Crystal Structures of Mitochondrial Processing Peptidase Reveal the Mode for Specific Cleavage of Import Signal Sequences

Alexander B. Taylor,1,2 Barbara S. Smith,1,2 Sakae Kitada,1 Katsuhiko Kojima,3 Hideki Miyaura,3 Zbyszek Otwinowski,2 Akio Ito,3 and Johann Deisenhofer1,2,4
1Howard Hughes Medical Institute
2Department of Biochemistry
University of Texas Southwestern Medical Center
Dallas, Texas 75390
3Department of Chemistry
Kyushu University
Fukuoka 812-8581
Japan

Summary

Background: Mitochondrial processing peptidase (MPP) is a metalloendopeptidase that cleaves the N-terminal signal sequences of nuclear-encoded proteins targeted for transport from the cytosol to the mitochondria. Mitochondrial signal sequences vary in length and sequence, but each is cleaved at a single specific site by MPP. The cleavage sites typically contain an arginine at position –2 (in the N-terminal portion) from the scissile peptide bond in addition to other distal basic residues, and an aromatic residue at position +1. Mitochondrial import machinery recognizes amphiphilic helical conformations in signal sequences. However, it is unclear how MPP specifically recognizes diverse presequence substrates.

Results: The crystal structures of recombinant yeast MPP and a cleavage-deficient mutant of MPP complexed with synthetic signal peptides have been determined. MPP is a heterodimer; its α and β subunits are homologous to the core II and core I proteins, respectively, of the ubiquinol-cytochrome c oxidoreductase complex. Crystal structures of two different synthetic substrate peptides cocrystallized with the mutant MPP each show the peptide bound in an extended conformation at the active site. Recognition sites for the arginine at position –2 and the +1 aromatic residue are observed.

Conclusions: MPP bound two mitochondrial import presequence peptides in extended conformations in a large polar cavity. The conformational changes differ from the amphiphilic helical conformation recognized by mitochondrial import components. Our findings suggest that the presequences adopt context-dependent conformations through mitochondrial import and processing, helical for recognition by mitochondrial import machinery and extended for cleavage by the main processing component. 

Introduction

Most mitochondrial proteins encoded in the nucleus are synthesized on cytoplasmic ribosomes with N-terminal extension peptides, which target them for transport into the mitochondria [1–4]. Matrix-targeted proteins are transferred to mitochondria by cytosolic factors, recognized by surface receptors on the mitochondrial outer membrane, and translocated through the outer membrane translocase complex (TOM complex) to the intermembrane space. Proteins destined for the matrix are then recognized by surface receptors on the mitochondrial inner membrane and translocated through the inner membrane translocase complex (TIM complex) [5]. The extension peptides of the mitochondrial precursor proteins are proteolytically removed during or subsequent to import into the mitochondria by three types of processing peptidases [1–4]. Mitochondrial processing peptidase (MPP; EC 3.4.24.64) [6, 7] cleaves off a large portion of the N-terminal presequence from precursor proteins, including the sequence that targets the proteins to the mitochondrial matrix. Some proteins require additional proteolytic processing.

MPP is a heterodimer with a mass of ~100 kDa composed of α and β subunits of roughly the same size. In accordance with Schechter and Berger [62], the enzyme binding sites are denoted S1, S2, ..., S, and S1′, S2′, ..., S′, away from the scissile peptide bond toward the N- and C-termini, respectively. Amino acid residues in the substrates are referred to as P1, P2, ..., P, and P1′, P2′, ..., P′, in accordance with the binding site. The β subunit contains an HxxEH zinc binding motif, which is an inversion of the thermolysin zinc binding motif, HExxH [8]. Although the α and β subunits of MPP have very similar sequences (i.e., 48% identical residues or conservative replacements between the Saccharomyces cerevisiae MPP subunits), only the β subunit contains the zinc binding motif. The α and β subunits of MPP are homologous to the core II and core I proteins, respectively, of mitochondrial ubiquinol-cytochrome c oxidoreductase [9–11], also known as the cytochrome bc, complex (BC1) or complex III of the respiratory chain. The crystal structure of BC1 revealed that the core proteins are located in the matrix, and are attached to the membrane-spanning domain [12–15]. The core I and core II proteins are structurally similar and form a heterodimer about an approximate two-fold symmetry axis [12–15].

Mitochondrial signal sequences vary in length and share little sequence similarity [1–4], but each is cleaved at a single specific site by MPP [16]. Common features for cleavage of the extension peptides by MPP are as follows: a proximal basic amino acid residue, usually arginine, at the P1 position [17–19]; distal N-terminal basic residues generally 3–10 residues from the proximal arginine [20–23]; and an aromatic or less often another type of bulky hydrophobic residue at position P1′.

