Amino Acid Sequence of the Blue Copper Protein Rusticyanin from *Thiobacillus ferrooxidans*†

Michael Ronk,‡ John E. Shively,‡ Elizabeth A. Shute,§ and Robert C. Blake II*,§

Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010, and Department of Biochemistry, Meharry Medical College, Nashville, Tennessee 37208

Received April 18, 1991; Revised Manuscript Received July 17, 1991

ABSTRACT: Rusticyanin is a small blue copper protein isolated from *Thiobacillus ferrooxidans*. The amino acid sequence of the rusticyanin has been determined by the structural characterization of tryptic and endoproteinase Asp-N peptides with use of amino terminal microsequencing, fast atom bombardment mass spectrometry, and electrospray triple-quadrupole mass spectrometry techniques. Amino acid analysis, carboxy-terminal sequence analysis, and circular dichroism spectroscopy were also performed on the protein. Amino acid sequence identity among rusticyanin and six other small blue copper proteins is apparent only in the limited C-terminal region of each protein bearing three of the four putative copper ligands. A structural model of the rusticyanin is proposed where the protein is principally a β-barrel comprised of six strands. This model is consistent with the circular dichroism data and computational predictions of the secondary structure of rusticyanin. A feature of the model is the hypothesis that Asp 73 may serve as a fourth copper ligand.

Thiobacillus ferrooxidans is a well-characterized member of the thiobacilli group of sulfur bacteria. The bacterium is an obligately acidophilic Gram-negative chemolithotroph that derives the energy required for growth and cell maintenance from aerobic respiration on reduced sulfur compounds and/or ferrous iron. The bacterium oxidizes sulfur compounds to sulfuric acid and ferrous ions to ferric. The former activity acidifies the bacterium's immediate environment, while the latter activity generates an aggressive oxidant capable of oxidizing and solubilizing sulfide minerals within the bacterium's natural habitat. The possibility of exploiting this bacterial activity to extract metals for commercial gain, a process known as bacterial leaching, has received increasing attention in recent years.

Rusticyanin is a small type I copper protein thought to be a principal component in the iron respiratory electron transport chain of *T. ferrooxidans* (Blake et al., 1991). Rusticyanin differs from other small type I copper proteins in two ways. First, as a component of the periplasmic space of an acidophilic bacterium, the protein is remarkably stable at acidic pH. Indeed, the rusticyanin is redox-active down to pH 0.2 (Blake & Shute, 1987). Second, the protein which contains 1 atom of copper per molecule of protein, has the unusually high reduction potential of 680 mV (Ingledew & Cobley, 1980). Values for the reduction potentials of other comparably sized blue copper proteins range from lows of 184 mV for stellacyanin (Cox et al., 1978) and 217 mV for amicyanin (Tobari & Hadara, 1981), through the medium range of 266–304 mV for the various azurins and pseudoazurins (Wherland & Pecht, 1978), to a high of 370 mV for plastocyanin (Katoh et al., 1962). It has been proposed that differences in copper ligation are responsible for the observed differences in reduction potentials (Cox et al., 1978; Holt et al., 1990). The reduction potential of the type I copper center in laccase from *Polyporus versicolor* has been measured at 785 mV (Reinhammar, 1972). Laccase is a multicopper oxidase containing three spectroscopically different copper centers, and it was suggested that interaction between the type I center and the other electron acceptors in the molecule might affect the reduction potential of the type I center (Reinhammar, 1972). It is therefore of interest to investigate and identify the structural features that might contribute to the rusticyanin's unique functional properties.

The present paper describes the determination of the amino acid sequence of rusticyanin utilizing amino acid analysis, amino-terminal microsequencing, carboxy-terminal sequencing, FAB[1] mass spectrometry, and electrospray quadrupole mass spectrometry. The results demonstrate a protein of 155 amino acids containing one cysteine and no arginine residues. Comparisons of the rusticyanin sequence with those of other small blue copper proteins reveal that appreciable identity exists only in a limited portion of the C-terminal region of each protein containing three ligands to the copper atom. Chou–Fasman (Chou & Fasman, 1978) and a Garnier–Robson (Garnier et al., 1978) predictions were employed to construct a model of the secondary and tertiary structures of the rusticyanin that resembles the X-ray crystallographic structures of three other small blue copper proteins. The circular dichroism data were consistent with the proposed model.

Materials and Methods

Materials. TPCK–trypsin and Tris were obtained from Sigma (St. Louis, MO). Sequencing grade endoproteinase Asp-N was from Boehringer Mannheim (Indianapolis, IN). 4-Vinylpyridine, protein sequencing grade trifluoroacetic acid, and β-mercaptoethanol were purchased from Aldrich (Milwaukee, WI). Ammonium bicarbonate, formic acid, and

† This research was supported by National Cancer Institute Grant CA33572 to the City of Hope Cancer Center and United States Department of Energy Grant DE-FG05-85ER13339 to Meharry Medical College.

* To whom correspondence should be addressed.
‡ Beckman Research Institute of the City of Hope.
§ Meharry Medical College.

[1] Abbreviations: FAB, fast atom bombardment; TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; EDTA, (ethylenedinitrilo)tetraacetic acid disodium salt dihydrate; HPLC, high-performance liquid chromatography; TFA, trifluoroacetic acid; PVDF, poly(vinylidene difluoride); CD, circular dichroism.

0006-2960/91/0430-9435$02.50/0

EDTA were from Mallinckrodt (Paris, KY). Ultrapure guanidine hydrochloride was from Schwarz Mann (Cleveland, OH). HPLC columns for peptide preparation and separation were purchased from Vydac (Hesperia, CA). HPLC columns for phenylthiohydantoin-amino acid identification were from Beckman (San Ramon, CA) (Ultrasphere ODS 5 μm, 2.0 × 250 mm) or ABI (San Jose, CA) (PTH C-18 5-μm cartridge, 2.1 × 220 mm). HPLC columns for thiohydantoin-amino acid identification were bought from Phenomenex (Torrance, CA). Poly(vinylidene difluoride) membranes were from Millipore (Bedford, MA). Sulfuric acid and hydrogen peroxide were from Baker (Phillipsburg, NJ). HPLC grade acetonitrile and methanol were purchased from E. M. Science (Gibbstown, NJ).

