Structure and function of a novel cytochrome c_4 from the purple photosynthetic bacterium *Thiocapsa roseopersicina.*

Rui Miguel Mamede Branca

Ph.D. Thesis

Supervisor: Dr. Csaba Bagyinka

Institute of Biophysics, Biological Research Center
of the Hungarian Academy of Sciences

Ph.D. School of Physics
University of Szeged

Szeged

2008

List of Original Publications

List of publications directly related to the subject of the thesis:

1. **Branca RMM**, Bodó G, Bagyinka C, Prokai L. (2007) De novo sequencing of a 21-kDa cytochrome c_4 from *Thiocapsa roseopersicina* by nanoelectrospray ionization ion-trap and Fourier-transform ion-cyclotron resonance mass spectrometry. *J Mass Spectrom* **42**, 1569-1582.
   IF = 3.574
2. **Branca RMM**, Bodó G, Várkonyi Z, Debreczeny M, Ősz J, Bagyinka C. (2007) Oxygen and temperature-dependent structural and redox changes in a novel cytochrome c_4 from the purple sulphur photosynthetic bacterium *Thiocapsa roseopersicina*. *Archives Biochem Biophys* **467**, 174-184.
   IF = 3.152
3. Tomcová I, **Branca RMM**, Bodó G, Bagyinka C, Smatanová IK. (2006) Cross-crystallization method used for the crystallization and preliminary diffraction analysis of a novel di-haem cytochrome c_4. *Acta Cryst* F **62**, 820-824.

Other publications:

1. Ősz J, Bodó G, **Branca RMM**, Bagyinka C. (2005) Theoretical calculations on hydrogenase kinetics: explanation of the lag phase and the enzyme concentration dependence of the activity of hydrogenase uptake. *Biophys J* **89**, 1957-1964.
   IF = 4.507

Conference proceedings:

1. **Branca RMM**, Bodó G, Bagyinka C, Prokai L. (2007) De Novo sequencing of a 21-kDa Cytochrome c_4 from *Thiocapsa roseopersicina* by Nano-ESI Ion Trap and FT-ICR Mass Spectrometry. *25th Informal Meeting on Mass Spectrometry, Nyíregyháza-Sóstó, Hungary*; p 33
2. **Branca RMM**, Bodó G, Bagyinka C, Prokai L. (2007) De Novo sequencing of a 21-kDa Cytochrome c_4 from *Thiocapsa roseopersicina* by Nano-ESI Ion Trap and FT-ICR Mass Spectrometry. *Regional Biophysics Conference, Balatonfüred, Hungary*; p 74
3. **Branca RMM**, Varkonyi Z, Bodó G, Ősz J, Debreczeny M, Bagyinka C. (2006) Cytochrome c_4 from *Thiocapsa roseopersicina*. *8th Internacional Conference on Membrane Redox Systems, Szeged, Hungary*; p 24

Contents

Abbreviations

Amino acid 3-letter and 1-letter codes

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

BSA – Bovine Serum Albumin
CD – Circular Dichroism
CID – Collision Induced Dissociation
DEAE – DiEthylAminoEthyl
DI – Direct Infusion
DSC – Differential Scanning Calorimetry
ECD – Electron Capture Dissociation
EDTA – Ethylenediaminetetraacetic acid
ESI – ElectroSpray Ionization
FAB – Fast Atom Bombardment
FFT – Fast Fourier Transformation
FPLC – Fast Protein Liquid Chromatography
FTICR – Fourier Transform Ion-Cyclotron Resonance
FTMS – Fourier Transform Mass Spectrometry
HPLC – High Performance Liquid Chromatography
IEF – IsoElectric Focusing
IT – Ion Trap
ITMS – Ion Trap Mass Spectrometry
ITO – Indium-Tin-Oxide
LC – Liquid Chromatography
M/ΔM – Mass Resolution
MALDI – Matrix-Assisted Laser Desorption Ionization
MS – Mass Spectrometry
MS/MS or MS^2 – Tandem Mass Spectrometry
NHE – Normal Hydrogen Electrode
NSI – NanoelectroSpray Ionization
PAGE – PolyAcrylamide Gel Electrophoresis
SDS – Sodium Dodecyl Sulfate
SVD – Singular Value Decomposition

Introduction

Cytochromes

Cytochromes are metalloproteins containing one or more heme rings as prosthetic groups. Hemes are protoporphyrin IX rings with an iron (Fe) centre (Fig.1).

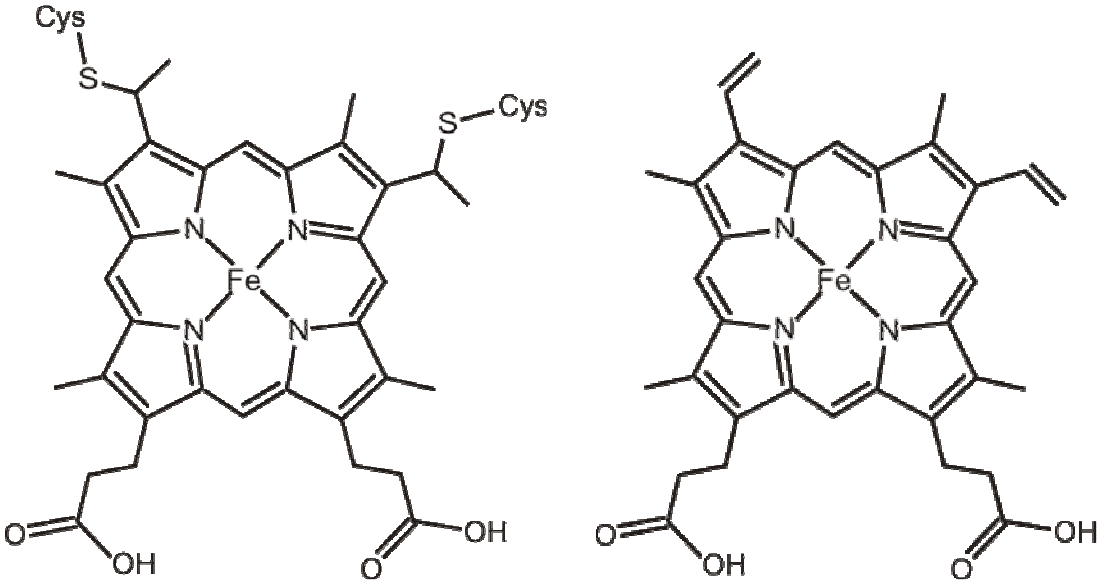

Figure 1. Heme *c* (left) and heme *b* (right). In cytochromes *c*, the vinyl groups bind covalently to cysteine residues of the polypeptide chain, firmly anchoring the heme group to the protein. The macrocyclic system of delocalised π electrons (arising from the system of conjugated double bonds) confers great stability to the liganded Fe and modulates the redox reactions.

Cytochromes are an important structural theme in nature, given their widespread presence in biological organisms. Eukaryotes and prokaryotes alike employ cytochromes in a wide variety of electron and/or proton transfer reactions. Bacteria, for example, benefit from the flexibility of this class of proteins by being able to grow on a wide range of energy sources [Page, *et al.* 1998, Pettigrew and Moore 1987].

Flexibility of the cytochrome motif implies that alterations to the amino acid sequence (and thus to the protein moiety) allow adjustment of not only docking sites (for selective partner recognition), but also of the heme iron redox potential. In Fig. 2 we can see an example of such versatility. Cytochrome c_{555} from *Aquifex aeolicus* expressed in *Escherichia coli* shows a midpoint redox potential of 221 ± 10 mV in the native form whereas the same protein with a Met78His site-directed mutation (i.e. 6th ligand methionine replaced by a histidine) shows a potential of -168 ± 10 mV [Aubert, et al. 2001].

6th ligand – methionine (+221 mV) **6th ligand – histidine (-168 mV)**

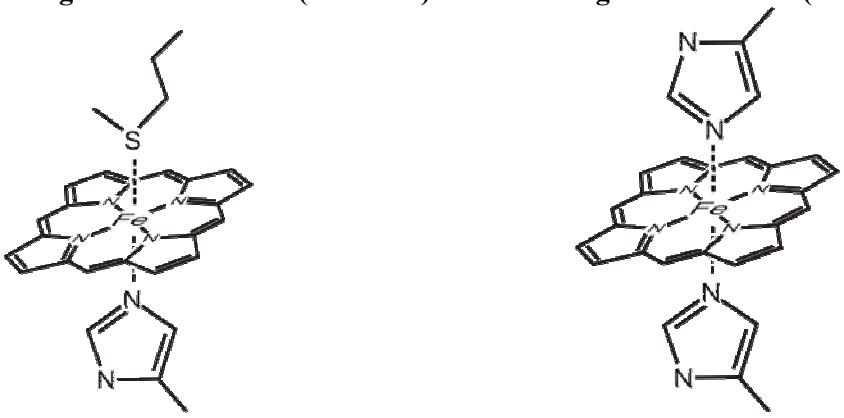

Figure 2. The native His/Met coordination of class I cytochromes *c* versus His/His coordination. Replacement of the native methionine by a histidine led to a midpoint potential decrease of nearly 400 mV!

While axial ligands do have a dramatic effect on the potential of the redox metal centre, other nearby amino acids that interact with the macro cycle ring also exert their influence. Another mutation (Met84His) produced in the same cytochrome c_{555} in an amino acid in the vicinity of the 6th axial ligand heme led to a small but significant drop in the potential to 181 ± 10 mV [Aubert, *et al.* 2001].

In fact, the axial coordination of the heme iron by amino acid side chains is only the more visible part of a broader network of interactions (be it hydrogen bonds, hydrophobic or electrostatic contacts) concerning surrounding protein residues and the heme (including the metal centre, delocalised macro cyclic π electron cloud, and the propionyl side-chains). Even slight alterations in this intricate heme environment have significant impact in the physical, chemical and spectral properties of the cytochrome [Rafferty, *et al.* 1990]. A particularly interesting example (because it permits the monitoring of structural alterations by spectroscopy) concerns the interaction between the π electron systems of a neighbouring phenylalanine residue (Phe87) and the heme macro cycle in *Sacharomyces cerevisiae* iso-1-cytochrome c [Pielak, *et al.* 1986]. The circular dichroic spectrum depends strongly on the proximity and orientation of the phenylalanine residue with respect to the plane of the heme ring. The extensive network of interactions around the heme ring contributes greatly to the compactness and overall protein structural stability; therefore, cytochromes are proteins particularly resistant to unfolding [Bartalesi, *et al.* 2004].

There are four major groups of cytochromes, classified accordingly to which type of heme prosthetic group is bonded and how it is attached [Moss 1988]. Thus, cytochromes *a* bind non-covalently an iron chelate of cytoporphyrin IX (or heme *a*).

Cytochromes *b* attach non-covalently a protoheme (the iron chelate of protoporphyrin IX [Caughey, *et al.* 1975]). Cytochromes *c* also bind a protoheme (Fig. 1), but covalently via thioether linkages between either or both of the vinyl side chains of heme *c* and cysteine residues of the protein. It should be noted that there are several variants of protoheme structures that may be bound to cytochromes *b* or *c*, therefore, regarding classification, the essential concept is whether or not the prosthetic group is covalently linked to the protein (Fig. 1). Finally, cytochromes *d* bind a tetrapyrrolic chelate of iron which has a lesser degree of double-bond conjugation than porphyrin. For example, the iron chelates of dihydroporphyrin (or chlorin [Timkovich, *et al.* 1985]), and tetrahydroporphyrin (isobacteriochlorin [Chang 1985]) are two forms of heme *d*.

Cytochromes c

Cytochromes *c* [Pettigrew and Moore 1987] are by far the most diversified and numerous family of cytochromes, and can be defined as proteins possessing one or more heme *c* (see Fig. 1) groups, which are linked to the polypeptide chain by sulphide covalent bonds with cysteine residues.[*] This covalent binding of the heme ring to the protein's cysteine residues is a unique feature amongst heme proteins [Wood 1983]. All *c*-type cytochromes thus have the signature CXXCH sequence in their primary structure. The two cysteines bind the vinyl groups of the heme ring, while the histidine serves as 5th axial ligand for the heme iron (Fe). The high thermal stability, high solubility, small size and a predominantly α-helical secondary structure allow the use of a variety of spectroscopic techniques for study. Finally, their intense red colour, which facilitates the purification, contributes to make this one of the most popular proteins among scientists [Bertini, *et al.* 2006].

Bacterial cytochromes *c* are either water-soluble proteins located in the periplasm of Gram-negative bacteria, or they are incorporated in a transmembrane protein whose heme-containing domain is on the external periplasmic side [Page, *et al.* 1998]. Therefore, they may be assumed to be absent from the cytoplasm, with few exceptions [Wood 1983]. Bacterial *c*-type cytochromes, upon translation, contain an N-terminal signal sequence recognized by the Sec system of protein export [Page, *et al.* 1998]. This

[*] http://www.chem.qmul.ac.uk/iubmb/etp/etp4.html#p443
http://metallo.scripps.edu/promise/CYTC.html

signal sequence is cleaved upon arrival to the periplasm. Bacteria generally possess a single set of biogenesis genes whose products are essential for heme attachment. Moreover, the biogenesis gene set is universal, providing heme attachment to every type of cytochrome *c* produced by the bacterium. Seemingly, heme attachment and folding occur only upon arrival to the periplasm [Page*, et al.* 1998]. Although heme attachment is predominantly mediated by enzymes called heme-lyases, there are observations of spontaneous heme attachment in thermophilic organisms [Page*, et al.* 1998] and in vitro for horse heart and *Paracoccus denitrificans* cytochrome c [Daltrop and Ferguson 2003].

Ambler was the first to attempt a systematic and comprehensive classification of *c*-type cytochromes [Ambler 1991] by grouping them into classes. Class I comprises soluble cytochromes *c* of mitochondria and bacteria, with the heme binding site (the CXXCH motif) near the N-terminus and a methionine (located about forty residues toward the C-terminus) as 6th ligand. The heme iron is typically low-spin and possesses a high redox potential (>200 mV). This rather broad class (to which the majority of cytochromes c belongs) includes:

- **Mitochondrial cytochrome *c*** (which is one of the most extensively studied proteins) is the prototypic *c*-type cytochrome. It is a monoheme protein with 103-112 residues, has a midpoint redox potential of 260 mV at physiological pH. A universal protein throughout the Eukaryota domain, it serves as electron carrier between complex III and the terminal oxidase (complex IV) in oxidative phosphorylation (Fig. 4 first panel). In animals, its release from the mitochondria is a signal for controlled cell death (apoptosis) [Lodish*, et al.* 2000a].
- **Bacterial cytochrome c_2** (also a monoheme) possesses the same tertiary folding as mitochrondrial cytochrome c but occasionally with several extra loops [Ambler 1991] and a slightly bigger size (varying from 100 to 120 residues)[*]. Cytochromes c_2 generally have somewhat higher potentials than mitochondrial cytochrome *c* (ranging between 260 and 450 mV; e.g. 345 mV in *Rhodobacter sphaeroides* [Venturoli*, et al.* 1998], 285 mV in *Blastochloris viridis*, 365 mV in *Rhodopseudomonas palustris* [Garau*, et al.* 2002]). Found mainly in the class of α-Proteobacteria (Table 1), cytochrome c_2 functions in the electron transport from

[*] NCBI-protein - http://www.ncbi.nlm.nih.gov/

cytochrome b/c_1 complex to the photosynthetic reaction centre in purple bacterial cyclic photophosphorylation (Fig. 4). It seems to be the only soluble carrier that is able to deliver electrons directly to the reaction centre photo-oxidized "special pair" (in reaction centres without tetraheme subunits).

- **Bacterial cytochrome c_5** is about 80-90 amino acids long, possesses a di-sulfide bridge in most species, and occasionally forms a dimer of two mono-heme subunits [Bertini, *et al.* 2006]. Both hemes have potentials somewhat higher than that of mitochondrial cytochrome *c*. Like cytochrome c_4, cytochrome c_5 has also been found in *Pseudomonas* species, and it is also thought to participate in a position close to the terminal oxidase in the oxidative phosphorylation electron transport chain [Ambler 1991].
- **Cytochrome c_6** is a mono-heme monomer, about 85 to 90 residues long (NCBI protein database), and has a redox potential of about 360 mV. Like cytochrome c_2 in bacteria, a similar function is attributed to cytochrome c_6 in chloroplasts of algae. Specifically, cytochrome c_6 transfers electrons at the high potential terminus of the eukaryotic photophosphorylation chain.
- **Bacterial cytochrome c_8**, formerly known as *Pseudomonas* cytochrome c_{551} type [Hochkoeppler, *et al.* 1997], includes now also cytochromes c_{552} and is about 80 residues long [Bertini, *et al.* 2006]. Cytochrome c_8 presents a potential of 285 mV in *Rhodoferax fermentans* [Hochkoeppler, *et al*. 1997]. Cytochrome c_8 was found to efficiently couple electron transfer from cytochrome b/c_1 complex to the photosynthetic reaction center in *Chromatium vinosum* [Samyn, *et al.* 1996]. It appears to be functionally diverse since, for example, in *Pseudomonas aeruginosa*, cytochrome c_{551} serves as an electron donor for the cd_1 nitrite reductase of this organism [Bertini, *et al.* 2006].
- **Cytochrome c_{555}** is thought to be exclusive of green photosynthetic bacteria (Chlorobiaceae family). It is around 80-90 residues in length and has a midpoint potential of about 150 mV. It provides electron transfer from the Rieske cytochrome *b* complex to the type I reaction center of green photosynthetic bacteria (Fig. 4)[*].
- **Bacterial cytochrome c_4** consists of two heme groups bound to a polypeptide chain about 190 amino acids long. These di-heme proteins are attached to the membrane on the periplasmic side [Pettigrew and Brown 1988, Hunter, *et al.* 1989,

[*] http://www.chem.qmul.ac.uk/iubmb/etp/etp4.html#p443

Wood 1983, Brown*, et al.* 1999]. There are reports showing that some mono-heme cytochromes present physicochemical properties and sequence homology reminiscent of the c_4 family [Brown*, et al.* 1999]. Cytochrome c_4 is characterized by a relatively high redox potential in both hemes (above ~250 mV) [Leitch*, et al.* 1985], a split α band, and a low α/β absorption ratio in the reduced form (it is ~1 in contrast with, for instance, mitochondrial cytochrome c where this ratio is ~2) in the visible absorption spectrum [Pettigrew and Brown 1988, Leitch*, et al.* 1985]. Heme-binding sites for both hemes are of the classical CXXCH type [Van Beeumen 1991], and the iron has His/Met axial coordination. The 6th axial methionine can easily be identified in the visible absorption spectra via the Met–Fe charge-transfer band at 695 nm, which appears exclusively in the oxidized state [Saraiva*, et al.* 1990, Santucci and Ascoli 1997]. The Soret band at (+)405/(-)416 nm in the visible circular dichroism (CD) spectrum was also assigned as a marker of the Met–Fe bond and can be detected in both redox states [Santucci and Ascoli 1997]. The first cytochrome c_4 was isolated from *Azotobacter vinelandii* [Tissieres 1956]. It has also been isolated from *Pseudomonas aeruginosa* [Hunter*, et al.* 1989], *Pseudomonas stutzeri* [Kadziola and Larsen 1997] and from *Acidithiobacillus ferrooxidans* (two different cytochromes c_4 from this species) [Giudici-Orticoni*, et al.* 2000]. In addition to the purified proteins, there are thirty-six sequences that have been identified in genomes as cytochromes c_4, including some genera of γ-proteobacteria (*Pseudomonas*, *Shewanella*, *Vibrio* and *Xanthomonas*) [Bertini*, et al.* 2006], and there are spectroscopic indications for more cytochromes c_4 in purple phototrophic bacteria (Table 1). Because this type of protein had only been purified from species of the order Pseudomonadales from the class of γ-Proteobacteria, and because these organisms lead an aerobic lifestyle, cytochrome c_4 was thought to participate in the oxidative phosphorylation chain; and, in view of its high redox potential, in a position close to the terminal oxidase (e.g. cytochrome *c* oxidase in *A. vinelandii*) [Rey and Maier 1997, Bertini*, et al.* 2006]. Our discovery of a cytochrome c_4 in a purple sulphur organism of the order Chromatiales [Branca*, et al.* 2007], which is predominantly anaerobic and photosynthetic, implies that this protein is capable of carrying out completely different roles in different electron transport systems.

species	c_2	c'	HiPIP	c_4	c_5	c_8	taxonomic group
Rhodopseudomonas palustris 37	+	+					α-proteobacteria; Rhizobiales; Bradyrhizobiaceae
Rhodopseudomonas palustris 6	+						α-proteobacteria; Rhizobiales; Bradyrhizobiaceae
Rhodopseudomonas rutila	+	+					α-proteobacteria; Rhizobiales; Bradyrhizobiaceae
Rhodobacter sulfidophila	+	+					α-proteobacteria; Rhodobacterales; Rhodobacteraceae
Rhodobacter sphaeroides	+	+					α-proteobacteria; Rhodobacterales; Rhodobacteraceae
Rhodobacter capsulatus	+	+					α-proteobacteria; Rhodobacterales; Rhodobacteraceae
Rhodopbacter blastica	+	+					α-proteobacteria; Rhodobacterales; Rhodobacteraceae
Rhodospirillum rubrum	+	+					α-proteobacteria; Rhodospirillales; Rhodospirillaceae
Rhodospirillum photometricum	+	+					α-proteobacteria; Rhodospirillales; Rhodospirillaceae
Rhodospirillum molischianum	+	+					α-proteobacteria; Rhodospirillales; Rhodospirillaceae
Rhodomicrobium vannielii	+		+				α-proteobacteria; Rhizobiales; Hyphomicrobiaceae
Rhodopseudomonas viridis	+						α-proteobacteria; Rhizobiales; Hyphomicrobiaceae
Rhodopseudomonas acidophila	+						α-proteobacteria; Rhizobiales; Bradyrhizobiaceae
Rhodopseudomonas globiformis	+		+				α-proteobacteria; Rhodospirillales; Acetobacteraceae
Rhodospirillum salexigens	+	+					α-proteobacteria; Rhodobacterales; Rhodobacteraceae
Rhodospirillum centenum	+						α-proteobacteria; Rhodospirillales; Rhodospirillaceae
Rhodopseudomonas marina	+	+	+				α-proteobacteria; Rhizobiales; Rhodobiaceae; Rhodobium
Rhodopseudomonas H1R		+			+		α-proteobacteria; Rhizobiales; Rhodobiaceae; Rhodobium
Rhodospirillum salinarum		+	+				α-proteobacteria; Rhodospirillales; Rhodospirillaceae
Rhodocyclus tenuis		+	+	+		+	β-proteobacteria; Rhodocyclales; Rhodocyclaceae
Rhodocyclus purpureus		+				+	β-proteobacteria; Rhodocyclales; Rhodocyclaceae
Rhodocyclus gelatinosus		+	+	+		+	β-proteobacteria; Burkholderiales
Ectothiorhodospira halophila		+	+		+		γ-proteobacteria; Chromatiales; Ectothiorhodospiraceae
Ectothiorhodospira halochloris					+		γ-proteobacteria; Chromatiales; Ectothiorhodospiraceae
Ectothiorhodospira abdelmalekii					+		γ-proteobacteria; Chromatiales; Ectothiorhodospiraceae
Ectothiorhodospira vacuolata		+	+	+			γ-proteobacteria; Chromatiales; Ectothiorhodospiraceae
Chromatium vinosum		+	+	+		+	γ-proteobacteria; Chromatiales; Chromatiaceae
Chromatium gracile		+	+	+			γ-proteobacteria; Chromatiales; Chromatiaceae
Thiocapsa roseopersicina		+	+	+			γ-proteobacteria; Chromatiales; Chromatiaceae
Thiocapsa pfennigii		+	+	+			γ-proteobacteria; Chromatiales; Chromatiaceae

α-proteobacteria without tetraheme
α-proteobacteria with tetraheme
β-proteobacteria with tetraheme
γ-proteobacteria with tetraheme

Table 1. Distribution of selected mobile redox proteins in purple phototrophic bacteria. The table is based on Bartsch 1991, with additional information from Bertini, *et al.* 2006, Samyn, *et al.* 1996, and Osyczka, *et al.* 2001. α-proteobacteria have been divided in two groups according to whether they possess a tetraheme cytochrome in the photosynthetic reaction center or not (see also Figure 4). A + sign indicates presence, confirmed by sequence or inferred from spectral data [Bartsch 1991]. Blank space means that the protein has not been observed.

Other cytochromes do not belong to Ambler's class I [Ambler 1991]:

- **Cytochrome c_1** is a low-spin monoheme cytochrome with His/Met coordination, and has a midpoint redox potential of 285 mV in *Rhodopseudomonas sphaeroides* [Gabellini, *et al.* 1982]. It is about 230-250 residues long*. This cytochrome does not belong to class I because it is an integral part of the transmembrane complex cytochrome b/c_1. Cytochrome b/c_1 complex is conserved in all bioenergetic electron transport chains, including aerobic oxidative phosphorylation (where it is oxidized by cytochrome c in mitochondria) and oxygenic and anoxygenic photophosphorylation (where it is oxidized by a variety of soluble periplasmic electron carriers, many of which are class I cytochrome *c*'s - see Fig. 4 last panel).
- **Cytochrome c_3** is a relatively small (about 12000 Da) tetraheme cytochrome with a His/His axial coordination and therefore a low redox potential (less than -200 mV [Tsapin, *et al.* 1996, Benosman, *et al.* 1989]). It is found in sulfate-reducing bacteria, where it is are thought to function as electron acceptor from periplasmic hydrogenases, thereby participating in the process of sulfite reduction [Bertini, *et al.* 2006].
- **Reaction centre associated tetraheme cytochrome** is an integral part of the photosynthetic transmembrane complex (Fig. 4, last panel) in many purple phototrophic bacteria. Incidentally, a reaction center of this type found in *Rhodopseudomonas viridis* (recently renamed to *Blastochloris viridis*) was the first membrane protein complex to have been successfully crystallized and to have had its 3D structure solved by X-ray diffraction [Deisenhofer and Michel 1989]. The tetraheme cytochrome is present in the reaction centre of only some phototrophic α-Proteobacteria, but seems to be ubiquitously present in phototrophic β- and γ-Proteobacteria (see Table 1 and related references). The tetraheme subunit faces the periplasmic side of the membrane, where it accepts electrons from a wide variety of soluble periplasmic electron carriers [Meyer, *et al.* 1993, Meyer and Cusanovich 2003, Lieutaud, *et al.* 2003, Lieutaud, *et al.* 2005]. Therefore, it is likely to be a fundamental crossroad junction point for distinct bacterial redox metabolisms [Osyczka, *et al.* 1998, Osyczka, *et al.* 1999 and Osyczka, *et al.* 2001] (Fig. 4, last panel). Tetraheme donates electrons to the photo-oxidized bacteriochlorophyl "special pair" within the reaction centre core. The four hemes

* NCBI-protein - http://www.ncbi.nlm.nih.gov/

present alternating high and low redox potentials, i. e. -60 mV, +320 mV, +20 mV, +380 mV [Baymann and Rappaport 1998] in a polypeptide chain of 336 amino acid residues* in *Blastochloris viridis*. The structure of the reaction centre associated cytochrome appears to be strongly species dependent. In *Rhodovulum sulfidophilum* [Masuda, *et al.* 2002] this cytochrome is a triheme; and in other cases there are indications for more than four heme redox centres: e.g. six hemes in *Chromatium minutissimum* [Chamorovsky, *et al.* 1998] and five hemes in *Chromatium vinosum* [Nitschke, *et al.* 1993].

- **Bacterial cytochrome *c*'** is a high-spin cytochrome due to its pentacoordination with a single histidine axial ligand. Usually occurring as a dimer, the monomer has a molecular mass of about 14 kDa and a midpoint redox potential close to 0 mV at pH 7. It is found in purple photosynthetic bacteria, and also in nitrate-reducing Pseudomonadales species, overall being the most widely distributed cytochrome in bacteria*.
- **Cytochrome *f*** is around 300 to 350 amino acids long (NCBI protein database) and has an analogous role in oxygenic photosynthetic membranes to that of cytochrome c_1 in mitochondria and purple bacteria (Fig. 4) [Prince and George 1995]. Cytochrome *f* is similarly inserted in a transmembrane complex (cytochrome *b*/*f*). It has the typical CXXCH heme binding motif of cytochromes *c*, with the histidine serving as 5th ligand. However, cytochrome *f* has an unusual 6th ligand, the α-nitrogen atom of the N-terminal tyrosine residue. It presents a very high midpoint redox potential, e. g. 297 mV in *P. laminosum* and 370 mV in *C. reinhardtii* [Ponamarev, *et al.* 2000]♣.
- **Flavocytochrome *c*** is a dimer in which the small subunit is a di-heme cytochrome *c* and the large subunit contains a flavin group. [Castillo, *et al.* 1994]. The cytochrome subunit is about 170 amino acids long [Chen, *et al.* 1994] and has a redox potential of about 10 mV [Zorin and Gogotov 1980]. Flavocytochrome has sulphide dehydrogenase activity *in vitro*, but it has been shown in *Chromatium vinosum* that the protein is not required for sulphide oxidation *in vivo* [Reinartz, *et al.* 1998], and consequently its function remains unclear.

---

* http://www.chem.qmul.ac.uk/iubmb/etp/etp4.html#p443
http://metallo.scripps.edu/promise/CYTC.html

♣ http://metallo.scripps.edu/promise/CYTF.html

Electron transfer chains (biological energy conversion pathways)

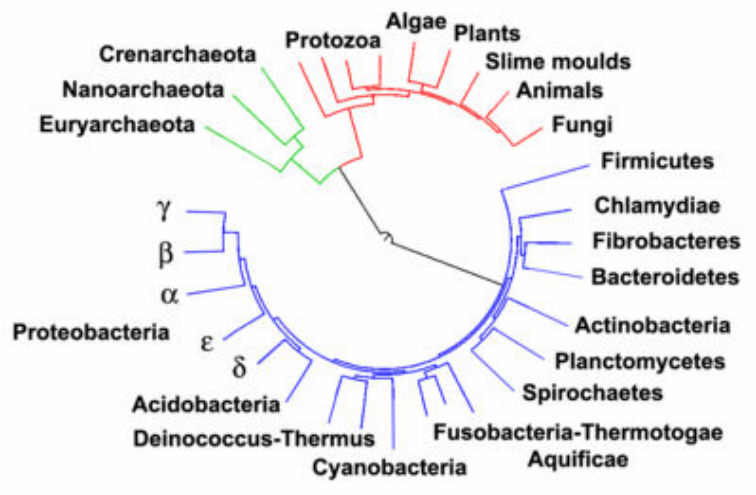

Figure 3. Phylogenetic tree of life [Ciccarelli, et al. 2006]. Purple bacteria are phototrophic Proteobacteria from the α, β or γ classes. Purple non-sulphur bacteria generally belong to either α or β classes, whereas purple sulphur bacteria are γ-proteobacteria.

Cytochromes constitute a universal motif in membrane-appended electron transport chains of biological organisms. As of 1991, the inventory of purified bacterial cytochromes *c* was very substantial, and additionally, there were indications for the presence of many more, particularly in purple bacterial species [Bartsch 1991] (Table 1). The diversity of redox carriers reflects the different variations in metabolism found in the superkingdom of Eubacteria (blue domain in Fig. 3), and probably is a consequence of the different environmental niches inhabited, particularly with respect to the variety of available electron donors. Purple bacteria show a striking interspecies diversity of cytochromes (including c_2, *c*', low spin iso *c*', flavocytochromes, low potential cytochrome *c*, c_{551}, c_{552}, c_4, c_5) in their repertoire of electron carrier proteins [Bartsch 1991] (Table 1). In Fig. 4 electron transport proteins are portrayed in their function as integral components of biological electron transport chains.

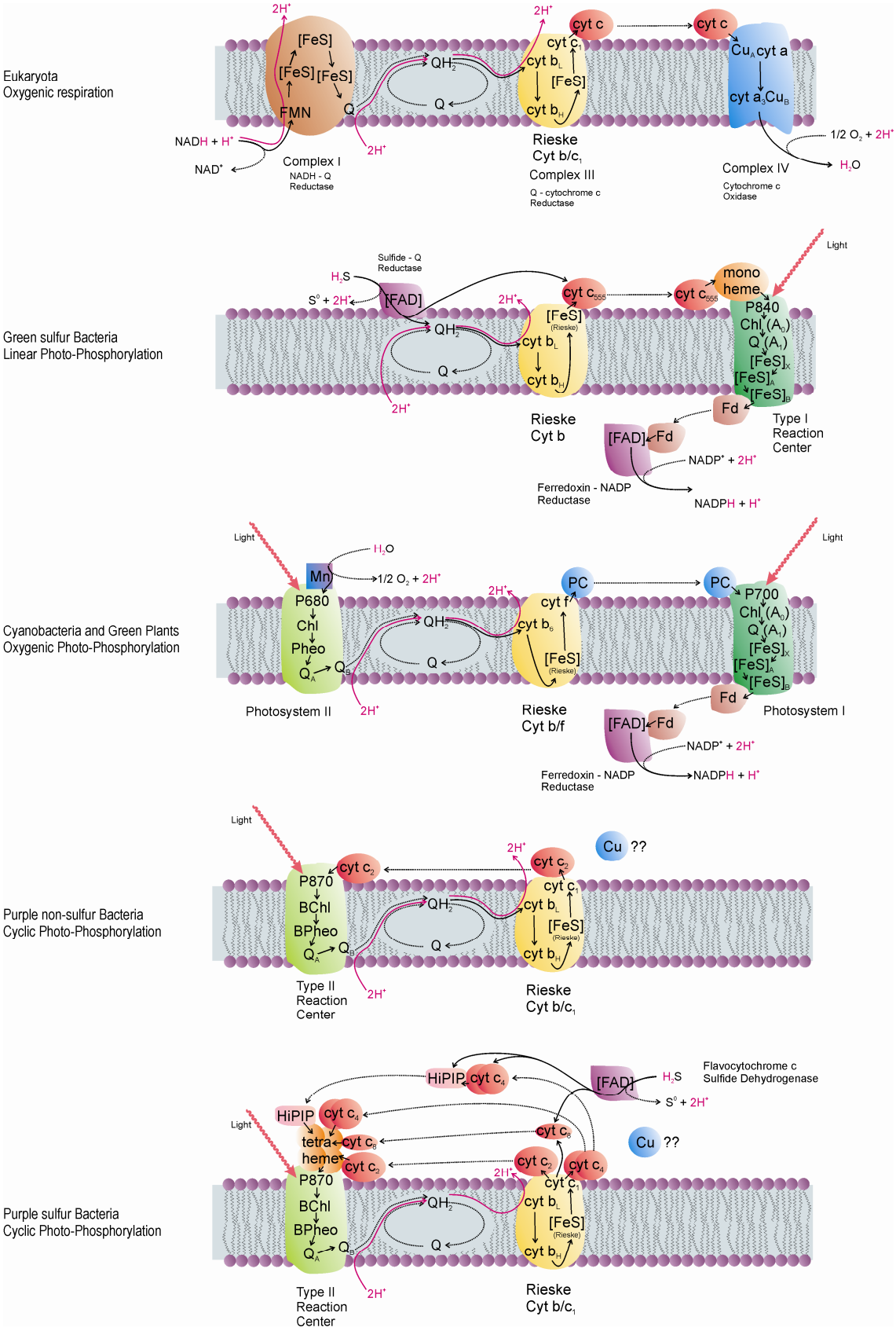

Figure 4. Electron transport chains appended on biological membranes.