Key words: crystal structure; metallopeptidase; mitochondrial signal sequence; substrate complex; zinc binding

Correspondence: johann.deisenhofer@UTSouthwestern.edu
### Table 1. Data Collection Statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>Resolution (Å)</th>
<th>λ (Å)</th>
<th>Total Reflections/ Uniq.</th>
<th>Completeness (%)</th>
<th>Rmerge (λ) (%)</th>
<th>L/σ</th>
<th>Sites</th>
<th>Phasing Power</th>
<th>Rdata (iso/ano)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>30–2.5</td>
<td>0.9796</td>
<td>1,062,534/167,094</td>
<td>99.4 (97.5)</td>
<td>8.1 (54.1)</td>
<td>23.6 (2.2)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>α/E73Q</td>
<td>50–2.55</td>
<td>0.9430</td>
<td>753,141/157,316</td>
<td>98.7 (97.9)</td>
<td>5.8 (42.8)</td>
<td>24.0 (3.8)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SAD Analysis</td>
<td>Xe</td>
<td>15–3.0</td>
<td>1.5418</td>
<td>355,699/93,284</td>
<td>92.2 (83.5)</td>
<td>4.7 (28.5)</td>
<td>27.8 (3.9)</td>
<td>24 Xe, 7 S</td>
<td>1.45</td>
</tr>
<tr>
<td></td>
<td>Overall figure of merit to 3.45 Å: 0.48</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAD Analysis</td>
<td>COX IV</td>
<td>50–2.7</td>
<td>0.9430</td>
<td>820,094/134,234</td>
<td>99.9 (100)</td>
<td>7.9 (64.3)</td>
<td>23.4 (3.5)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>MDH</td>
<td>50–3.0</td>
<td>0.9430</td>
<td>418,644/90,666</td>
<td>93.5 (95.9)</td>
<td>9.4 (48.1)</td>
<td>13.6 (2.3)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

* Statistics for the outer resolution shell are in parentheses.
* Rmerge = ∑(hkl) - <hkl>|∑(hkl) |
* Phasing power = √(Fobs² - Fcalc²) / √Fobs²
* Rdata = √(Fobs² - Fcalc²) / √Fobs²

[21]. MPP also prefers polar residues such as histidine, serine, and threonine at positions P2 and to elucidate the substrate binding characteristics of MPP, we determined four crystal structures which include recombinant yeast MPP, the mutant α/β/E73Q MPP (MPP composed of a wild-type α subunit and an E73Q mutant β subunit), α/E73Q MPP complexed with synthetic yeast malate dehydrogenase (MDH, residues 2–17), and α/β/E73Q MPP complexed with synthetic yeast cytochrome c oxidase subunit IV (COX IV, residues 2–25). Our findings indicate that mitochondrial import presequences bind in a large polar cavity of MPP in an extended conformation. The presequence structures bound to MPP differ from the amphiphilic α-helical structure observed for the synthetic precursor peptide bound to the Tom20 import component, which suggests that presequences adopt different conformations during mitochondrial import and processing.

## Results

### MPP and Cytochrome bc, Complex Core Proteins

The three-dimensional structure of recombinant yeast MPP was determined using X-ray crystallographic techniques. The yeast MPP α and β subunits have 34% overall sequence identity to the core II and core I subunits of bovine BC1, respectively. However, the phase problem could not be solved with the molecular replacement method using the core proteins as a search model. Phases were obtained using multiwavelength anomalous diffraction (MAD) data [32] collected from a selenomethionine variant of MPP. Data collection and refinement statistics are summarized in Tables 1 and 2, respectively. The four molecules in the asymmetric unit show distinct differences in crystalline order. Here, we discuss the structure of the best-ordered molecule.

The MPP fold is similar to the fold described for the core proteins of bovine BC1 [12, 14]. Each MPP subunit contains two domains of ~210 residues with nearly identical folding topology, related by an approximate two-fold rotation. The individual domains contain three α helices packed against one face of a six-stranded β sheet with five other α helices situated at one end of the sheet. One of the set of five α helices packs against the β sheet of the neighboring domain. The domains are connected by a linker of 16 residues in the middle domain and 22 residues in the β subunit. The subunits are related to each other by an approximate two-fold rotation (Figures 1a and 1b). The crystal structure of yeast BC1 revealed that the yeast core proteins have generally similar folds to the bovine core proteins, but significant differences are evident in the yeast core II protein, including the absence of an α helix of the C-terminal domain that packs against the β sheet of the N-terminal domain [15]. Accordingly, yeast MPP is more structurally similar to the bovine core proteins than to those in yeast. The primary structure of yeast MPP is also more similar...
Table 2. Refinement Statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>$R_{	ext{work}}$</th>
<th>$R_{	ext{free}}$</th>
<th>Number Protein Atoms</th>
<th>Number Heterogen Atoms</th>
<th>Number Solvent Atoms</th>
<th>Rmsd Bond Lengths</th>
<th>Rmsd Bond Angles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wild-type</td>
<td>0.245</td>
<td>0.279</td>
<td>27,804</td>
<td>34</td>
<td>223</td>
<td>0.013</td>
<td>1.6</td>
</tr>
<tr>
<td>α/E73Q</td>
<td>0.243</td>
<td>0.256</td>
<td>27,630</td>
<td>4</td>
<td>91</td>
<td>0.012</td>
<td>1.6</td>
</tr>
<tr>
<td>α/E73Q/COX IV</td>
<td>0.229</td>
<td>0.264</td>
<td>27,952</td>
<td>34</td>
<td>4</td>
<td>0.013</td>
<td>1.6</td>
</tr>
<tr>
<td>α/E73Q/MDH</td>
<td>0.217</td>
<td>0.256</td>
<td>27,891</td>
<td>34</td>
<td>0</td>
<td>0.014</td>
<td>1.6</td>
</tr>
</tbody>
</table>

to the bovine core proteins than to the yeast core proteins [15].