Growth of T. ferrooxidans and Purification of Rusticyanin. Bacterial cell culture and protein purification were performed on *T. ferrooxidans* (ATCC 23270) as described (Blake & Shute, 1987). The electrophoretically homogeneous rusticyanin (in 0.01 N H_2SO_4) was further purified by reverse-phase HPLC using a Vydac C_4 column (4.6 × 250 mm) prior to peptide preparation, sequencing, and electrospray mass spectrometry. The HPLC buffer system and detection used to purify the native protein were the same as that described in the peptide separation section of this paper.

Peptide Preparation. The protein was reduced and alkylated according to a previously published procedure (Friedman et al., 1970). Rusticyanin was reduced with β-mercaptoethanol (reagent to protein ratio, 5:1 mol/mol) in 0.25 M Tris, pH 8.5, containing 6 M guanidine hydrochloride and 1.0 mM EDTA. The reaction mixture was purged with argon and incubated for 2 h at room temperature in the dark. The reduced protein was then alkylated with 4-vinylpyridine (reagent to protein ratio, 25:1 mol/mol) in the same buffer and under the same conditions. Desalting was performed on reverse-phase HPLC, on the same C_4 column mentioned above, according to a method described previously (Pan et al., 1984). The reduced, alkylated, and desalted protein was digested by use of either TPCK–trypsin or endoproteinase Asp-N. TPCK–trypsin was further purified by reverse-phase HPLC according to Titani et al. (1982). The enzyme to rusticyanin ratio in all proteolytic digests was 1:30 (w/w). Digests were performed in 0.2 M ammonium bicarbonate (trypsin) or 0.2 M sodium phosphate (Asp-N), pH 7–8, at 37 °C, for 24 h. A total of 19 μg (1.1 nmol) of rusticyanin was reduced, alkylated, and desalted prior to digestion to produce tryptic map A. A total of 120 μg (8 nmol) each was used for tryptic map B and the Asp-N map.

Peptide Separation. Peptides were separated by reverse-phase HPLC on a C_4 column (2.1 × 250 mm for tryptic map A, 4.6 × 250 mm for tryptic map B and the Asp-N map). The peptides were eluted with a linear gradient from 100% solvent A (0.1% TFA in water) to 100% solvent B (0.1% TFA in water/90% CH_3CN) over 60 min, with a flow rate of 0.2 mL/min (2.1-mm column) or 1 mL/min (4.6-mm column). Detection was done at 214 nm. Rechromatography was performed on specific peptide peaks (N5, N15, N20, N22) with a Vydac C_{18} column (4.6 × 250 mm), with use of the same solvent system and detection. Various gradients were utilized for rechromatography depending on the retention time of the original peptide peak.

Amino-Terminal Microsequence Analysis. Automated Edman degradation was performed on a gas-phase microsequencer built at the City of Hope (Hawke et al., 1985) that includes the addition of a continuous-flow reactor (CFR) (Shively, et al., 1987). The phenylthiohydantoin-amino acid derivatives were identified by on-line reverse-phase HPLC, as described by Shively et al. (1987). Each sample was lyophilized to a volume of about 10 μL, CH_3CN was added to be 50% of the total volume, and each resulting sample was spotted onto a PVDF membrane (1 mm × 1 cm) for amino-terminal sequence analysis. A total of 12 μg (0.7 nmol) of intact rusticyanin was applied to the microsequencer, with initial yields of 300–400 pmol. Typically, 25–33% of each peptide was applied to the microsequencer. Initial yields on the microsequencer were in the 50–200-pmol range for peptides from tryptic map A, and in the 200–500-pmol range for peptides from tryptic map B and the Asp-N map.

Mass Spectrometry. FAB mass spectrometry was performed on the peptides as described previously (Haniu et al., 1988). Typically, 25–33% of each peptide was used for the FAB mass spectrometry.

Electrospray mass spectra were obtained with use of a Finnigan MAT TSQ-700 triple-sector quadrupole mass spectrometer equipped with an electrospray ion source and DECstation 2100 computer workstation. The HPLC-purified intact rusticyanin was lyophilized and then dissolved in 50% aqueous methanol (v/v) to yield a final concentration of 10 pmol/μL. The protein solution was transferred to the electrospray source with use of a syringe pump (Harvard Apparatus) at a flow rate of 1 mL/min. Repetitive scans of the mass range from 50 to 2000 u were accumulated over 3 min, with an single scan cycle time of 5 s.

Amino Acid Analysis. Amino acid analysis was performed on 17 μg (1 nmol) of the rusticyanin. The presence of cysteic acid was detected after performic acid oxidation. The sample was lyophilized to dryness, redissolved in 100 μL of performic acid (90% formic acid/10% H_2O_2), and incubated for 15 min at room temperature. The sample was then lyophilized to dryness again prior to hydrolysis. The hydrolysis and analysis procedures used are described elsewhere (Hefta et al., 1990).

Carboxy-Terminal Sequence Analysis. A total of 85 μg (5 nmol) of rusticyanin was lyophilized to about 10 μL, CH_3CN was added to be 50% of the total volume, and the sample was spotted onto PVDF (1 mm × 1 cm) for carboxy-terminal sequencing analysis. The analysis was performed on a City of Hope built microsequencer adapted to thiohydantoin chemistry (Bailey & Shively, 1991). The thiohydantoin chemistry used is described elsewhere (Bailey & Shively, 1990). Thiohydantoin amino acid residues were identified by on-line reverse-phase HPLC, with use of a Phenomenex column (Ultracarb 5 ODS (30), 2.0 × 250 mm), with detection performed at 265 nm. The chromatography conditions for the thiohydantoin separation are described elsewhere (Bailey et al., 1991).

Circular Dichroism Analysis. Circular dichroism spectra on a 550 μg/mL solution of rusticyanin (in 0.01 N H_2SO_4) were recorded at 0.2-nm intervals over the wavelength range of 178–250 nm on a Jasco J-600 spectropolarimeter. Samples were placed in a 0.1-mm cell and scanned at room temperature at a rate of 20 nm/min. The spectra were the average of eight scans, with the H_2SO_4 background subtracted.

RESULTS

Peptide Mapping. Proteolytic digestion attempted without reduction and alkylation of the rusticyanin produced incomplete digestion, as characterized by an HPLC peptide map in which the major peak corresponded to undigested rusticyanin. Tryptic and Asp-N digests of rusticyanin, after reduction and alkylation, produced the peptide maps shown (Figures 1 and 2). Portions of all the major peptide peaks from tryptic map A and the Asp-N map were applied to the microsequencer.

