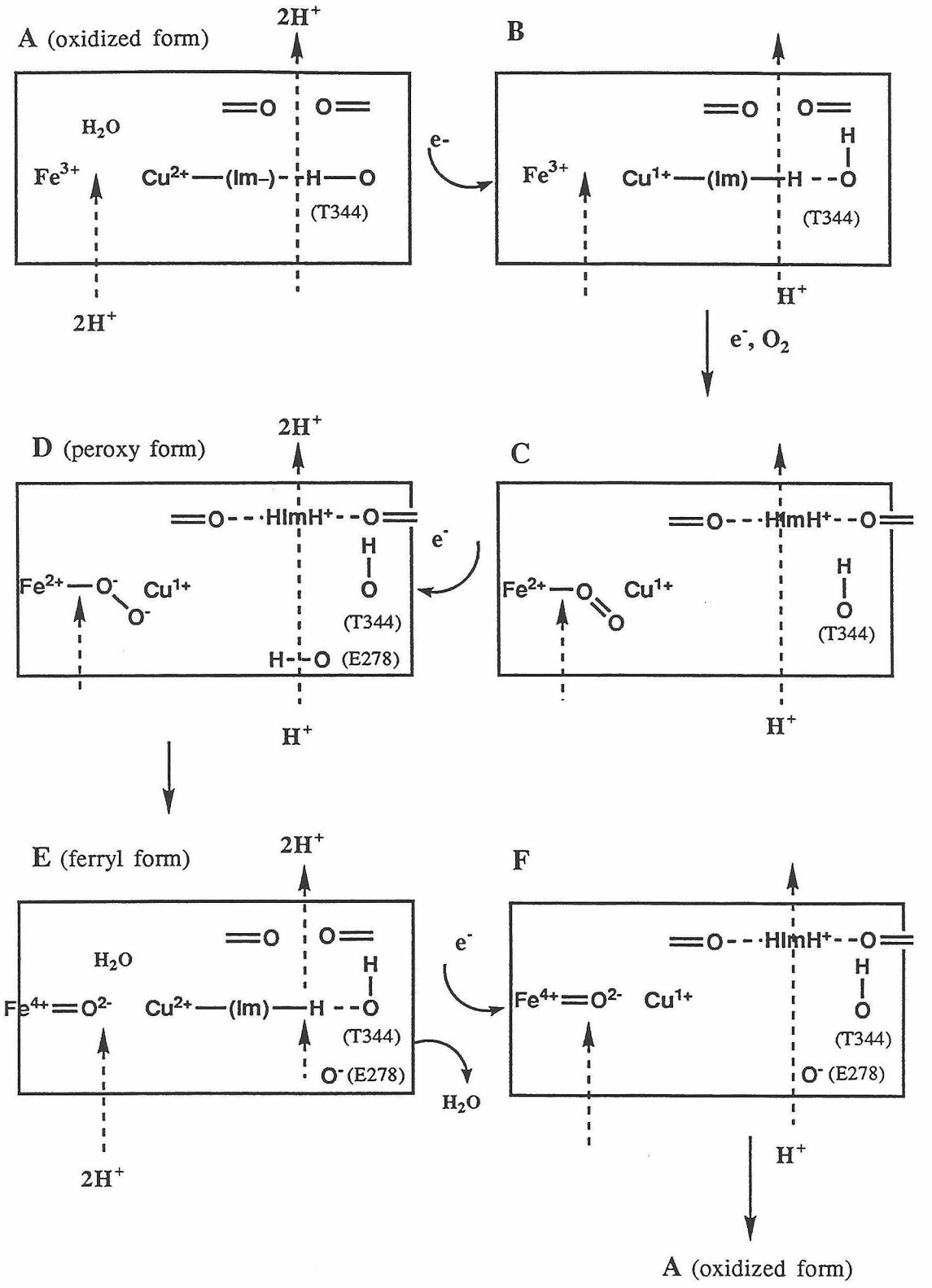
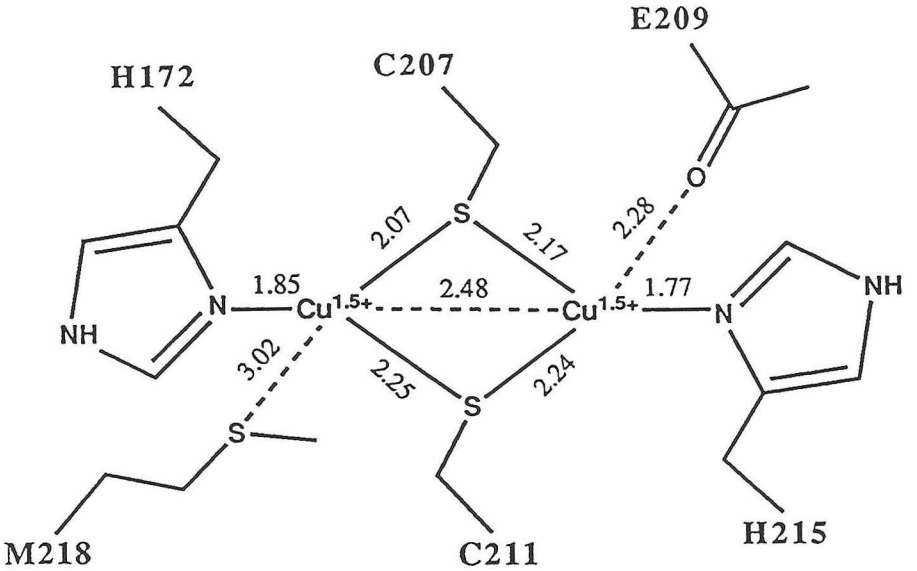

A (oxidized form)
2H+
H2O
Fe3+
Cu2+—(Im-)-H—O
(T344)
2H+
e-
B
Fe3+
Cu1+—(Im)—H - -O
(T344)
H+
e-, O2
D (peroxy form)
2H+
O- - -HImH+- -O
Fe2+—O- Cu1+
O-
(T344)
H-O (E278)
H+
e-
C
O- - -HImH+- -O
Fe2+—O Cu1+
O
(T344)
H+
E (ferryl form)
2H+
H2O
Fe4+=O2-
Cu2+—(Im)—H - -O
(T344)
O- (E278)
2H+
H2O
e-
F
O- - -HImH+- -O
Fe4+=O2-
Cu1+
(T344)
O- (E278)
H+
A (oxidized form)

Figure 1.5

A The ligand arrangment in the engineered binuclear Cu_A site of the CyoA structure. The distances are in Å. (This structure agrees with the purple centers found in the *P. denitrificans* and bovine oxidases.) **B** The mononuclear blue copper (type 1) protein azurin from *Pseudomonas aeruginosa.*

[*E. coli* CyoA structure from Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E. & Saraste, M. (1995) **Crystal structure of the membrane-exposed domain from a respiratory quinol oxidase complex with an engineered dinuclear copper center.** *Proc. Natl. Acad. Sci. USA 92*, 1995, 11955-11959.]

A


E209
H172
C207
O
2.07
S
2.17
2.28
1.85
2.48
1.77
NH
N
Cu1.5+
Cu1.5+
N
NH
3.02
2.25
2.24
S
S
M218
C211
H215

B

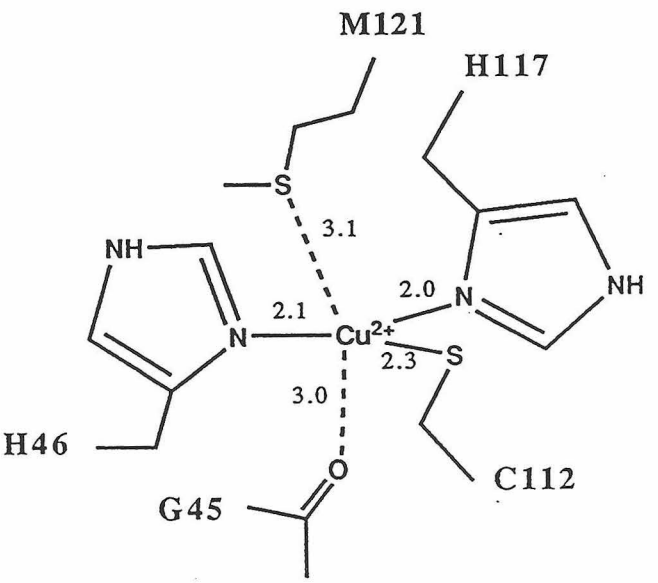

M121
H117
S
3.1
NH
2.0
N
NH
2.1
N
Cu2+
2.3
S
3.0
H46
O
C112
G45

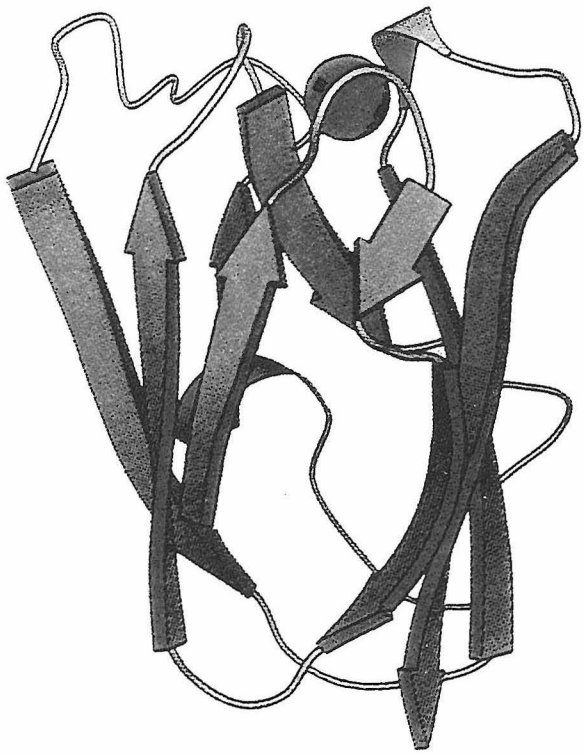
Figure 1.6

Ribbon structures of the small blue copper protein *Populus nigra* plastocyanin **A** and the fragment **B** from the *E. coli* quinol cytochrome oxidase (CyoA fragment) with an engineered purple center. Both share a cupredoxin fold.

[*Popolus nigra* plastocyanin from Karlsson, G. (1993) **Protein engineering on azurin: expression mutagensis and characterization of copper site mutants**, Chalmers University of Technolgy/ Göteborg University, Sweden, Thesis.]

[*E. coli* CyoA structure from Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E. & Saraste, M. (1995) **Crystal structure of the membrane-exposed domain from a respiratory quinol oxidase complex with an engineered dinuclear copper center.** *Proc. Natl. Acad. Sci. USA 92*, 1995, 11955-11959.]

A



B

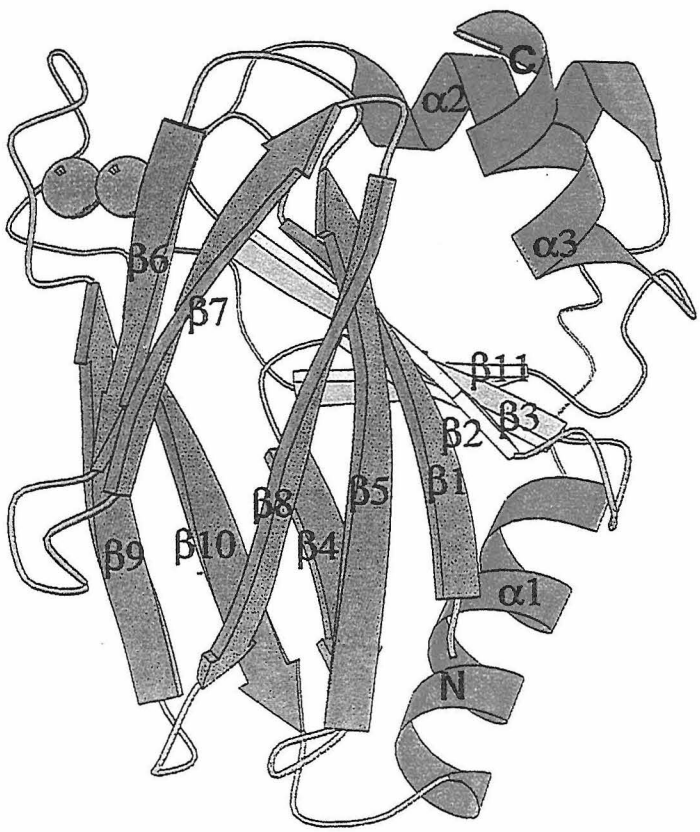
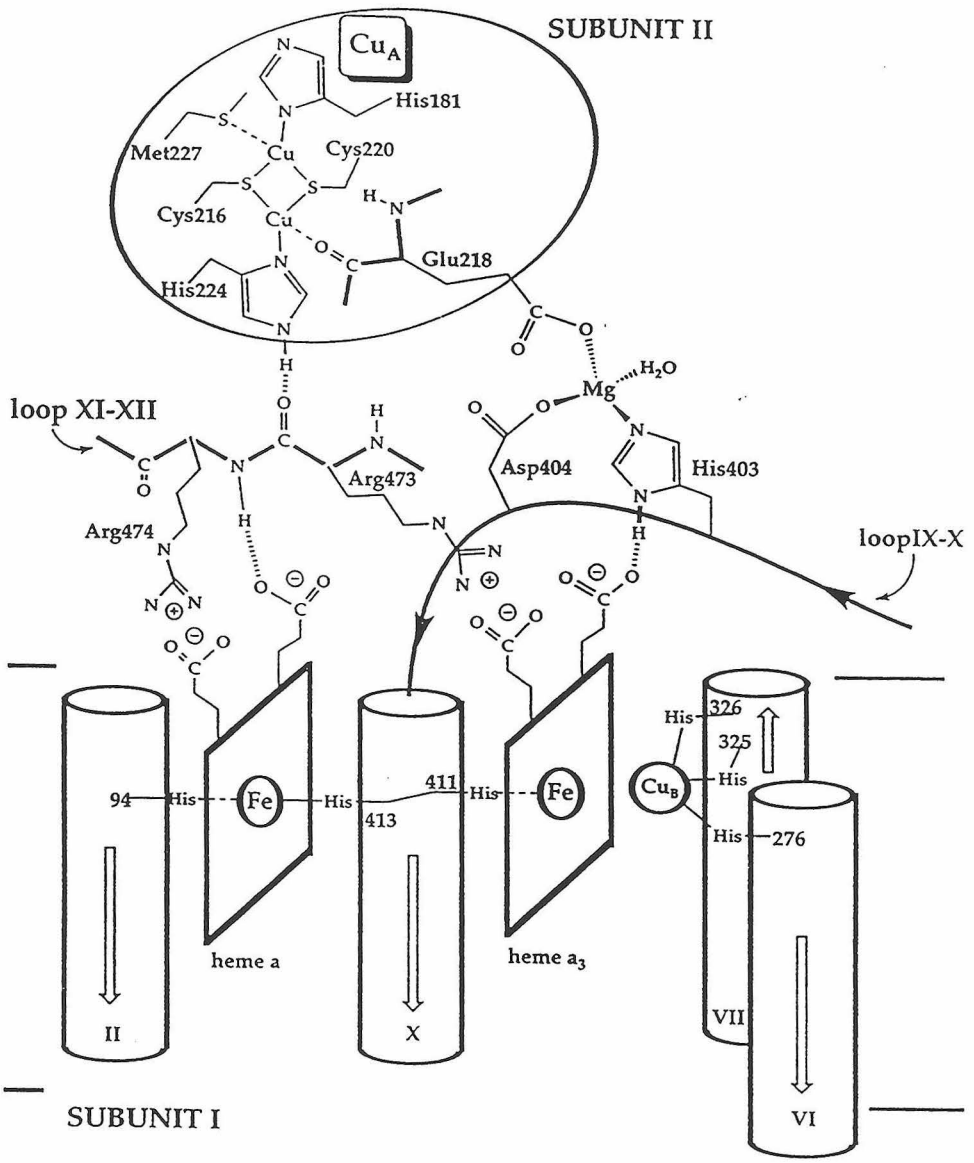


Figure 1.7

Diagram showing the proposed pathway for ET from Cu_A to cytochrome *a*. Cytochrome *c* (not shown) binds near H181. (See text for a detailed discussion of this pathway.)

[Ramirez, B. E., Malmstrom, B. G., Winkler, J. R. and Gray, H. B. (1995) **The currents of life: The terminal electron-transfer complex of respiration.** *Proc. Natl. Acad. Sci. USA 92*, 11949-11951.]


SUBUNIT II
Cu_A
His181
Met227
Cu
Cys220
Cys216
Cu
Glu218
His224
loop XI-XII
Arg473
Asp404
His403
Mg
H_2O
Arg474
loopIX-X
94
His
Fe
His
413
411
His
Fe
His
326
325
His
Cu_B
His
276
heme a
heme a_3
II
X
VII
VI
SUBUNIT I

Chapter 2

Overexpression and Preliminary Characterization of *Thermus* Copper A (Cu_A)

INTRODUCTION

Cytochrome *c* oxidases catalyze the four-electron reduction of dioxygen to water in the terminal step of aerobic respiration in eukaryotic organisms and some bacteria [for reviews, see Babcock & Wikström, 1992; Trumpower and Gennis, 1994]. Electrons enter the complex from the outside of the energy-transducing membrane through four consecutive bimolecular electron transfers from ferrocytochrome *c*. The subsequent reduction of dioxygen is coupled to the uptake of four protons from the mitochondrial matrix in eukaryotes or the cytoplasm in bacteria. In addition to the pH gradient generated across the membrane due to dioxygen reduction, redox free-energy is used to pump four additional protons per dioxygen reduced to the opposite side of the membrane. The proton gradient is then used by ATP synthase to produce ATP, a process originally described in Mitchell's chemiosmotic theory (Mitchell, 1961).

The simplest oxidases, found in bacteria, perform these functions with only two or three subunits; subunits I, II and III of the more complex mammalian oxidases are homologous to these simple bacterial oxidases, indicating that these three subunits contain the functional core of the enzyme (Saraste, 1990; Keightley *et al.*, 1995). The three-dimensional structures of the cytochromes aa_3 from *P. denitrificans* (Iwata *et al.*, 1995) and bovine mitochondria (Tsukihara *et al.*, 1995) confirm the suspected homology between bacterial and mitochondrial oxidases. Heme-copper oxidases are classified according to the types of bound hemes which are present (Trumpower & Gennis, 1994). Detailed electron transfer studies on the aa_3-type oxidases, which possess a cytochrome *a* and a cytochrome a_3-Cu_B redox center located on subunit I and a Cu_A site on subunit II, have shown that the Cu_A site is the initial electron acceptor from cytochrome *c* (Pan *et al.*, 1993; Brzezinski *et al.*, 1995). Protein labelling (Bisson *et al.*, 1982) and mutagenesis studies (Lappalainen *et al.*, 1995) have localized a portion of the cytochrome *c* binding site to the Cu_A domain of subunit II. Additionally, several acidic residues on subunit III of *P.*

denitrificans aa$_3$ oxidase are correctly positioned to interact with the lysines of cytochrome *c*, suggesting that the complete cytochrome *c* binding crevice is formed by subunits I, II and III (Iwata *et al.*, 1995). The electron on Cu_A is then transferred to cytochrome *a*, which is in rapid redox equilibrium with Cu_A (Oliveberg & Malmström, 1991). This electron is subsequently transferred to the bimetallic cytochrome a_3-Cu_B active site, where dioxygen is bound and reduced (Oliveberg *et al.*, 1989; see Wikström and Babcock, 1992 for review).

The thermophilic eubacterium *Thermus thermophilus* utilizes at least two cytochrome *c* oxidases, a *caa*$_3$- and a *ba*$_3$-type, depending on the growth conditions and oxygen availability (Fee *et al.*, 1986; Zimmermann *et al.*, 1988; Keightley *et al.*, 1995). The as-isolated cytochrome *ba*$_3$-protein is much simpler than the *caa*$_3$ oxidase (Fee *et al.*, 1993); it lacks subunit III, has a shortened subunit II and a longer subunit I (Keightley *et al.*, 1995), and it utilizes a *b* heme in the low-spin heme location (Zimmermann *et al.*, 1988). While most subunits II of heme-copper oxidases have two predicted transmembrane helices at the N-terminal region which anchor the monomer to the membrane, cytochrome *ba*$_3$ has only one (Keightley *et al.*, 1995). The conserved Cu_A motif, containing two histidines, two cysteines and one methionine, is found toward the C-terminal end of subunit II. This part of the protein projects into the periplasmic space and apparently represents an independent folding domain. By analogy to the more thoroughly characterized *aa*$_3$-type oxidases, the Cu_A site is believed to function as the primary electron acceptor of electrons from cytochrome *c* (Keightley *et al.*, 1995).

The traditional view of Cu_A described the site as a mononuclear redox center, similar to the blue copper proteins (Martin *et al.*, 1988). Recently, the C-terminal portion of subunit II from the *Paracoccus denitrificans aa*$_3$ (Lappalainen *et al.*, 1993) and the *Bacillus subtilis caa*$_3$ (von Wachenfeldt *et al.*, 1994) oxidases have been expressed. The optical spectra of these soluble fragments, free of the heme spectral interference which dominates the spectra of oxidases, clearly demonstrates that the Cu_A site is quite dissimilar

to more thoroughly characterized blue copper sites. Unlike the absorption spectra of the blue copper proteins, which are dominated by a ligand-to-metal charge transfer (LMCT) band near 600 nm (for review see Solomon *et al.*, 1992), the Cu_A absorption spectrum has prominent bands at 480 nm and 530 nm, and less intense bands centered at 360 nm and 790 nm.

A copper binding site similar to that of Cu_A is also found in nitrous oxide reductase (N_2OR[1]), which catalyzes the conversion of N_2O to N_2 in denitrifying bacteria. Antholine *et al.* (1988) were able to resolve a seven-line hyperfine pattern in the S-band EPR spectra of N_2OR, consistent with a two copper, mixed-valence site. They further proposed that the Cu_A site found in the cytochrome oxidases is binuclear. However, the EPR data for the bovine cytochrome *c* oxidase are ambigious, because direct observation of the Cu_A signal is hindered by interference from the low-spin cytochrome *a* signal. Soluble fragments of Cu_A have permitted EPR characterization of the Cu_A site without the complication of the heme signal and have yielded data which are suggestive of the N_2OR results. (See Figure 2.1.) The subunit II domains expressed so far have, however, the complications of either containing a significant amount of type 2 copper (van der Oost *et al.*, 1992; von Wachenfedt *et al.*, 1994) or of being unstable to moderate changes in solution conditions (Lappalainen *et al.*, 1993). For these reasons, we wanted to express a Cu_A domain from a thermostable bacterium in an attempt to find a system which would be more favorable for biophysical studies. A more stable, soluble fragment would allow detailed electron transfer and electrochemical studies, thus providing data that could enhance the understanding of the functional role of Cu_A. For example, a more stable, soluble fragment would allow detailed electron transfer and electrochemical studies (Slutter *et al.*, 1996).

MATERIALS AND METHODS

Bacterial strains. E. coli strain DH10B from Gibco BRL (Gaithersburg, MD) [genotype: F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) j80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K l^- *rps*L *nup*G] (Grant *et al.* 1990) was used for the ligation and sequencing steps. *E. coli* strain BL21(DE3) from Novagen (Madison, WI) [genotype: F^- *omp*T *hsd*S_B(r_B^- m_B^-) *gal dcm* (DE3)] (Studier and Moffatt, 1986) was used for the expression of the Cu_A fragment. *Thermust hermophilus* HB8 (No. 27634) was obtained from American Type Culture Collection (Rockville, MD).

Construction of pETCu$_A$. The gene fragment of the cytochrome *ba*$_3$ subunit II gene was amplified from *T. thermophilus* genomic DNA prepared using genomic DNA preparative columns (Qiagen, Chatsworth, CA). Oligonucleotides used in the PCR reaction were synthesized at the Microchemical Facility at the California Institute of Technology using an Applied Biosystems 380B DNA synthesizer. The sense oligonucleotide primer, 5'-d(CTTCGTCTTCATCGCC<u>CAT</u>**ATG**GCCTACA)-3', contained an Nde I site (underlined), the start codon (bold type), and the N-terminal portion of the Cu_A fragment. The antisense oligonucleotide primer, 5'-d(TTGCGCGCACCG<u>GGATCC</u>T**TCA**CTCCTTCA)-3', contained a Bam HI site (underlined), the stop codon (bold type), and the C-terminal portion of the fragment. The PCR reaction was optimized using the PCR Optimizer Kit™ (Invitrogen, San Diego, CA). Buffer J (5 x working concentration: 300 mM Tris-HCl, 75 mM $(NH_4)SO_4$, 10 mM $MgCl_2$, pH 9.5 @ 22°C) yielded the largest amplification in 30 cycles and was used to prepare the inserts. The PCR fragment was extracted once with one volume of chloroform to remove any traces of mineral oil, washed twice with 2 ml of 3 M sodium acetate, pH 5.3, in a Centricon 100 (Amicon, Beverly, MD) to remove the *Taq* polymerase, nucleotides and PCR buffer from the sample and washed twice with doubly distilled H_2O. After digesting the PCR fragment and pET9a vector (Novagen, Madison, WI) with Nde I

and BamH I, each was purified on a NuSieve (FMC, Rockland, MA) low melting temperature gel. The fragments were cut from the gel, melted at 70 °C, equilibrated at 45 °C and the agarose digested with Gelase (Epicenter, Madison, WI). This preparation was then ligated into the Nde I/BamH I fragment of pET9a to give the vector pETCu$_A$. The construct was sequenced with *Taq* polymerase and an Applied Biosystems DNA sequencer by Harold Kochounian at the DNA Sequencing Facility, Kenneth Norris Jr. Comprehensive Cancer Center, at the Univeristy of Southern California.

Expression and Purification of the Soluble Cu$_A$ Domain. 10 mL of culture medium (LB and 50 mg/ml kanamycin) were innoculated from a freshly streaked plate of BL21(DE3) cells containing the pETCu$_A$ plasmid. After incubation overnight at 37 °C, this culture was used to inoculate a 1 L flask of LB and 50 mg/ml kanamycin. This culture was incubated at 37 °C, typically for about 2 hours, until the A_{600} reached 0.4 to 0.6, and induced for 4-12 h using a final concentration of 0.4 mM IPTG. The cells were pelleted by centrifugation at 5,000 x g for 5 min. (At this point, the cell pellet can be frozen for future use.) The pellet from 1 L of culture was resuspended in 25 ml of 50 mM Tris-HCl, pH 8.0, 4 mg/ml lysozyme, 40 U/ml DNase I, 3 U/ml RNase A, 0.1% Triton X-100. PMSF to a final concentration of 2 mM to inhibit proteolysis was added, and the extract was incubated at 30 °C for at least 30 min. The cell debris was separated from the extract by centrifugation at 12,000 x g for 15 min at 4 °C. One volume of 50 mM sodium acetate, pH 4.6, was added to decrease the pH, and 15 l/ml volume of 100 mM Cu(II)(His)$_2$ was added to form the Cu$_A$ site. The solution turned purple at this point with some cloudy precipitate. Heat treatment of the protein at 65 °C for 10 min resulted in additional formation of color and additional precipitation of cell debris. This precipitate was pelleted by centrigufation at 12,000 x g for 30 min. The pH of the supernatant was readjusted to 4.6 with 50% acetic acid, incubated an additional 30 min on ice and clarified by a second 30-min centrifugation at 12,000 x g. The supernatant, which has a distinct purple color, was loaded onto a CM-52 or CM-Sepharose gravity column that had been equilibrated with 50 mM sodium

acetate, pH 4.6, at 4 °C, washed with several column volumes of equilibration buffer, and eluted with 50 mM sodium acetate and 1 M NaCl. The purple colored eluate was then dialyzed against 25 mM ammonium succinate, pH 4.6, at 4 °C prior to being lyophilized. The dried protein was stored until required or taken up in ~ 2 ml of doubly distilled H_2O per liter of original culture and stored frozen. Treated thusly, the chromophore appears to be stable indefinitely.

Preparation of the ^{63}Cu- and ^{65}Cu-enriched Cu_A protein. Isotopically enriched Cu (as CuO) was obtained from Oak Ridge National Laboratory: 99.70% ^{65}Cu plus 0.3 %^{63}Cu; and 99.89% ^{63}Cu plus 0.11% ^{65}Cu. The CuO was dissoved into a small amount of concentrated HCl to yeild a green soltion. This was diluted with a small amount of water to form a blue solution. It was then converted to the bis-histidine complex at pH 7 by addition of concentrated L-histidine followed by pH adjustment with concentrated NaOH. A portion of this solution was added to *E. coli* lysate containing apo-Cu_A protein, and the holoprotein was purified as described above. EPR samples enriched with ^{63}Cu and ^{65}Cu were prepared by freeze-drying and redissolving the protein into water or 40% ethyleneglycol to give a final buffer concentration of 100 mM ammonium succinate at pH 4.6. Protein concentrations were 1-2 mM in Cu_A as determined by double integration of the EPR spectra (see EPR methods section). In all cases, the samples were characterized by Cu and protein analyses, SDS-PAGE and optical and EPR spectra. We focused on the EPR spectra of the samples containing ethylene glycol because they showed narrower lines.

Protein analyses. Protein concentrations were regularly measured using the BCA protein assay kit from Pierce (Rockford, IL). Quantitative amino acid analyses and N-terminal sequencing were carried out at the University of New Mexico Protein Chemistry Facility as described by Pastuszyn in Keightley *et al.* (1995). SDS-PAGE was carried out with minor modifications according to the method of Downer *et al.* (1976) using a BioRad Mini-PROTEAN II electrophoresis cell (Hercules, CA). Non-denaturing gel electrophoresis was carried out using the same apparatus according to the method of

Gabriel (1972). Thin layer isoelectric focusing was carried out using pre-cast gels (FMC Isogel, pH range 3 - 7) in a BioRad Bio-Phoresis horizontal electrophoresis cell. Protein samples were subjected to electrospray ionization mass spectrometry after removal of sodium ions by passage over a pre-poured Pharmacia PD-10 column (Alameda, CA) equilibrated with 10 mM ammonium acetate. ESI-MS was carried out at the Scripps Research Institute Mass Spectrometry facility using a Perkin-Elmer SCIEX API III mass analyzer (Irvine, CA) with the orifice potential set at 100 volts (Siuzdak, 1994). The GCG program (Devereux *et al.*, 1984), PEPTIDESORT, was used to calculate expected properties of the protein from its amino acid composition.

Copper Analyses. Copper was released from the protein by precipitation and heating at 40 °C for several minutes in 4.4% TCA. Protein was removed by centrifugation and copper analyses were performed on the supernatant using the bathocupreine sulfonate (BCS) method, as described by Broman *et al.* (1962). More precise measurements of total copper were carried out by first denaturing the protein in a solution of 60% HCOOH, 30% 2-propanol and 10% water (hereafter called denaturing solvent), adding a trace of H_2O_2, then measuring the total Cu(II) by quantification of its EPR spectrum relative to a known standard at 77 K. Additional quantification of copper and other elements was obtained by induction coupled plasma mass spectrometry (ICP-MS) and by total-reflection X-ray fluoresence (TXRF) analysis (Pettersson and Wobrauschek (1995)). The ICP-MS spectra were recorded at the Scripps Institution of Oceanography and the TXRF spectra were recorded at Chalmers Tekniska Högskola.

X-band EPR Spectra. Spectra were recorded at liquid nitrogen temperatures using Bruker ER 200D-SRC X-band spectrometers either at the Scripps Research Institute in La Jolla or in Göteborg. Generally, the spin concentration in the protein samples = concentration of spins in standard x CDBLI(protein)/CDBLI(standard). CDBLI is the corrected double integral of the EPR signal. Signals were quantified as described by Aasa and Vänngard (1975) using Cu(II) in 2 M $NaClO_4$ at pH 2 as reference. Cytochrome ba_3

was obtained by the method of Keightley *et al.* (1995) and was maintained in solution with a Tris-EDTA buffer containing 0.1% Triton X-100.

Multi-frequency EPR Spectra. EPR spectra at 3.93 and 9.45 GHz were recorded with a Bruker ER 200D-SRC spectrometer equipped with an Oxford Instruments ESR-9 helium flow cryostat. 3.93 GHz spectra employed an ER 061 SR microwave bridge, an ER 6102 SR reentrant cavity and a homemade quartz insert. Spectra at 34 GHz were recorded with a Bruker ESP 380 spectrometer using an ER 050 QG bridge and an ER 5103 QT cavity. Low temperature was obtained with a Bruker Flexline ER 4118 CF helium cryostat. Quantitations of EPR spectra were performed under non saturation conditions as described by Aasa and Vänngård (1975).

