Recombinant Equine Cytochrome *c* in *Escherichia coli*: High-Level Expression, Characterization, and Folding and Assembly Mutants†

Jon N. Rumbley,* Linh Hoang, and S. Walter Englander

The Johnson Research Foundation, Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6059

Received July 31, 2002; Revised Manuscript Received September 17, 2002

ABSTRACT: To promote studies of cytochrome *c* (Cyt *c*) ranging from apoptosis to protein folding, a system for facile mutagenesis and high-level expression is desirable. This work used a generally applicable strategy for improving yields of heterologously expressed protein in *Escherichia coli*. Starting with the yeast Cyt *c* plus heme lyase construct of Pollock et al. [Pollock, W. B., Rosell, F. I., Twitchett, M. B., Dumont, M. E., and Mauk, A. G. (1998) *Biochemistry 37*, 6124−6131], an *E. coli*-based system was designed that consistently produces high yields of recombinant eucaryotic (equine) Cyt *c*. Systematic changes to the ribosome binding site, plasmid sequence, *E. coli* strain, growth temperature, and growth duration increased yields from 2 to 3 mg/L to as much as 105 mg/L. Issues related to purification, fidelity of heme insertion, equilibrium stability, and introduction and analysis of mutant forms are described. As an example, variants tailored for folding studies are discussed. These remove known pH-dependent kinetic folding barriers (His26 and His33 and N-terminus), reveal an additional kinetic trap at higher pH due to some undetermined residue(s), and show how a new barrier can be placed at different points in the folding pathway in order to trap and characterize different folding intermediates. In addition, destabilizing glycine mutants in the N-terminal helix are shown to affect the fractional yield of a heme inverted Cyt *c* isoform.

Cytochrome *c* (Cyt *c*)[1] is one of the most studied proteins, widely used to address such basic issues as energy transduction, electron transfer, redox potential modulation, protein−protein and protein-membrane interactions, membrane translocation, phylogenetic relationships, apoptosis, and protein folding. These studies have suffered from the inability to produce large quantities of recombinant protein and to pursue mutational investigations and isotope-labeled NMR studies.

Molecular biological techniques to improve yields of recombinant proteins are well documented. Nonetheless, few examples exist where the techniques available have been applied systematically to increase yields in a logical progression (*1*−*3*). We have developed an efficient high-level expression system for Cyt *c* in *Escherichia coli* by a systematic series of modifications. Yields resulting from each change are reported. These same changes may be expected to improve yields of any heterologously expressed protein.

Early attempts at heterologous expression of Cyt *c* in *E. coli* resulted in disappointingly low and variable yields of no more than 2 mg/L using the apoCyt *c* maturation machinery of the host strain (*4*−*6*). Marginally higher yields (2−6 mg/L) were obtained for eucaryotic Cyt *c*'s in a yeast system (*7*−*10*). The heterologous expression of *Pseudomonas aeruginosa* cytochrome c_{551} in *E. coli* using a procaryotic periplasmic signal sequence and a strain having enhanced heme availability produced 8 mg/L (*11*).

In a major advance, the Mauk group (*12*) was able to express yeast Cyt *c* in *E. coli* by coexpressing the yeast heme lyase, obtaining up to 8 mg/L in shaker flasks and 15 mg/L in a fermentor, although in some constructs plasmid instability tended to degrade yields. Morar et al. (*13*) used this system to obtain isotopically labeled protein with yields of 8 mg/L. Sanders and Lill reported 8 mg/g of cells (wet weight) of yeast cytochrome *c* using a similar approach (*14*). In our hands, the plasmid used by Pollock et al. (*12*) with the yeast gene replaced by the horse gene produced significantly less protein (2 mg/L), and plasmid instability was a continuing problem.

Here we describe an *E. coli*-based system, evolved from the tandem expression system of Pollock et al. (*12*), that consistently produces high yields of recombinant equine Cyt *c*, standardly 60−80 mg/L and up to 105 mg/L, in high purity. Studies were done to test for purity, stability, folding behavior, and structural perturbation. A number of variants were made that affect expression level, fidelity of heme insertion, stability, and folding kinetics.

MATERIALS AND METHODS

Construction of Expression Systems. The original strain HB2151 and plasmid pBPCYC1(wt)/3 for Cyt *c* expression

† This work was supported by research grants from the NIH and the Mathers Foundation.

* Corresponding author. Tel: 215-898-4509. Fax: 215-898-2415. E-mail: jr@hxiris.med.upenn.edu.

[1] Abbreviations: Cyt *c*, cytochrome *c*; WT, wild-type horse heart cytochrome *c*; pWT, pseudo-wild-type cytochrome *c* expressed in *Escherichia coli* with His26 and His33 changed to Asn; RBS, ribosome binding site; IPTG, isopropyl β-D-thiogalactoside; PMSF, phenylmethanesulfonyl fluoride; TFA, trifluoroacetic acid; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; GdmCl, guanidinium chloride; CD, circular dichroism; *m*, Gibbs free energy dependence on denaturant concentration; $[GdmCl]_{50\%}$, midpoint of the denaturant unfolding curve; NMR, nuclear magnetic resonance; COSY, correlation spectroscopy; TSP, sodium 3-(trimethylsilyl)propionate-*2,2,3,3-d*$_4$.

10.1021/bi026543y CCC: $22.00
Published on Web 10/31/2002

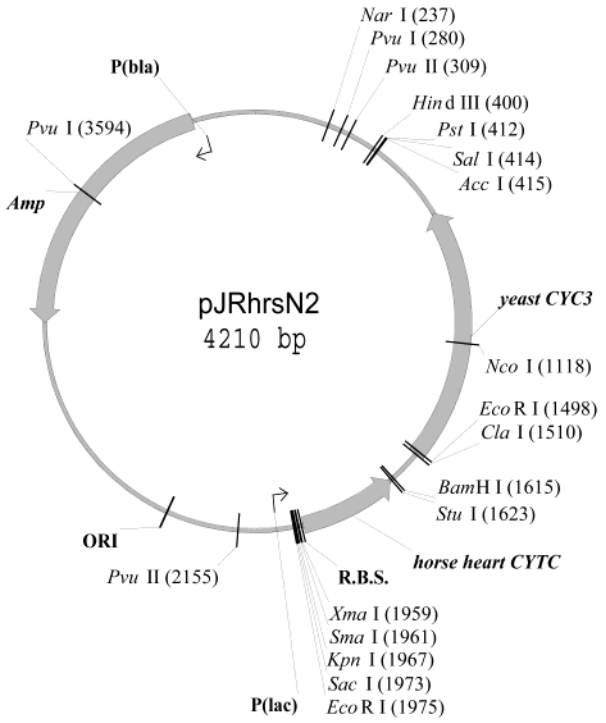


FIGURE 1: Plasmid map of the recombinant Cyt *c* expression vector pJRhrsN. The pUC-derived plasmid carries the tandem genes for horse heart Cyt *c* (CYTC) and yeast heme lyase (CYC3) and confers ampicillin resistance through the β-lactamase gene (Amp^r). Expression is driven by the *lac* promoter 5′ to the canonical ribosome binding site (RBS).

in *E. coli* were a gift from A. G. Mauk. This plasmid contains both *lac* and *trc* promoters to drive expression of the tandem genes for yeast Cyt *c* and yeast heme lyase and uses ampicillin resistance for selection. The original horse heart Cyt *c* gene, CYTC in plasmid pAB458 optimized for yeast expression, was a gift from G. McLendon.

