Cloning and Expression in Escherichia coli of the Cytochrome $c_{552}$ Gene from Thermus thermophilus HB8

EVIDENCE FOR GENETIC LINKAGE TO AN ATP-BINDING CASSETTE PROTEIN AND INITIAL CHARACTERIZATION OF THE cycA GENE PRODUCTS

(Received for publication, December 22, 1997, and in revised form, February 6, 1998)

J. Andrew Keightley‡§, Donita Sanders‡, Thomas R. Todaro‡, Andrzej Pastuszyn¶, and James A. Fee‡§

From the ‡Department of Biology, University of California at San Diego, La Jolla, California 92093, the §Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and the ¶Protein Chemistry Laboratory and Department of Biochemistry and Molecular Biology, University of New Mexico, Albuquerque, New Mexico 87131-5221

We report sequence of Thermus thermophilus HB8 DNA containing the gene (cycA) for cytochrome $c_{552}$ and a gene (cycB) encoding a protein homologous with one subunit of an ATP-binding cassette transporter. The cycA gene encodes a 17-residue N-terminal signal peptide with following amino acid sequence identical to that reported by (Titani, K., Ericsson, L. H., Hon-nami, K., and Miyazawa, T. (1985) Biochem. Biophys. Res. Commun. 125, 781–787). A modified cycA was placed under control of the T7 promoter and expressed in Escherichia coli. Protein identical to that predicted from the gene sequence was found in two heme C-containing fractions. Fraction $rC_{552}$, characterized by an $\alpha$-band at 552 nm, contains $60–70\%$ of a protein highly similar to native cytochrome $c_{552}$ and $30–40\%$ of a protein that contains a modified heme. Cytochrome $rC_{552}$ is monomeric and is an excellent substrate for cytochrome $ba_{3}$. Cytochrome $rC_{552}$ is characterized by an $\alpha$-band at 557 nm, contains $90\%$ heme C and $10\%$ of non-C heme, exists primarily as a homodimer, and is essentially inactive as a substrate for cytochrome $ba_{3}$. We suggest that $rC_{552}$ is a “conformational isomer” of $rC_{552}$ having non-native, axial ligands to the heme iron and an “incorrect” protein fold that is stabilized by homodimer formation.

Three cytochromes $c$ from Thermus thermophilus HB8 have been described (cf. Fee et al. for review (1)). One of these, the very basic cytochrome $c_{552}$, is a good substrate for the terminal oxidase, cytochrome $ba_{3}$ (−250 electrons/s at 25 °C (2)), and a poor substrate for the terminal oxidase cytochrome $ca_{3}$ (−1–2 electrons/s at 25 °C, Ref. 3). Mechanistic and biophysical examination of these respiratory enzymes is greatly impeded, because useful quantities are only obtained from large amounts of bacterial cell mass, and yields of the cytochrome $c_{552}$ from T. thermophilus are small; for example, <1 mg of pure cytochrome $c_{552}$ is obtained from 1 kg of Thermus cell paste (4), although this may depend on growth conditions and purification procedures (4, 5). Studies on the functional properties of the two Thermus cytochrome c oxidases are thus limited by the availability of and by the need for designed mutations in the cytochrome c. The prospect of having the three-dimensional structure of cytochrome $ba_{3}$ (6) and the report of an atomic resolution x-ray structure of native Thermus cytochrome $c_{552}$ (7) emphasize the potential of this system to provide novel insight on cytochrome c oxidase function. In the course of constructing a deletion series of a Thermus genomic DNA fragment for sequencing the $c_{552}$ locus, we observed that colonies from a certain time point in the series were distinctly reddish-brown in color (8). This led us to attempt engineered expression of the cytochrome $c_{552}$ gene in E. coli.

The c-type cytochromes have covalently attached heme groups usually linked via thioether bonds between two cysteine residues and the 2- and 4-vinyl groups of the heme. Bacterial cytochromes $c$ are located in the periplasmic space in Gram-negative bacteria, or in the case of Gram-positive bacteria, they are bound to the outer surface of the plasma membrane by means of a hydrophobic N-terminal segment (9). Cytochrome $c_{552}$ of T. thermophilus was previously shown to be located in the periplasmic space (10). Bacterial cytochrome $c$ synthesis involves several steps: (a) a pre-apocytochrome $c$ is synthesized in the cytoplasm, i.e. a protein having no heme attached and which possesses an additional N-terminal domain containing signals to transport the protein across the plasma membrane and to specifically proteolyze the signal sequence peptide. (b) The apoprotein and protoheme IX are transported across the plasma membrane, where the signal peptide is cleaved. (c) Once heme and apoprotein are in the periplasmic space, the heme is covalently attached to the protein. And finally, (d) the protein folds into the conformation of the mature holoprotein (cf. Ref. 11 for review). While many details remain unknown, the entire process occurs within a few seconds of protein synthesis (12).

As many as 16 genes and/or gene products are essential for cytochrome $c$ synthesis (cf. Refs. 11 and 13 for review). Included are ferrochelatase (HemH), an enzyme responsible for incorporating iron into protoporphyrin IX (14); heme lyase (CcmF and CcmH) to attach heme to apoprotein by catalyzing the formation of the thioether (see Refs. 11 and 15 for review and additional references); a thioredoxin-like molecule located in the periplasm maintains reduced cysteine SH in apocytochromes $c$ (CyeY or HelX, Refs. 16, 17); a disulfide isomerase that presumably helps maintain the apocytochrome $c$ in a state suitable...
for reaction with the lyase (DipZ) (18); unidentified "assembly factors" such as CycH in *Paracoccus denitrificans* (19); and genes that encode for ABC transporters (for review, see Refs. 11 and 20). For example, the helABCD operon of *Rhodobacter capsulatus* (17) and the cysFWXY operon of *Bradyrhizobium japonicum* (21) encode ABC-type ATP-dependent transport proteins that are essential for cytochrome c synthesis in their respective organisms (cf. Refs. 11 and 13 for review and additional citations). Finally, there are specialized gene products for synthesis of specific cytochromes c; e.g., cytochrome c of the bc₁ complex (19, 22), and there appear to be specific sequences within the cytochrome c molecule itself that contain information to guide the synthetic process (23, 24).

Given the complexity of cytochrome c maturation, one might suspect that heterologous expression would be inefficient and, indeed, there are numerous reports to that effect (25–31). There are, however, some notable exceptions. Thus, Ubbink et al. (32) describe expression of the complete (including code for the signal peptide) *Thiobacillus versutus* c₁₅₀₀ gene in *E. coli*. The cell mass obtained from each liter of culture medium yielded 1–2 mg of holoprotein, all of which appeared to reside in the periplasm. This allowed the authors to obtain enough protein for high-resolution NMR studies. Sambongi and Ferguson (33) made the remarkable observation that synthesis of holc *Paracoccus denitrificans* cytochrome c₅₅₀ in *E. coli* requires an N-terminal signal peptide to target the gene product to the periplasm, whereas *E. coli* can synthesize holc *Hydrogenobacter thermophilus* cytochrome c₅₉₂ without targeting to the periplasm. These authors attributed heme C formation to “spontaneous cytoplasmic attachment of heme to the (properly folded) thermostable protein,” revealing yet another complexity of cytochrome c synthesis in *E. coli*. We describe here the first heterologous expression system for large scale preparation of Thermus cytochrome c₅₅₂ from *E. coli* and our initial characterization of the gene products.