Optical Absorption and Circular Dichroism Spectra. Optical spectra of the Cu_A fragment were recorded on a Hewlett Packard 8452A Diode Array or SLM/AMINCO model DB3500 spectrophotometers in 1-cm cells. The azurin spectrum was taken on a Cary 4 UV-visible spectrophotometer. The protein concentrations used were 150 μM Cu_A domain or 100 μM azurin. Circular dichroism spectra were recorded on a Jasco J720 spectropolarimeter at 20 °C over the wavelength range 260-185 nm. A 1-mm pathlength cell was used and the cell compartment was continuously flushed with N_2. The total absorbance of the protein (0.05 mg/mL) and buffer (0.5 mM potassium phosphate buffer, pH 7.4) in the sample never exceeded 0.6 over the wavelength range used. All spectra are the average of ten recordings. The melting temperature of the fragment was measured by recording the change in CD signal at 218 nm using a Model 62A DS Aviv CD polarimeter equipped with a specialized thermoelectric device that permitted the sample temperature to be stepped, in this case, at intervals of 0.5 °C from 40 - 120 °C. The sample is retained in a sealed cuvette having a pathlength of 1 mm. Visible region CD data were collected at 20 °C over the wavelength range 310-700 nm in a 1-cm cell. The samples contained 0.25 mg/mL Cu_A in 0.5 mM potassium phosphate buffer, pH 7.4. The absorption of the far-UV CD samples was measured against air in the same 1-mm cell used to collect the CD

data.

Secondary Structure Prediction. The computer program used to estimate the amount of secondary structure from CD data compares the CD data to that of 22 proteins with known structures (Hennesy & Johnson, 1981; Johnson, 1990). To improve the accuracy of this algorithm, the homologous blue copper proteins azurin and plastocyanin have been included in this set of reference spectra. The statistical method of variable selection (Weisberg, 1985) was used to select the reference proteins most resembling the sample proteins in secondary structure.

Resonance Raman Spectra. Resonance Raman (RR) spectra were collected at the Oregon Graduate Institute of Science and Technology. Spectra were collected with excitation near 480, 530 and 790 nm (~150 mW) at 15 K. The spectra presented here were collected at 488 nm, 15 K. Sulfur isotope labelling (^{32}S and ^{34}S) in the *P. denitrificans* soluble Cu_A fragment has been used to assign the intense vibrations at 260 and 340 cm $^{-1}$ to a Cu-S(Cys) stretch. Comparison of this spectrum to others including that of *T. thermophilus* is discussed by Andrew *et al.* (1995).

^{1}H-NMR Spectra. Spectra were collected on a Bruker AMX-500 spectrometer. 10 mM of Cu_A fragment in 100 mM potassium phosphate buffer (pH 8) and 200 mM KCl was used. Initially, the high ionic strength was used to prevent aggregation and stabilize the reduced protein. However, the clear paramagnetically shifted protons indicate that aggregation may not be a problem. Temperature (296 to 323 K) was varied using the Bruker AMX-500 control unit.

RESULTS AND DISCUSSION

Expression and purification of the Thermus cytochrome ba_3 Cu_A domain. The nucleotide sequence encoding the first 32 N-terminal residues of subunit II of cytochrome ba_3 (Keightley *et al.* 1995) was removed by inserting a methionine start codon before A32.

This truncation disrupts a highly hydrophobic region (residues 18-38) that is presumed to form a helix anchoring subunit II to the membrane (Figure 2.2). Consequently, the remaining C-terminal portion of subunit II should contain only the soluble Cu_A domain. Including the start codon, the translated N-terminal amino acid sequence of the gene is MAYTLAT-- extending to the C-terminus (--GTIVVKE). The amino acid sequence of the *Thermus* Cu_A fragment is shown in Figure 2.3 and is aligned with the CyoA fragment amino acid sequence. The expected soluble protein fragment is 136 amino acids long with a molecular weight of 14,936. Because there are no signal sequences designed into the gene sequence, expression is directed to the cytoplasm.

Upon induction with IPTG, the cells produce only the apo form of the protein, even when grown in the presence of ~1 mM $Cu(II)(His)_2$. However, after cell lysis, the addition of Cu(II), usually as the bis-histidine complex although other complexes work as well, causes the cell extract to become purple[1]. The very simple purification procedure described in the Materials and Methods section typically yields ~30 mg pure holo protein per liter of culture medium.

It is also possible to prepare purified apoprotein by omitting the addition of Cu(II), and this can be done even when $Zn(II)(His)_2$ is added to the cell lysate. Interestingly, ICP-MS analyses of purified apoprotein showed that no metals were bound to the final product. This is an unexpected result given that overexpression of blue copper proteins, azurin for example, tends to yield a mixture of apoprotein, zinc protein and holoprotein. Holoprotein can be obtained from purified apoprotein by adding an excess of $Cu(II)(His)_2$ (or other Cu complexes); however, prior exposure to air can destroy the ability to bind Cu(II). The latter can be partially reversed by treatment with β-mercaptomethanol[2], suggesting that the formation of a disulfide in the active site prevents Cu (II) binding.

Protein characterization. SDS-PAGE analysis, Figure 2.4A, shows a single protein band of very high purity with an apparent M_r ~ 15,000. This is consistent with a predicted M_r of ~14,800 (see below). Similarly, Figure 2.4B shows an electrophoresis gel run

under non-denaturing conditions. This gel is significantly overloaded and shows several minority bands. However, densitometry scans of several gels indicate that the protein preparations are >95% pure. A thin-layer isoelectric focussing gel, Figure 2.4C, illustrates that the cupration reaction may not always proceed to completion. The upper band (pI = 6.0) corresponds to fully cuprated holoprotein while the lower (pI = 6.2) corresponds to the metal-free apoprotein; the program PEPTIDESORT predicts that the pI of the apoprotein should be 6.1. By following the preparative procedure described in the Materials and Methods section, and as evidenced by the absence of the apoprotein band in IEF gels (not shown), most of our preparations are free of apoprotein.

Our initial preparations of the Cu_A protein contained several different N-terminal sequences in varying amounts: TLATHTAGVIPA, THTAGVIPA and GVIPA indicating that proteolysis was occurring. The alignment in Figure 2.3 shows a large N-terminal overhang that does not contain secondary structural elements of the cupredoxin fold (Wilmanns *et al.*, 1995). After addition of PMSF during cell lysis, however, a homogeneous sample with the experimentally determined N-terminal sequence AYTLATH was obtained. The resulting protein should thus be 135 amino acids in length. Facile removal of the formyl-Met (Miller, 1987) and subsequent N-terminal proteolysis suggest this region of sequence may not be tightly packed in the soluble domain[3]. Table 2.1 presents a comparison of the observed amino acid composition with that predicted from translation of the gene sequence; they are, within error, indistinguishable.

Further evidence that the cloned DNA fragment encodes the desired protein was obtained from ESI-MS experiments. The predicted mass of the apoprotein having the N-terminal sequence AYTLATH-- is 14,804 daltons. Figure 2.5A shows the mass spectrum of the apoprotein, prepared by dissolution of the holoprotein into the denaturing solvent (HCOOH/2-propanol/water). The major band has a molecular weight of 14,804. The very weak bands at higher mass may reflect the presence of Na^+ and or Cu^{2+} ions bound to the protein even under these strongly denaturing conditions. Figure 2.5B shows the

mass spectrum of non-denatured holoprotein. The major band has a molecular weight of 14,928. The difference of 124 mass units between the holo- and apoproteins is due to the binding of two Cu ions to the former (see below). The minority peaks at 14,803 and 14,864 are apparently due to the presence of small amounts of apoprotein and apoprotein binding one Cu, respectively. The minority peaks at higher masses reflect binding of Na^+ ions to the holoprotein (see Figure legend).

Copper Analyses. While the electrospray mass spectrum of Figure 2.5B is consistent with two Cu ions per protein molecule, this was confirmed by measuring Cu/protein ratios. Initially we encountered considerable difficulty in removing the Cu from the protein. For example, even when the protein is precipitated with TCA, as is often done in metalloprotein analyses (Massey, 1957), substantial amounts of Cu are retained in the precipitate. However, when heated at 40 °C for several minutes in 4.4% TCA, the protein releases most of its copper. Quantitative ICP-MS also gave unexpectedly low Cu/protein ratios, presumably because protein in the very dilute (~μM) aqueous solutions used in these measurements adhered to the walls of the plastic tubes. Analytical data consistent with the ESI-MS result were obtained, however, using quantitative EPR spectroscopy after a dissolution of the protein into the denaturing solvent (see Materials and Methods). Four samples, judged from isoelectric focussing gels to be free of apoprotein, gave values of 2.00, 1.56, 1.88 and 1.88 moles of Cu per 14,800 g protein; the concentration of protein was determined by either BCA or/and quantitative amino acid analyses; similarly, two samples analyzed for Cu by TXRF gave values of 1.71 and 1.62. All our analytical data (from BCS, ICP-MS, EPR and TXRF) support the interpretation of the observed mass of the holoprotein (14,928 Da) as arising from the apoprotein plus two Cu ions: apoprotein (14,804) *plus* two Cu (2 x 63.5 = 127) *minus* two protons from the cysteine residues (-2) *equals* 14,929 Da. Note that the error in these mass measurements is 1 part in 10,000 (Siuzdak, 1994). Both ICP-MS and TXRF simultaneously measure a broad spectrum of elements, and significant amounts of other metals were not found in our samples.

Absorption and Visible CD Spectra. Figures 2.6A and 2.6B compare the absorption and visible CD spectra of the Cu_A domain with that of azurin from *Pseudomonas aeruginosa.* The *Thermus* soluble fragment shows absorption bands at 363, 480, 530 and 790 nm. The second derivative spectrum (not shown) reveals an additional peak at ~590 nm, which can be seen as a shoulder in the absorption spectrum. Based on the Cu analyses and quantitative EPR (see below), the extinction coefficient, ε, at 790 nm is 1900 $\pm$ 200 $[2Cu]^{-1}cm^{-1}$. The energies and extinction coefficients of these transitions are similar to those reported for other soluble Cu_A fragments from cytochrome oxidases, the Cu_A sites in two characterized N_2OR proteins and an engineered purple center. However, there are subtle differences in the position of these bands, suggesting that the properties of the site depends somewhat on the overall protein environment; Table 2.2 summarizes these features.

The visible CD spectrum of the ba_3-Cu_A domain shows features at 340 nm (+0.946), 381 nm (-0.638), 460 nm (+1.27) and 527 nm (-1.81) where the numbers in parentheses are the differential molar absorption coefficients ($\Delta\varepsilon = \varepsilon_l - \varepsilon_r$). There is also a strong shoulder at 425 nm and a negative going feature at ~690 nm. However, the latter is not shown, because the lamp available in the instrument only gives reliable signals up to ~700 nm. It is noteworthy that the two strong absorption bands around 500 nm in the absorption spectrum have opposite signs in the CD spectrum (see Discussion). The near-UV absorption spectrum shows a sharp peak at 276 nm with an observed extinction coefficient of 17,000 $\pm$ 400 $M^{-1}cm^{-1}$; because the value predicted from PEPTIDESORT is 14,770 $M^{-1}cm^{-1}$, it is possible the Cu_A center also has absorption bands in the near-uv region.

This is the first report of the CD spectrum of Cu_A, although the MCD spectrum has been described (Farrar *et al.*, 1995), and it deserves some comment. The spectra of azurin are actually quite well understood (Solomon *et al.*, 1992; Han *et al.*, 1993) while work has just begun on the origin of the various bands in the Cu_A spectra (Gray *et al.*, work in

progress). However, CNDO/S calculations of Larsson *et al.* (1995) suggest that the two absorption bands of nearly equal intensity at 480 and 530 nm arise from exciton splitting between the two Cu ions and predict, as is observed, that the corresponding CD bands should have opposite signs.

EPR Spectra. The X-band EPR spectrum of native cytochrome ba_3 from *T. thermophilus* recorded at ~80 K is shown in trace A of Figure 2.7. In this particular sample, there is a significant amount of the unknown impurity first described by Zimmermann *et al.* (1988) and seen in many of our preparations (Jim A. Fee, unpublished observations). However, the majority signal clearly arises from Cu_A. Trace B of Figure 2.7 shows the spectrum of the soluble Cu_A domain recorded at 77 K under similar instrument settings. Even if trace A is disturbed by the presence of the impurity, the closeness of the g-values of the majority signal with those of trace B allows one to conclude that the environment of the Cu_A center in the soluble domain is highly similar to that in the native protein. As reported elsewhere (Fee *et al.*, 1995), the signal is axial having g-values of 2.187 and ~2.00.

Double integration of the Cu_A signal and comparison to a standard Cu(II) solution yielded values of spin concentration, [S = 1/2]. This was done for ten different preparations for which independent Cu analyses were available (BCS). The [S = 1/2] /[Cu] ratio ranged from 0.32 to 0.68 with an average of 0.54 ± 0.09. In two additional samples, the Cu_A spin concentration was determined as described above and total Cu was obtained by dissolving the protein in the denaturing solvent, adding a trace of H_2O_2, then twice integrating the resulting EPR spectrum (see Materials and Methods section). The spectrum of the denatured protein shown in trace C of Figure 2.7 arises largely from a single specie having $g_{\parallel} = 2.40$, $g_{perpendicular} = 2.08$ and $A_{\parallel} = 125$ gauss. Both samples gave values of [S = 1/2; CuA]/[S = 1/2; Cu_{total}] = 0.40[4]. These data show that only half the Cu present in the protein is EPR detectable, a result predicted from the binuclear model of Kroneck and Antholine and co-workers (Antholine *et al.*, 1992).

Multi-frequency EPR spectra of ^{63}Cu- and ^{63}Cu-enriched derivatives. EPR spectra are dominated by the signal from Cu_A, but quantitation showed that the ^{63}Cu-enriched sample contained 15% of a type 2 Cu(II) signal while the ^{65}Cu sample contained very little extraneous signal. The experimental spectra are show in Figure 2.8 where they are displayed on a common g-value scale. At 34 GHz there is no evident hyperfine structure and is used primarily to obtain the g_z-value, which occurs at 2.187 and is marked with a vertical dashed line. At 9.45 GHz, the spectrum shows considerable hyperfine structure on the low-field site of g_z. As seen earlier by Fronsicz *et al.* and Antholine, Kroneck and co-workers, the hyperfine structure is resolved better at lower microwave frequencies. Thus, one can count five lines on the low-field side and four lines on the high-field site of the 3.93 GHz spectrum. The hyperfine structure is more pronounced in a second-derivative spectra using EXCEL. Figure 2.9A and B show the low- and high-field portions, respectively, of the 3.93 GHz spectra of ^{63}Cu- and ^{65}Cu- enriched Cu_A proteins. Because the experimental, first-derivative spectra show peaks in these regions, the second-derivative spectra have a first-derivative shape.

The nuclear magnetic moments of ^{63}Cu and ^{65}Cu are 2.226 and 2.385 nuclear magnetons respectively, giving a '65/63' ratio equal to 1.071. Thus, one expects the Cu hyperfine structure, centered around a common g-value, to have A-values with the same ratio. Figure 2.9 also shows spectra, derived from the ^{63}Cu spectra, but "stretched" about the g-position on the field axis to match the ^{65}Cu spectra as closely as possible, thereby producing a "simulated" ^{65}Cu spectrum. The center of the stretching is used as a variable in this simulation, thus giving an independent measure of the corresponding g-value. In both low- and high-field regions, the best "fit" is obtained with a stretching factor of 1.067, very close to the magnetic moment ratio for the two isotopes. This analysis shows that the observed splitting is derived primarily from the magnetic hyperfine interaction with copper nuclei, and rules out major contributions from copper quadrupole effects and other nuclei, such as ^{14}N or ^{1}H. The field axes of the experimental and simulated ^{65}Cu spectra coincides

with the dashed lines in Figure 2.9, thus showing the centers of stretching. On both sides of the spectrum, these occur in the middle of a 7-line multiplet. At the low-field side the center coincides wtith the g_z-value derived from the 34 GHz spectrum. At the high-field side the center should be at one or both of the 'pependicular' g-values. The same positions are marked with vertical dashed lines in Figure 2.9.

It is possible that the apparent hyperfine structure in the spectra could derive from either forbidden transitions and/or a rare combination of A- and g- tensor orientations. To test these possibilities, we attempted to simulate a 7-line multiplet in the z-direction of the 'powder' spectrum using a single Cu nucleus but with different axes of the g- and A-tensors but were unsuccessful. Further, microwave saturation studies (data not shown) indicate that forbidden transitions do not contribute to the number of lines. However, as predicted from the work of Antholine *et al.* (1988) the spectra can be fitted rather closely using two identical A-tensors (Karpefors, M., unpublished observations). In addition, X-band and Q-band ENDOR spectra show that the coupling constants could only differ by ~10 MHZ or less (~3 Gauss) (Doan, P., unpublished observations). Taken together, the present data therefore proves beyond any doubt that the multiplet must originate in the coupling to two closely equivalent Cu nuclei (resulting in a 1:2:3:4:3:2:1 septet).

The near identity of the two Cu A-values suggests that the unpaired electron is equally distributed over both Cu ions. This signifies that the complex is a Class III, mixed-valence center as discussed by Robin and Day (1967), and any proposed structure must satisfy this condition.

Resonance Raman Spectroscopy. Resonance Raman (RR) can be used to selectively examine the Cu-S(Cys) chromophore in the Cu_A site. Comparison of the RR spectra of the soluble Cu_A fragments from *B. subtilus*, *P. denitrificans* and *T. thermophilus* as well as the engineered purple centers from *P. aeruginosa* azurin and *T. versutus* amicyanin show a distinctive pattern of RR frequencies (Andrew *et al.*, 1995). Additionally, excitation near 480, 530 and 780 nm suggests that each of these bands has significant (Cys)S—›Cu

charge transfer character. These spectra are shown in Figure 2.11. All have two intense vibrations near 260 and 340 cm^{-1} and several weaker vibrations between 115 and 400 cm -1. Examination of the mass effect of sulfur-isotope substitution in the *P. denitrificans* Cu_A fragments indicates that both the 260 and 339 cm^{-1} bands have large isotope shifts of -4.1 cm^{-1} and -5.1 cm^{-1}, respectively (Figure 2.11A). The *T. thermophilus* and the azurin spectra have very intense 260 cm^{-1} peaks, similar to previously collected data on N_2OR (Andrew *et al.*, 1994).

Figure 2.12 presents a comparison of sulfur-isotope shifts for the mononuclear blue copper sites of azurin (Dave *et al.*, 1993) and plastocyanin (Qiu *et al.*, 1995) and the binuclear Cu_A site. In these blue copper proteins, several sulfur-dependent bands are generated by kinematic coupling of a single Cu-S stretch with cysteine perturbations of similar energy (Qiu *et al.*, 1995). Thus, Figure 2.12 B and C show a cluster of bands near (within ~30 cm^{-1}) the ~400 cm^{-1} ν(Cu-S) mode. The total sulfur-shift, distributed between the peaks of this cluster, is ~5 cm^{-1}. By contrast, the magnitude of the sulfur-shift increases to ~10 cm^{-1} in the Cu_A site as does the energy separation (~80 cm^{-1}) between the 260 and 339 cm^{-1} peaks. This suggests that these vibrational modes are due to two distances ν(Cu-S) modes rather than one. Andrew *et al.* (1995) note, that this behavior can be simulated with a binuclear site and two bridging thiolate ligands.

^{1}H-NMR of the Oxidized and Reduced Thermus Cu_A Fragment. The most unusual characteristic of the Cu_A site is the presence of hyperfine (paramagnetically) shifted protons in the oxidized form (Bren, K. and Slutter, C. unpublished results). Generally, hyperfine shifts are not seen in Cu complexes because the relaxation times are quite slow, making these peaks broad rather than sharp. However, the presence of ten sharp, hyperfine shifted peaks in the ^{1}H-NMR spectrum is consistent with an EPR signal only at low temperature (80 K) and not room temperature. Figure 2.13 shows the 500 MHz ^{1}H NMR spectrum for oxidized Cu_A fragment with the shifted peaks labeled a-j. All the resonances integrate to one proton each; and only proton, d, is exchangeable.

Currently, these peaks are unassigned; but, the strategy for making these assignments is straightforward. Deuterium-labeled ligands will be incorporated into the Cu_A site by the use of auxotrophs, *E. coli* strains which require specific amino acids to grow (the result of a knock-out mutation in the amino acid synthetic pathway). Presently, the methyl-deuterated methionine protein has been prepared (Sanders & Fee, unpublished results). Deuterium NMR investigation of this labeled Cu_A protein indicates that the methyl protons are not paramagnetically shifted and lie, instead, near the water peak. Similar preparations are underway to purify the cysteine and histidine deuterium-labeled proteins.

Figure 2.14 shows the temperature dependence of protons a-j. All of the temperature dependencies are Curie (i.e., increasing temperature shifts the resonances to the diamagnetic region). The observed shifts are fully reversible over the range examined. Moreover, these temperature dependencies are quite weak relative to other well studied centers like ferredoxin (see Bertini & Luchinat, 1992, and references therein). This suggests the presence of multiple low-lying S=1/2 states that are populated differently as the temperature increases.

Far-UV CD Spectra and Secondary Structure Prediction. The far-UV CD spectrum for the *T. thermophilus* fragment is shown in Figure 2.14A. The estimated percentages of five forms of secondary structure calculated from this spectrum are summarized in Table 2.3; listed for comparison are also the calculated percentages of each type of secondary structural element for two other Cu_A domains as well as those known from high-resolution X-ray data for azurin (Nar *et al.*, 1991) and plastocyanin (Colman *et al.*, 1978). The two blue proteins, which share considerable sequence similarity with the Cu_A domains, have been included in the basis set of related proteins that was used to predict the fold of this domain. The results are generally consistent with the newly reported X-ray structures (Iwata *et al.*, 1995; Tsukihara *et al.*, 1995) and with previous deductions of secondary structure in subunits II (Wittung *et al.*, 1994, Ramirez, unpublished observation. See Figure 2.13.) and suggest that the *T. thermophilus* fragment is a structural homolog of the

P. denitrificans subunit II, having a β-barrel fold with some additional α-helical structure[5].

Thermal and pH Stability. As shown in Figure 2.14B, protein secondary structure is retained up to at least 80 °C with denaturation being complete at ~110 °C. We have not studied the effect of temperature on the integrity of the Cu_A site, but preliminary studies during the cupration reaction suggest it is formed quite rapidly even at 65 °C and is stable at this temperature for at least one hour. The Cu_A center is particularly stable to acid pH (~3.0), but protein precipitation occurs at lower pH (~2.5) with some loss of copper. Above pH 9 the protein loses its color with apparent oxidation of the active site cysteines to a disulfide; the purple site can be partially regenerated after reduction with thiols[2].

Redox properties. Preliminary electron transfer studies have been carried out with the protein (Slutter *et al.*, in press and Chapter 3), and in a study that will be reported elsewhere (manuscript in preparation), the redox potential at pH 8.1 is 240 mV vs. NHE, similar to that observed for other Cu_A centers in cytochrome *c* oxidases (Wang *et al.*, 1986). It is sufficient to state that the protein exhibits well-behaved one-electron, Nernstian behavior and is not oxidized to the 2 Cu(II) form at available potentials.

FOOTNOTES

1. During overexpression of this fragment, purple-colored membranes were also found in the cell pellet, suggesting that the fragment still retained some hydrophobic residues. Because this hydrophobic patch may cause problems with self-association in solution, and our early work indicated that this region is sensitive to proteases and therefore less structured than the C-terminal portion, we constructed a second fragment with a larger portion of this region removed. The predicted N-terminal sequence of this fragment is MVIPAG, which is ten residues shorter than the original construct and corresponds to the most degraded protein fragment which was isolated in the original protein preparation minus the first glycine residue. During purification, there are actually two purple colored fractions, one of which is N-terminally blocked, presumably with an N-formyl group (Miller, 1987) while the other has the N-terminal sequence VIPAG.

2. The apoprotein was regenerated by reduction with β-mercaptoethanol followed by gel filtration on a PD-10 column equilibrated with 50 mM Tris-HCl, pH 8, and 10 mM $CuSO_4$. The colorless, reduced apoprotein immediately formed a distinct purple band upon entering the column. Initial attempts to regenerate the apoprotein without Cu(II) on the column were unsuccessful, as the time required to elute the protein from the column is sufficient for complete reoxidation.

3. This solvent also dissolves the cytochrome *c* oxidase from bovine heart mitochondria and releases the Cu as Cu(II). It was easy to demonstrate that the ratio of Cu to heme *a* is 1.5 in purified oxidase (R. Aasa, J. A. Fee and B. G. Malmström, unpublished observations). This is consistent with earlier work that indicated the presence of additional Cu in the bovine oxidase (Öblad *et al.*, 1979; Einarsdottir *et al.*, 1985; Steffens *et al.*, 1987).

4. Because the expected result is 0.5 with an approximate error of ±10 %, it is possible that some of the Cu_A protein in these samples was reduced.

5. While this manuscript was nearing completion, Dr. Matthias Wilmanns provided us with coordinates of his model for the purple-CyoA protein making it possible for us to briefly examine the model using molecular graphics. We have also compared the amino acid sequences of the soluble ba_3-Cu_A protein and the soluble CyoA (see also Mather *et al.*, 1992, for the alignment of CyoA sequence with other Cu_A containing proteins and Figure 2.3). After alignment of ba_3-Cu_A and CyoA sequences such that the metal liganding amino acids are in corresponding positions, it is evident that CyoA possesses a C-terminal extension of ~70 additional amino acids (34 of which are observable in the present X-ray data). Moreover, with this alignment, ba_3-Cu_A possesses an additional 33 amino acids at its N-terminus that are not represented in the purple CyoA structure. Because of the small size of the ba_3-Cu_A protein compared to CyoA, approximately 70 of the C-terminal residues of CyoA can have no counterpart in the structure of ba_3-Cu_A, and our analysis of the near-UV CD spectrum suggests that much of this segment probably exists in an α-helical form. The molecular graphics analysis indicates that the 57 C-terminal amino acids observable in the CyoA structure, that cannot be part of the ba_3-Cu_A structure, make no contacts with the Cu_A center and form a 'cap' that fits onto the 'β-barrel' which constitutes the core of the subunit II structure.

REFERENCES

Aasa, R. & Vänngård, T. (1975) *J. Magn. Reson. 19*, 308 - 315.

Aasa, R., Albracht, S. P. J., Falk, K. E., Lanne, B. & Vänngård, T. (1976) *Biochim. Biophys. Acta 422*, 260 - 272.

Andrew, C. R., Lappalainen, P., Saraste, M., Hay, M. T., Lu, Y., Dennison, C., Canters, G. W., Fee., J. A., Slutter, C. E., Nakamura, N. & Sanders-Loehr, J., (1995) *J. Am. Chem. Soc.* 117 (43) 10759-10760.

Andrew, C. R., Yeom, H., Valentine, J. S., Karlsson, B. G., Bonander, N., van Pouderoyen, G., Canters, G. W., Loehr, T. M. & Sanders-Loehr, J. (1994) *J. Am. Chem. Soc. 116*, 11489-98.

Antholine, W. E., Kastrau, D. H. W., Steffens, G. C. M., Buse, G., Zumft, W. G. & Kroneck, P. H. M. (1992) *Eur. J. Biochem. 209*, 875-888.

Babcock, G. T. & Wikström, M. (1992) *Nature 356*, 301-307.

Bertagnoli, H. & Kaim, W. (1995) *Angew. Chem. Intl. Ed. Engl. 34*, 771-773.

Bertini, I. & Luchinat, C. (1992) in *Physical Methods for Chemists* (ed. Drago, R.) p. 500-556, Sanders College Publishing, New York.

Bisson, R., Stetten, G. C. M., Capaldi, R. A. & Buse, G. (1982) *FEBS Lett. 144*, 359-363.

Blackburn, N. J., Barr, M. E., Woodruff, W. H., van der Oost, J. & de Vries, S. (1994) *Biochemistry 33*, 10401-10407.

Blair, D. F., Gelles, J. & Chan, S. I. (1986) *Biophys. J. 50*, 713 - 733.

Broman, L., Malmström, B. G., Aasa, R. & Vänngård, T. (1962) *J. Mol. Biol. 5*, 301-310.

Brzezinski, P., Sundahl, M., Ädelroth, P., Wilson, M. T., El-Agez, B., Wittung, P. & Malmström, B. G. (1995) *Biophys. Chem. 54*, 191-197.

Churg, A. K. & Warshel, A. (1986) *Biochemistry 25*, 1675-1681.

Colman, P. M., Freeman, H. C., Guss, J. M., Murata, M., Norris, V. A., Ramshaw, J. A. & Venkatappa, M. P. (1978) *Nature 272*, 319-324.

Dave, B. C., Germanas, J. P. & Czernuszewicz, R. S. (1993) *J. Am. Chem. Soc. 115*, 12175-12196.

Dennison, C., Vijenboom, E., de Vries, S., van der Oost, J. & Canters, G. (1995) *FEBS Lett. 365*, 92-94.

Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* 12, 387 - 395.

Downer, N. W., Robinson, N. C. & Capaldi, R. A. (1976) *Biochemistry 15*, 2930-2936.

Einarsdóttir, Ó. & Caughey, W. (1985) *Biochem. Biophys. Res. Commun. 129*, 840-847.

Farrar, J. A., Lappalainen, P., Zumft, W. G., Saraste, M. & Thomson, A. J. (1995) *Eur. J. Biochem. 232*, 294-303.

Fee, J. A., Kuila, D., Mather, M. W. & Yoshida, T. (1986) *Biochim. Biophys. Acta 853*, 153-185.

Fee, J. A., Sanders, D., Slutter, C. E., Doan, P. E., Aasa, R., Karpefors, M. & Vänngård, T. (1995) *Biochem. Biophys. Res. Commun. 212*, 77-83.

Fee, J. A., Yoshida, T., Surerus, K. K. & Mather, M. W. (1993) *J. Bioenerg. Biomembr. 25*, 103-114.Gabriel, O. (1972) *Methods Enzymol. 22*, 565-578.

Gelles, J., Blair, D. F. & Chan, S. I. (1986) *Biochim. Biophys. Acta 853*, 205-236.

Grant, S. G. N., Jessee, J., Bloom, F. R. & Hanahan, D. (1990) *Proc. Natl. Acad. Sci. USA* 87 4645-4649.

Gurbiel, R. J., Fann, Y.-C., Surerus, K. K., Werst, M. M., Musser, S. M., Doan, P. E., Fee, J. A. & Hoffman, B. M. (1993) *J. Am. Chem. Soc. 115*, 10888-10894.

Han, J., Loehr, T. M., Lu, Y., Valentine, J. S., Averill, B. A. & Sanders-Loehr, J. S. (1993) *J. Am. Chem. Soc. 115*, 4256-4263.

Hennessey, Jr., J. P. & Johnson, Jr., W. C. (1981) *Biochemistry 20*, 1085-1094.

Hulse, C. L. & Averill, B. A. (1990) *Biochem. Biophys. Res. Commun. 166*, 729-735.

Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995) *Nature 376*, 660-669.

Johnson, Jr., W. C. (1990) *Proteins 7*, 207-215.

Keightley, J. A., Zimmermann, B. H., Mather, M. W., Springer, P., Pastuszyn, A., Lawrence, D. M. & Fee, J. A. (1995) *J. Biol. Chem.,270*, 20345-20358.

Kelly, M., Lappalainen, P., Talbo, G., Haltia, T., van der Oost, J. & Saraste, M. (1993) *J. Biol. Chem. 268*, 16781-16787.

Langen, R., Chang, I.-J., Germanas, J. P., Richards, J. H., Winkler, J. H. & Gray, H. B. (1995) *Science 268*, 1733-1735.

Lappalainen, P., Aasa, R., Malmström, B. G. & Saraste, M. (1993) *J. Biol. Chem. 268*, 26416-26421.

Lappalainen, P., Watmough, N. J., Greenwood, C. & Saraste, M. (1995) *Biochemistry 34*, 5824-5830.

Larsson, S., Broo, A. & Sjölin, L. (1995a) *J. Phys. Chem. 99*, 4860-4865.

Larsson, S., Källebring, B., Wittung, P. & Malmström, B. G. (1995b) *Proc. Natl. Acad. Sci. USA 92*, 7167-7171.

Lübben, M., Castresana, J. & Warne, A. (1994) *Syst. Appl. Microbiol. 16*, 556-559.

Marcus, R. A. & Sutin, N. (1985) *Biochim. Biophys. Acta 811*, 265-322.

Martin, C. T., Scholes, C. P. & Chan, S. I. (1988) *J. Biol. Chem. 263*, 8420-8429.

Massey, V. (1957) *J. Biol. Chem. 229*, 763 - 770.

Mather, M. W., Springer, P. & Fee, J. A. (1991) *J. Biol. Chem. 266*, 5025-5035.