To construct the expression plasmid, the equine gene was PCR amplified from plasmid pAB458 with primers containing necessary *Sma*I and *Bam*HI restriction sites and cloned into expression plasmid pBPCYC1(wt)/3 using a standard T4 DNA ligase ligation reaction, creating pCYChrsN. This variant, pseudo-wild type (pWT), contained point mutations His26Asn and His33Asn, previously introduced by site-directed mutagenesis. The canonical ribosome binding site (RBS) of *E. coli* was inserted 5′ to horse CYTC by PCR amplifying the gene using a 5′ oligonucleotide containing *Sma*I and the RBS (5′-CCCGGGTAAGGAGGAAACAT-AATGGGTGACGTTGAA-3′) and cloning into pBPCYC1-(wt)/3 as above, creating plasmid pJRhrsN. Plasmid pJRhrsN2 was generated by subcloning the *Sma*I/*Hin*dIII fragment of pJRhrsN (pCYChrsN + optimized RBS) into pUC18 (Figure 1). The T7 promoter-based system was generated by cloning the RBS and tandem genes into vector pET21+ (Novagen) using *Sac*I and *Hin*dIII.

A version of the horse gene with codon usage optimized for *E. coli* was designed; 69 of 105 codons were changed including four amino acid changes (V3P, K72Q, K73A, and K79R). The N-terminus was extended to include a 6-His affinity tag followed by a factor Xa protease site and glutamine. These changes were accomplished in a single synthesis by Integrated DNA Technologies, Inc., and cloned into the pUC-based pJRhrsN2 expression vector via restriction sites *Eco*RI and *Bam*HI creating pJRopti1. The complete sequence of the optimized gene can be found in the Supporting Information.

Introduction of Site-Specific Mutations. Site-directed mutagenesis was done by PCR using the method of splicing by overlap extension (*15*, *16*). Briefly, two overlapping complementary oligonucleotides were designed, mutant primer 1 (MP1) and mutant primer 2 (MP2), each containing the new mutation, along with two oligonucleotides, outside primer 1 (OP1) and outside primer 2 (OP2), 5′ and 3′ to the cytochrome *c* gene complementary to the antisense and sense strands, respectively. In a two-step process OP1 and MP1 and OP2 and MP2 were used to amplify two halves of the sequence followed by a second amplification to combine the products into the full-length gene using oligonucleotides OP1 and OP2. The full-length product was cloned into the expression plasmid using restriction sites *Sma*I and *Bam*HI. To aid the identification of mutant constructs, a silent mutation adding a unique restriction site was included in the complementary oligonucleotides, MP1 and MP2, adjacent to the mutation of interest.

Cell Growth. Small-scale cell growth was performed in LB medium. Large-scale growth (greater than 1 L) was done aerobically in 2 L shaker flasks (1 L medium) with vigorous shaking (220 rpm) using Terrific broth and 100 μg/mL ampicillin. Cell growth and Cyt *c* expression were monitored at 600 and 410 nm, respectively, following inoculation with a 25 mL overnight culture which had been pelleted and resuspended prior to addition. To test for IPTG-induced expression, 0.8 mM IPTG (induced at $A_{600nm} = 1.0$) was used.

Purification of Recombinant Proteins. For purification of expressed Cyt *c*, cells were pelleted by centrifugation, resuspended in 50 mM Tris-HCl buffer at pH 8.0 containing protease inhibitors PMSF and benzamidine, and then broken in an Aminco French pressure cell (25000 psi). Cell debris was removed by centrifugation. An ammonium sulfate precipitation (55% saturation; 351 g/L) removed a number of host-specific proteins. Following centrifugation, the supernatant salt concentration was reduced by dialysis overnight against 10 mM potassium phosphate, pH 7.5. Cyt *c* was oxidized with 100 μM potassium ferricyanide, loaded onto a CM-Sepharose fast-flow column (Amersham Pharmacia Biotech) equilibrated with 25 mM potassium phosphate at pH 7.5, and gradient eluted using a 40−250 mM NaCl gradient in the same buffer. When necessary (determined from the baseline of denaturant melting curves; see below), Cyt *c* was further purified by reverse-phase HPLC (2.2 cm × 25 cm Vydac C18 column, 30−55% gradient of 90% acetonitrile in 0.1% TFA at pH 2.0, 1.4 mL/min). Acetonitrile was removed by dilution and filtration (Amicon concentrator).

Purification of the Codon-Optimized Cyt c. The codon-optimized HrsOpti-1 protein was grown, and cells were lysed as above. Na_2SO_4 was added to a final concentration of 1 M, and cell debris and protein precipitate were removed by centrifugation. The supernate was combined with a slurry of Ni^{2+}-NTA agarose (Qiagen), 1 mL of slurry to 4−5 mg of Cyt *c*, and was allowed to bind with gentle agitation for 1 h at 4 °C. The slurry was poured into a column preequilibrated with 25 mM Tris, 100 mM NaCl, and 10 mM $CaCl_2$, pH 8.0 (reaction buffer).

Table 1: Modifications for High-Level Cyt *c* Expression

strain and plasmid	growth[a]	modification	final yield (mg/L)[b]
HB2151 pBPCYC1(wt)/3[c]	37 °C, 36−48 h	coexpress yeast CYC1 with yeast heme lyase	7−8[c]
HB2151 or NM522 pCYChrsN	37 °C, 26 h	replace yeast CYC1 with horse CYTC	2−3
HB2151 or NM522 pJRhrsN	37 °C, 26 h	canonical RBS upstream of CYTC	8−10
NM522 pJRhrsN2	37 °C, 26 h	tandem genes for Cyt *c* and heme lyase in pUC18	19−21
NM522 pJRhrsN2	30 °C, 26 h	decrease growth temperature	26−30
NM522 pJRhrsN2	30 °C, 50 h	increase growth time	28−32
BL21(DE3) pJRhrsN2	37 °C, 50 h	in BL21(DE3), lacking lon and ompT proteases	30−34
BL21(DE3) pJRhrsN2	30 °C, 50 h	decrease growth temperature	60−65
BL21(DE3) pJRhrsNE4G	30 °C, 50 h	pWT with Glu4 to Gly mutation	100−105
BL21(DE3) pJRhrsNK8G	30 °C, 50 h	pWT with Lys8 to Gly mutation	70−80
BL21(DE3) pJRhrsNV11G	30 °C, 50 h	pWT with Val11 to Gly mutation	20−25
BL21(DE3) pJRT7hrsN	30 °C, 50 h	tandem genes for Cyt *c* and heme lyase in pET21+	18−20
BL21(DE3) pJRT7hrsN	30 °C, 50 h	IPTG induced at $A_{600nm} = 1.0$	45−50
BL21(DE3) pJRopti1	30 °C, 50 h	codon-optimized Cyt *c* in pJRhrsN2	30−35[d]

[a] In a platform shaker at 200 rpm after inoculation with a 25 mL overnight culture into Terrific broth. [b] All yields reported after purification to >95% purity (SDS−PAGE). [c] From Pollock et al. (*12*). [d] The amount of Cyt *c* in the cell extract was similar to that of pJRhrsN2; losses likely occurred during batch processing and trypsin proteolysis.