Figure 4. (continued) Whereas the positions and functions of the electron carriers on the top panels are well established, the last panel remains mostly tentative with respect to the precise positions and interrelations of the periplasmic soluble electron carriers as the function of most of these proteins still remains largely unclear. The provisional scheme was based on the information contained in Table 1 and on Meyer and Cusanovich 2003.
⟶ electron transfer; ⟶ proton transfer; ⇢ spatial movement; ⇝ photon;

NAD – Nicotinamide adenine dinucleotide
NADP – Nicotinamide adenine dinucleotide phosphate
FAD – Flavin adenine dinucleotide
FMN – Flavin mononucleotide
[FeS] – Iron sulphur cluster
Q – Quinone
Cyt – Cytochrome
P680, P700, P840, P870 – Reaction centre Pigments
Chl – Chlorophyll
BChl – BacterioChlorophyll
Pheo – Pheophytin
BPheo – Bacteriopheophytin
Mn – Manganese containing oxygen evolving centre
PC – Plastocyanin
Fd – Ferredoxin
Monoheme, tetraheme – cytochromes
Cu – Copper protein
HiPIP – High potential iron-sulphur protein

The main types of biological membrane-supported electron transport chains are depicted in Fig. 4. Several conserved features may be observed. Firstly, all share the coupling of electron transfer along the membrane with proton transfer across it; which is in keeping with the universal chemiosmotic theory of using a proton-motive force to generate ATP [Mitchell 1961]. Secondly, cytochrome b/c_1 (or b_6/f) transmembrane protein complex exists in some form or another in most living organisms (all eukaryotes and nearly all bacteria). Likewise, all electron transport chains use quinones (be it ubiquinone, menaquinone or plastoquinone [Voet and Voet 1995]) as lipophilic electron carriers in the electron donor position to cytochrome b/c_1. Incidentally, the one proton pump common to all electron transport chains lays therein this interplay between the quinone cycle and cytochrome b/c_1.

Electron transport chains may be linear, as in the first three panels of Fig. 4, or cyclic (whereby the proton pump runs inexhaustibly as long as photo-excitation at the reaction centre occurs) as in purple bacteria. Whilst cyclic photo-phosphorylation works as an unlimited supply of ATP, linear photo-phosphorylation is capable of generating

reducing power in the form of NADPH, which is readily usable by the cell in the synthesis of organic compounds, for example. In order to reduce the final electron acceptor ($NADP^+$), an initial electron donor (e. g. H_2O or H_2S) is required. Purple bacteria acquire reducing power by wiring the electron donors (which can be H_2S in purple sulphur bacteria) through the photo-phosphorylation apparatus to the final electron acceptors thereby converting a cyclic process into a linear one. Exactly how this is done remains unclear, but it is surely mediated by soluble electron carriers on the periplasmic side of the membrane.

Generally, the mobile electron carriers are overlaid on the polar surface of the membrane, in electron acceptor position from cytochrome b/c_1. Most of these carriers belong to the cytochrome *c* family, but in oxygenic photophosphorylation (the dominant form of photosynthesis on Earth, occurring in green plants and in cyanobacteria) the copper protein plastocyanin takes on this role instead (wherein it accepts electrons from the cytochrome b_6/f complex). The variety of soluble electron carriers is still not known to its full extent, but thus far mitochondrial cytochrome *c*, green sulphur bacteria cytochrome c_{555}, and purple bacteria cytochrome c_2 have been shown to perform this function [Meyer and Cusanovich 2003]. In purple bacteria it is thought that there are more candidates to the periplasmic electron carrier role, like cytochrome c_8 and High Potential Iron-Sulphur Protein (HiPIP) [Van Driessche*, et al.* 2003, Schoepp*, et al.* 1995, Nagashima*, et al.* 2002, Meyer and Cusanovich 2003].

Novel cytochrome c_4 from Thiocapsa roseopersicina

In the present work, we discuss the discovery and study of a novel cytochrome *c* from the species *Thiocapsa roseopersicina*. We have characterized the cytochrome through a combination of absorption spectroscopy and circular dichroism (CD) with redox potentiometry and also by using differential scanning calorimetry (DSC). The data gathered place this protein in the c_4 class of cytochromes. We have used Mass Spectrometry to determine its primary structure, which confirmed this cytochrome to be of the c_4 type. It is the first purified and identified cytochrome c_4 from an anaerobic phototrophic bacterium.

Anaerobicity and the temperature of the environment are crucial to the growth of *T. roseopersicina*, and we have therefore studied cytochrome c_4 under varying ambient conditions. Similarly to the hydrogenase from *T. roseopersicina*, cytochrome c_4 proved

to be a heat-tolerant protein if maintained under anaerobic conditions [Gogotov, *et al.* 1976, Zorin and Gogotov 1982]. We clarified that the oxygen initiates an unfolding of the protein at high temperatures, most likely through direct binding to the heme's sixth coordination site, which was left vacant after dissociation of the native Met ligand.

Cytochrome c_4 proteins are thought to participate in aerobic respiratory pathways, in a position close to the terminal oxidase of the electron transport chain [Pettigrew and Brown 1988, Hunter, *et al.* 1989]. The discovery of such a cytochrome in an anaerobic photosynthetic organism throws doubt upon this assumption. In *T. roseopersicina*, cytochrome c_4 must participate in photosynthesis instead.

We have proposed that cytochrome c_4 functions as a mobile electron carrier on the periplasmic side of the membrane in *Thiocapsa roseopersicina* (Fig. 4 last panel). This conclusion can probably be extended to most Chromatiaceae species since indications given in Table 1 suggest that cytochrome c_4 is present in several species of this family.

As a side note, several blue copper proteins are known to exist in the periplasm of some purple bacteria (Fig. 4) [Meyer and Cusanovich 2003], and while their role remains unknown, it is speculated that they may just as well participate in the electron transport chain in a similar way to the distant relative copper protein plastocyanin of cyanobacteria.

Mass Spectrometry and protein sequencing

With the arrival of large-scale genomic sequencing, primary structure of proteins is usually inferred from the DNA sequence. Even for organisms whose genome is unknown, such as *T. roseopersicina*, the mainstream approach is to search for the gene of the protein in question, determine the nucleotide sequence and translate that to amino acid sequences [Lodish, *et al.* 2000b]. However, this indirect approach has a major drawback in that it does not take into account posttranslational events such as polypeptide chain cleavages and amino acid modifications [Standing 2003]. Moreover, determination of a gene sequence from an organism with yet-unknown genome is an elaborate procedure [Lodish, *et al.* 2000b]. Direct "de novo" protein sequencing by mass spectrometry is a straightforward and unequivocal method.

Direct sequence determination generally requires that the protein be hydrolyzed to smaller component peptides that can be individually sequenced and later overlapped to yield the complete protein primary structure. [Kendrew 1994] This "bottom up"

approach is usually achieved through enzymatic digestion by using endoproteases such as trypsin. Edman degradation [Edman 1950], the first method that utilized this enzymatic strategy, requires considerable time, effort and material, mainly because every single peptide to be sequenced must be purified before analysis [Biemann and Papayannopoulos 1994]. Moreover, the degradation reaction cycles often run into chemical obstacles, such as for peptides with blocked N-terminus [Johnson and Biemann 1987].

Direct peptide sequencing by mass spectrometry (MS) became possible with the development of "soft" ionization techniques that are able to generate intact molecular ions without significant fragmentation, and through the use of tandem MS (MS/MS) [Standing 2003, Biemann and Martin 1987].

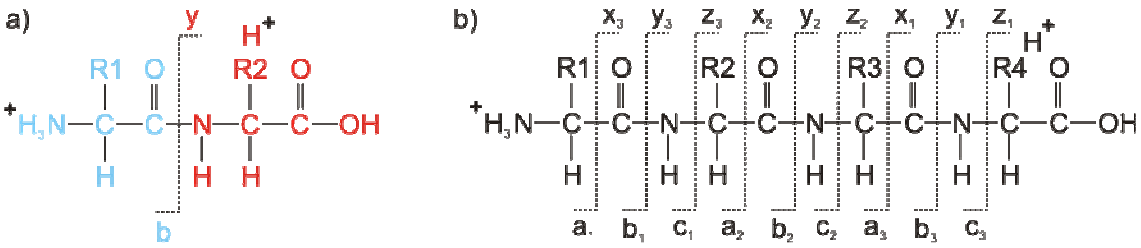

Figure 5. Peptide fragmentation nomenclature as developed by Roepstorff and Fohlman and later modified by Biemann [Roepstorff and Fohlman 1984, Biemann 1988, Biemann 1990].*
a) The dipeptide shown has an amino group on side chain R2 as well as on the N-terminus, and thus acquires two protons in acidic solution. The resulting doubly charged ion is a parent ion when inserted into the collision cell of a Mass Spectrometer, and has a high chance of fragmenting at the peptide bond position thereby yielding two product ions, the N-terminal b **ion and the C-terminal** y **ion.**
b) Fragmentation inside the collision cell is a random process, and even though breakage at the peptide bond is the most likely, other backbone bonds are also susceptible. The doubly charged tetrapeptide ion shown illustrates the different types of product ions that may be obtained. Thus, a, b **and** c **ions are N-terminal product ions whereas** x, y **and** z **ions are C-terminal product ions. Low energy fragmentation methods such as Collision Induced Dissociation (CID) preferably generate** a, b **and** y **product ions, whereas high energy methods like Electron Capture Dissociation (ECD) tend to produce** c **and** z **ions.**

For tandem MS, it is important first of all to isolate the population of one intact parent ion, and only then proceed with fragmentation inside the collision cell of the Mass Spectrometer (Fig. 5). The subsequent detection of the resulting product ions

* http://www.ionsource.com/tutorial/DeNovo/introduction.htm

yields the MS/MS (or MS^2) spectrum whose interpretation provides the amino acid sequence of the original parent peptide ion (Fig. 6).

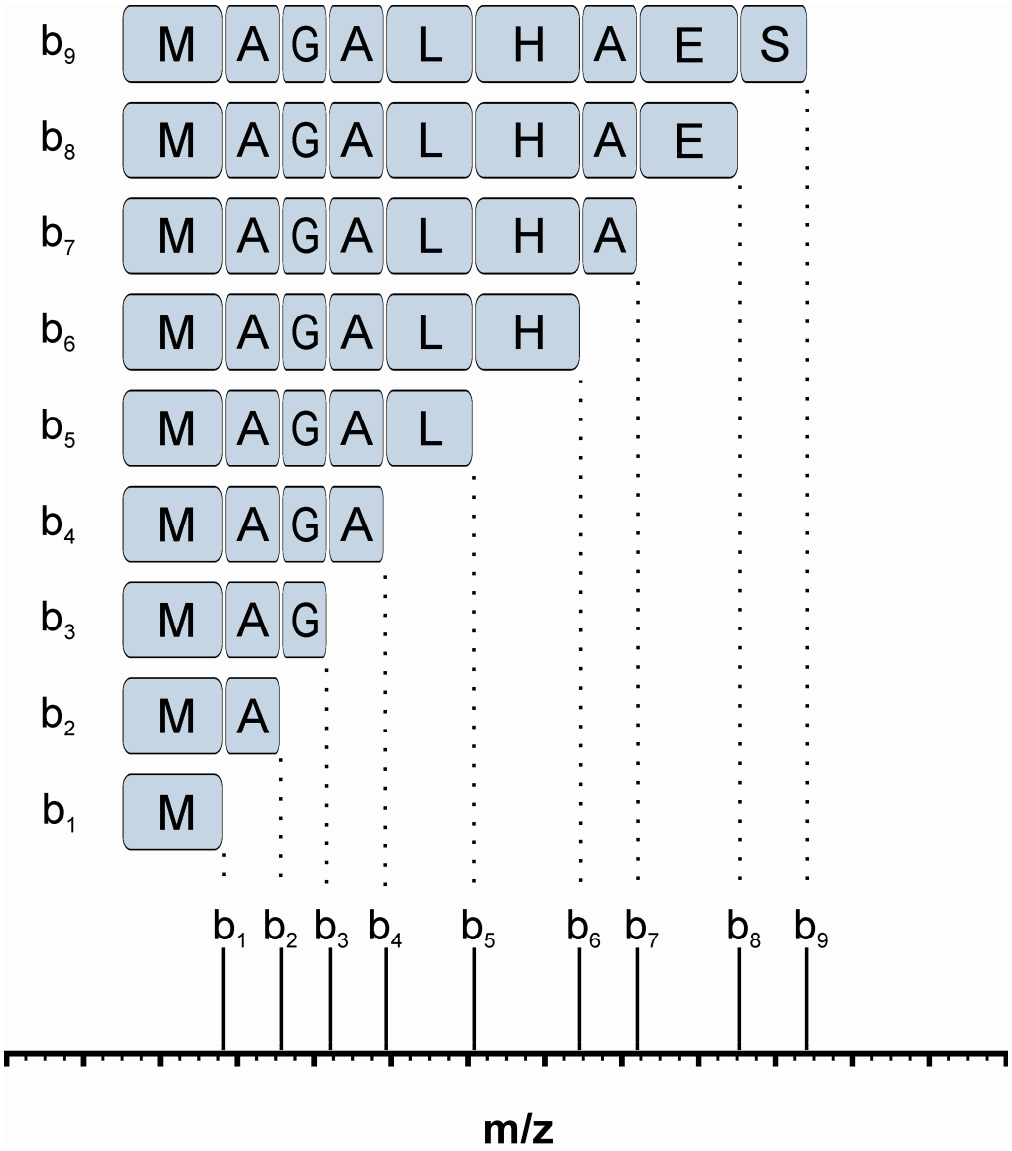

Figure 6. CID of the parent ion MAGALHAES would produce a wide range of product ions that ideally would include complete series of b **and** y **ions. The complete series of** b **ions and respective peaks in a hypothetical MS/MS spectrum are shown in the figure. The complementary series of** y **ions normally appears in conjunction with the** b **ion series in a real tandem mass spectrum.**

MS-based methods can cope better with peptide mixtures and with N-terminal modifications and provide higher sample throughput and sensitivity; therefore, they have become preferred over the classical Edman degradation methods. The first protein that was sequenced entirely by MS was a 11750-Da thioredoxin from *Chromatium vinosum* in 1987 [Johnson and Biemann 1987]. After digestion with trypsin, the resultant peptide mixture was separated by high-performance liquid chromatography

(HPLC), and the fractions were collected and analyzed by fast atom bombardment (FAB) MS and MS/MS. The interpretation of MS/MS spectra of the obtained 14 tryptic peptides (that spanned the entire protein) was done manually. A few related proteins of similar size were also sequenced later by the use of this strategy [Biemann and Papayannopoulos 1994, Mathews, *et al.* 1987, Hopper, *et al.* 1989]. The introduction of electrospray ionization (ESI) [Fenn, *et al.* 1989] and matrix-assisted laser desorption/ionization (MALDI) [Karas and Hillenkamp 1988] has permitted detection of very large, intact proteins through ionization with considerably less fragmentation. Another important development has been demonstrated in MS/MS technology, e.g. by the commercial availability of various hybrid instruments [Domon and Aebersold 2006] such as the linear ion-trap (IT) coupled with a Fourier-transform ion-cyclotron resonance (FTICR) mass spectrometer used in the present study. This instrument combines the flexibility of the linear IT with the extremely high accuracy and resolving power of the FTICR [Olsen and Mann 2004]. The possibility to perform online liquid chromatography (LC), and the automation brought about by improvements in computer-controlled data acquisition and processing, add to the speed and ease of use. In addition, the improvements in software designed to interpret MS/MS data in an automated fashion [Pevtsov, *et al.* 2006] have facilitated enormously the task of de novo peptide sequencing that had to be done manually in the 1980s and 1990s. Finally, the introduction of electron capture dissociation (ECD) [Horn, *et al.* 2000] technique for FTICR mass spectrometers yields the promise of "top down" sequencing (the analysis of intact proteins) [Zubarev, *et al.* 2000]. However, the "bottom up" approach remains the "proven" method when complete sequence information is required [Whiteaker, *et al.* 2004, Cohen, *et al.* 2005]. We have accomplished "de novo" sequencing of an entire 21-kDa cytochrome c_4 exclusively by ESI and tandem mass spectrometry. To our knowledge, this is the largest protein that has been fully sequenced by MS.

Aims of the study

To purify and characterize proteins with redox centers from purple sulfur photosynthetic bacteria in order to shed some light over the electron transport pathways within the metabolism of these organisms.

Particularly, to characterize a novel periplasmic soluble cytochrome *c* from *Thiocapsa roseopersicina*:

- To determine its primary structure, by using the recent advances in Mass Spectrometry.
- To determine the heme content of the cytochrome, and the respective midpoint potential of each redox site.
- To determine its sensitiveness to oxygen, given that the organism is photosynthetically active only under anaerobic conditions, but can survive in aerobic conditions; and to determine its thermal stability, given that temperature is another sensitive parameter to the survival of the organism. In sum, to elucidate the interrelation of these two parameters in their influence on the cytochrome structure.
- To elucidate the phylogenetics of this protein.

Finally, to use all of the gathered structural information in order to clarify the function of this cytochrome and on a broader sense to gain insight into the electron transport chains of phototrophic bacteria.

Materials and Methods

Purification procedure

Thiocapsa roseopersicina strain BBS was isolated from an estuary of the White Sea [Bogorov 1974] and was a kind gift from E. N. Kondrat'eva (Moscow State University, Russia). The cells were grown under standard photosynthetic conditions as described previously [Bagyinka, *et al.* 2003]. Cells were harvested in the late-logarithmic phase of growth and were stored at -20 ºC.

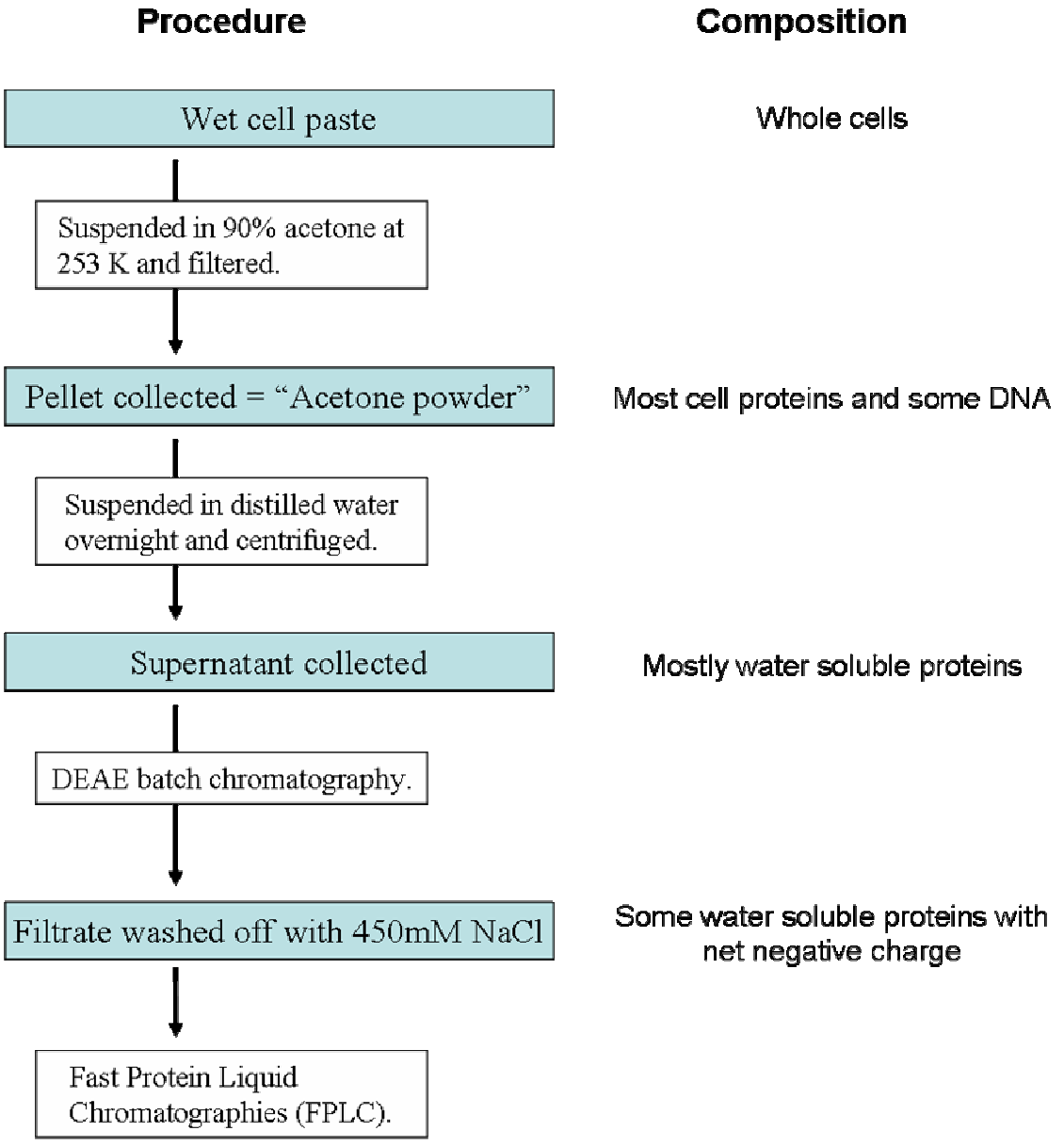

Figure 7. Stages of protein purification.

The wet cell paste was re-suspended in cold (-20 ºC) acetone/water (90%/10%) mixture and filtered. The acetone removed the lipophyl cell components (including carotenoids, which are abundant in purple bacteria). The procedure was repeated until the filtrated acetone became colourless. The pellet was finally washed with pure (100%) acetone, dried and stored at -20 ºC.

The pellet, which we denominate "acetone powder", contains mostly proteins, but also some DNA. For purification, 15 g of "acetone powder" was re-suspended in 1 L of distilled water and stirred overnight at 4 ºC. The slurry was centrifuged at 20 000 *g* for 2 h and the supernatant, which contains most of the water soluble proteins, was used in a batch chromatography with 60 g of Whatman DEAE (diethylaminoethyl) cellulose DE-52. After 1.5 h of binding while shaking, the beads were filtered off and re-suspended in about 300 mL of 450 mM NaCl and 20 mM Tris–HCl pH 7.5 solution. After shaking for 1 h, the suspension was filtered and the filtrate collected. This filtrate is abundant in metalloproteins, including the soluble hydrogenase, at least one blue copper protein and several cytochromes.

Ammonium sulfate $(NH_4)_2SO_4$ (final concentration 10%(w/w)) was added to the filtrate and it was loaded onto a Butyl-Sepharose column (Amersham XK 50, 5 cm diameter, 6 cm length). Fractions were eluted by a gradient from 10 to 0% $(NH_4)_2SO_4$ in 1 mM Tris–HCl pH 7.5 using the Amersham Biosciences FPLC (fast protein liquid chromatography) system. The next purification step was ion-exchange chromatography with 20 ml Q-Sepharose Fast Flow in an Amersham HR 16/20 column using 20 mM Tris–HCl pH 7.5 and 0–1 M NaCl gradient for elution. Another hydrophobic chromatography (1 mM Tris–HCl pH 7.5, 10–0% $(NH_4)_2SO_4$) followed using 8 ml ToyoPearl Phenyl-650S in an Amersham HR 10/15 column. Finally, 1 ml Q-Sepharose Fast Flow in an Amersham HR 5/10 column was used with a 20 mM Tris–HCl pH 8.5 buffer system and a 0–1 M NaCl gradient for elution.

It should be noted that a buffer capacity of 1mM Tris is not enough to maintain the pH near to 7.5 upon addition of 10% $(NH_4)_2SO_4$ due to the acidic character of the ammonium ion. The real pH of the 10% $(NH_4)_2SO_4$, 1mM Tris with nominal pH 7.5 solution is in fact about pH 6. What is important, however, is that this fact does not disturb the performance of the hydrophobic chromatographies.

The purity and molecular weight were determined by SDS–PAGE. We used 10% and 12% polyacrylamide gels as well as 10–20% polyacrylamide gradient gels [Maniatis*, et al.* 1982].

Heme content and concentration determination

The protein concentration was estimated both by measuring the optical density of the 280 nm absorption peak and by using the Bradford method [Bradford 1976]. In both cases BSA (bovine serum albumin, Sigma) was used as standard. The pyridine hemochromogen method was used to determine the exact heme concentration [Berry and Trumpower 1987] that was then compared to the estimated protein concentration, thus yielding the heme content. Accordingly, the protein concentration estimations were corrected with the knowledge of exact heme concentration and content.

Determination of primary structure by Mass Spectrometry

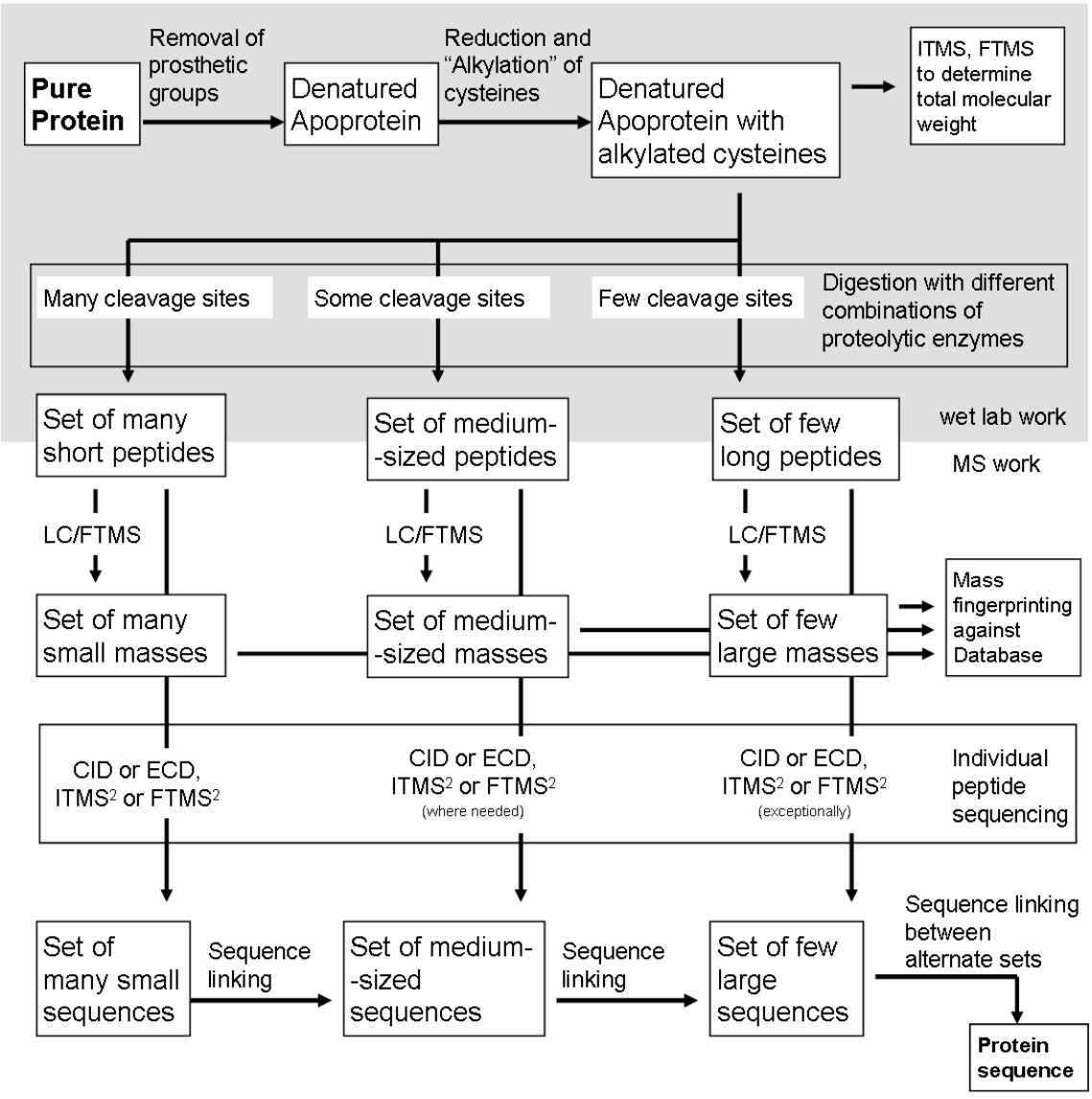

Figure 8. General outline of the "bottom up" approach for "de novo" protein sequencing by LC/DI-CID/ECD-IT-FTICR Mass Spectrometry.

Removal of heme groups

The first step in preparing the protein for mass spectrometric analysis was to remove the prosthetic heme groups [Ambler and Wynn 1973]. About 1 mg of pure cytochrome c_4 was added into a 5 ml solution containing 8 M urea, 1 M HCl and 2 mg $HgCl_2$, which was shaken overnight at 37 ºC (Fig. 9). The denatured apoprotein was then separated from the salts and the loose heme molecules by gel filtration (20 × 1 cm Sephadex G-25 column) in 50 mM ammonium acetate buffer. The ammonium acetate buffer was used because it does not contain any Na^+, K^+ or Cl^-, which would cause problems to the interpretation of mass spectra.

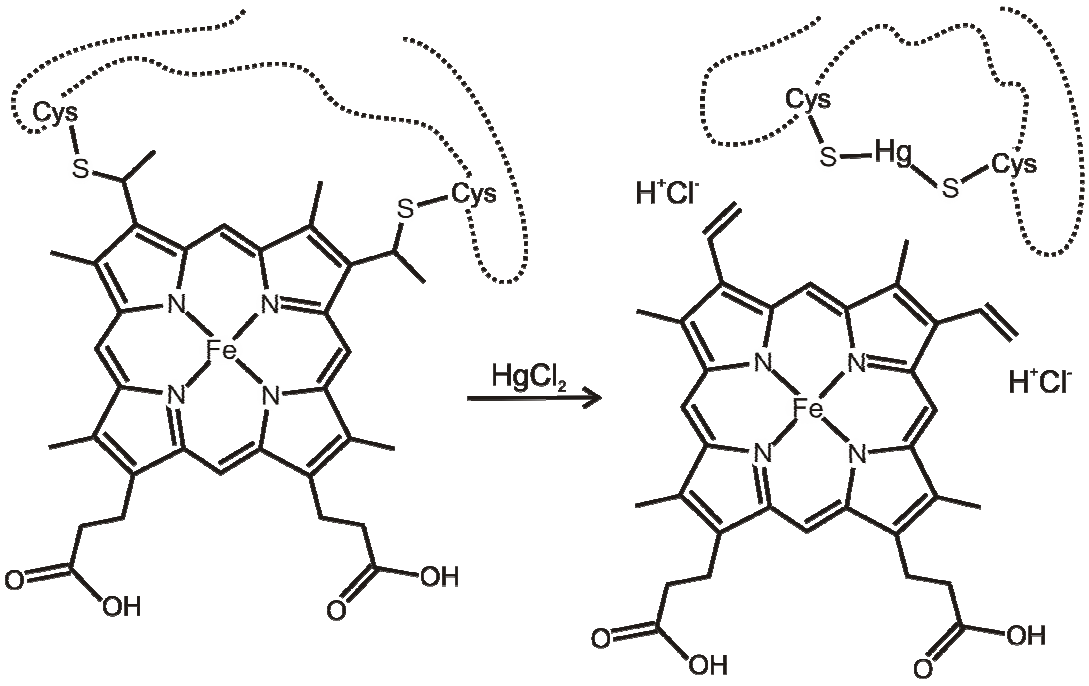

Figure 9. The denaturing effect of urea on the protein conformation exposes the heme-protein covalent bonds to reaction with mercury chloride, leading to heme detachment.

Reduction and S-carbamidomethylation ("alkylation")

About 20-fold excess in mass of dithiothreitol was added to reduce the denatured apoprotein, and incubation proceeded at 50 °C for 20 min. Dithiothreitol is a strong reductant with a redox potential of -330 mV at pH 7, and is capable of reducing all disulphide bonds at cysteinyl residues into thiol groups (Fig. 10).

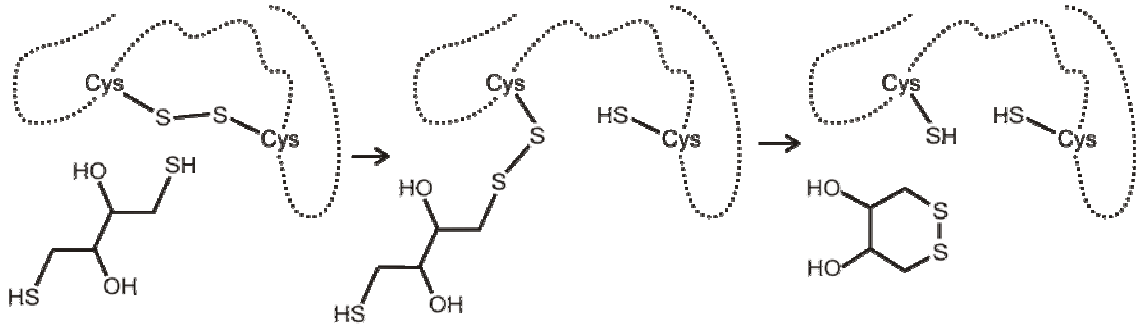

Figure 10. Reduction of a disulphide bond between two cysteines by dithiothreitol

After this reduction step, the protein was further treated with iodoacetamide (50-fold excess in mass) at room temperature for 30 min in darkness (Fig. 11). The resulting carbamidomethylated apoprotein solution was then desalted by repeated centrifugation in Microcon filters (PM5, Millipore, Billerica, MA, USA) at 8000 *g*.

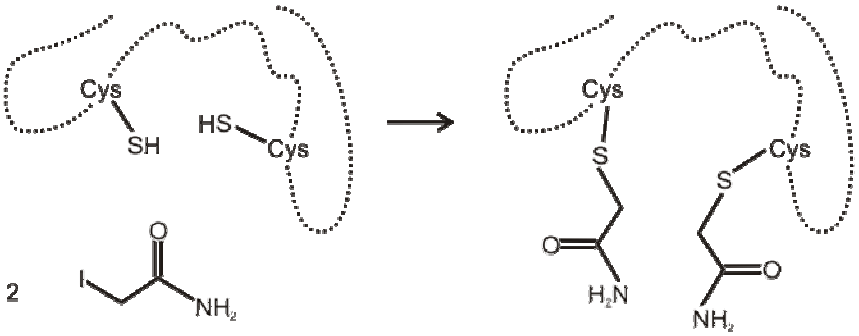

Figure 11. Carbamidomethylation of cysteinyl thiol groups by iodoacetamide.

Proteolytic digestions

Four different peptide mixtures (referred to in the text below as ”digest sets”) were prepared by digestion of the reduced and *S*-alkylated protein with three different proteolytic enzymes using a 1 : 50 enzyme : substrate weight ratio. The tryptic set was generated by incubation with trypsin (Promega, Madison, WI, USA) for 24 h at 37 °C in 50 mM ammonium bicarbonate (pH 7.8). The Lys-C set was prepared by incubation with endoproteinase Lys-C (Roche, Indianapolis, IN, USA) for 18 h at 37 °C in 25mM Tris (pH 8.5) and 1 mM EDTA. The Glu-C set was generated by incubation with endoproteinase Glu-C (Roche) for 18 h at 25 °C in 25mM ammonium bicarbonate buffer (pH 7.8). All digestions were terminated by adding 0.1% (v/v) acetic acid. The double digest set Glu-C + trypsin was obtained by digesting further half of theGlu-C set solution with trypsin. Each set contained approximately 50 µg of peptides in a volume of 500 µl (Fig. 12).