Miller, C. G. (1987) in "*Escherichia coli* and *Salmonella typhimurium*," Neidhardt, F. C. (Ed.) ASM Press, p. 680 - 691.

Mitchell, P. (1961) *Nature 191*, 144-148.

Nar, H., Messerschmidt, A., Huber, R., van de Kamp, M. & Canters, G. W. (1991) *J. Mol. Biol. 218*, 427-447.

Öblad, M., Selin, E., Malmström, B., Strid, L., Aasa, R. & Malmström, B. G. (1989)

Biochim. Biophys. Acta 975, 267-270.

Oliveberg, M., Brzezinski, P. & Malmström, B. G. (1989) *Biochim. Biophys. Acta 977*, 322-328.

Oliveberg, M. & Malmström, B. G. (1991) *Biochemistry 30*, 7053-7057.

Pan, L. P., Hibdon, S., Liu, R.-Q., Durham, B. & Millett, F. (1993) *Biochemistry 32*, 8492-8498.

Pettersson, R. P. & Wobrauschek, P. (1995) *Nucl. Instr. Meth. Phys. Res. A 355*, 665-667.

Qui, D., Dong, S., Ybe, J., Hecht, M. & Spiro, T. G. (1995) *J. Am. Chem. Soc. 117*, 6443-6446.

Ramirez, B. E., Malmström, B. G., Winkler, J. R. & Gray, H. B. (1995) *Proc. Natl. Acad. Sci. USA 92*, 11949-11951.

Riester, J., Kroneck, P. M. H. & Zumft, W. G. (1989) *Eur. J. Biochem. 178*, 751-762.

Robin, M. & Day, P. (1967) *Advan. Inorg. Radiochem. 10*, 247-422.

Sands, R. H. & Dunham, W. R. (1975) *Quart. Rev. Biophys. 7*, 443-504.

Saraste, M. (1990) *Q. Rev. Biophys. 23*,331-366.

Siuzdak, G. (1994) *Proc. Natl. Acad. Sci. USA 91*, 11290 - 11297.

Slutter, C. E., Langen, R., Sanders, D., Lawrence, S. M., Wittung, P., Di Bilio, A. J., Hill, M. G., Fee, J. A., Richards, J. H., Winkler, J. R. & Malmström, B. G. (1996) *Inorg. Chim. Acta*, in press.

Solomon, E. I., Baldwin, M. J. & Lowerey, M. D. (1992) *Chem. Rev. 92*, 521-542.

Steffens, G. C. M., Biewald, R. & Buse, G. (1987) *Eur. J. Biochem. 164*, 295-300.

Steffens, G. J. & Buse, G. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 613-619.

Studier, F. W. & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113-130.

Taniguchi, V. T., Sailasuta-Scott, N., Anson, F. C. & Gray, H. B. (1980) *Pure Appl. Chem. 52*, 2275-2281.

Trumpower, B. L. & Gennis, R. B. (1994) *Annu. Rev. Biochem. 63*, 675-716.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawaitoh, K., Nakashima, R., Yaono, R. & Yoshikawa S. (1995) *Science* 269 1069-1074.

van der Oost, J., Lappalainen, P., Musacchio, A., Warne, A., Lemieux, L., Rumbley, J., Gennis, R. B., Aasa, R., Pascher, T., Malmström, B. G. & Saraste, M. (1992) *EMBO J. 11*, 3209-3217.

von Wachenfeldt, C., de Vries, S. & van der Oost, J. (1994) *FEBS Lett. 340*, 109-113.

Wang, H., Blair, D. F., Ellis, W. R., Gray, H. B. & Chan, S. I. (1986) *Biochemistry 25*, 167 - 171.

Weisberg, S. (1985) *Applied Linear Regression*, p. 203-223, Wiley, New York.

Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E. & Saraste, M. (1995) *Proc. Natl. Acad. Sci. USA 92*, 11955-11959.

Wittung, P., Källebring, B. & Malmström, B. G. (1994) *FEBS Lett. 349*, 286-288.

Zimmermann, B. H., Nitsche, C. I., Fee, J. A., Rusnak, F. & Münck, E. (1988) *Proc. Natl. Acad. Sci. USA 85*, 5779-5783.

Figure 2.1

The unique EPR parameters of the Cu_A site compared to other known copper centers is summarized in this plot of $A_{\parallel}$ *vs.* $g_{\parallel}$. The type 1 (1) and type 2 (2) centers are located in a lower and upper group, respectively. The superimposed circle and square represent the Cu_A sites for cytochrome *c* oxidase and nitrous oxide reductase.

[Malmström, B. G. & Aasa, R. (1993) **The Nature of the Cu_A Center in Cytochrome *c* Oxidase**. *FEBS Lett., 325*, 49-52.]

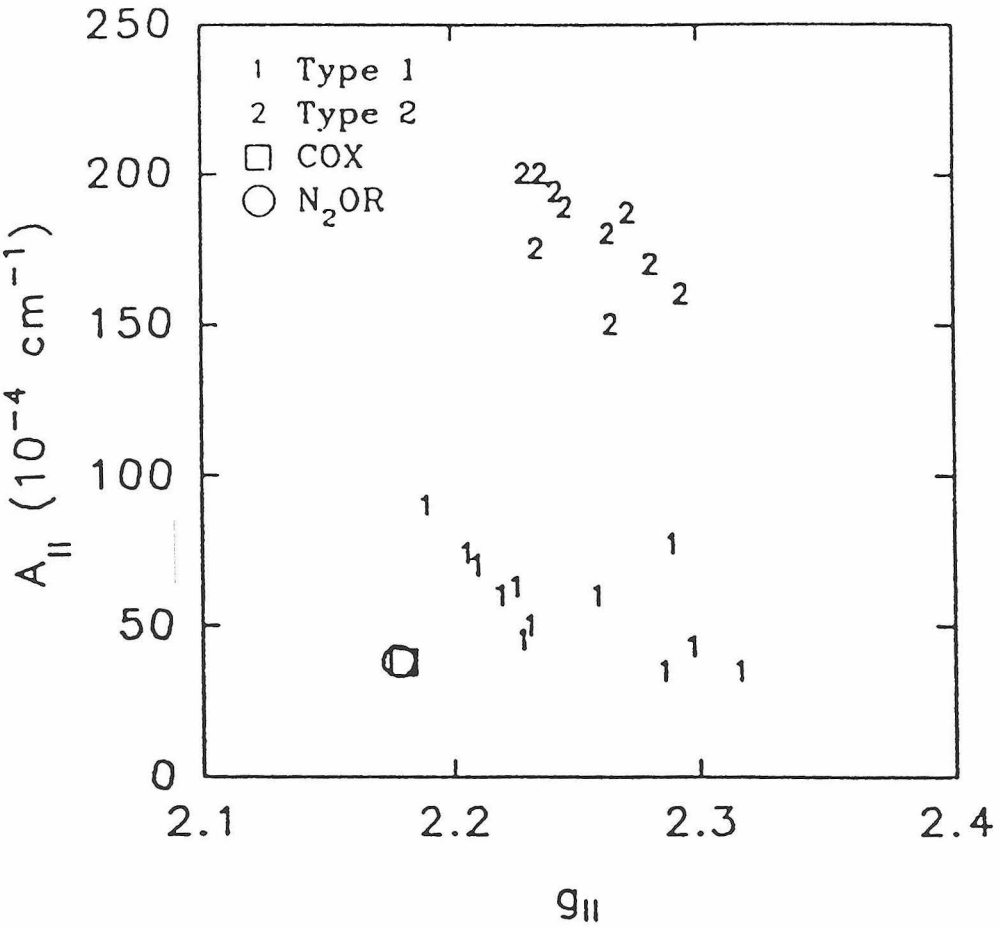

1 Type 1
2 Type 2
COX
N2OR
250
200
150
100
50
0
2.1
2.2
2.3
2.4
A∥ (10⁻⁴ cm⁻¹)
g∥

Figure 2.2

Primary amino acid sequence of subunit II of *T. thermophilus* cytochrome ba_3.

[Keightley, J. A., Zimmermann, B. H., Mather, M. W., Springer, P., Pastuszyn, A., Lawrence, D. M., & Fee, J. A. (1995) **Molecular Genetic and Protein Chemical Characterization of the Cytochrome ba_3 from *Thermus thermophilus* HB8***J. Biol. Chem.,270*, 20345-20358.]

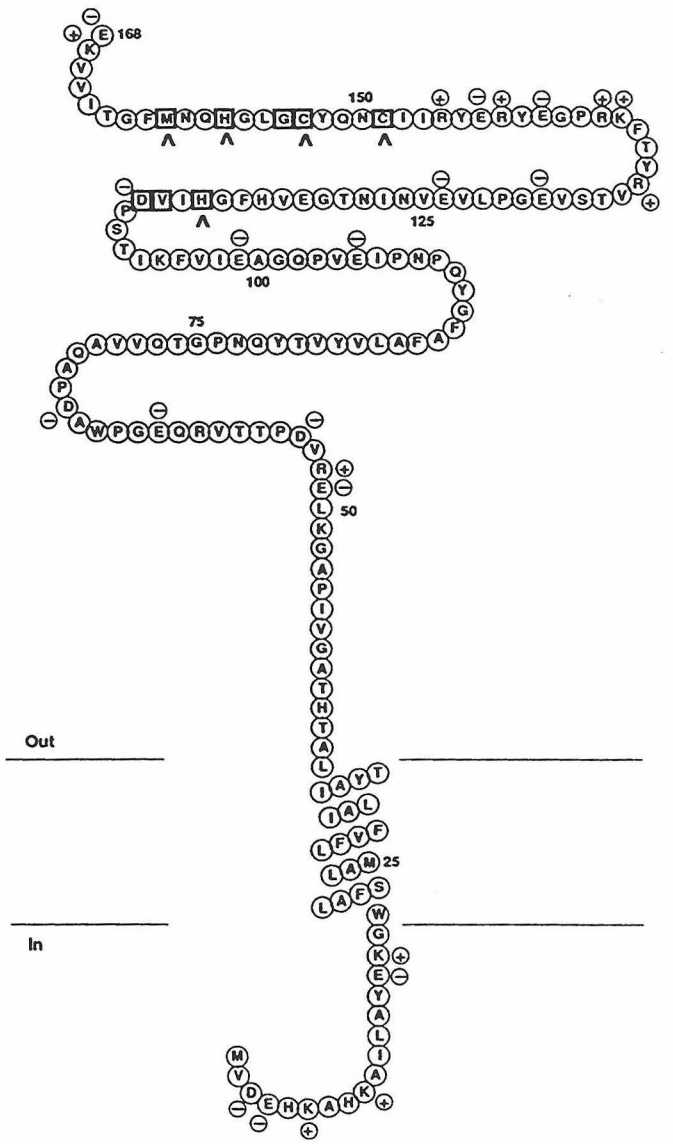

168
150
125
100
75
50
25
Out
In

Figure 2.3

Amino acid alignment of the *Thermus thermophilus* Cu_A fragment and engineered purple center in CyoA. The ligands are in bold face, the mutated ligands in CyoA are in italics.

[For additional alignments see also Keightley, J. A., Zimmermann, B. H., Mather, M. W., Springer, P., Pastuszyn, A., Lawrence, D. M., & Fee, J. A. (1995) **Molecular Genetic and Protein Chemical Characterization of the Cytochrome *ba*$_3$ from *Thermus thermophilus* HB8***J. Biol. Chem.,270*, 20345-20358.]

34	A	Y	T	L	A	T	H	T	A	G	V	I	P	A	G	K	L	E	R	V	54	*ba*3
	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		cyoA
55	D	P	T	T	V	R	Q	E	G	P	W	A	D	P	A	Q	A	V	V	Q	74	*ba*3
	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		cyoA
75	T	G	P	N	Q	Y	T	V	Y	•	V	L	A	F	A	F	G	Y	•	•	91	*ba*3
					.	.	.	|	:		|	:		:		:		:			
125	•	•	•	•	K	P	I	T	I	E	V	V	S	M	D	W	K	W	F	F	140	cyoA
92	•	•	•	•	•	•	•	•	Q	P	N	P	I	E	V	P	Q	G	A	E	103	*ba*3
										.	|	.	|	.	.	|	.	.	.	.	
141	I	Y	P	E	Q	G	I	A	T	V	N	E	I	A	F	P	A	N	T	P	160	cyoA
104	I	V	F	K	I	T	S	P	D	V	I	**H**	G	F	H	V	E	G	T	N	123	*ba*3
	:		|	|	:	|	|			|	:	:	:	|		:	.			.	
161	V	Y	F	K	V	T	S	N	S	V	M	*N*	S	F	F	I	P	R	L	G	180	cyoA
124	I	N	V	E	V	L	P	G	E	V	S	T	V	R	Y	T	F	K	R	P	143	*ba*3
		.	:		.	:	:	|			.		:	:	.	.		.		|		
181	S	Q	I	Y	A	M	A	G	M	Q	T	R	L	H	L	I	A	N	E	P	200	cyoA
144	G	E	Y	R	I	I	**C**	N	Q	Y	**C**	G	L	G	**H**	Q	N	**M**	•	F	162	*ba*3
	|	.	|			|	:			|	:	|		|		.	|		|	
201	G	T	Y	D	G	I	*S*	A	S	Y	*S*	G	P	G	*F*	S	G	**M**	K	F	220	cyoA
163	G	T	I	V	V	K	E	*	•	•	•	•	•	•	•	•	•	•	•	•	169	*ba*3
		.	|	.	.	.	:															
221	K	A	I	A	T	P	D	R	A	A	F	D	Q	W	V	A	K	A	K	Q	240	cyoA
	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
241	S	P	N	T	M	S	D	M	A	A	F	E	K	L	A	A	P	S	E	Y	260	cyoA
	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•		
257	N	Q	V	E	Y	F	S	N	V	K	P	D	L	F	A	D	V	I	N	K	280	cyoA
	•	•																				
279	F	M																			282	cyoA

Figure 2.4

Electrophoretic properties of purified ba_3-Cu_A protein. **A** SDS-PAGE: Total acrylamide concentration was 15% and the ratio to bis-acrylamide was 1:37.5. Protein was denatured in 0.35% SDS, 5% (β-mercaptoethanol, 2% glycerol and 6.25 mM Tris-HCl at 95 °C for 5 min. The gel was fixed and stained in a 45% methanol, 45% water and 10% acetic acid solution containing 0.25% Coomassie Brilliant Blue R250 from Bio-Rad. The left lane contained 10 μg purified Cu_A protein, while the right lane contained a total of 25 μg of the following molecular weight standards: serum albumin (66,200), ovalbumin (45,000) carbonic anhydrase (31,000), trypsin inhibitor (21,500), lysozyme (14,400) and aprotinin (6,500). **B** Non-denaturing PAGE: Stacking gel was 3.5% and the running gel was 15% acrylamide and the bis-acrylamide-acrylamide ratio was 1:37.5 in 20% glycerol, 50 mM Tris-HCl at pH 6.8; 20 μg of protein was used, and the gel was fixed and stained as described in **A**. **C** Thin layer gel isoelectric focusing, pH 3 - 7. In the right lane, standards and their isoelectric points are amylglucosidase (3.6), glucose oxidase (4.2), ovalbumin (4.8), β-lactoglobulin (A, 5.4; B,5.5), carbonic anhydrase (6.1), myoglobin (minor,7.0; major, 7.4); 10 μg Cu_A protein was used in the left lane. The holo-Cu_A protein has a pI_{obs} of 6.0 while the apoprotein has a pI_{obs} = 6.2.

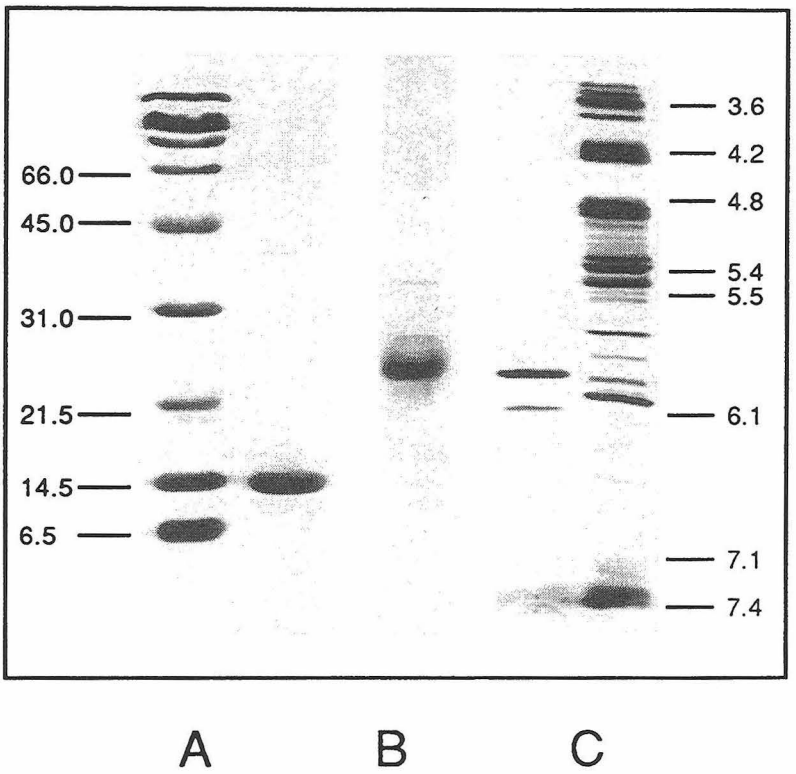

66.0
45.0
31.0
21.5
14.5
6.5
3.6
4.2
4.8
5.4
5.5
6.1
7.1
7.4
A
B
C

Table 2.1

Expected and observed composition of the cytochrome ba_3 - Cu_A Soluble Domain

Amino Acid	Expected[a]	Observed[b]
Asp[c]	9	11.5
Glu[d]	17	17.2
Ser	2	2.5
Gly	13	14.0
His	4	3.6
Arg	5	5.1
Thr	12	12.1
Ala	10	10.6
Pro	11	10.6
Tyr	7	6.6
Val	17	14.6
Met	1	0.2[e]
Cys	2	ND[f]
Ile	9	7.0
Leu	5	5.7
Phe	6	6.0
Lys	4	4.3
Trp	1	ND
Total	135	(121.0)[g]

[a]Based on the gene sequence from Keightley *et al.*, 1995.
[b]Determined experimentally as described in the Experimental section. The deviation in three different experiments was ~±10 %.
[c]Asn + Asp.
[d]Gln + Glu.
[e]In three separate analyses, this value has ranged from 0.2 to x.x. The value reported here is unique to this measurement.
[f]Not determined.
[g]Not corrected for non-determined amino acids.

Figure 2.5

Electrospray ionization mass spectra of *Thermus* cytochrome ba_3-Cu_A protein. Panel A shows the mass spectrum of the apoprotein. The experiment was done by first removing Na^+ ions from the solution by passsage over a short gel filtration column equilibrated with 10 mM ammonium acetate (see Materials and Methods section) then dissolving the protein into 60% HCOOH, 30% 2-propanol to a final concentration of ~2 mg/ml. The purple color disappeared in a few seconds, after which the sample was immediately admitted to the spectrometer. The principal component has a molecular mass of 14,804 Da followed by a peak due to apoprotein plus 1 Na^+ ion and a peak at 14,859 Da that is not assigned (see Text). Panel B shows the spectrum of the holoprotein. The protein solution was ~2 mg/ml in 10 mM ammonium acetate. The principal peak is at 14,928 Da and the 'ladder' of peaks trailing off at higher mass represent holoprotein having 1, 2, 3, etc. Na^+ ions bound: 14,950, 14,972, 14,994 and 15,015 Da. The lowest mass peak is the apoprotein at 14,803 Da, followed by two minor peaks due to apoprotein having 1 and 2 Na^+ bound. The small peak at 14,865 is probably due to the apoprotein having a single Cu bound (see Text).

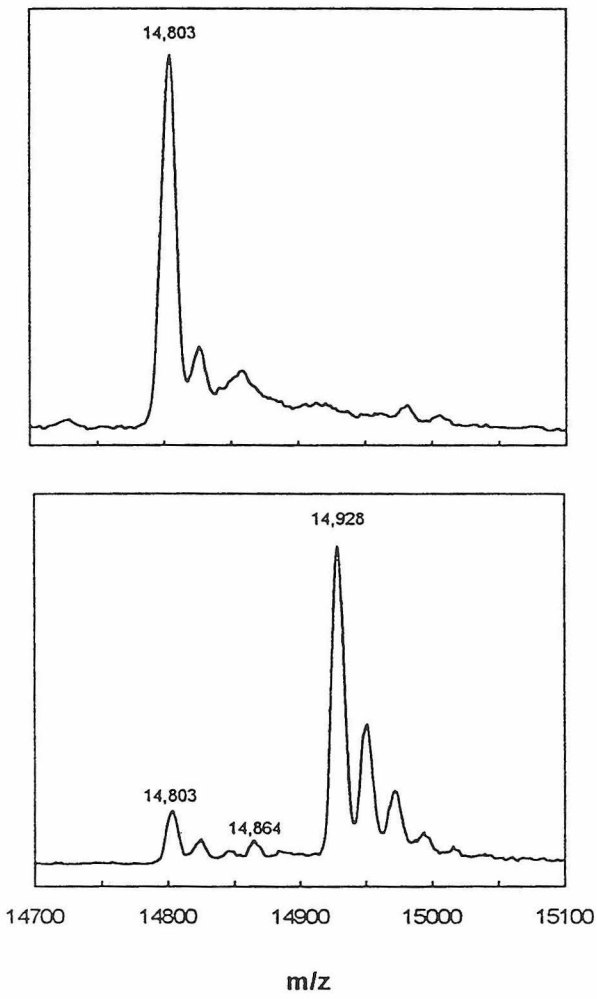

14,803
14,928
14,803
14,864
14700
14800
14900
15000
15100
m/z

Figure 2.6

Optical absorption and circular dichroism spectra of the blue copper protein azurin from *Pseudomonas aeruginosa* (dash-dot) and of the *Thermus* cytochrome ba_3-Cu_A protein (solid). The optical spectra are expressed as molar absorbance (ε) and the CD spectra are presented as differential molar absorption coefficients ($\Delta\varepsilon = \varepsilon_l - \varepsilon_r$).

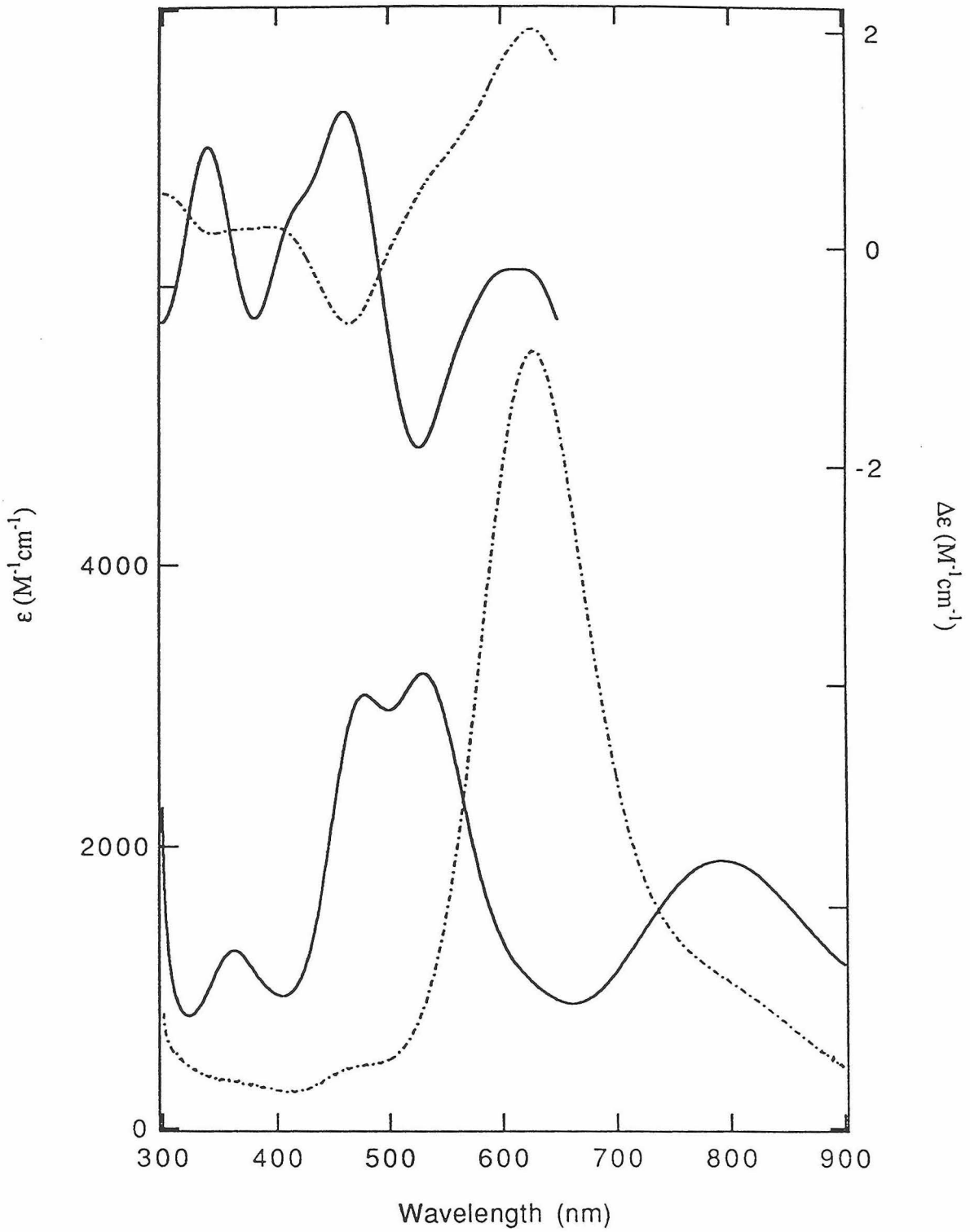

2
0
-2
4000
2000
0
300
400
500
600
700
800
900
Wavelength (nm)
ε (M-1cm-1)
Δε (M-1cm-1)

Table 2.2: Absorption Data for Cu_A Soluble Domains and Engineered Cu_A Sites

Protein	Source	Absorbance in nm			
Cu_A from aa_3 [a]	*Paracoccus denitrificans*	363	480	530	808
Cu_A from caa_3 [b]	*Bacillus subtilis*	365	480	530	790
Cu_A from ba_3 [c]	*Thermus thermophilus*	360	480	530	790
Cu_A engineered into CyoA [d]	*Escherichia coli*	358	475	536	765
Cu_A engineered into azurin [e]	*Pseudomonas aeruginosa*	350	485	530	765
Cu_A engineered into amicyanin [f]	*Thiobacillus versutus*	360	483	532	790
Cu_A from N_2OR [g]	*Achromobacter cycloclastes*	350	481	534	780
Cu_A from N_2OR [h]	*Pseudomonas stutzeri*	350	480	540	780

[a] Lappalainen *et al.*, 1993. [b] von Wachenfeldt *et al.*, 1994. [c] This work. [d] Kelly *et al.*, 1993. [e] Hay *et al.*, submitted. [f] Dennison *et al.*, 1995. [g] Hulse *et al.*, 1990. [h] Riester *et al.*, 1989.

Figure 2.7

X-band EPR spectra of *Thermus* cytochrome ba_3 and the cytochrome ba_3-Cu_A domain. Trace **A** shows the spectrum of oxidized cytochrome ba_3 recorded at 80 K. Trace **B** shows the spectrum of the oxidized cytochrome ba_3-Cu_A protein (77 K), and Trace **C** shows the latter after denaturation in 60% HCOOH/30% *iso*-propanol/10% H_2O (77 K). Spectrometer settings were typically as follows: frequency, 9.378 GHz; modulation amplitude, 10 - 12.5 Gauss; modulation frequency, 100 kHz; power, 2 mW; recording time, 200 s; time constant, 200 ms. Spectra are presented to show qualitative features. See text for details of quantitation.

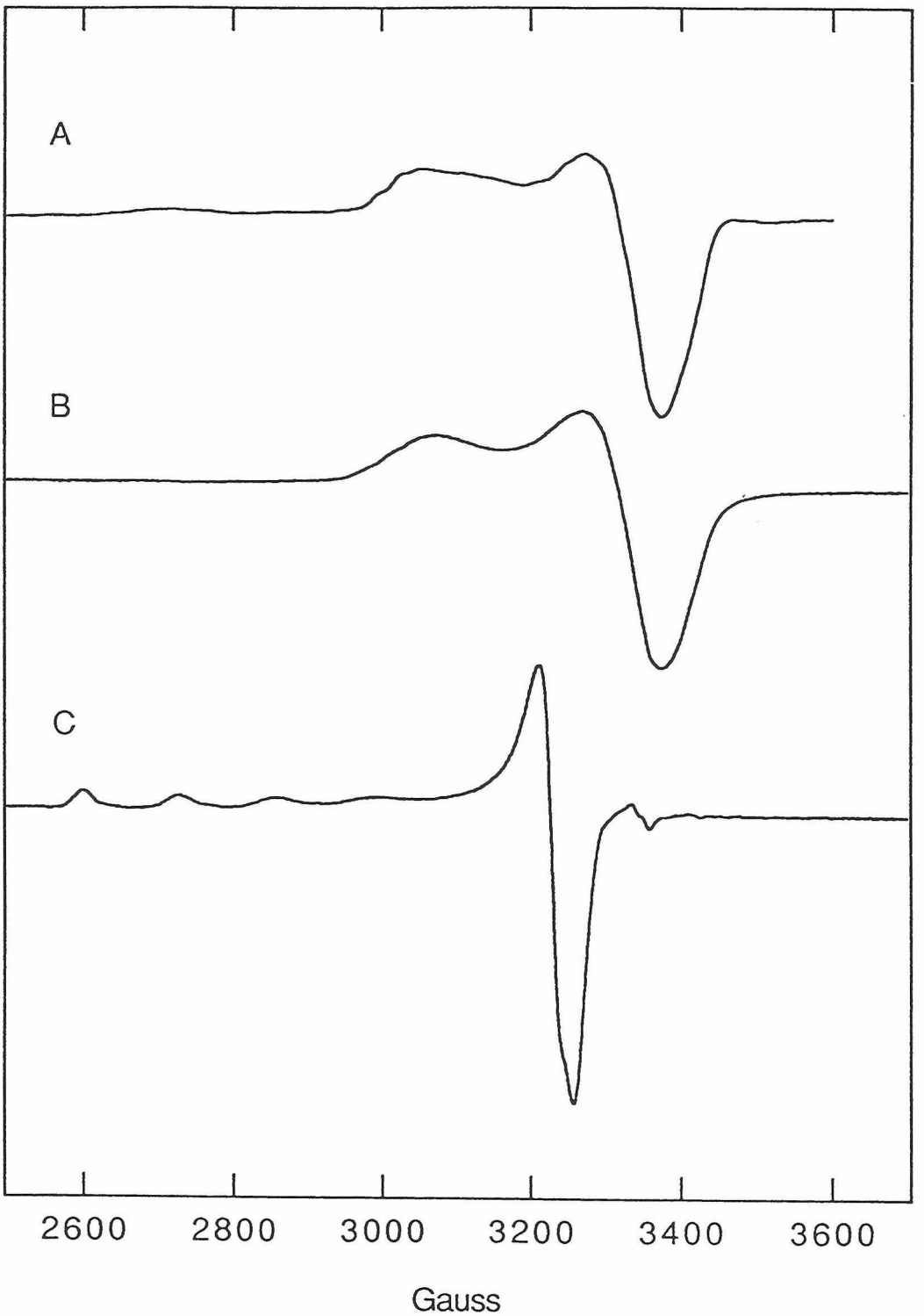

A
B
C
2600
2800
3000
3200
3400
3600
Gauss

Figure 2.8

Q-band (34.03 GHz), X-band (9.45 GHz) and S-band (3.93 GHz) EPR spectra of the Cu-enriched, soluble Cu_A-protein from the cytochrome ba_3 of *T. thermophilus*. The spectra are shown on a common g-value scale. The left dashed line indicates the position of g_z (= 2.187) and the right dashed line is placed on the fourth extremum in the perpendicular region. The narrow signal at g = 2.00 is due to a free-radical impurity. Experimental conditions for Q-, X- and S-band spectra, respectively: microwave power (mW), 0.66, 0.2, 2; modulation amplitude (mT), 1.8, 1.0, 0.63; time constant (ms), 82, 200, 100; sweep time (s), 41, 200, 100 and number of scans, 10, 1, 32. Protein concentration, 1.3 mM. All spectra were recorded at 20 K. The same tube was used in X- and S- band.