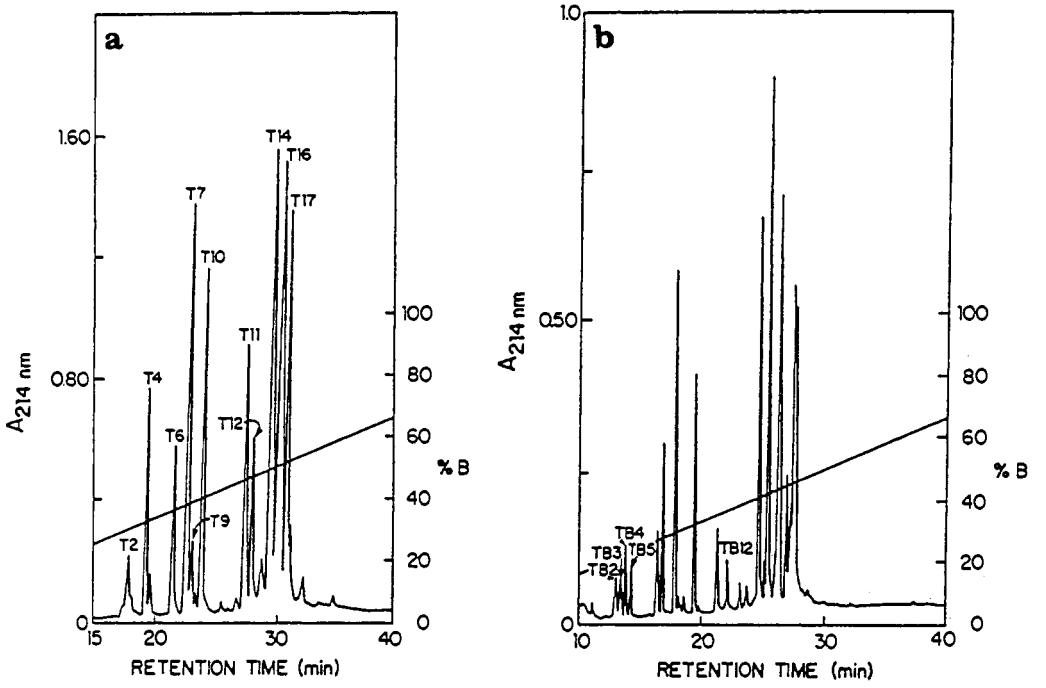

FIGURE 1: (a) Tryptic map A of rusticyanin from *T. ferrooxidans*. A total of 19 μg of reduced and alkylated protein was digested in 0.2 M ammonium bicarbonate, pH 7–8, for 24 h, at 37 °C. Separation was performed by reverse-phase HPLC using a Vydac C4 column (2.1 × 250 mm). (b) Tryptic map B of rusticyanin from *T. ferrooxidans*. A total of 120 μg of reduced and alkylated protein was digested in 0.2 M ammonium bicarbonate, pH 7–8, for 24 h, at 37 °C. Separation was performed by reverse-phase HPLC using a Vydac C4 column (4.6 × 250 mm).

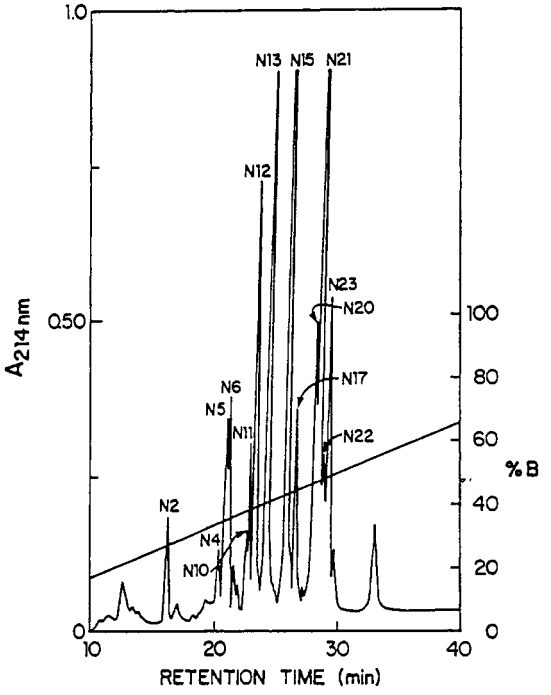

FIGURE 2: Endoproteinase Asp-N map of rusticyanin from *T. ferrooxidans*. A total of 120 μg of reduced and alkylated protein was digested in 0.2 M sodium phosphate, pH 7–8, for 24 h, at 37 °C. Separation was performed by reverse-phase HPLC using a Vydac C4 column (4.6 × 250 mm).

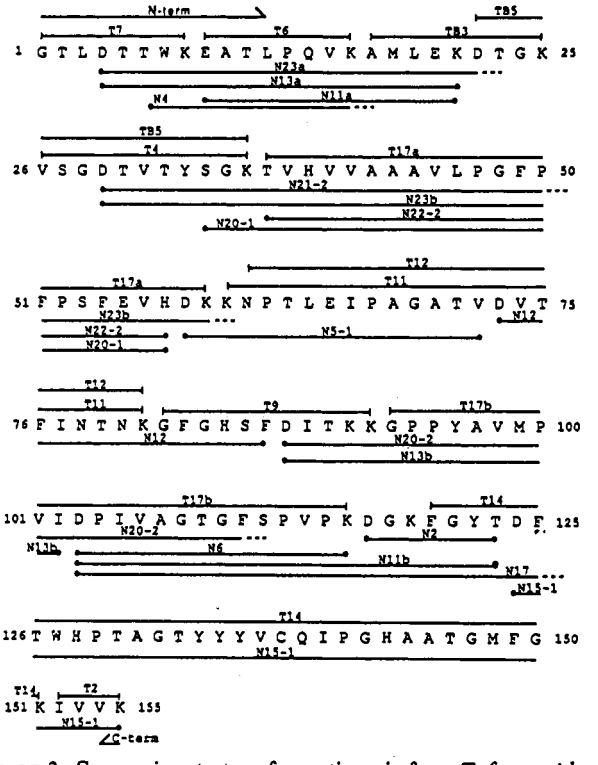

FIGURE 3: Sequencing strategy for rusticyanin from *T. ferrooxidans*. A bar with the letter T over it indicates tryptic peptides; a bar with the letter N over it indicates Asp-N peptides; forward half-arrows indicate N-terminal sequence data; and reverse half-arrows indicate C-terminal sequence data. Hyphenated numbers (i.e., N5-1) refer to peptides that were subjected to rechromatography. Lowercase letters (i.e., T17b) refer to HPLC peaks that yielded two discernible peptides upon microsequencing. Peptides that washed out during sequencing chemistry are indicated by termination in a dashed line.