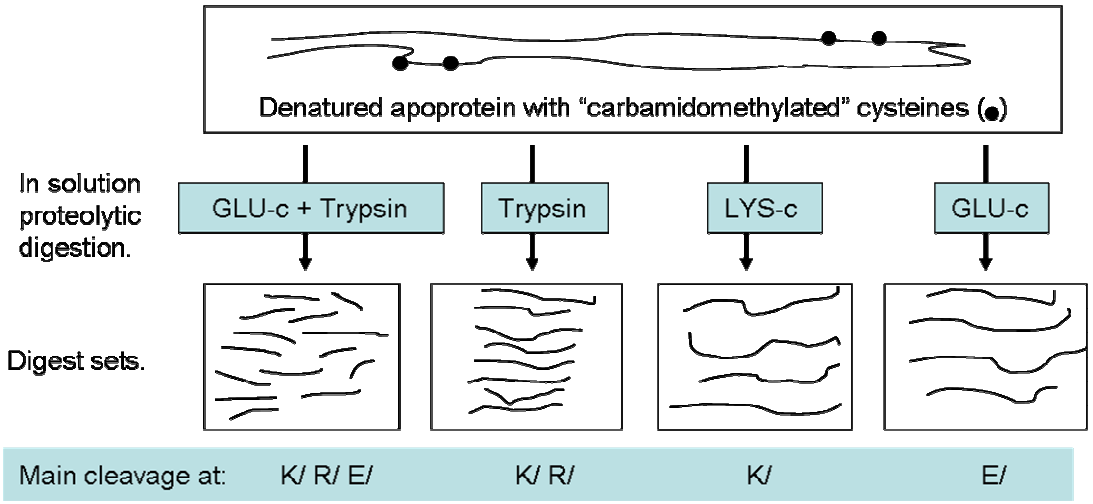

Figure 12. Proteolytic digestions planned in order to originate four distinct sets of peptides different in average length and number. K/ - cleavage at the carboxyl-side of lysine residues. R/ - cleavage at the carboxyl-side of arginine residues. E/ - cleavage at the carboxyl-side of glutamic acid residues.

Mass spectrometry

All measurements were performed on a linear IT-FTICR hybrid instrument (LTQ-FT, Thermo Fischer Scientific, San Jose, CA, USA).

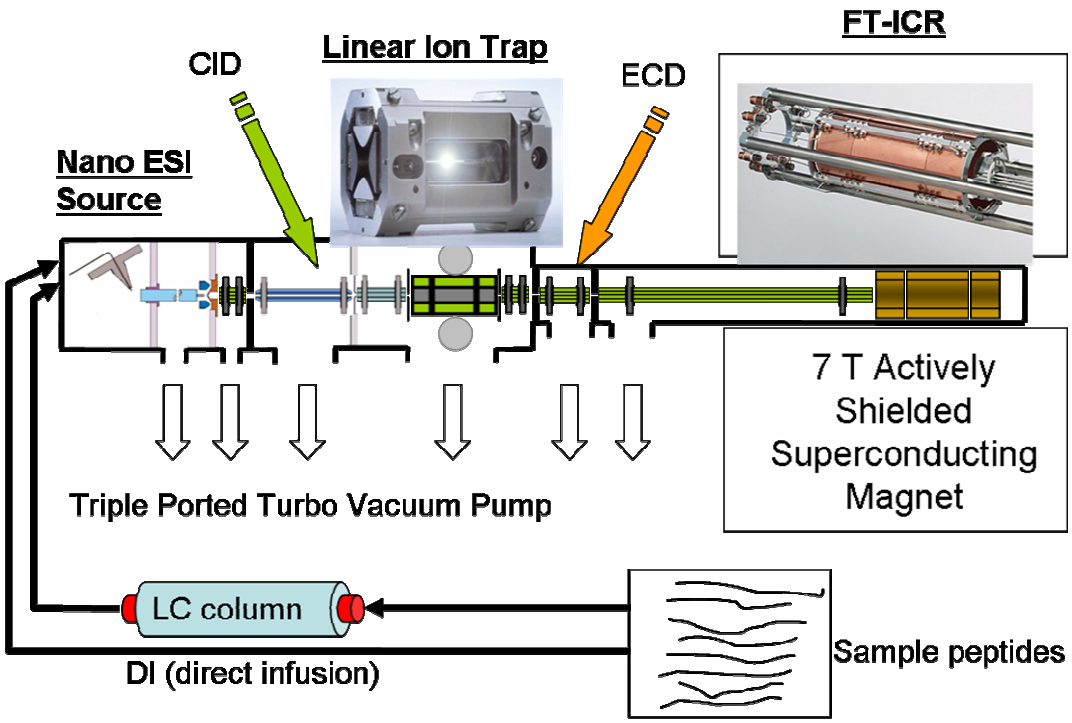

Figure 13. Diagram of the LTQ-FT Mass Spectrometer setup with sample feed by either direct infusion (DI) or online liquid chromatography (LC). Collision induced dissociation is performed in the ion trap chamber, whereas electron capture dissociation is done "en route" to the FT-ICR unit.

A sample flow of 1.4 µl/min was supplied to a nanoelectrospray ionization (NSI) source fitted with an uncoated silica picotip of 30 µm i.d. (NewObjective, Woburn, MA, USA). Data acquisition was controlled with the Xcalibur 2.0 software (Thermo Fischer Scientific). The molecular weight of denatured apocytochrome was determined by direct infusion (DI) (apoprotein solution in 40% methanol and 1% acetic acid) using both IT-MS and FTICR prior to reduction and carbamidomethylation of the cysteines. For the digest sets, the main mode of analysis was LC/NSI-MS and MS/MS. 10 µl of a digest set solution was desalted by C18 Zip-Tip (Millipore), after which the sample was dried in a Speedvac equipment (Savant, Thermo Fisher). The residue was then re-dissolved by adding 10 µl of water. The proteolytic fragments of each digestion set were separated by online gradient reversed-phase micro-LC (10 cm × 300 µm i.d. TARGA C18 column, Higgins Analytical, Mountain View, CA, USA) prior to NSI. The flow rate of about 1.4–1.5 µl/min was obtained by splitting a flow of 100 µl/min supplied by the LC pump (SurveyorMS, ThermoElectron, SanJose, CA). Solvent A was 0.1% acetic acid in water and solvent B was 0.1% acetic acid in acetonitrile. After column equilibration in 2% solvent B, the sample was injected. Following isocratic solvent delivery for 5 min, a linear gradient was used to raise solvent B to 60% in 60 min. High-resolution full-scan (FTICR) and product-ion mass spectra (MS/MS, in the linear ion trap) were obtained using the data-dependent acquisition feature of the Xcalibur software. Full-scan mass spectra were recorded in the FTICR unit (field strength of 7 T), and an interim 'snapshot' (by fast Fourier transformation (FFT) after about 0.1 s of transient acquisition to yield a spectrum with $M/\Delta M \sim 12\ 500$) was used to select the three most intense ions for consecutive collision-induced dissociation (CID), with He as target gas (35% relative collision energy) and MS/MS product ion scans in the IT. During the recording of these tandem mass spectra, the acquisition of the transient and subsequent FFT was continued to yield $M/\Delta M$ of about 100 000 at m/z 400 by the end of the approximately 1.6 s cycle. Subsequent analyses of the same sets were done using the Xcalibur's "exclusion" feature; i.e. ions from peptides whose sequence had been already determined from previous runs were put into a list to disregard for CID and MS/MS experiments. For each digest set, 4–5 data-dependent LC-NSI-MS/MS runs were taken. Alternatively, fragmentation of selected parent ions was done by ECD (~1.8 eV electron energy for 70 ms) in the ICR cell. The number of scans required for an acceptable quality ECD spectrum usually required a time window of only 20–30 s, which allowed us to combine ECD-FTMS with online LC. In contrast, LC-CID-IT

MS/MS with detection of products in the FTICR was generally unviable because the time window required for a good signal-to-noise ratio exceeded largely the elution time intervals. Thus, DI of the digest sets was employed as a complementary mode of sample introduction. For such experiments, 10 µl of the given digest solution was desalted by C18 Zip-Tip (Millipore) and then diluted to a total of 100 µl (40% methanol and 1% acetic acid), which was directly infused into the NSI source. Ions were manually selected, isolated, and then fragmented by CID in the IT. The product ions were then transferred to the FTICR cell where scans were accumulated for several minutes (usually up to 5 min per parent ion) until the spectral quality was deemed to give visually acceptable signal-to-noise ratio. Alternatively, fragmentation of selected parent ions was done by ECD in the FTICR.

De novo peptide sequence determination from MS/MS data

Most peptide sequences were determined with the aid of the DeNovoX program (Thermo Fischer Scientific, San Jose, CA, USA)*. The peptides for which the algorithm was unable to provide unambiguous solutions (often due to the low intensity of CID daughter ions) were sequenced manually by a process similar to that described by Zhang, *et al.* 2003. Each final sequence was validated by comparing the MS/MS product ion spectra to comprehensive tables of all possible sequence related ions predicted by Protein Prospector's MS-Product module (by Baker and Clauser)♣.

Potentiometry, UV-VIS and CD spectroscopy, calorimetry

Chemical reduction and oxidation of the samples

Chemically reduced and oxidized states were obtained by addition of ascorbic acid (100 µM – 1 mM) or sodium dithionite (100 µM – 1 mM) and potassium ferricyanide (100–500 µM), respectively.

* http://www.thermo.com/eThermo/CMA/PDFs/Articles/articlesFile_21648.pdf

♣ http://prospector.ucsf.edu

Redox titration

The redox potential of the cytochrome was determined by potentiometric titration [Wilson 1978, Dutton 1978] using a home-made stirred spectroelectrochemical cell (Fig. 14) with a volume of 3 ml and a 1 cm optical path.

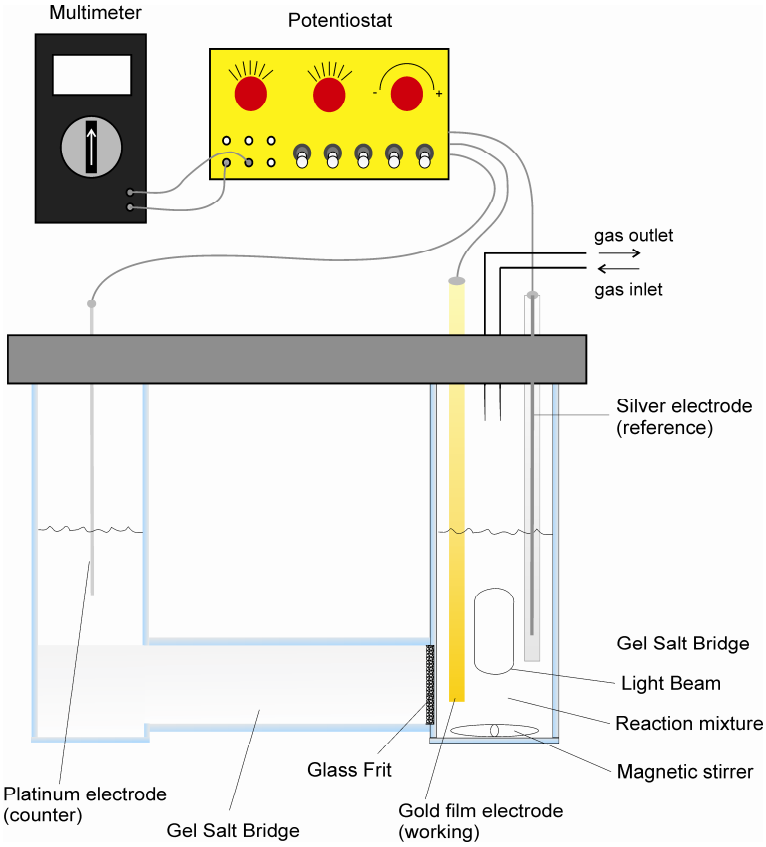

Figure 14. Diagram of our home-made redox titration setup.

The cell was installed on the spectrometer (either absorption or CD) and anaerobic conditions were maintained with a constant stream of 99.995% pure, water-saturated N_2 gas (Messer). The working electrode and the reference electrode were fitted directly into the cell, whereas the platinum counter-electrode was fitted into a vessel (containing 100 mM KCl and the same buffer used in the reaction mixture) connected to the cell by means of a salt bridge (1 M KCl in 10% polyacrylamide gel adjacent to a glass frit). Either indium-tin-oxide (ITO) or gold-covered glass plates were used as working electrodes. The home-made capillary Ag/AgCl reference electrode consisted of a silver wire coated with a thin layer of AgCl, inserted into a glass capillary and connected to the bulk solution by means of a salt bridge (1 M KCl in 5% polyacrylamide gel). Temperature was adjusted and maintained by the use of a water jacket. The reaction

mixture was similar to that described by Meyer, et al. 1991. It contained 7 μM cytochrome, 10 mM EDTA and 100 mM KCl in 20 mM Tris buffer at pH 7.7. As mediators, 1 mM iron-EDTA, 1 mM methyl viologen and 1 mM potassium ferricyanide were used. The enforced solution potential between the working and the platinum electrodes was provided by an Applied Photophysics (UK) Electrochemical Processor or by an EF438A Elektroflex (Szeged, Hungary) potentiostat. The equilibrium was declared to be set when the current across the cell damped down to the background current value (<0.5 μA). The potential between the working and reference electrodes was measured with a Maxwell Mx-25 302 multimeter. All potentials are given *vs.* the normal hydrogen electrode (NHE).

UV–VIS spectroscopy

Spectra were recorded with an ATI Unicam UV2 spectrometer equipped with a Grant thermostatic water bath. The spectrometer was controlled by the computer program Vision32 v1.25.

Circular dichroism spectroscopy

CD spectra were taken in a Jobin Yvon CD6 dichrograph equipped with a temperature-controlled sample holder. In the far-UV range (190–250 nm), the temperature dependences of both aerobic and anaerobic samples were measured in a 0.1 mm quartz cell, and the average of five scans was collected. Anaerobic samples were prepared in an anaerobic box (Bactron Anaerobic/Environmental Chamber, Sheldon, USA) under a 95% N_2 + 5% H_2 atmosphere; the quartz cell was sealed with silicone grease. Aerobic samples were prepared under normal atmospheric conditions. In the near-UV–visible range (250–700 nm), the optical path length was 2 mm for aerobic samples. The cell was sealed with a Teflon stopper to avoid evaporation of the sample, but the atmosphere was not changed. Anaerobic temperature dependence measurements were coupled with redox potential control, using the above-mentioned potentiostatic cell. All spectra were recorded in delta absorbance (ΔA) units with a resolution of 1 nm and an integration time of 1 s.

In temperature-dependent spectroscopic experiments, the samples were equilibrated at each temperature for 5 min. Overall, including the 5 min measuring time, the average heating rate was 0.5 ºC min^{-1}.

Secondary structure calculations

CD spectra in the range 190–250 nm were used for secondary structure calculations. The spectra were smoothed by the nine-point Savitzky Golay method. The CD-Pro software used classifies secondary structures into six types: regular and distorted α helices, regular and distorted β strands, turns and unordered structures [Sreerama and Woody 2000, Sreerama*, et al.* 2000]. The distorted α helix refers to the outer two and the distorted β strands refers to the outer one residues of each end of a segment [Sreerama and Woody 2004]. The amount of ‘‘distorted’’ structures can be used as an estimate of the number of such structures, whereas the ‘‘regular’’ portion is indicative of the overall length of the same structures.

Differential scanning calorimetry

DSC measurements were performed with a high-sensitivity scanning microcalorimeter MicroCal VP-DSC (Microcal LLC, Northampton, Massachusetts). Runs were routinely made in the temperature range indicated in the Results section, with a heating rate of 1 ºC min^{-1}. The specific heat capacity curves were deconvoluted into Gaussian functions. The transition temperature (Tm) was defined as the temperature at the maximum of the resulting Gaussian functions.

Mathematical analysis

All spectral analysis, including spectrum mix calculations, peak fittings, and transition function fittings, was carried out with the program SPSERV V3.5 (copyright Csaba Bagyinka). Use of the program is demonstrated in several papers [Böddi*, et al.* 1992, Böddi*, et al.* 1993, Böddi and Franck 1997, Kota*, et al.* 1999, Debreczeny*, et al.* 2003, Kelemen*, et al.* 2004]. When the peak height for molar absorptivity (ε) determination was calculated, the local baseline was subtracted. The absorption spectra were deconvoluted into Gaussian functions on the energy (frequency) scale. This results in an asymmetric band on the wavelength scale which has a tail on the red wing of the band. The total intensity of a band was calculated as the area of the individual peaks obtained by deconvolution of the measured absorption spectra. SVD analysis of the spectra was performed with the same program SPSERV V3.5. The data were analyzed

as described earlier [Bagyinka, *et al.* 2003, Kota, *et al.* 1999, Debreczeny, *et al.* 2003, Kelemen, *et al.* 2004, Shrager 1986].

The midpoint oxidation-reduction potential of cytochrome c_4 was determined by measuring the visible absorption spectrum at each applied potential. The whole α–β (500–600 nm) region of the absorption spectrum was deconvoluted into Gaussian distribution curves and the proportion of the oxidized state in each spectrum was determined. Redox ratios were also determined by reconstructing the actual spectrum by mixing of the spectra of the pure oxidized and reduced states or by mixing three spectra (the spectra of the fully oxidized and fully reduced states and the high-temperature spectrum) in the event of temperature dependence. The two methods gave the same results within the experimental error. Midpoint potentials were calculated by fitting the fraction of the oxidized state to the redox transition function derived from the Nernst equation:

$$f_{ox} = \left(1 + e^{-\frac{E - E^{o}}{\Gamma}}\right)^{-1}$$

where f_{ox} is the fraction of the oxidized state ($c_{ox}/(c_{ox} + c_{red})$), where c_{ox} and c_{red} are the concentrations of oxidized and reduced states, respectively; E is the system potential, E^{o} is the midpoint redox potential, and Γ = RT/nF is the width of the transition, where R is the universal gas constant, T is the temperature in K, n is the number of electrons transferred in the reaction, and F is the Faraday constant. At 25 ºC for n = 1, Γ = 25.68 mV, and at 60 ºC for n = 1, Γ = 28.76 mV.

Results

Purification of cytochrome c_4

The filtrate collected after the anion-exchange DEAE batch chromatography contains many negatively charged proteins, including several redox proteins of our interest. To obtain the pure protein of interest (in this case, cytochrome c_4), we relied on several steps of FPLC (Fig.15).

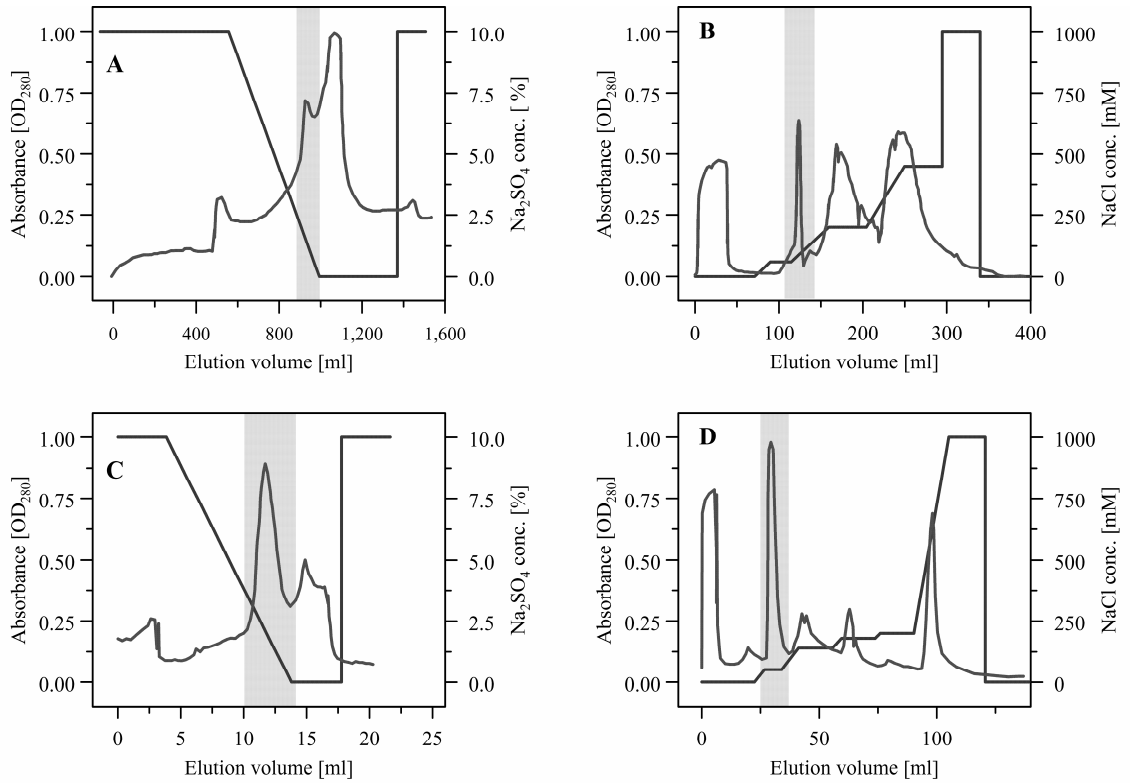

Figure 15. The final steps of cytochrome c_4 purification comprised several hydrophobic interaction (A, C) and anion exchange (B, D) column chromatographies. Fractions collected are highlighted in grey.

In the first hydrophobic chromatography, cytochrome c_4 eluted at around 1% $(NH_4)_2SO_4$ (Fig. 15a). As can be seen in the figure, the cytochrome peak was overlapping with the next peak in the elution order. Incidentally, this latter peak contains two other proteins of interest, the hydrogenase and the blue-copper protein. Often the two peaks were not resolved at all, and thus we collected both together in one single fraction, and proceeded to load the fraction onto the anion-exchange column without incurring in substantial performance loss (Fig. 15b). However, we have recently introduced a new hydrophobic interaction matrix for this first column chromatography

step, a Phenyl-Fractogel. It allows for a superior resolution, which permits complete separation of the cytochrome peak from the hydrogenase peak.

The plateaus in the gradient of the first anion-exchange chromatography were introduced to optimise the separation of the three main observable elution peaks (Fig. 15b). Cytochrome c_4 eluted in a well resolved peak at 100 mM NaCl (Fig. 15b). A second linear gradient hydrophobic chromatography followed with cytochrome elution in the 3–0% $(NH_4)_2SO_4$ interval (Fig. 15c). The final column chromatography also had plateaus in the elution gradient allowing cytochrome c_4 to elute in a well-resolved peak and in pure form at around 50 mM NaCl (Fig. 15d).

The purity was determined by SDS–PAGE (Fig.16).

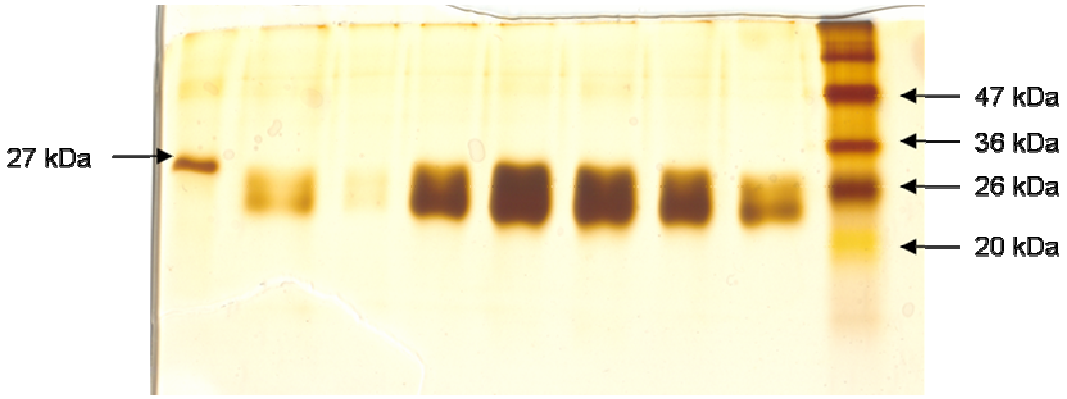

Figure 16. 10% polyacrylamide SDS gel of pure cytochrome fractions. The leftmost lane is a control protein of known size, whereas the right most lane contains a low molecular weight protein marker. The remaining lanes contain distinct fractions with different concentrations of pure cytochrome c_4. Notice how the cytochrome band tends to smear, especially when in high amounts, when it clearly overloads the lanes.

Molecular mass and heme content of cytochrome

Using SDS-PAGE gradient gels, we estimated the molecular weight of cytochrome c_4 to be in the range 22-25 kDa.

Cytochrome c_4 contains two hemes per molecule, as determined by the pyridine hemochromogen method, and later confirmed by Mass Spectrometry.

Determination of primary structure - Mass Spectrometry

Removal of prosthetic heme groups to obtain apocytochrome

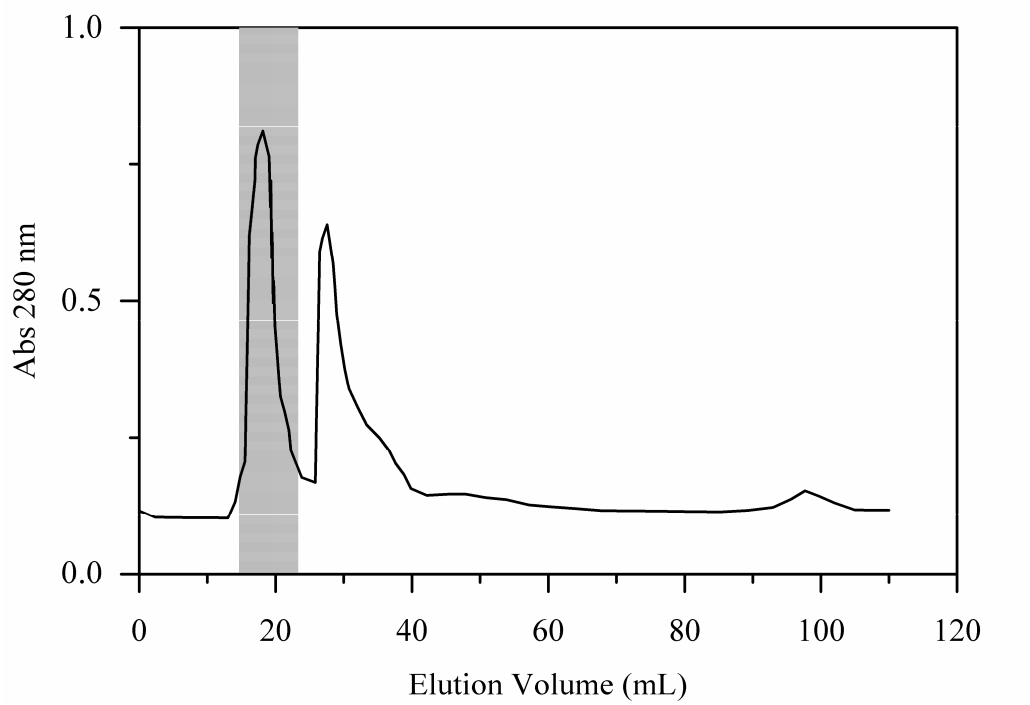

Figure 17. Gel filtration used to recover apocytochrome after overnight incubation in a solution containing reagents that denature the protein and break the covalent bonds between the heme rings and the polypeptide chain. Apocytochrome eluted in the highlighted peak, whereas hemes, salts and other small molecules eluted afterwards.

From the starting 1 mg of pure cytochrome c_4, approximately 0.5 mg of apoprotein was recovered. The apoprotein solution contained less than 5% holocytochrome as checked by comparing the ratios of the 280nm/Soret peaks in the absorption spectra.

Molecular weight determination of the intact apocytochrome

The first step to *de novo* sequencing of cytochrome c_4 was to determine the intact apoprotein total mass. Fig. 18 shows the mass spectrum of the apocytochrome prior to reduction and alkylation. The high-resolution spectra (Fig. 18b) obtained by FTICR yielded a monoisotopic mass of 20 749.0 Da. This value was a useful guide in the final assembly of the peptide sequences into the complete protein sequence [Johnson and Biemann 1987].

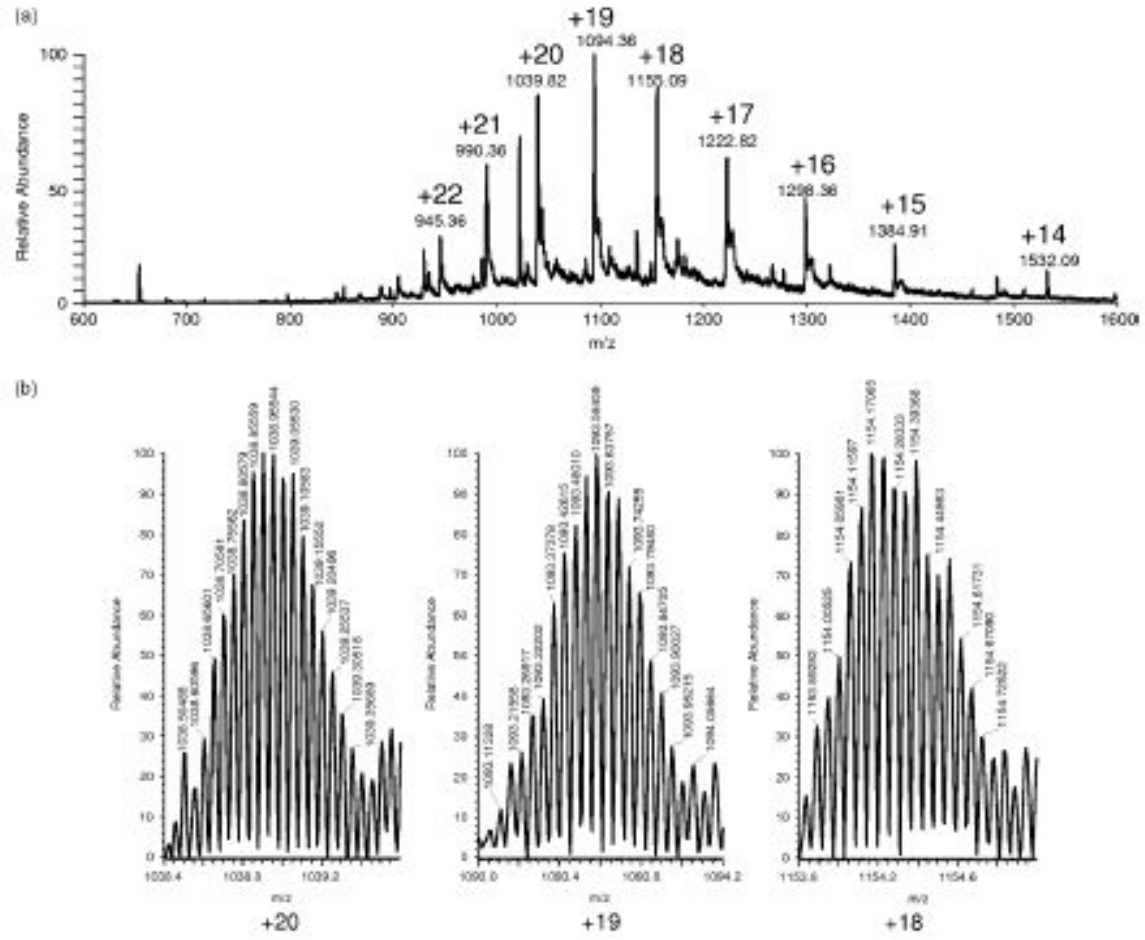

Figure 18. a) ITMS of intact apocytochrome. b) FTMS of the most intense ions (+20, +19, +18). The higher resolution provided by the FT-ICR unit shows that each individual peak in a) is in fact a cluster of isotopic peaks.

Compilation of peptide masses

In keeping with the general procedure depicted in Fig. 8, we analysed by MS the peptide mixtures resulting from the proteolytic digestions (Fig. 12). The primary spectra of the peptide mixtures of each set were obtained first by LC/FTMS and later without LC separation (i.e. by direct infusion - DI) as well. The Electro Spray Ionization method usually produces multiply charged peptide ions [Loo*, et al.* 1993]. We have compiled the most abundant ions into comprehensive lists of peptide masses (Table 2). The compilation was performed manually and was primarily derived from the LC/FTMS full-scan data, and later completed through the use of DI/FTMS data.

From the lists of accurate masses (Table 2), and following our general outline of protein sequencing by mass spectrometry (Fig. 8), we attempted to perform Mass-fingerprinting. The method consisted in searching for the measured exact masses of the Trypsin digest set in online databases. The search did not yield any useful hits.