Individually, the subunits of MPP superimpose well with their counterparts in bovine BC1. The yeast MPP β subunit superimposes onto the core I protein of BC1 with a root-mean-square deviation (rmsd) of 1.8 Å for 429 of 443 MPP Cα atoms. The MPP α subunit superimposes onto the core II protein with an rmsd of 2.3 Å for 415 of 457 MPP Cα atoms. The difference in residues 284–301 of α-MPP (Figures 1a and 1b) compared to core II is of particular interest since these residues are part of a glycine-rich loop, which is highly conserved in α-MPP but not conserved in BC1 or present in β-MPP. This loop reaches toward the active site of β-MPP.

Although the individual subunits of MPP and the core proteins superimpose, the MPP heterodimer does not superimpose as well with the core I/II heterodimer of BC1. When only β-MPP and core I are superimposed, α-MPP and core II are clearly misaligned. This misalignment is oriented about an axis of rotation that is near the main interface between the subunits. A 15° rotation about this axis will superimpose the core II protein onto α-MPP. The difference in subunit interactions is also illustrated by a comparison of the orientations of intra-subunit two-fold rotation axes of MPP and bovine BC1 core proteins in Figure 1c. The residues that form the interfaces between core proteins or MPP are not well conserved.

Substrate Binding Region

The active site of MPP is located in a large central cavity situated between the α and β subunits that is lined with hydrophilic amino acids, including many glutamate and aspartate residues, from both subunits. A surface map representation of MPP colored according to electrostatic potential reveals that this cavity is negatively charged.
predicted by the presence of the inverted zinc binding motif (HxxEHx76E) in the primary sequence [8, 34]. The inversion in the zinc binding motif is a result of reverse main chain orientation of the α helices that contain the zinc binding residues compared to those of the metalloprotease thermolysin. Although MPP lacks significant sequence similarity to thermolysin, the enzymes exhibit both functional and structural convergence in the active site [35]. The water molecule coordinated to Zn$^{2+}$ in MPP is within hydrogen bonding distance to Glu-73, which is predicted to polarize this water molecule, thereby aligning it for nucleophilic attack on the carbonyl carbon of the peptide bond of the substrate in the same fashion as in the thermolysin model of catalysis [36]. Furthermore, two active site residues are proposed to stabilize the oxyanion of the tetrahedral intermediate formed in thermolysin [36], but no analogous residues are observed in the vicinity of the active site of MPP.

**MPP Complexed with Synthetic Peptide Substrates**

Crystal structures of a mutant MPP complexed with two different synthetic peptide substrates have been obtained which reveal the orientation and mode of substrate binding at the active site. The mutant MPP has a glutamine substitution at residue Glu-73, which reduces the enzymatic activity of MPP to undetectable levels [34]. The uncomplexed α/βE73Q MPP has a Zn$^{2+}$ ion bound in the active site and its structure does not differ significantly from the wild-type MPP structure. The mutant and wild-type structures have an rmsd of 0.14 Å for all C$_\text{α}$ atoms and an rmsd of 0.37 Å for all atoms in βHis-70, βGlu-73, βHis-74, and βGlu-150, which are involved in zinc binding and catalysis. The signal peptide COX IV 2–25 (LSLRQSIRFFKPATRT-LCSSRYLL, cleavage site between threonine and leucine) was cocrystallized with α/βE73Q MPP and resulted in a noncovalently bound substrate complex that may provide a model for the Michaelis complex of this enzyme. In the structure at 2.7 Å resolution, residues 7–19 of the peptide are observed in the electron density, while the remaining residues 2–6 and 20–25 are disordered. The substrate

![Figure 2. The Central Cavity of MPP](image)

(a) Electrostatic surface representation of MPP contoured at ±15 kcal calculated by GRASP [58]. Positive charge is shown as blue and negative charge as red. The flexible loop (residues α284–α301) is circled.

(b) A cutaway view of the surface representation of MPP revealing the electrostatic potential of the central cavity.

charged, which contrasts with the more neutral surface potential over the exterior surface of the enzyme (Figures 2a and 2b). Signal peptide substrates are rich in positively charged amino acid residues; thus, the presence of negatively charged residues in the substrate binding region may provide opportunities for the formation of stabilizing electrostatic interactions between the enzyme and its substrate. Additionally, the highly polar cavity may disfavor the amphiphilic α-helical structure that has been determined previously for several peptide substrates in membrane-mimetic environments. The ribbon drawing (Figure 1b) and surface representation (Figure 2a) of MPP show that the accessibility of the active site to substrate is partially blocked by the glycine-rich loop formed by residues 284–301 of α-MPP. The electron density is weak for much of this loop, indicating flexibility. A recent study has shown that MPP containing a deletion of part of this loop has a significantly reduced affinity for substrate peptide as well as a reduction in catalytic activity [33]. These results, in addition to the observed proximity of the loop to the active site, suggest that the glycine-rich loop of α-MPP is involved in substrate binding and/or product release, possibly dependent upon its flexibility.

