Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5355–5358, June 1993
Biochemistry

The β subunit of the mitochondrial processing peptidase from rat liver: Cloning and sequencing of a cDNA and comparison with a proposed family of metallopeptidases

(protease-enhancing protein/leader peptides)

VÁCLAV PAČES*, LEON E. ROSENBERG†, WAYNE A. FENTON‡, AND FRANTISEK KALOUSEK

Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510

Contributed by Leon E. Rosenberg, March 11, 1993

ABSTRACT **Most nuclearly encoded mitochondrial proteins are synthesized with amino-terminal leader peptides that are removed by the mitochondrial processing peptidase (MPP) after translocation. Earlier we reported cloning and sequencing of a cDNA for the larger subunit (MPP α subunit) of this enzyme from rat liver mitochondria. We have now completed the cloning and sequencing of a cDNA encoding the smaller subunit of the enzyme (MPP β subunit) from the same source. The cDNA consists of 1570 bp: 17 bp of 5′-untranslated sequence, 1467 bp of coding sequence, and 86 bp of 3′-untranslated sequence. The predicted protein consists of 489 amino acid residues, including a 45-amino acid leader peptide at the amino terminus and a 444-amino acid mature protein. The amino acid sequences of four tryptic peptides derived from purified MPP β subunit precisely match those predicted by the cDNA sequence, as does the predicted mature amino terminus. The amino-terminal sequence is typical of a mitochondrial leader peptide, with eight positively charged arginine residues and a single negatively charged aspartate residue. When the amino acid sequence of rat MPP β subunit is compared with sequences in the protein data bases, significant homology is found with the protease-enhancing protein of *Neurospora crassa*, the smaller subunit of MPP from *Saccharomyces cerevisiae*, and the core I protein of bovine ubiquinol:cytochrome c reductase. Lower homology is found with other members of a recently proposed class of endoproteases, which includes human insulinase and protease III from *Escherichia coli*.**

The transport of proteins from cytoplasm to mitochondria has been studied in a variety of systems (1). Most mitochondrial proteins are encoded by nuclear genes and are synthesized as larger precursors with amino-terminal leader peptides of 10 to 70 amino acid residues (1). These sequences are rich in positively charged amino acids and contain sufficient information to target these proteins to mitochondria. It has been shown that most precursors are cleaved during import in either one or two steps and that the two-step cleavage requires two different proteolytic enzymes (2). Most leader peptides are removed in one step by the mitochondrial processing peptidase (MPP), whereas others are first processed by MPP to an intermediate form that is cleaved by a second enzyme, mitochondrial intermediate peptidase, to generate the mature amino terminus (2, 3).

The first protease has been isolated from *Neurospora crassa*, where it consists of two dissociable components, the larger of which has some catalytic activity and has been called MPP and the smaller of which increases the activity of the complex and has been referred to as the protease-enhancing protein (PEP) (4). In *Saccharomyces cerevisiae* (5) and rat liver (2, 6, 7), MPP apparently is a true heterodimer of two dissimilar subunits, α (≈55 kDa) and β (≈50 kDa). In these species, it is not yet clear whether protease activity resides solely in the dimer or whether one of the two subunits may be active by itself. The larger α subunit of rat liver MPP (α-MPP) has been purified, and a cDNA encoding it has been cloned and sequenced (6). The mature protein shows 36% identity with the larger subunits of *S. cerevisiae* and *N. crassa* MPP.

We now report isolation of the β subunit of rat liver MPP (β-MPP), the cloning and sequencing of a cDNA encoding this protein,§ and an initial analysis of its relationship to members of a proposed family of metalloendopeptidases characterized by the lack of a conventional consensus metal-binding site (8).

MATERIALS AND METHODS

Enzyme Purification and Peptide Sequencing. MPP was prepared from rat liver mitochondria as described (2) with a minor modification. To remove contaminants that migrated with β-MPP on SDS/PAGE, only the first 30% of the activity peak after DEAE-Biogel was pooled. After chromatography on ω-aminooctyl-agarose, the active fraction was virtually free from impurities, as judged by SDS/PAGE and amino-terminal amino acid analysis of the β-MPP band excised from the gel. Generation of tryptic peptides from the β-MPP band and isolation and sequencing of these peptides were done as described (6). Of the several peptides sequenced, two of them—namely, the amino-terminal peptide and peptide P (see Fig. 1)—were used to design degenerate oligonucleotide primers. Peptide P was chosen because it matched an amino acid sequence conserved between *Neurospora* PEP (amino acids 398–408) and the smaller subunit of yeast MPP (amino acids 387–397; see Fig. 2). The primer sequences were 5′-CARGCHGCHCCHCARGTBGTBYTBAAYGTBCC-3′ [amino-terminal peptide (sense)] and 5′-RTCYTCRTTD-ATNGGNGTNSWNCCRTCNARYTG-3′ [peptide P (antisense)].