The factor Xa site introduced to remove the 6-His affinity tag was not accessible to factor Xa under any conditions or incubation times tested. Limited trypsin proteolysis was used to remove the His tag and elute the protein. Trypsin was added at 1:50 (w/w) [~1:100 (mol/mol)] and incubated at 25 °C with gentle stirring for 15 min. The liberated Cyt *c* was eluted with 3 volumes of reaction buffer into 90% acetonitrile and 10% of 0.1% TFA (buffer B) to give a final concentration of 25% buffer B. Trypsin activity was quenched by reducing the pH to 2.3. The trypsin digestion step was repeated four times, resulting in about 70% recovery.

Samples were pooled, concentrated, loaded on a Vydac C18 column preequilibrated with 25% buffer B and 75% buffer A (0.1% TFA, pH 2.05), and eluted isocratically (36% buffer B). Fractions containing Cyt *c* were pooled and dialyzed against 25 mM KH_2PO_4, pH 7.0, to remove the acetonitrile. The N-terminal glutamine was cyclized by raising the KH_2PO_4 concentration to 0.5 M and heating at 50 °C for 5 h (*17*, *18*). Cyclization was verified by MALDI-TOF mass spectrometry. The absence of multiply cleaved fragments was also verified by SDS−PAGE and mass spectrometry. No fragmented Cyt *c* appeared in the final sample.

Protein Purity and Stability. Cyt *c* purity was established by SDS−PAGE analysis on 16% Tris-Tricine gels, by mass spectrometry, and by reverse-phase C18 chromatography. Protein stability was determined by denaturant melting analysis using guanidinium chloride (GdmCl), measured by circular dichroism (CD) at 222 nm and fluorescence (Aviv Model 202CD with autotitration) using 5−10 μM Cyt *c* in 25 mM potassium acetate and 500 mM KCl, pH 5.0. Melting midpoints were determined, analytically solving for $[GdmCl]_{50\%}$ (*19*, *20*). Stability at zero denaturant was obtained by multiplying the midpoint GdmCl concentration by the *m* value of 4.5 kcal mol^{-1} M^{-1} established previously from native state hydrogen exchange measurements (*21*, *22*). This is the preferred method of ΔG determination due to the sensitivity of classical Santoro−Bolen fits to the pretransition baseline slope, known to occur in heterologously expressed horse Cyt *c*. Also, the significant population of intermediate forms in the melting transition region (*22*) leads to an underestimate of the global *m* value and therefore in ΔG, which is more pronounced in some mutant forms. The midpoint determination is insensitive to these issues. The presence of Cyt *c* protein with some unspecified damage that reduces stability (thought to be misinserted heme although still covalently bound) was revealed by a premelting baseline slope. Preparations having more than 5% signal change in the baseline were purified further by C18 chromatography.

Protein Folding Kinetics. Folding kinetics were measured by absorbance at 695 nm with a Biologic SFM4 stopped-flow spectrometer. This Met80 sulfur to heme charge transfer band registers final native state acquisition. The initially denatured protein (4.2 M GdmCl and 100 mM potassium acetate at pH 5.0 or 100 mM potassium phosphate at pH 7.0) was diluted 1:10 into folding solvent (same salts without GdmCl). Kinetic folding measurements were done at 10 °C using 150 μM protein (final concentration).

NMR Measurements. COSY spectra were collected on a Varian Inova 500 MHz NMR spectrometer using 5−8 mM Cyt *c* in 100 mM K_2HPO_4, pH 7.0, 10% D_2O, and 1 mM TSP. Spectra were processed on a SGI Indigo2 workstation using Felix 2.30 software (Molecular Simulations).

RESULTS

Optimization of Yield. Manipulations described here are listed in Table 1. Figure 1 shows a plasmid map of the high-yielding construct pJRhrsN2.

Pollock et al. (*12*) showed that yeast holoCyt *c* could be expressed in *E. coli* (15 mg/L in a fermentor) when the yeast heme lyase is expressed concurrently. We used standard cloning procedures to replace the yeast gene (CYC1), located 5′ to the yeast heme lyase, with the equine gene (CYTC). To promote subsequent protein folding studies, the His26 and His33 codons in WT Cyt *c* were replaced with Asn because histidine residues in the neutral form introduce a kinetic barrier into the folding pathway. We refer to the His to Asn substituted protein as pseudo-wild-type (pWT) Cyt *c*. This construct (pCYChrsN) yielded only 2−3 mg/L of Cyt *c*, and one in four flasks failed to express altogether. Systematically lower expression of horse heart Cyt *c* in yeast has been noted in the literature before but has not been explained.

Recombinant yield can be improved by using a ribosome binding site complementary to the *E. coli* 16S rRNA Shine−

Dalgarno sequence followed by an A/T-rich sequence of seven or eight nucleotides before the start codon (*23*, *24*). Introduction of the canonical *E. coli* ribosome binding site (TAAGGAGG) and the A/T-rich sequence (AAACATA) into vector pCYChrsN (original sequence, AGGAAACAGACC), creating pJRhrsN, improved expression of pWT Cyt *c* about 4-fold in good cases. However, yields still varied significantly between preparations, and within flasks of a given preparation, and one in four flasks showed little or no expression.

Restriction analysis of plasmid from flasks failing to express significant amounts of Cyt *c* and inspection of the original vector suggested that the two promoters, *trc* and *lac*, used by the Mauk group (*12*) to increase synthesis, create repeat sequences that allow homologous recombination and expulsion of the promoters 5′ to CYTC. The common *recA*$^-$ strains, JM109 and HB101, did not solve the problem and in fact failed to reach high cell densities or produce appreciable amounts of Cyt *c*. To eliminate the redundant sequences, we cloned the tandem genes for equine Cyt *c* and yeast heme lyase into the pUC18 vector (using *Sma*I and *Hin*dIII) under the control of the *lac* promoter while retaining the modified ribosome binding site (pJRhrsN2, Figure 1).