A metal ion modeled as Zn$^{2+}$ is localized in the active site cavity. It is coordinated by βHis-70, βHis-74, and βGlu-150, with βGlu-73 positioned to act as a general base catalyst on a water molecule that occupies a fourth coordination site of the Zn$^{2+}$ ion (Figure 3), as was pre-
forms a short β strand from its residues 16–18 that hydrogen bonds with a β strand of MPP (residues β98–β103) in an antiparallel fashion. Described in a C-terminal to N-terminal direction, the peptide then passes between two α helices and associates loosely with another β strand (residues β322–β329) in the same subunit; beyond this point, the structure of the N terminus of the substrate is disordered. The peptide bond would be cleaved by wild-type MPP between residues 17 and 18; thus, residues on both sides of the uncleaved peptide bond are observed in the crystal structure. NMR has shown that the COX IV signal sequence region has an α-helical structure at the N-terminal portion of the peptide (residues 4–11), while the C-terminal region does not form a regular secondary structure in a micellar environment [37]. It should be noted that the crystallization conditions for MPP and the peptide complexes include 0.2 mM n-dodecylmaltoside, which provides a micellar environment. Nevertheless, COX IV is observed bound to αβE73Q MPP in an extended conformation in the crystal structure (Figure 4a). The carbonyl oxygen of the scissile peptide bond points toward the active-site Zn²⁺ ion. The arginine at position P₂ is highly conserved among mitochondrial import sequences. The side chain of this amino acid occupies a negatively charged S₂ site containing βGlu-160 and βAsp-164 and forms a salt bridge with βGlu-160. These acidic residues are highly conserved among β-MPPs. The S₁ β site appears to accommodate the bulky hydrophobic (often aromatic) residues of substrates at position P₁. In the case of COX IV, the leucine at position P₁ is located near βPhe-77 in the S₁ β site. βPhe-77 is also highly conserved among β-MPPs.

The signal peptide MDH 2–17 (LSRVAKRA-FSSTVANP, cleavage site between alanine and phenylalanine) was cocrystallized with αβE73Q MPP and, unlike the αβE73Q MPP/COX IV complex structure, formed a product complex in the crystal structure instead of forming the anticipated substrate complex. Residual activity of the mutant MPP, acting over the long time period before the harvesting of the crystals, could account for the observed proteolysis. To test the activity of αβE73Q MPP under conditions similar to crystallization, including overnight incubation, the mutant enzyme was incubated with COX IV and MDH peptides following the previously established procedure [20]. The MDH peptide was 1.5% cleaved by αβE73Q MPP and cleavage was not detected for the COX IV peptide. Residues 7–8 of the product form main chain hydrogen bonds with a β strand of MPP (residues β98–β103) in an antiparallel fashion. Additionally, the P₂ arginine is found in the negatively charged S₂ site formed from the β subunit residues βGlu-160 and βAsp-164. As observed with the COX IV peptide, the P₂ arginine of MDH forms a salt bridge with βGlu-160. The structure at 3.0 Å resolution reveals that one of the C-terminal oxygens of the cleaved MDH peptide acts as a ligand to the Zn²⁺ ion, occupying the position of the initial water ligand (Figure 4b). This observation is consistent with the final step proposed for thermolysin-catalyzed proteolysis, in which the product is coordinated to the Zn²⁺ ion at the position previously occupied by a water molecule [36]. The mutation of the catalytic βGlu-73 to glutamine provides a hydrogen bond from the amide group of the side chain to the free carboxylate oxygen of the product, which may stabilize the product bound at the active site. The coordination pattern of βGlu-150 on the Zn²⁺ ion in both the COX IV and MDH complexes was assigned as bidentate based on the electron-density maps. However, these structures were determined at medium to low resolution, which leaves open the possibility that βGlu-150 is actually involved in monodentate coordination of the Zn²⁺ ion. Since an NMR structure is not available for MDH residues 2–17, the amino acid sequence was submitted

Figure 4. Structures of Presequence Peptides Bound to αβE73Q MPP Shown in Different Scales and Orientations
(a) Simulated annealing omit map (Fo-Fc) calculated to 2.7 Å resolution of COX IV bound to αβE73Q MPP (COX IV was omitted from the model) and contoured at 3.5 σ.
(b) Simulated-annealing omit map (Fo-Fc) calculated to 3.0 Å resolution of MDH bound to αβE73Q MPP (MDH was omitted from the model) and contoured at 4.0 σ.
Figure 6. Ribbon Drawing of MPP Illustrating a Proposed "Substrate Binding Scaffold" Formed from the Edges of Four \( \beta \) Sheets. Strands 1 and 2 (red) are shown in \( \beta \)-MPP (blue); strands 3 and 4 (green) are shown in \( \alpha \)-MPP (yellow). The \( \beta \) strands are shown as arrows; the remainder of the polypeptide is shown as a C trace drawing for clarity. All figures were created using MOLSCRIPT [59], BOBSCRIPT [60], GL_RENDER [L. Esser, personal communication], and/or POV-Ray [61].