**EXPERIMENTAL PROCEDURES**

*Molecular Biology Methods—Genomic DNA was isolated from *T. thermophilus* DSM 25608 (ATCC 27643) as described previously (34). E. coli strains were cultured in L-broth (35) modified by omitting glucose and lowering the sodium chloride to 0.5%, with 1.5% agar (Difco) for plates. Plasmid preparations were carried out by standard procedures (36). Probe design for genomic hybridization was based on the method of most probable codons (37, 38; see "Results") using the GCG (39) program, SORT, was used to calculate expected properties of the isolated gene sequence. Mass spectra were calculated using the GCG (39) program, PEPTIDE-III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47). Reconstructions of mass spectra from ion spectra were made using the Perkin-Elmer SCIEX API III mass analyzer with the orifice potential set at 100 V (47).

**Expression and Purification of Recombinant Cytochrome c₅₅₂—10 ml of culture medium (LB and 50 mg/ml ampicillin) were inoculated from an M13 streaked plate of *E. coli* strain BL21-DE3 cells containing the pETC₅₅₂ plasmid. After incubation overnight at 37 °C, this culture was used to inoculate 1 liter of LB medium, containing 50 mg/ml ampicillin, in a 2.8-liter Fernbach flask. This culture was incubated at 37 °C overnight on a rotary shaker set at 125 rpm. The cells were pelleted by centrifugation at 5000 × g for 5 min. The cell pellet was distinctively reddish-brown color, indicating heme protein overproduction. The pellet from 1 liter of culture was resuspended in 25 ml of 50 mM Tris-HCl, pH 8.0, 4 mg/ml lysozyme, 40 units/ml DNase I, 3 units/ml RNase A, 0.1% Triton X-100. p-toluenesulfonyl fluoride was added (2 mM) to inhibit proteolysis, 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 × g for 15 min at 4 °C. The supernatant, which had the characteristic reddish-brown color, was loaded onto a CM-52 gravity column that had been equilibrated with 25 mM Tris-HCl, 0.1 mM EDTA, pH 8.0 (25 mM TE buffer (TE buffer: 25 mM Tris-HCl, 0.1 mM EDTA, pH 8.0)) at 4 °C, washed overnight with several column volumes of equilibration buffer, and eluted with a gradient of 0–1 M NaCl in 10 mM TE buffer. Fractions containing the reddish-orange protein were combined and concentrated using Centricon-10 concentrators (Amicon) with YM-3 or YM-10 membranes. The concentrated protein was desalted (10 mM TE buffer) using a PD-10 (Amersham Pharmacia Biotech) and passed over a EconoPac CM Column (Bio-Rad) attached to an Amersham Pharmacia Biotech fast protein liquid chromatography purification system. The column was washed with several column volumes of the same buffer, and the protein was eluted with a gradient of 0–2 M NaCl in 10 mM TE buffer. The final step of the procedure involves fast protein liquid chromatography purification using a HiLoad 16/60 Superdex 75 column (Amersham Pharmacia Biotech) that was equilibrated with TE buffer (see Fig. 3). The protein was concentrated as described above.

1 The abbreviations used are: PCR, polymerase chain reaction; nt, nucleotide; ABC, ATP-binding cassette; Phc, pyridine hemochrom; IPTG, isopropyl-β-D-thiogalactoside.
RESULTS

The amino acid sequence of *T. thermophilus* HB8 cytochrome c₅₅₂ was previously determined by Edman degradation of proteolytically generated peptides (48). We used reverse translation and took advantage of the highly skewed codon usage of *T. thermophilus* (cf. Refs. 49 and 50) to design a nondegenerate oligonucleotide probe for the gene that was complementary to the protein sequence, QGQIEVKGMKYNG: 5'-CCGTTTC-TACTTCATCGCCCTGCACCCCTGGTTCTGAG-3'. This probe was used to identify and isolate a 1.6-kilobase pair KpnI fragment as described under "Experimental Procedures." The entire fragment was sequenced on both strands and is presented in Fig. 1; the sequence of the probe was found to differ in only 4 of the 38 positions (at the positions underlined above).

The DNA sequence reveals two open reading frames. The first open reading frame begins at position 154 and ends at position 600 with the stop codon TAA. The translated nucleotide sequence from 205 to 597 corresponds to the chemically determined amino acid sequence of native cytochrome c₅₅₂ (48), and we designate this *cycA*. Translation from position 154 through 204 indicates the presence of a 17-amino acid signal peptide (underlined in Fig. 1; cf. Ref. 51). The region upstream of the initiation codon (position 154) contains elements that probably serve as a promoter, including −10 and −35 regions and a ribosome binding site, also underlined in Fig. 1 (cf. Ref. 52 for a review of *Thermus* promoters). Only 9 bases downstream of the TAA stop codon in *cycA*, a second open reading frame begins at nucleotide 610 and ends with a termination codon (TAG) at position 1404; we designate this region as *cycB*. The DNA sequences in the *cycA* and *cycB* regions were subjected to codon usage analysis using a table of codon usage generated from 14 previously sequenced *T. thermophilus* genes (cf. Ref. 42). The average codon preference value for *cycA* (nt 154 to nt 600) is 1.3084, and the value for *cycB* (nt 610 to nt 1404) is 1.2317, values typical of other *Thermus* genes (cf. Refs. 42 and 50). In contrast, average codon preferences of the other two possible frames in these two regions range from 0.3858 to 0.7692. These results reveal that both *cycA* and *cycB* adhere closely to *Thermus* usage and are thus properly assigned as open reading frames. The sequence of *cycA* predicts exactly the amino acid sequence of cytochrome c₅₅₂ determined chemically by Titani *et al.* (48).

The translated amino acid sequence from *cycB* is also shown in Fig. 1. A TFASTA search (53) against the Swissprot data base retrieved numerous proteins that share amino acid sequence similarity, with the highest scoring hits being members of the protein superfamily known as ABC transporters (data not shown). These results reveal that both *cycA* and *cycB* adhere closely to *Thermus* usage and are thus properly assigned as open reading frames. The sequence of *cycA* predicts exactly the amino acid sequence of cytochrome c₅₅₂ determined chemically by Titani *et al.* (48).

The translated amino acid sequence from *cycB* is also shown in Fig. 1. A TFASTA search (53) against the Swissprot data base retrieved numerous proteins that share amino acid sequence similarity, with the highest scoring hits being members of the protein superfamily known as ABC transporters (data not shown). These results reveal that both *cycA* and *cycB* adhere closely to *Thermus* usage and are thus properly assigned as open reading frames. The sequence of *cycA* predicts exactly the amino acid sequence of cytochrome c₅₅₂ determined chemically by Titani *et al.* (48).

