Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 6078–6082, June 1994
Biochemistry

N-Arginine dibasic convertase, a metalloendopeptidase as a prototype of a class of processing enzymes

ADRIAN R. PIEROTTI*, ANNIK PRAT, VALÉRIE CHESNEAU, FLORENCE GAUDOUX, ANNE-MARIE LESENEY, THIERRY FOULON, AND PAUL COHEN†

Laboratoire de Biochimie des Signaux Régulateurs Cellulaires et Moléculaires, Université Pierre et Marie Curie, Unité de Recherches Associée au Centre National de la Recherche Scientifique 1682, 96 Boulevard Raspail, 75006 Paris, France

Communicated by I. Robert Lehman, March 2, 1994

ABSTRACT *N*-Arg dibasic convertase is a metalloendopeptidase from rat brain cortex and testis that cleaves peptide substrates on the N terminus of Arg residues in dibasic stretches. By using both an oligonucleotide and antibodies to screen a rat testis cDNA library, a full-length cDNA was isolated. The sequence contains an open reading frame of 1161 codons corresponding to a protein of 133 kDa that exhibits 35% and 48% similarity with *Escherichia coli* protease III (pitrilysin, EC 3.4.99.44) and rat or human insulinase (EC 3.4.99.45), respectively. Moreover, the presence of the HXXEH amino acid signature (XX = FL) clearly classifies *N*-Arg dibasic convertase as a member of the pitrilysin family of zinc-metalloendopeptidases. In addition, a Cys residue that may be responsible for the thiol sensitivity of the insulinase and *N*-Arg dibasic convertase was proposed. The protein sequence contains a distinctive additional feature consisting of a stretch of 71 acidic amino acids. We hypothesize that this metalloendopeptidase may be a member of a distinct class of processing enzymes.

Limited proteolysis at basic amino acids is a general process of bioactivation of macromolecular precursors. Recent progress has suggested that these mechanisms may not be solely related to prohormone processing in the trans-most Golgi network (1–3) but may also intervene in modulation of peptide messengers at the exocellular level. The first gene encoding a processing enzyme, *KEX2*, was cloned by complementation experiments in *Saccharomyces cerevisiae* (4, 5). On the basis of its similarity to subtilisin and to furin (6), a human homolog of the Kex2 protein, a family of prohormone convertases (PCs) has been identified by PCR techniques. Their involvement in processing of a number of propeptides and proproteins was inferred mainly from cotransfection experiments (7–9), and for PC1, by the use of antisense mRNA (10).

Characterization of putative processing endoproteases by classical biochemical techniques has led to the identification of a number of activities, selective for basic residues in precursors, that belong to the four classes of proteases (metallo-, serine, aspartyl, and thiol enzymes; for review, see ref. 11). This suggested that more than one processing endoprotease family could exist (12). To our knowledge, none of these basic-residue-specific enzymes has been cloned.

Recently, a metalloendopeptidase was completely purified from rat testis and shown to cleave a number of peptide substrates on the N terminus of Arg residues in dibasic moieties (13). This enzyme was also present in rat brain cortex and its functional properties appeared undistinguishable from those of the somatostatin-28 convertase activity previously identified in this tissue (14, 15). By using microsequencing of tryptic fragments of the purified enzyme to design an oligonucleotide probe and polyclonal antibodies raised against the purified protein (13) to screen a rat testis cDNA library, the cDNA encoding this protease was cloned and its complete amino acid sequence was deduced.‡ The data indicate that *N*-Arg dibasic (NRD) convertase is a metalloendopeptidase related to protease III from *Escherichia coli* (pitrilysin, EC 3.4.99.44; ref. 16) and to insulin degrading enzyme (IDE or insulinase, EC 3.4.99.45; refs. 17–19). In addition, its sequence exhibits similarities with other proteins from the "pitrilysin" family [M16 from the nomenclature of Rawlings and Barrett (20)].