cDNA Amplification and Identification. Poly(A)$^+$ RNA from rat liver was prepared as described (9) and used for synthesis of cDNA with random primers (cDNA Cycle kit; Invitrogen, San Diego). The single-stranded cDNA was used as a template in a 100-μl PCR amplification with 500 pmol of each of the primers corresponding to the amino-terminal

Abbreviations: MPP, mitochondrial processing protease; α-MPP and β-MPP, MPP α and β subunits, respectively; PEP, protease-enhancing protein.

*Present address: Institute of Molecular Genetics, Czech Academy of Sciences, CS-166 10 Prague, Czech Republic.
†Present address: Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, NJ 08543.
‡To whom reprint requests should be addressed.
§The sequence reported in this paper has been deposited in the GenBank data base (accession no. L12965).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

peptide and peptide P of the protein. The reaction conditions were as follows: 94°C, 1 min; 45°C, 1 min; 72°C, 2 min; 25 cycles. The high homology among the peptides derived from the rat β subunit and those from corresponding regions in the yeast small subunit and *Neurospora* PEP suggested that a PCR product of ≈1100 bp should be generated with these primers. A product of this size was found upon agarose gel electrophoresis (data not shown). The band was excised, and the DNA was eluted from the gel and amplified again under the same conditions. Two units of T4 DNA polymerase were added to the completed PCR mixture, and it was incubated 15 min at 37°C. The blunt-ended products were ligated with *Eco*RV-digested pBluescript (Stratagene). *Escherichia coli* DH5α cells (BRL) were transformed with the ligation mixture and plated on indicator plates, and white colonies were screened for the presence of the insert by PCR amplification using T3 and T7 sequencing primers (Stratagene), corresponding to sites in the vector flanking the insert. Selected plasmids were prepared by an alkaline lysis protocol (10), and the inserts were sequenced from both ends by a dideoxynucleotide chain-termination method (11) using Sequenase (United States Biochemical) according to the supplier's protocols.

Colony Screening and Cloning of the 5′ End. The PCR-generated partial cDNA was labeled with [α-^{32}P]dCTP (Amersham) using a random primer DNA labeling kit (Boehringer Mannheim) and used to screen a rat liver cDNA library in the λ ZapII vector (Stratagene), following the suppliers' protocols. Several clones were isolated and sequenced. Although all had long open reading frames, none of them contained a full-size β-MPP cDNA. All began downstream of the 5′ terminus of the original PCR product; several began at the same position. The 3′ end of several of these clones extended beyond the termination codon to a poly(A) tail.