A second tryptic map (map B) was judged necessary due to certain peptides from map A producing ambiguous or incomplete sequence analysis results. Only the peaks from map B needed to clear up the uncertainties were sequenced. The small differences in the elution profile between tryptic maps A and B are probably due to resolution differences in the two HPLC columns used (2.0 mm i.d. for A vs 4.6 mm i.d. for B).

Sequence Determination. The resulting sequencing strategy is shown (Figure 3). All tryptic fragments, except for the tripeptide at residues 117–119, were recovered from the maps. All Asp-N fragments were recovered except for the tripeptide at residues 1–3. The sequences from these two peptides were

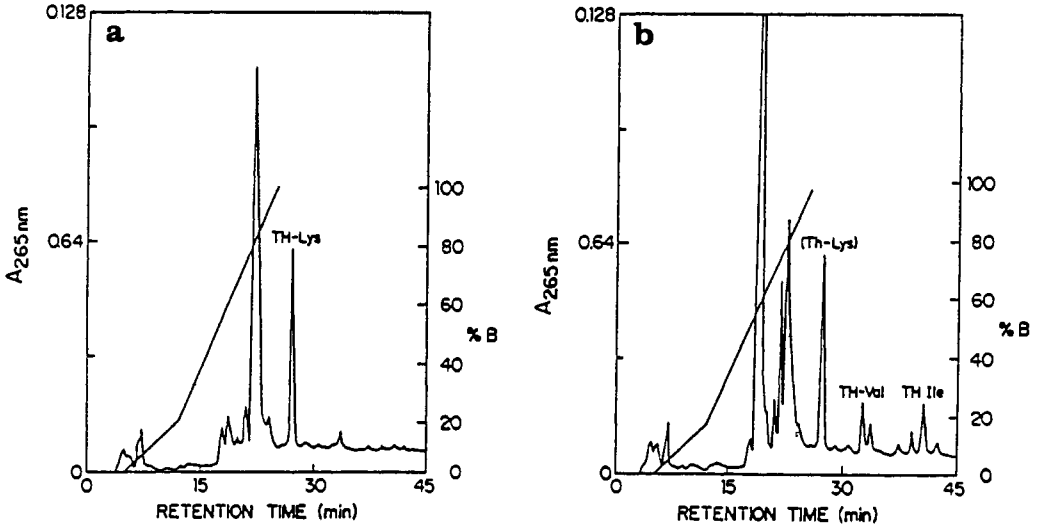

FIGURE 4: (a) Cycle 1 of the carboxy-terminal sequence of rusticyanin from *T. ferrooxidans*. A total of 85 μg of the rusticyanin was applied to the sequencer. The amino acid residues were identified by on-line reverse-phase HPLC, with use of a Phenomenex column (Ultracarb 5 ODS (30), 2.0 × 250 mm), at 265 nm. Th, thiohydantoin. (b) Cycle 2 of the carboxy-terminal sequence of rusticyanin from *T. ferrooxidans*. A total of 85 μg of the rusticyanin was applied to the sequencer. The amino acid residues were identified by on-line reverse-phase HPLC, with use of a Phenomenex column (Ultracarb 5 ODS (30), 2.0 × 250 mm), at 265 nm. Th, thiohydantoin. (Th-Lys), carry-over from previous cycle.

accounted for by their inclusion in other proteolytic fragments. Overlap was produced at all tryptic sites in the protein by the Asp-N digestion. The Asp-N digestion produced nonspecific cleavages at Glu, Thr, and Ser residues (for peptides N11a, N22-2, and N20-1, respectively).

FAB mass spectrometry was performed on the rusticyanin peptides for which amino acid sequence data had been obtained (Table I). It should be noted that the mass spectrometry data gives no indication of unusual or unexpected posttranslational modifications that might contribute to the extraordinary acid stability of the rusticyanin. The peptides lacking mass spectrometry data either were trivial extensions of other peptides (peptide T16a differs from T17b by one lysine residue at the amino terminus) or may have been peptides of indeterminate length produced by incomplete proteolysis by the Asp-N protease.

Peptides T14 and N15-1, both of which contain rusticyanin's only cysteine, each displayed evidence, both by microsequencing and FAB mass spectrometry results, of a (pyridylethyl)cysteine. This agrees with the amino acid analysis showing the presence of only one Cys residue in the rusticyanin molecule.

The entire sequence of *T. ferrooxidans* rusticyanin was deduced from the amino-terminal microsequencer analysis except for confirmation of the C-terminus of the protein. Sequence and mass spectrometry data from peptide N15-1 indicated that the protein terminated in a lysine. This left us with the problem of not knowing whether this was a combined tryptic and Asp-N site or not; a small extension beginning with an aspartic acid could have followed the lysine residue. A C-terminal peptide of as many as four residues could possibly have been excluded from the gradient during HPLC separation of the peptides.

To help solve this problem, we utilized chemistry being developed in this lab (Bailey & Shively, 1990), which allowed for the formation of thiohydantoin amino acid residues at the C-terminus of the protein. The first two cycles produced by this method are shown (Figure 4). The initial yield for lysine was 250 pmol/5 nmol, with carry-over on the next cycle being essentially 100%. Initial yield for valine on the second cycle was 40 pmol/5 nmol. Sequencing yields diminished rapidly after three cycles; this is due to the protein not being covalently bound to the sequencing support. The low initial yield and high carry-over are due to the chemistry not yet being optimized. The thiohydantoin isoleucine found on cycle 2 is

Table I: FAB Mass Spectrometry Data of *T. ferrooxidans* Rusticyanin Peptides[a]

peptide	obsd MH^+ monoisotopic mass	calcd MH^+ monoisotopic mass
T2	458.3	458.3
T4	1113.4	1113.5
T6	885.6	885.5
T7	921.5	922.0
T9	1236.9	1236.6
T10	1108.4	1108.5
T11	2343.0	2343.3
T12	2215.1	2215.2
T14*	3692.8	3692.7
T16a[b]	no mass obtained	2537.5
T16b	same as peptide T17a	
T17a	2463.8	2464.3
T17b	2409.2	2409.3
TB2	same as peptide T2	
TB3	992.5	992.5
TB4	same as peptide T4	
TB5	1514.5	1514.7
TB12	same as peptide T11	
N2	no mass obtained	787.4
N4	no mass obtained	1961.0
N5-1	1553.8	1553.9
N6	1384.7	1384.8
N11a	no mass obtained	1547.8
N11b	2153.4	2153.1
N12	1683.7	1683.8
N13a	2089.1	2089.1
N13b	1628.7	1628.9
N15-1*	3664.4	3663.8
N17	no mass obtained	5852.9
N20-1	2492.8	2493.3
N20-2	no mass obtained	2994.6
N21-2	no mass obtained	3072.6
N22-2	2221.1	2221.2
N23a	no mass obtained	2735.4
N23b	no mass obtained	4607.4

[a] T, tryptic peptides; N, Asp-N peptides, MH^+ monoisotopic mass for the peptides marked with an asterisk includes the mass of the (pyridylethyl)cysteine. [b] Peptide 17b is identical with the Lys-extended (at the N-terminus) peptide 16a.