A. List of peptide masses and respective ions for the GluC + Trypsin digest set.

TRYP+GLUc

peptide mass	z=1	z=2	z=3	z=4
337.1750	**338.1823**	169.5948	113.3989	85.3010
418.2328	**419.2401**	210.1237	140.4182	105.5655
458.2740	**459.2813**	230.1443	153.7653	115.5758
489.2435	**490.2508**	245.6290	164.0884	123.3181
521.3213	**522.3286**	261.6679	174.7811	131.3376
537.2911	**538.2984**	269.6528	180.1043	135.3300
589.2240	**590.2313**	295.6193	197.4153	148.3133
590.2548	**591.2620**	296.1347	197.7589	148.5710
601.2496	**602.2569**	301.6321	201.4238	151.3197
606.2505	**607.2578**	304.1326	203.0908	152.5699
652.3254	**653.3327**	**327.1700**	218.4491	164.0886
750.3660	**751.3733**	**376.1903**	251.1293	188.5988
756.3290	**757.3363**	**379.1718**	253.1169	190.0895
763.3574	**764.3647**	382.6860	255.4598	191.8466
780.3840	**781.3913**	**391.1993**	261.1353	196.1033
796.3789	797.3862	**399.1967**	266.4669	200.1020
807.4127	**808.4199**	**404.7136**	270.1448	202.8604
903.4008	**904.4080**	**452.7077**	302.1409	226.8575
919.3957	**920.4030**	**460.7051**	307.4725	230.8562
938.5589	**939.5662**	**470.2867**	313.8603	235.6470
954.5538	**955.5611**	**478.2842**	319.1919	239.6457
991.5338	992.5411	**496.7742**	331.5186	248.8907
1004.4168	**1005.4241**	**503.2157**	335.8129	252.1115
1005.4008	1006.4081	**503.7077**	336.1409	252.3575
1109.6233	1110.6306	**555.8189**	370.8817	278.4131
1124.5866	**1125.5939**	**563.3006**	375.8695	282.1539
1132.5513	**1133.5586**	**567.2829**	**378.5244**	284.1451
1230.5802	**1231.5875**	**616.2974**	411.2007	308.6523
1253.6768	1254.6841	**627.8457**	**418.8995**	314.4265
1265.5636	1266.5709	**633.7891**	422.8618	317.3982
1376.7300	**1377.7373**	**689.3723**	459.9173	345.1898
1393.7565	1394.7638	**697.8855**	465.5928	349.4464
1436.5925	1437.5998	**719.3035**	479.8714	360.1554
1451.7660	1452.7733	**726.8903**	**484.9293**	363.9488
1453.6191	1454.6263	**727.8168**	**485.5470**	364.4120
1454.6031	**1455.6104**	**728.3088**	**485.8750**	364.6580
1467.7609	1468.7682	734.8877	**490.2609**	367.9475
1531.7015	1532.7088	766.8580	**511.5745**	383.9327
1730.8336	1731.8409	866.4241	**577.9518**	433.7157
1747.7762	1748.7834	**874.8954**	**583.5993**	437.9513
1946.9082	1947.9155	**974.4614**	**649.9767**	**487.7343**
2038.9313	2039.9386	**1020.4729**	**680.6510**	**510.7401**
2357.1409	2358.1482	**1179.5777**	**786.7209**	590.2925
2374.1674	2375.1747	**1188.0910**	**792.3964**	**594.5491**
2375.1514	2376.1587	**1188.5830**	**792.7244**	**594.7951**

B. List of peptide masses and respective ions for the Trypsin digest set.

TRYPsin

peptide mass	z=1	z=2	z=3	z=4
302.1703	**303.1775**	152.0924	101.7307	76.5498
415.2543	**416.2616**	208.6344	139.4254	104.8209
418.2328	**419.2401**	210.1237	140.4182	105.5655
444.2445	**445.2518**	223.1295	149.0888	112.0684
723.3551	**724.3624**	**362.6849**	242.1257	181.8461
763.3574	**764.3647**	**382.6860**	255.4598	191.8466
780.3840	781.3913	**391.1993**	261.1353	196.1033
796.3789	797.3862	**399.1967**	266.4669	200.1020
846.3364	**847.3437**	**424.1755**	283.1194	212.5914
912.4454	**913.4526**	**457.2300**	305.1557	229.1186
920.3732	**921.3805**	**461.1939**	307.7983	231.1006
929.4719	**930.4792**	**465.7432**	**310.8312**	233.3753
935.4712	**936.4785**	468.7429	312.8310	234.8751
938.5589	**939.5662**	**470.2867**	313.8603	235.6470
957.4767	**958.4840**	**479.7456**	320.1662	240.3765
1004.4168	**1005.4241**	**503.2157**	335.8129	252.1115
1005.4008	**1006.4081**	**503.7077**	336.1409	252.3575
1031.5135	1032.5208	**516.7640**	344.8451	258.8856
1076.4743	**1077.4816**	**539.2444**	359.8320	270.1258
1109.6233	1110.6306	**555.8189**	370.8817	278.4131
1124.5866	1125.5939	**563.3006**	375.8695	282.1539
1132.5513	1133.5586	**567.2829**	**378.5244**	284.1451
1155.5772	**1156.5844**	**578.7959**	386.1997	289.9016
1159.5721	1160.5794	**580.7933**	387.5313	290.9003
1253.6768	**1254.6841**	**627.8457**	**418.8995**	314.4265
1265.5636	1266.5709	**633.7891**	422.8618	317.3982
1322.6354	1323.6427	**662.3250**	441.8857	331.6661
1436.5925	1437.5998	**719.3035**	479.8714	360.1554
1453.6191	1454.6263	**727.8168**	485.5470	364.4120
1454.6031	1455.6104	**728.3088**	**485.8750**	364.6580
1485.7385	1486.7458	**743.8765**	496.2534	372.4419
1563.7780	1564.7853	**782.8963**	**522.2666**	**391.9518**
1688.8457	1689.8530	845.4301	**563.9558**	423.2187
1946.9082	1947.9155	**974.4614**	**649.9767**	**487.7343**
1948.9742	1949.9815	**975.4944**	**650.6653**	488.2508
1966.0007	1967.0080	984.0076	**656.3409**	492.5075
2337.1562	2338.1635	**1169.5854**	**780.0594**	585.2963
2353.1511	2354.1584	1177.5829	**785.3910**	589.2951
2357.1409	2358.1482	**1179.5777**	**786.7209**	590.2925
2374.1674	2375.1747	**1188.0910**	**792.3964**	594.5491
2375.1514	2376.1587	1188.5830	**792.7244**	**594.7951**
2622.1703	2623.1776	**1312.0925**	**875.0641**	**656.5499**

C. List of peptide masses and respective ions for the LysC digest set.

LYSc

peptide mass	z=1	z=2	z=3	z=4	z=5
763.3574	**764.3647**	382.6860	255.4598	191.8466	153.6788
780.3840	**781.3913**	**391.1993**	261.1353	196.1033	157.0841
938.5589	**939.5662**	**470.2867**	313.8603	235.6470	188.7191
1140.5928	1141.6000	**571.3037**	381.2049	286.1555	229.1258
1155.5772	**1156.5844**	**578.7959**	386.1997	289.9016	232.1227
1253.6768	1254.6841	**627.8457**	**418.8995**	314.4265	251.7426
1265.5636	1266.5709	**633.7891**	422.8618	317.3982	254.1200
1532.7736	1533.7808	**767.3941**	**511.9318**	**384.2007**	307.5620
1946.9082	1947.9155	**974.4614**	**649.9767**	487.7343	390.3889
2337.1562	2338.1635	1169.5854	**780.0594**	585.2963	468.4385
2357.1409	2358.1482	**1179.5777**	**786.7209**	590.2925	472.4355
2374.1674	2375.1747	**1188.0910**	**792.3964**	**594.5491**	475.8408
2375.1514	2376.1587	1188.5830	**792.7244**	594.7951	476.0376
3327.5149	3328.5222	1664.7647	**1110.1789**	**832.8860**	**666.5103**
3679.7796	3680.7869	1840.8971	**1227.6005**	**920.9522**	736.9632

D. List of peptide masses and respective ions for the GluC digest set.

GLUc

peptide mass	z=1	z=2	z=3	z=4	z=5	z=6
1008.5314	1009.5387	**505.2730**	337.1844	253.1401	202.7136	169.0958
1120.5302	**1121.5374**	**561.2724**	374.5173	281.1398	225.1133	187.7623
1204.5724	1205.5797	**603.2935**	**402.5314**	302.1504	241.9218	201.7693
1265.5636	1266.5709	**633.7891**	422.8618	317.3982	254.1200	211.9345
1376.7664	**1377.7736**	**689.3905**	459.9294	345.1989	276.3605	230.4683
1393.7565	**1394.7638**	**697.8855**	465.5928	349.4464	279.7586	233.3000
1491.6408	**1492.6480**	**746.8277**	498.2209	373.9175	299.3354	249.6141
1747.7762	1748.7834	**874.8954**	**583.5993**	437.9513	350.5625	292.3033
1862.9050	1863.9123	**932.4598**	**621.9756**	466.7335	373.5883	311.4914
1963.9891	1964.9964	**983.0018**	**655.6703**	492.0046	393.8051	328.3388
2038.0782	2039.0855	1020.0464	**680.3667**	**510.5268**	**408.6229**	340.6870
2054.0731	2055.0804	1028.0438	685.6983	**514.5256**	**411.8219**	343.3528
2190.0845	2191.0917	**1096.0495**	**731.0354**	548.5284	439.0242	366.0214
2840.3671	2841.3743	1421.1908	947.7963	**711.0990**	**569.0807**	474.4018
3305.6550	3306.6623	1653.8348	1102.8923	**827.4210**	**662.1383**	551.9498
3306.6390	3307.6463	1654.3268	1103.2203	**827.6670**	**662.3351**	**552.1138**
3787.8675	3788.8748	1894.9410	1263.6298	947.9742	**758.5808**	632.3185
3788.8515	3789.8588	1895.4330	1263.9578	948.2202	**758.7776**	632.4825
4024.9272	4025.9345	2013.4709	1342.6497	**1007.2391**	**805.9927**	**671.8285**
5035.4206	5036.4279	2518.7176	1679.4808	1259.8624	**1008.0914**	**840.2440**

Table 2. Lists of peptide masses for the four digest sets. The peptide masses were calculated based on the ions highlighted in grey background, which were observed in the primary FTMS mass spectra. The non-highlighted ion masses are hypothetical only, and the respective peaks have not been found. The z=1 ions highlighted in orange background in B were found only in mass spectra from DI mode and not on LC mode.

The difficulty of sequencing each peptide depends largely on the peptide length. Table 3 reveals the expected trend between the various digest sets: with more cleavage sites, the number of peptides increased and the average length decreased.

Digest set:	Trypsin + Glu-C	Trypsin	Lys-C	Glu-C
Main cleavage sites:	R/ K/ E/	R/ K/	K/	E/
Number of sequenced peptides:	45	42	15	20
Average peptide mass (kDa):	1.1±0.5	1.3±0.6	1.8±0.9	2.3±1.2

Table 3. Total number and average size of individual peptides that have been sequenced from the four proteolytic digest sets.

Peptide sequencing by MS/MS – mainly by LC-CID-ITMS^2

The bulk of cytochrome c_4 proteolytic peptides were sequenced by automated analysis (DeNovoX software) of the MS/MS spectra acquired in the Ion Trap (IT) during online LC separation of the peptide mixtures. The software was developed by its manufacturers to handle CID-MS/MS data from ITs that do not provide accurate masses. Therefore, DeNovoX does not distinguish between Gln (Q) and Lys (K) whose mass difference is 0.036 Da, and usually treats them as equivalent (both listed as Q by the software).

It was fairly straightforward to determine the sequence of peptides up to 14 amino acids in length (which translates to roughly 1500 Da in peptide mass). Larger peptides posed mounting difficulties. Unfortunately, we didn't have enough smaller peptides (less than ~1500 Da) in our digest sets to get full coverage of the protein sequence, and therefore we had to undertake the difficult task of determining the sequence of a few large peptides.

In Fig. 19, the CID-IT-MS/MS spectrum of the m/z 726.8905 doubly charged precursor ion (peptide mass 1451.7667 Da, taken from the Glu-C + Trypsin set) is shown. This is a typical spectrum for peptides smaller than 1500 Da. The fragmentation pattern is complete (all the **b** and **y** ions are observable) although the ions originating close to the N-terminus ($\mathbf{y_{12}}$ and $\mathbf{b_2}$) show rather low intensity. The considerably high intensity of the $\mathbf{b_{11}}$ ion and the low intensity of the $\mathbf{b_{12}}$ ion are explained by the fragmentation propensities of the proline residue, which largely favours dissociation at its N-terminus [Loo*, et al.* 1993, Tabb*, et al.* 2003].

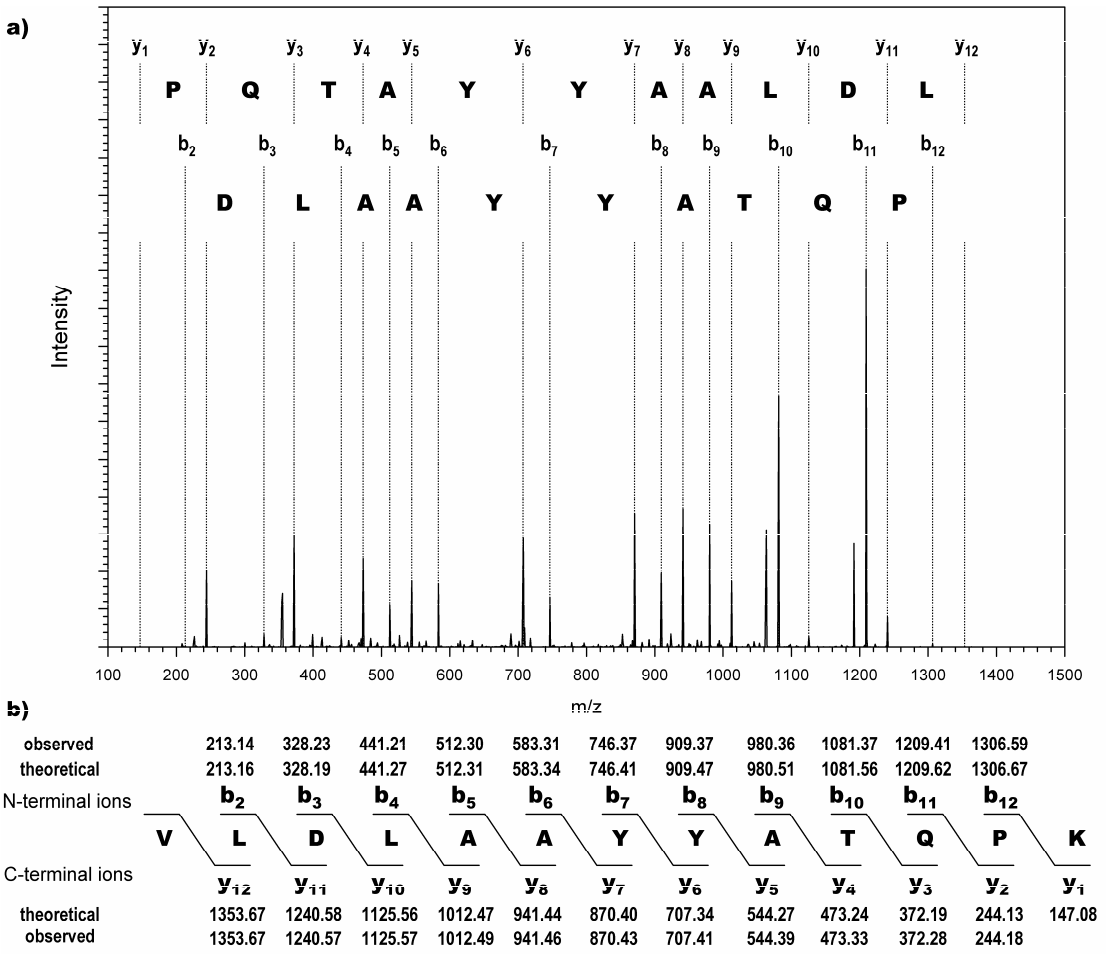

Figure 19. a) CID-IT-MS/MS of doubly charged parent ion of 726.8903 m/z with detection of products in the IT.
b) Theoretical versus observed m/z of CID product ions for the peptide VLDLAAYYATQPK. All product ions shown are singly charged and m/z values are for monoisotopic ions (all carbon atoms being ^{12}C) only.

DeNovoX scored the sequence VLDLAAYYATQPQ with the highest probability. Since full-scan mass spectrum of this peptide set was taken in the FTICR, we know the parent peptide mass with high accuracy and, therefore, we can distinguish between Q (Gln) and K (Lys). Because trypsin had been one of the enzymes used to generate this set of peptides (Glu-C + Trypsin), the likely sequence was VLDLAAYYATQPK. This solution was confirmed by comparing the observed parent peptide mass to the theoretical one (Table 4). The final sequence was validated by matching a theoretical table of product ions (Baker and Clauser)* with the actual LC/CID-IT-MS/MS spectrum (Fig. 19).

* http://prospector.ucsf.edu

Long peptide sequencing by MS/MS – via low-abundance "component" peptides

A slightly larger peptide of 1747.7782 Da, taken from the Glu-C + trypsin set, is shown in Fig. 20. The Pro residues exhibited the usual N-terminal bias, [Tabb*, et al.* 2003] fragmenting abundantly on the amino side (**b_8**, **y_{10}**, **b_{13}** and **y_5**) while scarcely fragmenting on the carboxylic side (**b_9** and **y_9** are very weak, while **b_{14}** and **y_4** are absent).

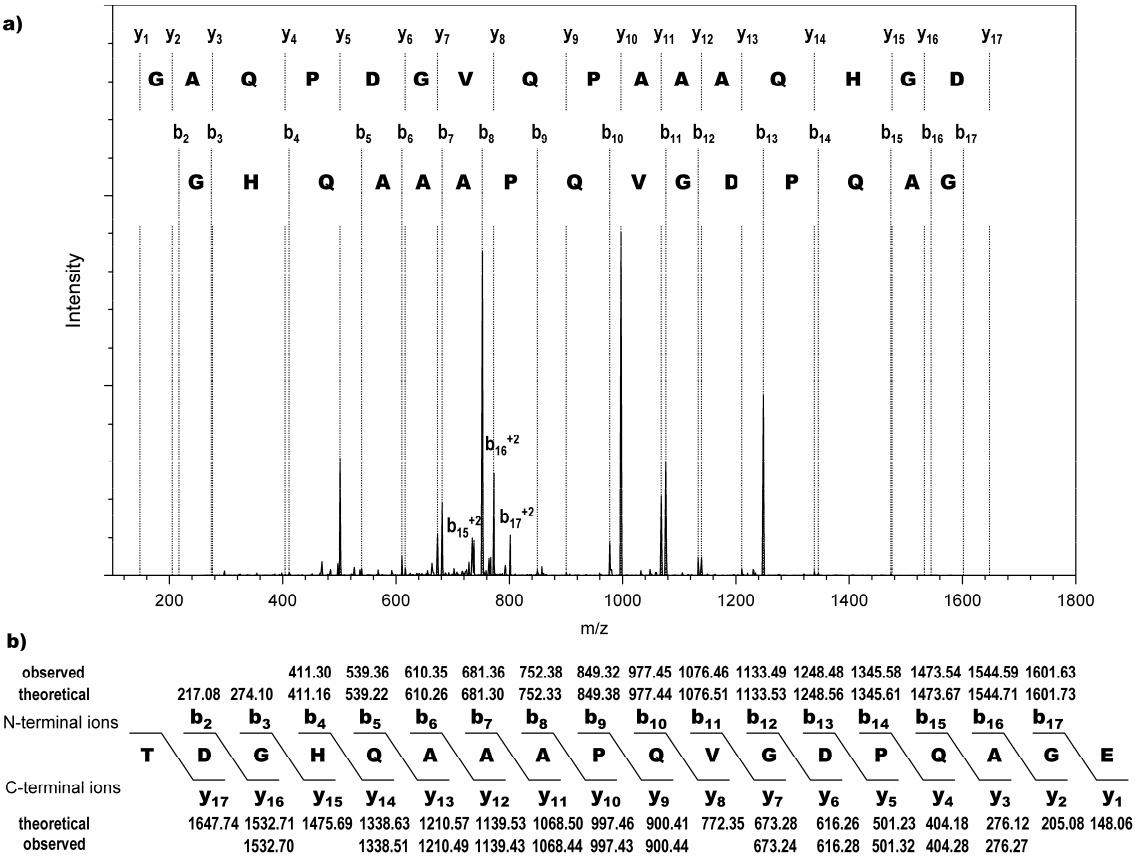

Figure 20. a) CID-IT-MS/MS of the doubly charged parent ion of 874.8957 m/z with detection of the products in the IT. The ion showed very uneven CID fragmentation propensities; many predicted product ions were missing, and many others showed very weak intensity. Additionally, doubly charged product ions became a common occurrence and tended to flood the lower half of the spectrum. b) Predicted versus observed CID product ions for the peptide TDGHQAAAPQVGDPQAGE. The absence of the b_2, b_3, y_{15} and y_{17} ions left the N-terminal four amino acids unresolved. The missing N-terminus was found only after the smaller related peptide TDGHQAAAPQVGD had been resolved (see text for details). All product ions shown in the table are singly charged and m/z values are for monoisotopic ions.

CID fragmentation of this larger peptide brought about a more complex MS/MS spectrum. Doubly charged product ions, which were negligible in the previous example (Fig. 19), now appeared more abundantly (note the doubly charged **b_{15}**, **b_{16}** and **b_{17}** ions)

and confused the sequencing task immensely, since in the IT-MS/MS data, multiply charged ions are indistinguishable from singly charged ones. With the detection of product ions by FTICR, the charge of each daughter ion can be determined precisely (as shown below in Fig. 21) and, therefore, these multiply charged peaks are transformed from a drawback into an advantage.

Another serious problem was that several predicted product ions were missing from the spectrum and others had very weak intensity (Fig. 20). This incompleteness of fragmentation tended to affect mainly the terminal segments of the peptide. In the particular case of the 1747.7782 Da peptide, we found the partial sequence [410.1550]QAAAPQVGDPQAGE (the number 410.1550 contained in square brackets signifies the mass in Da of the missing segment in the peptide's amino acid sequence) from the MS/MS spectrum shown in Fig. 20.

In some cases, we have benefited from a peculiarity of Endoproteinase Glu-C, which is its secondary cleavage activity at the carboxylic side of aspartic acid. This secondary activity occurs at a rate ~3000 times lower than at the main cleavage site (carboxylic side of glutamic acid) [Sorensen*, et al.* 1991]. As a result, digestion of a protein by Glu-C usually yields a small population of peptides specifically produced by the cleavage at D. We, therefore, sought for a parent ion corresponding to the mass of the [410.1550]QAAAPQVGD peptide in the full-scan FTICR mass spectrum recorded online with LC. A doubly charged parent ion of 633.7893 (peptide mass 1265.5636 Da) was found (Table 2). Although the parent ion had a relatively low intensity (compared to that of the various predominant ion species) because it was the result of the low-rate cleavage activity of Glu-C, its reasonably small size (peptide mass of 1265.5640 Da) allowed a complete CID fragmentation, which in turn permitted the elucidation of TDGH as the N-terminal sequence with the accurate mass of 410.1550 Da. The "dynamic exclusion" feature in instrument control (see Materials and Methods) was particularly useful in obtaining MS/MS data on the low-intensity parent ions.

We also observed a very significant secondary activity of trypsin (or rather, of the product of autolysis, pseudotrypsin), the cleavage between a methionine and an alanine.* It produced the peptides QMTPM (606.2513 Da) and AMPLTDQE (903.4016 Da) by cleaving in the middle of the QMTPMAMPLTDQE peptide (1491.6398 Da,

* http://www.expasy.ch/tools/peptidecutter/peptidecutter enzymes.htmlTryps

Table 4). The former two peptides were 10 times more abundant than the latter, indicating that this type of reaction proceeded extensively.

Other non-specific cleavages also occurred, albeit to a smaller extent. Several fragments of the peptide ADGV[641.2414]GPQGNSLVPLWPK (2375.1547 Da) were observed: ADGV[641.2414]GPQGN (1454.6031 Da), SLVPLWPK (938.5589 Da) and GNSLVPLWPK (1109.6240 Da). The cleavages occurred at the amino terminals of a serine and of a glycine. These cleavage sites are not documented for trypsin or pseudotrypsin; however, together with proline, serine and glycine are the residues with the highest N-terminal bias, i.e. propensity of peptides to fragment on the N-side of these residues [Tabb*, et al.* 2003]. Considering that 0.1% acetic acid was used to terminate the enzyme digestions and that the solutions remained at 4 °C for several days while the measurements were being performed, it is suggested that in-solution acid-catalyzed hydrolysis originated these cleavages.

Long peptide sequencing by MS/MS – via DI-CID-FTMS^2 data

On the whole, minor cleavages were useful because they produced smaller fragments from the large peptides, such as the rather long ADGV[641.2414]GPQGNSLVPLWPK. Nevertheless, the internal segment of 641.2414 Da remained obscure. The LC/CID-IT-MS/MS spectrum of this peptide (not shown) was complex and difficult to interpret, as expected from its molecular mass, 2375.1547 Da. Part of the problem is due to the fact that the triply charged parent ion produced singly, doubly and triply charged daughter ions, whose charge was indistinguishable by MS/MS with the detection of product ions in the linear IT. We, therefore, chose to use the Direct Infusion (DI) mode of operation combined with IT-CID (with 3 m/z parent-ion isolation width) and detection of product ions in the FTICR (Fig. 21). The resolving power of the FTICR allowed the exact determination of the charge and monoisotopic mass of each product ion with an accuracy that ensured unequivocal assignments. The missing segment was thus determined to be CLACH, which is one of the heme-binding motifs of this di-heme cytochrome.

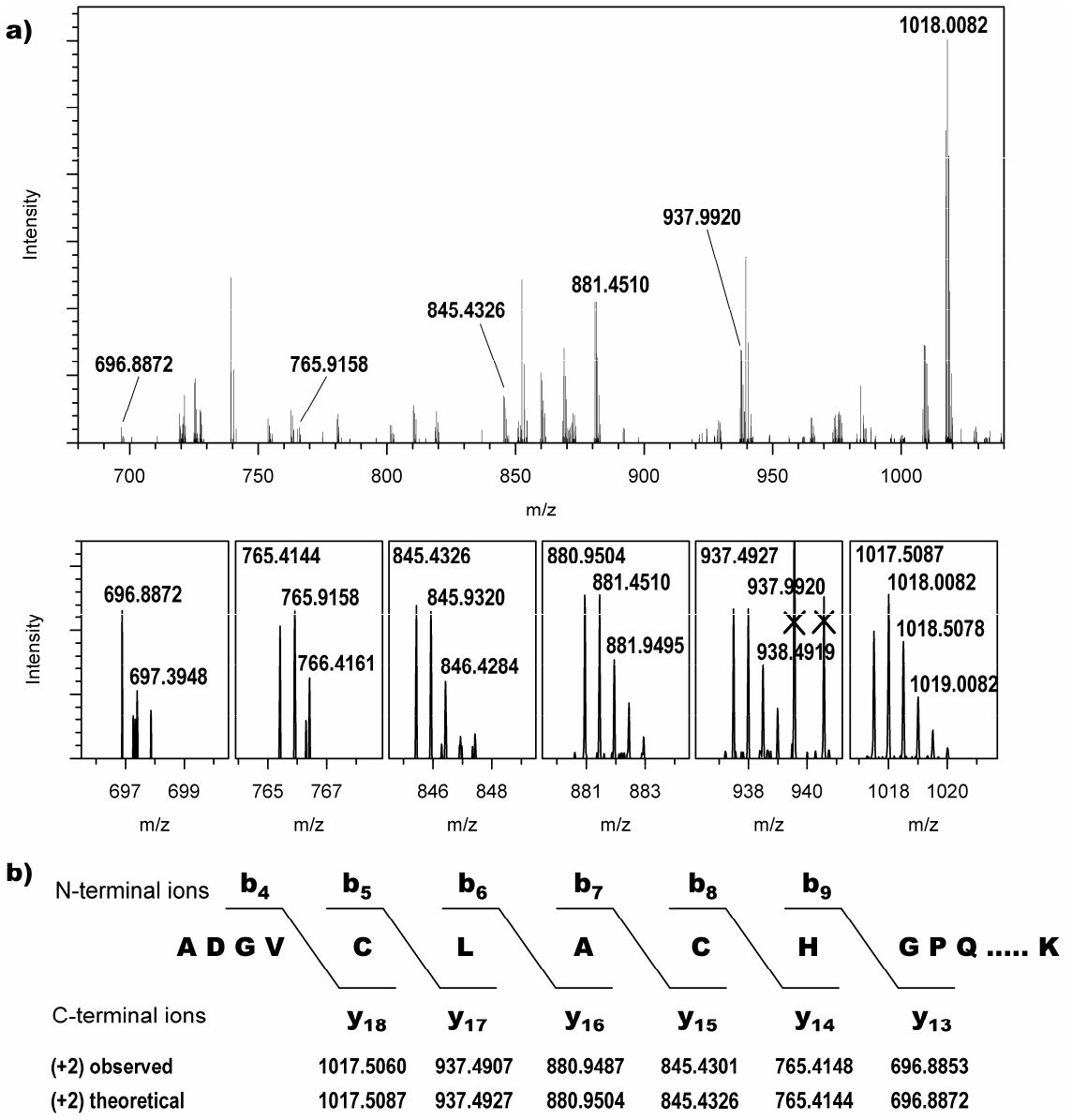

Figure 21. a) Detail of MS/MS spectrum for ADGVCLACHGPQGNSLVPLWPK (2375.1547 Da) from the triply charged precursor ion of m/z 792.7244 with sample introduction by direct infusion, CID in the IT (3.0 m/z isolation width of the parent ion) and detection of the product ions by FTICR. On the lower panels, selected doubly charged y ions corresponding to the CLACH region are shown. FTICR resolved the doubly charged y ions. In the m/z 937.5–940.5 region, singly charged ions were crossed out because they were not discussed in the text.
b) Theoretical versus observed CID product ions of the triply charged parent ion. Only the doubly charged y ions spanning the CLACH region are shown. The difference between the observed and theoretical monoisotopic m/z values is <3 ppm for all ions.

Long peptide sequencing by MS/MS – a particularly difficult segment

Generally, the LC mode proved to be more sensitive than the DI mode, the former revealing ions of very low intensity that remained hidden in the noise in spectra from the latter (Table 4). There were, however, three exceptions where peptides that were not observable in LC mode appeared in the DI mode. This was probably because of either very strong or very weak binding to the C18 LC column, effectively preventing these peptides from eluting in the useful gradient range of 2–60% acetonitrile. But even in the DI mode, the three peptides displayed a very weak intensity and were only found after their existence had already been indicated by sequences obtained in alternative digest sets.

The fact that the tryptic peptides of 415.2543 and 444.2445 Da were missing from the data-dependent LC/NSI-MS and MS/MS experiments of the Glu-C + trypsin and trypsin sets posed some problems. As the assembly of the various peptides into a full protein sequence proceeded, we realized that we did not have all the small peptides composing the 2840.3666 Da peptide from the Glu-C set (Table 4). Its CID-MS/MS with FTICR detection of the product ions yielded modest but important information. We found the two C-terminal amino acids to be QE (Gln-Glu), and also two very strong complementary **b** and **y** ions roughly in the middle of the peptide. These two ions were surely arising from cleavage at the amino side of a proline residue (as shown above in two other peptides: Figs. 19 and 20). Since all the other Glu-C peptides had already been sequenced and linked, we knew also that the N-terminal and C-terminal of this peptide were HFR and LSDQE, respectively, and that the sequence ANDPNGMMR had to be somewhere in the middle. With the knowledge of the precise position of the amino side of Pro, we fixed the ANDPNGMMR segment and deduced the size of the missing gaps: HFR[397.2437]ANDPNGMMR[426.2339]LSDQE.

Only after inferring their existence, did we find the ions corresponding to those gaps in the DI experiments on the trypsin digest set, and could manually perform CID on these ions thereby obtaining their CID-IT/MS/MS data. The gaps were found to be GALR and GAAAR.

We sought for additional experimental data to confirm these sequences. Therefore, we used ECD as an alternative to CID. Good ECD data could be obtained in both LC and DI modes and was particularly useful to confirm sequences of long peptides such as the above-mentioned HFRGALRANDPNGMMRGAAARLSDQE. The ECD spectrum

of the respective quintuply charged m/z 569.2807 ion was matched against the theoretical series of **c**- ions (Fig. 22) and sufficient completeness was found to sustain full validation.

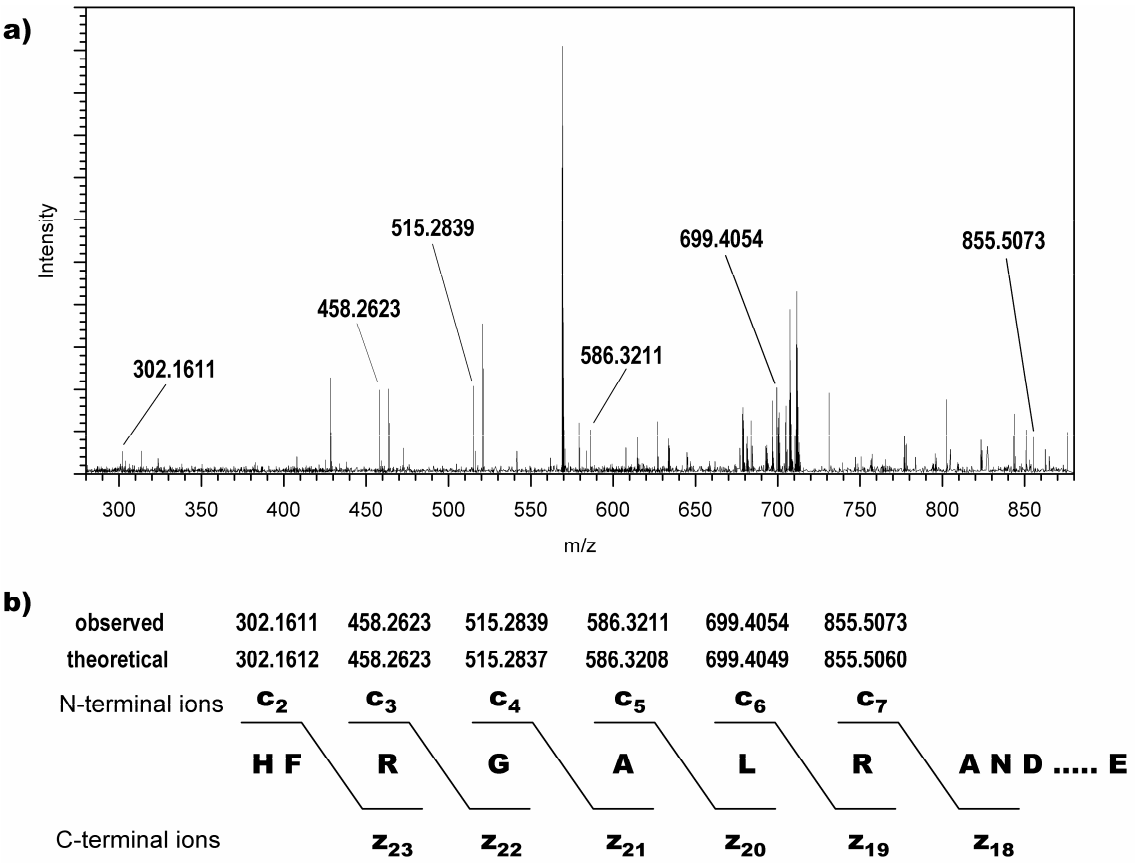

Figure 22. a) ECD mass spectrum of the quintuply charged ion with monoisotopic m/z of 569.0798 (the most abundant isotope peak contained one ^{13}C and thus was 569.2801 m/z). Singly charged c ions corresponding to the RGALR segment are marked by the accurate m/z (monoisotopic).
b) Theoretical versus observed ECD product ions of the quintuply charged peptide HFRGALRANDPNGMMRGAAARLSDQE. The difference between observed and theoretical masses is <2 ppm for all ions. All masses refer to the monoisotopic ions.

Sequencing very short peptides

In general, DeNovoX fared well with MS/MS spectra as long as CID fragment ion series were mostly complete (true for virtually all peptides up to ~1500 Da in size), normally providing the correct solution among its top suggestions. However, DeNovoX revealed some difficulties in scoring the solutions for peptides up to six or seven amino acids long (Table 4). It usually suggested a few different solutions, but all of them with weak scores. Regardless, we regularly found the correct solution by manually verifying each suggested sequence. This problem disappeared for peptides larger than seven or

eight amino acids, in which case DeNovoX usually suggested the correct solution with an outstanding score.

Chemical modifications

Four different types of peptide modifications were found: the apparent gain of an ammonia molecule, gain of an oxygen atom, loss of an ammonia molecule and the deamidation of asparaginyl residue to aspartyl residue (Table 4). These modifications were fairly common, since 13 modified peptides and 2 N to D conversions were observed. Deamidation of N to D is a frequent and well known phenomenon, and it occurs at a high rate when the asparaginyl residue is followed by a glycinyl in the sequence [Robinson 2002], such as in both our observations (peptide masses of 1004.4168 / 1005.4008 and 1453.6191 / 1454.6031 in Table 4).

Table 4.