The translated amino acid sequence from *cycB* is also shown in Fig. 1. A TFASTA search (53) against the Swissprot data base retrieved numerous proteins that share amino acid sequence similarity, with the highest scoring hits being members of the protein superfamily known as ABC transporters (data not shown). These results reveal that both *cycA* and *cycB* adhere closely to *Thermus* usage and are thus properly assigned as open reading frames. The sequence of *cycA* predicts exactly the amino acid sequence of cytochrome c₅₅₂ determined chemically by Titani *et al.* (48).

The translated amino acid sequence from *cycB* is also shown in Fig. 1. A TFASTA search (53) against the Swissprot data base retrieved numerous proteins that share amino acid sequence similarity, with the highest scoring hits being members of the protein superfamily known as ABC transporters (data not shown). These results reveal that both *cycA* and *cycB* adhere closely to *Thermus* usage and are thus properly assigned as open reading frames. The sequence of *cycA* predicts exactly the amino acid sequence of cytochrome c₅₅₂ determined chemically by Titani *et al.* (48).
Expression of Thermus Cytochrome c\textsubscript{552} Gene in E. coli

\textbf{Fig. 2.} Amino acid sequence alignment showing \textit{cycB} compared with six homologous proteins implicated in cytochrome \textit{c} biogenesis. The sequences (accession numbers are indicated in parentheses) are from: \textit{T. thermophilus} (M93437) \textit{cycB}, \textit{E. coli} (P33931) \textit{cycA}, \textit{B. japonicum} (M60674) \textit{cycB}, \textit{R. capsulatus} (X63462) \textit{cycA}, \textit{P. denitrificans} (U83136) \textit{cycA}, \textit{D. radiodurans} (P52218) \textit{cycA} (cf. Ref. 11 for original citations). All are homologs of the ATP binding component of putative heme transport systems. The sequences which comprise the nucleotide binding fold or “Walker motif” (54) are overlined. Conserved amino acids in this alignment are indicated by asterisks, the presence of conservative (+) or nonconservative (−) substitutions is so indicated, and when inserts are necessary to preserve alignments, these are indicated with a \textit{dash}.  

\begin{table}[htb]
\centering
\begin{tabular}{ccc}
\hline
\textbf{Species} & \textbf{Accession} & \textbf{Sequence} \\
\hline
\textit{T. thermophilus} & M93437 & EGVTLTVTVHQCQGAEAFLLAVNHALLREQLGELVRQGEEYARPPCMARAFLGLHRNLSSLGES \\
\textit{E. coli} & P33931 & KGQLVLAXEHFQGLCILKAEEILGKFAGVDDVGWG \\
\textit{B. japonicum} & M60674 & RGLLTAMITMLISREDRLTGGVA \\
\textit{R. capsulatus} & X63462 & AGGALAMTHILDLGSLERVLALDFEPPPENGGRRAGHDQGDAGFL \\
\textit{P. denitrificans} & U83136 & EGVTLTVTVHQCQGAEAFLLAVNHALLREQLGELVRQGEEYARPPCMARAFLGLHRNLSSLGES \\
\textit{D. radiodurans} & P52218 & KGQLVLAXEHFQGLCILKAEEILGKFAGVDDVGWG \\
\hline
\end{tabular}
\end{table}

Across membranes (cf. Ref. 20 for a general review). As reviewed by Thyony-Meyer (11), a subset of this family is essential for cytochrome \textit{c} maturation. Fig. 2 shows a complete alignment of the CycB amino acid sequence with those of the ATP-binding subunit of selected ABC transporters that have been implicated in cytochrome \textit{c} maturation in several bacteria. The designated \textbf{A} and \textbf{B}-sites (Fig. 2) form the ATP binding fold (20, 54), and the sequence similarity in these regions as well as throughout the molecule is evident. CycB from \textit{Thermus} thus appears to be a homolog of known cytochrome \textit{c} maturation proteins and therefore probably has the same function in \textit{T. thermophilus} (see “Discussion”). 

Two palindromic sequences were identified using the GCG (39) program STEMLOOP as undetected in Fig. 1. The first palindromic sequence occurs at nt 606 to nt 617 with a complementary sequence at nt 633 to nt 622. This includes 2 G=U pairings, which suggests that this palindrome may form in the mRNA and thus may regulate a termination of translation. Since the start codon of \textit{cycB} is located in the leading strand of this potential structure, it may provide part of an attenuation signal to modulate CycB formation.\textsuperscript{2} Potentially the strongest stemloop-forming sequence begins at nt 1435 to nt 1448 with a complementary region nt 1466 to nt 1453. This occurs 30 nucleotides after the termination codon of \textit{cycB}, it contains a single A-C mismatch and is probably involved in termination of transcription.

Although functional in \textit{Thermus} cells, the suggested promoter elements upstream of \textit{cycA} (Fig. 1) do not direct the expression of detectable amounts of cytochrome \textit{c}\textsubscript{552} in \textit{E. coli}. Other \textit{T. thermophilus} H88 promoters containing similar –10 and –35 elements do function in \textit{E. coli} (cf. Refs. 55–57), however, and transcription from a 23 S rRNA-5 S rRNA rRNA\textsuperscript{10} promoter with similar –10 and –35 elements was found to promote the initiation of transcription at the identical nucleotide position in both \textit{E. coli} and \textit{Thermus} (58). That holocytochrome \textit{c} can be expressed from \textit{cycB} in \textit{E. coli} if preceded by a suitable promoter was first evidenced by the appearance of reddish-colored colonies among clones bearing a set of nested deletions that were prepared for DNA sequencing (8). This observation suggested to us that placing \textit{cycA} under the control of another promoter might permit expression of useful quantities of recombinant cytochrome \textit{c}\textsubscript{552} in \textit{E. coli}.

Accordingly, a plasmid was prepared by PCR amplification of a DNA fragment containing \textit{cycA} in which the forward primer includes an \textit{NdeI} site and an alanine codon just upstream of the codon for glutamine at nucleotide position 205, while the reverse primer contains a \textit{BamHI} site and the normal stop codon for \textit{cycA} (see “Experimental Procedures”). This fragment was then ligated into the Studier vector (59), pET17b, that had been digested with \textit{NdeI} and \textit{BamHI}; by this strategy, \textit{cycA} was placed immediately downstream of the T7 RNA polymerase promoter. The standard strategy in this expression system is to activate the expression of T7 RNA polymerase from the host

\begin{footnotesize}
\begin{itemize}
\item A significant effort was made to detect an RNA transcript in \textit{Thermus} cells containing both \textit{cycA} and \textit{cycB} using Northern analyses (36). By rapidly cooling \textit{Thermus} cells grown at 70 °C to ice temperature, we were able to detect very low levels of an RNA molecule that bound an oligonucleotide probe specific for \textit{cycB} but could not detect any RNA using an oligonucleotide probe specific for \textit{cycB} (J. A. Fee, J. A. Keightley, and D. Sanders, unpublished results).
\end{itemize}
\end{footnotesize}
Expression of Thermus Cytochrome $c_{552}$ Gene in E. coli

gene through an IPTG-inducible promoter, which then transcribes an open reading frame placed downstream of the T7 promoter in a pET construct (59). The complete DNA sequence of the insert showed that no mutations had been introduced during the construction process. The predicted gene product thus begins with MAQADGAKI and ends with KKLGLK (cf. Fig. 1). Lacking a signal sequence, any protein formed is likely to be restricted to the cytoplasm (cf. Ref. 11).