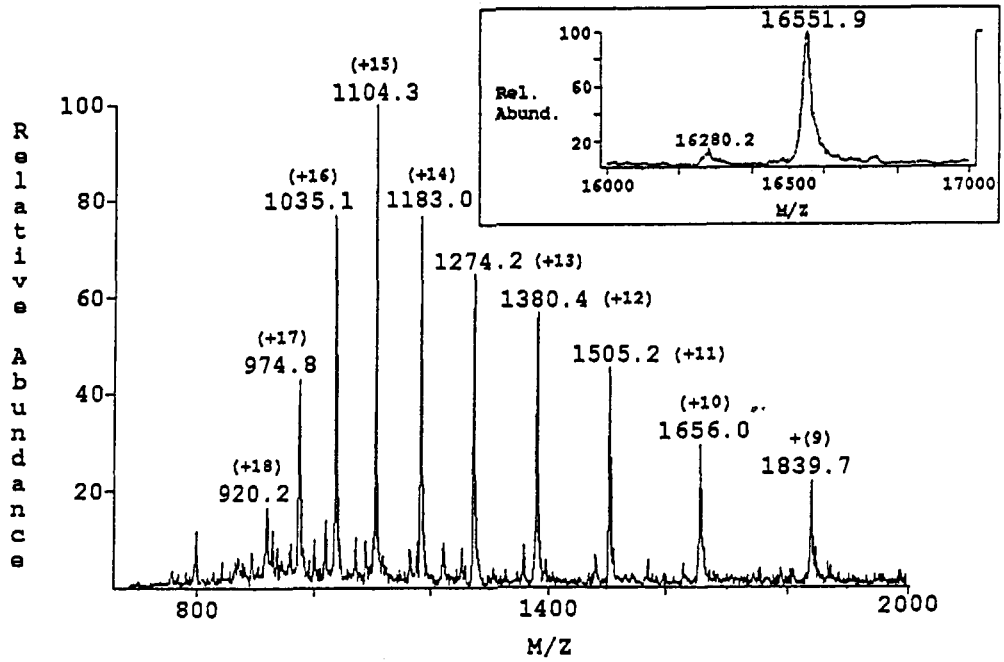

FIGURE 5: Positive-ion electrospray mass spectrum of *T. ferrooxidans* rusticyanin showing ions formed by multiple attachment of protons to the molecule. The m/z value of each ion is indicated along with the number of charges in parentheses. (Inset) The deconvoluted spectrum obtained with use of the BIOMASS data reduction system.

thought to be a "preview" amino acid, which have been noted when excessive amounts of protein are subjected to the C-terminal chemistry. The amino-terminal sequence and FAB mass spectrometry data on peptide N15-1 discount the possibility of an isoleucine at this position. The analysis shows the protein ends in a lysine residue, indicating that peptide N15-1 is the C-terminal peptide.

Several months after performing the carboxy-terminal analysis we were able to perform electrospray mass spectrometry on the intact rusticyanin in our laboratory. The positive-ion electrospray mass spectrum is shown (Figure 5). Multiply charged ions were observed for the molecule with 9–18 attached protons. The inset portion of Figure 5 shows the deconvoluted spectrum for the neutral rusticyanin molecule, obtained with use of Finnigan MAT's BIOMASS deconvolution software. The higher mass peak (average mass = 16551.9 u) in the deconvoluted spectrum is 0.9 mass unit higher than the calculated mass for the rusticyanin without its copper atom. It was not surprising that the copper atom was absent from the rusticyanin molecule, since the protein was purified by reverse-phase HPLC and lyophilized to complete dryness before being dissolved for the electrospray analysis. The smaller mass peak (average mass = 16280.2 u) probably represents the rusticyanin minus residues 1–3. At the time of the electrospray analysis, the protein had been stored in 0.1 N H_2SO_4 for more than a year, so an acid-catalyzed cleavage of the peptide bond at Asp 4 is a possible explanation for the truncated mass. The electrospray mass spectrometry data is entirely consistent with our elucidation of the amino acid sequence of the rusticyanin. It also further points out the absence of posttranslational modifications of the rusticyanin.

Amino acid composition of rusticyanin obtained experimentally shows agreement with the amino acid composition calculated from the sequence data (Table II). All discrepancies are plus-or-minus one residue, which is significant only with methionine, of which there are only three in the protein by sequence, two by amino acid analysis. However, since the typical yield for methionine by amino acid analysis in our laboratory is in the 60–70% range, it is likely there are three

Table II: Comparison of the Amino Acid Composition Obtained Experimentally from *T. ferrooxidans* Rusticyanin and the Amino Acid Composition Calculated from the Sequence Data

	exptl composition		calcd composition
	mol %	residues/mol	residues/mol
Ala	8.1	12	12
Cys*	0.8	1	1
Asx	8.7	13	12
Glx	4.5	7	6
Phe	6.6	10	10
Gly	11.3	17	17
His	3.4	5	5
Ile	3.9	6	7
Lys	9.0	14	14
Leu	3.6	5	5
Met	1.6	2	3
Pro	9.8	15	14
Arg	0.1	0	0
Ser	3.2	5	5
Thr	12.1	18	19
Val	10.3	16	17
Trp			2
Tyr	3.7	6	6

[a] The asterisk indicates determination after performic acid oxidation.

Met residues in the protein. The amino acid composition calculated from our sequence data is also consistent with a previously published composition (Cox & Boxer, 1978).