[Fee, J. A., Sanders, D., Slutter, C. E., Doan, P. E., Aasa, R., Karpefors, M. and Vänngård, T. (1995) **Multi-frequency EPR Evidence for a Binuclear CuA Center in Cytochrome *c* Oxidase: Studies with a ^{63}Cu- and ^{65}Cu-Enriched, Soluble Domain of the Cytochrome ba_3 Subunit II from *Thermus thermophilus***, *BBRC* 212 (1) 77-83.]

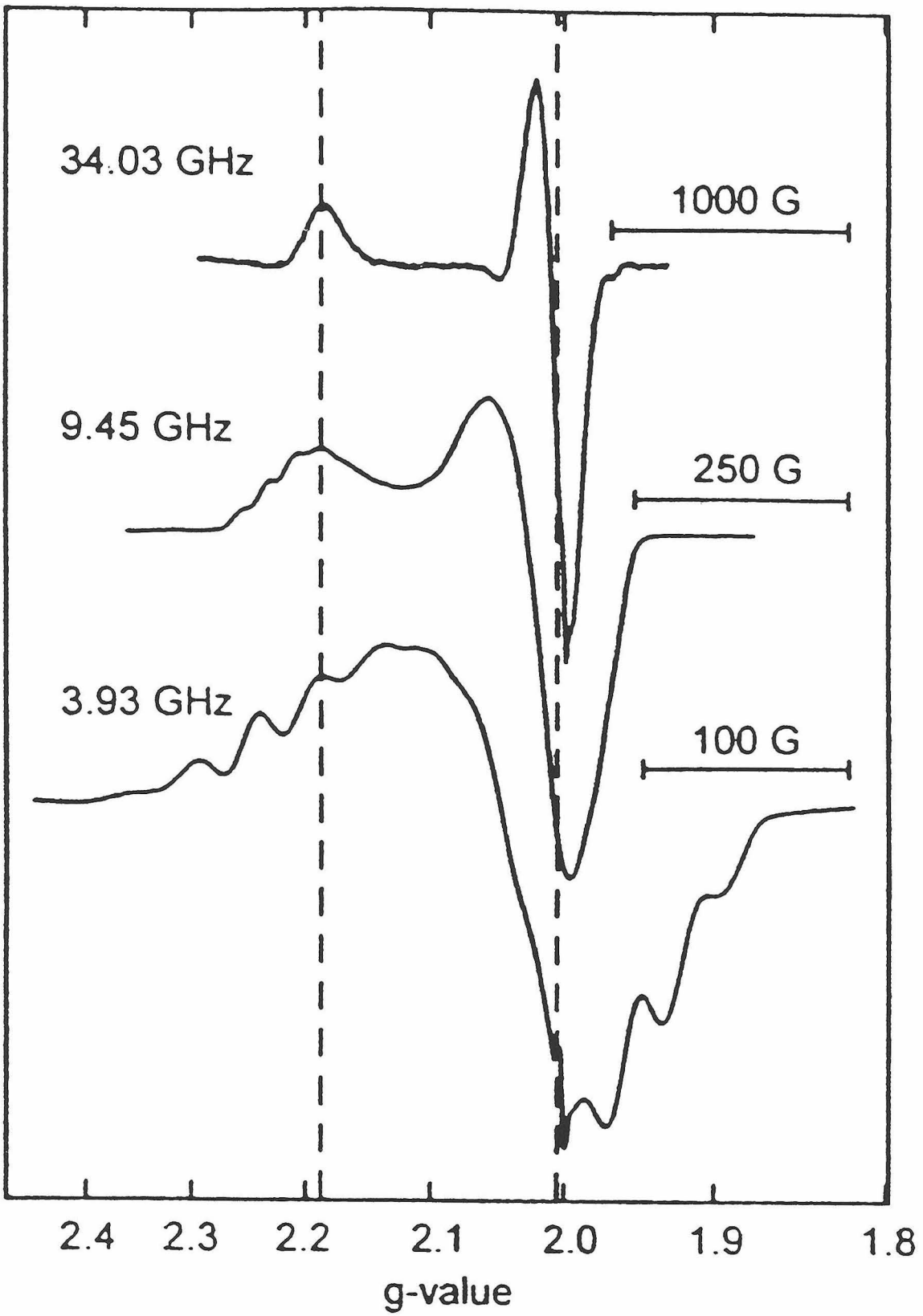
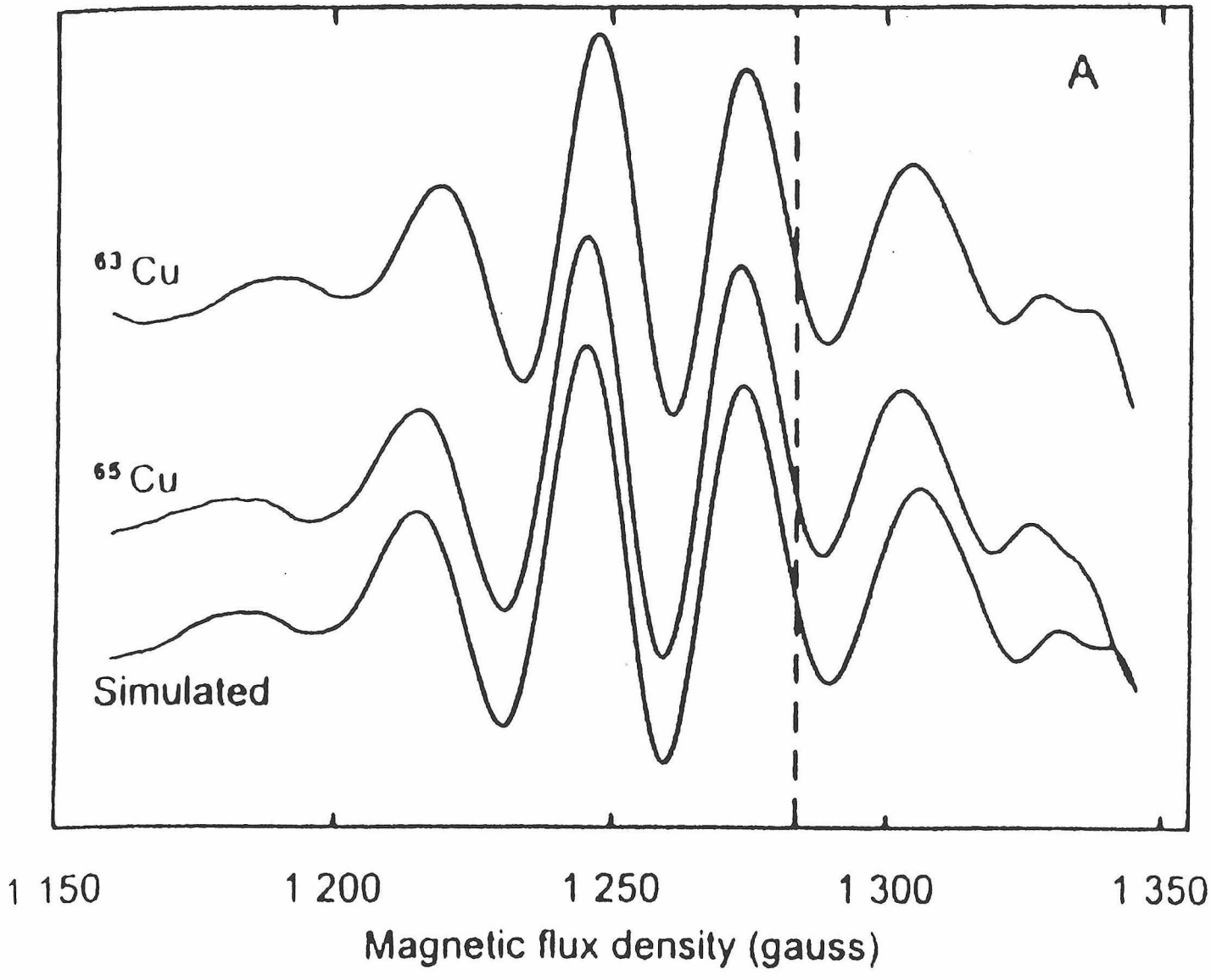

34.03 GHz
1000 G
9.45 GHz
250 G
3.93 GHz
100 G
2.4
2.3
2.2
2.1
2.0
1.9
1.8
g-value

Figure 2.9

Experimental and simulated S-band (3.93 GHz) EPR spectra of ^{63}Cu and ^{65}Cu EPR spectra of the solube Cu_A-protein from the cytochrome ba_3 of T. thermophilus presented as second derivatives. **A** The low field portion of ^{63}Cu, ^{65}Cu and "simulated" ^{65}Cu spectra (see results section). The deviation between the experimental and "simulated" ^{65}Cu spectrum, seen in the parallel region at around 1300 Gauss, is caused by contributions from the outermost hyperfine lines around the other *g*-values. **B** The high field region of the spectra shown in **A**. The very left of the spectra in **B** is dominated by the free radical also seen in Figure 2.5. The dashed lines show where the field axes of the experimental and "simulated" ^{65}Cu spectra coincide (see results). Experimental conditions are the same as in Figure 2.8.

[Fee, J. A., Sanders, D., Slutter, C. E., Doan, P. E., Aasa, R., Karpefors, M. and Vänngård, T., (1995) **Multi-frequency EPR Evidence for a Binuclear CuA Center in Cytochrome *c* Oxidase: Studies with a ^{63}Cu- and ^{65}Cu-Enriched, Soluble Domain of the Cytochrome ba_3 Subunit II from *Thermus thermophilus***, *BBRC* 212 (1) 77-83.]


A
^{63}Cu
^{65}Cu
Simulated
1 150
1 200
1 250
1 300
1 350
Magnetic flux density (gauss)

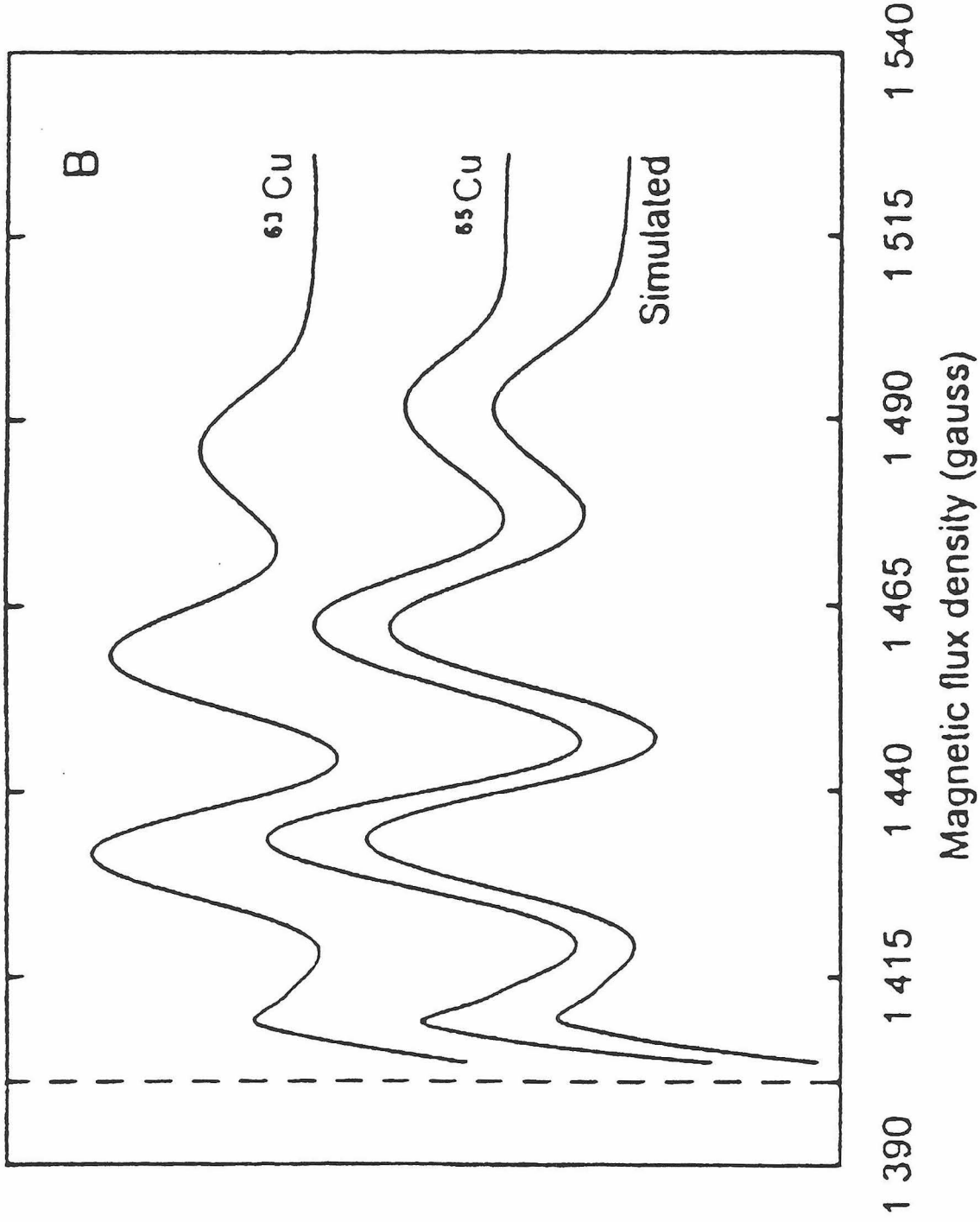

B
63 Cu
65 Cu
Simulated
1 390
1 415
1 440
1 465
1 490
1 515
1 540
Magnetic flux density (gauss)

Figure 2.10

Resonance Raman spectra of Cu_A sites upon 488-nm (~150-mW) excitation at 15 K. **A** Cytochrome *c* oxidase (CCO) fragment from *Paracoccus denitrificans* (2.0 mM in Cu_A) in 20 mM Bis-Tris (pH 6.5) from bacteria grown in ^{32}S (—) or ^{34}S-substituted (---) Na_2SO_4 prepared as in Lappalainen *et al.* (1993). **B** CCO fragment from *T. thermophilus* (1.8 mM in Cu_A) in 30 mM Tris-HCl, prepared as in Slutter *et al.* (1996). **C** CCO fragment from *Bacillus subtilis* (1.5 mM in Cu_A) in 20 mM Tris-HCl (pH 8.0), spectrum from Andrew *et al.* (1994). **D** Cu_A construct in *Pseudamonas aeruginosa* azurin (0.4 mM in Cu_A) in 50 mM NH_4OAc (pH 5.2) prepared as in Hay *et al.* (1996). **E** Cu_A construct in *Thiobacillus versutus* amicyanin (1.7 mM in Cu_A) in 50 mM HEPES buffer (pH 7.0), prepared as in Dennison *et al.* (1995).

[Andrew, C. R., Lappalainen, P., Saraste, M., Hay, M. T., Lu, Y., Dennison, C., Canters, G. W., Fee., J. A., Slutter, C. E., Nakamura, N. and Sanders-Loehr, J. (1995) **Engineered Curedoxins and Bacterial Cytochrome *c* Oxidases Have Similar Cu_A Sites: Evidence from Resonance Raman Spectroscopy**, *J. Am. Chem. Soc.* 117 (43) 10759-10760.]

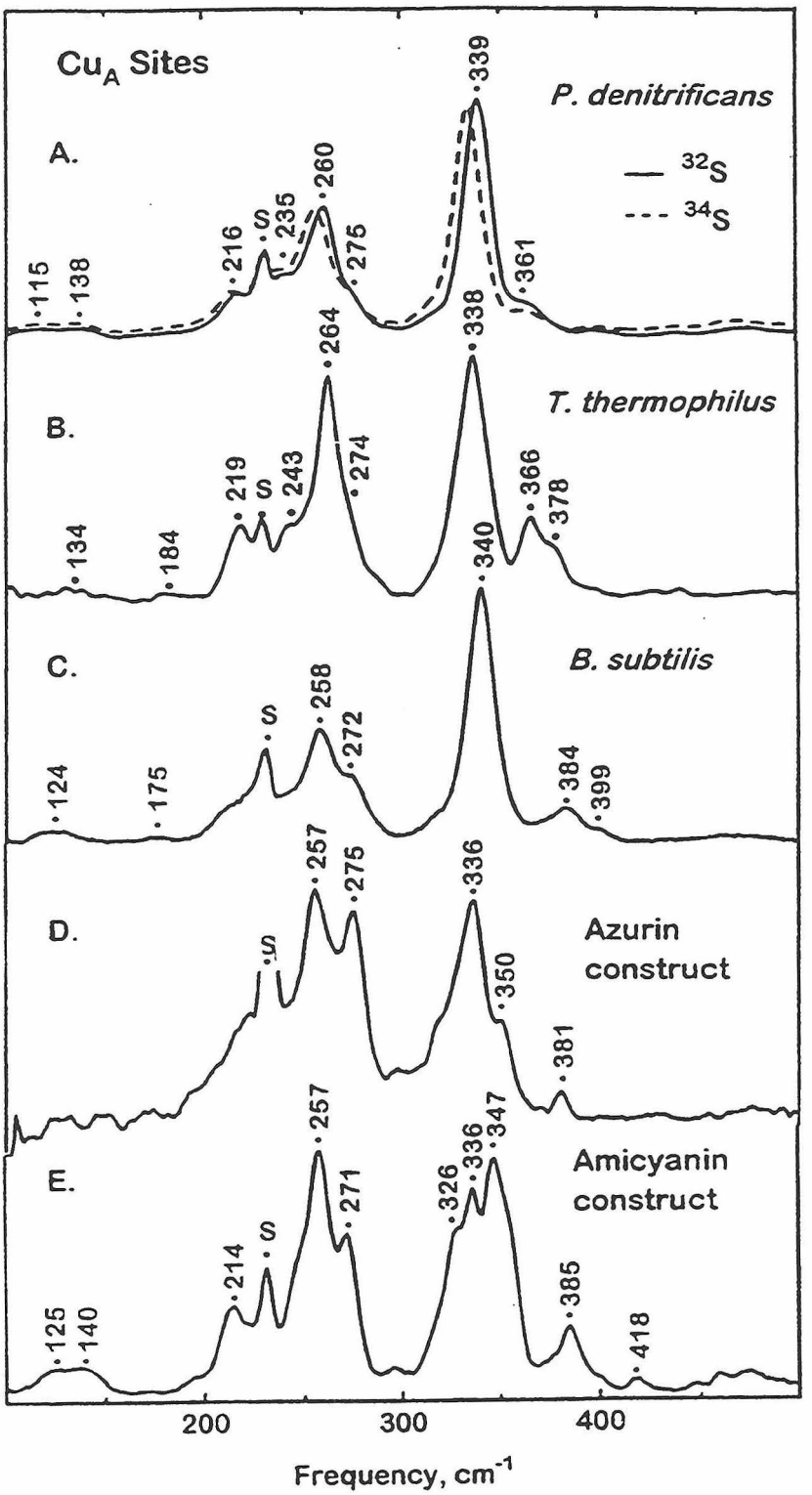

Cu$_A$ Sites
A.
P. denitrificans
^{32}S
^{34}S
115
138
216
S
235
260
275
339
361
B.
T. thermophilus
134
184
219
S
243
264
274
338
366
378
C.
B. subtilis
124
175
S
258
272
340
384
399
D.
Azurin construct
S
257
275
336
350
381
E.
Amicyanin construct
125
140
214
S
257
271
326
336
347
385
418
200
300
400
Frequency, cm^{-1}

Figure 2.11

Sulfur isotope downshifts ($^{32}S \rightarrow ^{34}S$) for different vibrational modes. Based on RR spectra for **A** Cu_A fragment from *P. denitrificans* (see Figure 2.7), **B** azurin from *P. aeruginosa* (Dave *et al.*, 1993) and **C** plastocyanin from poplar (Qiu *et al.*, 1995).

[Andrew, C. R., Lappalainen, P., Saraste, M., Hay, M. T., Lu, Y., Dennison, C., Canters, G. W., Fee., J. A., Slutter, C. E., Nakamura, N. and Sanders-Loehr, J. (1995) **Engineered Curedoxins and Bacterial Cytochrome *c* Oxidases Have Similar Cu_A Sites: Evidence from Resonance Raman Spectroscopy**, *J. Am. Chem. Soc.* 117 (43) 10759-10760.]

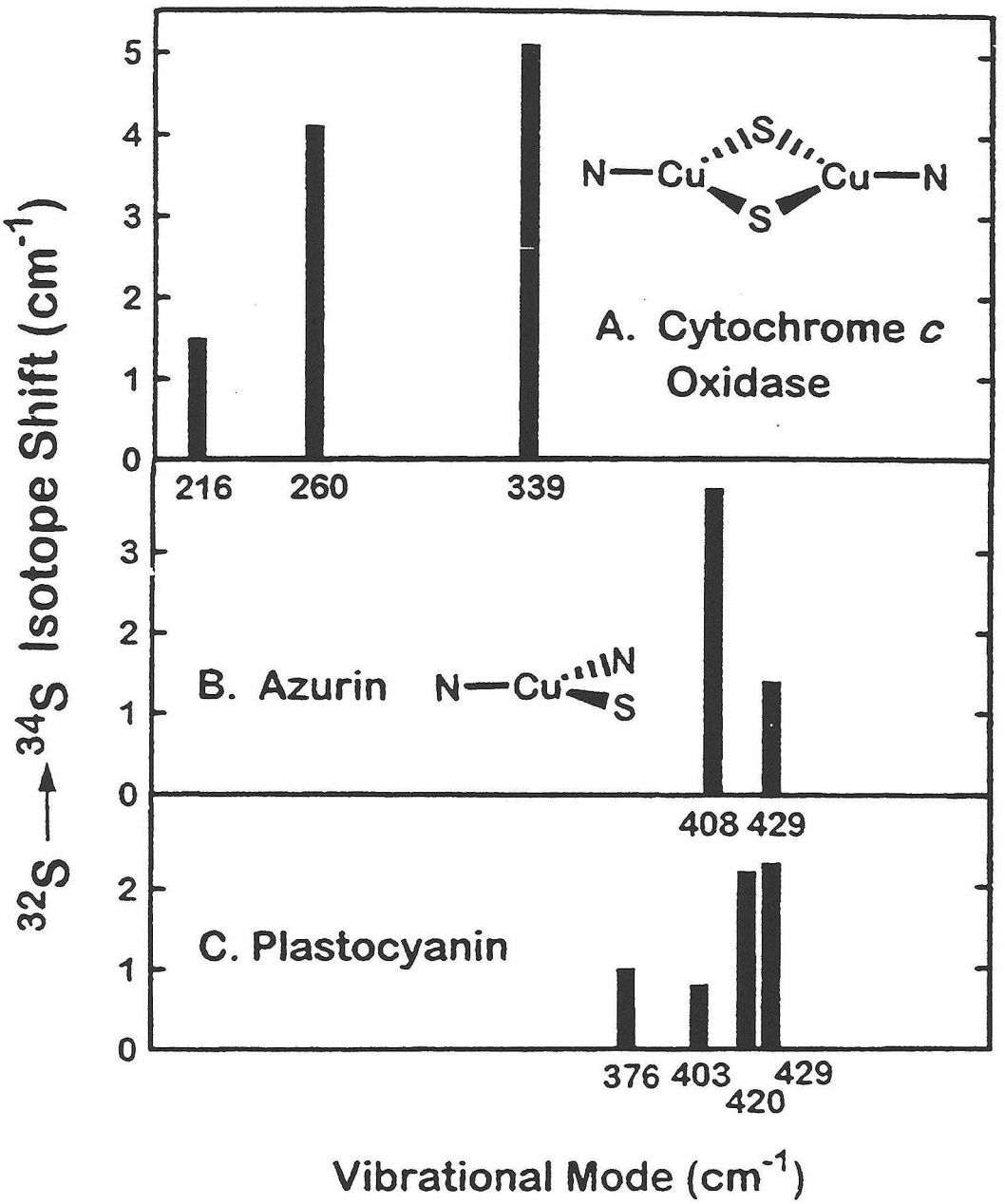
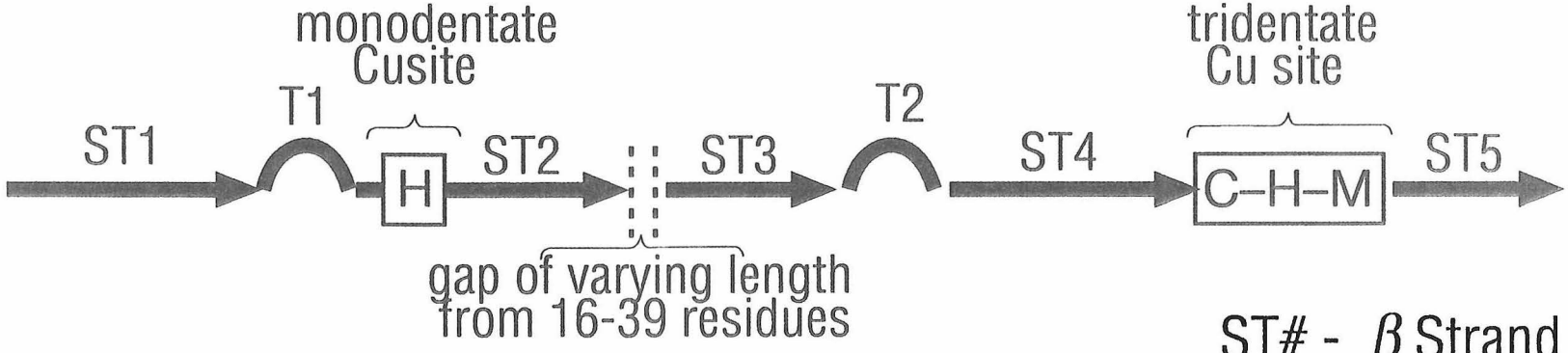

A. Cytochrome c Oxidase
N—Cu S Cu—N
S
216
260
339
B. Azurin
N—Cu N
S
408 429
C. Plastocyanin
376 403 420 429
$^{32}S \longrightarrow {}^{34}S$ Isotope Shift (cm^{-1})
Vibrational Mode (cm^{-1})

Figure 2.12

Schematic depicting a common secondary structural element for small blue (type 1) copper proteins aligned with the predicted secondary structural elements for Cu_A domains. (Ramirez, 1994, unpublished observations.) Note the similar placement of ligands within these elements for the purple and blue centers.

Secondary Structural Alignment of Azurin Proteins



ST# - β Strand

T# - β Turn

Speculative Secondary Structure for Copper A Domains

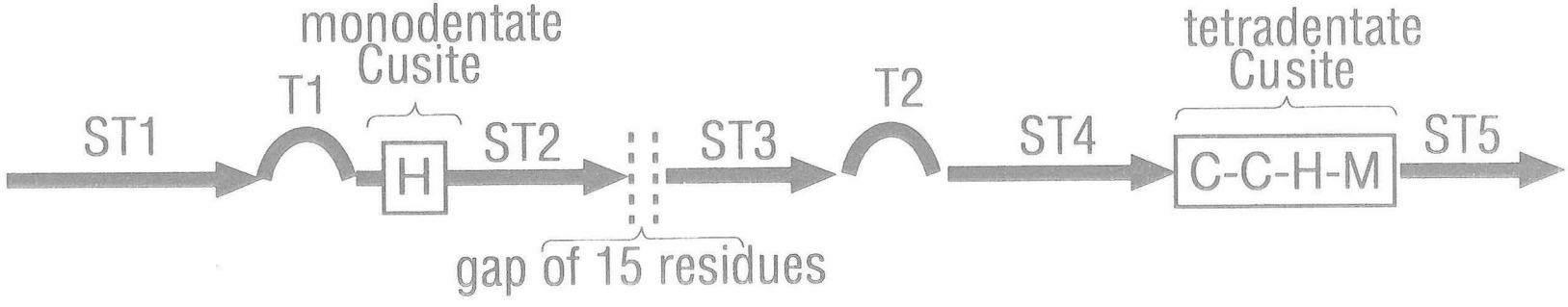


Figure 2.13

500 MHz ^1H NMR spectrum (T = 296 K) of the soluble, oxidized [Cu(1.5)Cu(1.5)] Cu_A fragment from *T. thermophilus* cytochrome *ba*$_3$. The Cu_A protein was 10 mM in 100 mM potassium phosphate and 200 mM KCl in 90% H_2O/10%D_2O. The spectrum was acquired on a Bruker AMX-500 spectrometer using presaturation to suppress the water signal. The insets show the resolved resonances (labeled a-j) with a 8x-greater vertical scale. Signal d is due to an exchangeable proton. All resonances integrate to one proton. Chemical shifts are referenced to internal DSS. (Bren *et al.*, 1995, unpublished results.)

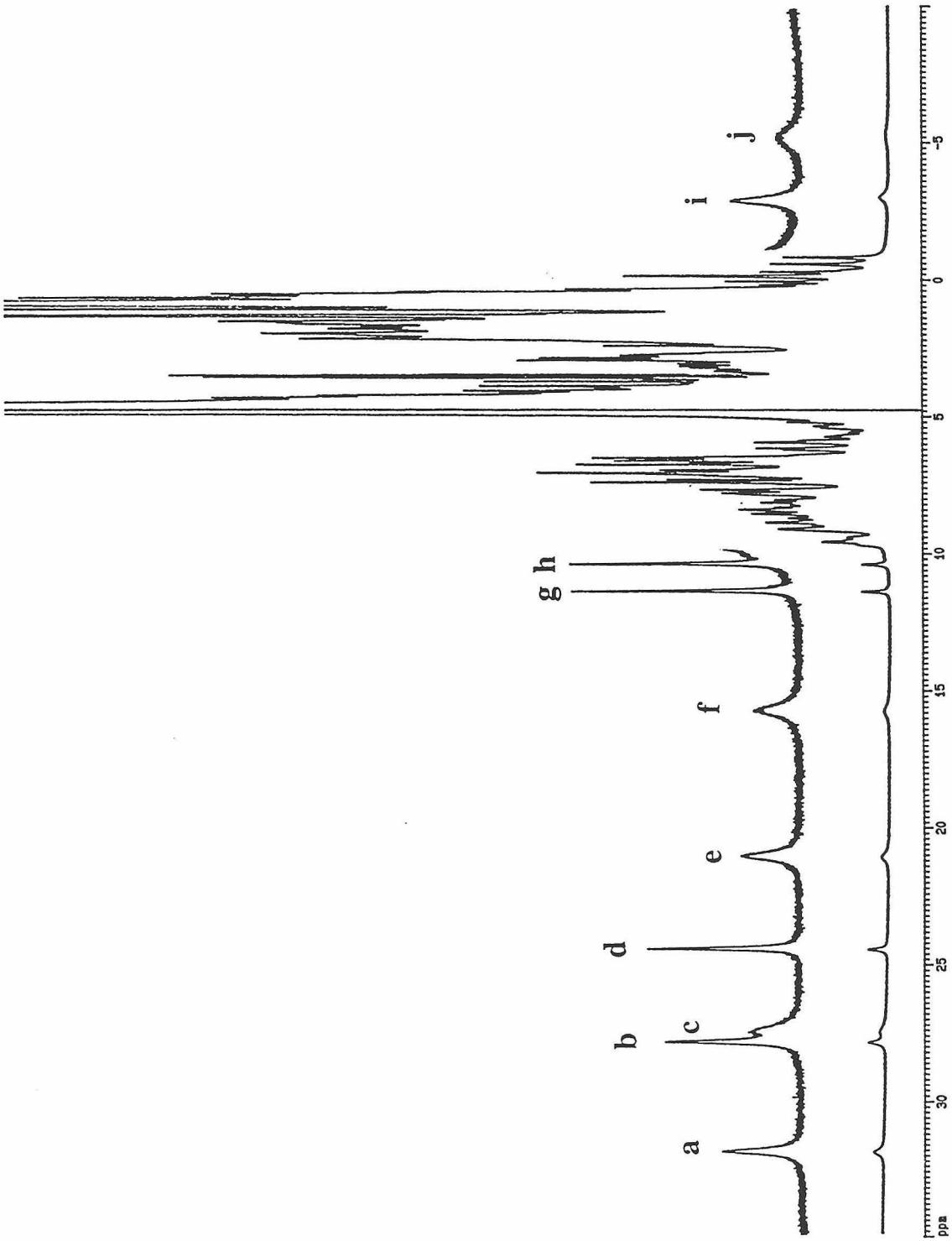

a
b
c
d
e
f
g
h
i
j
ppm
30
25
20
15
10
5
0
-5

Figure 2.14

Curie plot for the resolved ^1H NMR resonances of the oxidized Cu_A fragment. Experimental conditions are as explained in the caption for Figure 2.13, except that the temperature was varied from 296 to 323 K using the temperature control unit of the spectrometer. At each temperature, the sample was allowed to equilibrate for ten minutes before data collection was commenced. The changes in the spectrum were observed to be fully reversible in this temperature range. (Bren *et al.*, 1995, unpublished results.)