Expression in NM522 reached a plateau by 26 h and consistently yielded 20 mg/L of Cyt *c* at 37 °C, with no plasmid instability. Growth temperature and duration were important variables and dramatically affected pJRhrsN2 expression in BL21(DE3). In BL21(DE3) basal level expression plateaued at 48 h and reached a yield of 30 mg/L at 37 °C and 60−65 mg/L (purified) at 30 °C. These levels persisted in culture for more than 72 h. The increase in yield may be due to the deficiency of the *lon* and *ompT* proteases in the BL21 cells. IPTG induction failed to increase expression in any strain. Table 1 shows that yields from some variants were significantly higher. The yield in minimal media reached 6 mg/L. A T7-based system was constructed by cloning the RBS and two genes into pET21+ (using *Sac*I and *Hin*dIII), creating pJRT7hrsN for expression in BL21-(DE3) cells. No improvement was observed in rich media. Unlike the *lac* promoter-based plasmid, IPTG was effective in inducing expression from this more tightly regulated plasmid. This is the likely explanation for a doubling of yield found in minimal media (12 mg/L purified).

A codon-optimized horse Cyt *c* gene (HrsOpti-1) based on *E. coli* codon utilization bias was constructed by total gene synthesis with the expectation that it would provide higher yields. The optimized gene was cloned into pJRhrsN2 between *Eco*RI and *Bam*HI where the second *Eco*RI site had been previously removed [Figure 1, *Eco*RI (1498)], creating pJRopti1. NM522 and BL21(DE3) cells containing this plasmid reached the same yield obtained from pJRhrsN2. IPTG induction failed to increase the yield. Evidently, codon availability is not limiting. Nonetheless, the HrsOpti-1 construct offers other advantages (see below).

Appearance of a Contaminating Species. In most cases ammonium sulfate fractionation and CM-Sepharose chromatography yielded Cyt *c* protein of greater than 95% purity by SDS−PAGE. Mutations near the site of heme attachment led to a specific contaminant. Figure 2 shows denaturant melting curves for WT Cyt *c*, pWT, and the mutants E4G, K8G, and V11G which approach the heme attachment site

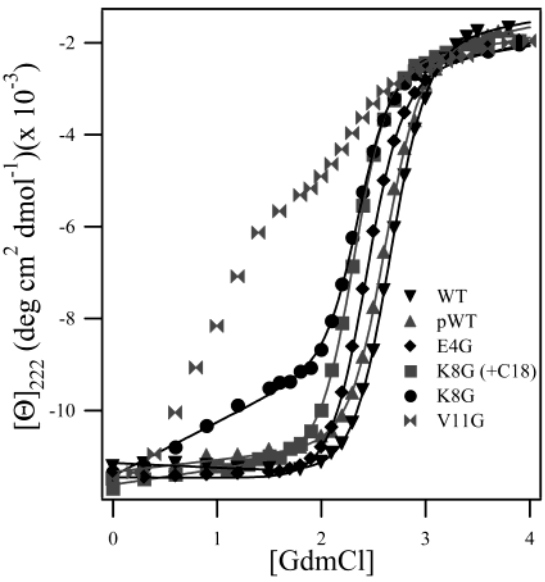


FIGURE 2: Equilibrium denaturation of pWT Cyt *c* and mutants in the N-terminal helix. Helical content was measured by CD at 222 nm in the presence of 100 mM potassium acetate and 500 mM potassium chloride, pH 5.0 and 25 °C. Varying amounts of a contaminating species can be seen in the pretransition baselines. K8G melting curves before and after removal of the contaminant (inverted heme) by C18 reverse-phase chromatography are shown.

(Cys14 and Cys17). The sloping pretransition baseline in the melting curve is due to the copurification of a less stable contaminant that is heme containing (spectrum), identical in size (mass spectrometry), and has similar charge (native PAGE and ion-exchange chromatography) to pWT Cyt *c*. The contaminant level varies among different mutations and preparations. For K8G the pretransition slope accounts for 25% of the CD signal. The mutant V11G, nearer the heme attachment site, contained 70% contaminant. Other variants more distant from Cys14 and Cys17 reproducibly showed only 0−5% contaminant. Hydroxyapatite, gel filtration chromatography, and molecular weight cutoff filters, with or without premelting concentrations of GdmCl, did not separate the contaminating species. SDS−PAGE and mass spectrometry of even the worst cases yielded a single molecular weight species coincident with Cyt *c*. This would be expected if the contaminating species itself were a modified Cyt *c* resulting from its overexpression. N-Terminal mutants are specifically affected, suggesting a relationship to heme insertion by proximity.

Preparations showing significant (>5%) CD changes in the premelting baseline were further purified by reverse-phase C18 chromatography (see Materials and Methods), which was able to separate the two isoforms. Removal of the contaminating species, which appears as a leading shoulder on the main Cyt *c* peak (Figure 3) eliminates the pretransition baseline, as shown for K8G in Figure 2.

Recombinant Structure and Stability. The pWT construct differs from WT at three positions. His26 and His33 were changed to asparagines. The N-terminus, which is normally blocked by N-acetylation in the mammalian WT protein, exists as a free amino group in the recombinant protein since the N-terminal acetyltransferase enzyme is absent in *E. coli*.

NMR COSY spectra show that pWT and WT structures are very similar (Figure 4a). Many chemical shifts are identical. Observed changes in chemical shift are understand-

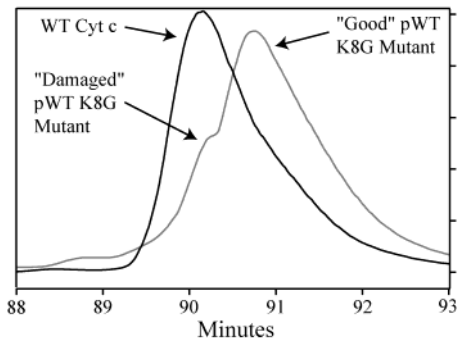


FIGURE 3: C18 reverse-phase chromatographic purification of K8G Cyt *c*. Prior to loading the pH was lowered to 2.0 and acetonitrile added to 25%. Protein was loaded at the 10 min mark followed immediately by an elution gradient of 30−55% of 90% acetonitrile in 0.1% TFA, pH 2.0, at a flow rate of 1.4 mL/min. A fraction size of 1 mL allowed for adequate separation of the contaminating misisomerized heme species (shoulder). No such contaminating species is observed in WT samples.

able in terms of the residue changes made or proximity to the free N-terminus (see Discussion). Other shifts occur due to the loss of hydrogen bonds from His26 in WT Cyt *c* to Asn31 and Pro44.

The global stability of pWT Cyt *c* (GdmCl denaturation measured by CD and tryptophan fluorescence) is lower than the wild-type protein by 0.25 kcal/mol at pH 5.0 (Figure 2) and 1.2 kcal/mol at pH 7.5 [using the melting midpoint and an *m* value of 4.5 kcal mol^{-1} M^{-1} previously determined from native state hydrogen-exchange measurements (see Materials and Methods)].

Kinetic Folding. Figure 5 compares the kinetic folding of WT and pWT Cyt *c* when the protein in high denaturant is diluted into native conditions. At pH 5.0, WT and pWT fold rapidly (10 ms) in a two-state manner (not shown). At pH 7.0, WT folding is slower and three state because its histidine residues compete with the Met80 sulfur as alternative axial heme ligands at elevated pH (*25*−*30*), inserting a kinetic barrier that slows folding to 520 ms (Figure 5). In pWT, the His to Asn mutations remove the His to heme misligation-dependent barrier, but a new barrier due to misligation by the free N-terminus is then revealed, which slows folding to 100 ms. This is also the case for yeast iso-1-Cyt *c* (*31*). The pK_a values for misligation to the heme under unfolding conditions (4.5 M GdmCl) are 5.6 (*27*, *28*, *32*) for the histidines in WT and 6.6 for the free N-terminal amino group in pWT. Clearly, pWT has the desired quality of a greater dynamic range for fast error-free folding experiments but is not ideal for future barrier insertion mutational analysis.