Circular Dichroism Analysis. A typical CD spectrum obtained for *T. ferrooxidans* rusticyanin is shown (Figure 6). When compared to typical spectra for various protein conformations (α-helix, β-sheet, β-turn, and unordered), the rusticyanin spectrum shows an atypical set of bands: a relatively strong negative peak at about 213 nm and a weaker negative peak at 210 nm. Another negative peak appears at 200 nm, and a weak positive peak is shown at about 193 nm. The spectrum is similar to that expected for β-sheet, which has a negative peak at 217 nm and positive one around 195 nm; however, both of the bands appear to be blue-shifted by 2–4 nm. Similar shifts in the major bands of protein CD spectra have been found previously (Hennessey & Johnson, 1981). Spinach plastocyanin, a small blue copper protein with a high content of β-sheet (Colman et al., 1978), has been found

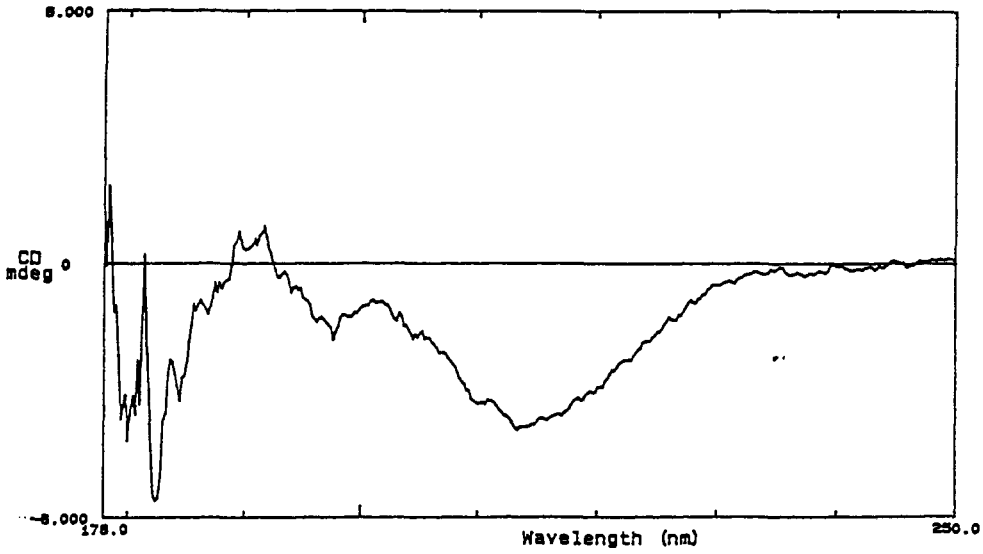

FIGURE 6: Circular dichroism spectrum of rusticyanin from *T. ferrooxidans*. The spectrum is the average of eight scans and was recorded on a 550 mg/mL solution of rusticyanin (in 0.01 N sulfuric acid) at 0.2 nm over the wavelength range of 178–250 nm.

to have a far-UV CD spectrum resembling that of rusticyanin (Draheim et al., 1986). However, its major bands are positive at 198 nm and negative at 225 nm, both of which are red-shifted from the ideal absorption wavelengths for β-sheet structure. The negative peak at 210 nm in the rusticyanin spectrum is similar to one of the negative bands generally attributed to α-helix (208 and 220 nm), although it is very weak and the negative band at 220 nm is not present. The transition at 200 nm is similar to that caused by unordered form, though the magnitude of the peak is somewhat less than expected.

The fractional compositions of secondary structure present in rusticyanin were calculated from the experimental spectrum by use of algorithms based on work by Greenfield and Fasman (Greenfield & Fasman, 1969). This calculates a nonrestrained least-squares fit for α-helix, β-sheet, and random coil, optimizing curve shape over the CD intensity. Four spectra, each averaged over eight scans, were used to produce estimates of 5% α-helix, 50% β-sheet, and 45% random coil for the native rusticyanin, with an average correlation coefficient of 0.949.

Discussion

The structures of the three small blue copper proteins azurin (*Alcaligenes denitrificans*) (Norris et al., 1983), plastocyanin (Lombardy poplar) (Colman, et al., 1978), and pseudoazurin (*Alcaligenes faecalis*) (Petratos et al., 1987) have been previously determined by X-ray crystallography. Identity between the sequence of plastocyanin and the sequences of azurin and and pseudoazurin is only 9 and 20%, respectively. Even with this low level of sequence identity, all three of these blue copper proteins show a high degree of similarity in their secondary and tertiary structures. The three proteins are in the form of a β-barrel, comprised of eight strands, with the copper atom located in the interior of the barrel. Pseudoazurin consists of two β-sheets stacked face-to-face; β-sheet I being made of strands 1, 3, and 6, β-sheet II being made of strands 2, 4, 5, 7, and 8 (Petratos et al., 1987). The last 30 amino acids of pseudoazurin form two α-helices, a structure not found in plastocyanin (Petratos et al., 1987). Azurin is also formed by two β-sheets face-to-face, with strands 1, 3, and 6 making up one sheet and strands 4, 5, 7, and 8 making up the other (Norris et al., 1983). Azurin contains a single α-helix in a 30 amino acid stretch between β-strands 4 and 5 (Norris et al., 1983). Plastocyanin has the same general β-sheet structure as pseudoazurin and azurin (Colman et al., 1978). The two facing β-sheets in plastocyanin are composed of strands 1–4 and 6–8 (Colman et al., 1978). In all three proteins, the four copper ligands—one Met, one Cys, and two His residues—are conserved. Three of these invariant residues, Cys, Met, and a His, are clustered in a region of conserved hydrophobic residues near the C-terminus of each protein. It has been postulated that this hydrophobic patch is involved in electron transfer either as an attractive region for negatively charged inorganic reductants and oxidants (Colman et al., 1978; Norris et al., 1983) or as the lining for "hydrophobic channels" for electron transport (Colman et al., 1978). The fourth invariant copper ligand, another His, is typically found in the N-terminal region of the proteins.

Alignment of rusticyanin and azurin produces 18 of 129 residues invariant, including the three putative copper ligands near the C-terminus of the proteins, and indicates an N-terminal extension of 26 amino acids in rusticyanin. Alignment of rusticyanin and plastocyanin shows 14 of 99 residues invariant, including the same putative copper ligands, and an N-terminal extension of 52 residues for rusticyanin. Alignment of rusticyanin and pseudoazurin shows 13 of 123 residues invariant, including the three putative copper ligands. In this last alignment, rusticyanin has a 61-residue N-terminal extension, and pseudoazurin shows an extension of 14 amino acids at the C-terminus. These levels of sequence identity are consistent with those shown between the three blue copper proteins discussed above.