Sequence	Theoretical mass	Observed mass	Error (ppm)	Glu-C +Trypsin	Trypsin	Lys-C	Glu-C
QR	302.1703	302.1702	-0.2		X		
YR	337.1750	337.1753	0.9	X			
GALR	415.2543	415.2546	0.7		X		
FPR	418.2328	418.2332	0.8	X	X		
GAAAR	444.2445	444.2448	0.7		X		
VLDL	458.2740	458.2744	0.8	X			
QTLE	489.2435	489.2441	1.3	X			
YLVK	521.3213	521.3217	0.7	X			
SLYR	537.2911	537.2918	1.3	X			
Q*MTPM	589.2240	589.2245	0.9	X			
LSDQE	590.2548	590.2556	1.4	X			
WGNPE	601.2496	601.2507	1.8	X			
QMTPM	606.2505	606.2513	1.2	X			
LMDFK	652.3254	652.3260	0.9	X			
GESLYR	723.3551	723.3548	-0.5		X		
LAGQHPE	750.3660	750.3668	1.0	X			
TPGAADPE	756.3290	756.3297	0.9	X			
Q*LMDFK	763.3574	763.3579	0.6	X	X	X	
QLMDFK	780.3840	780.3846	0.8	X	X	X	
QLMDFK°	796.3789	796.3796	0.9	X	X		
YLQGLSQ	807.4127	807.4134	0.9	X			
AN*GVcLAc	846.3364	846.3375	1.3		X		
AMPLTDQE	903.4008	903.4016	0.9	X			
Q*TLEHFR	912.4454	912.4446	-0.8		X		
AMPLTDQE°	919.3957	919.3966	1.0	X			
ANEQMTPM	920.3732	920.3724	-0.9		X		
QTLEHFR	929.4719	929.4712	-0.8		X		
QYLQGLSQ	935.4712	935.4713	0.1		X		
SLVPLWPK	938.5589	938.5594	0.5	X	X	X	
SLVPLWPK°	954.5538	954.5544	0.6	X			
GAADPELASK	957.4767	957.4758	-1.0		X		
LAAVSQYLQ	991.5338	991.5348	1.0	X			
ANDPNGMMR	1004.4168	1004.4178	1.0	X	X		
ANDPDGMMR	1005.4008	1005.4008	0.0	X	X		
YLVKQLMD	1008.5314	1008.5308	-0.6				X
LSDQELAAVS	1031.5135	1031.5128	-0.7		X		
RANEQMTPM	1076.4743	1076.4732	-1.0		X		
GNSLVPLWPK	1109.6233	1109.6240	0.6	X	X		
SLYRWGNPE	1120.5302	1120.5296	-0.5				X
LAAYYATQPK	1124.5866	1124.5880	1.2	X	X		
LSAQHADYTK	1132.5513	1132.5517	0.4	X	X		
AGQHPEYLVK	1140.5928	1140.5902	-2.2			X	
TPGAADPELASK	1155.5772	1155.5762	-0.8		X	X	
LSDQELAAVSQ	1159.5721	1159.5716	-0.4		X		
HADYTKQTLE	1204.5724	1204.5715	-0.8				X
AMPLTDQEVLD	1230.5802	1230.5816	1.1	X			
LAGQHPEYLVK	1253.6768	1253.6785	1.3	X	X	X	
TDGHQAAAPQVGD	1265.5636	1265.5640	0.3	X	X	X	X
LSDQELAAVSQY	1322.6354	1322.6350	-0.3		X		
LAAVSQYLQGLSQ	1376.7300	1376.7316	1.2	X			X
LAAVSQYLQGLSQ[a]	1393.7565	1393.7604	2.8	X			X
AN*GVcLAcHGPQGN	1436.5925	1436.5940	1.0	X	X		
VLDLAAYYATQPK	1451.7660	1451.7667	0.5	X			
ANGVcLAcHGPQGN	1453.6191	1453.6190	0.0	X	X		
ADGVcLAcHGPQGN	1454.6031	1454.6042	0.8	X	X		

(Continued)

VLDLAAYYATQPK°	1467.7609	1467.7603	-0.4	X			
AMPLTDQEVLDLAA	1485.7385	1485.7376	-0.6		X		
QMTPMAMPLTDQE	1491.6408	1491.6398	-0.6				X
GHQAAAPQVGDPQAGE	1531.7015	1531.7029	0.9	X			
FPRLSAQHADYTK	1532.7736	1532.7753	1.1			X	
LSDQELAAVSQYLQ	1563.7780	1563.7774	-0.4		X		
cHGPQGNSLVPLWPK	1688.8457	1688.8450	-0.4		X		
GHQAAAPQVGDPQAGEAK	1730.8336	1730.8360	1.4	X			
TDGHQAAAPQVGDPQAGE	1747.7762	1747.7782	1.2	X			X
LAAYYATQPKTPGAADPE	1862.9050	1862.9035	-0.8				X
TDGHQAAAPQVGDPQAGEAK	1946.9082	1946.9101	1.0	X	X	X	
LSDQELAAVSQYLQGLSQ	1948.9742	1948.9729	-0.7		X		
VLDLAAYYATQPKTPGAAD	1963.9891	1963.9867	-1.2				X
LSDQELAAVSQYLQGLSQ[a]	1966.0007	1965.9997	-0.5		X		
YLVKQLMDFKQRRANE	2038.0782	2038.0776	-0.3				X
TGVPAcSGcHGPQGAGQSLAK	2038.9313	2038.9324	0.5	X			
YLVKQLMDFKQRRANE°	2054.0731	2054.0731	0.0				X
VLDLAAYYATQPKTPGAADPE	2190.0845	2190.0826	-0.8				X
AMPLTDQEVLDLAAYYATQPK	2337.1562	2337.1552	-0.4		X	X	
AMPLTDQEVLDLAAYYATQPK°	2353.1511	2353.1494	-0.7		X		
AN*GVcLAcHGPQGNSLVPLWPK	2357.1409	2357.1454	1.9	X	X	X	
ANGVcLAcHGPQGNSLVPLWPK	2374.1674	2374.1677	0.1	X	X	X	
ADGVcLAcHGPQGNSLVPLWPK	2375.1514	2375.1547	1.4	X	X	X	
WGNPETGVPAcSGcHGPQGAGQSLAK	2622.1703	2622.1669	-1.3		X		
HFRGALRANDPNGMMRGAAARLSDQE	2840.3671	2840.3666	-0.2				X
AKANGVcLAcHGPQGNSLVPLWPKLAGQHPE	3305.6550	3305.6521	-0.9				X
AKADGVcLAcHGPQGNSLVPLWPKLAGQHPE	3306.6390	3306.6401	0.3				X
GESLYRWGNPETGVPAcSGcHGPQGAGQSLAK	3327.5149	3327.5157	0.2			X	
AGRRANEQMTPMAMPLTDQEVLDLAAYYATQPK	3679.7796	3679.7809	0.3			X	
PQAGEAKANGVcLAcHGPQGNSLVPLWPKLAGQHPE	3787.8675	3787.8681	0.2				X
PQAGEAKADGVcLAcHGPQGNSLVPLWPKLAGQHPE	3788.8515	3788.8521	0.2				X
TGVPAcSGcHGPQGAGQSLAKFPRLSAQHADYTKQTLE	4024.9272	4024.9241	-0.8				X
TDGHQAAAPQVGDPQAGEAKANGVCLACHGPQGNSLVPLW PKLAGQHPE	5035.4206	5035.4185	-0.4				X

Table 4. Theoretical versus observed masses of all the sequenced peptides and their occurrence in the digest sets of the cytochrome c_4 from *Thiocapsa roseopersicina*. The average error of the peptide mass determinations was 0.8 ± 0.5 ppm. Chemical modifications: M° = M + Oxygen; M^a = M + NH_3; M* = M – NH_3 (see Tabb, *et al.* 2003). DeNovoX scored poorly the solutions for the small peptides highlighted in grey (see text and Harrison, *et al.* 2006 for explanation).

Establishing the full protein sequence

Partial protein sequences started revealing themselves as soon as overlapping peptides were obtained from individual peptide sequencing. As the linked portions grew larger, they provided some guidance on to which peptides we should focus our attention. Table 4 shows all the peptides that have been fully sequenced and that are sufficient to provide full coverage and even redundancy to guarantee certainty as to the correctness of the final protein amino acid sequence. There were many more ions (albeit of low intensity) in the mass spectra than the ones that translated into the peptide list of Table 4. The alignment of the peptides from the various sets is shown in Fig. 23.

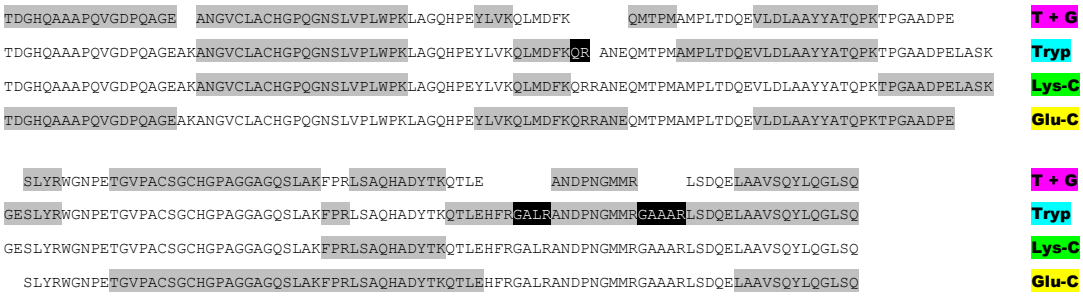

Figure 23. Amino acid sequence of cytochrome c_4 from *Thiocapsa roseopersicina*. Peptides from the four different digest sets are aligned. The highlighted tryptic peptides QR, GALR and GAAAR were not found in the LC/ESI-MS and MS/MS data, and were found only in the DI mode data after their presence was indicated by alternative digest sets (see text). The visible gaps in the Trypsin+Glu-C and Glu-C sets correspond to peptides that were not found in any MS spectrum.

In order to eliminate the possibility of errors due to homeometric peptides (different peptides with similar theoretical mass spectra [Frank, *et al.* 2007]), sequences of a sufficient number of peptides to span the entire protein primary structure were validated by matching their DI-CID-MS/MS spectra (with detection of product ions by FTICR) to theoretical product ion tables (such as the one in Fig. 21b). Frank, *et al.* 2007 have shown that the probability of occurrence of homeometric peptides becomes negligible with high-resolution tandem mass spectra.

Finishing the sequence - differentiation of leucine and isoleucine

We addressed the leucine/isoleucine ambiguity (Xle) only after the entire protein sequence had been assembled. Conventional ECD induces secondary reactions in radical **z** ions, which often suffer side-chain losses leading to the formation of **w** ions. [Savitski, *et al.* 2007] Having different side chains, a Leu residue suffers a mass loss of 43.0542 Da whereas Ile loses 29.0397 Da. This type of reaction has been estimated to occur for about 44% of the cases in conventional ECD. [Savitski, *et al.* 2007] We, therefore, searched the ECD mass spectra for **w** ions containing Xle residues on their N-terminal by matching the spectra to theoretical product ion tables (Baker and Clauser)[*]. In total, 9 out of the 19 Xle's were resolved in this way. The remaining 10 Xle residues were assigned by homology comparison using the SwissProt blast tool.[♣] The final protein sequence is shown in Fig. 24.

[*] http://prospector.ucsf.edu
[♣] http://www.expasy.org/tools/blast/

TDGHQAAAPQ VGDPQAGEAK ANGVCLACHG PQGNSLVPLW PKLAGQHPEY
IVKQLMDFKQ RRANEQMTPM AMPLTDQEVL DLAAYYATQP KTPGAADPEL
ASKGESLYRW GNPETGVPAC SGCHGPAGGA GQSLAKFPRL SAQHADYTKQ
TLEHFRGALR ANDPNGMMRG AAARLSDQEL AAVSQYLQGL SQ

Figure 24. Amino acid sequence of cytochrome c_4 from *Thiocapsa roseopersicina*. Leucines or isoleucines resolved by ECD are underlined, whereas the assignments made by homology are highlighted with grey background.

Confirmation of molecular weight

Interestingly, after the protein sequence had been completed, the calculated mass for the apoprotein (with reduced cysteines) was about 400 Da less than the initially estimated 20749.0 Da, i.e. 20348.9 Da (both monoisotopic masses). This difference was found to be a consequence of the treatment used to remove the heme groups; in acidic conditions, mercury chloride was used to break the covalent bonds between the heme vinyl chains and the cysteine sulfurs. After heme detachment, two mercury (Hg) ions remained coordinated to the heme binding motifs CXXCH [Van Beeumen 1991] (one mercury ion per heme binding site – see Fig. 9). We confirmed the presence of two Hg by simulating isotope distribution profiles for the major ESI ions (Fig. 25, 18–20 positive charges; compare with Fig. 18b). The mercury ions were removed only later upon reduction by dithiothreitol.

Adding the mass of two heme rings (2 × 616.5) to the apoprotein mass of 20348.9 yields the above-mentioned value of 21581.9 Da for holoprotein monoisotopic mass.

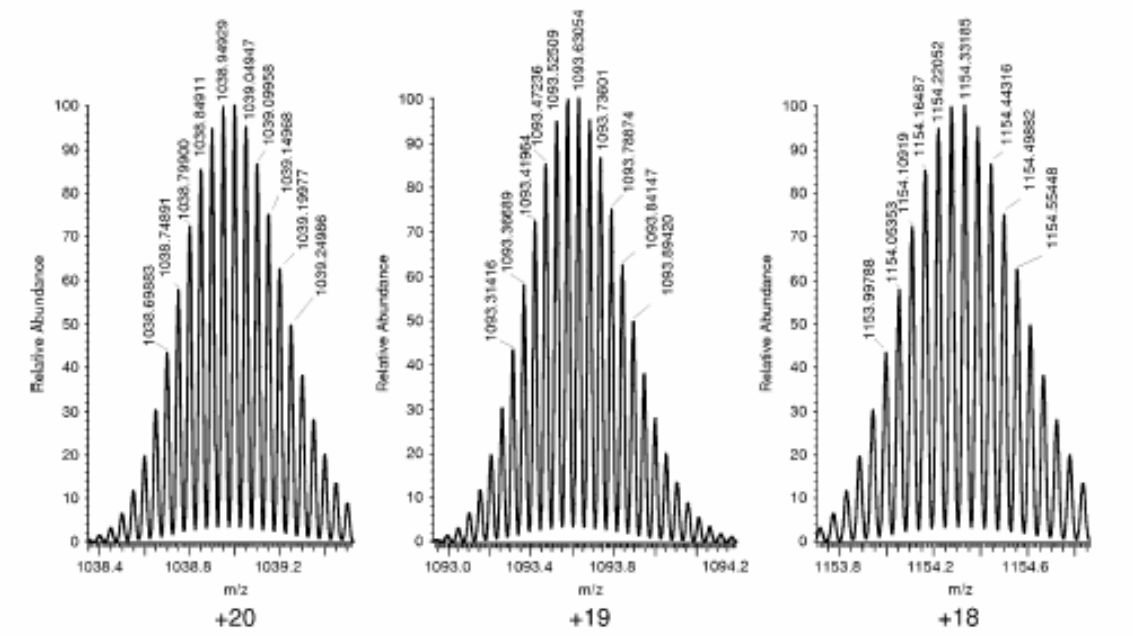

Figure 25. Simulated isotope distribution (M/ΔM = 50 000) of the ions with 20, 19 and 18 positive charges based on calculated mass from sequence (Fig. 24) and adduct with two Hg. See Fig. 18b for measured profiles; m/z values were within 5 ppm of the calculated values.

Phylogenetics of cytochrome c_4 *from Thiocapsa roseopersicina*

The amino acid sequence of cytochrome c_4 from *Thiocapsa roseopersicina* was compared the sequences of other proteins. We have used the online tool BLAST 2.0 on the Swiss-Prot/TrEMBL database* to establish the degrees of sequence similarity between our newly discovered cytochrome and cytochromes of diverse bacterial species. Table 5 displays the top hits of the online database search.

* http://www.expasy.org/tools/blast/

species	cyt	identity	positive	gaps	taxonomy	metabolism
Pseudomonas fluorescens	**c**	51%	65%	3%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Hahella chejuensis	**c553**	50%	66%	2%	γ-proteobacteria; Oceanospirillales	O_2 respiration
Idiomarina baltica	**c4**	50%	64%	5%	γ-proteobacteria; Alteromonadales	O_2 respiration
Reinekea	**c**	50%	58%	6%	unclassified γ-proteobacteria	fac. anaerobic
Pseudomonas entomophila	**c4**	49%	64%	2%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Pseudoalteromonas tunicata	**c4**	49%	63%	5%	γ-proteobacteria; Alteromonadales	fac. anaerobic
Pseudomonas mendocina	**c**	48%	64%	2%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Pseudomonas putida	**c**	48%	64%	2%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Alkalilimnicola ehrlichii	**c**	48%	63%	1%	γ-proteobacteria; Chromatiales	fac. anaerobic
Pseudomonas alcaliphila	**c551**	48%	63%	2%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Nitrococcus mobilis	**c4**	48%	62%	2%	γ-proteobacteria; Chromatiales	NO_2^- oxidation
Psychromonas ingrahamii	**c**	48%	61%	1%	γ-proteobacteria; Alteromonadales	O_2 respiration
Aeromonas salmonicida	**c4**	47%	64%	4%	γ-proteobacteria; Aeromonadales	fac. anaerobic
Photobacterium sp.	**c4**	47%	64%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Vibrio angustum	**c4**	47%	64%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Dechloromonas aromatica	**c**	47%	63%	4%	β-proteobacteria; Rhodocyclales	ClO_4^{2-} reduction
Nitrosococcus oceani	**c**	47%	62%	1%	γ-proteobacteria; Chromatiales	aerobic/NH_3 urea oxidation
Marinomonas bacterium	**c4**	47%	62%	1%	γ-proteobacteria; Oceanospirillales	O_2 respiration
Nitrosospira multiformis	**c**	47%	62%	4%	β-proteobacteria; Nitrosomonadales	NH_3 oxidation
Shewanella baltica	**c**	47%	61%	7%	γ-proteobacteria; Alteromonadales	fac. anaerobic
Azotobacter vinelandii	**c4**	47%	60%	6%	γ-proteobacteria; Pseudomonadales	O_2 respiration/N_2 fixation
Oceanobacter sp.	**c4**	47%	58%	1%	γ-proteobacteria; Oceanospirillales	O_2 respiration
Alteromonadales bacterium	**c4**	46%	65%	5%	γ-proteobacteria; Alteromonadales	
Vibrio fischeri	**c4**	46%	64%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Pseudomonas aeruginosa	**c4**	46%	63%	2%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Rhodoferax ferrireducens	**c**	46%	63%	3%	β-proteobacteria; Burkholderiales	fac. anaerobic
Photobacterium profundum	**c4**	46%	63%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Delftia acidovorans	**c**	46%	62%	3%	β-proteobacteria; Burkholderiales	O_2 respiration
Aeromonas hydrophila	**c4**	46%	62%	4%	γ-proteobacteria; Aeromonadales	fac. anaerobic
Alteromonas macleodii	**c4**	46%	61%	5%	γ-proteobacteria; Alteromonadales	O_2 respiration
Pseudomonas stutzeri	**c4**	46%	61%	6%	γ-proteobacteria; Pseudomonadales	O_2 respiration
Shewanella amazonensis	**c**	46%	61%	7%	γ-proteobacteria; Alteromonadales	fac. anaerobic
Thiobacillus denitrificans	**cox**	46%	60%	5%	β-proteobacteria; Hydrogenophilales	fac. anaerobic/S oxidation
Burkholderia phymatum	**c**	46%	59%	5%	β-proteobacteria; Burkholderiales	O_2 respiration
Vibrio parahaemolyticus	**c4**	45%	65%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Pseudoalteromonas haloplanktis	**c4**	45%	64%	5%	γ-proteobacteria; Alteromonadales	O_2 respiration
Idiomarina loihiensis	**c4**	45%	63%	5%	γ-proteobacteria; Alteromonadales	O_2 respiration
Vibrio cholerae	**c4**	45%	63%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Vibrio vulnificus	**c4**	45%	62%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Shewanella oneidensis	**c**	45%	62%	5%	γ-proteobacteria; Alteromonadales	fac. anaerobic
Neptuniibacter caesariensis	**c4**	45%	61%	1%	γ-proteobacteria; Oceanospirillales	O_2 respiration
Vibrio splendidus	**c4**	45%	61%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Shewanella sp.	**c**	45%	61%	7%	γ-proteobacteria; Alteromonadales	fac. anaerobic
Burkholderia xenovorans	**c**	45%	57%	5%	β-proteobacteria; Burkholderiales	O_2 respiration
Burkholderia thailandensis	**c**	45%	57%	7%	β-proteobacteria; Burkholderiales	O_2 respiration
Burkholderia pseudomallei	**c**	45%	56%	7%	β-proteobacteria; Burkholderiales	O_2 respiration
Burkholderia mallei	**c**	45%	56%	7%	β-proteobacteria; Burkholderiales	O_2 respiration
Vibrio alginolyticus	**c4**	44%	65%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Vibrio harveyi	**c4**	44%	64%	5%	γ-proteobacteria; Vibrionales	fac. anaerobic
Psychromonas bacterium	**c4**	44%	62%	1%	γ-proteobacteria; Alteromonadales	anaerobic
Legionella pneumophila	**c4**	44%	61%	2%	γ-proteobacteria; Legionellales	O_2 respiration
Pseudoalteromonas atlantica	**c**	44%	60%	5%	γ-proteobacteria; Alteromonadales	O_2 respiration
Moritella sp.	**c4**	44%	58%	5%	γ-proteobacteria; Alteromonadales	anaerobic
Alcanivorax borkumensis	**c4**	44%	58%	8%	γ-proteobacteria; Oceanospirillales	O_2 respiration
Marinobacter algicola	**c553**	43%	63%	2%	γ-proteobacteria; Alteromonadales	O_2 respiration
Nitrosomonas europaea	**c**	43%	58%	5%	β-proteobacteria; Nitrosomonadales	NH_3 oxidation
Nitrosomonas eutropha	**c**	43%	56%	5%	β-proteobacteria; Nitrosomonadales	NH_3 oxidation

Table 5. Species containing cytochromes most closely related in sequence to cytochrome c_4 from *Thiocapsa roseopersicina*. Identity – percentage of identical amino acids in the sequence. Positive – percentage of similar, albeit distinct amino acids in the sequence. Gaps – percentage of insertions and deletions in the sequence. Search was performed with the sequence shown in Fig. 24.

UV-Vis and CD spectroscopy

Assignment of the peaks in UV–Vis spectroscopy

The UV–Vis electronic absorption spectra at pH 8.5 and 25 °C of the purified sample and of the two redox states at a concentration of 38 μM, and of a redox potential controlled sample (300 mV) at a concentration of 17 μM, at 60 °C, are shown in Fig. 26.

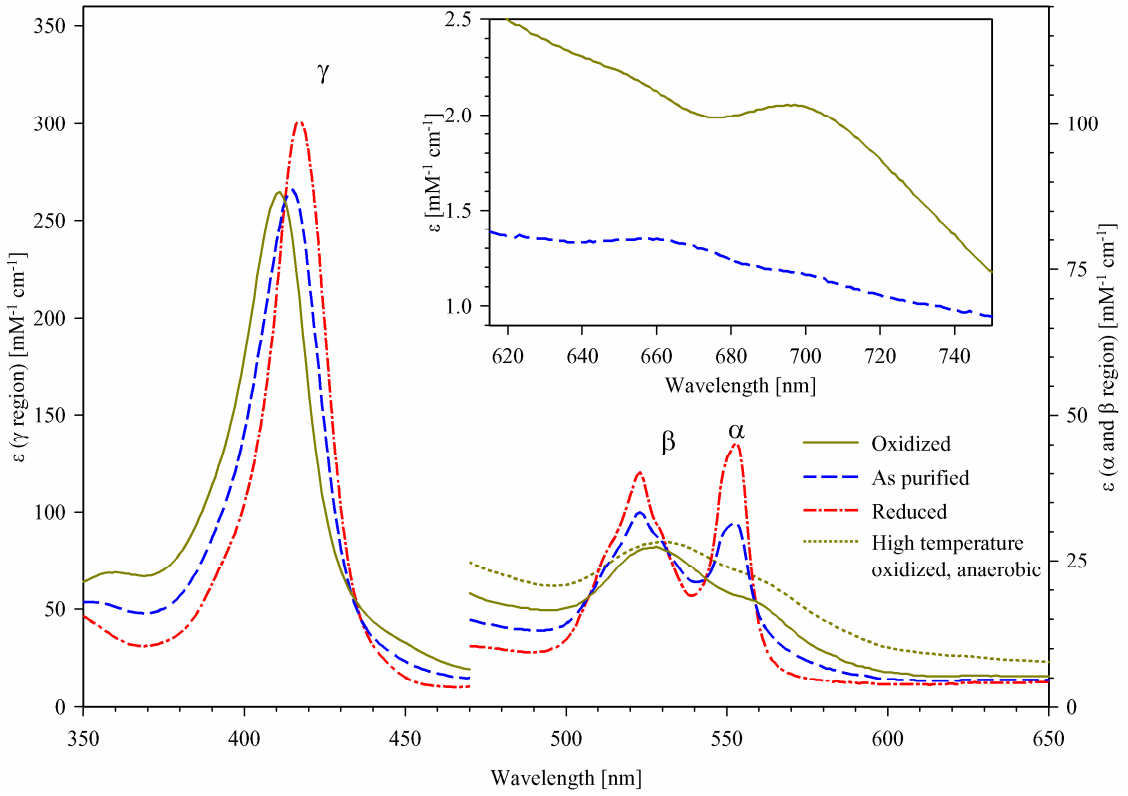

Figure 26. Absorption spectra of cytochrome c_4 from *Thiocapsa roseopersicina* in different redox states and temperatures. Reduced spectrum at high temperature (not shown) overlaps with the spectrum taken at room temperature. Inset: Enlarged portion of the spectra in the red region (620–750 nm) measured on a more concentrated sample.

The spectra are typical of low-spin c-type cytochromes [Saraiva, *et al.* 1995]. In the reduced state, α, β and γ (Soret) bands appear at 553 nm (43.5 mM^{-1} cm^{-1}), 523 nm (39 mM^{-1} cm^{-1}) and 417 nm (289 mM^{-1} cm^{-1}), respectively. The ratio of the absorption maxima of α and β bands is α/β = 1.13. In the oxidized state, the γ (Soret) band shifts to 411 nm and both α and β bands red-shift and become characteristically broad (Table 6).

	Reduced *T.r.* cytochrome c_4		Oxidized *T.r.* cytochrome c_4	
	λ [nm]	ε_{max} [mM^{-1} cm^{-1}]	λ [nm]	ε_{max} [mM^{-1} cm^{-1}]
γ	417	289	411	202
β	523	38.4	529	20.8
α	553	43.5	559	13.8
	Reduced horse heart cytochrome		Oxidized horse heart cytochrome	
	λ [nm]	ε_{max} [mM^{-1} cm^{-1}]	λ [nm]	ε_{max} [mM^{-1} cm^{-1}]
γ	415	122	411	97.8
β	521	16.2	529	8
α	550	28	559	7

Table 6. UV–Vis absorption parameters (band wavelengths and respective molar absorptivities). Comparison of cytochrome c_4 from *Thiocapsa roseopersicina* with horse heart mitochondrial cytochrome *c*. *

In addition, the visible spectrum of a more concentrated solution (106 μM) of oxidized cytochrome (Fig. 26, inset) exhibits two weak bands. The 695 nm band (which could be deconvoluted into two different peaks at 698 and 713 nm) indicates an axial methionine ligand [Saraiva*, et al.* 1990, Santucci and Ascoli 1997]. A low intensity band is also discernable at 651 nm. Upon cytochrome reduction, it is red-shifted to 663 nm. This band was assigned to the iron high-spin state [Miles*, et al.* 1993]. The very low intensity of this high-spin marker indicates that, in both redox states, only a small amount of hemes exists in 5-coordination or with a weak-field ligand at the sixth coordination site under these conditions (25 °C and pH 8.5).

Assignment of the peaks in CD spectroscopy

The far-UV CD revealed the regular and characteristic protein CD spectra (Fig. 27). These spectra were used for secondary structure calculations.

The CD spectra in the visible region for both redox states corroborated the His–Fe–Met heme coordination (Fig. 27). A positive (405 nm) and a negative (419 nm) Soret band were found in both redox states [Santucci and Ascoli 1997].

* Horse heart cytochrome *c* data are from László Zimányi, Institute of Biophysics, Biological Research Centre.

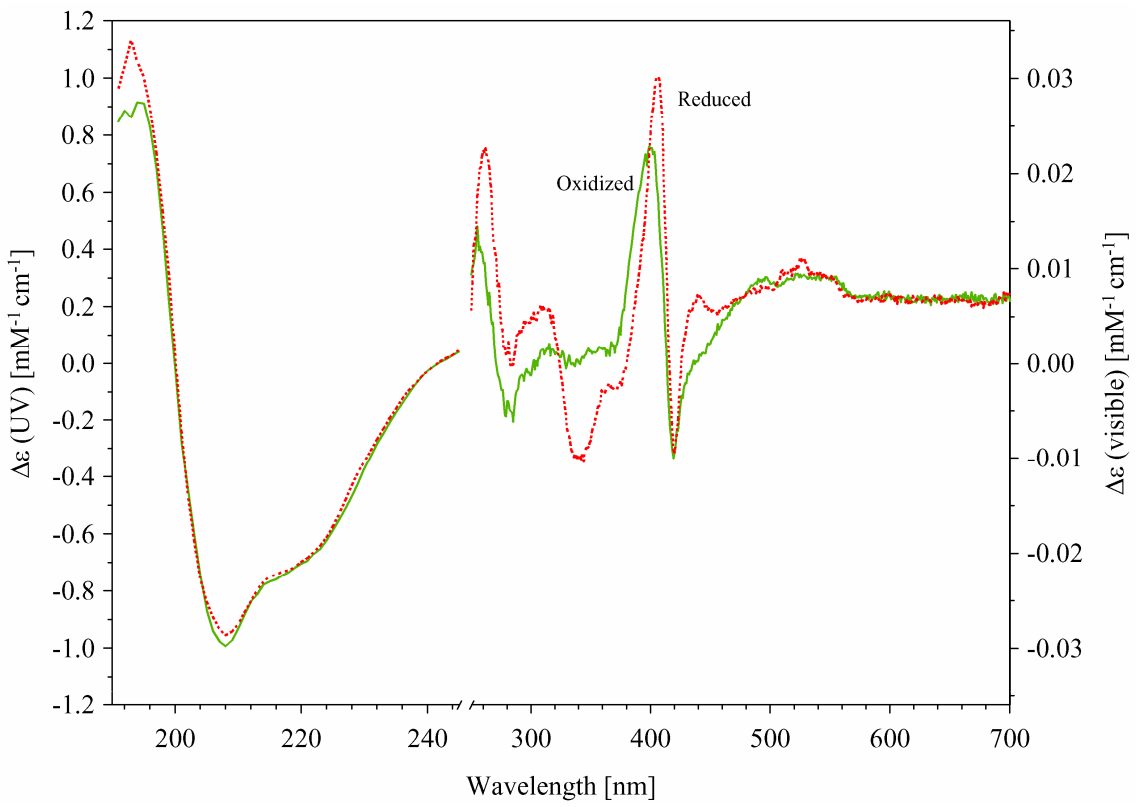

Figure 27. CD spectra of cytochrome c_4 at pH 8.5 in different redox states.

Potentiometry

Redox titration curves using α and β bands of cytochrome c_4 from *T. roseopersicina* at pH 7.7, at 25 °C are presented in Fig. 28. Fitting one redox transition function to the data yielded E^o = 252 ± 1 mV and Γ = 27.7 ± 1.4 mV (0.93 ± 0.05 electrons) for the oxidative direction and E^o = 254 ± 1 mV, and Γ = 24.8 ± 1.2 mV (1.03 ± 0.05 electrons) for the reductive direction. There was no observable hysteresis; the redox reaction was fully reversible. The deviation from n = 1 is minimal, which indicates that the redox potentials of the two hemes present in the cytochrome molecule differ by less than 60 mV [Meyer, *et al.* 1991].

In order to calculate the midpoint potential of each individual heme, we fixed n = 1 for both hemes and fitted a linear combination of two equally-contributing redox transition functions to the oxidative direction data set (Fig. 28a). This resulted in midpoint potentials of 237 ± 5 and 268 ± 6 mV.

The cytochrome was also redox-titrated at 60 °C (Fig. 28b).

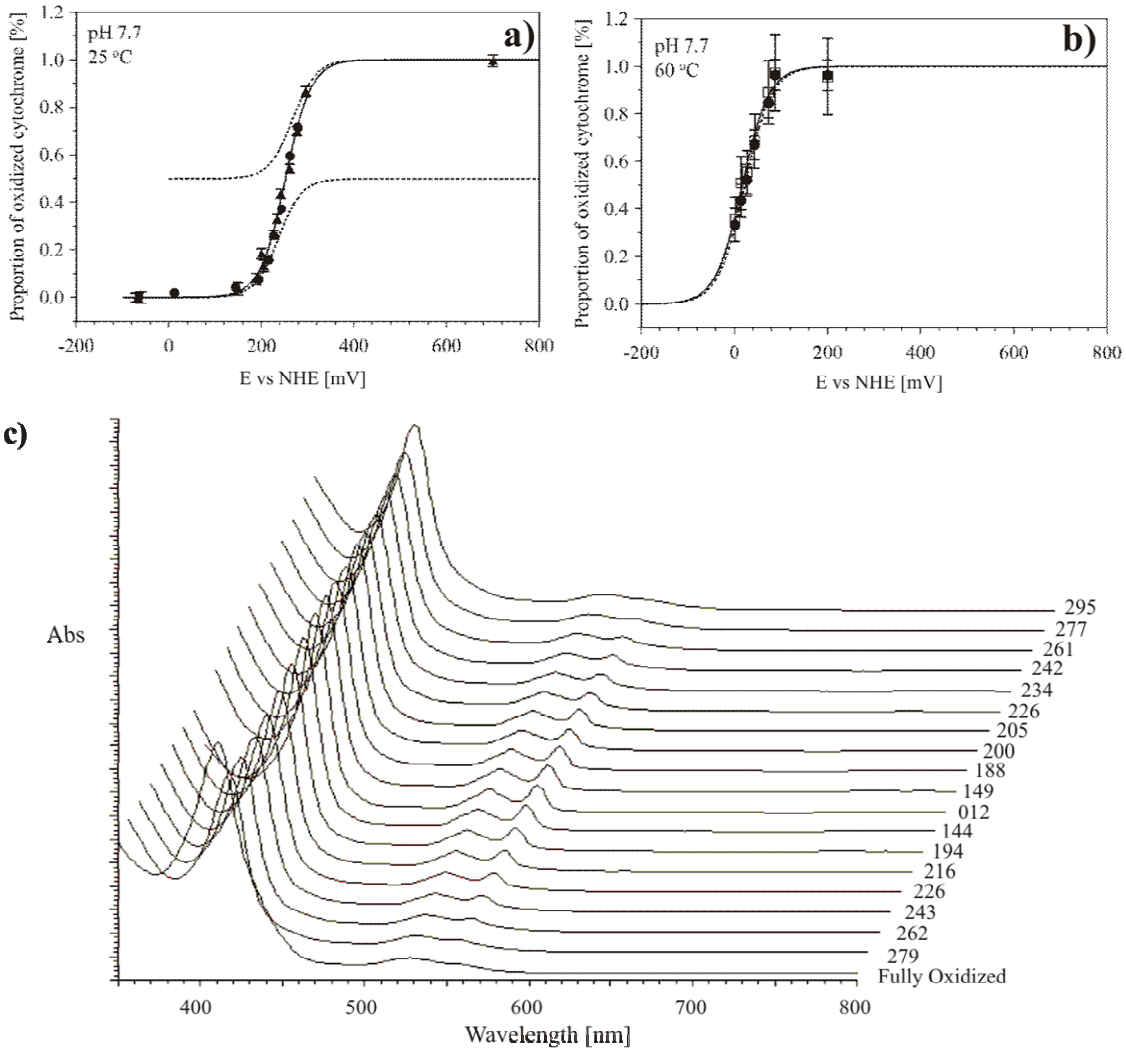

Figure 28. Redox titration of cytochrome c_4 from *Thiocapsa roseopersicina* at 25 and 60 °C and at pH 7.7. The absorption spectrum in α and β peak region (500–600 nm) and red region (620–750 nm) was evaluated. In the 500–600 nm region each spectrum was mixed with the spectra of fully reduced, fully oxidized and high temperature oxidized (where applicable) states. The proportion of the spectrum of the oxidized state was plotted against the redox potential relative to the normal hydrogen electrode. In the red region a Gaussian function was fitted to the 650 nm band of the absorption spectrum. The calculated peak intensity was plotted against the redox potential relative to the normal hydrogen electrode.
a) Redox titration at pH 7.7 and 25 °C. ● measured points in the reductive direction; ▲ measured points in the oxidative direction; — sum of fitted Nernst functions; - - - individual Nernst functions.
b) Redox titration at pH 7.7 and 60 °C. ● measured points for the 500–600 nm region; □ measured points for the red region; —, - - - fitted Nernst functions.
c) Full spectra of redox titration at pH 7.7 and 25°C. Each spectrum is labelled with the respective ambient potential in mV vs. NHE. Titration proceeded at first in reductive direction and subsequently in oxidative direction.