To obtain the 5′ end of the coding sequence and at least some of the 5′-untranslated region, we carried out a PCR protocol referred to as "rapid amplification of cDNA ends" (RACE) (12). First-strand cDNA was prepared as above, starting with 1 μg of poly(A)$^+$ rat liver RNA. Excess primers were removed by dilution and concentration with a Centricon 30 centrifugal ultrafiltration device (Amicon), repeated once. Terminal transferase (BRL) was used to add multiple dA residues to the 3′ ends of the cDNAs, essentially as described (12), except that the reaction time was extended to 90 min, instead of 2 min, to add enough residues. Two specific antisense oligonucleotide primers were synthesized based on the known sequence near the 5′ end of the longest clone, along with an adapter primer and an adapter-dT_{17} primer, as described (12). One percent (5 μl) of the tailed cDNA was used as template in a 100-μl PCR reaction with 20 pmol of adapter-dT_{17} primer, 50 pmol of adapter primer, and 50 pmol of the downstream, specific primer (corresponding to nt 431–453 in Fig. 1). The cycle pattern was as follows: 94°C, 1 min; 50°C, 30 sec; 72°C, 2 min; 25 repetitions. Then 1 μl of this reaction was used as template in a second amplification; this time 50 pmol of the adapter primer and 50 pmol of the upstream, specific primer (nt 258–284 in Fig. 1) were used. The PCR conditions were identical to those used before. When an aliquot of the reaction was analyzed by agarose gel electrophoresis, a prominent, broad band was observed at 330–360 bp (data not shown). The PCR products were recovered by extraction and precipitation, cleaved with *Eco*RI (site in the adapter primer) and *Cla* I (site in the cDNA), and cloned into pBluescript. Individual clones with the largest inserts were selected, and their inserts were sequenced as above. The sequences of four independent clones were compiled to correct for errors introduced in the PCR amplifications.

```
   1 CCCTATCCACAGCAGAGATGGCGGCGGCGGCTGTATCCCGGACGTTGTTGCCGGTGGCCGGACGGCGCCTGTGGGGATTTACGCGAAGGC   90
   1                  M  A  A  A  A  V  S  R  T  L  L  P  V  A  G  R  R  L  W  G  F  T  R  R  L   25
  91 TTCCGCTCCGCGCCGCCGCTGCTCAGCCATTGTACTTTGGAGGGGACCGACTCAGAAGTACACAGGCTGCTCCACAAGTTGTTCTGAATG  180
  26  P  L  R  A  A  A  A  Q  P  L  Y  F  G  G  D  R  L  R  S  T↑ Q  A  A  P  Q  V  V  L  N  V   55
 181 TTCCTGAGACACAAGTCACTTGTTTGGAAAATGGACTCAGGGTAGCTTCTGAAAACTCTGGGATCTCAACATGCACCGTTGGGCTGTGGA  270
  56  P  E  T  Q  V  T  C  L  E  N  G  L  R  V  A  S  E  N  S  G  I  S  T  C  T  V  G  L  W  I   85
 271 TCGATGCTGGCAGTCGGTATGAGAATGAAAAGAACAATGGCACCGCTCACTTTCTGGAGCACATGGCTTTCAAGGGCACTAAAAAGAGGT  360
  86  D  A  G  S  R  Y  E  N  E  K  N  N  G  T  A  H  F  L  E  H  M  A  F  K  G  T  K  K  R  S  115
 361 CCCAGTTAGACCTTGAACTTGAGATTGAGAACATGGGTGCCCATCTCAATGCCTATACCTCCAGAGAGCAGACTGTCTACTACGCCAAAG  450
 116  Q  L  D  L  E  L  E  I  E  N  M  G  A  H  L  N  A  Y  T  S  R  E  Q  T  V  Y  Y  A  K  A  145
 451 CCTTCTCCAAAGATTTGCCAAGAGCTGTAGAAATTCTTGCTGACATAATTCAGAACAGTACATTGGGAGAGGCAGAGATTGAACGTGAGC  540
 146  F  S  K  D  L  P  R  A  V  E  I  L  A  D  I  I  Q  N  S  T  L  G  E  A  E  I  E  R  E  R  175
 541 GTGGAGTCATCCTCAGAGAAATGCAGGAAGTTGAGACCAACTTGCAAGAAGTTGTCTTTGATTATCTGCACGCCACAGCCTATCAAAACA  630
 176  G  V  I  L  R  E  M  Q  E  V  E  T  N  L  Q  E  V  V  F  D  Y  L  H  A  T  A  Y  Q  N  T  205
 631 CCGCGCTCGGACGGACAATTCTGGGACCAACCGAGAACATCAAATCTATAAGCCGTAAGGACTTAGTGGACTACATAACCACACATTATA  720
 206  A  L  G  R  T  I  L  G  P  T  E  N  I  K  S  I  S  R  K  D  L  V  D  Y  I  T  T  H  Y  K  235
 721 AGGGACCAAGAATCGTGCTGGCTGCTGCTGGAGGTGTTTGCCATAACGAACTGCTGGAGTTAGCGAAGTTTCATTTTGGTGACTCTTTGT  810
 236  G  P  R  I  V  L  A  A  A  G  G  V  C  H  N  E  L  L  E  L  A  K  F  H  F  G  D  S  L  C  265
 811 GCGCACACAAAGGAGATGTACCAGCTCTGCCTCCCTGTAAATTCACTGGAAGCGAGATTCGGGTGAGGGATGACAAGATGCCACTGGCGC  900
 266  A  H  K  G  D  V  P  A  L  P  P  C  K  F  T  G  S  E  I  R  V  R  D  D  K  M  P  L  A  H  295
 901 ACCTCGCAGTGGCTATTGAAGCAGTTGGTTGGACGCACCCAGACACGATCCGTCTCATGGTCGCAAACACACTGATAGGCAACTGGGACC  990
 296  L  A  V  A  I  E  A  V  G  W  T  H  P  D  T  I  R  L  M  V  A  N  T  L  I  G  N  W  D  R  325
 991 GCTCTTTCGGAGGAGGAATGAATTTATCGAGCAAGCTGGCCCAGCTCACGTGCCATGGCAATCTCTGCCACAGCTTCCAGTCCTTCAACA 1080
 326  S  F  G  G  G  M  N  L  S  S  K  L  A  Q  L  T  C  H  G  N  L  C  H  S  F  Q  S  F  N  T  355
1081 CCTCCTACACAGACACAGGATTGTGGGGGCTCTATATGGTTTGTGAACAAGCCACAGTTGCTGACATGCTGCACGCTGTACAAAAGGAAT 1170
 356  S  Y  T  D  T  G  L  W  G  L  Y  M  V  C  E  Q  A  T  V  A  D  M  L  H  A  V  Q  K  E  W  385
1171 GGATGCGTCTGTGCACAGCTGTTAGTGAGAGTGAGGTCGCACGCGCCAAAAACCTTCTCAAGACAAACATGCTGCTGCAGCTTGATGGGT 1260
 386  M  R  L  C  T  A  V  S  E  S  E  V  A  R  A  K  N  L  L  K  T  N  M  L  L  Q  L  D  G  S  415
1261 CAACTCCGATCTGTGAAGACATCGGTAGGCAGATGCTATGCTACAACAGGAGAATCCCCATCCCTGAGCTTGAGGCCAGGATCGATGCTG 1350
 416  T  P  I  C  E  D  I  G  R  Q  M  L  C  Y  N  R  R  I  P  I  P  E  L  E  A  R  I  D  A  V  445
1351 TGGACGCGGAGATGGTTCGAGAAGTGTGCACCAAGTACATTTATGGCAAAAGCCCGGCCATTGCTGCTCTCGGTCCTATTGAGCGCTTAC 1440
 446  D  A  E  M  V  R  E  V  C  T  K  Y  I  Y  G  K  S  P  A  I  A  A  L  G  P  I  E  R  L  P  475
1441 CAGATTTTAACCAGATTTGTAGTAACATGCGCTGGACTCGTGACTGATTTTTGTGATGAATTATCCTACACATACATATTTATAAAACTG 1530
 476  D  F  N  Q  I  C  S  N  M  R  W  T  R  D  *                                               489
1531 AGACCTAGAAGTTTTGAAAGACTGCTCATCTTCAAGTGAA                                                   1570
```