Discussion

The crystal structure of recombinant yeast MPP and complexes of the \( \alpha/\beta \)E73Q mutant MPP with substrate peptides has provided the basis for a model of recognition and binding of the mitochondrial matrix targeting signal sequences. The large cavity formed between the \( \alpha \) and \( \beta \) subunits (Figure 2b) is distinctly negatively charged due to a high incidence of aspartate and glutamate residues and can therefore be expected to accommodate the multiple positively charged amino acid residues common to mitochondrial matrix targeting signal sequences. Recent studies of the effect of ionic strength on MPP activity and affinity for substrates indicate electrostatic recognition of matrix targeting signals by MPP [39], which further underscores the importance of the negatively charged character of the central cavity. Sequence analysis and biophysical studies suggest that the presequences of several mitochondrial proteins can adopt an amphiphilic \( \alpha \)-helical structure. In contrast, our observations show two synthetic signal peptides bound to \( \alpha/\beta \)E73Q mutant MPP in an extended conformation (Figures 4a, 4b, and 5a). Both substrate peptides form main chain hydrogen bonds with the edge of the \( \beta \) sheet located at the active site (strand 1 in Figure 6). The edges of three other \( \beta \) sheets are available (strands 2–4 in Figure 6) within the cavity to offer main chain hydrogen bonds to longer matrix presequences. It can be concluded from the MDH product complex structure that short signal sequences of matrix preproteins interact only with strands 1 and 2, while longer extension pep-
tides of greater than 20 residues may interact with strands 3 and/or 4. Indeed, biochemical experiments suggest that mutation of negatively charged residues on an α helix positioned between the β sheets formed from strands 3 and 4 adversely affects the binding of longer signal peptide substrates of aspartate aminotransferase and adrenodoxin precursors [40]. In the crystal structures of bovine and chicken mitochondrial BC1 [13, 14], the signal peptide of the Rieske [2Fe-2S] protein is observed interacting with the BC1 counterparts of these β strands. The biochemical studies and observations from crystal structures of MPP and the core I and core II proteins complexed with peptides suggest that the edges of the β sheets of MPP at the interior of the cavity formed between the α and β subunits provide a substrate binding scaffold. The formation of main chain hydrogen bonds between extended presequence substrates and the scaffold β strands alone is not strongly dependent on the substrate sequence. The microenvironment at each strand, including negatively charged residues, seems to determine substrate binding by charge complementarity to positively charged residues characteristic of presequence substrates.

The crystal structures of α/β E73Q MPP complexed with COX IV and MDH have revealed elements in MPP that recognize residues in the region of the proteolytic cleavage site (Figure 5a). β-Glu-160 and β-Asp-164 form a negatively charged S2 site near the zinc binding site, which accommodates the P2 arginine commonly found in substrates. MPP has a preference for arginine at the P2 position of substrates, followed by lysine and to a lesser extent, alanine [20, 23]. Although a P2 arginine is a strong signal for recognition as a substrate, it is not decisive for determining cleavage (e.g., bovine cytochrome P-450[SCC] has a cleavable signal sequence that contains a P2 alanine). We observe that the P2 arginines of both MDH and COX IV form salt bridges with β-Glu-160. Recent biochemical studies show that α/β E160A, α/β D164A, and α/β E160A, D164A mutant MPPs, have reduced activity toward substrates with a P2 arginine but show no loss of activity toward substrates with a P2 alanine (S.K. and A.I., unpublished data). Accordingly, P2 arginine may be cooperatively recognized by β-Glu-160 and β-Asp-164 since both are required for full activity, while a substrate with a P2 alanine is unaffected by a loss of negative charge at the S2 site. We observe β-Phe-77 in the S2′ site, which accommodates bulky hydrophobic (often aromatic) P1 residues of substrates. In the α/β E73Q MPP/COX IV structure, the side chain of P1′ leucine binds near the side chain of β-Phe-77. MPP also shows a preference for polar residues such as histidine, serine, and threonine at positions P2 and P1′, but the C terminus of COX IV (P1′-P2) was disordered in the structure. The electron density was weak for the P1′ cysteine and not observed for the P1′ serine, so we do not have a clear picture to model possible interactions of the C-terminal region from the scissile peptide bond with MPP. The P1′-P1″ of synthetic COX IV may also behave differently than in a MPP/COX IV full-length preprotein complex.

Although MPP has an inverted zinc binding sequence compared to the prototypical zinc binding motif of thermolysin (HExXH), the active site region of MPP (Figures 3 and 5a) resembles that of thermolysin. It has been shown previously in a modeling study that the putative zinc binding residues and catalytic glutamate residue of the BC1 core I protein superimpose with the zinc binding residues of thermolysin (Protein Data Bank [PDB] entry 4TMN) with an rmsd of 1.56 Å [35]. Additionally, weak processing activity has been detected in crystalline bovine heart BC1 [41]. The zinc binding and water activating residues from the active sites of MPP and thermolysin superimpose with the same rmsd of 1.55 Å for the four Cα atoms compared. Furthermore, the alignment of the four residues superimposes carbobenzoxyl-L-Phe′-L-Leu-L-Ala (the superscript P indicates a phosphonamide moiety replacing the peptide bond), a substrate analog bound to thermolysin [42], onto the COX IV peptide bound to MPP with striking similarity. The backbone atoms of the thermolysin substrate analog and the COX IV peptide all orient in the same directions to form hydrogen bonds with the edge of the β sheet at the active sites of both enzymes, further illustrating the functional and structural convergence of these protease families. The arrangement of the substrate in MPP forms an antiparallel β sheet interaction at the active site comparable to what was predicted for thermolysin substrates [43]. Additionally, the model of thermolysin catalysis indicates that the carbonyl oxygen between residues 113 and 114 and the carbonyl oxygen from the side chain of Asn-112 at the S1′ site provide hydrogen bond acceptors to stabilize the amide group at position P1′. MPP has analogous hydrogen bond acceptors at the S1′ site in the carbonyl oxygen between β101 and β102 as well as in the carbonyl oxygen from the side chain of βAsn-100. Overall, we conclude from these comparisons that the model of catalysis in thermolysin can be applied to MPP. The active site of MPP differs from that of thermolysin in the absence of analogous residues to Tyr-157 and His-231, which are proposed to stabilize the oxyanion of a tetrahedral intermediate. It is conceivable that solvent molecules could donate hydrogen bonds to fulfill this role in MPP.