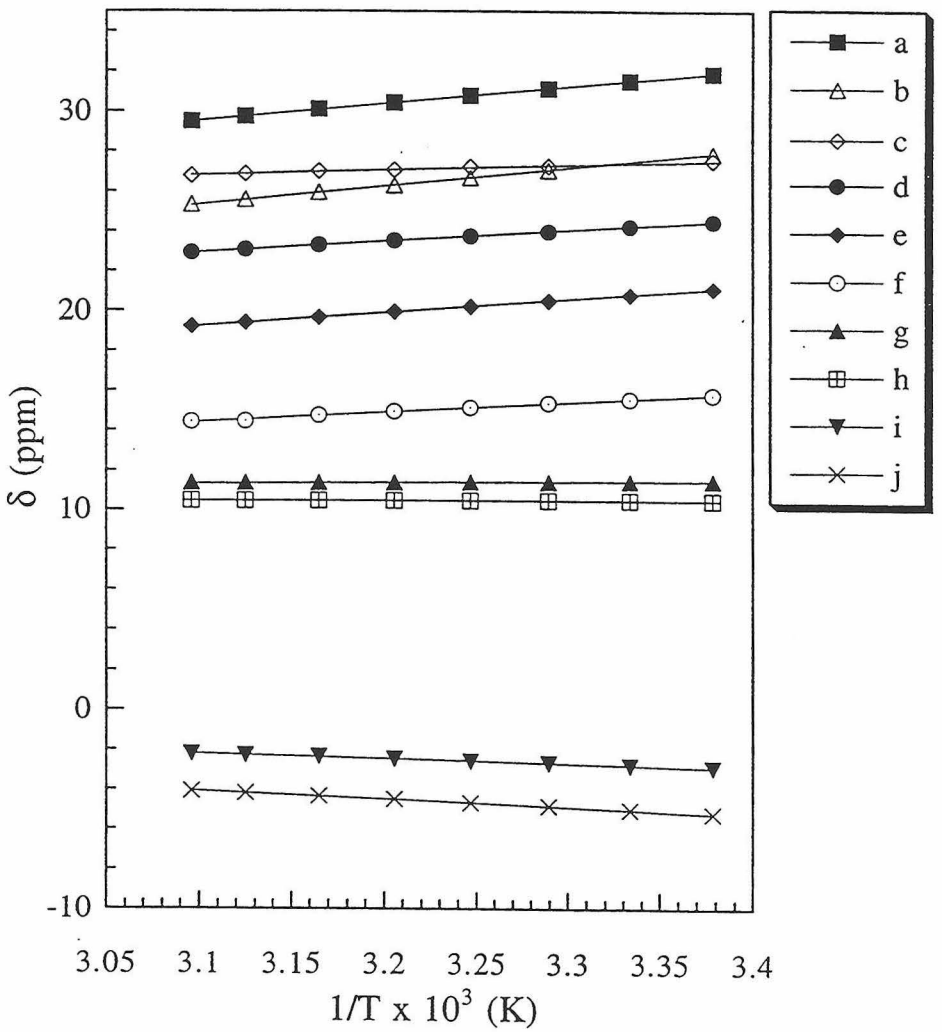
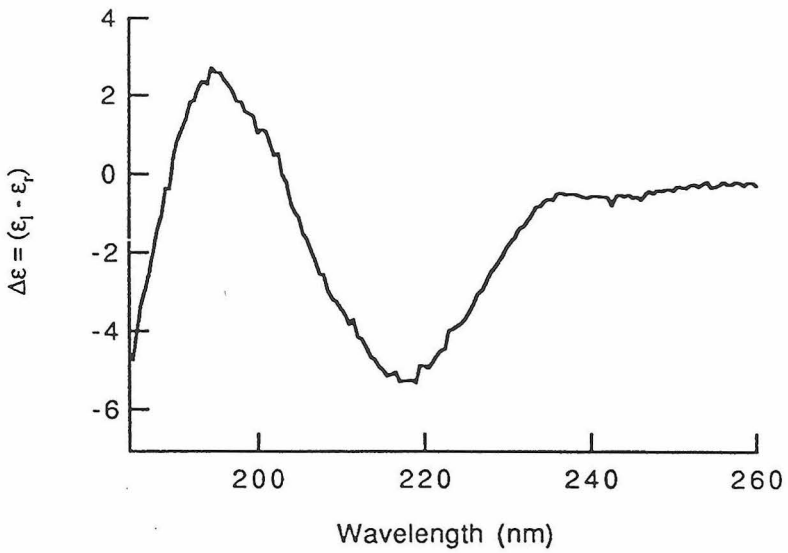

30
20
10
0
-10
δ (ppm)
3.05
3.1
3.15
3.2
3.25
3.3
3.35
3.4
1/T x 10³ (K)
a
b
c
d
e
f
g
h
i
j

Figure 2.15

Far UV-CD spectrum spectra of the *Thermus* cytochrome ba_3-Cu_A domain (upper panel) and the effect of temperature on the CD at 218 nm (lower panel). The spectrum was obtained from a solution containing ~0.05 mg protein/mL in 0.5 mM potassium phosphate buffer at pH 7.4. The pathlength was 1 mm. The effect of temperature (lower panel) shows the change in dichroism at 218 nm as a function of solution temperature. See text for details.


4
2
0
-2
-4
-6
Δε = (ε_l - ε_r)
200
220
240
260
Wavelength (nm)

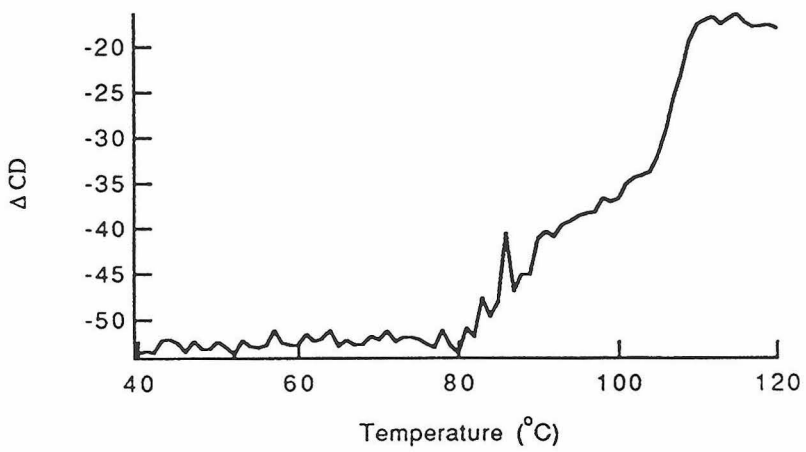

-20
-25
-30
-35
-40
-45
-50
Δ CD
40
60
80
100
120
Temperature (°C)

Table 2.3

Secondary Structure of the *Thermus* Cu_A domain from Circular Dichroism Spectra

Protein	α-helix	Antiparallel β-sheet	Parallel β-sheet	β-turn	Other	Sum
Azurin (X-ray)	10	33	11	22	24	100
Plastocyanin (X-ray)	4	38	18	28	12	100
Paracoccus Cu_A	14±6	31±10	17±6	27±5	12±8	101
Thermus Cu_A	17±7	20±11	11±8	13±5	39±8	100
E. coli CyoA	10±3	33±4	11±3	22±2	25±4	101

Chapter 3

Electron-transfer studies with the Cu_A domain of *Thermus thermophilus* cytochrome ba_3

INTRODUCTION

Cytochrome *c* oxidase is the terminal complex in the electron transport chain of aerobic respiration. This complex catalyzes the oxidation of four consecutive ferrocytochrome *c* molecules and the concurrent reduction of dioxygen to water. In the *aa*3-type oxidases, electrons from ferrocytochrome *c* initially reduce the copper A (Cu_A) and the heme *a* sites. These electrons are then transferred to the bimetallic active site. Here dioxygen, bound between the iron of a high-spin heme, heme *a*3, and a type 2 copper site (Cu_B), is reduced to water.

The mitochondrial cytochrome oxidases are 13 subunit, transmembrane proteins. Because the size and complexity of these proteins make them difficult to manipulate, many investigations have relied on the homologous bacterial oxidases. These oxidases have only three subunits, the functional core of the mitochondrial oxidases. In all the known oxidases of the *aa*3-type, subunit I contains the bimetallic and heme *a* sites, and the Cu_A site is located on subunit II.

Labeling studies of bovine cytochrome oxidase in the presence and absence of cytochrome *c* binding have shown that the cytochrome *c* /oxidase complex protects Asp 112, Glu 114 and the conserved Glu 198 of subunit II (Millett *et al.*, 1983) (bovine oxidase sequence numbering). Sequentially, Glu 198 lies between the cysteine ligands of the Cu_A site, suggesting that Cu_A is proximal to the primary cytochrome *c* binding site and functions as the initial acceptor of electrons from ferrocytochrome *c*.

The enigmatic spectroscopic properties of Cu_A (see Malmström, 1990, for a review) have engendered numerous questions about the structure of this unique copper site. The UV/Vis absorption spectrum of cytochrome oxidase is dominated by the intense absorption bands of the bound heme groups. However, these spectra also contain a fairly

intense absorption band centered around 790 nm due to the presence of the Cu_A site. Recently, soluble fragments of the Cu_A domain of cytochrome oxidases from *Paracoccus denitrificans* (Lappalainen *et al.*, 1993), *Bacillus subtilis* (Wachenfeldt *et al.*, 1994) and *Thermus thermophilus* (Slutter *et al.*, 1995) have been produced by overexpressing the protein encoded by truncated cytochrome oxidase genes. The optical spectra of these fragments, free from the heme absorption bands, reveal two closely spaced bands of nearly equal intensity at 480 and 530 nm and a weaker band at 360 nm.

The EPR spectra of the type 1 blue copper proteins have small hyperfine coupling constants relative to the type 2 copper proteins or synthetic copper complexes. However, relative to the type 1 copper center, the EPR for the cytochrome oxidase Cu_A site has unusually small hyperfine coupling constants. In fact, this hyperfine splitting is only partially resolvable in X-band EPR spectra, but at S-band frequencies the Cu_A EPR spectrum reveals at least five isolated hyperfine lines. Antholine *et al.* (1992) have suggested that the Cu_A site is binuclear and should exhibit seven hyperfine lines in the EPR spectrum. Recent work with the soluble domain from *T. thermophilus* cytochrome ba_3 fully supports the conclusion that Cu_A is a mixed-valence [Cu(II)/Cu(I)] cluster in which the unpaired electron is distributed equally over both ions (Fee *et al.*, 1995).

The recent X-ray structures of *P. denitrificans* cytochrome aa_3 (Iwata *et al.*, 1995), bovine cytochrome aa_3 (Tsukihara *et al.*, 1995) and the engineered purple center in the CyoA fragment (Wilmanns *et al.*, 1995) reveal that Cu_A is a binuclear site in which two cysteine thiolates bridge two Cu to form a four-membered ring. In addition, each Cu has one strong coordinate bond to an imidazole ring of histidine (Gurbiel *et al.*, 1993) and a much weaker coordinate bond to either the sulfur of the conserved methionine or to the carbonyl oxygen of a nearby peptide bond. In all the structures, both Cu appear to be bound in a highly distorted geometry and are separated by ~2.5 Å, which is consistent with EXAFS data (Blackburn *et al.*, 1994; see also Bertagnoli & Kaim, 1995). The observations reported here, and in the preliminary communication (Fee *et al.*, 1995), show

that cytochrome ba_3 from *T. thermophilus* also contains a Cu_A center. While this enzyme is currently the most divergent of the heme-copper oxidases (Lübben *et al.*, 1994), comparative analysis of its amino acid sequence (Keightley *et al.*, 1995) suggests that the overall three-dimensional structure of the ba_3 -Cu_A protein will be similar to that found in the recent X-ray structures. Finally, the pioneering work of Kroneck, Antholine and Zumft and their co-workers (Antholine *et al.*, 1992) demonstrates that nitrous oxide reductase also contains a Cu_A center. This 'purple copper' center is thus widely distributed in Nature.

The biological function of the Cu_A center involves transfer of electrons from reduced cytochrome *c* to the other centers in cytochrome *c* oxidase. Sequence analyses (Steffens *et al.*, 1979) and mutagensis studies (van der Oost *et al.*, 1992; Kelley *et al.*, 1993) indicate a possible evolutionary relationship between soluble blue copper proteins and the soluble, C-terminal domain of subunits II. Moreover, the blue copper proteins are efficient electron transfer proteins with redox potentials comparable to those measured for Cu_A in the soluble domain (240 mV, 10 mM Tris, pH 8.0, 200 mM KCl, Hill & Slutter, 1995). This raises the question of why Nature chose a binuclear purple site, rather than a mononuclear blue site, to serve as the primary electron acceptor in cytochrome *c* oxidases. While mononuclear sites have redox potentials suitable for this role, preliminary electrochemical characterization (Imoos, C. and Hill, M., unpublished results) suggests that Cu_A may have a significantly lower reorganization energy than that of mononuclear copper sites. The rate of electron transfer between two redox centers in a protein or within a protein-protein complex is determined by three factors (Marcus & Sutin, 1985): (1) the driving force, i.e., the ΔG^o of the reaction; (2) the reorganization energy, which depends on changes in the nuclear coordinates accompanying electron transfer; and (3) the electronic coupling between the donor and acceptor, which itself is determined by the donor-acceptor distance and the structure of the intervening medium (Langen *et al.*, 1995). Because access to the solvent would increase the reorganization energy (Churg & Warshel, 1986), it is generally believed that electron transfer redox sites are shielded from solvent, and this necessarily

implies that distances are not minimized in natural electron transfer partners. Nevertheless, the driving force for most biological electron transfers is rather modest, and for electron transfer from cytochrome *c* to Cu_A, it is close to zero, the reduction potentials being 255 mV (Taniguchi *et al.*, 1980) and ~240 mV (*cf.* Slutter *et al.*, 1996), respectively. In the *P. denitrificans* oxidase structure (Iwata *et al.*, 1995), Trp-121 and Asp-178 of subunit II shield the Cu_A site from solvent. With the low driving force and the larger distance, an appreciable electron transfer rate from cytochrome *c* to Cu_A may only be achieved if the reorganization energy of the latter is quite small. It is reasonable to suggest that one of the rack-induced properties (Malmström, 1994) of Cu_A is a greatly lowered reorganization energy. Indeed, a theoretical analysis based on electronic spectroscopic properties indicates that this energy decreases from about 0.4 eV in a blue site to less than 0.2 eV in a purple site (Larsson *et al.*, 1995).

Previous studies of electron tunneling in ruthenium-modified azurins (Langen *et al.*, 1995; Regan *et al.*, 1995) have shown that good electronic coupling over rather large distances could be mediated along β-strands. In the *P. denitrificans* oxidase structure (Iwata *et al.*, 1995), the distances from the edge of the conjugated π-electron system of the Cu_A ligand, His-224, to the edge of heme *a* and heme a_3 are ~12 Å and ~15 Å, respectively. Therefore, the fact that the Cu_A domain is not properly oriented to take advantage of this possibility is somewhat surprising. In the *P. denitrificans* oxidase structure, the Cu_A ligand His-224 is hydrogen bonded to Arg-473 in loop XI-XII of subuit I, and several other residues of this loop are in contact with the propionate groups of heme *a*, suggesting a possible electron transfer path from Cu_A to cytochrome *a* (Iwata *et al.*, 1995).

In this paper, we present the first electron transfer studies on a soluble Cu_A fragment from cytochrome ba_3, from *Thermus thermophilus*. This enzyme is similar to the aa_3-type oxidases, except that a *b*-type cytochrome replaces the cytochrome *a*. In prior work, Nilsson (1992) demonstrated that the positively charged tris(2,2'-bipyridyl)ruthenium(II) ion ($Ru(bpy)_3^{2+}$) mimics the electrostatic binding interaction

between cytochrome oxidase and cytochrome *c* at low ionic strength and high pH. Furthermore, he found that the metal-to-ligand charge transfer (MLCT) excited state of this ion ($^*Ru(bpy)_3^{2+}$) rapidly injects electrons into the Cu_A domain of cytochrome oxidase. Inhibiting charge recombination by the addition of aniline to scavenge the transient $Ru(bpy)_3^{2+}$ permitted the observation of electron transfer from reduced Cu_A to oxidized cytochrome *a*. We have used $Ru(bpy)_3^{2+}$ to study photoinduced electron injection into the soluble *T. thermophilus* Cu_A domain. At low ionic strength, we observe second-order kinetics with low reactant concentrations the rates tend to saturate, indicating complex formation. As a preliminary to future work with ruthenium-modified cytochrome *c*, binding studies of horse heart cytochrome *c* with the *T. thermophilus* Cu_A domain are also presented. These experiments demonstrate that these unnatural partners form a complex.

MATERIALS AND METHODS

Expression and Purification. The cytochrome ba_3 oxidase gene of *Thermus thermophilus* has been cloned (Keightly *et al.*, 1995). The construction and expression of the soluble fragment will be presented elsewhere (Slutter *et al.*, 1995) and in Chapter 2.

Electrochemistry. All electrochemical experiments were carried out using a Bioanalytical Systems (BAS) Model CV50-W potentiostat. Measurements were performed in 10 mM Tris buffer (pH 8) and 200 mM KCl with a thiosemicarbazide-modified gold electrode (Hill, 1991). Potentials were recorded vs. SCE and are reported vs. NHE. Cyclic voltammetry yielded a one-electron reduction for the Cu_A fragment; a plot of the peak current vs. the square root of the scan rate gave a straight line, confirming a diffusion-controlled process.

Electron Transfer Rate Measurements. The fully oxidized Cu_A fragment in 5 mM

Tris, pH 8.1, 40μm $Ru(bpy)_3^{2+}$, with or without 1M NaCl, was degassed and equilibrated under nitrogen in vacuum cells with 1-cm quartz cuvette side arms. The samples were excited with 2.5 mJ, 20 ns pulses at 480 nm generated by a XeCl excimer-pumped dye laser (Lambda Physik LPX 210i, FL-3002). Luminescence-decay kinetics were recorded at 630 nm, and transient absorption was measured at selected wavelengths. Kinetics traces represent an average of at least 1000 laser shots. Data were analyzed with nonlinear least-squares routines, fitting to both single exponential and biexponential functions.

Binding of cytochrome c. Horse heart cytochrome *c* (Type VI) was purchased from Sigma Chemical Co. Optical absorption spectra were recorded on an SLM AMINCOR DB-3500 UV/Vis spectrophotometer. Numerical manipulations of the spectra were performed with the wave analysis program IGOR (Wavemetrics).

RESULTS AND DISCUSSION

Cyclic Voltammetry and redox potential measurements. The cyclic voltammogram (CV) of Cu_A in 0.1 M Tris (pH=8) at a thiosemicarbazide-modifed (Hill *et al.*, 1987) gold electrode is shown in Figure 3.1. At 25 °C, Cu_A exhibits a reversible reduction (ΔE_p = 60 mV, ν = 10 mV/sec) with E° = 0.24 V *vs.* NHE; plots of i_{pc} (peak cathodic current) and i_{pa} (peak anodic current) *vs.* scan rate$^{1/2}$ yield straight lines (Figure 3.2), indicating a diffusion-controlled process (Bard *et al.*, 1980). Attempts to oxidize the Cu_2^{3+} complex gave irreversible anodic currents, presumably due to the Cu(II)-catalyzed oxidation of the bridging thiolates to disulfide species.

Thermodynamic parameters for the Cu_2^{3+}/Cu_2^{2+} electrode reaction were determined by variable-temperature electrochemistry. Using a nonisothermal cell configuration (Yee *et al.*, 1979), potentials for the Cu_A reduction were measured between -4 and 38 °C (Figure 3.3); these data yield a reaction entropy, ΔS°_{rc} (where ΔS°_{rc} =

$S^\circ_{(Cu_2^{2+})} - S^\circ_{(Cu_2^{3+})}$), of -5.4 eu. Correcting for the entropy of the NHE reference electrode (assuming $S^\circ_{H_2} = 31.2$ eu, and $S^\circ_{H^+} = 0$. Latimer, 1973), we calculate a standard entropy change for the complete cell reaction of $\Delta S^\circ = -21$ eu, giving $\Delta H^\circ = -11.9$ kcal/mol and $\Delta G^\circ = -5.6$ kcal/mol.

Thermodynamic parameters for the electrochemical reductions of a number of metalloproteins have been obtained (Taniguchi *et al.*, 1980). The negative enthalpies normally associated with these processes are a reflection of protein-ligand interactions that stabilize the lower oxidation states of transition metals (e.g., Cu(I) vs. Cu(II)); negative entropies have been attributed to increased solvent ordering within the protein, and/or enhanced hydrophobic interactions that occur upon reducing the charge at the metal center (Taniguchi *et al.*, 1980 and references therein; Sailasuta *et al.*, 1979). Our values for the Cu_A site are consistent with this picture: the weak Met227 S-Cu and Glu218 O-Cu bonds (2.7 and 3.0 Å, respectively for *P. denitrificans*) should favor the reduced (Cu_2^{2+}) state, and hydrophobic sidechains (e.g., Trp 121) are close enough to the binuclear site to shield the Cu_2^{2+} center from interactions with other polar functional groups. While in qualitative agreement with thermodynamic parameters for other redox-active metalloproteins, the ΔS°_{rc} value for Cu_A is unusually small. For example, several cytochromes and azurins exhibit ΔS°_{rc} values of ~ -10 to -20 eu; and the ET reaction entropies for ferredoxins can be even more negative (Lui *et al.*, 1994).

Given the proximity of Met227 to one of the Cu atoms and Glu218 to the other, it is remarkable that the binuclear center does not collapse into a C(II)-Cu(I) valence-trapped state, with the "soft" Met(S) bound to the Cu(I) and the "hard" glu(O) bound to Cu(II): (Note that the inherent advantages of the binuclear site--low reorganization energy and the ability to delocalize an electron over a large region of space (Recent work suggests a λ of 0.25 eV for the binuclear center. Larsson *et al.*, 1995.)--would be lost for a valence-trapped system.) To maintain a delocalized copper center, the ligand conformation around the binuclear site must be sufficiently rigid to preclude even a slight distortion (e.g., Glu

218 (O) - Cu bond shortening) that would stabilize Cu(II).

Current models (Guckert *et al.*, 1995) suggest that the high reduction potentials of blue-copper proteins result from the unusually long Met(S)-Cu bond imposed by the protein binding site; that the potentials are not even higher is a consequence of very strong Cys (s) - to - Cu(II) π bonding. As π-bonding is precluded from the bridging cysteine sulfur atoms of Cu_A, a localized Cu(II) ion would be extremely unstable. Indeed, based upon the onset of anodic current from the Cu_A oxidation, we estimate that the first Cu(II) reduction would occur at a potential greater than 1V.

Electron transfer with Thermus Cu_A and $*Ru(bpy)_3^{2+}$. The luminescent excited state $Ru(bpy)_3^{2+}$ is quenched by the soluble *T. thermophilus* Cu_A domain. In the presence of Cu_A (20 - 200 μM), the $*Ru(bpy)_3^{2+}$ decay remains single-exponential at both low (5 mM Tris, pH 8.1) and high (5 mM Tris, pH 8.1, 1 M NaCl) ionic strength. At low Cu_A concentrations (20-200 μM), the $*Ru(bpy)_3^{2+}$ decay rates display Stern-Volmer behavior with second-order quenching rate constants of 2.9×10^9 M^{-1} s^{-1} at low ionic strength, and 1.3×10^9 M^{-1} s^{-1} at high ionic strength. Under low salt conditions, however, the quenching-rate constant reaches a maximum value (saturates). This behavior can be interpreted in terms of complex formation between the $Ru(bpy)_3^{2+}$ ion and the soluble Cu_A domain.

Luminescence-decay measurements alone cannot distinguish between electron- and energy-transfer quenching mechanisms. In order to ascertain whether electron transfer was responsible for $*Ru(bpy)_3^{2+}$ quenching by oxidized Cu_A site, we measured the formation of $Ru(bpy)_3^{3+}$ using transient absorption spectroscopy. The absorption changes at 400 nm, induced by photoinitiated electron transfer at low and high ionic strength conditions, are consistent with $Ru(bpy)_3^{3+}$ formation and a transient difference spectrum (350-500 nm) is in excellent agreement with a steady-state [$Ru(bpy)_3^{3+}$ - $Ru(bpy)_3^{2+}$] difference spectrum (Figure 3.4). At higher protein concentrations (~ 250 M Cu_A), the formation of reduced Cu_A could be detected at 530 nm and the signal amplitude was

consistent with that expected on the basis of the $Ru(bpy)_3^{3+}$ signal. The magnitude of this signal compared to that of $*Ru(bpy)_3^{2+}$ suggests that only about 20% of the bimolecular quenching leads to charge-separated products. From these data and assuming unit cage-escape efficiency, we have estimated a second-order ET rate constant of 2.2×10^8 M^{-1} s^{-1} at high ionic strength (5 mM Tris, pH 8.1, 1 M NaCl) (Figure 3.5). This can be compared with electron-transfer rates of 10^7-10^8 M^{-1} s^{-1} observed in stopped-flow experiments with oxidized cytochrome oxidase and reduced cytochrome *c* (see Malatesta *et al.*, 1995). Stopped-flow studies with a soluble Cu_A domain and its natural partner, cytochrome c_{550}, from *P. denitrificans* have given rates of 1.5×10^6 M^{-1} s^{-1} (Lappalainen *et al.*, 1995). This ~ten-fold reduction in rate constant has been attributed to differences in binding between the soluble domain and the intact oxidases.

Under saturating conditions ($[Cu_A]$ = 250 μM, 5 mM Tris, pH 8.1), we estimate an intracomplex $*Ru(bpy)_3^{2+} \rightarrow Cu_A$ ET rate of 1.3×10^5 s^{-1}. The contribution of electron transfer to the intracomplex quenching is approximately 5%. The apparently efficient energy-transfer pathway is consistent with the significant overlap between the luminescence spectrum of $Ru(bpy)_3^{2+}$ and the Cu_A absorption spectrum. The rate constant for $Cu_A \rightarrow Ru(bpy)_3^{3+}$ charge-recombination within the complex is 2.1×10^5 s^{-1}.

The high driving force for the $*Ru^{2+} \rightarrow Cu_A$ ET reaction ($-\Delta G^\circ$ = 1.1 eV), estimated from the $Ru(bpy)_3^{3+/2+*}$ (E° = — 850 mV; Brunschwig & Sutin, 1990) and Cu_A (E° = — 240 mV) reduction potentials, raises the possibility that reaction occurs in the inverted driving-force regime (Marcus & Sutin, 1985). The $*Ru^{2+} \rightarrow Cu_A$ ET rate constant is close to that found for electron transfer from $Fe^{2+} \rightarrow Cu_A$ in a complex of intact oxidase with ruthenium-modified cytochrome c (Pan *et al.*, 1993). The driving force for the $Fe^{2+} \rightarrow Cu_A$ reaction, however, is 1 eV lower than that for $*Ru^{2+} \rightarrow Cu_A$ ET. One explanation would be that $Ru(bpy)_3^{2+}$ binds at a site that is poorly coupled for ET to Cu_A. However, a similar rate was also found for electron transfer from the triplet state of Zn-cytochrome *c* in its electrostatic complex with the whole oxidase (Brzezinski *et al.*, 1995);

in this case, the driving force is very close to that for electron injection from $*Ru(bpy)_3^{2+}$. In the Zn-cytochrome c/ oxidase system, the reorganization energy was estimated to be 0.57 eV and, consequently, ET from the Zn-porphyrin triplet lies in the inverted region. Unless the reorganization energy is very much larger in our case, it is likely that $*Ru^{2+} \rightarrow Cu_A$ ET is in the inverted region as well.

Horse heart cytochrome c binding to the Thermus fragment. The most likely binding site for positively charged $Ru(bpy)_3^{2+}$ and $Ru(bpy)_3^{3+}$ is the region of proposed cytochrome *c* binding. Mutagenesis and binding studies with the cytochrome c_{550} and the *P. denitrificans* Cu_A domain have mapped out important binding interactions (Lappalainen *et al.*, 1995). The D206N mutation (*P. denitrificans* numbering) yielded a six-fold decrease in the injection rate. Additionally, mutations Q148S, E154Q, D221N and E246Q decrease this rate by 35-60%. The E246 residue corresponds to the E198 of bovine oxidase and is located between the two cysteine ligands of the Cu_A site. If the binding location of $Ru(bpy)_3^{2+}$ and/or $Ru(bpy)_3^{3+}$ is analogous to that of cytochrome *c*, then the distance to the Cu_A site has been estimated to be 16 Å (Brzezinski *et al.*, 1995).

In order to provide a foundation for electron-transfer experiments with cytochromes *c* and the Cu_A domain, the interaction between horse heart cytochrome *c* and the *T. thermophilus* Cu_A domain was explored with optical absorption spectrometry. We expect that, if an interaction occurs between the two proteins, small changes will be induced in the strongly absorbing heme spectrum, and these will appear in the [cytochrome *c* + Cu_A minus cytochrome *c*] difference spectrum. To test this hypothesis, a 50 μM solution of oxidized cytochrome *c* was placed in each of two quartz cuvettes (2-mm pathlength). A solution of Cu_A protein was added to one of the cuvettes and an equal volume of the relevant buffer was added to the other, and the spectra in Figure 3.6 was obtained by subtracting the contribution of Cu_A at each point in the titration. Addition of Cu_A protein caused perturbation in the Soret region of the cytochrome *c* with a positive-going peak at ~402 nm and a negative-going peak at ~415 nm. At this concentration of

cytochrome *c*, the reaction is essentially complete even with sub-stoichiometric amounts of Cu_A. At Cu_A/cytochrome *c* ratios greater than unity, additional spectral changes are observed, indicating some complexity in the system. In experiments carried out with a ~10 μM cytochrome (not shown), the interaction of ba_3-Cu_A protein with horse heart cytochrome *c* was found to have a dissociation constant of ~5 μM, and, as expected for a predominantly electrostatic interaction, high ionic strength attenuates the binding between these two proteins.

In conclusion, our results indicate that the *T. thermophilus* Cu_A domain can participate in ET reactions with small inorganic redox reagents. Electron injection from $*Ru(bpy)_3^{2+}$ in a $Ru(bpy)_3^{2+}/Cu_A$ proceeds with a rate constant comparable to that found for electron injection from Zn-cytochrome *c* at a similar driving force. It is possible that both of these high-driving-force photoinduced ET reactions are slowed by the inverted effect. We have also demonstrated that complex formation occurs between the Cu_A fragment and horse heart cytochrome *c*, thus laying the groundwork for future electron-transfer studies with ruthenium-modified cytochrome *c*.

REFERENCES

Antholine, W. E., Kastrau, D. H. W., Steffens, G. C. M., Buse, G., Zumft, W. G. & Kroneck, P. M. H. (1992) *Eur. J. Biochem.* 209, 875-888.

Bard, A. J. & Faulkner, L. R., *Electrochemical Methods*, John Wiley & Sons, New York, N. Y. (1980).

Bertagnoli, H. & Kaim, W. (1995) *Angew. Chem. Intl. Ed. Engl. 34*, 771-773.

Blackburn, N. J., Barr, M. E., Woodruff, W. H., van der Oost, J. & de Vries, S. (1994) *Biochemistry* 33, 10401-10407.

Brunschwig, B. S. & Sutin, N. (1970) *J. Am. Chem. Soc.* 100, 7568-7575.

Brzezinski, P., Sundahl, M., Ädelroth, P., Wilson, M. T., El-Agez, B., Wittung, P. & Malmström, B. G. (1995) *Biophys. Chem.* 54, 191-197.

Churg, A. K. & Warshel, A. (1986) *Biochemistry 25*, 1675-1681.

Farrar, J. A., Lappalainen, P., Zumft, W. G., Saraste, M. & Thomson, A. J. (1995) *Eur. J. Biochem.* 232, 294-303.

Fee, J. A., Kuila, D., Mather, M. W. & Yoshida, T., (1986) *Biochem. Biophys. Acta* 853, 153-185.

Fee, J. A., Sanders, D., Slutter, C. E., Doan, P. E., Aasa, R., Karpefors, M. & Vänngård, T. (1995) *Biochem. Biophys. Res. Commun.* 212, 77-83.

Gurbiel, R. Z., Fann, Y.-C., Surerus, K. K., Werst, M. M., Musser, S. M., Doan, P.E., Chan, S. I., Fee, J. A. & Hoffman, B. M., (1993) *J. Am. Chem. Soc.* 115, 10888-10894.

Hill, H. A. O., Page, D. J. & Walton, N. J. (1987) *J. Electroanal. Chem. 217*, 129-140.

Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. (1995) *Nature 376*, 660-669.

Keightley, J. A., Zimmermann, B. H., Mather, M. W., Springer, P., Pastuszyn, A., Lawrence, D. M. & Fee, J. A. (1995) *J. Biol. Chem.,270*, 20345-20358.