Throughout the redox potential range applied, it was always possible to combine the measured spectrum using the spectra of the fully oxidized and the fully reduced states

measured at 25 °C (Fig. 26), and the spectrum of the high temperature oxidized state. This state was obtained when the redox potential was fixed to an oxidative potential value (i.e. 200 mV), the sample was kept anaerobic, and temperature was increased to 60 °C. This state has a distinct spectrum (see below and Fig. 26).

The 650 nm bands were also used to estimate the redox transition at high temperature. As the reduced cytochrome was converted to the oxidized form, the 650 nm band appeared and exhibited a typical redox transition, i.e. the 650 nm band followed the redox transition throughout the applied potential (Fig. 28b and Table 7).

Temperature	Wavelength of detection	pH	Titration direction	Single transition		Double transitions			
°C	nm			E^0 [mV]	n	E^0 [mV]	n	E^0 [mV]	n
25	550	7.7	ox.	252±1	0.93±0.05				
25	550	7.7	red.	254±1	1.03±0.05	237±5	1 (50%)	268±6	1 (50%)
60	550	7.7	ox.	35±6	1.5±0.3				
60	650	7.7	ox.	34±9	1.5±0.5				

Table 7. Redox potentials and Nernst numbers of cytochrome c_4 from *Thiocapsa roseopersicina* at different temperatures, as calculated from different regions of the visible absorption spectra.

Cytochrome conformational states as a function of temperature

UV–Vis absorption spectra

Aerobic measurements.

The temperature dependences of the absorption of mostly reduced and mostly oxidized cytochrome under aerobic conditions are presented in Fig. 29. The oxidation of the samples was achieved chemically.

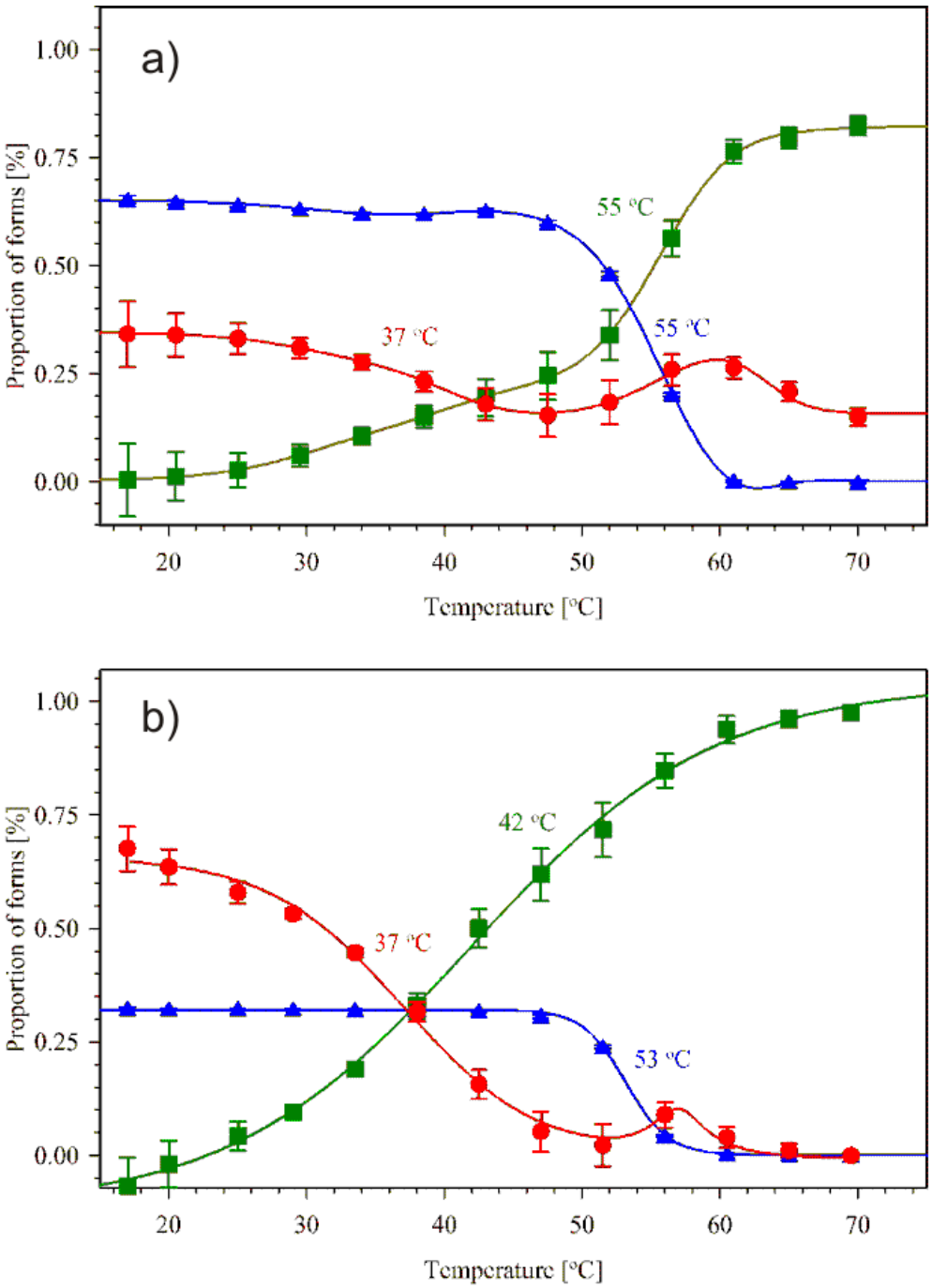

Figure 29. Temperature dependence of different redox forms of cytochrome c_4 at pH 8.5, under aerobic conditions. a) Mostly reduced (as purified), b) chemically oxidized samples. The absorption spectrum in α and β peak region (500–600 nm) was evaluated. Each spectrum was combined by using the spectra of the fully oxidized and the fully reduced states measured at 25 ºC, and the spectrum of the high-temperature oxidized state obtained from high-temperature redox titration. The proportion of each component spectrum was plotted against temperature. ▲, proportion of the spectrum of the reduced state; ●, proportion of the spectrum of the oxidized state; ■, proportion of the spectrum of the high-temperature oxidized state.

The spectra at each temperature were deconvoluted as a mix of three basic spectra (fully oxidized, fully reduced and high-temperature forms) fitting the α and β regions of

the spectrum. The changes in the proportions of each component are presented in Fig. 29.

The main feature of the temperature dependence of cytochrome is that both the reduced and the oxidized forms of the sample disappear on increase of the temperature, giving way to a new conformer – a high-temperature form (Fig. 26). The reduced cytochrome is more resistant to heat than the oxidized form. The transition point for the oxidized cytochrome is around 37 °C, while for the reduced cytochrome it was measured as 53 °C and 55 °C. It is clear from the transition curves that the reduced cytochrome first undergoes oxidation and only the oxidized cytochrome is converted into the high-temperature form. This is evidenced by a small bump in the temperature dependence curve of the oxidized form at the end of the transition of the reduced form in the case of the chemically oxidized sample (Fig. 29b) and by the increase and subsequent decrease in the oxidized form in the case of the mostly reduced sample (Fig. 29a).

The transition points of the high-temperature form follow the transition of the originally predominant oxidation state of the enzyme. It was 42 °C in the case of the chemically oxidized sample, and 55 °C in the case of the mostly reduced sample.

The UV–Vis electronic absorption spectrum of a highly concentrated (120 μM) cytochrome was also monitored under aerobic conditions in the temperature range 17–47 °C, with focus on the 600–750 nm peaks. The 695 nm band disappeared upon heating, with a midpoint temperature of 38 °C (Fig. 30). The process was mostly, but not fully reversible (88% of the 695 nm peak area was recovered). The 650 nm band disappeared too, but the transition point could not be determined because of the low intensity of the band.

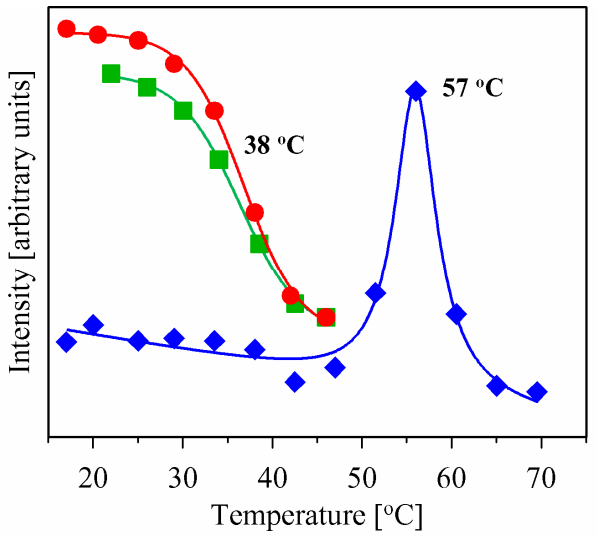

Figure 30. Temperature dependence of the spectral parameters of the absorption spectra for mostly reduced (as purified) and chemically oxidized cytochrome c_4 at pH 8.5, under aerobic conditions. The absorption spectrum in the red region (620–750 nm) was evaluated. A Gaussian function was fitted to the 650 and 695 nm bands of the absorption spectra. The calculated peak intensities were plotted against temperature. ●, peak intensity of the 695 nm band of the chemically oxidized protein measured on increase of the temperature; ■, peak intensity of the 695 nm band of the chemically oxidized protein measured on decrease of the temperature; ♦, measured peak intensity of the 650 nm band of the mostly reduced (as purified) protein on increase of the temperatures; —, fitted transition curves.

When the same wavelength region of a sample with 80% reduced state was monitored over a larger temperature span (15–70 °C), the 650 nm band appeared transiently at around 57 °C (Fig. 30), indicating the appearance of a transient intermediate high-spin state [Miles*, et al.* 1993].

Anaerobic and potentiostated measurements.

When the redox potential was kept reductive (i.e. -100 mV) for all temperatures, the cytochrome remained reduced independently of the temperature. There was no sign of a distinct high-temperature reduced state. At an oxidative potential (i.e. 300 mV), the oxidized state was converted to a high-temperature oxidized form with a midpoint temperature of 34 °C.

Visible CD spectra

Aerobic measurements.

The CD spectra in the Soret region of both the oxidized and the reduced state were measured in the temperature interval 25–70 °C at pH 8.5 (Fig. 31).

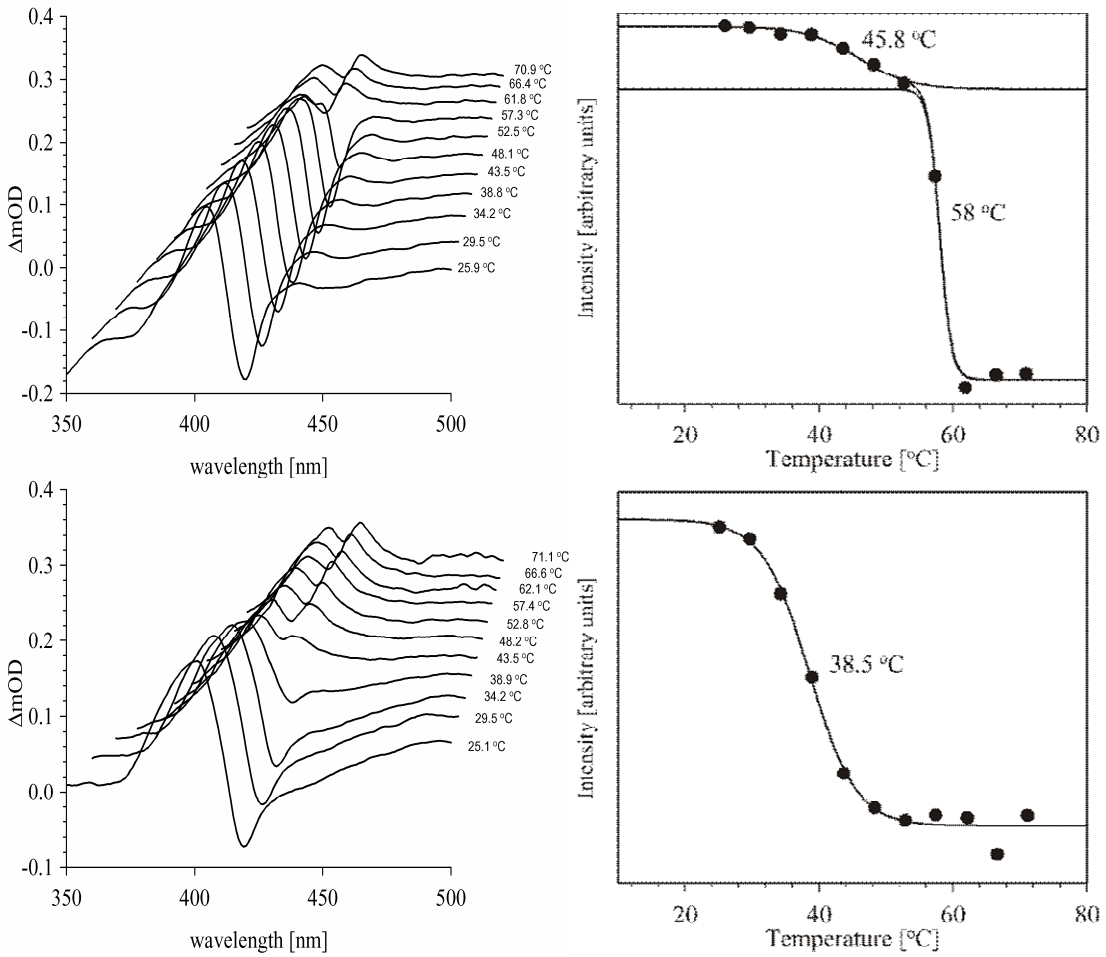

Figure 31. Soret region of the CD spectrum of cytochrome c_4 (30 µM) as a function of temperature and under aerobic conditions. Top: mostly reduced (as purified) protein. Bottom: chemically oxidized protein. Left: spectral series depicting alterations occurring in the heme pocket. Right: state transition curves calculated from the CD spectra using the second V vector of the SVD transformation.

The spectra were deconvoluted by the SVD method and the transition temperatures were determined from the 1st and 2nd V matrix vectors [Shrager 1986]. The cytochrome in the two oxidation states was transformed into the high-temperature form, but at different temperatures. Two transitions could be observed for the reduced cytochrome, at 45.8 and 58 °C (Fig. 31a), while the oxidized protein displayed only one transition, at 38.5 °C (Fig. 31b). The structure of the CD signals changed during the transitions: the positive signal at 405 nm and the negative signal at 416 nm were replaced by a signal with a double positive peak, one at around 393 nm and the other at around 417 nm. All these changes in the spectrum were mostly, but not fully reversible (up to 80 °C), indicating that the Met–Fe bonds were restored in the majority of the cytochrome molecules.

Anaerobic and redox potential controlled measurements.

The temperature dependence of the CD spectra in the Soret region (350–450 nm) was measured in the range 20–80 °C, while the redox potential was set to -300 mV (not shown). The sample was cooled back down to 20 °C, and the temperature dependence measurement was repeated at 500 mV (Fig. 32). The sample was cooled down again to 20 °C and two spectra were taken, first at 500 mV, and then at -300 mV. Both spectra were identical to the original spectra of the oxidized and reduced states, respectively.

The Soret signal in the enforced reduced state remained unchanged throughout the temperature range 20–80 °C, indicating that the Met–Fe bond remained intact. In the enforced oxidized state, the spectrum changed considerably, exhibiting a sigmoid transition with a midpoint at 32 °C (Fig. 32).

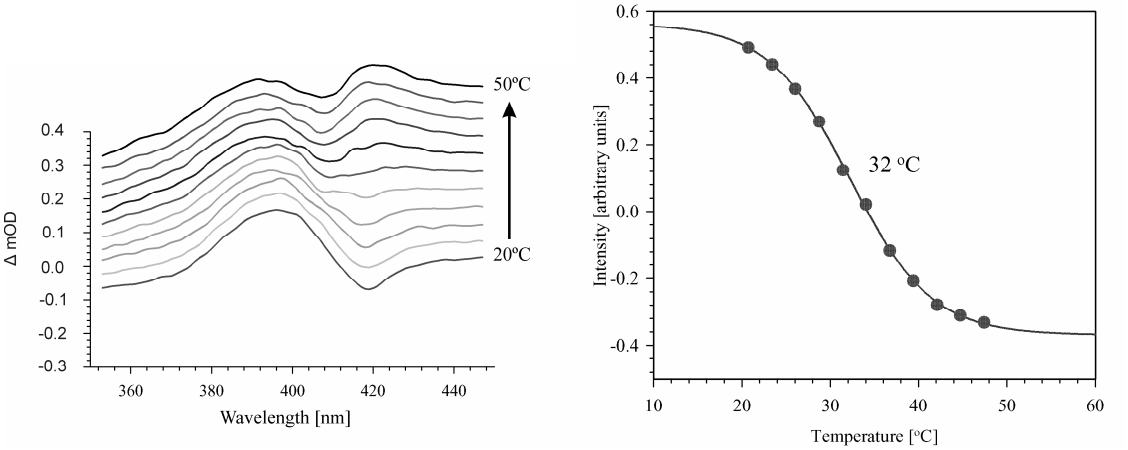

Figure 32. Soret region of the CD spectrum of cytochrome c_4 (30 µM) as a function of temperature, under anaerobic conditions and at a potential of 500 mV.
Left: CD spectra in the Soret band region (CD-vis, 350–450 nm).
Right: The spectral series were transformed by using the second V vector of the SVD method, and the resulting transition curve was plotted against temperature.

At an oxidative potential (500 mV), the oxidized state was converted to the high-temperature oxidized form. The high-temperature spectrum was similar, but not identical to the high-temperature spectrum in the aerobic CD measurement; it also presented two positive peaks, at the same positions. A similar spectrum was earlier observed for horse heart cytochrome c when the protein was embedded in phospholipid vesicles [Oellerich, et al. 2002 Oellerich, et al. 2004].

Secondary structure calculations from far-UV CD spectra

The proportion of the different secondary structure elements, calculated from the far-UV CD spectra of cytochrome c_4 as a function of temperature, is presented in Fig. 33.

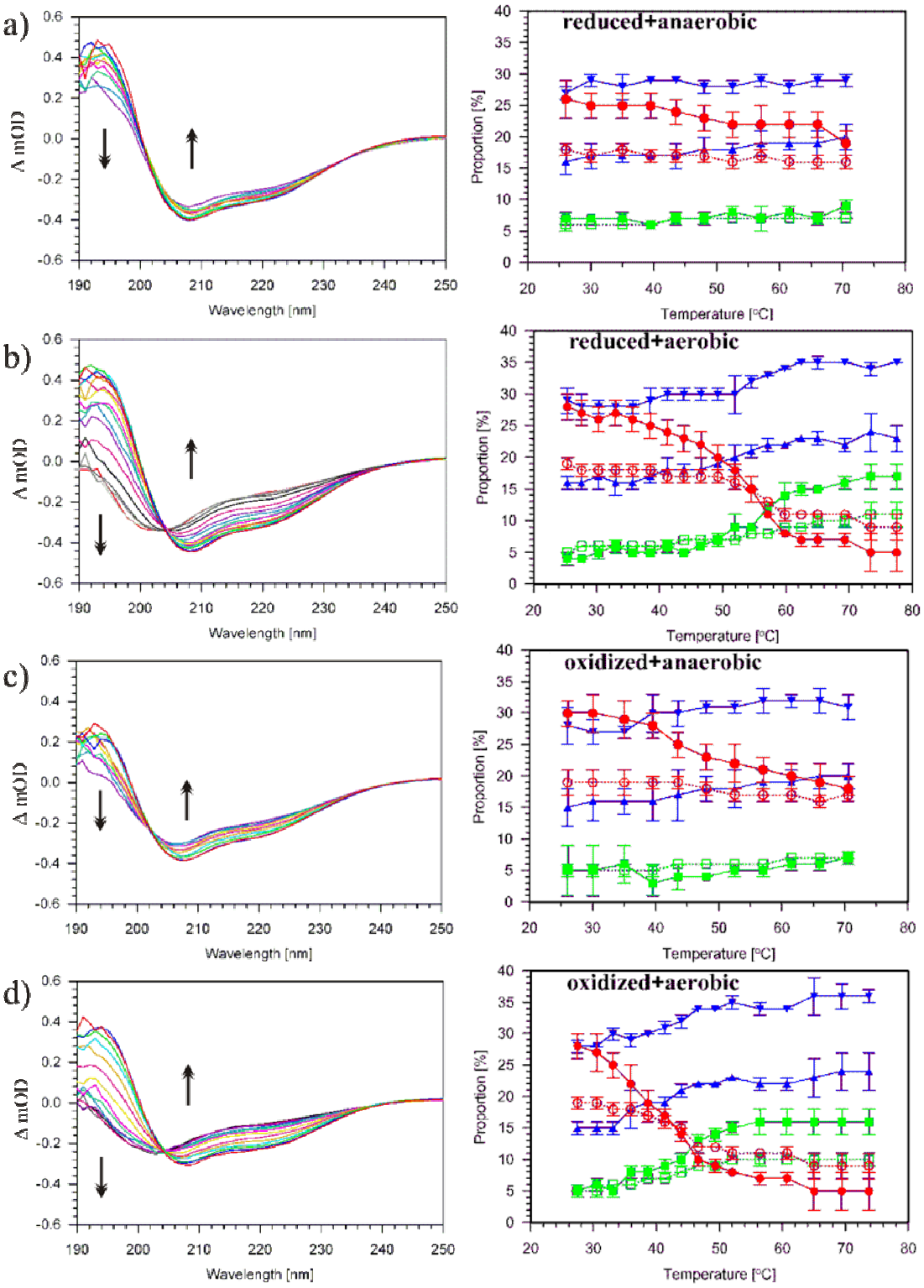

Figure 33. Temperature dependence of secondary structure data (right side) calculated from the far-UV range (190–250 nm) of the CD spectra (left side). Arrows in spectra depict direction of alterations with increase of the temperature. Six different structural elements (regular helix, distorted helix, regular strand, distorted strand, turns and unordered structures) were calculated, and their proportions plotted against the temperature.

-●- α helix (regular); -○- α helix (distorted); -■- β strand (regular); -□- β strand (distorted); -▲- turns; -▼- unordered. a) reduced anaerobic sample, b) reduced aerobic sample, c) oxidized anaerobic sample, d) oxidized aerobic sample.

The far-UV CD spectrum at 25 °C is essentially the same for the reduced and oxidized states under both aerobic and anaerobic conditions (spectra are like shown in Fig. 27); and therefore the relative amounts of secondary structures are practically identical in all cases at 25 ºC (Fig. 33) . The calculated secondary structure yielded roughly 45% α helix and a very low (~10%) β sheet content. These results are consistent with the recently obtained crystal structures of *P. stutzeri* and *A. ferrooxidans* cytochrome c_4 (PDB# 1ETP [Kadziola and Larsen 1997] and 1H1O [Abergel*, et al.* 2003]).

Under anaerobic conditions, the proportion of the β structures did not change in either the oxidized or the reduced state of the protein (Fig. 33a and c). The same held true for the number of α helices (the helix distorted value is practically constant throughout the temperature range). In both oxidation states, however, the length of the α helices decreased as the temperature increased. In the reduced state, the α helices were converted solely to turn structures, but this change was rather small. In the oxidized state, both the turn and the random structures gained from this transition and the transition itself was more pronounced.

Under aerobic conditions, the two redox states of the protein behaved differently (Fig. 33b and d). In the reduced state, there were clearly two transitions. In the first (transition point at around 44 °C), the number of α helices was stable, but their length decreased rapidly. The helices were converted mostly to turn structures, similarly to the anaerobic case. Unordered and β structures remained practically intact in this temperature range. In the second transition, with a transition point of 56 °C, both the number and the length of α helices decreased rapidly, while all the other structures gained from this transition. It is interesting that not the number, but the length of the β strands is involved in this conversion. In the oxidized state, there is just one transition, with 42 °C as the midpoint; α helices decreased rapidly (in both length and number) and all the other structures increased. In this case, there was also a slight increase in the number of β strands.

Differential scanning calorimetry

The enthalpic transitions measured under aerobic conditions by DSC for both chemically adjusted redox states of the cytochrome at pH 8.5 are summarized in Table 8. The oxidized form presents only one strong peak, at 42 °C. The reduced sample exhibits two main peaks (44–50 and 60– 65 °C), which we could deconvolute into four transitions at different temperatures.

oxidized			reduced			
Transition temperature	42 °C		44 °C	50 °C	60 °C	64 °C
First run	401.9			549.3		
Second run	382.0		225.5	205.3	381.7	308.5
Third run	319.3		365.2		156.1	3.9
Fourth run	109.6		n. a.	n. a.	n. a.	n. a.

Table 8. Transition heats (kJ mole^{-1}) of oxidized and reduced cytochrome c_4 from *Thiocapsa roseopersicina* under aerobic conditions as calculated from differential scanning calorimetry.

Discussion

Classification of the newly purified cytochrome c from T. roseopersicina

Before my work started, four different cytochromes had been purified and characterized from *Thiocapsa roseopersicina* [Zorin and Gogotov 1980, Trüper and Rogers 1971, Korsunsky, *et al.* 1982 Korsunsky, *et al.* 1984]. A cytochrome of 180 kDa was purified by Trüper and Rogers 1971, a homotetramer cytochrome (4×11 kDa) by Korsunsky, *et al.* 1984 and a cytochrome of 55 kDa also by Korsunsky, *et al.* 1982. The fourth is a flavocytochrome c_{552} purified by Zorin and Gogotov 1980, and possesses a di-heme subunit with a similar mass (~21 kDa) to the cytochrome c_4 discussed here, but with a midpoint redox potential of 10 mV vs NHE. Cytochrome c_4 was found to have a high redox potential for both hemes (237 and 268 mV at pH 7.7 at 25 ºC), which is typical for the cytochrome c_4 family. The observed split α band and the α/β absorption peak ratio for cytochrome c_4 in the fully reduced form support this classification. The ratio is low as compared, for instance, with horse heart mitochondrial cytochrome *c* (Table 6), or the flavocytochromes from *Shewanella putrefaciens* [Morris, *et al.* 1994] and from *Chromatium vinosum* [Meyer, *et al.* 1991], or cytochromes c_3 from *Desulfovibrio* [Meyer, *et al.* 1971], or cytochrome c_6 from *Chlorella fusca* [Inda, *et al.* 1997]. This low α/β absorption peak ratio is characteristic of cytochrome c_4 as in *Azotobacter vinelandii* [Leitch, *et al.* 1985], *Pseudomonas stutzeri* [Conrad, *et al.* 1995], and *Acidithiobacilus ferrooxidans* [Giudici-Orticoni, *et al.* 2000].

Furthermore, the secondary structure predictions (45% α helix, 10% β sheet and 15% turn content) are consistent with the 3D structure of *P. stutzeri* cytochrome c_4 (PDB# 1ETP [Kadziola and Larsen 1997]), indicating that *T. roseopersicina* cytochrome c_4 possesses a similar fold. Finally, the most conclusive evidence to classify the cytochrome which is subject study here comes from its primary structure determined by Mass Spectrometry (Fig. 24). The blast search* conducted on the protein sequence (Table 5) reveals that most hits are cytochrome c_4 proteins.

All these characteristics are in line with the classification of cytochrome c_4 from *T. roseopersicina*.

* http://www.expasy.org/tools/blast/

Primary structure determination by Mass Spectrometry

In the late 1980's, Biemann and co-workers achieved the outstanding feat of sequencing proteins entirely by Mass Spectrometry [Johnson and Biemann 1987, Mathews*, et al.* 1987, Hopper*, et al.* 1989]. Up until that point, this job was performed mostly by the method of Edman degradation [Edman 1950]. The proteins sequenced in Biemann's laboratories were of the thioredoxin family and all had ~11 kDa size. To achieve this, they benefited from a "soft" ionization method, the Fast Atom Bombardment, which permits the generation of intact molecular ions without significant fragmentation. This is a fundamental pre-requisite to "bottom-up" sequencing. It is possible to perform tandem Mass Spectrometry on any given peptide ion only if the ion remains intact upon ionization.

Our challenge resided in the fact that the largest proteins sequenced so far were only 11 kDa in size and we now intended to sequence a much larger protein (~21 kDa, roughly double size). Given the many improvements in recent years in the field of tandem Mass Spectrometry, we attempted and succeeded in using this method to determine the amino acid sequence of our newly purified cytochrome. However, there is still ample space for further development, and the areas where this may be achieved came to light as we tackled our concrete problem of solving the primary structure of *T. roseopersicina*'s cytochrome c_4 primary structure. We conceived a general method for the task (see outline in Fig. 8) and followed the procedure step by step.

Digestion with proteolytic enzymes

The ~11 kDa proteins sequenced by Biemann and co-workers were digested mainly by trypsin [Johnson and Biemann 1987, Mathews*, et al.* 1987, Hopper*, et al.* 1989]. The peptide mixture was separated by fractionated HPLC prior to MS, and each tryptic peptide was analysed individually. This was feasible because the digestion yielded fewer than 20 peptides and these were sufficient to get full coverage. In order to obtain overlap for the final alignment, either α-chymotrypsin or endoproteinase Glu-C (proteinase V8) were additionally used.

The cytochrome we set out to sequence was ~21 kDa in size, and thus we have used a more expanded digestion strategy involving trypsin, endoproteinase Glu-C and endoproteinase Lys-C (Fig. 12). The digestion strategy was designed to allow a gradual

build-up of smaller (and easy to sequence) peptides into larger ones that can be overlapped to yield the complete sequence of the protein.

Compilation of lists of peptide masses belonging to each digest set

Though an apparently straightforward procedure, it required several days to manually build a list of masses for all four sets of peptide mixtures. This task had not been a serious problem for previous protein primary structure determinations [Biemann and Papayannopoulos 1994, Johnson and Biemann 1987, Mathews*, et al.* 1987, Hopper*, et al.* 1989] because those proteins were only about 11 kDa in size. With a protein size of 21 kDa, the number of peptides generated by the proteolytic enzymes is dramatically higher (mostly because of the combinatorial nature of partial cleavages). We are currently writing an algorithm that can perform this tedious operation automatically. Obviously, not all the existing peptides need to be sequenced absolutely, but a sufficiently large number is required (and some redundancy recommended) to guarantee completeness and confidence on the primary structure of the protein being sequenced. [Biemann and Papayannopoulos 1994]

The great interspecies variability of the primary structure of cytochromes found in bacteria rendered the technique of peptide mass fingerprinting (Fig. 8) and database searching useless (as opposed to similar work done for instance in animal species). [Cohen*, et al.* 2005] However, it was helpful to compare the peptide masses of each digest set: some peptides were present in multiple sets; in other cases, two peptide masses of one set when added yielded, with subtraction of the mass for H_2O, the exact mass of a peptide of another set. All this proved to be useful information that facilitated the final assembly of the peptide sequences into the full protein sequence.

Peptide sequencing by MS/MS

Generally, the difficulty of sequencing each peptide depends largely on the peptide length. Firstly, it proved to be a practical impossibility to generate complete series of sequence ions (namely, the critical **b** and **y** ions) by CID for MS/MS beyond a certain length (compare Figs. 19 and 20). The other problem arising with increasing parent ion mass was the growing complexity of the MS/MS spectra, particularly because the

appearance of multiply charged product ions (+2, +3 or more) became common (see Fig. 20).

On the other hand, very small peptides also posed some problems. Some of the peptides comprising less than 6 or 7 residues (smaller than ~700 Da) produced MS/MS spectra that were difficult to interpret by the DeNovoX software (see Table 4), and thus, each of these problematic peptides required manual interpretation work. The issue probably arises because non-direct sequence **b** ions are a common occurrence for MS/MS spectra of relatively small peptides, as shown for the pentapeptide YAGFL [Harrison*, et al.* 2006], and DeNovoX was not programmed to take non-direct sequence ions into account. Thus, for example, the YAGFLoxa $\mathbf{b_5}$ ion can undergo cyclization, and after various proton transfer reactions, reopen to form isomer $\mathbf{b_5}$ ions with scrambled sequence (such as AGFLYoxa, GFLYAoxa, FLYAGoxa and LYAGFoxa).

The digestion strategy schematized in Fig. 12 is an attempt to generate sufficient peptides of a workable length (700-1500 Da) to cover the entire protein sequence. As shown in Table 3, the digest sets produced by cleavage at a higher number of sites (Glu-C+Trypsin and Trypsin sets) contained peptides more amenable to individual sequencing due to their smaller size [Cohen*, et al.* 2005]. We therefore concentrated our initial sequencing work on peptides of these two sets, and only moved on to the other digest sets to cover gaps and to obtain overall alignment.

Ideally, with a comprehensive collection of proteolytic enzymes, the task of sequencing long peptides should not be necessary. Unfortunately, having available only endoproteinases Lys-C and Glu-C, and trypsin, a certain number of long peptides had also to be sequenced in order to obtain full coverage of the protein sequence. We decided to employ the high resolving power of the FT-ICR and the available electron capture dissociation (ECD) high-energy fragmentation method to tackle the sequences of the few long peptides. We have used a variety of approaches according to the specific characteristics of each peptide.