When E. coli cells (strain BL21-DE3) containing this plasmid are grown on agar plates in the absence of the T7 RNA polymerase inducer (IPTG) they appear as reddish-colored colonies. Similarly, when the cells are grown aerobically in 1-liter batches, again in the absence of IPTG, the cell pellets are distinctly red-orange in color. However, induction of T7 RNA polymerase with IPTG in liquid cultures causes the formation of inclusion bodies in the host cytoplasm and leads to no detectable holocytochrome $c$ formation. These observations indicate that a balance exists under noninducing conditions between protein synthesis initiated by a somewhat “leaky” IPTG-inducible promoter in the DE3 lysogen (cf. Ref. 59) and the ability of the cell to provide the necessary heme for cytochrome synthesis. We see no evidence for apocytochrome $c$ in our preparations (see below). When cell pellets from cultures grown under expressing conditions (without IPTG) are treated with a lysozyme/EDTA mixture (60) and centrifuged, the resulting supernatant has only a faint color, suggesting that little if any heme protein was transported to the periplasm (not shown).

The ability of the cell to provide the necessary heme for cytochrome synthesis is consistent with the appearance of a 30-kDa dimer at 65 ml (Fig. 3) and a 15-kDa monomer at 79 ml (Fig. 3). The ratio of absorbance at 552 to that at 557 is about 3:1, whereas for cytochrome $c_{552}$ from the adsorbed material. Accordingly, a purification procedure was developed based on the cationic nature of cytochrome $c_{552}$ at neutral pH (the pI predicted by PEPTIDESORT is 10.2, in agreement with that observed for native protein (5) and its relatively small molecular mass (~15 kDa) (see “Experimental Procedures”). As shown in Fig. 3, the elution profile of the final gel filtration column reveals the presence of two cytochrome $c$-containing fractions. The void volume of this column is ~24 ml with an exclusion limit of ~100,000 Da. The first peak, eluting at ~65 ml, shows an $a$-band absorbance in the reduced state at 557 nm (Fig. 3) and is designated cytochrome $r_{C557}$, while the major peak, eluting at ~79 ml, has an $a$-band at 552 nm and is designated cytochrome $r_{C552}$. (We have not done a complete standardization of the column, but the elution profile is consistent with the appearance of a 30-kDa dimer at 65 ml and a 15-kDa monomer at ~80 ml.) The ratio of absorbance at 408 nm to the absorbance at 280 nm ($r = A_{408}/A_{280}$) is a good indicator of purity, and these are relatively constant across the two peaks. For $r_{C557}$ the purity ratio is $r = 3.1$, whereas for $r_{C552}$ the ratio is $r = 4.5$. Yoshida et al. (4) reported a value of 5.1 for native cytochrome $c_{552}$ from Thermus cells, and our best preparations of $r_{C552}$ have values of $r = 4.5$. As evidenced by gel electrophoresis (Fig. 4), N-terminal sequence analyses (Table I) and electrospray mass spectrometry (Figs. 5 and 6), each of these fractions contain highly pure cytochrome $c$ proteins that are encoded for by the Thermus cycA gene.

Fig. 4A shows SDS gels stained with Coomassie Blue. The sample of native protein shows a majority band near 14.5 kDa and weak impurity bands at ~21.5 and ~40 kDa. The $r_{C552}$ sample shows a single band at ~14.5 kDa and no obvious impurity bands. Similarly $r_{C552}$ shows a single band at ~14.5 but has a band of lesser intensity near 30 kDa. Thus, the two recombinant proteins appear to have molecular masses close to that of the native protein and consistent with translation of the cloned gene. Fig. 4B shows SDS gels that were stained for heme based peroxidase activity according to the method of Thomas et al. (61). A and B are part of a single 15% SDS-polyacrylamide gel electrophoresis gel that was run according to the method of Downer et al. (45) and cut in half prior to staining. Each lane in A was loaded with 2.5 µg of protein, and each lane in B was loaded with 10 µg of protein. Prior to loading onto the gel, the protein/SDS solutions were boiled for 5 min. C is a nondenaturing 15% acrylamide gel equilibrated at pH 4.5, run with reversed polarity according to the method of Gabriel (46) and stained with Coomassie Blue. Each lane contained 2 µg of protein.

$^{3}$ J. A. Keightley, unpublished observations.
the 14.8-kDa protein that survives the denaturing influence of the SDS even at 100 °C (see below).

While native cytochrome c552 is N-terminally blocked (48), the recombinant protein was designed to begin with the MAQ sequence so as to avoid this potentially complicating factor and to remove the signal sequence. Partial N-terminal sequence information on both of these fractions (Table I) show sequence beginning with alanine. The proteins are very pure with respect to the N terminus, and the individual amino acids were obtained in high percent yield (data not shown). Prior to N-terminal sequence determination, the samples were reduced with dithiothreitol, reacted with with CH2COOH and purified by high performance liquid chromatography. These procedures should convert any free cysteine-SH groups to the carboxymethyl derivative (44). The left column of Table I shows the yield of the individual amino acids from cytochrome methyl derivative (44). The right column of Table I shows the yield of the individual amino acids from cytochrome c557 as the sequencing reactions progress. It is noteworthy that the first predicted cysteine, however, the yield dropped by 10-fold, increased again for the next two amino acids (also predicted by the gene sequence), and fell again precipitously at the next predicted cysteine. The last amino acid reported is the histidine predicted to serve as the proximal ligand to the heme (7). These data demonstrate that both cysteine groups are present in pETC552 (data not shown). This is to be expected if all the heme in the sample is covalently bound to the protein.

The electrospray ion spectrum of cytochrome c552 from T. thermophilus and recombinant cytochromes c552 and c557 obtained from E. coli. A shows the mass spectrum for native cytochrome c552 in 10 mM NH4-acetate. The primary peak is at 14,771 daltons followed by a ladder of peaks resulting from the binding of 1–5 Na+ ions. B shows the mass spectrum of cytochrome c552 in 10 mM NH4-acetate. The first peak is at 14,858, and the dominant peak is at 14,888 daltons followed by a peak at 14,910 corresponding to binding of one Na+ ion. The predicted molecular mass of c552 is 14,245 + 617 = 14,862 daltons. C shows the mass spectrum of cytochrome c552 in the denaturing solvent: 30% formic acid, 60% isopropanol, 10% water. Injection of a concentrated cytochrome solution into this solvent causes its normally orange color to change to brown-green. The electrospray experiment was carried out immediately thereafter. The dominant peak in the spectrum is now at 14,874 with lesser peaks at 14,566 and 14,891 daltons. D shows the mass spectrum of cytochrome c557 in 10 mM NH4-acetate. The shape of this spectrum is unchanged on dissolution of the protein into the denaturing solvent. The primary peak is at 14,858 daltons. See text for details.