Removing Kinetic Barriers and Further Optimization. To eliminate pH-dependent folding kinetics altogether, we designed a protein that retains the His to Asn mutations of pWT but also blocks the N-terminus, called HrsOpti-1. HrsOpti-1 was constructed by total gene synthesis and incorporated the following changes. (1) To simplify cloning of new mutations and facilitate the combining of preexisting mutations, unique restriction sites were placed every 40−60 base pairs when possible (without changing the amino acid sequence or introducing poor codon usage). (2) To facilitate purification, a 6-His affinity tag was added to the N-terminus, followed by a factor Xa protease sequence to enable His-tag removal during purification. (3) To produce a blocked N-terminus, a glutamine residue was placed directly following the protease sequence and before the normal N-terminal glycine residue. N-Terminal glutamine spontaneously cyclizes to form pyroglutamate (*17*, *18*), blocking the N-terminus to heme iron interaction. A modified

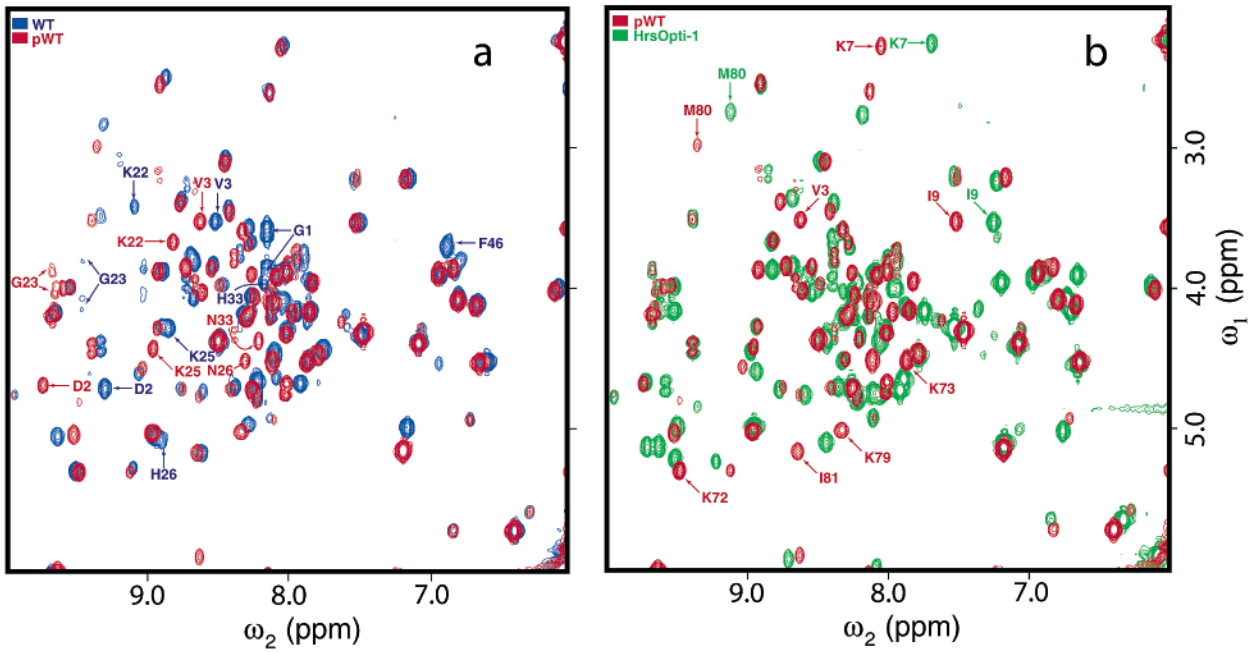


FIGURE 4: $^1H-^1H$ COSY spectra of oxidized WT, pWT, and HrsOpti-1 Cyt *c*. Spectra were recorded in 50 mM potassium phosphate buffer (pH 7.0) with 10% D_2O at concentrations between 2 and 4 mM Cyt *c*. (a) Overlay of WT and pWT showing chemical shift changes specific to changes to His26 (Asn in pWT) and His33 (Asn in pWT) as well as the lack of acetylation of the free N-terminus in pWT. Farther reaching changes are likely related to the loss of His26 side chain H-bonds to Asn31 and Pro44 across two ω loops. (b) Overlay of pWT and HrsOpti-1 spectra showing the expected chemical shift changes due to V3P, K72Q, K73A, and K79R as well as a number of additional small shifts. Residues 72 and 73 exist in a highly conserved PKKY sequence in Cyt *c* in an ω loop extending from residue 70 to residue 85, which may contribute to the numerous small differences.

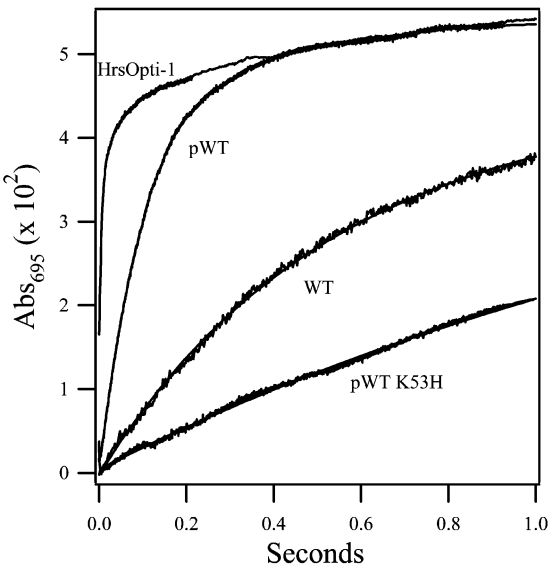


FIGURE 5: Kinetic folding at pH 7.0, 10 °C. To initiate folding, denatured Cyt *c* in 4.2 M GdmCl was diluted to 0.7 M GdmCl and 150 μM protein with 100 mM potassium phosphate and 500 mM potassium chloride. Final native state acquisition was measured by absorbance at 695 nm. Kinetic traps described in the text slow folding of WT (His26 and His33; 520 ms), pWT (N-terminus; 100 ms), and K53H (His53; 3.1 s) compared to HrsOpti-1 (no blocking; 8 ms). Similar measurements at pH 5.0 resulted in a folding rate of 8−10 ms for all variants (not shown).

gene without the 6-His tag, factor Xa site, and terminal Gln was also made as a control (HrsOpti-2). (4) The normally occurring Val3 in WT was replaced with Pro in order to increase global stability, as suggested by helical peptide studies and statistical analysis of known structures (for review, see ref *33*), and especially to selectively enhance the stability of the most stable folding unit in Cyt *c*, involving the N- and C-terminal helices (*21*). (5) The normally occurring lysines at positions 72, 73, and 79, which are implicated as major contributors to the Cyt *c* alkaline transition, were changed to Gln, Ala, and Arg, respectively. These changes were suggested from alignment of 100 sequences retrieved by Blast search (*34*). NMR COSY spectral comparison with pWT is shown in Figure 4b. The changes (V3P, K72Q, K73A, and K79R) lead to local chemical shift changes; other small chemical shift changes appear throughout the protein. This may be related to the changes made to the nearly fully conserved PKKY (to PQAY) sequence. Nonetheless, the structure is largely the same as WT and pWT Cyt *c*.