FIG. 1. Nucleotide and deduced amino acid sequence of β-MPP from rat liver. The sequenced peptides used to design primers for PCR-aided cloning and to confirm the cloned sequence are underlined; peptide P consisted of amino acids 409–423. The arrow indicates the cleavage site between the leader peptide and mature β-MPP.

Comparison of Rat Liver β-MPP with Other Proteins. Computer programs from the Genetics Computer Group (version 7) (Madison, WI), and the Protein Identification Resource (National Biomedical Research Foundation, Washington, DC) were used to compare the nucleotide and amino acid sequences of β-MPP with sequences found in the GenBank and SwissProt data libraries.

RESULTS AND DISCUSSION

Nucleotide and Amino Acid Sequence of β-MPP. The nucleotide sequence of the cDNA encoding rat liver β-MPP and the deduced amino acid sequence of the protein are given in Fig. 1. The total length of 1570 bp includes 17 bp of 5′-untranslated region, 1467 bp of coding sequence, and 86 bp of 3′-untranslated region. The deduced amino acid sequence is 489 amino acids long, with a 45-amino acid leader peptide and a 444-amino acid mature protein. Also indicated in Fig. 1 are the peptides for which predicted sequences were confirmed by those obtained from the isolated protein and those used to generate PCR primers for the initial cloning. The calculated molecular size of the precursor protein is 54.3 kDa, whereas that of the mature protein is 49.4 kDa, slightly less than that estimated from SDS/PAGE (50–52 kDa) (6). The mature protein is mildly acidic; the predicted isoelectric point is 5.9. It has a hydrophobic amino terminus and a highly hydrophilic region in the vicinity of amino acid 95, although the general distribution of residues is consistent with that of a soluble, globular protein. The leader peptide is strongly positively charged (eight arginines, one aspartate) and has the general appearance of a mitochondrial leader sequence. The predicted cleavage site (Fig. 1), Arg-Ser-Thr- ↓ -Gln-Ala, is typical of many sites recognized by MPP itself (13, 14), suggesting that, after being imported, pre-β-MPP is cleaved by preexisting MPP. A similar prediction has been made with respect to pre-α-MPP (6).