Some peptides possessed, in the middle of their sequence, sites that were vulnerable to the secondary activities of the proteolytic enzymes used. Others simply fragmented spontaneously in solution (through acid-catalyzed hydrolysis) or upon ionization in the mass spectrometer. Either way, low-abundance ions, which were components of long peptides, were present and were easy to sequence owing to the high sensitivity of the FT-ICR unit. Thus, through the short component peptide we reached the sequence of the long peptide.

Additionally, the use of Direct Infusion of sample allowed the combination of CID of parent ions in the Ion Trap with the detection of product ions in the FT-ICR. Since the FT-ICR detector has a much higher resolution than the Linear Ion Trap (100,000 M/ΔM v.s. 4,000 M/ΔM), high resolution tandem MS spectra of long peptides could be obtained, facilitating sequence determination. In addition, the use of DI mode coupled with IT-CID-FTMS^2 provides a high resolution verification of the solution sequences (Fig. 21 b), and therefore this mode was used to verify the final protein sequence in its full extension.

Finally, the available alternative method of parent ion fragmentation, ECD, was useful in additional verification of provisional sequences for long peptides.

Chemical modifications present in the peptides

All modifications (see Table 4) probably took place during sample handling and would occur very slowly in a protein with native fold [Robinson 2002]. For the purpose of sequencing, modified peptides were considered artefacts, even though the process of deamidation, for example, might have biological significance in the aging of proteins and in the regulation of protein turnover within the living cell [Robinson 2002].

It is known that ESI may generate $[M + 16 + nH]^{n+}$ ions by oxidation at either methionyl, tryptophanyl or tyrosyl residues [Morand*, et al.* 1993]. Five peptides were found with this +16 modification and had either M, W or Y residues in their sequence (Table 4).

Gain of an ammonia molecule was probably due to replacement of a proton (H^+) with ammonium (NH_4^+); consequently $[M + NH_4]^+$, $[M + H + NH_4]^{2+}$, etc. were formed at the NSI source. During the heme removal process, 50 mM ammonium acetate buffer (pH 7) was used, and some carboxyl ammonium adducts were probably formed. The C-terminal carboxyl group of the protein appeared to be particularly prone to the H^+/NH_4^+ replacement because the only examples of this modification are two peptides arising from the C-terminus of the protein (1393.7604 and 1965.9997 Da – Table 4).

The loss of an ammonia molecule occurred on the side chains of glutaminyl or asparaginyl residues [Tabb*, et al.* 2003], and this reaction resulted in the formation of cyclic imides (Fig. 34a) [Geiger and Clarke 1987].

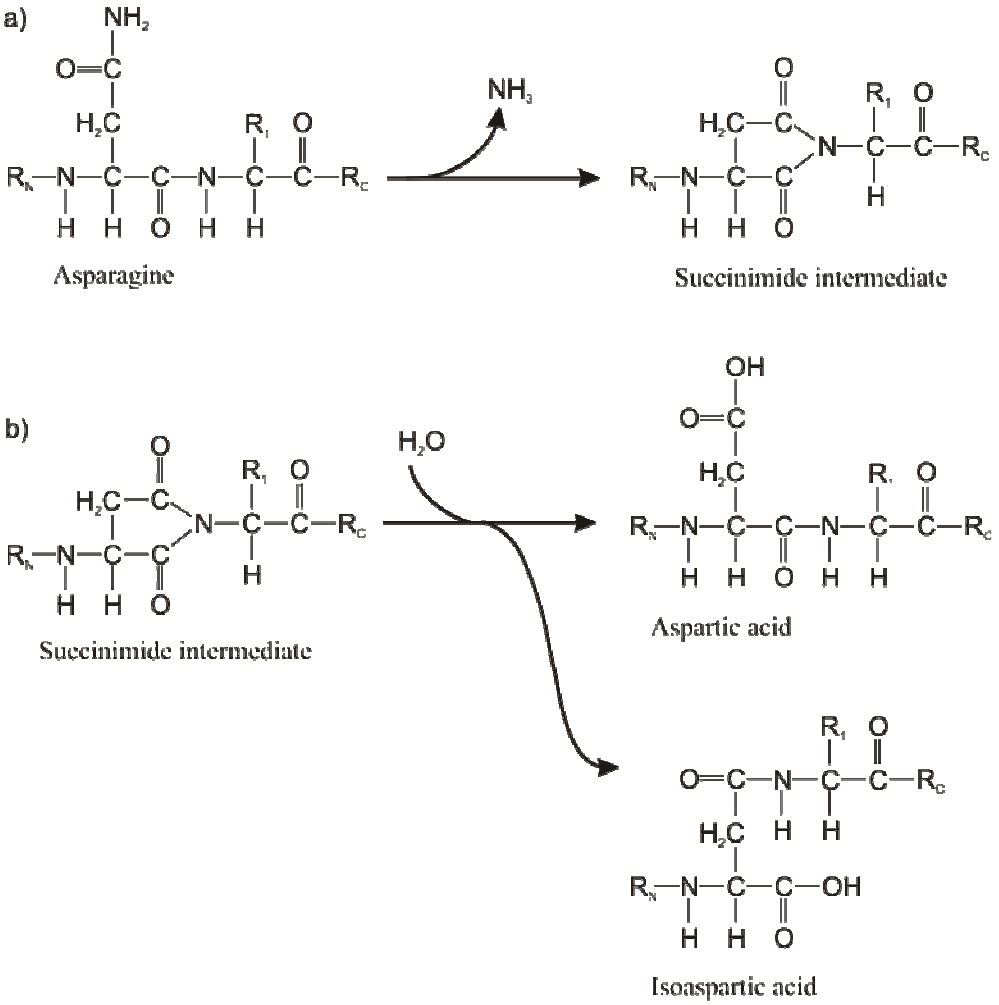

Figure 34. Deamidation reaction. a) Asparagine residues lose ammonia leading to the production of cyclic imides (or succinimide intermediates). b) Cyclic imides break open to yield aspartic or isoaspartic acid residues.

The cyclic imides or succinimide intermediates can be hydrolyzed to yield side chains with carboxylic groups (Fig. 34b). The complete reaction explains the deamidation of Gln/Asn to Glu/Asp [Geiger and Clarke 1987]. In accordance to Robinson 2002, deamidation took place extensively when the given asparagine residue was followed on its C-terminal by a glycyl residue (whereby R_1 in Fig. 34 would be a hydrogen atom). This tendency occurs because a glycine on the carboxylic side allows the second part of the mechanism (Fig. 34b) to proceed sterically unhindered.

Mass Spectrometry – a competitive method to individual gene sequencing

It is estimated that around 10 million different bacterial species are populating the earth. [Curtis, *et al.* 2002, DeLong and Pace 2001] Only around 30 000 bacteria are represented in the GenBANK, of which only 13 000 are formally characterized. Less

than 1000 bacterial genomes are known today, and the level of genomic knowledge is not better for eukaryotes. For the overwhelming majority of organisms whose genome is unknown, the mainstream approach of protein sequence determination has been an indirect strategy of seeking and sequencing the DNA of the gene that codes for the protein in question. The direct mass spectrometry strategy here outlined can be often faster and more straightforward than the indirect gene sequencing approach.

We have determined the entire amino acid sequence of the 21 kDa cytochrome c_4 from *T. roseopersicina*. To our knowledge, this is currently the largest protein that has been completely sequenced by mass spectrometry alone. We took advantage of the flexibility and high accuracy provided by the hybrid linear IT/FTICR instrument. Other mass spectrometers that provide similar features (high resolving power, high mass accuracy and alternative methods for ion activation/dissociation preferably in LC timescale) may also provide suitable platforms for *de novo* protein sequencing. The availability of ECD as a complement to CID was particularly important to obtain full sequence coverage. When facing the challenge of *de novo* sequencing of a protein, the ideal strategy would involve a preliminary step of amino acid composition determination, the determination of the exact protein mass and probably obtaining partial sequence information on the intact protein by a top down strategy. Armed with this knowledge, one can then select suitable proteolytic enzymes in order to maximize the coverage of the protein sequence by peptides with an amenable size to MS/MS (ideally up to ~1500 Da).

In order to obtain the full protein sequence, we have used less than 10 μg of protein for all the MS runs, but it is very likely that a substantially smaller amount of sample would suffice with procedure optimization. This sensitivity allows us to foresee the possibility of sequencing from a single protein spot in a two-dimensional isoelectric focusing sodium dodecyl sulfate (2D IEF-SDS) protein gel. The speed and ease of use afforded by online LC-MS and automated MS and MS/MS data acquisition, as well as automated MS/MS spectra interpretation, permitted the completion of this work in about two months. One of the major time limitations in the whole process was the *de novo* peptide sequencing algorithm that manifested ambiguities and therefore required manual sequence verification for specific peptides. With continuing development expected in the area of sequencing algorithms in the near future, one can envision that the determination of a protein's primary structure would be reliably achieved in even less time.

Phylogenetics of cytochrome c_4 *from* Thiocapsa roseopersicina

T. roseopersicina belongs to Chromatiaceae family, Chromatiales order, γ-proteobacteria class and therefore, as expected, the majority of the hits shown in Table 5 are cytochromes belonging to γ-proteobacterial organisms. Of these, most are cytochromes c_4; but there are also a few unspecified cytochromes *c*, two cytochromes c_{553} and one cytochrome c_{551}. However, even these cytochromes may also be, and probably are of the c_4 type. The reason for unspecified classification is that almost all proteins on the database are inferred from genome sequences, and thus no physico-chemical properties are known for them. From the sequence information thus it is only obvious that these are cytochromes *c*, and without knowledge of their redox potential, for example, inclusion into the c_4 group is uncertain. Furthermore, the versatility of the cytochrome theme means that sequence information alone is manifestly insufficient to give insight into a protein's functional role.

Some proteins of β-proteobacterial organisms also appeared; all of them unspecified cytochromes *c*, except one from *Thiobacillus denitrificans*, which is classified as a cytochrome *c* oxidase (cox). Since not a single β-proteobacterial cytochrome c_4 showed up, most likely the cytochromes c therein are not of the c_4 type. The high similarity between proteins of distantly related organisms is probably due to the widespread phenomenon of lateral gene transfer, which ensures the universality of highly versatile structures, such as cytochromes *c*.

There were only three hits from the Chromatiales order: *Alkalilimnicola ehrlichei*, *Nitrococcus mobilis*, and *Nitrosococcus oceani*. The first two are part of the Ectothiorhodospiraceae family and only *N. oceani* is from the Chromatiaceae family. There was however an abundance of hits in the Alteromonadales (16 hits), Vibrionales (10 hits), Pseudomonadales (8 hits) and Oceanospirillales (5 hits) orders. In addition, there were two species of Aeromonadales and one of the Legionellales order.

The explanation for this strange hit distribution lies in the skewed and uneven database (which is only a reflection of a biased orientation of research towards medically relevant species). Only four species of Chromatiales and seven of Oceanospirillales have had their genome sequenced in comparison to fifty species of Alteromonadales, sixty four of Vibrionales and seventy five of Pseudomonadales [*]. The

*NCBI taxonomy database

http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Root&id=1&lvl=3&lin=f&keep=1&srchmode=1&unlock

latter three orders have been favoured in research because they include many pathogenic species. This explains also the relatively low percentage of identity found (51% was the maximum value). It is to be expected that, as more species of Chromatiales become sequenced, higher identity percentages will be found.

Taking the unevenness of the database into consideration, 3 hits out of a total of 4 database presences for Chromatiales and 5 hits out of a total of 7 database presences for Oceanospirillales reveals the true affiliation of *T. roseopersicina's* cytochrome c_4. It seems that cytochromes c_4 have a fundamental presence across these two orders which live in neighbouring habitats. It may be noted that many species of these two orders live in seas and lakes: some Chromatiales like *T. roseopersicina* live in the anaerobic depths, whereas others such as *N. oceani* and Oceanospirillales species like *H. chejuensis* live in the upper aerobic layers [Holt, *et al.* 1994].

Temperature-dependent conformational transitions of cytochrome c_4

Under anaerobic conditions, cytochrome c_4 is a heat tolerant protein. The protein is operable at 60 ºC, as evidenced from high-temperature redox titration (Fig. 28b and Table 7). The temperature-dependent measurements were fully reversible when they were performed anaerobically (data not shown). There is a moderate but well-observable change in the secondary structure in both redox states of the protein (Fig. 33a and c). α helices are getting gradually shorter as the temperature increases, and this process appears to strain the Met–Fe bonds. Because these bonds are much weaker in the oxidized than in the reduced form, the oxidized form is more sensitive to temperature [Uchiyama*, et al.* 2004]. Consequently, if the cytochrome is in oxidized state (Fig. 32), disruption of the Met–Fe bond occurs at 32 ºC (which allows a larger secondary structure alteration to follow with a midpoint transition at 48 ºC, Fig. 33c); whereas in the reduced state the bond remains intact up to 80 ºC (the CD Vis spectrum of reduced cytochrome c_4 does not change under anaerobic conditions). Such extra stability of the reduced state has been reported previously for cytochrome c_{551} from *P. aeruginosa* [Uchiyama*, et al.* 2004]. Therefore, at sufficiently high temperature, oxidation triggers the detachment of the 6th axial ligand methionine from the iron, which then remains in the 5-coordinated state (Fig. 28b). His–Fe–Met to His–Fe-Empty

coordination switch upon oxidation/reduction of a heme has already been reported for a mono-heme cytochrome *c* from *S. putrefaciens* [Bartalesi, *et al.* 2002]. Here, the existence of the coordination switch at high temperature explains the ~200 mV drop in midpoint redox potential from the titration at 25ºC to the one at 60ºC (Fig. 28).

Under aerobic conditions, the thermal reversibility of cytochrome c_4 is not full: there is always some loss in absorption or CD signals (e.g. Fig. 30). This is consistent with the ‘‘low thermal reversibility’’ observed in cytochrome c_4 from *P. stutzeri* [Andersen, *et al.* 2002].

In the chemically oxidized sample, there is a large conformational change with a midpoint of ~38 ºC (Figs. 29, 30, 31b and 33d). This conformational change can also be seen in the DSC data but with a slight delay (42 ºC), probably because of the faster heating rate. The characteristic feature of this transition is a substantial loss of α helix and increase in the other conformations (Fig. 33d), as the Fe–Met bond breaks (Fig. 30). Since the loss of α helix is more dramatic than in the anaerobic case (and this transformation only occurs after the iron had become 5-coordinated), no other possibility remaining, we propose that the cause of the difference is the oxygen binding to the vacant sixth coordination site. One possible mechanism of secondary structure denaturation might proceed through the subsequent reaction of the coordinated oxygen molecule with the nearby methionine side chain leading to oxidation of the methionyl residue.

In the mostly reduced (as purified) sample, the first change to occur in the protein (35–52 ºC) is again a gradual decrease in length of α helices (from which the turn and unordered structures gain), while the number of α helices and both the number and length of the β structures remain constant (Fig. 33b). The DSC experiments also revealed a broad transition in this range, which could be deconvoluted into two different transitions at 44 and 50 ºC (Table 8). This gradual change can also be seen in the CD-Vis transition at 46 ºC (Fig. 32a). The CD signal of the chromophore is sensitive to the changes in the nearby aromatics in the methionine loop [Pielak, *et al.* 1986]. Because α helices are the structures involved in this transition and the methionine loop environment has been affected, we can assume therefore that the methionine which coordinates the iron is located in an α helix. This is in line with the structure of cytochrome c_4 from *P. stutzeri* (PDB# 1ETP [Kadziola and Larsen 1997]) and *A. ferrooxidans* (PDB# 1H1O [Abergel, *et al.* 2003]), where the first chromophore is

coordinated by Met66, which is located in an α helix, while the second heme is ligated by Met167, which is also a member of an α helix. After the gradual, slow transition, related to α helix shortening, there follows a sharper one, at 55–58 ºC (Fig. 33b). Here cytochrome c_4 gets rapidly oxidized and the Fe–Met bond breaks (Figs. 29 and 31a). This correlates with a further, fast conformational change, both the length and the number of α helices diminishing, while all other structures (including β sheets) increase. The transformation can also be seen on DSC, but, due to the higher heating rate, at higher temperatures (60–65 ºC). In contrast to the anaerobic case where the iron permanently becomes 5-coordinated and has high spin (Fig. 28b), under aerobic conditions this is only temporary (Fig. 30). The sixth coordination site is then presumably occupied by oxygen, as occurs in the case of oxidized cytochrome c_4.

The various protein conformations and heme configuration states of cytochrome c_4 are presented in Fig. 35. In all conformational changes discussed, we have considered only fully reduced or fully oxidized states of di-heme cytochrome c_4. The mixed valence state in which one of the hemes is reduced and the other oxidized was not noticeable in the experiments, probably because of the relatively close midpoint redox potential of the two hemes (237±5 mV and 268±6 mV). However, it is likely that oftentimes there was a minor population of this state present.

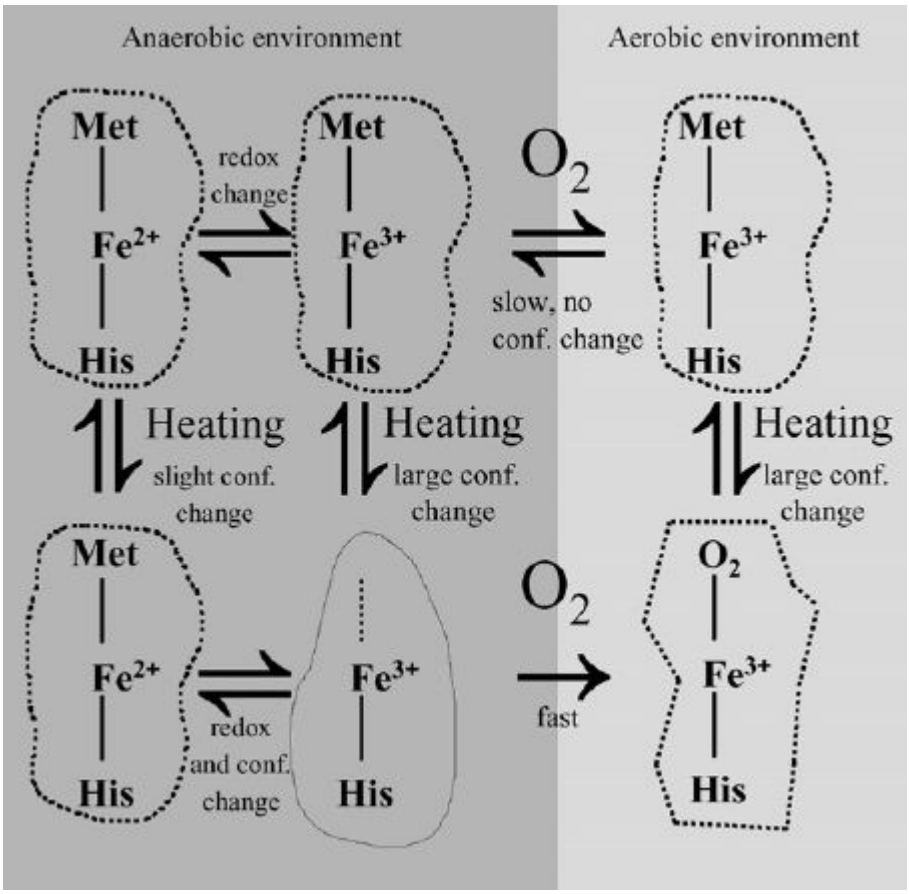

Figure 35. Proposed redox and temperature transition model for cytochrome c_4 from *Thiocapsa roseopersicina*. At room temperature under anaerobic conditions, both the oxidized and the reduced form of the protein have His–Fe–Met coordination. The secondary structure does not change during the redox transition. Upon heating, the secondary structure of the reduced form changes slightly, but the coordination of the iron atom is not involved. In order to attain a 5-coordinated high-temperature reduced form, the temperature should exceed 100 °C and high pressure should be applied [Uchiyama, *et al.* 2004]. In the oxidized form, the change in the secondary structure induces Fe–Met bond dissociation; the high-temperature oxidized form produced is 5-coordinated. During the redox transition at high temperature under anaerobic conditions, not merely the oxidation state, but also the coordination of the iron atom changes. When oxygen is also present (aerobic environment), it readily occupies the empty sixth iron coordination site, which causes additional secondary structure changes in the protein. This results in a partially unfolded protein matrix, with a 6-coordinated iron atom. The reversibility from this state is low.

Redox properties of cytochrome c_4

The measured redox potentials for cytochrome c_4 from *T. roseopersicina* at pH 7.7 (237 ± 5 and 268 ± 6 mV) are within the range of previously observed redox potentials for cytochromes c_4 (317 and 263 mV for *A. vinelandii*; 322 and 268 mV for *P. aeruginosa*; 300 and 190 mV for *P. stutzeri*, all at pH 7 [Leitch, *et al.* 1985]). These high potentials indicate that the two hemes have methionine as the sixth axial ligand. The Nernst value n = 1 (Table 7) indicates that the two hemes are independent.

During the redox transition at pH 7.7, at 60 ºC, the protein should undergo conformational changes (Fig. 33a and c - compare secondary structures at 60 ºC). The heme environment also changes extensively from 6-coordinated iron (reduced state) to 5-coordinated iron (oxidized state). It has been shown that either a change in the coordination of the iron or a change in the protein conformation (the heme side-chain packing) can itself cause a large redox potential shift [Raphael and Gray 1991]. It has also been shown that changing the ambient potential induces conformational changes in cytochrome *c* [Hildebrandt and Stockburger 1989]. The protein conformational and the heme environment differences themselves would explain why the midpoint potential at high temperature is lower by about 200 mV as compared with that at 25 ºC (Fig. 28).

Possible biological role

Despite the fact that *T. roseopersicina* grows predominantly under anaerobic conditions, and two proteins characterized from it (cytochrome c_4 and hydrogenase [Zorin and Gogotov 1982]) are able to operate at very high-temperatures under anaerobic conditions, *T. roseopersicina* does not survive high-temperature (it dies above 32 ºC).

Cytochrome c_4 has been previously suggested to perform electron transfer close to a terminal oxidase of the respiratory chain [Rey and Maier 1997, Bertini, *et al.* 2006]. Such a role has been suggested for cytochrome c_4 in *A. vinelandii* [Rey and Maier 1997], where it serves to augment respiration in low O_2 tension environments, especially when the bacterium must meet the heavy demands of the N_2-fixation metabolism. This is currently the only hypothesis for the role of cytochrome c_4, i.e. it is bound to cytochrome *c* oxidase in a membrane complex whose arrangement would resemble the organization of cbb_3 oxidases [Bertini, *et al.* 2006].

Thiocapsa roseopersicina is a purple sulfur photosynthetic bacterium. It oxidizes sulfur compounds (H_2S, elemental sulfur, thiosulphate etc.) during photosynthesis, which requires anaerobic conditions. Under darkness and aerobic conditions, however, *T. roseopersicina* can switch from photosynthesis to a dark chemolithoautotrophic way of life [Kondrat'eva, *et al.* 1975, Schaub and Van Gemerden 1994]. In this metabolic state, the bacterium, unable to perform photoreactions, is still able to assimilate CO_2 into organic compounds by using thiosulphate (as well as sulfide and elemental sulfur) as an energy source. Therefore, it seems that only the light-dependent metabolism requires anaerobicity. *T. roseopersicina* growing under anaerobic conditions synthesizes quite a large amount of cytochrome c_4 [Tomčová, *et al.* 2006]. Consequently, cytochrome c_4 probably exerts its function within the light dependent metabolism. The most natural place would be within the photosynthetic electron transfer chain. Having a redox potential of over 200 mV, cytochrome c_4 might be located in a position linking the b/c_1 complex to the bacteriochlorophyll "special pair" in the photosynthetic reaction center P870 [Voet and Voet 1995] (Fig. 4). Thus, cytochrome c_4, in conjunction with High Potential Iron-sulfur Protein (HiPIP, which has also been found in *Thiocapsa roseopersicina* [Ciszewska, *et al.* 1989]), is likely to fulfil the role of periplasmic electron carrier in some photosynthetic γ-proteobacteria (see also Table 1). Cytochromes c_5 and c_8 are also important in some species. It is tempting to speculate that this paraphernalia of periplasmic soluble electron carrier proteins serves as a link between the universal cyclic photophosphorylation process of photosynthetic purple bacteria and the metabolism that taps into the reducing power sources available in the environment specific to each species. Morevoer, this hypothesis provides an explanation for the di-heme nature of cytochrome c_4, in that the two hemes may serve as two distinct docking sites for distinct redox partners of distinct electron transport pathways.

As a summary, whereas earlier studies on cytochrome c_4 point to a role within oxygenic respiratory chains [Rey and Maier 1997, Bertini, *et al.* 2006], the discovery of a cytochrome c_4 in *T. roseopersicina* indicates that this type of cytochrome can also be part of an anaerobic electron transport chain, namely purple bacterial photophosphorylation.

Summary of novel findings

- We have purified a novel cytochrome from the organism *Thiocapsa roseopersicina*. This was the first purified and identified cytochrome c_4 from a photosynthetic bacterium.
- We have determined the primary structure of cytochrome c_4 from *T. roseopersicina*. This was the largest protein (about 21 kDa in size) whose amino acid sequence has been determined solely by Mass Spectrometry.
- Using recent developments in the field, we have implemented a protein sequencing methodology that allows high-resolution Mass Spectrometry to compete with indirect DNA-based methods.
- We have characterized cytochrome c_4 from *T. roseopersicina* in terms of UV-Vis and CD spectroscopy, correlating spectral profiles to structural features of the protein.
- We have determined the midpoint redox potential of the two heme groups of cytochrome c_4 from *T. roseopersicina*.
- We have discovered that cytochrome c_4 is heat tolerant and indeed shows a very resilient secondary structure (especially the reduced state) if under anaerobic conditions, which are the normal growth conditions for *T. roseopersicina*. If under aerobic conditions and in oxidized state, the secondary structure and thus the overall fold of the protein is irreversibly affected by heating in a process where oxygen binding to the vacant 6^{th} axial positions of the heme groups plays a decisive role.
- Our survey through the phylogenetics of cytochrome c_4 from *T. roseopersicina* has revealed that this type of cytochrome is widespread in γ-proteobacterial organisms, particularly in Oceanospirillales and Chromatiales orders.
- Up until now cytochromes c_4 were believed to participate in oxygenic electron transport chains. In contrast, the work here presented indicates that in purple sulphur bacteria (Chromatiales), cytochrome c_4 plays a role in anaerobic phototrophic electron transport chains instead.

Acknowledgements

I am extremely grateful to my supervisor, Csaba Bagyinka, for his permanent availability, for reinforcing my conviction that physical work is also extremely important (I literally had to cart bricks and lumber towards my PhD), and above all, for treating me as an equal peer in all scientific discussions and conversations.

I would like to thank my colleagues in the Biological Research Center, Judit Ősz, who assisted me greatly when I first came to Szeged, Rózsa Verebély, for her kind and ever present assistance, and Gabriella Bodó, for the solidarity through the long days of hard work in the laboratory.

Special thanks to Várkonyi Zsuzsanna and her family, for the limitless hospitality, and for the invaluable help in difficult situations.

Many thanks to László Prokai in Fort Worth, Texas, USA, for teaching me about the enourmosly powerful method that is Mass Spectrometry.

To my parents, who gave me continuous support from the far edge of Europe, and to my brother Fernando who daringly crossed the continent bridging Portugal to Hungary.

To Erika Bereczki, who helped me through the decisive moments of this laborious and arduous process, summed up in three letters: PhD.

I am grateful to the Institute of Biophysics of the Biological Research Center in Szeged for hosting me through these years of scientific endeavour.

I want to thank the financial support to our laboratory from the Hungarian Science Foundation (under the grants OTKA T049276 and OTKA T049207), and from AUTOESKORT Ltd.

Finally, I am greatly indebted the Portuguese Science and Technology Foundation for providing me personally with the PhD scholarship (under the programme POCTI, SFRH/BD/13128/2003). May the Foundation continue with their fundamental contribution to the advancement of Science in Portugal.

References

Abergel C, Nitschke W, Malarte G, Bruschi M, Claverie JM, Giudici-Orticoni M-T (2003). The structure of *Acidithiobacillus ferrooxidans* c_4 cytochrome: A model for complex-induced electron transfer tuning. *Structure*; **11**: 547-555

Ambler RP, Wynn M (1973). The amino acid sequences of cytochromes c_{551} from three species of pseudomonas. *Biochem J*; **131**: 485-498

Ambler RP (1991). Sequence variability in bacterial cytochromes c. *Biochim Biophys Acta*; **1058**: 42-47

Andersen NH, Norgaard A, Jensen TJ, Ulstrup J (2002). Sequential unfolding of the two domain protein *Pseudomonas stutzeri* cytochrome c_4. *J Inorg Biochem*; **88**: 316-327

Aubert C, Guerlesquin F, Bianco P, Leroy G, Tron P, Stetter KO, Bruschi M (2001). Cytochromes c555 from the hyperthermophilic bacterium *Aquifex aeolicus*. 2. Heterologous production of soluble cytochrome c555s and investigation of the role of methionine residues. *Biochemistry*; **40**: 13690-13698

Bagyinka C, Osz J, Szaraz S (2003). Autocatalytic oscillations in the early phase of the photoreduced methyl viologen-initiated fast kinetic reaction of hydrogenase. *J Biol Chem*; **278**: 20624-20627

Bartalesi I, Bertini I, Hajieva P, Rosato A, Vasos PR (2002). Solution structure of a monoheme ferrocytochrome c from *Shewanella putrefaciens* and structural analysis of sequence-similar proteins: Functional implications. *Biochemistry*; **41**: 5112-5119

Bartalesi I, Bertini I, Di Rocco G, Ranieri A, Rosato A, Vanarotti M, Vasos PR, Viezzoli MS (2004). Protein stability and mutations in the axial methionine loop of a minimal cytochrome c. *J Biol Inorg Chem*; **9**: 600-608

Bartsch RG (1991). The distribution of soluble metallo-redox proteins in purple phototrophic bacteria. *Biochim Biophys Acta*; **1058**: 28-30

Baymann F, Rappaport F (1998). Electrostatic interactions at the donor side of the photosynthetic reaction center of *Rhodopseudomonas viridis*. *Biochemistry*; **37**: 15320-15326

Benosman H, Asso M, Bertrand P, Yagi T, Gayda JP (1989). Epr study of the redox interactions in cytochrome c_3 from *Desulfovibrio vulgaris* miyazaki. *Eur J Biochem*; **182**: 51-55

Berry EA, Trumpower BL (1987). Simultaneous determination of hemes a, b, and c from pyridine hemochrome spectra. *Anal Biochem*; **161**: 1-15

Bertini I, Cavallaro G, Rosato A (2006). Cytochrome c: Occurence and functions. *Chem Rev*; **106**: 90-115

Biemann K, Martin SA (1987). Mass spectrometric determination of the amino acid sequence of peptides and proteins. *Mass Spectrom Rev*; **6**: 1-75

Biemann K (1988). Contributions of mass spectrometry to peptide and protein structure. *Biomed Environ Mass Spectrom*; **16**: 99-111

Biemann K (1990). Appendix 5. Nomenclature for peptide fragment ions (positive ions). *Methods Enzymol*; **193**: 886-887

Biemann K, Papayannopoulos I (1994). Amino acid sequencing of proteins. *Acc Chem Res*; **27**: 370-378

Böddi B, Ryberg M, Sundqvist C (1992). Identification of four universal protochlorophyllide forms in dark-grown leaves by analyses of the 77 k fluorescence emission spectra. *J Photochem Photobiol B*; **12**: 389-401

Böddi B, Ryberg M, Sundqvist C (1993). Analysis of the 77 k fluorescence emission and excitation spectra of isolated etioplast inner membranes. *J Photochem Photobiol B*; **21**: 125-133

Böddi B, Franck F (1997). Room temperature fluorescence spectra of protochlorophyllide and chlorophyllide forms in etiolated bean leaves. *J Photochem Photobiol B*; **41**: 73-82

Bogorov LV (1974). The properties of *Thiocapsa roseopersicina*, strain BBS, isolated from an estuary of the white sea. *Mikrobiologiia*; **43**: 326-332

Bradford MM (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem*; **72**: 248-254

Branca RMM, Bodó G, Várkonyi Z, Debreczeny M, Osz J, Bagyinka C (2007). Oxygen and temperature-dependent structural and redox changes in a novel cytochrome c_4 from the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. *Arch Biochem Biophys*; **467**: 174-184

Brown K, Nurizzo D, Besson S, Shepard W, Moura J, Moura I, Tegoni M, Cambillau C (1999). Mad structure of *Pseudomonas nautica* dimeric cytochrome c_{552} mimicks the c_4 dihemic cytochrome domain association. *J Mol Biol*; **289**: 1017-1028

Castillo MCG, Finnegan MG, Conover RC, Knaff DB, Johnson MK (1994). Spectroscopic characterization of flavocytochrome c_{552} from the photosynthetic purple sulfur bacterium *Chromatium vinosum*. *Biochim Biophys Acta*; **1184**: 273-278

Caughey WS, Smythe GA, O'Keeffe DH, Maskasky JE, Smith MI (1975). Heme a of cytochrome c oxicase. Structure and properties: Comparisons with hemes b, c, and s and derivatives. *J Biol Chem*; **250**: 7602-7622

Chamorovsky SK, Zakharova NI, Remennikov SM, Sabo Y, Rubin AB (1998). The cytochrome subunit structure in the photosynthetic reaction center of *Chromatium minutissimum*. *FEBS Lett*; **422**: 231-234

Chang CK (1985). On the structure of heme d_1. An isobacteriochlorin derivative as the prosthetic group of dissimilatory nitrite reductase? *J Biol Chem*; **260**: 9520-9522

Chen Z-W, Koh M, Van Driessche G, Van Beeumen J, Bartsch RG, Meyer TE, Cusanovich MA, Mathews FS (1994). The structure of flavocytochrome c sulfide dehydrogenase from a purple phototrophic bacterium. *Science*; **266**: 230-232

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*; **311**: 1283-1287

Ciszewska H, Bagyinka C, Tigyi G, Kovács KL (1989). Purification and properties of high potential iron-sulphur protein from *Thiocapsa roseopersicina*. *Acta Biochem Biophys Acad Sci Hung*; **24**: 361-375

Cohen AM, Mansour AAH, Banoub JH (2005). "De novo" sequencing sequencing of atlantic cod vitellogenin tryptic peptides by matrix assisted laser desorption-ionization quadrupole time-of-flight tandem mass spectrometry: Similarities with haddock vitellogenin. *Rapid Commun Mass Spectrom*; **19**: 2454-2460

Conrad LS, Karlsson JJ, Ulstrup J (1995). Electron transfer and spectral alpha-band properties of the di-heme protein cytochrome c_4 from *Pseudomonas stutzeri*. *Eur J Biochem*; **231**: 133-141

Curtis TP, Sloan WT, Scannell JW (2002). Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci U S A*; **99**: 10494-10499

Daltrop O, Ferguson SJ (2003). Cytochrome c maturation. The in vitro reactions of horse heart apocytochrome c and Paracoccus dentrificans apocytochrome c_{550} with heme. *J Biol Chem*; **278**: 4404-4409

Debreczeny M, Ball V, Boulmedais F, Szalontai B, Voegel JC, Schaaf P (2003). Multilayers built from two component polyanions and single component polycation solutions: A way to engineer films with desired secondary structure. *J Phys Chem B*; **107**: 12734-12739

Deisenhofer J, Michel H (1989). Nobel lecture. The photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridis*. *Embo J*; **8**: 2149-70

DeLong EF, Pace NR (2001). Environmental diversity of bacteria and archaea. *Syst Biol*; **50**: 470-478

Domon B, Aebersold R (2006). Review - mass spectrometry and protein analysis. *Science*; **312**: 212-217

Dutton PL (1978). Redox potentiometry: Determination of midpoint potentials of oxidation-reduction components of biological electron-transfer systems. *Methods Enzymol*; **54**: 411-435

Edman P (1950). *Acta Chem Scand*; **4**: 283

Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*; **246**: 64-71

Frank AM, Savitski MM, Nielsen ML, Zubarev RA, Pevzner PA (2007). De novo peptide sequencing and identification with precision mass spectrometry. *J Proteome Res*; **6**: 114-123

Gabellini N, Bowyer JR, Hurt E, Melandri BA, Hauska G (1982). A cytochrome b/c1 complex with ubiquinol-cytochrome c_2 oxidoreductase activity from *Rhodopseudomonas sphaeroides* ga. *Eur J Biochem*; **126**: 105-111

Garau G, Geremia S, Randaccio L (2002). Relationship between hydrogen-bonding network and reduction potential in c-type cytochromes. *FEBS Lett*; **516**: 285-286

Geiger T, Clarke S (1987). Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *J Biol Chem*; **262**: 785-794

Giudici-Orticoni M-T, Leroy G, Nitschke W, Bruschi M (2000). Characterization of a new dihemic c_4-type cytochrome isolated from *Thiobacillus ferrooxidans*. *Biochemistry*; **39**: 7205-7211

Gogotov IN, Zorin NA, Kondrat'eva EN (1976). Purification and properties of hydrogenase from phototrophic bacterium *Thiocapsa roseopersicina*. *Biokhimiia*; **41**: 836-842

Harrison AG, Young AB, Bleiholder C, Suhai S, Paizs B (2006). Scrambling of sequence information in collision-induced dissociation of peptides. *J Am Chem Soc*; **128**: 10364-10365

Hildebrandt P, Stockburger M (1989). Cytochrome c at charged interfaces. 1. Conformational and redox equilibria at the electrode/electrolyte interface probed by surface-enhanced resonance raman spectroscopy. *Biochemistry*; **28**: 6710-6721

Hochkoeppler A, Ciurli S, Kofod P, Venturoli G, Zannoni D (1997). On the role of cytochrome c_8 in photosynthetic electron transfer of the purple non-sulfur bacterium *Rhodoferax fermentans*. *Photosynth Res*; **53**: 13-21

Holt JG, Krieg NR, Sneath PHA, Staley JT, Williams ST (1994). *Bergey's manual of determinative bacteriology.* 9th ed. Williams & Wilkins, Baltimore, Maryland, USA.