Stability of HrsOpti-1. HrsOpti-1 is 0.68 kcal/mol more stable than the WT protein and 0.93 kcal/mol more stable than pWT (pH 5.0). Contributions to the increase in stability may arise from blocking N-terminal to heme-iron interaction in the unfolded state, helix stabilization by Pro3 and the blocked N-terminus, and the removal of positive charges at positions 72 and 73, which we observe to be slightly destabilizing in the pWT protein (unpublished results).

Folding Kinetics of HrsOpti-1. HrsOpti-1 with the blocked N-terminus folds rapidly at both pH 5.0 and pH 7.0 (8 ms) due to the elimination of both the histidine misligation barrier identified in WT Cyt *c* and the free N-terminal block seen in pWT. Figure 5 shows folding data at pH 7.0. The N-terminal pyroglutamate further extends the pH-insensitive range for folding experiments up to 7.0, which will be useful for subsequent mutagenesis studies. Nonetheless, it is interesting that HrsOpti-1 under unfolding conditions still displays a heme liganding interaction with pK_a 7.7, suggesting misligation to the heme iron by tyrosine or other lysine residues. Hammack et al. (*31*) made a similar observation in yeast iso-1-Cyt *c*.

Inserting a New Kinetic Barrier and Blocking the Folding Pathway. By introducing histidines into pWT, His to heme misligation will cause intermediates to populate at specific (designed) steps in the folding pathway. In the context of pJRhrsN2, Lys53 was changed to His. Like WT Cyt *c*, mutant K53H is unable to fold rapidly to the native state when histidine is in the neutral form (3.1 s; Figure 5). Fast folding could be restored upon histidine protonation (10 ms, pH <5.5; not shown). This shows that misligation is not specific to His26 and His33 and the ω loop that contains them (residues 20−36). Lys53 is located in a large ω loop (residues 40−58) that shields the heme propionates from bulk solvent and has been shown to be an independent folding unit on the folding pathway (*21*, *35*, *36*).

DISCUSSION

Factors Limiting Expression. The methods applied here to improve yields of heterologous Cyt *c* expression in *E. coli*, strain selection, adjustment of growth conditions, promoter selection, ribosome binding site optimization, and removal of sequence instability, all contributed to successful expression of high levels of recombinant protein, as summarized in Table 1, and are generally applicable to any gene of interest.

Cyt *c* yield dropped from 15 to 2 mg/L when yeast CYC1 was replaced by the horse sequence in pCYChrsN. Assuming that expression was limited by codon usage, we made codon-optimized HrsOpti-1. The lack of improvement in yield over construct pJRhrsN2 suggests that transcription and translation are not limiting factors. This is supported by the observations that yield did not improve with the use of a much stronger promoter (T7) and that IPTG induction did not raise expression above basal levels.

A plausible explanation is that heme synthesis and/or attachment become(s) limiting. A 10-fold increased level of *Hydrogenomonas thermophilus* Cyt c_{552} was observed when expressed in an *E. coli* strain that produces elevated amounts of heme (*37*). Sanders and Lill concluded that their improved recombinant yeast Cyt *c* yields were due to increased soluble heme lyase following the addition of a histidine tag, which eliminated the accumulation of heme lyase in inclusion bodies (*14*). In addition, we observed mutation-dependent variations that appear to be related to heme lyase recognition and fidelity.

Misincorporation of Heme. The recombinant Cyt *c* was observed to copurify with variable amounts of a contaminating species, detected indirectly in denaturation experiments. It could only be separated by carefully adjusted C18 HPLC chromatography. Its properties are consistent with a subpopulation with heme inversion as found for recombinant *Thermus thermophilus* Cyt c_{552} by McRee et al. (*38*). This form of the protein cannot interconvert to form native Cyt *c* due to the covalent nature of the incorrectly inserted heme. The cooperativity of its melting transition indicates a folded structure. The stability of the heme-inverted forms can be

estimated from an enriched fraction of the K8G isoform (not shown) and from the V11G melting curve in Figure 2, 6.2 and 4.1 kcal/mol, respectively. Therefore, destabilization due to heme inversion in the context of either K8G or V11G is 4.4 and 4.2 kcal/mol, respectively. When glycines were incorporated at positions nearer to the heme insertion site (Cys14 and Cys17), progressively greater amounts of heme-inverted protein were observed (0−5% in pWT and E4G, 20−30% in K8G, 70% in V11G). Because glycines are helix destabilizing, these observations may suggest that N-terminal helix formation is required for heme lyase fidelity. It is known that apoCyt *c* acquires significant helical content on association with membranes (*39*, *40*), which may contribute to heme lyase recognition and heme insertion in vivo.

Nature of Structural and Stability Differences. The recombinant and WT cytochromes show some revealing differences.

In the recombinant pWT protein, the N-terminal methionine is removed, but the normal terminal glycine remains in a nonacetylated condition (seen by mass spectrometry). This produces some local NMR chemical shift changes affecting immediately adjacent amide hydrogens and, to a lesser extent, those further into the helix, suggesting some fraying at the N-terminus. Changes in chemical shift were seen for a limited number of residues immediately adjacent to the H26N and H33N mutation sites and for residues affected by the loss of the hydrogen bonds between the His26 side chain and the Asn31 backbone NH and Pro44 backbone CO. The hydrogen bonds bridge two ω loops in WT Cyt *c*. This situation provides the most likely explanation for the pH-dependent disparity in stability ($\Delta\Delta G_{WT-pWT}$ goes from 1.20 kcal/mol at pH 7.5 to 0.25 kcal/mol at pH 5.0) since this bond in WT is weakened at low pH. In addition, loss of N-capping by the *N*-acetyl group (*41*, *42*) and the ability of the free amino group to misligate to the heme iron in the unfolded state may also contribute.

Recombinant HrsOpti-1 is more stable than WT by 0.68 kcal/mol at pH 5.0 even though the NMR shows more distributed chemical shift changes (Figure 4b). In HrsOpti-1, lysines 72, 73, and 79 were changed to Gln, Ala, and Arg, respectively, to suppress the alkaline transition.