Kelley, M., Lappalainen, P., Talbo, G., van der Oost, J. & Saraste, M. (1993) *J. Biol.*

Chem. 268, 16781-16787.

Langen, R., Chang, I.-J., Germanas, J. P., Richards, J. H., Winkler, J. H. & Gray, H. B. (1995) *Science 268*, 1733-1735.

Larsson, S., Källenbrin, B., Wittung, P. & Malmström, B. G. (1995) *Proc. Natl. Acad. Sci. USA* 92, 7167-7171.

Lappalainen, P., Aasa, R., Malmström, B. G. & Saraste, M. (1993) *J. Biol. Chem.* 268 26416-26421.

Lappalainen, P. & Saraste, M. (1994) *Biochim. Biophys. Acta* 1187, 222-226.

Lappalainen, P., Watmough, N. J., Greenwood, C. & Saraste, M. (1995) *Biochemistry* 34, 5924-5930.

Latimer, W. M. *Oxidation Potentials*, 2nd ed., Plenum Press, New York, N. Y. (1973).

Lübben, M., Castresana, J. & Warne, A. (1994) *Syst. Appl. Microbiol. 16*, 556-559.

Lui, S. M. & Cowan J. A. (1994) *J. Am. Chem. Soc. 116*, 11538-11549.

Malatesta, F., Antonini, G., Sarti, P. & Brunori, M. (1995) *Biophys. Chem.* 54, 1-33.

Malmström, B. G. (1994) *Eur. J. Biochem. 223*, 711-720.

Malmström, B. G. (1990) *Chem. Rev.* 90, 1247-1260.

Marcus, R. A. & Sutin, N. (1985) *Biochim. Biophys. Acta* 811, 265-322.

Millett, F., de Jong, C., Poulsen, L. & Capaldi, R. A. (1983) *Biochemistry* 22 546-552.

Nilsson, T. (1992) *Proc. Natl. Acad. Sci.* USA 89 6497-6501.

Pan, L. P., Hibdon, S., Liu, R.-Q., Durman, B. & Millett, F.(1993) *Biochemistry* 32, 8492-8498.

Sailasuta, N., Anson, F. C. & Gray, H. B. (1979) *J. Am. Chem. Soc. 101*, 455-462.

Saraste, M. (1990) *Q. Rev. Biophys.* 23, 331-366.

Slutter, C. E., Sanders, D., Wittung, P., Malmström, B. G., Aasa, R., Richards, J. H., Gray, H. B. & Fee, J. A.(1995) *Biochemistry*, submitted.

Steffens, G. C. M., Biewald, R. & Buse, G. (1987) *Eur. J. Biochem. 164*, 295-300.

Steffens, G. J. & Buse, G. (1979) *Hoppe-Seyler's Z. Physiol. Chem. 360*, 613-619.

Taniguchi, V. T., Sailasuta-Scott, N., Anson, F. C., & Gray, H. B. (1980) *Pure Appl. Chem. 52*, 2275-2281.

Tsukihara, T., Aoyama, H., Yamashita, E., Tomizaki, T., Yamaguchi, H., Shinzawa-Itoh, K., Nakashima, R., Yaono, R. & Yoshikawa, S. (1995) *Science 269*, 1069.

Trumpower, B. L. & Gennis, R. B. (1994) *Annu. Rev. Biochem.* 63, 675-716.

von Wachenfeldt, C., de Vries, S. & van der Oost, J. (1994) *FEBS Lett.* 340, 109-113.

Wilmanns, M., Lappalainen, P., Kelly, M., Sauer-Eriksson, E. and Saraste, M. (1995) *Proc. Natl. Acad. Sci. USA 92*, 11955-11959.

Yee, E. L., Cave, R. J., Guyer, K. L., Tyma, P. D. & Weaver, M. J. (1979) *J. Am. Chem. Soc. 101*, 1131-1135.

Figure 3.1

The cyclic voltammogram (CV) of *Thermus* Cu_A in 100 mM Tris buffer (pH 8) and 200 mM KCl. At 25 °C, ΔE_p = 60 mV,ν = 10 mV/sec with E° = 240 mV *vs*. NHE.

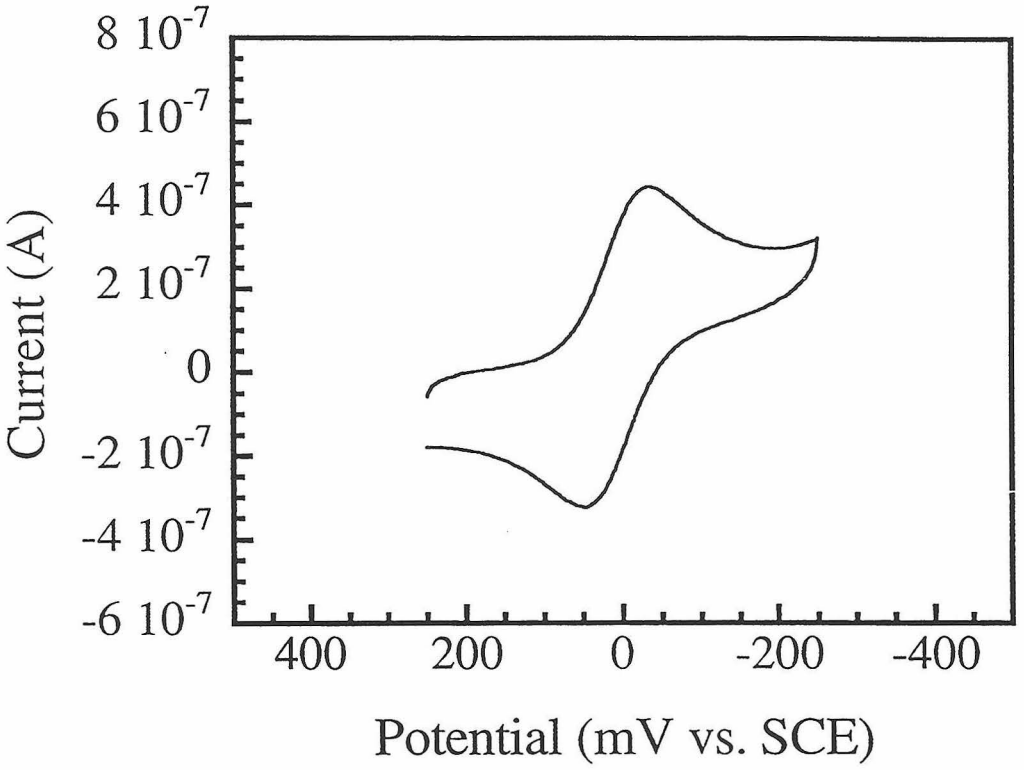

8 10-7
6 10-7
4 10-7
2 10-7
0
-2 10-7
-4 10-7
-6 10-7
Current (A)
400
200
0
-200
-400
Potential (mV vs. SCE)

Figure 3.2

Plot of i_{pc} (peak cathodic current) and i_{ac} (peak anodic current) *vs*. scan rate$^{1/2}$. A staight line is indicative of a diffusion-controlled process.

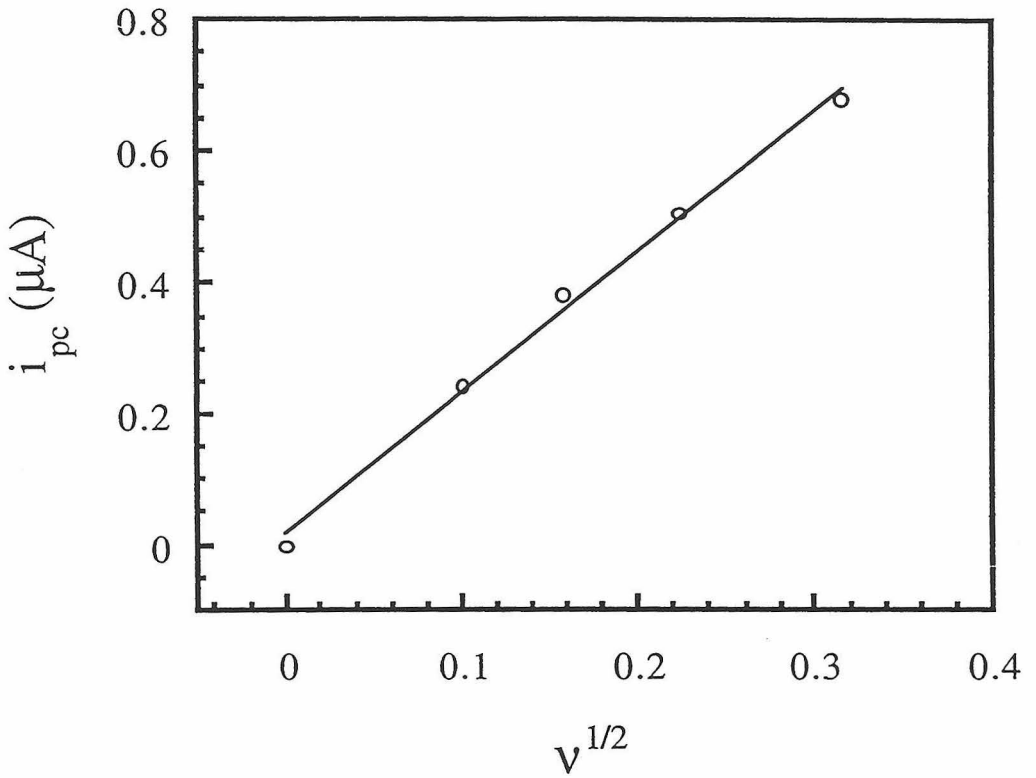

0.8
0.6
0.4
0.2
0
i_{pc} (μA)
0
0.1
0.2
0.3
0.4
$\nu^{1/2}$

Figure 3.3

Temperature dependence of the redox potential using the same conditions for data collection as in Figure 3.1.

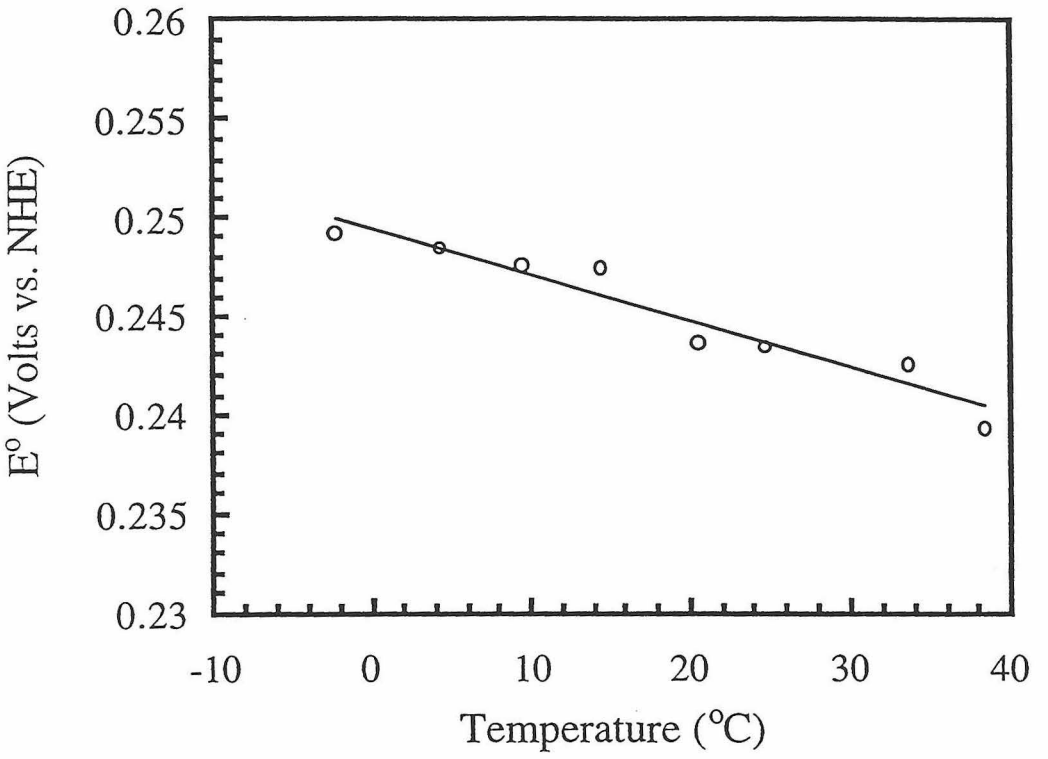

0.26
0.255
0.25
0.245
0.24
0.235
0.23
E° (Volts vs. NHE)
-10
0
10
20
30
40
Temperature (°C)

Figure 3.4

Transient difference spectrum (solid circles) recorded after 480-nm laser excitation of $Ru(bpy)_3^{2+}$ (40 μM) in the presence of the *T. thermophilus* Cu_A (80 μM) domain at low ionic strength (5 mM Tris, pH 8.1). Solid line is the predicted difference spectrum for formation of reduced Cu_A and $Ru(bpy)_3^{3+}$.

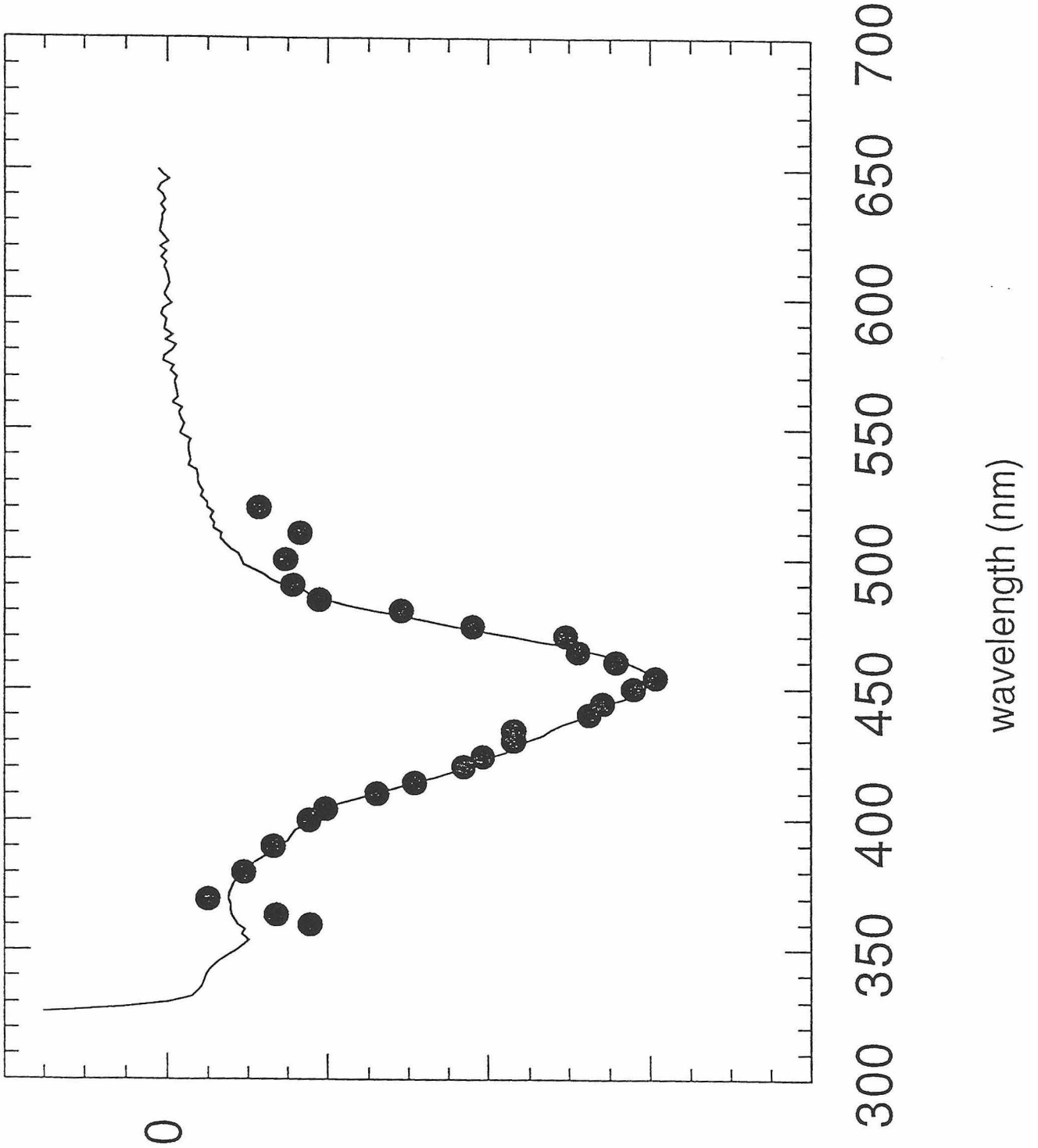

0
300
350
400
450
500
550
600
650
700
wavelength (nm)

Figure 3.5

Dependence of the $^{*}Ru(bpy)_3^{2+}$ ET quenching rate constant on *T. thermophilus* Cu_A domain concentration at high ionic strength (5 mM Tris, pH 8.1, 1 M NaCl).

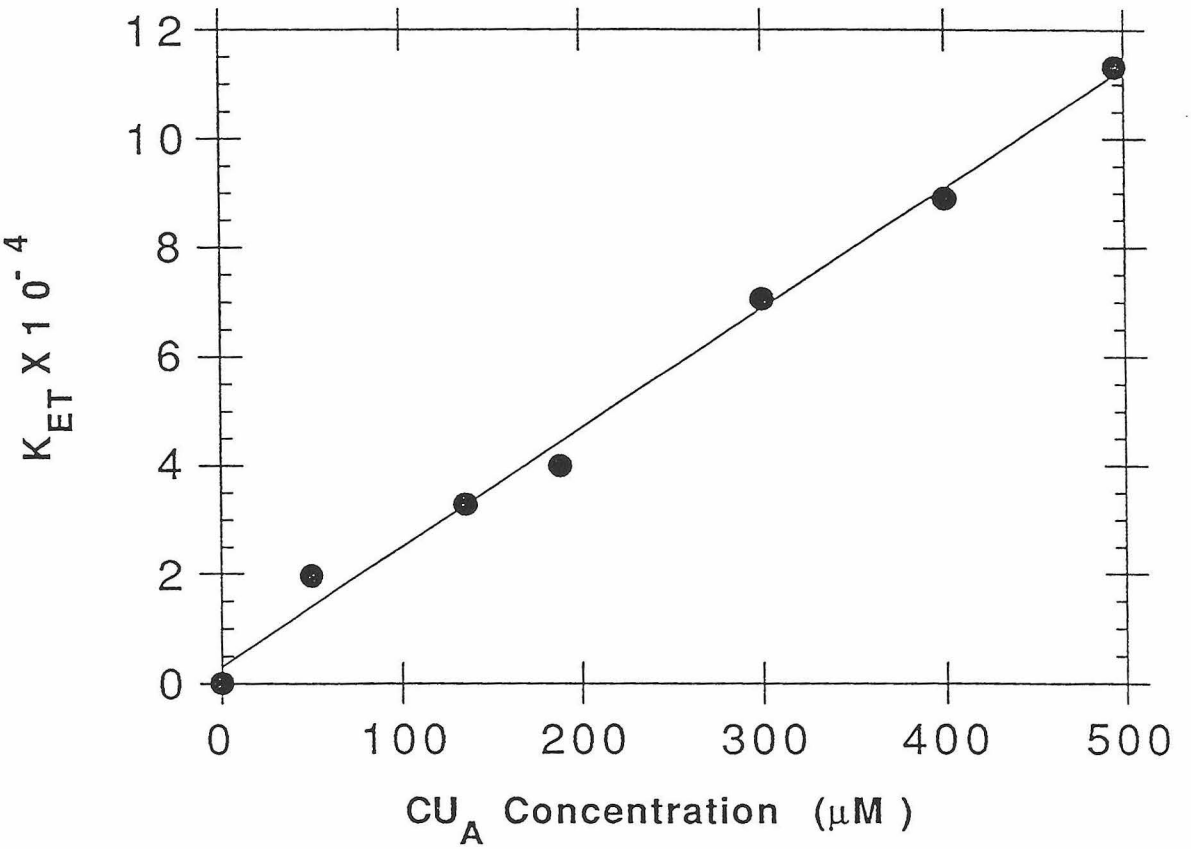
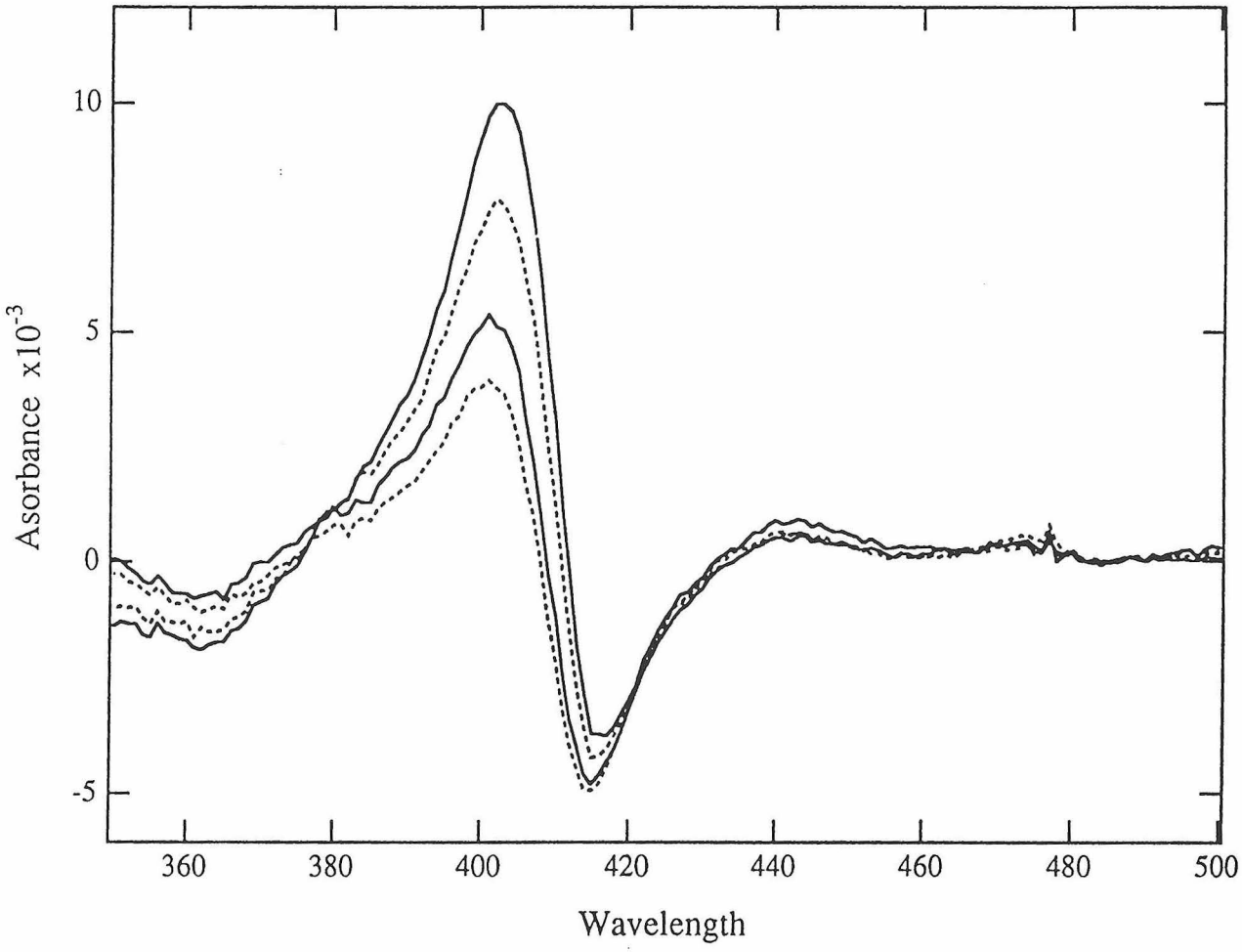

12
10
8
6
4
2
0
K_{ET} X 10^{-4}
0
100
200
300
400
500
CU_A Concentration (μM)

Figure 3.6

Effect of the Cu_A domain on the optical absorption spectrum of cytochrome *c*. The titration was carried out in 2 mm quartz cuvettes, each initially containing 0.50 ml of 50 μM oxidized cytochrome *c* in 5 mM Tris/0.1 mM EDTA/50 mM $K_3Fe(CN)_6$, pH 8.0. Aliquots (2.7 μL) of a concentrated Cu_A solution (933 μM in 50 mM ammonium sucinate pH 4.6 buffer) were added to the sample cuvette and equal volumes of ammonium succinate buffer to the reference cuvette. Spectra were recorded in the 350-850 nm range and stored in digital form. The difference spectra shown correspond to [(cytochrome *c* + Cu_A) - Cu_A] and were obtained by subtracting a spectrum of Cu_A, recorded in the absence of cytochrome *c*, normalized to the absorbance of the Cu_A plus cytochrome *c* solution at 482 nm. Spectra are shown at Cu_A/cytochrome *c* ratios of 0.4, 0.5, 0.67 and 0.77.


10
5
0
-5
Asorbance x10^{-3}
360
380
400
420
440
460
480
500
Wavelength

Appendix

pETCu$_A$: Full Restriction Map and Sequence Data

Figure A.1

Unique restriction sites in $pETCu_A$.

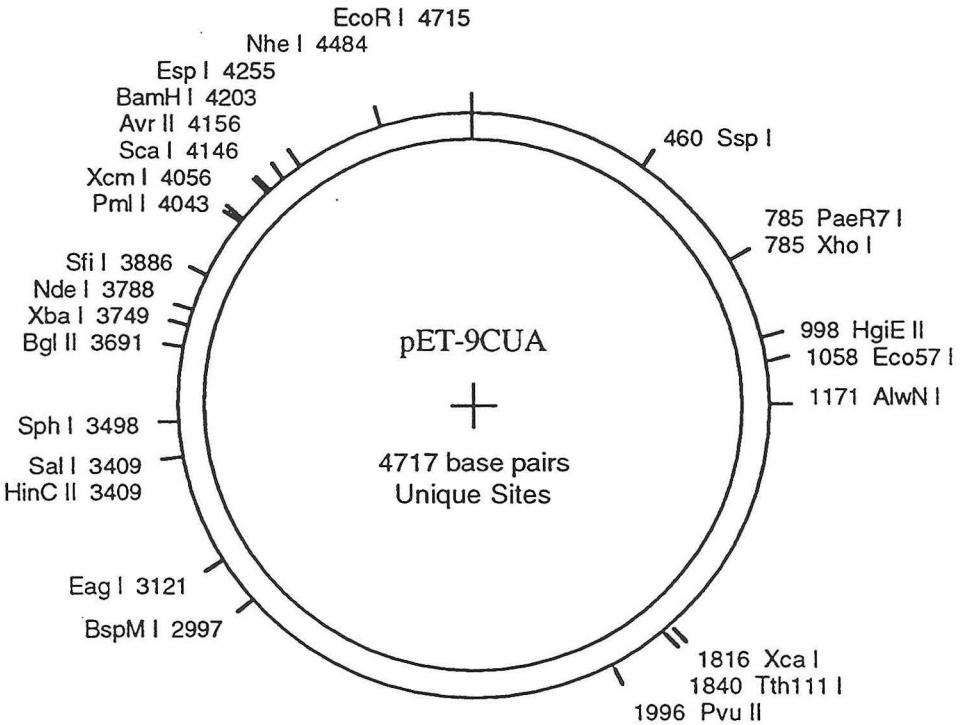

EcoR I 4715
Nhe I 4484
Esp I 4255
BamH I 4203
Avr II 4156
Sca I 4146
Xcm I 4056
Pml I 4043
Sfi I 3886
Nde I 3788
Xba I 3749
Bgl II 3691
Sph I 3498
Sal I 3409
HinC II 3409
Eag I 3121
BspM I 2997
460 Ssp I
785 PaeR7 I
785 Xho I
998 HgiE II
1058 Eco57 I
1171 AlwN I
1816 Xca I
1840 Tth111 I
1996 Pvu II
pET-9CUA
4717 base pairs
Unique Sites

Figure A.2

Amino acid translation of the gene encoding the *Thermus* Cu_A fragment.

pET-9CUA [3791 to 4201] -> 1-phase Translation

```
DNA  sequence    4716 b.p.   TTCTTAGAAAAA ... CAAACATGAGAA   circular


3791/1                                  3821/11
atg gcc tac acc ctg gcc acc cac acc gcc ggg gtc att ccc gcc gga aag ctt gag cgc
Met ala tyr thr leu ala thr his thr ala gly val ile pro ala gly lys leu glu arg
3851/21                                 3881/31
gtg gac ccc acc acg gta agg cag gaa ggc ccc tgg gcc gac ccg gcc caa gcg gtg gtg
val asp pro thr thr val arg gln glu gly pro trp ala asp pro ala gln ala val val
3911/41                                 3941/51
cag acc ggc ccc aac cag tac acg gtc tac gtc ctg gcc ttc gcc ttc ggc tac cag ccg
gln thr gly pro asn gln tyr thr val tyr val leu ala phe ala phe gly tyr gln pro
3971/61                                 4001/71
aac ccc att gag gtg ccc caa ggg gcg gag atc gtc ttc aag atc acg agc ccg gac gtg
asn pro ile glu val pro gln gly ala glu ile val phe lys ile thr ser pro asp val
4031/81                                 4061/91
atc cac ggc ttt cac gtg gag ggc acc aac atc aac gtg gag gtg ctc ccg ggc gag gtc
ile his gly phe his val glu gly thr asn ile asn val glu val leu pro gly glu val
4091/101                                4121/111
tcc acc gtg cgc tac acc ttc aaa agg ccc ggg gag tac cgc atc atc tgc aac cag tac
ser thr val arg tyr thr phe lys arg pro gly glu tyr arg ile ile cys asn gln tyr
4151/121                                4181/131
tgc ggc cta ggc cac cag aac atg ttc ggc acg atc gtg gtg aag gag tga
cys gly leu gly his gln asn met phe gly thr ile val val lys glu OPA
```

Figure A.3

Full restriction map and sequence of $pETCu_A$.