A comparison of the region about the copper ligands near the C-terminus of several small blue copper proteins shows a high amount of identity (Figure 7), including the above-mentioned hydrophobic residues. The necessity of reduction and alkylation prior to proteolytic digestion of rusticyanin, even though there is no disulfide bond present in the molecule, is evidence of Cys 138 being involved in binding of the copper atom. Without the coordination bond between Cys 138 and the copper atom being irreversibly broken, the protein molecule is not completely denatured and therefore does not undergo complete proteolytic digestion. This is in agreement with the absorbance spectra of rusticyanin, which shows characteristics of charge transfer between a cystyl residue and a protein-bound Cu(II) (Blake & Shute, 1987). It is also in agreement with recent extended X-ray absorption fine structure studies done on rusticyanin (Holt et al., 1990). This EXAFS work also

TFR	134	Y	Y	Y	V	C	Q	I	P	G	H	A	A	T	G	M	F	G
AZ	108	Y	A	Y	F	C	S	F	P	G	H	W	A	M		M	K	G
PAZ	74	Y	G	F	K	C	A		P		H	Y	M	M	G	M	V	A
BCP	74	Y	L	V	K	C	T		P		H	Y	A	M	G	M	I	A
PL	80	Y	S	F	Y	C	S		P		H	Q	G	A	G	M	V	G
AM	82	Y	D	Y	I	C	T		P		H	P		F		M	K	G
ST	84	Y		Y	I	C	G	V	P	K	H	C	D	L	G	Q	K	V	H

FIGURE 7: Alignment of *T. ferrooxidans* rusticyanin with several blue copper proteins in the C-terminal region containing three of the copper ligands. Abbreviations: AZ, azurin (*A. denitrificans*); PAZ, pseudoazurin (*Pseudomonas* AM1); BCP, blue copper protein (*A. faecalis*); PL, plastocyanin (*Populus nigra* var. *italica*); AM, amicyanin (*Pseudomonas* AM1); ST, stellacyanin (*Rhus vernicifera*). The conserved copper ligands are shown in a bold box. Other conserved residues are shown in a box. Residues in the putative hydrophobic binding patch and hydrophobic channel are highlighted by a bar at the bottom of the figure.

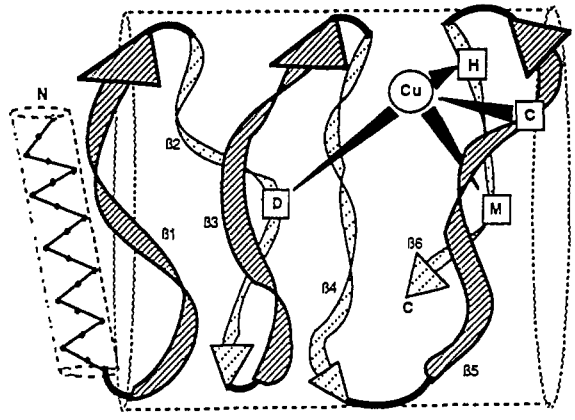

FIGURE 8: Proposed model for rusticyanin from *T. ferrooxidans*. The strands of the β-barrel are represented by the striated ribbons, with the darker ribbons running on top. The strands are numbered 1–6, as they appear in the amino acid sequence.

produced evidence of a Cu–S bond in addition to that formed by Cys 138. The Cu–S bond detected by EXAFS may correspond to Met 148, which is located in the highly conserved C-terminal region.

The histidine residue serving as a copper ligand in the amino-terminal region of the other small blue copper proteins is not conserved in rusticyanin, although there are two histidines toward the amino terminus in rusticyanin (His 39 and His 57) that could serve as possible ligands. The above-mentioned EXAFS study (Holt et al., 1990) reports evidence of two Cu–N bonds, indicating that one of these His residues could serve as a copper ligand. However, recent saturation mutagenesis work on azurin from *Pseudomonas aeruginosa* suggests another possible fourth ligand in rusticyanin. Replacement of the codon for His 46 in *P. aeruginosa* azurin with the codons for other naturally occurring amino acids has shown that an aspartic acid replacement will allow the mutant protein to retain its copper center.[2] This work raises the possibility of Asp 73 in rusticyanin, which aligns with His 46 in *P. aeruginosa* azurin, serving as the fourth copper ligand in rusticyanin. Evidence that the copper ligands are not conserved among all blue copper proteins has been previously published (Bergman et al., 1977). Stellacyanin, a blue copper protein from the Japanese lacquer tree, contains no methionine, while the other three copper ligands are conserved. Since both stellacyanin and rusticyanin have markedly different reduction potentials from the other small blue copper proteins, it would be of interest to speculate whether the ligation of the copper atom in rusticyanin with Asp 73 might account for its higher reduction potential.

Both Chou–Fasman (Chou & Fasman, 1978) and Garnier–Robson (Garnier et al., 1978) computerized analyses of the amino acid sequence of rusticyanin predict an α-helix structure involving the first 15–20 amino acids at the N-terminus. Both algorithms also can be interpreted to predict six β-strands in the rusticyanin molecule. The first β-strand (β1) is tentatively comprised of residues 27–46. The second β-strand (β2) is predicted to be made of residues 61–79, β3 of residues 97–107, β4 of residues 119–128, β5 of residues 135–140, of β6 of residues 150–155. Each β-strand is predicted to be separated by regions of turn comprised of about 5–10 residues. Chou–Fasman predicts conformation values of 11% helix, 55% extended, and 34% turn. The Garnier–Robson predicts conformation values of 18% helix, 35% extended, 16% turn, and 30% coil.

The weak band at 210 nm shown in the CD spectrum (Figure 6) indicates that rusticyanin contains a small amount of α-helix. The second negative peak normally found with α-helix has possibly been obscured by the strong negative band at 213 nm, which is probably an indicator of the presence of a significant amount of β-sheet in the native protein. The relative weakness of the negative peak at 200 nm may be due to the interaction of β-form and unordered form, as has been previously postulated in the protein α-bungarotoxin (Chen et al., 1982). There are two plausible factors that could contribute to the blue shifting of the two peaks (positive at 193 nm and negative at 213 nm) associated with the β-form conformation of rusticyanin. First, as has been reviewed elsewhere (Yang et al., 1986), β-forms and β-turns have a much broader range of conformations than α-helices, leading to a broader range of CD spectra for proteins containing significant amounts of these conformations. Second, as was found in a study of cobra neurotoxin (Chen et al., 1977), the contributions of aromatic residues in the far-UV region of a CD spectrum cannot be ignored when dealing with β-form and β-turn. A drawback to the Greenfield and Fasman algorithm when it is used to analyze an atypical CD spectra of a protein such as rusticyanin is that it does not take into account the contributions of β-turn to the spectrum. Taking this into consideration, analysis of the experimental CD data by the Greenfield and Fasman algorithm indicated the Garnier–Robson prediction for helix was high, but otherwise the predictions of secondary structure were in reasonable agreement with the analysis of the CD spectra.