Removing and Inserting Kinetic Folding Traps. Cyt *c* folding is inherently fast (10 ms in 0.7 M GdmCl at pH 5.0), but kinetic traps due to heme misligation by non-native ligands can slow folding. The pH titration of denatured WT Cyt *c* (in 4.2 M GdmCl) reveals a His to heme misligation with pK_a 5.6 (His26 or His33). At higher pH, misligation imposes a kinetic trap that slows folding by 50-fold (520 ms). In pWT Cyt *c* with His26 and His33 replaced by Asn, folding is faster. However, pH titration of denatured pWT shows yet another heme misligation, due to the unblocked N-terminus, with pK_a 6.6 that slows folding by 10-fold at higher pH (100 ms).

To remove these kinetic traps, we designed HrsOpti-1 with no exogenous histidines and a blocked N-terminus. Introduction of glutamine at the N-terminus of HrsOpti-1 leads to the formation of pyroglutamate (cyclic glutamate), eliminating its interaction with the heme iron in the unfolded state. As expected, kinetic misfolding was absent. Folding was fast and independent of pH (8 ms at pH 5.0 or 7.0). Evidently, the free N-terminus causes the 10-fold reduction in the folding rate of pWT. In support, HisOpti-2, which has all of the same changes as HrsOpti-1 except for the blocked N-terminus, again folds in 90 ms at pH 7.0.

Unexpectedly, yet another trapped species becomes populated at pH >7.0. In unfolded HrsOpti-1 pH titration reveals a ligand to heme pK_a of 7.7, apparently due to misligation by a tyrosine or to lysines other than the lysines usually suggested to be involved in the alkaline transition, namely, Lys72, Lys73, and Lys79, which are absent in this construct. Nevertheless, the changes made to pWT do considerably extend the useful range of pH-independent folding by eliminating trapped intermediates due to heme misligation by histidines (pK_a 5.6) and the free N-terminus (pK_a 6.6).

Finally, it is noteworthy that the recombinant Cyt *c* system provides the unique opportunity for reimposing a new kinetic barrier at any point in the folding pathway by placing individual histidine mutations at any position in the protein. Initial results for the Lys53His mutant (Figure 5) show a 310-fold slowing of the folding rate. The block in the pathway results from a misfolding necessary to ligate the newly inserted His53 to the heme iron. To proceed on the folding pathway, the polypeptide chain must first remove the misfold and then try again to fold correctly. The particular intermediate that accumulates at that point in the pathway may then be identified by hydrogen-exchange pulse labeling (*43*).

NOTE ADDED IN PROOF

A previous description of horse heart Cyt *c* expression in *E. coli* was presented by Patel et al. (*44*). Their expression system yielded 15 mg/L pure Cyt c.

ACKNOWLEDGMENT

We thank Drs. W. Brent R. Pollock and A. Grant Mauk for providing the initial Cyt *c* expression system and for valuable discussion and Dr. G. McLendon for providing the horse CYC1 gene.

SUPPORTING INFORMATION AVAILABLE

A graphical representation of the synthesized HrsOpti-1 gene including RBS and restriction endonuclease sites as well as the gene sequence with the translated Cyt *c* amino acid sequence and associated restriction sites is available free of charge via the Internet at http://pubs.acs.org.

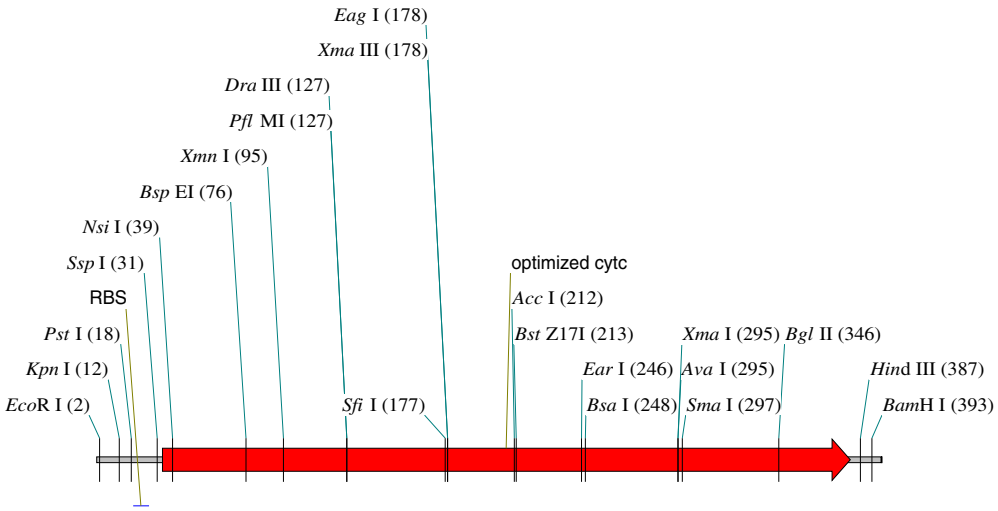
REFERENCES

1. Murby, M., Uhlen, M., and Stahl, S. (1996) *Protein Expression Purif. 7*, 129−136.
2. Ejdeback, M., Young, S., Samuelsson, A., and Karlsson, B. G. (1997) *Protein Expression Purif. 11*, 17−25.
3. Nygaard, F. B., and Harlow, K. W. (2001) *Protein Expression Purif. 21*, 500−509.
4. McEwan, A. G., Kaplan, S., and Donohue, T. J. (1989) *FEMS Microbiol. Lett. 50*, 253−258.
5. Ubbink, M., Van Beeumen, J., and Canters, G. W. (1992) *J. Bacteriol. 174*, 3707−3714.
6. Sambongi, Y., and Ferguson, S. J. (1994) *FEBS Lett. 340*, 65−70.
7. Scarpulla, R. C., and Nye, S. H. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6352−6356.
8. Clements, J. M., O'Connell, L. I., Tsunasawa, S., and Sherman, F. (1989) *Gene 83*, 1−14.
9. Hickey, D. R., Jayaraman, K., Goodhue, C. T., Shah, J., Fingar, S. A., Clements, J. M., Hosokawa, Y., Tsunasawa, S., and Sherman, F. (1991) *Gene 105*, 73−81.