The electro spray mass spectrum from Thermus cells in aqueous solution has its primary peak at 14,771 Da followed by a ladder of peaks resulting from the binding of one to several Na+ ions (Fig. 5A). The predicted molecular mass of the molecule, based on the amino acid sequence of Titani et al. (48), after correcting for the formation of an N-terminal pyroglutamate residue (14,157 Da) and adding 617 Da for the iron-protoporphyrin IX, is 14,774 daltons. The difference between the predicted and observed values is just outside the expected error limits of ± 2 Da, but this is not pursued further here. There is no evidence for apoprotein in the mass spectrum recorded on a broader scale (not shown). Furthermore, dissolving the protein in a highly denaturing solvent prior to recording the ion spectrum (see below) reveals no apoprotein (not shown). This is to be expected if all the heme in the sample is covalently bound to the protein.

The electrospray ion spectrum of cytochrome c552 from a solution of 10 mM NH4-acetate, consists of only two major peaks at 2483(6+) and 2128(7+) and a minor peak at 1862(8+). The mass spectrum of cytochrome c552 (Fig. 5B) reveals a relatively small peak is at 14,858, a dominant peak at 14,890 Da, and a lesser peak at 14,910 Da, the latter corresponding to the binding of one Na+ ion to the 14,890-Da material. A broader reconstruction of the mass spectrum from the ion spectrum failed to reveal the presence of apoprotein (14,245

| Table I |
| N-terminal sequence determination of carboxymethylated recombinant cytochromes c552 and c557 |
| Cytochrome c552 | Cytochrome c557 |
|---------------------------------|-----------------|
| Ala 272 | 207 |
| Glu 272 | 176 |
| Ala 261 | 204 |
| Asp 248 | 193 |
| Gly 182 | 127 |
| Ala 240 | 173 |
| Lys 270 | 139 |
| Ile 182 | 125 |
| Tyr 124 | 31 |
| Ala 203 | 121 |
| Glu 159 | 94 |
| Cys# 12 | 0 |
| Ala 163 | 101 |
| Gly 95 | 64 |
| Cys# 5 | 0 |
| His 89 | 63 |

*a Carboxymethylated cysteine chromatographs and is detected at the position of serine.

*Fig. 5. Electrospray mass spectra of native cytochrome c552 from T. thermophilus and recombinant cytochromes c552 and c557 obtained from E. coli. A shows the mass spectrum for native cytochrome c552 in 10 mM NH4-acetate. The primary peak is at 14,771 daltons followed by a ladder of peaks resulting from the binding of 1–5 Na+ ions. B shows the mass spectrum of cytochrome c552 in 10 mM NH4-acetate. The first peak is at 14,858, and the dominant peak is at 14,888 daltons followed by a peak at 14,910 corresponding to binding of one Na+ ion. The predicted molecular mass of c552 is 14,245 + 617 = 14,862 daltons. C shows the mass spectrum of cytochrome c552 in the denaturing solvent: 30% formic acid, 60% isopropanol, 10% water. Injection of a concentrated cytochrome solution into this solvent causes its normally orange color to change to brown-green. The electrospray experiment was carried out immediately thereafter. The dominant peak in the spectrum is now at 14,874 with lesser peaks at 14,566 and 14,891 daltons. D shows the mass spectrum of cytochrome c557 in 10 mM NH4-acetate. The shape of this spectrum is unchanged on dissolution of the protein into the denaturing solvent. The primary peak is at 14,858 daltons. See text for details.

*From Thermus cells in aqueous solution has its primary peak at 14,771 Da followed by a ladder of peaks resulting from the binding of one to several Na+ ions (Fig. 5A). The predicted molecular mass of the molecule, based on the amino acid sequence of Titani et al. (48), after correcting for the formation of an N-terminal pyroglutamate residue (14,157 Da) and adding 617 Da for the iron-protoporphyrin IX, is 14,774 daltons. The difference between the predicted and observed values is just outside the expected error limits of ± 2 Da, but this is not pursued further here. There is no evidence for apoprotein in the mass spectrum recorded on a broader scale (not shown). Furthermore, dissolving the protein in a highly denaturing solvent prior to recording the ion spectrum (see below) reveals no apoprotein (not shown). This is to be expected if all the heme in the sample is covalently bound to the protein.

The electrospray ion spectrum of cytochrome c552 from a solution of 10 mM NH4-acetate, consists of only two major peaks at 2483(6+) and 2128(7+) and a minor peak at 1862(8+). The mass spectrum of cytochrome c552 (Fig. 5B) reveals a relatively small peak is at 14,858, a dominant peak at 14,890 Da, and a lesser peak at 14,910 Da, the latter corresponding to the binding of one Na+ ion to the 14,890-Da material. A broader reconstruction of the mass spectrum from the ion spectrum failed to reveal the presence of apoprotein (14,245...
The predicted molecular mass for \( rC_{552} \) is 14,245 + 617 = 14,862 daltons which may correspond to the relatively small peak at 14,858 Da, although it is just outside the error limits of ± 2 Da. The most likely explanation of the mass spectrum is that 1–3 water molecules remain tightly bound to this protein even after the ionization/evaporation event of the electrospray experiment. This notion is supported by the observation that when cytochrome \( rC_{552} \) is diluted into the strongly denaturing solvent (cf. Ref. 42), 30% formic acid, 60% isopropyl alcohol, 10% water, the majority peak is shifted from 14,890 in water to 14,875 Da (Fig. 5). This 15-Da decrement is assigned to the loss of a single water molecule. Interestingly, the peak at 14,890 Da remains in Fig. 5C, but at reduced intensity, suggesting that the water is quite tightly bound to the protein. In contrast to native cytochrome \( c_{552} \), dissolution in the denaturing solvent produces a small amount of material (−5%) appearing at 14,245 Da that corresponds to the predicted mass of the apoprotein. This is consistent with the finding that −4−5% of the cysteine residues in this fraction are not covalently attached to heme (Table I). A reconstruction of the mass spectrum over the range 10–50 kDa (not shown) reveals only a weak band at 29,724 Da, which is assigned to cytochrome \( rC_{557} \). As shown in Fig. 5D, the mass spectrum of cytochrome \( rC_{557} \) in aqueous solution reveals a single peak at 14,858 Da which is very close to that predicted from the gene sequence plus one iron-protoporphyrin IX (14,862 Da). However, this is not the dominant species in the mass spectrum (see below). As with the \( rC_{552} \) protein, there is no evidence for apoprotein in this sample. Similarly, while the ion spectrum (not shown) changes dramatically on dissolution of the protein into the denaturing solvent, the reconstructed spectrum in this range (not shown) is essentially the same as that obtained in aqueous solution (see below). Also, there is no evidence for formation of apoprotein upon denaturation, confirming the N-terminal sequence data, which shows that essentially all the heme in \( rC_{557} \) is covalently bound (Table I).