On the basis of the similarity of the structures of the three blue copper proteins discussed above, the similarity in the level of sequence identity between rusticyanin and these three proteins, the CD data from the native rusticyanin, and the secondary structural prediction for rusticyanin, we suggest a plausible model for *T. ferrooxidans* rusticyanin. We predict an α-helix extension of 21 amino acids at the N-terminus and a β-barrel formed by two face-to-face β-sheets comprised of six β-strands, each strand separated by short stretches of β-turn, with the copper atom located in the interior of the β-barrel. A two-dimensional schematic of the model is shown (Figure 8). As shown, the two β-sheets would be formed of strands 1–3 and 4–6. However, further data (X-ray crystal or NMR) would be necessary to determine the true β-strand

[2] Personal communication from Juris Germanas, Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA 91125.

orientation and β-sheet formation. In the model, an attempt has been made to suggest the three-dimensional form of the β-barrel by use of a dashed line in the form of a cylinder. The fourth coordinative bond of the copper atom shown with Asp 73 is based upon the work of Germanas et al.[2] The correct positioning of the N-terminal α-helix is not known. We speculate that this helical structure may replace the apparent function of the Cys 3–Cys 26 disulfide bond found in the N-terminus of azurin (Petratos et al., 1987). Hydrogen bonding between the side chains of the residues comprising the helix and those of one or more of the β-strands may serve to hold the N-terminal extension in close to the rest of the β-barrel structure of the protein.

The strengths of the model shown in Figure 8 are (a) it is consistent with the structural data presented in this paper and elsewhere and (b) it provides a clear hypothesis for further experimental investigations. For instance, efforts to generate crystals of rusticyanin for X-ray crystallographic studies are currently underway. In the meantime, it would be of interest to investigate the Asp 46 mutant of *P. aeruginosa* azurin to determine whether it (a) displayed a much higher reduction potential than the native azurin or (b) exhibited an EPR spectrum more closely resembling rusticyanin than native azurin.

Acknowledgments

We thank Dr. Kay Rutherfurd for her assistance and advice during the circular dichroism analysis portion of this work, Dr. Gottfried Feistner for his assistance during the electrospray mass spectrometry analysis, and Jim Sligar for providing the amino acid analysis data.

References

Bailey, J. M., & Shively, J. E. (1990) *Biochemistry 29*, 3145–3156.

Bailey, J. M., & Shively, J. E. (1991) In *Techniques in Protein Chemistry II* (Villafranca, J. J., Ed.) pp 115–129, Academic Press, San Diego.

Bailey, J. M., Shenoy, N. R., Ronk, M., & Shively, J. E. (1991) *Protein Science* (in press).

Bergman, C., Gandvik, E.-K., Nyman, P. O., & Strid, L. (1977) *Biochem. Biophys. Res. Commun. 77*, 1052–1059.

Blake, R. C., II, & Shute, E. A. (1987) *J. Biol. Chem. 262*, 14983–14989.

Blake, R. C., II, White, K. J., & Shute, E. A. (1991) *Biochemistry* (following paper in this issue).

Chen, Y.-H., Lo, T.-B., & Yang, J. T. (1977) *Biochemistry 16*, 1826–1830.

Chen, Y.-H., Tai, J.-C., Huang, W.-J., Lai, M.-Z., Hung, M.-C., Lai, M.-D., & Yang, J. T. (1982) *Biochemistry 21*, 2592–2600.

Chou, P. Y., & Fasman, G. D. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol. 47*, 45–148.

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

Cox, J. C., & Boxer, D. H. (1978) *Biochem. J. 174*, 497–502.

Cox, J. C., Assa, R., & Malmstrom, B. G. (1978) *FEBS Lett. 93*, 157–160.

Draheim, J. E., Anderson, G. P., Duane, J. W., & Gross, E. L. (1986) *Biophys. J. 49*, 891–900.

Friedman, M., Krull, L. H., & Cavins, J. F. (1970) *J. Biol. Chem. 245*, 3868–3871.

Garnier, J., Osguthorpe, D. J., & Robson, B. (1978) *J. Mol. Biol. 120*, 97–120.

Greenfield, N., & Fasman, G. D. (1969) *Biochemistry 8*, 4108–4115.

Haniu, M., Yuan, H., Chen, S., Takashi, I., Lee, T. D., & Shively, J. E. (1988) *Biochemistry 27*, 6877–6883.

Hawke, D. H., Harris, D. C., & Shively, J. E. (1985) *Anal. Biochem. 147*, 315–330.

Hefta, S. A., Paxton, R. J., & Shively, J. E. (1990) *J. Biol. Chem. 265*, 8618–8626.

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

Holt, S. D., Piggot, B., Ingledew, W. J., Feiters, M. C., & Diakun, G. P. (1990) *FEBS Lett. 269*, 117–121.

Ingledew, W. J., & Cobley, J. G. (1980) *Biochim. Biophys. Acta 590*, 141–158.

Katoh, S., Shiratori, I., & Takamiya, A. (1962) *J. Biochem. 51*, 32–40.

Norris, G. E., Anderson, B. F., & Baker, E. N. (1983) *J. Mol. Biol. 165*, 501–521.

Pan, Y.-C. E., Wideman, J., Blacher, R., Chang, M., & Stein, S. (1984) *J. Chromatogr. 297*, 13–19.

Petratos, K., Banner, D. W., Beppu, T., Wilson, K. S., & Tsernglou, D. (1987) *FEBS Lett. 218*, 209–214.

Reinhammar, B. R. (1972) *Biochim. Biophys. Acta 275*, 245–259.

Shively, J. E., Miller, P., & Ronk, M. (1987) *Anal. Biochem. 163*, 517–529.

Tobari, J., & Hadara, Y. (1981) *Biochem. Biophys. Res. Commun. 101*, 502–508.

Wherland, S., & Pecht, I. (1978) *Biochemistry 17*, 2585–2591.

Yang, J. T., Wu, C.-S. C., & Martinez, H. M. (1986) *Methods Enzymol. 130*, 208–269.