Hopper S, Johnson RS, Vath JE and Biemann K (1989). Glutaredoxin from rabbit bone marrow. Purification, characterization, and amino acid sequence determined by tandem mass spectrometry. *J Biol Chem*; **264**: 20438-20447

Horn DM, Zubarev RA, McLafferty FW (2000). Automated de novo sequencing of proteins by tandem high-resolution mass spectrometry. *Proc Natl Acad Sci USA*; **97**: 10313-10317

Hunter DJB, Brown KR, Pettigrew GW (1989). The role of cytochrome c_4 in bacterial respiration. Cellular location and selective removal from membranes. *Biochem J*; **262**: 233-240

Inda LA, Medina M, Saraiva LM, Gómez-Moreno C, Teixeira M, Peleato ML (1997). Characterisation of cytochrome c_6 from *Chlorella fusca*. *Photosynth Res*; **54**: 107-114

Johnson RS, Biemann K (1987). The primary structure of thioredoxin from *Chromatium vinosum* determined by high-performance tandem mass spectrometry. *Biochemistry*; **26**: 1209-1214

Kadziola A, Larsen S (1997). Crystal structure of the dihaem cytochrome c_4 from *Pseudomonas stutzeri* determined at 2.2Å resolution. *Structure*; **5**: 203-216

Karas M, Hillenkamp F (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem*; **60**: 2299-2301

Kelemen L, Rizk S, Debreczeny M, Ogier J, Szalontai B (2004). Streptococcal antigen i/ii binds to extracellular proteins through intermolecular β-sheets. *FEBS Lett*; **566**: 190-194

Kendrew J (1994). *The encyclopedia of molecular biology. Protein sequencing.* 1st ed. Blackwell Science Ltd., Cambridge.

Kondrat'eva EN, Petushkova IP, Zhukov VG (1975). Growth and oxidation of sulfur compounds by *Thiocapsa roseopersicina* in darkness. *Mikrobiologiia*; **44**: 389-394

Korsunsky OF, Smolygina LD, Laurinavichene TV, Gogotov IN (1982). The low potential c-type cytochrome of *Thiocapsa roseopersicina*. *Biokhimiia*; **47**: 355-360

Korsunsky OF, Smolygina LD, Gogotov IN (1984). Purification and properties of membrane-bound cytochrome c from *Thiocapsa roseopersicina*. *Biokhimiia*; **49**: 409-414

Kota Z, Debreczeny M, Szalontai B (1999). Separable contributions of ordered and disordered lipid fatty acyl chain segments to nuch2 bands in model and biological membranes: A fourier transform infrared spectroscopic study. *Biospectroscopy*; **5**: 169-178

Leitch FA, Brown KR, Pettigrew GW (1985). Complexity in the redox titration of the dihaem cytochrome c4. *Biochim Biophys Acta*; **808**: 213-218

Lieutaud C, Nitschke W, Vermeglio A, Parot P, Schoepp-Cothenet B (2003). HIPIP in *Rubrivivax gelatinosus* is firmly associated to the membrane in a conformation efficient for electron transfer towards the photosynthetic reaction centre. *Biochim Biophys Acta*; **1557**: 83-90

Lieutaud C, Alric J, Bauzan M, Nitschke W, Schoepp-Cothenet B (2005). Study of the high-potential iron sulfur protein in *Halorhodospira halophila* confirms that it is distinct from cytochrome c as electron carrier. *Proc Natl Acad Sci U S A*; **102**: 3260-3265

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell JE (2000a). *Molecular cell biology. 23.8. Cell death and its regulation.* 4th ed. W. H. Freeman & Co., New York.

Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell JE (2000b). *Molecular cell biology. 3.5. Purifying, detecting, and characterizing proteins. 7.3. Identifying, analyzing, and sequencing cloned DNA.* 4th ed. W. H. Freeman & Co., New York.

Loo JA, Edmonds CG, Smith RD (1993). Tandem mass spectrometry of very large molecules. 2. Dissociation of multiply charged proline-containing proteins from electrospray ionization. *Anal Chem*; **65**: 425-438

Masuda S, Tsukatani Y, Kimura Y, Nagashima KV, Shimada K, Matsuura K (2002). Mutational analyses of the photosynthetic reaction center-bound triheme cytochrome subunit and cytochrome c_2 in the purple bacterium *Rhodovulum sulfidophilum*. *Biochemistry*; **41**: 11211-11217

Mathews WR, Johnson RS, Cornwell KL, Johnson TC, Buchanan BB, Biemann K (1987). Mass spectrometrically derived amino acid sequence of thioredoxin from

Chlorobium, an evolutionarily prominent photosynthetic bacterium. *J Biol Chem*; **262**: 7537-7545

Meyer TE, Bartsch RG, Kamen MD (1971). Cytochrome c_3. A class of electron transfer heme proteins found in both photosynthetic and sulfate-reducing bacteria. *Biochim Biophys Acta*; **245**: 453-464

Meyer TE, Bartsch RG, Caffrey MS, Cusanovich MA (1991). Redox potentials of flavocytochromes c from the phototrophic bacteria, *Chromatium vinosum* and *Chlorobium thiosulfatophilum*. *Arch Biochem Biophys*; **287**: 128-134

Meyer TE, Bartsch RG, Cusanovich MA, Tollin G (1993). Kinetics of photooxidation of soluble cytochromes, HIPIP, and azurin by the photosynthetic reaction center of the purple phototrophic bacterium *Rhodopseudomonas viridis*. *Biochemistry*; **32**: 4719-4726

Meyer TE, Cusanovich MA (2003). Discovery and characterization of electron transfer proteins in the photosynthetic bacteria. *Photosynth Res*; **76**: 111-126

Miles CS, Manson FD, Reid GA, Chapman SK (1993). Substitution of a haem-iron axial ligand in flavocytochrome b_2. *Biochim Biophys Acta*; **1202**: 82-86

Mitchell P (1961). Coupling of phosphorylation to electron and hydrogen transfer by a chemiosmotic type of mechanism. *Nature*; **191**: 144-148

Morand K, Talbo G, Mann M (1993). Oxidation of peptides during electrospray ionization. *Rapid Commun Mass Spectrom*; **7**: 738-743

Morris CJ, Black AC, Pealing SL, Manson FDC, Chapman SK, Reid GA, Gibson DM , Ward FB (1994). Purification and properties of a novel cytochrome – flavocytochrome c from *Shewanella putrefaciens*. *Biochem J*; **302**: 587-593

Moss GP (1988). Nomenclature of tetrapyrroles. Recommendations 1986 IUPAC-IUB joint commission on biochemical nomenclature. *Eur J Biochem*; **178**: 277-328

Nagashima KV, Matsuura K, Shimada K, Vermeglio A (2002). High-potential iron-sulfur protein (HIPIP) is the major electron donor to the reaction center complex in photosynthetically growing cells of the purple bacterium *Rubrivivax gelatinosus*. *Biochemistry*; **41**: 14028-14032

Nitschke W, Jubault-Bregler M, Rutherford AW (1993). The reaction center associated tetraheme cytochrome subunit from *Chromatium vinosum* revisited: A reexamination of its EPR properties. *Biochemistry*; **32**: 8871-8879

Oellerich S, Wackerbarth H, Hildebrandt P (2002). Spectroscopic characterization of nonnative conformational states of cytochrome c. *J Phys Chem B*; **106**: 6566-6580

Oellerich S, Lecomte S, Paternostre M, Heimburg T, Hildebrandt P (2004). Peripheral and integral binding of cytochrome c to phospholipids vesicles. *J Phys Chem B*; **108**: 3871-3878

Olsen JV, Mann M (2004). Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc Natl Acad Sci USA*; **101**: 13417-13422

Osyczka A, Nagashima KV, Sogabe S, Miki K, Yoshida M, Shimada K, Matsuura K (1998). Interaction site for soluble cytochromes on the tetraheme cytochrome subunit bound to the bacterial photosynthetic reaction center mapped by site-directed mutagenesis. *Biochemistry*; **37**: 11732-44

Osyczka A, Nagashima KV, Sogabe S, Miki K, Shimada K, Matsuura K (1999). Comparison of the binding sites for high-potential iron-sulfur protein and cytochrome c on the tetraheme cytochrome subunit bound to the bacterial photosynthetic reaction center. *Biochemistry*; **38**: 15779-15790

Osyczka A, Nagashima KV, Sogabe S, Miki K, Shimada K, Matsuura K (2001). Different mechanisms of the binding of soluble electron donors to the photosynthetic reaction center of *Rubrivivax gelatinosus* and *Blastochloris viridis*. *J Biol Chem*; **276**: 24108-24112

Page MD, Sambongi Y, Ferguson SJ (1998). Contrasting routes of c-type cytochrome assembly in mitochondria, chloroplasts and bacteria. *Trends Biochem Sci*; **23**: 103-108

Pettigrew GW, Brown KR (1988). Free and membrane-bound forms of bacterial cytochrome c_4. *Biochem J*; **252**: 427-435

Pevtsov S, Fedulova I, Mirzaei H, Buck C, Zhang X (2006). Performance evaluation of existing "de novo" sequencing algorithms. *J Proteome Res*; **5**: 3018-3028

Pielak GJ, Oikawa K, Mauk AG, Smith M, Kay CM (1986). Elimination of the negative soret cotton effect of cytochrome c by replacement of the invariant phenylalanine using site-directed mutagenesis. *J Am Chem Soc*; **108**: 2724-2727

Ponamarev MV, Schlarb BG, Howe CJ, Carrell CJ, Smith JL, Bendall DS, Cramer WA (2000). Tryptophan-heme pi-electrostatic interactions in cytochrome f of oxygenic photosynthesis. *Biochemistry*; **39**: 5971-5976

Prince RC, George GN (1995). Cytochrome f revealed. *Trends Biochem Sci*; **20**: 217-218

Rafferty SP, Pearce LL, Barker PD, Guillemette JG, Kay CM, Smith M, Mauk AG (1990). Electrochemical, kinetic, and circular dichroic consequences of mutations at position 82 of yeast iso-1-cytochrome c. *Biochemistry*; **29**: 9365-9369

Raphael AL, Gray HB (1991). Semisynthesis of axial-ligand (position 80) mutants of cytochrome c. *J Am Chem Soc*; **113**: 1038-1040

Reinartz M, Tschape J, Bruser T, Truper HG, Dahl C (1998). Sulfide oxidation in the phototrophic sulfur bacterium *Chromatium vinosum*. *Arch Microbiol*; **170**: 59-68

Rey L, Maier RJ (1997). Cytochrome c terminal oxidase pathways of *Azotobacter vinelandii*: Analysis of cytochrome c_4 and c_5 mutants and up-regulation of cytochrome c-dependent pathways with N_2 fixation. *J Bacteriol*; **179**: 7191-7196

Robinson NE (2002). Protein deamidation. *Proc Natl Acad Sci USA*; **99**: 5283

Roepstorff P, Fohlman J (1984). Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom*; **11**: 601

Samyn B, De Smet L, Van Driessche G, Meyer TE, Bartsch RG, Cusanovich MA, Van Beeumen JJ (1996). A high-potential soluble cytochrome c_{551} from the purple phototrophic bacterium *Chromatium vinosum* is homologous to cytochrome c_8 from denitrifying pseudomonads. *Eur J Biochem*; **236**: 689-696

Santucci R, Ascoli F (1997). The soret circular dichroism spectrum as a probe for the heme Fe(III)-met(80) axial bond in horse cytochrome c. *J Inorg Biochem*; **68**: 211-214

Saraiva LM, Liu MY, Payne WJ, LeGall J, Moura JJG, Moura I (1990). Spin-equilibrium and heme-ligand alteration in a high-potential monoheme cytochrome (cytochrome c_{554}) from *Achromobacter cycloclastes*, a denitrifying organism. *Eur J Biochem*; **189**: 333-341

Saraiva LM, Besson S, Moura I, Fauque G (1995). Purification and preliminary characterization of three c-type cytochromes from *Pseudomonas nautica* strain 617. *Biochem Biophys Res Commun*; **212**: 1088-1097

Savitski MM, Nielsen ML, Zubarev RA (2007). Side-chain losses in electron capture dissociation to improve peptide identification. *Anal Chem*; **79**: 2296-2302

Schaub BEM, Van Gemerden H (1994). Simultaneous phototrophic and chemotrophic growth in the purple sulfur bacterium *Thiocapsa roseopersicina*. *FEMS Microbiol Ecol*; **13**: 185-196

Schoepp B, Parot P, Menin L, Gaillard J, Richaud P, Vermeglio A (1995). In vivo participation of a high potential iron-sulfur protein as electron donor to the

photochemical reaction center of *Rubrivivax gelatinosus*. *Biochemistry*; **34**: 11736-11742

Shrager I (1986). Chemical transitions measured by spectra and resolved using singular value decomposition. *Chemom Intell Lab Syst*; **1**: 59-70

Sorensen SB, Sorensen SB, Breddam K (1991). *FEBS*; **294**: 195-197

Sreerama N, Venyaminov SY, Woody RW (2000). Estimation of protein secondary structure from circular dichroism spectra: Inclusion of denatured proteins with native proteins in the analysis. *AnalBiochem*; **287**: 243-251

Sreerama N, Woody RW (2000). Estimation of protein secondary structure from circular dichroism spectra: Comparison of contin, selcon, and cdsstr methods with an expanded reference set. *Anal Biochem*; **287**: 252-260

Sreerama N, Woody RW (2004). Computation and analysis of protein circular dichroism spectra. *Methods Enzymol*; **383**: 318-351

Standing KG (2003). Peptide and protein de novo sequencing by mass spectrometry. *Curr Opin Struct Biol*; **13**: 595-601

Tabb DL, Smith LL, Breci LA, Wysocki VH, Lin D, Yates III JR (2003). Statistical characterization of ion trap tandem mass spectra from doubly charged tryptic peptides. *Anal Chem*; **75**: 1155-1163

Timkovich R, Cork MS, Gennis RB, Johnson PY (1985). Proposed structure of heme d, a prosthetic group of bacterial terminal oxidases. *J Am Chem Soc*; **107**: 6069-6075

Tissieres A (1956). Purification, some properties and the specific biological activity of cytochromes c_4 and c_5 from *Azotobacter vinelandii*. *Biochem J*; **64**: 582-589

Tomčová I, Branca RMM, Bodó G, Bagyinka C, Smatanová IK (2006). Cross-crystallizationmethod used for the crystallization and preliminary diffraction analysis of a novel di-haem cytochrome c_4. *Acta Cryst*; **F62**: 820-824

Trüper HG, Rogers LA (1971). Purification and properties of adenylyl sulfate reductase from the phototrophic sulfur bacterium, *Thiocapsa roseopersicina*. *J Bacteriol*; **108**: 1112-1121

Tsapin AI, Nealson KH, Meyers T, Cusanovich MA, Van Beuumen J, Crosby LD, Feinberg BA, Zhang C (1996). Purification and properties of a low-redox-potential tetraheme cytochrome c_3 from *Shewanella putrefaciens*. *J Bacteriol*; **178**: 6386-6388

Uchiyama S, Ohshima A, Nakamura S, Hasegawa J, Terui N (2004). Complete thermal-unfolding profiles of oxidized and reduced cytochromes c. *J Am Chem Soc*; **126**: 14684-14685

Van Beeumen JJ (1991). Primary structure diversity of prokaryotic diheme cytochromes c. *Biochim Biophys Acta*; **1058**: 56-60

Van Driessche G, Vandenberghe I, Devreese B, Samyn B, Meyer TE, Leigh R, Cusanovich MA, Bartsch RG, Fischer U, Van Beeumen JJ (2003). Amino acid sequences and distribution of high-potential iron-sulfur proteins that donate electrons to the photosynthetic reaction center in phototropic proteobacteria. *J Mol Evol*; **57**: 181-199

Venturoli G, Drepper F, Williams JC, Allen JP, Lin X, Mathis P (1998). Effects of temperature and deltago on electron transfer from cytochrome c_2 to the photosynthetic reaction center of the purple bacterium *Rhodobacter sphaeroides*. *Biophys J*; **74**: 3226-3240

Voet D, Voet JG (1995). *Biochemistry. Chapter 22. Photosynthesis.* 2nd ed. John Wiley and sons, New York.

Whiteaker JR, Warscheid B, Pribil P, Hathout Y, Fenselau C (2004). Complete sequences of small acid-soluble proteins from *Bacillus globigii*. *J Mass Spectrom*; **39**: 1113-1121

Wilson DF (1978). Determination of oxidation-reduction potentials. *Methods Enzymol*; **54**: 396-410

Wood PM (1983). Why do c-type cytochromes exist? *FEBS Lett*; **164**: 223-6

Zhang B, Krutchinsky AN, Chait BT (2003). "De novo" peptide sequencing by MALDI-quadrupole-ion trap mass spectrometry: A preliminary study. *J Am Soc Mass Spectrom*; **14**: 1012-1021

Zorin NA, Gogotov IN (1980). Purification and properties of cytochrome c_{552} from purple sulfur bacterium *Thiocapsa roseopersicina*. *Biokhimiia*; **45**: 1497-1502

Zorin NA, Gogotov IN (1982). Stability of hydrogenase from the purple sulfur bacteria *Thiocapsa roseopersicina*. *Biokhimiia*; **47**: 827-833

Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW (2000). Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem*; **72**: 563-573

Summary of the Thesis (English)

Introduction

In the present work, we discuss the discovery and study of a novel cytochrome *c* from the purple sulfur bacterium *Thiocapsa roseopersicina*. We have characterized the cytochrome through a combination of absorption spectroscopy and circular dichroism (CD) with redox potentiometry and also by using differential scanning calorimetry (DSC). The data gathered place this protein in the c_4 class of cytochromes. We have used Mass Spectrometry to determine its primary structure, which confirmed this cytochrome to be of the c_4 type. It is the first purified and identified cytochrome c_4 from an anaerobic phototrophic bacterium.

Aims of the study

To purify and characterize proteins with redox centers from purple sulfur photosynthetic bacteria in order to shed some light over the electron transport pathways within the metabolism of these organisms.

Particularly, to characterize a novel periplasmic soluble cytochrome *c* from *Thiocapsa roseopersicina*:

- To determine its primary structure, by using the recent advances in Mass Spectrometry.
- To determine the heme content of the cytochrome, and the respective midpoint potential of each redox site.
- To determine its sensitiveness to oxygen, given that the organism is photosynthetically active only under anaerobic conditions, but can survive in aerobic conditions; and to determine its thermal stability, given that temperature is another sensitive parameter to the survival of the organism. In sum, to elucidate the interrelation of these two parameters in their influence on the cytochrome structure.
- To elucidate the phylogenetics of this protein.

Finally, to use all of the gathered structural information in order to clarify the function of this cytochrome and on a broader sense to gain insight into the electron transport chains of phototrophic bacteria.

Materials and Methods

The protein components from the cells of *Thiocapsa roseopersicina* strain BBS were extracted in a procedure that employed cold acetone. A first round of anion exchange batch liquid chromatography was followed by four rounds of alternating hydrophobic and anion exchange fast protein liquid chromatographies (FPLC). The purity and molecular weight were determined by SDS–PAGE.

The pyridine hemochromogen method was used, in combination with the Bradford method, to determine the heme content and the exact protein concentration.

In preparation for mass spectrometry, four different digest sets were obtained by digestion of the protein with combinations of three different proteolytic enzymes. All measurements were performed on a linear IT-FTICR hybrid instrument. In most measurements, the proteolytic fragments of each digestion set were separated by online gradient reversed-phase micro-Liquid Chromatography prior to NSI, full-scan mass spectra were recorded in the FTICR unit, and MS/MS product ion scans in the IT. Most peptide sequences were determined by interpretation of the MS/MS spectra with the aid of a computer algorithm.

The redox potential of the cytochrome was determined by potentiometric titration using a home-made stirred spectroelectrochemical cell which included a three electrode system. The enforced solution potential between the working and the platinum electrodes was provided by a potentiostat.

CD spectra were measured both in the far-UV range (190–250 nm) and near-UV–visible range (250–700 nm) as a function of the temperature in both anaerobic and aerobic conditions. CD spectra in the range 190–250 nm were used for calculation of secondary structures.

UV-VIS absorbance measurements and DSC measurements were performed to provide additional information in temperature dependence studies.

Results and Discussion

Cytochrome c_4 contains two hemes per molecule, as determined by the pyridine hemochromogen method, and later confirmed by Mass Spectrometry.

After analysis by MS of the peptide mixtures resulting from the proteolytic digestions, we have compiled the most abundant ions into comprehensive lists of peptide masses. Each peptide was then individually sequenced by interpretation of the respective MS/MS data. It was fairly straightforward to determine the sequence of peptides up to 14 amino acids in length (which translates to roughly 1500 Da in peptide mass). Larger peptides posed mounting difficulties. Linking up yielded the complete amino acid sequence of cytochrome c_4 from *Thiocapsa roseopersicina*. To our knowledge, this is currently the largest protein that has been completely sequenced by mass spectrometry alone.

```
TDGHQAAAPQ VGDPQAGEAK ANGVCLACHG PQGNSLVPLW PKLAGQHPEY
IVKQLMDFKQ RRANEQMTPM AMPLTDQEVL DLAAYYATQP KTPGAADPEL
ASKGESLYRW GNPETGVPAC SGCHGPAGGA GQSLAKFPRL SAQHADYTKQ
TLEHFRGALR ANDPNGMMRG AAARLSDQEL AAVSQYLQGL SQ
```

The direct mass spectrometry strategy used in this work can be often faster and more straightforward than the indirect gene sequencing approach.

Cytochrome c_4 proved to be a heat-tolerant protein if maintained under anaerobic conditions. We clarified that oxygen initiates an irreversible unfolding of the protein at high temperatures, most likely through direct binding to the heme's sixth coordination site, which was left vacant after dissociation of the native methionine ligand. It was shown that the methionine-Fe bond (i. e. the 6^{th} axial ligation) is closely linked to the protein moiety (i.e. the protein's secondary structure) and plays a crucial role in the overall folding of the protein.

Cytochrome c_4 proteins are thought to participate in aerobic respiratory pathways, in a position close to the terminal oxidase of the electron transport chain. The discovery of such a cytochrome in an anaerobic photosynthetic organism throws doubt upon this assumption. In *T. roseopersicina*, cytochrome c_4 must participate in photosynthesis instead. More specifically, since the redox titration revealed midpoint potentials of 237 ± 5 mV (1^{st} heme) and 268 ± 6 mV (2^{nd} heme), we suggest to place this cytochrome in the position of electron transport between cytochrome b/c_1 and the tetraheme cytochrome of the reaction centre.

Our phylogenetic survey revealed that cytochrome c_4 is widespread across γ-proteobacteria, and thus it is likely to have this role of periplasmic electron carrier in several species of photosynthetic bacteria.

In purple non-sulfur bacteria, monoheme cytochrome c_2 usually carries out this function, hence it is appealing to speculate that purple sulphur bacteria use instead a diheme cytochrome c_4 in order to link sulphur metabolism with the photophosphorylation cycle.

Summary of the Thesis (Hungarian)

Bevezetés

A dolgozatban egy citokrom c_4 felfedezéséről és tulajdonságairól számolunk be, melyet a *Thiocapsa roseopersicina* bíbor kénbaktériumból tisztítottunk. A citokróm c_4 fehérjét kombinált redox és abszorpciós, illetve CD spektroszkópiai módszerrel karakterizáltuk, s felhasználtunk DSC méréseket is. A mérési adatok alapján a fehérjét a c_4 citokrómok családjába sorolhatjuk. Kifejlesztettünk egy új tömegspektroszkópiai módszert, mellyel meghatároztuk a fehérje szekvenciáját. A szekvencia alátámasztotta, hogy valóban egy citokróm c_4 fehérjét izoláltunk. Ez az első tisztított és azonosított citokróm c_4 fehérje, melyet anaerob fototróf baktériumból izoláltak.

Célkitűzések

Elsődleges célunk az volt, hogy redox fehérjéket tisztítsunk a *Thiocapsa roseopersicina* fotoszintetizáló bíbor kénbaktériumból., hogy megértsük az elektron transzport lánc szerkezetét és működését.

Ezen belül figyelmünket egy citokróm fehérjére fordítottuk, s meghatároztuk

- az elsődleges szerkezetét a legújabb tömegspektrometriai módszerekkel,
- a fehérje hem tartalmát, s az egyes redox centrumok redox potenciálját,
- oxigén érzékenységét, ami annak fényében különösen érdekes, hogy az organizmus csak anaerob körülmények között fotoszintetizál,
- érzékenységét a magas hőmérsékletre, mivel az organizmus magas hőmérsékleten életképtelen,
- az oxigén és a hőmérséklet hatását a fehérje szerkezetére,
- valamint megvizsgáltuk a fehérje filogenetikai kapcsolatait is.

Végül az összes fent említett információ birtokában következtetéseket vontunk le a citokróm funkciójáról, s elhelyeztük azt a fototróf baktériumok elektron transzport láncában is.

Anyagok és Módszerek

A fehérjéket a *Thiocapsa roseopersicina* sejtekből acetonos kezeléssel vontuk ki. A keletkezett fehérje port 12 órán keresztül desztillált vízben kevertettük, hogy az oldhtaó fehérjéket megkapjuk. A tisztítás első lépése egy anioncserélő batch kromatográfia, melyet hidrofób kromatográfia és két újabb anioncserélő kromatográfia követett. A kromatografálásokat egy FPLC készülék segítségével végeztük. A fehérje frakciók tisztaságát és az egyes fehérjék molekulasúlyat SDS-PAGE segítségével határoztuk meg.

A citokróm hem tartalmát a piridin-hemokromogén módszerrel, a fehérje koncentrációkat pedig a Bradford módszerrel határoztuk meg.

A tömegspektrometriai mérésekhez használt emésztési stratégia során a fehérjét három különböző proteolitikus enzim kombinációjával hasítottuk, mely négy különböző emésztett fragmentumot eredményezett. Minden mérést az IT-FTICR hibrid készüléken végeztünk. A legtöbb mérés esetén az emésztéseket a készülékhez kapcsolt reverz-fázisú micro-folyadék kromatográffal választottuk el. A frakciók ezután kerültek a tömegspektrográfba. A teljes tömegspektrumokat az FTICR-ban mértük meg, a kiválasztott ionokat (MS/MS) pedig az IT-ban analizáltuk. A peptid szekvenciák többségét az MS/MS spektrumokból állapítottuk meg a műszerhez tartozó szoftver segítségével.

A citokróm redox potenciálját egy házi készítésű spektro-elektrokémiai cellában mértük meg. A kívánt redox potenciál értéket egy potenciosztát segítségével állítottuk be.

A CD spektrumokat a távoli UV (190-250 nm) és a közeli UV – látható (250-700 nm) tartományban vettük fel. A hőmérsékletfüggő méréseket anaerob és aerob körülmények között végeztünk.

Hőmérsékletfüggő abszorpciós (UV-látható) és DSC mérések szolgáltattak további információkat.

Eredmények és megbeszélésük

A citokróm c_4 két hemet tartalmaz molekulánként, amit először a piridin hemokromogén módszerrel állapítottunk meg, s amit a tömegspektroszkóppal kapott szekvencia adatok alátámasztottak.

A proteolitikus emésztésből származó peptidekeveréket tömegspektrográffal analizáltuk. A leggyakrabban előforduló ionokat táblázatba foglaltuk. Minden egyes peptidet ezek után MS/MS analízissel szekvenáltunk. A maximum 14 aminósavat tartalmazó peptideket (kb. 1500 Da tömegig) viszonylag egyszerű volt szekvenálni. Az ennél hosszabb peptidek szekvenálása jóval nehezebb feladatnak bizonyult. Az összes peptidszekvencia összekapcsolása eredményezte a *Thiocapsa roseopersicina* citokróm c_4 szekvenciáját. Legjobb tudomásunk szerint pillanatnyilag ez a leghosszabb fehérje amit kizárolag tömegspetroszkópiai módszerekkel szekvenáltak.

```
TDGHQAAAPQ VGDPQAGEAK ANGVCLACHG PQGNSLVPLW PKLAGQHPEY
IVKQLMDFKQ RRANEQMTPM AMPLTDQEVL DLAAYYATQP KTPGAADPEL
ASKGESLYRW GNPETGVPAC SGCHGPAGGA GQSLAKFPRL SAQHADYTKQ
TLEHFRGALR ANDPNGMMRG AAARLSDQEL AAVSQYLQGL SQ
```

Az általunk kidolgozott tömegspektroszkópiai módszer gyakran gyorsabb és egyszerübb mint a fehérje szekvencia meghatározása géntechnológiai módszerrel.

A citokróm c_4 hőtűrő fehérjének bizonyult, amennyiben anaerob körülmények között tartjuk. Bebizonyítottuk, hogy magas hőmérsékleten az oxigén irreverzibilis konformációs változást okoz , valószínüleg bekötődik a hem hatodik ligand pozíciójába, mely magas hőmérsékleten felszabadul a metionin kötésből. Kimutattuk, hogy a metionin-Fe kötésnek meghatározó szerepe van a fehérje szerkezet fenntartásában valamint kulcsszerepet játszik a fehérje denaturálódásában.

A citokróm c_4–nek eddig a légzési elektron transzport láncban volt feltételezett szerepe, közel az elektron transzport lánc terminális oxidáló végéhez. Egy ilyen citokróm felfedezése egy anaerob fotoszintetizáló organizmusban kétségessé teszi ezt a feltevést. A *Thiocapsa roseopersicinában* a citokróm c_4 a fotoszintetikus elektron transzport láncban kell helyet foglaljon. Mivel a redox potenciálja 237 ± 5 mV, illetve 268 ± 6 mV a két hem esetében, úgy gondoljuk, hogy a citokróm c_4 a citokróm b/c_1 és a reakciócentrum tetrahem citokrómja között foglal helyet.

Filogenetikai vizsgálódásaink azt mutatták, hogy a citokróm c_4 igen elterjedt a γ-proteobaktériumok között. Ez valószínűsíti, hogy periplazmikus elektron hordozó szerepe általános lehet.

A bíbor nem-kén baktériumok esetében ezt a szerepet a citokróm c_2 játssza. Kézenfekvő feltételezni, hogy a bíbor kénbaktériumok esetében a dihem citokróm c_4 köti össze a kén metabolizmust a fotofoszforilációs ciklussal.

Tézispontok

- Kitisztítottunk egy új citokrómot a *Thiocapsa roseopersicina*ból. Ez az első fotoszintetikus baktériumból tisztított és azonosított citokróm c_4.
- Meghatároztuk a citokróm c_4 szekvenciáját. A *Thiocapsa roseopersicina* citokróm c_4 a legnagyobb fehérje melynek szekvenciáját kizárólag tömegspektroszkópiai módszerekkel meghatározták.
- Kidolgoztunk egy új, tömegspektroszkópot felhasználó szekvenálási módszert, mely versenyképes az indirekt, DNS alapú szekvenálási módszerrel.
- Meghatároztuk a citokróm c_4 CD és abszorpciós spektroszkópiai tulajdonságait, s korreláltuk azokat a fehérje struktúrális tulajdonságaival.
- Meghatároztuk a *T. roseopersicina* citokróm c_4 redox potenciál értékét.
- Felfedeztük, hogy a citokróm c_4 hőtűrő fehérje, s redukált állapotban, oxigénmentes környezetben igen stabil szerkezettel rendelkezik. Oxigén jelenlétében, oxidált állapotban melegítés hatására irreverzibilis konformációs változások zajlanak le, melynek során az oxigén a hem csoport felszabaduló hatodik ligand pozíciójába köt be, s ezzel inicializálja a konformációs változásokat.
- Filogenetikai vizsgálódásaink során megállapítottuk, hogy ez a fajta citokróm igen elterjedt a γ-proteobaktériumok körében, különösen az Oceanospirillales és a Chromatiales rendekben.
- A citokróm c_4-et eddig a légzési elektrontranszport láncban helyezték el. Ezzel szemben mi kimutattuk, hogy a bíbor kénbaktériumokban a citokróm c_4 az anaerob fotoszintetikus elektrontranszport láncban foglal helyet.

Filename: Rui Branca - PhD Thesis (extra figure version).doc
Directory: D:\Os meus documentos\PhD Thesis
Template: C:\Documents and Settings\Rui\Application Data\Microsoft\Templates\Normal.dot
Title: Structure and function of a novel cytochrome c4
Subject:
Author: Rui Branca
Keywords:
Comments:
Creation Date: 18-06-2008 14:54:00
Change Number: 100
Last Saved On: 24-06-2008 15:42:00
Last Saved By: Rui Branca
Total Editing Time: 577 Minutes
Last Printed On: 24-06-2008 15:42:00
As of Last Complete Printing
Number of Pages: 105
Number of Words: 30,945 (approx.)
Number of Characters: 167,108 (approx.)