Secondary Structure Prediction. The secondary structure of rat β-MPP was predicted by using both the Chou–Fasman (15) and Garnier–Osguthorpe–Robinson (16) algorithms (data not shown). The protein appears to have a somewhat more ordered secondary structure than *Neurospora* PEP or the yeast β subunit. Perhaps the most conspicuous feature is an α-helix predicted for the region between amino acids 165 and 205. This α-helix is predicted to be conserved among all three of these proteins and in the bovine core I protein as well. Interestingly, the first part of this region (residues 168–186) is highly charged (seven glutamates, three arginines) and appears homologous to the putative negatively charged α-helix predicted previously for rat α-MPP and the homologous yeast and *Neurospora* proteins (6). The presence of this amphiphilic, highly charged helix in both subunits of MPP gives further support to the hypothesis that it has an important role in this enzyme, perhaps as part of the site that binds the positively charged, leader peptide substrates.

Comparison with Related Enzymes and with a Family of Metallopeptidases. The amino acid sequence of rat β-MPP was compared with proteins that had been shown (8, 17) to have a similar function and/or structure (Fig. 2). The comparison was designed both to assess the overall homology to these proteins and to determine whether structural features believed to underlie the function of homologous enzymes were present in the rat protein. It has been proposed recently that *S. cerevisiae* β-MPP and *N. crassa* PEP are members of a family of metallopeptidases that includes the *S. cerevisiae* and *N. crassa* large MPP subunits, human insulinase, *E. coli* protease III, and subunits I and II of yeast ubiquinol:cytochrome *c* reductase (8). The recently published amino acid sequences of the core I and II proteins of bovine ubiquinol:cytochrome *c* reductase are also homologous to these pro-

FIG. 2. Comparison of rat β-MPP with *Neurospora* PEP, yeast smaller MPP subunit (MAS1/MIF1 protein), and bovine ubiquinol:cytochrome *c* reductase core I protein. The alignment was generated by the program PILEUP from the Genetics Computer Group and does not include the leader peptides (the mature terminus of MIF1 was deduced by cleavage-site homology). The core I alignment was not significant beyond residue 265, and the remaining residues are not included in the figure. Residues identical in all four (or all three beyond amino acid 265) sequences are shaded. Amino acid sequences were recovered or translated from GenBank and are designated β-MPP (rat β-MPP, this manuscript), PEP (*N. crassa* small subunit, no. M20928), MAS1 (= MIF1, *S. cerevisiae* small subunit, no. X07649), and BoCORE1 (core I protein of bovine ubiquinol:cytochrome *c* reductase, no. X59692).

Table 1. Comparison of rat β-MPP amino acid sequence with sequences of putatively related proteins

	α-MPP	MAS1	MAS2	PEP	MPP	$CoreI_Y$	$CoreI_B$	IDE	ProtIII
Segment comparison score, SD units	8.110	26.377	5.881	30.031	8.620	7.614	25.606	2.878	3.022
Similarity/identity, %	55/32	65/45	50/28	70/52	51/30	51/28	70/52	44/20	46/24

Data are the output of the RELATE program from the Protein Identification Resource, in which the complete amino acid sequences of the indicated proteins as translated from GenBank were used. Values are the segment comparison scores in SD units, calculated using a fragment length of 10, the Dayhoff mutation data matrix (250 accepted point mutations) (19), and 100 random runs per comparison. Values >3 or 4 suggest some homology, values >6 or 7 indicate significant homology, and values >10–12 indicate strong homology across large regions of sequence. α-MPP, α subunit of rat MPP (no. M38282); MAS2 (= MIF2), large subunit of *S. cerevisiae* MPP (no. M36596); MPP, large subunit of *N. crassa* MPP (no. J05484); $CoreI_Y$, core I protein of *S. cerevisiae* ubiquinol:cytochrome *c* reductase (no. J02636); $CoreI_B$, bovine core I, as in Fig. 2; IDE, human insulinase (no. M21188); ProtIII, *E. coli* protease III (no. X06227); other sequences are as in Fig. 2.

teins, especially in their amino-terminal portions (18). Table 1 shows that rat β-MPP is clearly most closely related to *N. crassa* PEP and *S. cerevisiae* β-MPP and also to the core I protein of bovine ubiquinol:cytochrome *c* reductase. The similarity of rat β-MPP with this last protein is very high up to amino acid 310 of β-MPP (77% similarity, 58% identity). The remaining carboxyl-terminal portions share much less homology; in addition, the bovine protein is ≈115 amino acids shorter. These observations suggest that in mammals, β-MPP and core I are products of different genes, unlike the situation in *N. crassa* in which these are identical proteins, encoded by a single gene (20).