The mechanism of substrate entry into the large cavity of MPP containing the active site is unknown but the crystal structure of MPP does offer clues. A highly conserved glycine-rich loop structure characteristic to α-MPP is observed in close proximity to the active site at the exterior of the enzyme (Figures 1a, 1b, and 2a). Deletion of this loop results in diminished affinity for substrate and catalytic efficiency for MPP [33]. The glycine-rich loop structure may be involved in recruiting the substrate presequences to the active site and/or in releasing the product from the enzyme. The presence of several glycine residues in this loop, in addition to the crystallographic evidence of weak electron density for this loop in the wild-type MPP and uncomplexed α/β E73Q MPP structures, suggest that it is a flexible surface loop in the enzyme. As such, it may be able to expose or limit access of the substrates to the active site. The loop may also interact with part of the mature protein near the scissile peptide bond. In the α/β E73Q MPP/peptide complex structures, the electron density for the glycine-rich loop is absent for part of the loop near the active site (residues α286–α292), suggesting that binding of substrates may trigger conformational change in the glycine-rich loop.
The results of our structural studies provide evidence that the structures required for recognition by mitochondrial import components and proteolytic processing are different. It is believed that the formation of amphiphilic α-helices in the N-terminal presequences of proteins targeted to the mitochondria is a critical factor in identification by the import machinery. This is illustrated in a recent NMR structure of a presequence peptide bound to a component of the mitochondrial outer membrane translocase complex [30]. In this structure, the pre-sequence adopted the anticipated α-helical conformation. Our structures of presequences bound to a point mutant of MPP reveal that the substrate peptides exhibit an extended conformation, thereby supporting the hypothesis that elements in the signal sequences are used differently for import and for processing. Therefore, mitochondrial import signal sequences appear to have context-dependent structures. The signal sequence adopts an amphiphilic α-helical conformation at the point of recognition by mitochondrial translocation components. Once the preprotein is translocated from the cytosol through the outer and inner mitochondrial membranes to the matrix, a cleavable signal sequence is recognized in an extended conformation by processing machinery where it is removed.

Biological Implications

Nuclear-encoded proteins destined for the mitochondrial matrix undergo numerous steps from translation to translocation to the matrix. An N-terminal presequence functions as a signal for import to the matrix. Mitochondrial precursor proteins are transported from the cytosol through the mitochondrial outer membrane and inner membrane into the matrix where the signal sequence is processed by mitochondrial processing peptidase (MPP). MPP specifically recognizes a wide variety of mitochondrial precursor proteins and cleaves them at a single site. Therefore, MPP plays an important role in the maturation of mitochondrial precursors and in mitochondrial biogenesis and function. Mitochondrial import signal sequences vary in length and sequence but contain some common elements including an overall basic charge, typically an arginine at position –2 (in the N-terminal portion) from the cleavage site, and an α-helical conformation. Our structures of presequences bound to a point mutant of MPP reveal that the substrate peptides exhibit an extended conformation, thereby supporting the hypothesis that elements in the signal sequences are used differently for import and for processing. Therefore, mitochondrial import signal sequences appear to have context-dependent structures. The signal sequence adopts an amphiphilic α-helical conformation at the point of recognition by mitochondrial translocation components. Once the preprotein is translocated from the cytosol through the outer and inner mitochondrial membranes to the matrix, a cleavable signal sequence is recognized in an extended conformation by processing machinery where it is removed.

Experimental Procedures

Protein Expression and Purification

Recombinant wild-type and α/E73Q mutant yeast MPP were over-expressed in Escherichia coli BL21(DE3) cells with a hexahistidine tag on the C terminus of the α subunit and purified according to a published procedure [40]. The purification scheme was modified to use 0.2 mM n-dodecylmaltoside (Boehringer Mannheim) in the place of 0.1% Tween 20. Purified MPP was stored at –80°C until use. A seleno-methionine variant of MPP was overexpressed in BL24(DE3) cells in minimal medium in the presence of 30 mg/L L-seleno-methionine. The seleno-methionine variant was purified using the published procedure for the wild-type protein with the addition of 1 mM of the reducing agent tris(2-carboxyethyl)phosphine hydrochloride (Calbiochem) to all buffers. Expressed MPP contains the following mutations: α/E177G, α/E217G, β/S20A, β/S84P, and β/Q350R. All are assumed PCR errors with the exception of β/S20A, which is a cloning artifact. These mutations do not affect the enzymatic activity.

Peptide Synthesis

Yeast malate dehydrogenase peptide substrate (residues 2–17, LSRAVKAFTSVSTYAN) and yeast cytochrome c oxidase peptide IV substrate (residues 2–25, LSRGSIFFPPARLTCYSSRYLL) were synthesized using standard solid-phase Fmoc (N-9-fluorenylmethoxycarbonyl) synthesis methodology by Anaspec Incorporated (San Jose, CA) and the Howard Hughes Medical Institute Biopolymers Facility at the University of Texas Southwestern Medical Center (Dallas, TX), respectively.