Positions of Restriction Endonucleases sites (unique sites underlined)

```
                    SfaN I
  Dde I          Taq I
  |              |   |
TTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAA  80
AAGAATCTTTTTGAGTAGCTCGTAGTTTACTTTGACGTTAAATAAGTATAGTCCTAATAGTTATGGTATAAAAACTTTTT
  |      .       |  .|         .         .         .         .         .         .
  3              17
                     21
                                                        ScrF I
                                                        EcoR II
                                                        BstN I
                                                     Sau3A I
                                                     Mbo I                    Ple I
                                 Mnl I               Dpn I                    Hinf I
                              Sec I          Fok I   Alw I                 Mme I
                           Hph I         PflM I      BstY I              Hinf I
                           |  | |         |    |      ||  |              |  |  |
GCCGTTTCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCG  160
CGGCAAAGACATTACTTCCTCTTTTGAGTGGCTCCGTCAAGGTATCCTACCGTTCTAGGACCATAGCCAGACGCTAAGGC
         .         .       |  | |       .|    |   .   ||  | .         .   |  |  |
                           107           121         134               154
                              110             126      135                  157
                                112                    135                      160
                                                       135                      160
                                                       135
                                                          138
                                                          138
                                                          138
                           Mse I                                   Nla III
     Mme I               Ase I     Mnl I                        Hph I    Mae III
     |                   ||        |                            |   |    |
ACTCGTCCAACATCAATACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGAC  240
TGAGCAGGTTGTAGTTATGTTGGATAATTAAAGGGGAGCAGTTTTTATTCCAATAGTTCACTCTTTAGTGGTACTCACTG
     |   .         .    ||   .    |    .         .         .      |  .|    |   .
     166               185       195                              227      236
                        186                                           231
          Hph I
        Msp I                 Nsi I
        Hpa II             Alu I                          Hae III
    Hinf I               HinD III                       Hae I
    |   | |              ||    |                        ||
GACTGAATCCGGTGAGAATGGCAAAAGCTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCAT  320
CTGACTTAGGCCACTCTTACCGTTTTCGAATACGTAAAGAAAGGTCTGAACAAGTTGTCCGGTCGGTAATGCGAGCAGTA
    |   |.|        .    ||   .|        .         .       ||.         .         .
    245               265                               298
        249            266                                299
        249                 271
          251
                                                                Sau3A I
                                                                Mbo I
                                                   Dde I        Dpn I
                                              HinP I            Pvu I
           SfaN I                             Hha I     BsmA I     BstU I     Mse I
           |                                  |   |     |          | ||       |
CAAAATCACTCGCATCAACCAAACCGTTATTCATTCGTGATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAA  400
GTTTTAGTGAGCGTAGTTGGTTTGGCAATAAGTAAGCACTAACGCGGACTCGCTCTGCTTTATGCGCTAGCGACAATTTT
         . |       .         .         .  |   |  .  |      .   | ||  .    |    .
           332                            363       373        384        395
                                          363                  386
                                              367               387
                                                                387
                                                                387
```

```
                                HinP I
                                Hha I
                   Taq I     Msp I                  SfaN I                Hinf I
                Hinf I       Hpa II               HinP I               EcoN I
         Tth111 II   Bsm I  Cfr10 I               Hha I     Ssp I  Hph I
         |      |  | |      ||  |                | |       |      |  |  |
GGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGCCAGCGCATCAACAATATTTTCACCTGAATCAGG  480
CCTGTTAATGTTTGTCCTTAGCTTACGTTGGCCGCGTCCTTGTGACGGTCGCGTAGTTGTTATAAAAGTGGACTTAGTCC
         |      |  | |      ||  |      •          | |       |      |  |  |      •
         410         422    429                 450       460    467
                417          430                 450                 470
                   420       430                   452                   473
                                433
                                433
                                    Sau3A I
                                    Mbo I
                                    Dpn I
                                   Alw I
                               Sec I
                               ScrF I
                               Nci I
                               Msp I
                               Hpa II
                               Bcn I
                              Xma I
                              Sma I
                              Sec I
                   Bsm I      ScrF I                     SfaN I
              ScrF I          Nci I                    Nsi I
              EcoR II         Bcn I               Mae III
    Mbo II    BstN I          Ava I          Hph I     Nla III        Rsa I
    |         |   |           ||   ||        |    |     || |           |
ATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATGCATCATCAGGAGTACGGATAA  560
TATAAGAAGATTATGGACCTTACGACAAAAGGGCCCCTAGCGTCACCACTCATTGGTACGTAGTAGTCCTCATGCCTATT
    |    •    |   |•          •||   ||  •      |   •|    || |•          •|        •
    485       495             511            526       536             551
              495             511                 531
              495             511                       537
                  499         511                         539
                              511
                              511
                              511
                               512
                               512
                               512
                               512
                               512
                               512
                                   516
                                    517
                                    517
                                    517
              Mbo II
           Mme I  Mnl I
  Tth111 II    Ear I                Bsr I                    Mae III
  |        |   |  |                 |                        |
AATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATCATTGGCAACG  640
TTACGAACTACCAGCCTTCTCCGTATTTAAGGCAGTCGGTCAAATCAGACTGGTAGAGTAGACATTGTAGTAACCGTTGC
  |      • |   |  |•          •       | •          •          •  |      •         •
  563         576                     598                       623
           572    579
              576
                                 SfaN I
                                HinP I                 Taq I
          Nla III               Hha I                Cla I
          |                     | |                  ||
CTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCC  720
GATGGAAACGGTACAAAGTCTTTGTTGAGACCGCGTAGCCCGAAGGGTATGTTAGCTATCTAACAGCGTGGACTAACGGG
         •|        •          • | |     •          •  ||     •          •          •
          651                  672                  693
                               672                   694
                                 674
```

```
                                                                  Taq I
                                                                  Xho I
                                                                  PaeR7 I
                                                                  Ava I
          Bsp1286 I                                              Mnl I
          Ban II                          Nla III          Fnu4H I
       BstU I                          Fok I                BstU I
      Nru I                           SfaN I  Mme I   Mse I   Hae III      Mae II
      ||  |                           ||   |  |        |    || | |||        |
GACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTCGAGCAAGACGTTT  800
CTGTAATAGCGCTCGGGTAAATATGGGTATATTTAGTCGTAGGTACAACCTTAAATTAGCGCCGGAGCTCGTTCTGCAAA
      ||• |       •         •        ||•  |  |   •   |    || | |||   •    |    •
      728                            758     766     774     782          795
       729                            759                 779
          732                             763              780
          732                                                 784
                                                               785
                                                               785
                                                               785
                                                                786
                                                                Nla III
                                                               BspH I
                                                               ||
CCCGTTGAATATGGCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCC  880
GGGCAACTTATACCGAGTATTGTGGGGAACATAATGACAAATACATTCGTCTGTCAAAATAACAAGTACTGGTTTTAGGG
         •         •         •         •         •         •    ||   •         •
                                                               865
                                                                866
                                                     Sau3A I
                                                     Mbo I      Sau3A I
                                                     Dpn I      Mbo I
                                            Sau3A I   Mbo II   Dpn I          BstU I
   Mae II                Hga I             Mbo I  BstY I      Alw I         HinP I
Mse I                 Dde I                Dpn I  Alw I      BstY I         Hha I
|  |                  |   |                |      || |       ||             ||
TTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGC  960
AATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGGCATCTTTTCTAGTTTCCTAGAAGAACTCTAGGAAAAAAAGACG
|  |     •         • |   |   •         •     |   •  || |   •   ||    •        ||
881                  902                      926    933        944            959
   884                   906                  926    933         945           959
                                              926       936      945            960
                                                     934        945
                                                     934        945
                                                     934
                                                          Sau3A I
                                                          Mbo I
                                                          Dpn I
                                                         Alw I
          Tth111 II                                     Msp I
         Fnu4H I                        NspB II        Hpa II
         Bbv I   Tth111 II              HgiE II      Tth111 II       Alu I
         | |     |                     | |           |     | ||      |
GCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTT  1040
CGCATTAGACGACGAACGTTTGTTTTTTTGGTGGCGATGGTCGCCACCAAACAAACGGCCTAGTTCTCGATGGTTGAGAA
         | |     | •         •        | |          •|     | ||      |  •         •
         970     978                  998          1011            1027
         970                            1000              1017
           972                                           1017
                                                           1019
                                                            1020
                                                            1020
                                                            1020
                  Eco57 I
             Bsr I          HinP I                                   Hae III
         Mae III            Hha I                    Mae I           Hae I
         |  |    |          |                        |               ||
TTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTT  1120
AAAGGCTTCCATTGACCGAAGTCGTCTCGCGTCTATGGTTTATGACAGGAAGATCACATCGGCATCAATCCGGTGGTGAA
         |  |    | •        | •          •         • |       •        ||         •
         1050                1068                    1092            1109
            1053             1068                                     1110
                 1058
```

```
                                                   Fnu4H I
                                                AlwN I  Fnu4H I
                                               Bsr I    Bbv I
                          Mnl I             Mae III   Bbv I  Bsr I
                          |                 |   ||    |  |   |
CAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGT  1200
GTTCTTGAGACATCGTGGCGGATGTATGGAGCGAGACGATTAGGACAATGGTCACCGACGACGGTCACCGCTATTCAGCA
         •         •      | •          •     |   ||    |  |•  |      •         •
                          1148              1166      1176   1183
                                               1170     1179
                                                1171    1179
                                                     1176
         Mme I
     ScrF I                                 NspB II
     Nci I                                  Fnu4H I
     Msp I                     Msp I        Bbv I                        HgiA I
     Hpa II  Ple I             Hpa II    HinP I                          Bsp1286 I
     Bcn I   Hinf I        Mae III       Hha I                           ApaL I
     |   |   |             |   |         | ||                            |
GTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAG  1280
CAGAATGGCCCAACCTGAGTTCTGCTATCAATGGCCTATTCCGCGTCGCCAGCCCGACTTGCCCCCCAAGCACGTGTGTC
     |  •|   |    •        |•  |       • | ||    •         •          •|         •
     1207    1215          1229         1242                           1271
     1207    1215              1233     1242                           1271
     1207                      1233       1244                         1271
     1207                                 1244
     1207                                  1245
         1211
                                                          HinP I
                                                          Hha I
   Alu I                      Dde I              Alu I      Hae II
   |                          |                  |          ||
CCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGG  1360
GGGTCGAACCTCGCTTGCTGGATGTGGCTTGACTCTATGGATGTCGCACTCGATACTCTTTCGCGGTGCGAAGGGCTTCC
   |     •         •         •|        •         |         •||        •         •
   1284                       1311               1330       1341
                                                            1342
                                                            1342
                                                               Sec I
                                                               ScrF I
                                                        Mnl I  EcoR II      ScrF I
                     Msp I  Fnu4H I                 HinP I    BstN I       EcoR II
                     Hpa II         Mme I           Hha I      Alu I       BstN I
                     |      |       |               |     |    |    |       |
GAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCT  1440
CTCTTTCCGCCTGTCCATAGGCCATTCGCCGTCCCAGCCTTGTCCTCTCGCGTGCTCCCTCGAAGGTCCCCCTTTGCGGA
         •           |      |  •    |     •         |•     |    |    |    •       | •
                     1380           1395            1409       1420               1438
                     1380   1387                    1409            1425          1438
                                                        1415        1425          1438
                                                                    1425
                                                                    1425
                                       Drd I     Taq I
                                   Mnl I       Hga I         SfaN I            Nla IV
                                   |    |      |  |          |                 |
GGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGC  1520
CCATAGAAATATCAGGACAGCCCAAAGCGGTGGAGACTGAACTCGCAGCTAAAAACACTACGAGCAGTCCCCCCGCCTCG
         •         •         • |    |  •   |  |  •        | •          •     |   •
                                1472       1484          1498              1516
                                     1477      1487
                                          Hae III
                                         Hae I
                                       ScrF I                    Nla III
                          Hae III      EcoR II                   NspH I
                       Fnu4H I         BstN I        Hae III     Nsp7524 I
                      BstU I        Nla IV           Hae I       Afl III
                      || |          |   | ||         ||          ||
CTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGC  1600
GATACCTTTTTGCGGTCGTTGCGCCGGAAAAATGCCAAGGACCGGAAAACGACCGGAAAACGAGTGTACAAGAAAGGACG
         •         •|| |     •    |   |•||       • ||      •    ||   •         •
                    1541          1555             1572        1585
                     1542             1559           1573       1585
                       1544           1559                      1585
                                      1559                       1586
                                        1561
                                         1562
```

```
                                                                  Fnu4H I
                                                                  Bbv I
          Hinf I                                  Alu I          Fnu4H I                HinP I
                                                                                         Hha I
GTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGC  1680
CAATAGGGGACTAAGACACCTATTGGCATAATGGCGGAAACTCACTCGACTATGGCGAGCGGCGTCGGCTTGCTGGCTCG
          1611                                    1646           1660                   1679
                                                                    1663                 1679
                                                                    1663

                                HinP I
     Ple I                      Hha I
     Hinf I                     Hae II
Fnu4H I                    Mbo II
Bbv I              Mnl I   Ear I      SfaN I                 SfaN I
GCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACAC  1760
CGTCGCTCAGTCACTCGCTCCTTCGCCTTCTCGCGGACTACGCCATAAAAGAGGAATGCGTAGACACGCCATAAAGTGTG
1681               1698    1707       1718                   1739
1681                       1707
     1686                      1711
     1686                       1712
                                1712
                Dde I                                                                Tth111 I
          HgiA I                                          Xca I                      Mae III
          Bsp1286 I                     Fnu4H I       Bsr I                        Mae II
          ApaL I      Rsa I          SfaN I       Mse I   Acc I                     BsaA I
CGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTG  1840
GCGTATATACCACGTGAGAGTCATGTTAGACGAGACTACGGCGTATCAATTCGGTCATATGTGAGGCGATAGCGATGCAC
          1771        1782           1796        1808    1816                     1835
          1771                          1799          1813                          1836
          1771                                        1816                            1838
                 1778                                                                   1840
                                                                                Fok I
                                                                               SfaN I
              HinP I                          HinP I                          Msp I
              Hha I                           Hha I                           Hpa II
            Fnu4H I                          BstU I                          ScrF I
            Bbv I                              Hga I                         Nci I
Bsr I  Nla III                          NspB II        Drd I                 Bcn I
ACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCT  1920
TGACCCAGTACCGACGCGGGGCTGTGGGCGGTTGTGGGCGACTGCGCGGGACTGCCCGAACAGACGAGGGCCGTAGGCGA
1841   1848                             1877           1892                  1908
           1852                                1882                            1908
           1852                                  1884                          1908
              1855                                1885                          1909
              1855                                1885                          1909
                                                                                   1912
                                                                                    1913
                                Fnu4H I                                                 Fnu4H I
                                Bbv I                                                   Bbv I
                        ScrF I                                                         Pvu II
                        Nci I                                                       Mnl I
                        Msp I     Nla III                                        BstU I  Alu I
                        Hpa II   NspH I                                          HinP I  NspB II
            Mae III     Bcn I    Nsp7524 I                                       Hha I  Fnu4H I
          Alu I     BsmA I   Alu I         Mnl I  Hph I    Hph I    BstU I  Bbv I
TACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCT  2000
ATGTCTGTTCGACACTGGCAGAGGCCCTCGACGTACACAGTCTCCAAAAGTGGCAGTAGTGGCTTTGCGCGCTCCGTCGA
          1929      1939     1948          1962   1969     1978     1987    1995
              1933       1943     1952                                  1988   1995
                         1943     1952                                  1988    1996
                         1943      1953                                  1989    1997
                         1943                                                1992
                         1943                                                    1996
                               1949                                                1998
                               1949                                                1998
```

```
                                                    Hga I
                                  Hinf I            BstU I
      Alu I                Xmn I                  Fok I      Alu I          Gsu I
      |                    |    |                  |   ||     |              |
GCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCT  2080
CGCCATTTCGAGTAGTCGCACCAGCACTTCGCTAAGTGTCTACAGACGGACAAGTAGGCGCAGGTCGAGCAACTCAAAGA
      | •          •       |  • |        •          •   |   ||•    |     •         |•
      2008                 2027                        2054       2065          2079
                                2032                      2058
                                                           2059
                                          Mse I
                                       Nla III
                                    Hae III                        Mae III   Mnl I
        Mse I                      Sau96 I                    Tth111 II   SfaN I
        |                          ||  |   |                  |      |      |   |
CCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCC  2160
GGTCTTCGCAATTACAGACCGAAGACTATTTCGCCCGGTACAATTCCCGCCAAAAAAGGACAAACCAGTGACTACGGAGG
        |          •          •   ||  | • |        •          |      |  • |   |   •
        2090                       2114                       2140        2152
                                    2115                            2146      2156
                                       2118
                                          2122
                                                     SfaN I
                                                    Fok I
                     Nla III                      Mnl I                Mae III
                     |                            | ||                 |
GTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGA  2240
CACATTCCCCCTAAAGACAAGTACCCCCATTACTATGGCTACTTTGCTCTCTCCTACGAGTGCTATGCCCAATGACTACT
        •          •|        •          •          •| ||      •          |          •
                     2181                           2211                2230
                                                      2213
                                                       2214
        Msp I
        Hpa II
       ScrF I
       Nci I
       Bcn I                                                   Sau96 I
    Nla III                                              Fnu4H I
   NspH I   Mae III  Mae II                            SfaN I   Ava II
   Nsp7524 I   Bsr I        Mnl I      Bsr I       Fok I    Nla IV            Dde I
   ||   ||  |  |      |      |          |           ||  |    ||               |
TGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTC  2320
ACTTGTACGGGCCAATGACCTTGCAACACTCCCATTTGTTGACCGCCATACCTACGCCGCCCTGGTCTCTTTTTAGTGAG
   ||   ||  |  |   • |      |•         |          •||  |    ||        •       | •
   2244        2256         2269       2280        2291     2300              2318
   2244     2253     2262                            2292     2301
    2245                                                2295
        2249                                                 2301
        2249
        2249
         2250
         2250
                                                                          Msp I
                                                                          Hpa II
                                                                       Sau3A I
                                                                       Mbo I
             HinP I                                            Fok I    Dpn I
             Hha I                                            SfaN I    Alw I
            Hae II   Mse I                                 Fnu4H I    BstY I
            Eco47 III                                      Bbv I     SfaN I  BspM II
            ||       |                                     |  ||      |    || ||
AGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGG  2400
TCCCAGTTACGGTCGCGAAGCAATTATGTCTACATCCACAAGGTGTCCCATCGGTCGTCGTAGGACGCTACGTCTAGGCC
          •  ||     • |       •          •          •     |  ||      | •  || ||  •
            2333                                           2376       2388    2396
            2333     2342                                  2376            2393
             2334                                             2379         2394
             2334                                              2380        2394
                                                                           2394
                                                                           2394
                                                                              2397
                                                                              2397
```

```
                HinP I
                Hha I                                          Mbo II
               Hae II      BstU I                              Bbv II    Nla III   Dde I
AACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGC  2480
TTGTATTACCACGTCCCGCGACTGAAGGCGCAAAGGTCTGAAATGCTTTGTGCCTTTGGCTTCTGGTAAGTACAACAACG
          2416        2428                                   2460      2470      2480
           2417                                              2460
           2417
                Fnu4H I
              Fnu4H I
              Bbv I                                  Hinf I
        Mae II    Bbv I         Mae II   BstU I  Hph I              Bsr I
TCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGC  2560
AGTCCAGCGTCTGCAAAACGTCGTCGTCAGCGAAGTGCAAGCGAGCGCATAGCCACTAAGTAAGACGATTGGTCATTCCG
          2492      2502          2516     2525    2533              2551
           2499                                        2536
           2499
              2502
                                                          Sau96 I
                 Mnl I                                    Nla IV
              Sau96 I                                     Ava II
              PpuM I                                   ScrF I
              Nla IV                                   EcoR II
              Eco0109 I                   HinP I       BstN I
           ScrF I                         Hha I     Hae III
           Nci I                    Sau3A I        Msc I
           Msp I                    Mbo I          Hae I
           Hpa II                   Dpn I          Eae I          Fnu4H I
           Bcn I               HgiA I   Nla III  Sec I   PpuM I    Bbv I    SfaN I
        Mae I  Ava II          Bsp1286 I  Fsp I  Dsa I   Eco0109 I     Ava I
AACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAG  2640
TTGGGGCGGTCGGATCGGCCCAGGAGTTGCTGTCCTCGTGCTAGTACGCGTGGGCACCGGTCCTGGGTTGCGACGGGCTC
        2573   2580          2595       2606   2613    2621          2635
           2577                2595     2604     2613    2621      2631     2640
           2577                     2601           2616          2631
           2577                     2601           2616
           2577                     2601           2616
           2577                           2607      2617
             2579                         2607        2619
             2579                                     2619
             2579                                     2619
              2580                                       2622
                 2583                                    2622
                                                         2622
      BstU I      Gsu I                                        Bsm I
    Fnu4H I   Fnu4H I                          Sty I         HinP I
  HinP I      Bbv I             BstU I         Sec I         Hha I
  Hha I    Fnu4H I            Hga I            PflM I       Fsp I
ATGCGCCGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTCACAGTTCT  2720
TACGCGGCGCACGCCGACGACCTCTACCGCCTGCGCTACCTATACAAGACGGTTCCCAACCAAACGCGTAAGTGTCAAGA
  2643     2652              2671                2691         2704
  2643        2655              2673             2691          2705
    2645      2655                               2691          2705
      2647        2659                                           2707
                                                       Msp I
                                                      Nae I
                                                   Fnu4H I
                                                 Nla IV
                 PflM I           Hinf I         Ban I  Hpa II            Mnl I
              Nla IV            Hph I          Mnl I   Cfr10 I          Taq I
CCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGGTGCCGCCGGCTTCCATTCAGGTCGAGGT  2800
GGCGTTCTTAACTAACCGAGGTTAAGAACCTCACCACTTAGGCAATCGCTCCACGGCGGCCGAAGGTAAGTCCAGCTCCA
              2736              2754           2769    2777             2794
                 2740              2757          2771   2778               2796
                                                 2771
                                                    2774
                                                       2777
                                                        2778
```

```
                                                        HinP I
     Nla IV                                             Hha I
   Msp I                                               Nla IV
   Hpa II                                              Nar I
  ScrF I                                               Hae II
  Nci I                                                Bbe I
  Bcn I                                                Ban I
Sau96 I          BstU I          Mnl I                Aha II
Hae III   Nla III   Hga I  BstU I                    Fnu4H I          Nla III
| || |    |       |  |      |     |                 | ||             |
GGCCCGGCTCCATGCACCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGT  2880
CCGGGCCGAGGTACGTGGCGCTGCGTTGCGCCCCTCCGTCTGTTCCATATCCCGCCGCGGATGTTAGGTACGGTTGGGCA
| || |   •|      | •|      | •   |     •          •  | ||   •       | •          •
2801      2811      2821   2828                     2853           2868
2801             2818            2834                 2855
  2803                                                2855
  2803                                                2855
  2803                                                2855
   2804                                               2855
   2804                                               2855
     2806                                              2856
                                                       2856
                                     NspB II     Sau3A I
           Sec I                   Sau3A I        Mbo I
     HgiA I   Mnl I                Mbo I     Bsr I
     Bsp1286 I   Fnu4H I           Dpn I  Sau96 I   Taq I                  BstU I
  Nla III  Bgl I                Mae III     Ava II  Dpn I                  Fnu4H I
  |  |     || |  |              |    |  |   |  |    | |                  | |
TCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGCGGTCCAGTGATCGAAGTTAGGCTGGTAAGAGCCGCG  2960
AGGTACACGAGCGGCTCCGCCGTATTTAGCGGCACTGCTAGTCGCCAGGTCACTAGCTTCAATCCGACCATTCTCGGCGC
  |  |   • || |  | •          •  |    | •|   |  | •  | |    •          •    | |  •
  2883     2892                 2913        2925    2933                  2955
     2886        2898               2918   2925      2935                   2957
     2886     2895                  2918      2928
           2893                     2918           2933
                                       2921        2933
   Sau3A I                              ScrF I
   Mbo I                                EcoR II     Hae III        Fok I
   Dpn I                                BstN I     Hae I          SfaN I    Fnu4H I
   Alw I    Alu I                    BspM I        Nla III     BstU I      SfaN I
   |        |                        |    |        | ||         |    ||      |  |
AGCGATCCTTGAAGCTGTCCCTGATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCC  3040
TCGCTAGGAACTTCGACAGGGACTACCAGCAGTAGATGGACGGACCTGTCGTACCGGACGTTGCGCCCGTAGGGCTACGG
   |     •  |       •          •      |  • |        •| ||     •  |    ||•     |  | •
   2964     2973                     2997          3011        3023        3035
   2964                                   3002       3013            3028      3038
   2964                                   3002        3014           3029
   2964                                   3002
                                                BstU I                      Hga I
                                  Fok I      Hga I                         BstU I
 Msp I        Hinf I          Hae III      BstU I                         HinP I
 Hpa II    Mbo II             Hae I      Mnl I  Nru I             Mae II    Hha I
 |         |  |               ||  |      |  ||  ||                |        |||
GCCGGAAGCGAGAAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGT  3120
CGGCCTTCGCTCTTCTTAGTATTACCCCTTCCGGTAGGTCGGAGCGCAGCGCTTGCGGTCGTTCTGCATCGGGTCGCGCA
 |       • |  |    •          ||  |     •|  ||  ||•          •    |    •    |||  •
 3042      3052               3070       3081   3088              3105      3115
 3042          3055             3071         3084                            3115
                                   3074       3085                             3116
                                                3089                            3117
     Bgl I
  Fnu4H I  Msp I                                          Mae III
 Hae III   Hpa II                                      Bsr I
Gdi II    Nae I                                     Sau96 I
Eag I     Cfr10 I     Hae III                       Ava II
Eae I  Nla III       Hae I              Mae II      Nla IV                   Mnl I
|||  | |  ||          ||                 |          ||  |  |                  |
CGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGG  3200
GCCGGCGGTACGGCCGCTATTACCGGACGAAGAGCGGCTTTGCAAACCACCGCCCTGGTCACTGCTTCCGAACTCGCTCC
|||  | | •||        • ||      •          •|        •  ||  |  |          •       | •
3121   3128          3142                3161        3173                      3197
3121      3131        3143                            3174
3121      3131                                        3174
 3122      3132                                          3177
  3123     3132                                            3180
     3126
```

```
                                    Sau3A I      Gsu I
                                    Mbo I     HinP I              Mnl I
                                    Dpn I     Hha I            Sau96 I
        Hinf I                   Hae III     BstU I            Ava II
        |                        |   |       ||  |             |  |
GCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGAC  3280
CGCACGTTCTAAGGCTTATGGCGTTCGCTGTCCGGCTAGTAGCAGCGCGAGGTCGCTTTCGCCAGGAGCGGCTTTTACTG
        |·         ·         · |   |   ·    ||  |·         · |  |    ·         ·
        3209                   3232        3245             3262
                                    3236      3246            3262
                                    3236      3246               3265
                                    3236         3249
          Msp I
          Hpa II
         Nae I
         Cfr10 I
       Fnu4H I
       Bbv I
     HinP I
     Hha I   Nla IV
    Hae II   Ban I                        Mbo II                                BstU I
    Eco47 III                   Nla III  Bbv II          Fnu4H I       Nla III
    || |  || |                 |        |               |             |      |
CCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCC  3360
GGTCTCGCGACGGCCGTGGACAGGATGCTCAACGTACTATTTCTTCTGTCAGTATTCACGCCGCTGCTATCAGTACGGGG
    || | ·|| |     ·         ·   |     ·  |      ·        |·          ·  |      |
    3285                        3314     3323            3339          3353
    3285     3294               3323                                   3360
     3286    3294
     3286
       3288
       3288
          3291
          3291
           3292
           3292
                                                Taq I
                                                Sal I
       Msp I                                    HinC II                  EcoN I
HinP I              Bsr I                       Acc I                Ple I
Hha I  Hpa II  Alu I                       SfaN I   Hga I            Hinf I
|      |       |    |                      |      || |             |    |
GCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGACGCTCTCCCTTATGCGACTCCTGCATT  3440
CGCGGGTGGCCTTCCTCGACTGACCCAACTTCCGAGAGTTCCCGTAGCCAGCTGCGAGAGGGAATACGCTGAGGACGTAA
|      | ·     |   ·|        ·          ·  |     || |        ·        |·  |      ·
3361   3368    3376                       3403     3412             3429
3361                3381                       3409                3429
       3368                                    3409                    3433
                                               3409
                                                3410
                                                                     HinP I
                                                                     Hha I
                                                                    Nla IV
                                                                    Nar I
                                                          Nla III   Hae II
          Bsr I                  HgiA I    Fnu4H I      Sph I      Bbe I
      Fnu4H I              Hae III       Fnu4H I        NspH I     Ban I
      Bbv I              Mnl I    Bsp1286 I              Nsp7524 I  Aha II
      |    |             | |      |       |  |           ||         ||
AGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAA  3520
TCCTTCGTCGGGTCATCATCCAACTCCGGCAACTCGTGGCGGCGGCGTTCCTTACCACGTACGTTCCTCTACCGCGGGTT
      |   ·|         ·   | |   ·  |       |  |       ·        ||·          · ||      ·
      3446               3464     3473                       3498          3512
      3446                 3466          3480                3498          3512
           3451                    3473      3483            3498          3512
                                                              3499          3512
                                                                            3512
                                                                            3512
                                                                             3513
                                                                             3513
```

```
          Sec I
          Dsa I
        Hae III
       Gdi II
       Eae I
      Msp I                                        Nla III
      Hpa II                                      BspH I                    Sau3A I
     ScrF I     Hae III                         HinP I                       Mbo I
     Nci I     Sau96 I                          Hha I   Bsp1286 I            Dpn I
     Bcn I    Nla IV                           Hae II   Ban II        Bsp1286 I
    Sec I     EcoO109 I                        Eco47 III              Ban II  Mbo II
    ||||| |   |||                              ||   ||  |             |      | |
CAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTT  3600
GTCAGGGGGCCGGTGCCCCGGACGGTGGTATGGGTGCGGCTTTGTTCGCGAGTACTCGGGCTTCACCGCTCGGGCTAGAA
    ||||| |   ||| •         •         •      ||  •||  |    •         |•    | |  •
    3526      3536                           3566                     3589    3597
     3527     3536                           3566     3575            3589
     3527      3537                           3567    3575                    3595
     3527       3538                          3567                            3595
      3528                                        3571                         3595
      3528                                         3572
       3529
       3529
        3530
          3532
          3532
                                                   Hph I
                                                  Msp I
                                                  Hpa II
                                                 Cfr10 I
                          HinP I                HinP I        Hae III
                          Hha I                 Hha I        Gdi II
                         Nla IV                Nla IV        Eae I
                         Nar I                 Nar I        Msp I
                         Hae II                Hae II       Hpa II          Msp I
                         Bbe I                 Bbe I       Nae I            Hpa II
                         Ban I                 Ban I       Cfr10 I     Hga I
       Hph I             Aha II                Aha II   SfaN I      SfaN I
       |                 ||                    || || |  |  ||||     |  |   |
CCCCATCGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCG  3680
GGGGTAGCCACTACAGCCGCTATATCCGCGGTCGTTGGCGTGGACACCGCGGCCACTACGGCCGGTGCTACGCAGGCCGC
       | •         •     ||  •         •      || || |  |  ||||     | •|   |    •
       3608              3626                 3647     3656        3668
                         3626                 3647        3659        3671
                         3626                 3647        3659            3675
                         3626                 3647         3660           3675
                         3626                 3647         3660
                         3626                 3647          3661
                          3627                 3648         3661
                          3627                 3648          3662
                                                 3650
                                                  3651
                                                  3651
                                                    3653
                 Sau3A I
            Sau3A I
            Mbo I
           BstY I
       Taq I     Mbo I
     Sau3A I     Dpn I
     Mbo I       Alw I
     Dpn I     Taq I         Mse I                                  Mae I
    Alw I  Dpn I            Ase I   Ple I                          Xba I
   Mnl I   Bgl II       BstU I      Hinf I       BsmA I           Mnl I
   | || |  ||   | |     |     ||     |            |               | ||
 TAGAGGATCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGAAATAAT  3760
 ATCTCCTAGCTCTAGAGCTAGGGCGCTTTAATTATGCTGAGTGATATCCCTCTGGTGTTGCCAAAGGGAGATCTTTATTA
   | || | •||   | | •  |     ||     |  •          |          •      | ||          •
   3683    3691        3703         3717         3730             3747
     3685   3692             3709    3717                            3749
      3686      3696          3710                                    3750
      3686        3698
      3686        3698
        3688      3698
           3691
            3692
            3692
                 3698
```

```
                                                Hae III
                                               Msc I
                                               Hae I          Sec I
                                               Eae I          ScrF I
                                             ScrF I           Nci I                Alu I
                                   Hae III  EcoR II           Msp I               HinD III
          Mse I                   Hae I     BstN I           Hpa II         Msp I
     Mse I                  Nde I          Sec I             Bcn I          Hpa II
     |     |                |   ||        || ||              |              |    ||
TTTGTTTAACTTTAAGAAGGAGATATACATatggcctacaccctggccacccacaccgccggggtcattcccgccggaaa  3840
AAACAAATTGAAATTCTTCCTCTATATGTAtaccggatgtgggaccggtgggtgtggcggccccagtaagggcggcctttt
     |   • |        •        | • ||       •|| ||     •         |•          •    |    ||
     3766                    3788          3801                3819           3834
          3772                   3792       3802                3819           3834
                                  3793      3802                3819                3839
                                            3802                3819                 3840
                                              3804              3819
                                              3804              3819
                                              3804
                                               3805
                                                Sfi I
                                               Sau96 I
                                             Sec I
                                             ScrF I           PflM I
                                             EcoR II     Msp I
          Sau96 I                            BstN I      Hpa II                  Sau96 I
          Nla IV                            Sec I       ScrF I                   Nla IV
          Ava II                        Sau96 I         Nci I                    Hae III
      BstU I                            Nla IV  Hae III  Sau96 I              Msp I
     HinP I        Sec I                Hae III  Bgl I   Hae III              Hpa II
     Hha I         Dsa I               EcoO109 I       Bcn I                 Cfr10 I
     ||    |       |                   ||  ||  |||     || |  |               || |
gcttgagcgcgtggaccccaccacggtaaggcaggaaggcccctgggccgacccggcccaagcggtggtgcagaccggcc  3920
cgaactcgcgcacctggggtggtgccattccgtccttccggggacccggctgggccgggttcgccaccacgtctggccgg
     || •  |      •|        •       || •||  |||  • || |  | •          •   || |  •
     3847          3861              3877            3892                 3914
     3847          3861               3878     3887    3895                 3915
      3848                            3878    3886     3895                 3915
          3853                        3878          3892                      3917
          3853                           3881       3892                      3917
          3853                            3882       3893                      3917
                                          3882       3893
                                          3882            3898
                                          3882
                                             3885
                                              3886
                          Hae III
                         Hae I                                      Sty I
                        ScrF I                                  Bsp1286 I
                        EcoR II                                Nla IV            Sau3A I
       Rsa I      Mae II                                      Ban I             Mbo I
     Bsr I     Acc I   BstN I                                Mnl I   Sec I       Dpn I
     |  |      |   |   | ||                                  | ||    |           |
ccaaccagtacacggtctacgtcctggccttcgccttcggctaccagccgaaccccattgaggtgccccaaggggcggag  4000
ggttggtcatgtgccagatgcaggaccggaagcggaagccgatggtcggcttggggtaactccacggggttccccgcctc
     |  | •    |   |•  | ||   •         •         •         | ||    | •         |
     3925      3935    3943                                  3980    3988        4000
        3928        3939                                      3982              4000
                        3943                                  3982              4000
                        3943                                   3983
                          3945                                      3988
                           3946
```

```
                                                                                                ScrF I
                                                                                                Nci I
                                                                                                Msp I
                                                                                                Hpa II
                                                                                                Bcn I
                                                                                               Xma I
                                                                                               Sma I
                                Msp I      Sec I                                               Sec I
                                Hpa II     Dsa I                                               ScrF I
                           Bsp1286 I   Sau3A I                                               Nci I
                     Sau3A I  ScrF I   Mbo I              Mnl I  Xcm I                       Bcn I
                     Mbo I    Nci I    Dpn I         Mae II  Nla IV                 HgiA I
                Mbo II       Ban II  Mae II          Pml I    Ban I              Mnl I   Ava I
               Bbv II  Dpn I    Bcn I    Alw I        BsaA I  Bsp1286 I     Mae II  Bsp1286 I
               ||      |     |  ||   |   |  |         ||    | ||   |         |    |  |    ||
atcgtcttcaagatcacgagcccggacgtgatccacggctttcacgtggagggcaccaacatcaacgtggaggtgctccc  4080
tagcagaagttctagtgctcgggcctgcactaggtgccgaaagtgcacctcccgtggttgtagttgcacctccacgaggg
   ||    • |     | •||   |   |  |      •  ||    |•||   |   •    |    |  |    ||•
   4004    4012     4021     4030         4043    4051          4065    4073
    4005         4018    4026             4043     4052              4070    4078
            4012     4021     4030          4044    4052                 4073
            4012     4021     4030               4049   4056                 4078
                  4018        4030                                          4078
                      4022       4033                                       4078
                      4022       4033                                       4078
                                                                            4078
                                                                            4078
                                                                             4079
                                                                             4079
                                                                             4079
                                                                             4079
                                                                             4079
                                      Sec I
                                      ScrF I
                                      Nci I
                                      Msp I
                                      Hpa II
                                      Bcn I
                                     Xma I
                                     Sma I                                          Hae III
                                     Sec I                                         Hae I
                                     ScrF I                                      Mae I
                                     Nci I                                      Sty I
                                     Bcn I                                      Sec I
                                     Ava I                            Rsa I    Avr II
       BsmA I     HinP I           Sau96 I        SfaN I          Sca I   Hae III
    Mnl I         Hha I   EcoN I   Hae III   Rsa I              Bsr I   Fnu4H I
    |  |          |       |        | ||      |    |             | ||    | | || ||
gggcgaggtctccaccgtgcgctacaccttcaaaaggcccggggagtaccgcatcatctgcaaccagtactgcggcctag  4160
cccgctccagaggtggcacgcgatgtggaagttttccgggcccctcatggcgtagtagacgttggtcatgacgccggatc
    |  | •        |•      |  •     | ||•     |   •|         •   | ||  • | | || ||
    4085          4099    4107     4116      4126                4144    4152
       4088       4099             4116           4131            4146    4154
                                     4118                          4147     4156
                                     4118                                   4156
                                     4118                                   4156
                                     4118                                    4157
                                     4118                                      4159
                                     4118                                       4160
                                     4118
                                      4119
                                      4119
                                      4119
                                      4119
                                      4119
                                      4119
```

```
                                                Msp I
                                                Hpa II
                                              Sau3A I
                                              Mbo I
                                              Dpn I
                                              Alw I
          Nla III                            Nla IV
         NspH I        Sau3A I               BstY I
         Nsp7524 I     Mbo I                 BamH I
         Afl III       Dpn I                 Alw I  Fnu4H I                    Dde I
       Xmn I          Pvu I  Hph I          Mnl I    Bbv I                   Alu I
       | ||           ||     |              | ||  |  |                       | |
gccaccagaacatgttcggcacgatcgtggtgaaggagtgaGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAGTT  4240
cggtggtcttgtacaagccgtgctagcaccacttcctcactCCTAGGCCGACGATTGTTTCGGGCTTTCCTTCGACTCAA
       | ||          · ||     |·           | ||  |  |·          ·          · | |     ·
       4168           4182   4189          4200     4209                      4232
         4170          4183                 4202   4209                        4234
         4170          4183                 4202
         4170          4183                 4202
          4171                              4202
                                             4203
                                             4203
                                             4203
                                             4203
                                                4206
                                                4206
                                             Hae III
                                            Sau96 I
    Fnu4H I                                Nla IV
    Bbv I      Dde I                       Eco0109 I
 Fnu4H I      Esp I                    Sty I
 Bbv I      NspB II        Mae I       Sec I   Mnl I            Mnl I
 |  |       | ||           |           |   ||| |                |
GGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAG  4320
CCGACGACGGTGGCGACTCGTTATTGATCGTATTGGGGAACCCCGGAGATTTGCCCAGAACTCCCCAAAAAACGACTTTC
 |  |     · | ||     ·      |    ·       |  ·||| |     ·          ·|         ·          ·
 4242       4252           4266        4277    4285             4301
 4242         4254                     4277
    4245       4255                       4281
    4245                                  4281
                                           4282
                                            4283
                                                 BstU I
                EcoR V                        Sau3A I
              Msp I                           Mbo I
              Hpa II                          Dpn I
 Mnl I       BspM II                       Nla III       Taq I    Nla IV
 |           ||  |                         |  |  |       |        |
GAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGC  4400
CTCCTTGATATAGGCCTATAGGTGTCCTGCCCACACCAGCGGTACTAGCGCATCAGCTATCACCGAGGTTCATCGCTTCG
|         · ||  |   ·          ·          · |  |  | ·     |     ·  |       ·          ·
4321        4332                            4362          4375     4383
             4333                              4365
             4333                              4365
                4336                           4365
                                                  4368
                  Hae III
                 Gdi II
                 Eae I                                  HinP I
              Fnu4H I                   HgiA I          Hha I                        Hae II
        Bsr I   Fnu4H I      Mme I    Bsp1286 I       Fsp I          SfaN I         Eco47 III
        |     |  |||          |        |              ||             |              |
GAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATA  4480
CTCGTCCTGACCCGCCGCCGGTTTCGCCAGCCTGTCACGAGGCTCTTGCCCACGCGTATCTTTAACGTAGTTGCGTATAT
        | ·  |  ||| ·        | ·      |   ·          · ||       ·      |   ·          |
        4408    4416          4428     4436             4452           4466           4480
             4413                       4436              4453                         4480
                 4417                                      4453
                 4417
                  4418
```

```
                                                                  Msp I
                                                                  Hpa II
      Fnu4H I                                                    ScrF I
      Bbv I                                                      Nci I
   Mae I                                                         Bcn I      Msp I
  Nhe I                                                         Sau96 I     Hpa II
HinP I                Bsr I       Mme I                         Hae III    Cfr10 I
Hha I               Mae III  SfaN I            EcoR V       Mnl I       Rsa I
| ||  |             |  |      |     |           |            | | ||      | ||
GCGCTAGCAGCACGCCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATA  4560
CGCGATCGTCGTGCGGTATCACTGACCGCTACGACAGCCTTACCTGCTATAGGGCGTTCTCCGGGCCGTCATGGCCGTAT
| ||  | •         |  |      |•    |    •       |  •         |•| ||      | ||      •
4481                4500     4509               4527        4539       4550
4481                   4503        4515                      4541       4552
  4483                                                        4541        4553
   4484                                                         4543       4553
      4487                                                      4543
      4487                                                      4543
                                                                 4544
                                                                 4544

                      Sec I
                      ScrF I
                      EcoR II   Nla IV
                      BstN I    Ban I   Fok I
                   Fok I   Mae III    Mnl I          HinP I                  Nla IV
                  SfaN I  Hph I     Sec I           Hha I                   Ban I
                  ||  |   ||    |   | | |           |                       |
ACCAAGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGC  4640
TGGTTCGGATACGGATGTCGTAGGTCCCACTGCCACGGCTCCTACTGCTACTCGCGTAACAATCTAAAGTATGTGCCACG
         •        ||  |   || •  |   | |•|         •  |      •          •     |   •
                  4579    4587      4597            4613                    4636
                   4580    4588       4599           4613                    4636
                      4583      4593    4601
                      4583      4593
                      4583
                      4583

                                                Taq I
                                            Alu I                          EcoR I
                                           HinD III                Tth111 II
                  Mse I                  Mse I    Cla I     Alu I     Nla III
                  |                      |  ||    ||        |       |   |    |
CTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA  4716
GACTGACGCAATCGTTAAATTGACACTATTTGATGGCGTAATTTCGAATAGCTACTATTCGACAGTTTGTACTCTT
         •        | •          •          |  ||    ||        |•      |   |•   |
                  4658                    4680     4689      4699      4709
                                             4683                    4705
                                              4684                          4714
                                                   4690
```