Rat β-MPP is considerably less homologous, overall, to other members of the proposed endopeptidase family (Table 1). The residues suggested by Rawlings and Barrett (8) to be involved in metal binding—namely, His-101, Glu-181, and His-198—are conserved, however. There is another potential metal-binding site in the rat protein—namely, His-101-Phe-Leu-Glu-His. This inversion of the more usual, thermolysin-like His-Glu-Xaa-Xaa-His zinc-binding motif (21) is also present in the yeast β subunit and *N. crassa* PEP. Interestingly, this site is not precisely conserved in the α-MPP subunits or in yeast and bovine core I, but it is present in insulinase (22) and protease III (23). Recently, *in vitro* mutagenesis of these residues in protease III has revealed that they are important for both zinc binding and catalytic activity (24). Substitution of arginine for either of the histidine residues abolishes both zinc binding and activity, whereas substitution of glutamine for glutamate results in the complete loss of activity without apparent effect on metal binding. The roles of these residues (and the corresponding ones in α-MPP) in metal-ion binding and catalysis will be interesting to explore, particularly given that MPP, unlike insulinase and protease III, has a preference for Mn^{2+} (instead of Zn^{2+}) as its divalent cation (6). This difference may explain the divergence of this site in MPP from that in the other members of this proposed family of proteases.

Note. While this manuscript was being edited, we became aware of the recent paper by Kitada *et al.* (25) reporting the cloning of a partial cDNA for rat β-MPP, which they refer to as "P52," that is missing some 5′-coding and untranslated sequences.

We thank Dr. Grazia Isaya for helpful discussions. This work was supported by a grant from the National Institutes of Health (DK09527).

1. Pfanner, N. & Neupert, W. (1990) *Annu. Rev. Biochem.* **59,** 331–353.
2. Isaya, G., Kalousek, F., Fenton, W. A. & Rosenberg, L. E. (1991) *J. Cell Biol.* **113,** 65–76.
3. Kalousek, F., Hendrick, J. P. & Rosenberg, L. E. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7536–7540.
4. Hawlitschek, G., Schneider, N., Schmidt, B., Tropschung, M., Hartl, F.-U. & Neupert, W. (1988) *Cell* **53,** 795–806.
5. Yang, M., Jensen, R. E., Yaffe, M. P., Opplinger, W. & Schatz, G. (1988) *EMBO J.* **7,** 3857–3862.
6. Kleiber, J., Kalousek, F., Swaroop, M. & Rosenberg, L. E. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 7978–7982.
7. Ou, W.-J., Ito, A., Okazaki, H. & Omura, T. (1989) *EMBO J.* **8,** 2605–2612.
8. Rawlings, N. D. & Barrett, A. J. (1991) *Biochem. J.* **275,** 389–391.
9. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
10. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed., pp. 368–369.
11. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.
12. Frohman, M. A., Dush, M. K. & Martin, G. R. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8898–9002.
13. Hendrick, J. P., Hodges, P. E. & Rosenberg, L. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4056–4060.
14. Gavel, Y. & von Heijne, G. (1990) *Protein Eng.* **4,** 33–37.
15. Chou, P. W. & Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–148.
16. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
17. Pollock, R. A., Hartl, F.-U., Cheng, M. Y., Osterman, J., Horwich, A. & Neupert, W. (1988) *EMBO J.* **7,** 3493–3500.
18. Gencic, S., Schägger, H. & von Jagow, G. (1991) *Eur. J. Biochem.* **199,** 123–131.
19. Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345–362.
20. Schulte, U., Arretz, M., Schneider, H., Tropschung, M., Wachter, E., Neupert, W. & Weiss, H. (1989) *Nature (London)* **339,** 147–149.
21. Matthews, B. W. (1988) *Acc. Chem. Res.* **21,** 333–340.
22. Affholter, J. A., Fried, V. A. & Roth, R. A. (1988) *Science* **242,** 1415–1418.
23. Finch, P. W., Wilson, R. E., Brown, K., Hickson, I. D. & Emmerson, P. T. (1986) *Nucleic Acids Res.* **14,** 7695–7703.
24. Becker, A. B. & Roth, R. A. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 3835–3839.
25. Kitada, S., Niidome, T., Nagano, T., Ogishima, T. & Ito, A. (1993) *Biochem. Biophys. Res. Commun.* **190,** 289–293.