Crystallization and X-Ray Data Collection

The recombinant wild-type, α/E73Q mutant, and seleno-methionine variant of MPP were crystallized by the hanging-drop vapor diffusion method at 21°C. The crystal drops contained 2 μl of protein (8–14 mg/ml in 50 mM HEPES [pH 7.5], 30% [v/v] glycerol, 0.2 mM n-dodecylmaltoside) and 2 μl of reservoir solution (6%–9% [w/v] PEG 10,000, 35% [v/v] ethylene glycol, 6% [v/v] 2-methyl-2,4-pentanediol, 25 mM sodium citrate [pH 6.0], 2% [w/v] benzamidine, 0.2 mM n-dodecylmaltoside, 2 mM sodium azide). Single crystals grew in a trapezoidal, ingot-like morphology to average dimensions of 0.20 mm × 0.20 mm × 0.15 mm. Crystals were harvested typically after 2–3 weeks. The crystals have unit cell constants of a = 133 Å, b = 176 Å, c = 201 Å and the symmetry of space group P2_12_12_1. The Matthews parameter suggested 3–6 molecules per asymmetric unit [44].

The crystals were flash-cooled in liquid propane prior to data collection. A xenon derivative was prepared using the Hampton Research Xenon Chamber with a wild-type crystal pressurized to 180 psi for 15 min followed by flash-cooling in liquid propane. The α/E73Q mutant MPP/peptide complex crystals were prepared by cocrystallization of the α/E73Q mutant MPP with the same reservoir solution listed above with the addition of a solution of peptide dissolved in 50 mM HEPES (pH 7.5) and 0.2 mM n-dodecylmaltoside such that the molar ratio of peptide to MPP was 25–50 to 1. The peptide was added only to the crystal drop.

The largest wild-type crystals diffracted to 2.9 Å resolution with 0.2 mM n-dodecylmaltoside (Boehringer Mannheim) in the place of 0.1% Tween 20. Purified MPP was stored at –80°C until use. A seleno-methionine variant of MPP was overexpressed in BL23(DE3) cells in minimal medium in the presence of 30 mg/L L-seleno-methionine. The seleno-methionine variant was purified using the published procedure for the wild-type protein with the addition of 1 mM of the reducing agent tris(2-carboxyethyl)phosphine hydrochloride (Calbiochem) to all buffers. Expressed MPP contains the following mutations: α/E177G, α/E217G, β/S20A, β/S84P, and β/Q350R. All are assumed PCR errors with the exception of β/S20A, which is a cloning artifact. These mutations do not affect the enzymatic activity.

Peptide Synthesis

Yeast malate dehydrogenase peptide substrate (residues 2–17, LSRAVKAFTSVSTYAN) and yeast cytochrome c oxidase peptide IV substrate (residues 2–25, LSRGSIFFPPARLTCYSSRYLL) were synthesized using standard solid-phase Fmoc (N-9-fluorenylmethoxycarbonyl) synthesis methodology by Anaspec Incorporated (San Jose, CA) and the Howard Hughes Medical Institute Biopolymers Facility at the University of Texas Southwestern Medical Center (Dallas, TX), respectively.

Crystallization and X-Ray Data Collection

The recombinant wild-type, α/E73Q mutant, and seleno-methionine variant of MPP were crystallized by the hanging-drop vapor diffusion method at 21°C. The crystal drops contained 2 μl of protein (8–14 mg/ml in 50 mM HEPES [pH 7.5], 30% [v/v] glycerol, 0.2 mM n-dodecylmaltoside) and 2 μl of reservoir solution (6%–9% [w/v] PEG 10,000, 35% [v/v] ethylene glycol, 6% [v/v] 2-methyl-2,4-pentanediol, 25 mM sodium citrate [pH 6.0], 2% [w/v] benzamidine, 0.2 mM n-dodecylmaltoside, 2 mM sodium azide). Single crystals grew in a trapezoidal, ingot-like morphology to average dimensions of 0.20 mm × 0.20 mm × 0.15 mm. Crystals were harvested typically after 2–3 weeks. The crystals have unit cell constants of a = 133 Å, b = 176 Å, c = 201 Å and the symmetry of space group P2_12_12_1. The Matthews parameter suggested 3–6 molecules per asymmetric unit [44].

The crystals were flash-cooled in liquid propane prior to data collection. A xenon derivative was prepared using the Hampton Research Xenon Chamber with a wild-type crystal pressurized to 180 psi for 15 min followed by flash-cooling in liquid propane. The α/E73Q mutant MPP/peptide complex crystals were prepared by cocrystallization of the α/E73Q mutant MPP with the same reservoir solution listed above with the addition of a solution of peptide dissolved in 50 mM HEPES (pH 7.5) and 0.2 mM n-dodecylmaltoside such that the molar ratio of peptide to MPP was 25–50 to 1. The peptide was added only to the crystal drop.