Restriction Endonucleases site usage

Aat II	-	BstN I	11	HinC II	1	Ple I	5
Acc I	3	BstU I	25	HinD III	3	Pml I	1
Afl II	-	BstX I	-	Hinf I	15	PpuM I	2
Afl III	2	BstY I	6	HinP I	33	Pst I	-
Aha II	4	Bsu36 I	-	Hpa I	-	Pvu I	2
Alu I	17	Cfr10 I	8	Hpa II	33	Pvu II	1
Alw I	12	Cla I	2	Hph I	13	Rsa I	6
AlwN I	1	Dde I	9	Kpn I	-	Rsr II	-
Apa I	-	Dpn I	23	Mae I	6	Sac I	-
ApaL I	2	Dra I	-	Mae II	12	Sac II	-
Ase I	2	Dra III	-	Mae III	15	Sal I	1
Asp718	-	Drd I	2	Mbo I	23	Sau3A I	23
Ava I	5	Dsa I	4	Mbo II	9	Sau96 I	17
Ava II	7	Eae I	6	Mlu I	-	Sca I	1
Avr II	1	Eag I	1	Mme I	8	ScrF I	28
BamH I	1	Ear I	2	Mnl I	34	Sec I	23
Ban I	10	Eco47 III	4	Msc I	2	SfaN I	26
Ban II	4	Eco57 I	1	Mse I	13	Sfi I	1
Bbe I	4	EcoN I	3	Msp I	33	Sma I	3
Bbv I	21	EcoO109 I	5	Nae I	4	SnaB I	-
Bbv II	3	EcoR I	1	Nar I	4	Spe I	-
Bcl I	-	EcoR II	11	Nci I	17	Sph I	1
Bcn I	17	EcoR V	2	Nco I	-	Spl I	-
Bgl I	3	Esp I	1	Nde I	1	Ssp I	1
Bgl II	1	Fnu4H I	41	Nhe I	1	Stu I	-

BsaA I	2	Fok I	11	Nla III	25	Sty I	4
Bsm I	3	Fsp I	3	Nla IV	25	Taq I	12
BsmA I	4	Gdi II	4	Not I	-	Tth111 I	1
Bsp1286 I	13	Gsu I	3	Nru I	2	Tth111 II	7
BspH I	2	Hae I	12	Nsi I	2	Xba I	1
BspM I	1	Hae II	11	Nsp7524 I	5	Xca I	1
BspM II	2	Hae III	31	NspB II	6	Xho I	1
Bsr I	17	Hga I	10	NspH I	5	Xcm I	1
BssH II	-	HgiA I	7	PaeR7 I	1	Xma I	3
BstB I	-	HgiE II	1	PflM I	4	Xmn I	2
BstE II	-	Hha I	33				

Enzyme	Site	Use	Site position (Fragment length) Fragment order											
AlwN I	cagnnn/ctg	1	1171(4716)	1										
Avr II	c/ctagg	1	4156(4716)	1										
BamH I	g/gatcc	1	4202(4716)	1										
Bgl II	a/gatct	1	3691(4716)	1										
BspM I	acctgc 4/8	1	2997(4716)	1										
Eag I	c/ggccg	1	3121(4716)	1										
Eco57 I	ctgaag 16/14	1	1058(4716)	1										
EcoR I	g/aattc	1	4714(4716)	1										
Esp I	gc/tnagc	1	4254(4716)	1										
HgiE II	accnnnnnnggt	1	998(4716)	1										
HinC II	gty/rac	1	3409(4716)	1										
Nde I	ca/tatg	1	3788(4716)	1										
Nhe I	g/ctagc	1	4483(4716)	1										
PaeR7 I	c/tcgag	1	785(4716)	1										
Pml I	cac/gtg	1	4043(4716)	1										
Pvu II	cag/ctg	1	1996(4716)	1										
Sal I	g/tcgac	1	3409(4716)	1										
Sca I	agt/act	1	4146(4716)	1										
Sfi I	ggccnnnn/nggcc	1	3886(4716)	1										
Sph I	gcatg/c	1	3498(4716)	1										
Ssp I	aat/att	1	460(4716)	1										
Tth111 I	gacn/nngtc	1	1840(4716)	1										
Xba I	t/ctaga	1	3749(4716)	1										
Xca I	gta/tac	1	1816(4716)	1										
Xho I	c/tcgag	1	785(4716)	1										
Xcm I	ccannnnn/nnnntgg	1	4056(4716)	1										
Afl III	a/crygt	2	1585(2585)	1	4170(2131)	2								
ApaL I	g/tgcac	2	1271(500)	2	1771(4216)	1								
Ase I	at/taat	2	185(3524)	1	3709(1192)	2								
BsaA I	yac/gtr	2	1835(2208)	2	4043(2508)	1								
BspH I	t/catga	2	865(2706)	1	3571(2010)	2								
BspM II	t/ccgga	2	2396(1936)	2	4332(2780)	1								
Cla I	at/cgat	2	693(3996)	1	4689(720)	2								
Drd I	gacnnnn/nngtc	2	1477(415)	2	1892(4301)	1								
Ear I	ctcttc 1/4	2	576(1131)	2	1707(3585)	1								
EcoR V	gat/atc	2	4336(191)	2	4527(4525)	1								
Msc I	tgg/cca	2	2616(1188)	2	3804(3528)	1								
Nru I	tcg/cga	2	728(2360)	1	3088(2356)	2								
Nsi I	atgca/t	2	271(266)	2	537(4450)	1								
PpuM I	rg/gwccy	2	2579(42)	2	2621(4674)	1								
Pvu I	cgat/cg	2	386(3796)	1	4182(920)	2								
Xmn I	gaann/nnttc	2	2027(2141)	2	4168(2575)	1								
Acc I	gt/mkac	3	1816(1593)	2	3409(526)	3	3935(2597)	1						
Bbv II	gaagac 2/6	3	2460(863)	2	3323(681)	3	4004(3172)	1						
Bgl I	gccnnnn/nggc	3	2892(234)	3	3126(761)	2	3887(3721)	1						
Bsm I	gaatgc 1/-1	3	422(77)	3	499(2208)	2	2707(2431)	1						
EcoN I	cctnn/nnnagg	3	470(2963)	1	3433(674)	3	4107(1079)	2						
Fsp I	tgc/gca	3	2606(98)	3	2704(1748)	2	4452(2870)	1						
Gsu I	ctggag 16/14	3	2079(580)	3	2659(590)	2	3249(3546)	1						
HinD III	a/agctt	3	265(3574)	1	3839(844)	2	4683(298)	3						
Sma I	ccc/ggg	3	511(3567)	1	4078(40)	3	4118(1109)	2						
Xma I	c/ccggg	3	511(3567)	1	4078(40)	3	4118(1109)	2						
Aha II	gr/cgyc	4	2855(657)	2	3512(114)	3	3626(21)	4	3647(3924)	1				
Ban II	grgcy/c	4	732(2843)	1	3575(14)	4	3589(429)	3	4018(1430)	2				
Bbe I	ggcgc/c	4	2855(657)	2	3512(114)	3	3626(21)	4	3647(3924)	1				
BsmA I	gtctc 1/5	4	373(1566)	2	1939(1791)	1	3730(358)	4	4088(1001)	3				
Dsa I	c/crygg	4	2613(919)	2	3532(329)	3	3861(172)	4	4033(3296)	1				

Eco47 III	agc/gct		4	2333(952) 2	3285(281) 4	3566(914) 3	4480(2569) 1
Gdi II	yggccg	-5/-1	4	3121(408) 3	3529(132) 4	3661(756) 2	4417(3420) 1
Nae I	gcc/ggc		4	2777(354) 3	3131(160) 4	3291(368) 2	3659(3834) 1
Nar I	gg/cgcc		4	2855(657) 2	3512(114) 3	3626(21) 4	3647(3924) 1
PflM I	ccannnn/ntgg		4	121(2570) 1	2691(49) 4	2740(1158) 2	3898(939) 3
Sty I	c/cwwgg		4	2691(1297) 2	3988(168) 3	4156(121) 4	4277(3130) 1
Ava I	c/ycgrg		5	511(274) 4	785(1850) 1	2635(1443) 2	4078(40) 5
				4118(1109) 3			
EcoO109 I	rg/gnccy		5	2579(42) 5	2621(915) 2	3536(341) 4	3877(404) 3
				4281(3014) 1			
Nsp7524 I	r/catgy		5	1585(367) 4	1952(292) 5	2244(1254) 2	3498(672) 3
				4170(2131) 1			
NspH I	rcatg/y		5	1585(367) 4	1952(292) 5	2244(1254) 2	3498(672) 3
				4170(2131) 1			
Ple I	gagtc	4/5	5	160(1055) 3	1215(471) 4	1686(1743) 1	3429(288) 5
				3717(1159) 2			
BstY I	r/gatcy		6	134(799) 3	933(11) 6	944(1449) 1	2393(1298) 2
				3691(511) 5	4202(648) 4		
Eae I	y/ggccr		6	2616(505) 3	3121(408) 4	3529(132) 6	3661(143) 5
				3804(613) 2	4417(2915) 1		
Mae I	c/tag		6	1092(1481) 1	2573(1177) 3	3750(407) 4	4157(109) 6
				4266(218) 5	4484(1324) 2		
NspB II	cmg/ckg		6	1000(245) 5	1245(632) 4	1877(119) 6	1996(925) 3
				2921(1331) 2	4252(1464) 1		
Rsa I	gt/ac		6	551(1231) 2	1782(2146) 1	3928(198) 5	4126(21) 6
				4147(403) 4	4550(717) 3		
Ava II	g/gwcc		7	2301(279) 4	2580(42) 7	2622(303) 3	2925(249) 5
				3174(88) 6	3262(591) 2	3853(3164) 1	
HgiA I	gwgcw/c		7	1271(500) 5	1771(824) 2	2595(291) 7	2886(587) 4
				3473(600) 3	4073(363) 6	4436(1551) 1	
Tth111 II	caarca	11/9	7	410(153) 5	563(409) 4	972(6) 7	978(33) 6
				1011(1129) 2	2140(2565) 1	4705(421) 3	
Cfr10 I	r/ccggy		8	429(2348) 1	2777(354) 5	3131(160) 7	3291(359) 4
				3650(9) 8	3659(255) 6	3914(638) 2	4552(593) 3
Mme I	tccrac	20/18	8	157(9) 8	166(406) 3	572(194) 5	766(445) 2
				1211(184) 6	1395(3033) 1	4428(87) 7	4515(358) 4
Dde I	c/tnag		9	3(364) 7	367(535) 3	902(409) 6	1311(467) 4
				1778(540) 2	2318(162) 8	2480(1754) 1	4234(21) 9
				4255(464) 5			
Mbo II	gaaga	8/7	9	485(91) 9	576(360) 6	936(771) 2	1707(753) 3
				2460(592) 4	3052(271) 8	3323(274) 7	3597(408) 5
				4005(1196) 1			
Ban I	g/gyrcc		10	2771(84) 7	2855(439) 3	3294(218) 5	3512(114) 6
				3626(21)10	3647(335) 4	3982(70) 8	4052(541) 2
				4593(43) 9	4636(2851) 1		
Hga I	gacgc	5/10	10	906(578) 3	1484(398) 4	1882(177) 8	2059(612) 2
				2671(150) 9	2821(264) 6	3085(32)10	3117(295) 5
				3412(259) 7	3671(1951) 1		
BstN I	cc/wgg		11	138(357) 6	495(930) 2	1425(13)11	1438(121) 8
				1559(1060) 1	2619(383) 5	3002(800) 3	3802(80) 9
				3882(61)10	3943(640) 4	4583(271) 7	
EcoR II	/ccwgg		11	138(357) 6	495(930) 2	1425(13)11	1438(121) 8
				1559(1060) 1	2619(383) 5	3002(800) 3	3802(80) 9
				3882(61)10	3943(640) 4	4583(271) 7	
Fok I	ggatg	9/13	11	126(633) 4	759(1154) 2	1913(141) 7	2054(159) 6
				2213(78) 9	2291(89) 8	2380(649) 3	3029(45)10
				3074(1506) 1	4580(21)11	4601(241) 5	
Hae II	rgcgc/y		11	1341(370) 6	1711(622) 3	2333(83) 8	2416(439) 4
				2855(430) 5	3285(227) 7	3512(54)10	3566(60) 9
				3626(21)11	3647(833) 2	4480(1577) 1	
Alw I	ggatc	4/5	12	135(381) 6	516(417) 5	933(12)11	945(74) 9
				1019(1375) 1	2394(570) 4	2964(721) 2	3685(13)10
				3698(332) 7	4030(172) 8	4202(1)12	4203(648) 3
Hae I	wgg/ccw		12	298(811) 3	1109(452) 5	1561(11)12	1572(1044) 1
				2616(397) 6	3013(57)10	3070(72) 9	3142(650) 4
				3792(12)11	3804(141) 8	3945(214) 7	4159(855) 2
Mae II	a/cgt		12	795(89) 7	884(952) 2	1836(426) 5	2262(230) 6
				2492(24)10	2516(589) 4	3105(56) 9	3161(778) 3
				3939(87) 8	4026(18)12	4044(21)11	4065(1446) 1
Taq I	t/cga		12	17(403) 5	420(274) 8	694(92)10	786(701) 2

				1487(1307) 1	2794(141) 9	2935(475) 4	3410(278) 7
				3688(8)12	3696(679) 3	4375(315) 6	4690(43)11
Bsp1286 I	gdgch/c		13	732(539) 4	1271(500) 5	1771(824) 2	2595(291) 8
				2886(587) 3	3473(102) 9	3575(14)13	3589(394) 6
				3983(35)10	4018(33)11	4051(22)12	4073(363) 7
				4436(1012) 1			
Hph I	ggtga	8/7	13	107(120) 9	227(24)12	251(216) 8	467(59)10
				526(1443) 1	1969(9)13	1978(555) 3	2533(221) 7
				2754(854) 2	3608(45)11	3653(536) 4	4189(398) 5
				4587(236) 6			
Mse I	t/taa		13	186(209) 8	395(379) 4	774(107) 9	881(927) 2
				1808(282) 5	2090(32)11	2122(220) 7	2342(1368) 1
				3710(56)10	3766(6)13	3772(886) 3	4658(22)12
				4680(222) 6			
Hinf I	g/antc		15	154(6)15	160(85)12	245(172)10	417(56)14
				473(742) 2	1215(396) 4	1611(75)13	1686(346) 5
				2032(504) 3	2536(221) 8	2757(298) 6	3055(154)11
				3209(220) 9	3429(288) 7	3717(1153) 1	
Mae III	/gtnac		15	236(295) 6	531(92)11	623(427) 4	1050(116) 9
				1166(63)14	1229(609) 3	1838(95)10	1933(213) 8
				2146(84)13	2230(23)15	2253(660) 2	2913(267) 7
				3180(1320) 1	4500(88)12	4588(364) 5	
Alu I	ag/ct		17	266(761) 2	1027(257) 9	1284(46)14	1330(90)11
				1420(226)10	1646(283) 7	1929(19)15	1948(49)13
				1997(11)17	2008(57)12	2065(908) 1	2973(403) 5
				3376(464) 3	3840(392) 6	4232(452) 4	4684(15)16
				4699(283) 8			
Bcn I	ccs/gg		17	511(1)15	512(695) 3	1207(701) 2	1908(35)14
				1943(306) 7	2249(328) 6	2577(226) 9	2803(724) 1
				3527(292) 8	3819(73)11	3892(129)10	4021(57)12
				4078(1)16	4079(39)13	4118(1)17	4119(424) 5
				4543(684) 4			
Bsr I	actgg	1/-1	17	598(455) 4	1053(117)12	1170(13)17	1183(630) 2
				1813(28)15	1841(415) 5	2256(24)16	2280(271) 7
				2551(377) 6	2928(249) 9	3177(204)11	3381(70)14
				3451(474) 3	3925(219)10	4144(264) 8	4408(95)13
				4503(811) 1			
Nci I	cc/sgg		17	511(1)15	512(695) 3	1207(701) 2	1908(35)14
				1943(306) 7	2249(328) 6	2577(226) 9	2803(724) 1
				3527(292) 8	3819(73)11	3892(129)10	4021(57)12
				4078(1)16	4079(39)13	4118(1)17	4119(424) 5
				4543(684) 4			
Sau96 I	g/gncc		17	2114(187) 8	2301(279) 3	2580(42)13	2622(179) 9
				2801(124)11	2925(249) 6	3174(88)12	3262(275) 4
				3537(316) 2	3853(25)14	3878(7)17	3885(10)16
				3895(22)15	3917(199) 7	4116(166)10	4282(259) 5
				4541(2289) 1			
Bbv I	gcagc	8/12	21	970(206) 7	1176(3)18	1179(65)13	1244(419) 4
				1663(18)17	1681(171) 8	1852(97)12	1949(46)14
				1995(3)19	1998(378) 5	2376(123)11	2499(3)20
				2502(129)10	2631(24)16	2655(633) 3	3288(158) 9
				3446(763) 2	4209(33)15	4242(3)21	4245(242) 6
				4487(1199) 1			
Dpn I	ga/tc		23	135(252) 8	387(130)12	517(409) 3	926(8)21
				934(11)20	945(75)14	1020(1374) 1	2394(207) 9
				2601(317) 5	2918(15)18	2933(31)15	2964(272) 7
				3236(359) 4	3595(91)13	3686(6)22	3692(6)23
				3698(302) 6	4000(12)19	4012(18)17	4030(153)11
				4183(20)16	4203(162)10	4365(486) 2	
Mbo I	/gatc		23	135(252) 8	387(130)12	517(409) 3	926(8)21
				934(11)20	945(75)14	1020(1374) 1	2394(207) 9
				2601(317) 5	2918(15)18	2933(31)15	2964(272) 7
				3236(359) 4	3595(91)13	3686(6)22	3692(6)23
				3698(302) 6	4000(12)19	4012(18)17	4030(153)11
				4183(20)16	4203(162)10	4365(486) 2	
Sau3A I	/gatc		23	135(252) 8	387(130)12	517(409) 3	926(8)21
				934(11)20	945(75)14	1020(1374) 1	2394(207) 9
				2601(317) 5	2918(15)18	2933(31)15	2964(272) 7
				3236(359) 4	3595(91)13	3686(6)22	3692(6)23
				3698(302) 6	4000(12)19	4012(18)17	4030(153)11
				4183(20)16	4203(162)10	4365(486) 2	
Sec I	c/cnngg		23	110(401) 4	511(1)21	512(913) 2	1425(1188) 1
				2613(78)11	2691(202) 8	2893(633) 3	3526(6)20

				3532(269) 6	3801(18)18	3819(42)14	3861(20)17
				3881(1)22	3882(106)10	3988(45)12	4033(45)13
				4078(40)15	4118(1)23	4119(37)16	4156(121) 9
				4277(306) 5	4583(14)19	4597(229) 7	
BstU I	cg/cg		25	384(345) 5	729(50)20	779(181) 8	960(581) 2
				1541(343) 6	1884(103)15	1987(2)25	1989(69)17
				2058(370) 4	2428(97)16	2525(122)13	2647(26)22
				2673(145) 9	2818(10)23	2828(129)11	2957(66)18
				3023(61)19	3084(5)24	3089(27)21	3116(129)12
				3245(115)14	3360(343) 7	3703(145)10	3848(520) 3
				4368(732) 1			
Nla III	catg/		25	231(305) 4	536(115)16	651(112)17	763(103)19
				866(720) 1	1586(262) 5	1848(105)18	1953(165)11
				2118(63)22	2181(64)21	2245(225) 7	2470(134)13
				2604(207) 8	2811(57)23	2868(15)25	2883(128)14
				3011(117)15	3128(186)10	3314(39)24	3353(146)12
				3499(73)20	3572(599) 2	4171(191) 9	4362(347) 3
				4709(238) 6			
Nla IV	ggn/ncc		25	1516(39)19	1555(745) 2	2300(279) 4	2579(43)17
				2622(114)10	2736(35)21	2771(35)22	2806(49)16
				2855(318) 3	3173(121) 9	3294(218) 5	3512(24)24
				3536(90)12	3626(21)25	3647(206) 7	3853(25)23
				3878(39)20	3917(65)15	3982(70)14	4052(150) 8
				4202(79)13	4281(102)11	4383(210) 6	4593(43)18
				4636(1596) 1			
SfaN I	gcatc	5/9	26	21(311) 6	332(120)13	452(87)15	539(135)12
				674(84)17	758(740) 1	1498(220)10	1718(21)23
				1739(57)21	1796(116)14	1912(240) 9	2152(62)20
				2214(78)18	2292(87)16	2379(9)25	2388(252) 8
				2640(388) 3	3028(7)26	3035(368) 4	3403(253) 7
				3656(12)24	3668(463) 2	4131(335) 5	4466(43)22
				4509(70)19	4579(158)11		
ScrF I	cc/ngg		28	138(357) 4	495(16)23	511(1)26	512(695) 1
				1207(218)10	1425(13)24	1438(121)13	1559(349) 5
				1908(35)21	1943(306) 7	2249(328) 6	2577(42)18
				2619(184)12	2803(199)11	3002(525) 2	3527(275) 8
				3802(17)22	3819(63)15	3882(10)25	3892(51)17
				3943(78)14	4021(57)16	4078(1)27	4079(39)20
				4118(1)28	4119(424) 3	4543(40)19	4583(271) 9
Hae III	gg/cc		31	299(483) 3	782(328) 6	1110(434) 5	1544(18)25
				1562(11)27	1573(542) 1	2115(502) 2	2617(184) 9
				2801(213) 8	3014(57)19	3071(51)20	3122(21)24
				3143(89)16	3232(234) 7	3466(64)18	3530(8)29
				3538(124)13	3662(131)12	3793(12)26	3805(73)17
				3878(8)30	3886(9)28	3895(22)23	3917(29)22
				3946(170)10	4116(38)21	4154(6)31	4160(123)14
				4283(135)11	4418(123)15	4541(474) 4	
Hha I	gcg/c		33	363(70)22	433(17)33	450(222) 8	672(287) 4
				959(109)17	1068(174)11	1242(100)19	1342(67)23
				1409(270) 5	1679(33)29	1712(143)14	1855(30)30
				1885(103)18	1988(346) 3	2334(83)20	2417(190)10
				2607(36)28	2643(62)24	2705(151)13	2856(259) 6
				3115(131)16	3246(40)27	3286(75)21	3361(152)12
				3513(54)26	3567(60)25	3627(21)32	3648(199) 9
				3847(252) 7	4099(354) 2	4453(28)31	4481(132)15
				4613(466) 1			
HinP I	g/cgc		33	363(70)22	433(17)33	450(222) 8	672(287) 4
				959(109)17	1068(174)11	1242(100)19	1342(67)23
				1409(270) 5	1679(33)29	1712(143)14	1855(30)30
				1885(103)18	1988(346) 3	2334(83)20	2417(190)10
				2607(36)28	2643(62)24	2705(151)13	2856(259) 6
				3115(131)16	3246(40)27	3286(75)21	3361(152)12
				3513(54)26	3567(60)25	3627(21)32	3648(199) 9
				3847(252) 7	4099(354) 2	4453(28)31	4481(132)15
				4613(466) 1			
Hpa II	c/cgg		33	249(181) 9	430(82)21	512(505) 2	1017(190) 8
				1207(26)27	1233(147)13	1380(529) 1	1909(34)26
				1943(307) 4	2250(147)14	2397(180)10	2577(201) 7
				2778(26)28	2804(238) 5	3042(90)19	3132(160)11
				3292(76)22	3368(160)12	3528(123)17	3651(9)32
				3660(15)30	3675(144)15	3819(15)31	3834(59)23
				3893(22)29	3915(107)18	4022(57)24	4079(40)25
				4119(87)20	4206(127)16	4333(211) 6	4544(9)33