10. Koshy, T. I., Luntz, T. L., Garber, E. A., and Margoliash, E. (1992) *Protein Expression Purif. 3*, 441−452.
11. Hasegawa, J., Shimahara, H., Mizutani, M., Uchiyama, S., Arai, H., Ishii, M., Kobayashi, Y., Ferguson, S. J., Sambongi, Y., and Igarashi, Y. (1999) *J. Biol. Chem. 274*, 37533−37537.
12. Pollock, W. B., Rosell, F. I., Twitchett, M. B., Dumont, M. E., and Mauk, A. G. (1998) *Biochemistry 37*, 6124−6131.
13. Morar, A. S., Kakouras, D., Young, G. B., Boyd, J., and Pielak, G. J. (1999) *J. Biol. Inorg. Chem. 4*, 220−222.
14. Sanders, C., and Lill, H. (2000) *Biochim. Biophys. Acta 1459*, 131−138.
15. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989) *Gene 77*, 51−59.
16. Horton, R. M., Cai, Z. L., Ho, S. N., and Pease, L. R. (1990) *BioTechniques 8*, 528−535.
17. Blomback, B. (1967) *Methods Enzymol. 11*, 398−411.
18. Khandke, K. M., Fairwell, T., Chait, B. T., and Manjula, B. N. (1989) *Int. J. Pept. Protein Res. 34*, 118−123.
19. Clarke, J., and Fersht, A. R. (1993) *Biochemistry 32*, 4322−4329.
20. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry 27*, 8063−8068.
21. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science 269*, 192−197.
22. Mayne, L., and Englander, S. W. (2000) *Protein Sci. 9*, 1873−1877.
23. Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) *Nucleic Acids Res. 22*, 4953−4957.
24. Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., and Gold, L. (1992) *Mol. Microbiol. 6*, 1219−1229.
25. Babul, J., and Stellwagen, E. (1972) *Biochemistry 11*, 1195−1200.
26. Brems, D. N., and Stellwagen, E. (1983) *J. Biol. Chem. 258*, 3655−3660.
27. Elove, G. A., Bhuyan, A. K., and Roder, H. (1994) *Biochemistry 33*, 6925−6935.
28. Sosnick, T. R., Mayne, L., Hiller, R., and Englander, S. W. (1994) *Nat. Struct. Biol. 1*, 149−156.
29. Yeh, S. R., Takahashi, S., Fan, B., and Rousseau, D. L. (1997) *Nat. Struct. Biol. 4*, 51−56.
30. Pierce, M. M., and Nall, B. T. (1997) *Protein Sci. 6*, 618−627.
31. Hammack, B., Godbole, S., and Bowler, B. E. (1998) *J. Mol. Biol. 275*, 719−724.
32. Colon, W., Wakem, L. P., Sherman, F., and Roder, H. (1997) *Biochemistry 36*, 12535−12541.
33. Aurora, R., and Rose, G. D. (1998) *Protein Sci. 7*, 21−38.
34. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol. 215*, 403−410.
35. Xu, Y., Mayne, L., and Englander, S. W. (1998) *Nat. Struct. Biol. 5*, 774−778.
36. Rumbley, J., Hoang, L., Mayne, L., and Englander, S. W. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 105−112.
37. Sinha, N., and Ferguson, S. J. (1998) *FEMS Microbiol. Lett. 161*, 1−6.
38. McRee, D. E., Williams, P. A., Sridhar, V., Pastuszyn, A., Bren, K. L., Patel, K. M., Chen, Y., Todaro, T. R., Sanders, D., Luna, E., and Fee, J. A. (2001) *J. Biol. Chem. 276*, 6537−6544.
39. Bryson, E. A., Rankin, S. E., Carey, M., Watts, A., and Pinheiro, T. J. (1999) *Biochemistry 38*, 9758−9767.
40. Bryson, E. A., Rankin, S. E., Goormaghtigh, E., Ruysschaert, J. M., Watts, A., and Pinheiro, T. J. (2000) *Eur. J. Biochem. 267*, 1390−1396.
41. Chakrabartty, A., Doig, A. J., and Baldwin, R. L. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 11332−11336.
42. Doig, A. J., Chakrabartty, A., Klingler, T. M., and Baldwin, R. L. (1994) *Biochemistry 33*, 3396−3403.
43. Roder, H., Elove, G. A., and Englander, S. W. (1988) *Nature 335*, 700−704.
44. Patel, C. N., Lind, M. C., and Pielak, G. J. (2001) *Protein Expression Purif. 22* 220−224.

BI026543Y

SUPPORTING INFORMATION



opticytc
397 bp

Synthetic Horse Heart Cyt c sequence (HrsOpti-1)

```
      +1                                    M  H  H   H  H  H   H  I  E   G  R  Q  G   D  P
          EcoRI KpnI   PstI            SspI   NsiI                                      BspEI
        ~~~~~~~~~~~~~~~~~~          ~~~~~~~~~~~~~~                                    ~~~~~~
      1 GAATTCGGTA CCCTGCAGTA AGGAGGAAAT ATTATGCATC ACCATCACCA CCATATCGAA GGTCGTCAGG GCGATCCGGA
        CTTAAGCCAT GGGACGTCAT TCCTCCTTTA TAATACGTAG TGGTAGTGGT GGTATAGCTT CCAGCAGTCC CGCTAGGCCT

      +1   K  G  K   K  I  F  V   Q  K  C   A  Q  C   H  T  V  E   K  G  G   K  N  K   T  G  P
                                                      PflMI
                                                   ~~~~~~~~~~~~~
                     XmnI                             DraIII
                  ~~~~~~~~~~~                       ~~~~~~~~~
     81 GAAAGGTAAG AAAATCTTCG TGCAGAAATG CGCCCAGTGC CACACCGTGG AAAAAGGTGG CAAAAACAAA ACGGGTCCGA
        CTTTCCATTC TTTTAGAAGC ACGTCTTTAC GCGGGTCACG GTGTGGCACC TTTTTCCACC GTTTTTGTTT TGCCCAGGCT

      +1 N  L  N  G   L  F  G   R  K  T  G   Q  A  P   G  F  T   Y  T  D  A   N  K  N   K  G  I
                         XmaIII
                         ~~~~~~~
                      SfiI                                    BstZ17I
                  ~~~~~~~~~~~~~~~                             ~~~~~~~
                        EagI                                    AccI
                      ~~~~~~~                                 ~~~~~~~
    161 ACCTGAACGG CCTGTTCGGC CGTAAAACCG GCCAAGCACC GGGCTTCACG TATACCGATG CGAACAAGAA TAAAGGCATC
        TGGACTTGCC GGACAAGCCG GCATTTTGGC CGGTTCGTGG CCCGAAGTGC ATATGGCTAC GCTTGTTCTT ATTTCCGTAG

      +1  T  W  K  E   E  T  L   M  E  Y   L  E  N  P   Q  A  Y   I  P  G   T  R  M  I   F  A
                                                                  XmaI
                                                                ~~~~~~
                                                                 SmaI
                                                                ~~~~~~
                                                                 AvaI
                                                                ~~~~~~
    241 ACCTGGAAAG AAGAGACCCT GATGGAATAC CTGGAAAATC CACAGGCGTA CATCCCGGGC ACCCGTATGA TCTTTGCGGG
        TGGACCTTTC TTCTCTGGGA CTACCTTATG GACCTTTTAG GTGTCCGCAT GTAGGGCCCG TGGGCATACT AGAAACGCCC
                  EarI
               ~~~~~~~
                   BsaI
                ~~~~~~

      +1   I  K  K   K  T  E  R   E  D  L   I  A  Y   L  K  K  A   T  N  E
                                BglII                                    HindIIIBamHI
                              ~~~~~~                                   ~~~~~~~~~~~~
    321 CATCAAGAAA AAGACCGAAC GCGAAGATCT GATCGCGTAT CTGAAAAAAG CGACCAACGA ATAATAAGCT TGGATCC
        GTAGTTCTTT TTCTGGCTTG CGCTTCTAGA CTAGCGCATA GACTTTTTTC GCTGGTTGCT TATTATTCGA ACCTAGG
```