The largest wild-type crystals diffracted to 2.9 Å resolution with 0.2 mM n-dodecylmaltoside (Boehringer Mannheim) in the place of 0.1% Tween 20. Purified MPP was stored at –80°C until use. A seleno-methionine variant of MPP was overexpressed in BL23(DE3) cells in minimal medium in the presence of 30 mg/L L-seleno-methionine. The seleno-methionine variant was purified using the published procedure for the wild-type protein with the addition of 1 mM of the reducing agent tris(2-carboxyethyl)phosphine hydrochloride (Calbiochem) to all buffers. Expressed MPP contains the following mutations: α/E177G, α/E217G, β/S20A, β/S84P, and β/Q350R. All are assumed PCR errors with the exception of β/S20A, which is a cloning artifact. These mutations do not affect the enzymatic activity.
crystal complexes with a peptide, all have a characteristically high Wilson B factor greater than 45. This is evident in visual inspection of the data showing a significant drop in overall intensity of data in the resolution range greater than ~5 Å. Data collection statistics are summarized in Table 1.

**Phasing and Structure Refinement**
It was anticipated that the structure of MPP would be solved by the molecular replacement method using the core I and core II proteins from BC1 as a search model. However, multiple attempts to solve the phase problem using the molecular replacement method failed to yield a complete solution for the asymmetric unit. MPP crystals were screened for heavy atom derivatives to no avail due to lack of isomorphism with wild-type data sets. MAD phasing with a seleno-methionine variant seemed a suitable approach since it avoids the nonisomorphism problem; however, in the case of MPP, a large number of selenium positions (76) had to be identified. Direct methods and Patterson search methods were unsuccessful in identifying enough selenium positions to calculate usable phases from MPP seleno-methionine variant data, possibly due to the poor quality (Rwp values greater than 10%) and low resolution limits of the MAD data (Table 1).

Independent phase information was deemed necessary to locate the selenium positions. A data set for an MPP crystal pressurized with xenon prior to flash-cooling was collected on a rotating anode X-ray generator but was not found to be useful as an isomorphous heavy atom derivative. The data were reprocessed using the absorption correction implemented in HKL2000 [46] and, as a result, the anomalous signal was extended from 6.5 Å in the original processing to 4.5 Å in the reprocessed data set. Patterson search methods identified 27 heavy atom positions in an anomalous difference Patterson map calculated for the xenon derivative. The CNS suite [48] was used to calculate phases to 5.0 Å by the single wavelength anomalous diffraction (SAD) method, followed by solvent flattening. The map revealed four MPP molecules in the asymmetric unit, indicating a Matthews coefficient of 3.0 Å³/Da with a calculated solvent content of 59%. The phased translation function implemented in CNS calculated a reasonable fit for only one of the four molecules in the asymmetric unit using BC1 core I and core II proteins. The remaining three molecules were fit manually using the graphics program O [49]. The model underwent rigid-body refinement and phases calculated for the resultant model were combined with the xenon SAD phases using CNS. The combined phases were then used to calculate a difference anomalous Fourier map for the selenium absorption peak wavelength data, from which 72 of the expected 76 selenium sites were identified. The selenium parameters were refined and phases were calculated to 3.45 Å with CNS, using the inflection point data set as the reference data set. The selenium MAD phases were applied to a 2.5 Å data set and improved and extended using 4-fold NCS averaging. The averaged and phased model extension were carried out with multidomain density modification implemented in DM [50] using the 4-fold averaging of four masks, each defining one of four individual ~210 residue domains of the MPP molecule. This produced an interpretable map with which the BC1 core I and core II proteins served as a peptide backbone template for building the MPP model. The model revealed that 7 of the 27 heavy atom sites that were previously identified in the xenon derivative data were actually sulfur atom positions. Four of the sulfur atoms identified were from methionine residues and three from cysteine residues. Although anomalous diffraction study and inductively coupled plasma (ICP) analysis of MPP crystals indicate both Zn²⁺ and Ni²⁺ are present in the crystals (data not shown), we modeled the metal ion in the zinc binding site as Zn²⁺. We do not expect the presence of Ni²⁺ to affect our conclusions since MPP has been shown to have activity when Zn²⁺ is replaced with other divalent metal cations [51].

The MPP model was refined using simulated annealing in CNS after a nearly complete model was built [52]. Further refinement included 4-fold NCS restraints, overall anisotropic B factor calculation, and bulk solvent correction. The R/REOM and R/REOM MPP/peptide complex models were rebuilt from the wild-type model after undergoing rigid-body refinement and simulated annealing.

Due to the low resolution of the data, bond distances for the Zn²⁺ ligands were restrained according to distances compiled from similar Zn²⁺ ligands in the Cambridge Structural Database [53]. A test set reserved for crossvalidation during refinement [54] was calculated to include at least 2000 reflections for each data set using DATAMAN [55] prior to refinement. Refinement statistics are summarized in Table 2. Rmsd calculations were carried out using the programs DALI [56] or LSQMAN [57]. Intersubunit rotation angles were calculated by the DOMOV server (http://bioinfo1.mbfys.lu.se/cgi-bin/Domov/domov.cgi) provided by the Bioinformatics Group at Lund University.

**Acknowledgments**
We thank the members of the Deisenhofer laboratory for helpful discussions, in particular, Y. Chelliah for technical assistance and Drs. C. Brautigam and E. Istvan for critical reading of the manuscript, and staff members of the synchrotron beamlines for their assistance.

Received: March 14, 2001
Accepted: May 18, 2001

**References**


Accession Numbers

The atomic coordinates for the structures reported in this work have been deposited in the Protein Data Bank with the ID codes 1HR6, 1HR7, 1HR8, and 1HR9.