The monomeric species in the mass spectrum of \( rC_{557} \) (Fig. 5D) represents only a very small fraction of the protein reaching the mass detector. Instead, the protein appears to oligomerize. The molecular ion spectrum of cytochrome \( rC_{557} \) is shown in Fig. 6 and, as noted above, is very different from that of cytochrome \( rC_{552} \). Here there are 12 ions ranging from 1982.0 (15+ on the dimer) to 1144.0 (est. 26+ on the dimer) occurring in unit charge steps. A broad scale reconstruction of the mass spectrum using the program Bio-MultiView reveals peaks at \( n \times 14,858 \), where \( n \) appears to increase without bound. Thus, there is an extremely small peak at 14,858 Da (that shown in Fig. 5D), a strong peak at 29,716 Da, a much weaker peak at 44,575 Da, followed by a stronger peak at 59,433 Da, a weak peak at 74,291 Da, and so on. The molecular ion spectrum, however, supports only the presence of dimers. This can be seen, for example, by using the formula \( M_{app} = \left\{ (m/z) \times (m/z) - z \right\} \) and assuming that the ion peak at 1982.0 is due to the tetramer (\( M_{app} = 59,430 \) Da) and has 30+ charges. It is predicted then that an ion peak must occur at ~1918 having 31+ charges. No such peak is visible in Fig. 6. Similar ion peaks are expected for the tetramer mass at all odd charge values; by inspection, these are not present. Therefore, the mass spectrum is only consistent with the presence of a dimer having mass 29,716, which is twice that expected for the monomer (see above). This suggests that dimerization is the physical basis for the separation of \( rC_{557} \) from \( rC_{552} \) on the gel filtration column (Fig. 3).

To further test for the presence of multimers in aqueous solutions of \( rC_{557} \), samples were subjected to gel electrophoresis under nondenaturing conditions. Fig. 4C shows reversed polarity electrophoresis at pH 4.5 in 15% acrylamide gel (46). In the left lane, the native cytochrome \( c_{552} \) from Thermus shows essentially a single band (and a minor impurity), and the recombinant \( rC_{552} \) fraction (central lane) shows a dominant single band with electrophoretic mobility very similar to that of native protein and a weak band of lesser mobility. In the case of \( rC_{557} \) (right lane), only a weak band is observed at the mobility of native and \( rC_{552} \) proteins, but it shows a strong band with approximately the same electrophoretic mobility as the weak band in \( rC_{552} \). Were there a mixture of dimers, tetramers, etc., these would appear as a ladder of bands, which is not the case. Thus, as anticipated from the electrospray and gel filtration observations, this experiment indicates that “as isolated” \( rC_{557} \) exists predominantly as a homodimer, whereas \( rC_{552} \) exists primarily as a monomer, with the possibility of
Expression of Thermus Cytochrome c552 Gene in E. coli

The pyridine hemochrome (phc) spectra of native cytochrome c552 and recombinant cytochromes rC552 and rC557 are shown in Fig. 9, A and B. These provide additional insight into the origin of the spectral differences described above. In the Soret region, the native Thermus cytochrome c552 shows a single band at 413 nm, whereas cytochrome rC552 shows a somewhat broader band near 415 nm with a distinct shoulder near 433 nm. The cytochrome rC557 fraction shows the major peak near 416 nm, being approximately as broad toward the blue as the rC552 spectrum but having distinctly less of the 433 nm material evident toward the red. In the visible region (Fig. 9B), the phc spectrum of native Thermus cytochrome c552 shows a β-band at 522 nm and a sharp α-band at 550 nm with no evidence for a longer wavelength band. By contrast, the phc spectrum of rC552 shows bands at 522, 550, and ~576 nm, while rC557 shows maxima in the visible region at 521 and 552 nm and has a much smaller contribution near 576 nm than does the spectrum of c552. The fact that the heme C components in the phc spectrum of rC552 and rC557 are essentially identical to each other when the protein is denatured is consistent with the above conclusion that the protein in these two fractions is the same.5 The small differences in the native and recombinant phc spectra are likely due to the different N terminus, a neutral pyrogglutamate in native protein versus a glutamine and a neutral N-terminal alanine (in the strongly basic solution) in the recombinant proteins.

The absorption spectra indicate the presence of some non-heme C material in the recombinant cytochrome c preparations. As noted above, essentially all the heme in both fractions is covalently bound, as confirmed by the ability of acid-acetone extraction (63) to remove only traces of color from the rC552 protein (data not shown). Some heme could be removed from rC552 by the silver sulfate method of Paul (64) (provided the protein was first dialyzed against dilute acetic acid to remove chloride), as indicated by the fact that the phc spectrum of the resulting extract was similar to that of starting material (not shown). However, treatment of the extracted heme by the method of Grinstein (65), which simultaneously removes the iron and produces the dimethyl ester, reveals only the presence of protoporphyrin IX (as evidenced by the characteristic four-

![Figure 8](image_url)

**Fig. 8.** Inverted second derivative absorption spectra of native cytochrome c552 from T. thermophilus and recombinant proteins rC552 and rC557 from E. coli. The Soret region of oxidized (dashed line) and reduced (solid line) proteins are shown in A, while the visible region of the reduced proteins is shown in B. Data processing was as described under “Experimental Procedures.”

Some of the spectral differences are better visualized in the second derivative absorption spectra of the Soret (oxidized and reduced) and visible (reduced) regions, as shown in Fig. 8. This presentation defines the Soret absorption maxima mentioned above and illustrates the characteristic splitting of the α-band of native c552 (548.4 and 553.4 nm), first recognized by Honnami and Oshima (5),4 and shows that this distinguishing feature is absent in rC552, which exhibits only a single, second derivative absorption band at 550 nm. Interestingly, the second derivative of the absorption spectrum of cytochrome rC557 suggests a split α-band that is shifted several nanometers to the red (552.8 and 558.2 nm). The broad shoulder near 570 nm, evident in the spectra of Fig. 7, does not survive the loss in

4 The α-band (or Qα transition) results from the promotion of an electron from the vibrational ground state of the nondegenerate a g (p) Dirac orbital to the vibrational ground state of the doubly degenerate e g (p) and e u (p) orbitals of the porphyrin (76, 77). The characteristic A-type magnetic circular dichroism spectrum in this region of the spectrum is caused by the field induced splitting of e g (p) (78, 79). In addition, small structural changes, which induce structural differences along the x and y directions of the porphyrin, can also split the e u (p) orbitals, resulting in the appearance of two transitions in the α-band region, the so-called split-α spectrum seen in many cytochromes c (80). Than et al. (7) remark that the heme in their crystals of native cytochrome c552 is distorted to a “saddle shape,” which could be the origin of the observed splitting of the native c552 α-band.

5 Cytochromes c from certain protozoa (81, 82) and C14A human cytochrome c expressed in yeast (83) show α-bands near 556–558 nm in their reduced forms with pyridine hemochrome α-bands from 553–554 nm. In addition, the phc α-band is shifted to 550–551 nm on heating in the presence of hydrazine. Amino acid sequence analysis of cytochrome c557 from Crithidia oncopelti and cytochrome c594 from Euglena gracilis showed that an alanine was substituted in the place of the cysteine nearest the N terminus (81). Because heating with hydrazine has no effect on the α-band of heme C in the recombinant cytochromes being described here (J. A. Keightley, D. Sanders, T. R. Todaro, A. Pastuszyn, and J. A. Fee, unpublished results), and the data of Table I show that both cysteine residues are blocked, the unique spectral properties of rC557 are not due to failure to form a thioether linkage between the protein and the heme.

---

**Table II**

Optical absorption features (nm) of native Thermus cytochrome c552 and recombinant cytochromes rC552 and rC557 and of pyridine hemochrome

|               | Soret band | Visible region, reduced |
|---------------|------------|------------------------|
|               | Oxidized   | Reduced                |
| c552          | 409        | 417 (413)              |
| rC552         | 410        | 417 (415, 413 s)       |
| rC557         | 416        | 423 (416)              |
| rC557         | 525        | 552 (549)              |
| rC557         | 522        | 552 (551, 576)         |
| rC557         | 526        | 557 (551, 576 s)       |

4 The numbers in parentheses correspond to the position of the pyridine hemochrome (see Fig. 9).
line spectrum with peaks at 505, 538, 577, and 625 nm (65). We will defer identification of this material until it can be separated from the protein in a form suitable for chromatographic purification (Ref. 66; however, see “Discussion”).

The above results show that the cytochrome rC552 obtained from E. coli is not a “molten image” (Psalms 106:19) of native cytochrome c552. However, the data of Fig. 10 show that cytochrome rC552 is an excellent substrate for cytochrome bα3, having 85% of native c552 activity, while cytochrome rC557 is quite a poor substrate, having only 5% of native cytochrome c552 activity. Thus, when used as a substrate for Thermus cytochrome bα3, we obtain the following Michaelis-Menten parameters: native cytochrome c552 has Vmax = 70 s⁻¹ and Km = 10 μM while cytochrome rC552 has a Vmax of 60 s⁻¹ and Km = 15 μM. Soulimane et al. (2) have reported a somewhat higher value for Vmax (~250 s⁻¹) and a similar value for Km (~17 μM).

**DISCUSSION**

Cytochrome c synthesis in E. coli is a complicated process that appears to accommodate only certain foreign cytochrome c genes, and there are few examples of successful expression. While we were surprised that E. coli bearing pETC552 produced such high levels of holocytocrome c552, our enthusiasm has been tempered by the knowledge that this cytochrome is actually a complicated mixture of different forms. Identification of the structural gene for the cytochrome c in the same operon with an ABC transporter was also an unexpected finding.

We will first deal with the implications of cycB. The currently suggested function of the ABC transporters in cytochrome c maturation is to selectively transport protocen IX from the cytosol into the periplasmic space and to bring it into contact with a specific apocytochrome c, and Thöny-Meyer (11) has pointed out that the CcmB component of the ABC transporter (not yet identified in Thermus) may contain a heme binding site. Among the currently known genes involved in cytochrome c maturation, none is transcribed as part of an operon that includes the cytochrome c structural gene (11). While further work is needed to establish involvement of this ABC transporter component in Thermus cytochrome c biosynthesis, our finding of an operon that encodes both the structural gene of the cytochrome c and a component of the ABC transporter strongly suggests that the latter is involved in the maturation process.

We turn now to the cytochrome c expressed in E. coli. As reported by Titani et al. (48), the mature holocytocrome c552 isolated from T. thermophilus has a blocked N terminus, probably a pyroglutamate. The DNA sequence reveals that the cytochrome is synthesized in Thermus cells as a pre-protein
having an N-terminal signal peptide. The sequence of this signal peptide, (M)K'R'TLMAFLLGGLALA Q, is very similar in length and composition to other known pre-apolocytochrome c sequences (27, 28, 32, 67, 68), which typically have 1–2 positively charged amino acids at the N terminus followed by 14–20 hydrophobic amino acids; proteolytic cleavage occurs prior to the first hydrophilic residue (69), in this case ↓ Q. Very likely, this sequence is responsible for directing the protein into the periplasm of Thermus cells (10). The mechanism of pyroglutamate (48) formation in the mature cytochrome c<sub>552</sub> is not known, and, unfortunately, because residues 1–3 are disordered in the currently available crystals (7), it is impossible to confirm the presence of a pyroglutamyl N terminus.

The truncated cycA gene in pETC552 begins with the sequence MAQ followed by the sequence of the mature protein. In both fractions, the N-terminal Met is completely removed. As evidenced by the absence of heme-free protein in our preparations, essentially all of the cytochrome c apoprotein that is synthesized appears to react with protoheme IX. This is occurring in the cytoplasm and is thus consistent with the idea of Sambongi and Ferguson (33), that a properly folded and reduced apoprotein can react spontaneously with protoheme IX to form the thioether linkages of heme C. However, there would appear to be two folding paths, one of which results in monomeric rC<sub>552</sub> and the other dimeric rC<sub>557</sub>. In addition, part of the heme in both fractions is chemically modified (see below).

The expression system yields ~3 mg of purified cytochrome rC<sub>552</sub> and ~0.8 mg of cytochrome rC<sub>557</sub> per liter of culture medium. This level of production is adequate to obtain the quantities of material required for most modern biophysical studies. However, because there are several cytochrome c related products in the final protein solutions, the use of this material will be somewhat restricted. The majority species in fraction rC<sub>552</sub> appears to be a “native-like” cytochrome c. Because rC<sub>552</sub> lacks the “split-α” absorption band which is characteristic of the native protein in its reduced state (5), the unique structural details responsible for splitting the ε<sub>2a</sub>(π) excitation excited state (cf. Footnote 4) are evidently not present in the recombinant protein. Nevertheless, two observations, that will be published elsewhere, suggest that this material probably has an atomic structure very similar to that of native cytochrome c<sub>552</sub>. First, high resolution NMR studies reveal a pattern of paramagnetically shifted <sup>1</sup>H NMR lines similar to that previously reported for native Thermus cytochrome c<sub>552</sub> (Ref. 70). Second, single crystals of cytochrome rC<sub>552</sub> have been grown that have the same morphology and x-ray diffraction properties as those obtained with native protein.<sup>6</sup>

The most important complication is that the cytochrome rC<sub>552</sub> contains a significant amount of non-heme C chromophore. While the atomic structure of this chromophore remains to be determined, we know the following from our current data. The material is covalently attached to the protein, probably at the cysteine residues; it has a molecular mass very close to that of protoheme IX (within ±3–4 Da); it is removed from the protein under the same conditions that remove the protoheme IX, namely reaction with Ag<sup>+</sup> ions; either it does not survive the strongly acidic conditions needed to remove the iron from protoporphyrin IX, or it does not release its Fe under these conditions;<sup>5</sup> it is redox active and accounts for ~30–40% of the oxidizing equivalents in the sample while heme C accounts for the remainder<sup>10</sup>, and it reacts with hydrazine at elevated temperatures to create a new chromophore having an absorption band at 602 nm.<sup>11</sup> Unfortunately, we have not found growth conditions that suppress formation of this material. Fortunately, the unusual heme is not a serious impediment to examining the steady-state kinetic/proton translocation properties of cytochrome ba<sub>3</sub>. Thus, on a per milligram protein basis, the rC<sub>552</sub> protein is ~85% as active as native cytochrome c<sub>552</sub>, suggesting that the protein molecules to which the unusual heme is bound are at least partially active. Alternatively, the change from the neutral N-terminal pyroglutamate to the positively charged N-terminal Ala may affect the interaction with cytochrome ba<sub>3</sub>. While not perfect, the expression system, as it is currently used, will provide the substrate needed for many functional studies with cytochrome ba<sub>3</sub> that were not previously feasible.

What is the nature of the non-heme C material? Barker and co-workers (71, 72), in attempting to transmute cytochromes b to cytochromes c by the judicious placement of cysteine residues near the heme, discovered a noncovalently attached green heme in their recombinant preparations, which they recently identified as iso-spirographis heme, in which the 4-vinyl of protoheme IX is oxidized to a formyl group.<sup>12</sup> The unusual heme is thought to form, during expression in E. coli, by a heme-mediated oxidation of its own vinyl group when in close proximity to the free thiol group of the protein (71). The pyridine hemochrome spectrum of the free iso-spirographis heme has absorption maxima at 434, 538, and 582 nm (73), while the pyridine hemochrome spectrum of the unusual heme in fraction rC<sub>552</sub> shows absorption bands near 433 nm and at 576 nm. Because the electrospray spectra show essentially one protein species, the non-C heme material has a molecular mass very close to that of protoheme IX. A mono-formyl heme would have a mass only 2 Da larger than protoheme, a difference that would not be resolved in our spectra. This and the similarity of the phe spectra support a working hypothesis that the unusual chromophore is a “formyl” heme. The N-terminal sequences and mass spectral data indicate that the unusual heme is covalently attached to the protein in a manner that protects both cysteine residues from reacting with iodoacetate, perhaps as a thiohemiacetal bond. Efforts are currently underway to separate this material from the true recombinant cytochrome c<sub>552</sub> and to characterize it by a combination of resonance Raman, high resolution NMR, EPR, and other techniques.

The third interesting product of the gene expression system is the rC<sub>557</sub> fraction. All the analytical results indicate that the two proteins are identical in composition, but the red-shifted α-band in the spectrum of the reduced protein indicates that the heme C is in a significantly different environment. We hypothesize, therefore, that cytochrome rC<sub>557</sub> is a “conformational isomer” of cytochrome rC<sub>552</sub>, which has ligands to the heme iron different from those of the native protein (His<sub>14</sub> and Met<sub>69</sub>) and that this conformation has a strong propensity to dimerize.

Sosnick et al. (74) have rationalized biphasic (“fast” and “slow”) folding in horse heart cytochrome c, with the fast phase due to molecules having the heme properly coordinated in the unfolded state (His<sub>14</sub> and Met<sub>80</sub>), while the slow phase is due to molecules that have suffered “adventitious misorganization in early chain condensation” caused by non-native ligation of the heme in the unfolded state (either His<sub>28</sub> or His<sub>33</sub>). The presence of rC<sub>557</sub> in our preparations can be explained by the scheme of Sosnick et al. (74) if a portion of the molecules is improperly coordinated in the unfolded state (His<sub>16</sub> and pos-

---

<sup>7</sup> K. Bren, J. A. Fee, et al., unpublished observations.
<sup>8</sup> P. Williams, D. McRee, E. Stura, and J. A. Fee, unpublished observations.
<sup>9</sup> A. Pastuszyn and J. A. Fee, unpublished data.
<sup>10</sup> T. Todaro and J. A. Fee, unpublished data.
<sup>11</sup> J. A. Fee, unpublished results.
<sup>12</sup> G. Mauk, personal communication.
sibly His-33, Met-64, or His-87 in place of Met-70), followed by thermodynamic trapping in this state by formation of a dimer. Bryngelson et al. (75) have pointed out that misalignment during cytochrome c folding would introduce considerable roughness into the energy landscape for the folding of this protein, which otherwise appears to be quite smooth. The low activity of rC_{557} as a substrate for cytochrome ba_{3} suggests that the dimer interface involves a portion of the surface area that normally interacts with the oxidase. Once the structural coordinates are made available (7), it may be possible to deduce the structure of this dimer. A fourth complication is that the rC_{557} fraction also contains unusual heme, though by comparison with the rC_{552} fraction, at a much lower level.

In summary, thermostable cytochromes c can be expressed heterologously in E. coli cells, but the resulting product is not the desired mature, homogeneous holoprotein. Indeed, there appears to be a multiplicity of closely related products, some of which, at least, perform the expected biological function.

Acknowledgments—We thank Dr. Gary Siuzdak and his staff at TSRI for discussions. Dr. Steve Marsch at Cal Tech for sequencing the PCR product, Dr. Claire Slutter for assistance in the initial design of the PCR primers, Drs. Grant Mak and Kevin Smith for valuable advice regarding the non-heme C material, Dr. Graham Palmer for bringing the work of Sutherland and Klein (79) to our attention, and Professor Jack Richards of CalTech for his general interest in and support of this project.

REFERENCES

1. Fee, J. A., Kuila, D., Mather, M. W., and Yoshida, T. (1986) Biochim. Biophys. Acta 853, 153–185

2. Soulimane, T., Gohlke, U., Huber, R., and Buse, G. (1995) Biochim. Biophys. Acta 1237, 490–497

3. Miller, C. G. (1987) in The Porphyrins (Neidle, S. D., ed) Vol. 7, pp. 945–951, Academic Press, New York

4. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

5. von Heijne, G., and Abrahmsen, L. (1989) EMBO J. 8, 3685–3692

6. Miller, C. G. (1987) in The Porphyrins (Neidle, S. D., ed) Vol. 7, pp. 363–400, Academic Press, New York

7. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

8. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

9. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

10. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

11. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

12. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

13. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

14. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

15. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

16. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

17. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

18. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

19. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

20. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

21. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

22. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

23. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

24. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

25. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

26. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

27. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

28. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

29. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

30. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

31. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

32. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

33. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599

34. Bhattacharya, A. and Sanyal, P. P. (1988) J. Mol. Biol. 205, 591–599