The *in vitro* highly restricted specificity of NRD convertase for Arg residues in dibasic processing signals and its belonging to the M16 family, which contains other enzymes involved in maturation, suggest that this enzyme is the prototype of a distinct family of processing endoproteases.

MATERIALS AND METHODS

Isolation and Characterization of cDNA Clones Encoding NRD Convertase. Four tryptic fragments were sequenced after digestion of previously purified NRD convertase following native PAGE. One fragment, GMQLIYLPPSPLLAE, was used to design the following degenerate inosine-containing oligonucleotide: 5′-GGIGGIAG(A/G)TAIATIA(G/A)(T/C)TGCATICC-3′. Two additional peptides were obtained by endolysine C treatment (13). A rat testis cDNA library in the λ ZapII vector containing 1.5×10^6 recombinants was purchased from Stratagene. Screening with the oligonucleotide probe was carried out as follows: 200 μl of an overnight culture of *E. coli* XL-1 Blue cells was mixed with $\approx 5 \times 10^4$ plaque-forming units of the library. After preadsorption of the phages for 15 min at 37°C, the cells were plated onto NZ-amine medium plates (21) and incubated at 37°C for 6 h. After chilling the plates for 1 h, duplicate nitrocellulose filters were blotted and the DNA was fixed by alkaline denaturation. Filters were baked in a vacuum oven (45 min, 80°C) before being probed. Filters were prehybridized for 1 h in hybridization solution [6× standard saline citrate (SSC)/5× Denhardt's solution/0.1% SDS] at 42°C, and then [γ-^{32}P]ATP end-labeled oligonucleotide (10^6 cpm/ml) was added and incubated overnight at 42°C. After washing in 0.1× SSC/1% SDS at 37°C for 30 min and autoradiography of the filters, positive phage plaques were isolated and rescreened to obtain single purified phage isolates. A similar aliquot of the library was plated onto NZ-amine medium plates for screening using polyclonal antibodies raised

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.

Abbreviations: NRD convertase, *N*-arginine dibasic convertase; IDE, insulin degrading enzyme; MPP, mitochondrial matrix-processing peptidase.

*Present address: Glasgow Caledonian University, Department of Biological Sciences, Cowcaddens Road, Glasgow G4 OBA, Scotland.
†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. L27124).

```
tgaggagtaggaggaggaggtgggcctggggagcgggatgtccagcgctacgacttgctggtga ATG CTG AGG AGA GTC GCG GTT GCT GCA GTC TTT GCC  100
                                                                  M   L   R   R   V   A   V   A   A   V   F   A    12
ACC GGG AGG AAG TTG CGG TGT GAG GCC GGG CGG GAC GTC ACG GCT GTG GGA CGA ATC GAA GCA CGG GGT CTG TGC GAA GAA TCA  184
 T   G   R   K   L   R   C   E   A   G   R   D   V   T   A   V   G   R   I   E   A   R   G   L   C   E   E   S    40
GCG AAA CCG TTT CCT ACG CTG ACC ATG CCT GGC AGA AAC AAG GCG AAG TCC ACC TGC AGT TGC CCG GAC CTG CAG CCT AAT GGA  268
 A   K   P   F   P   T   L   T   M   P   G   R   N   K   A   K   S   T   C   S   C   P   D   L   Q   P   N   G    68
CAG GAC TTG GGC GAG AGC GGC CGG GTT GCT CGT CTA GGA GCG GAT GAA TCT GAG GAG GAG GGA CGG TCT CTC AGT AAT GTC GGG  352
 Q   D   L   G   E   S   G   R   V   A   R   L   G   A   D   E   S   E   E   E   G   R   S   L   S   N   V   G    96
GAC CCT GAG ATC ATC AAG TCT CCC AGC GAT CCC AAG CAG TAC CGA TAC ATC AAA TTA CAG AAT GGC TTG CAG GCT CTT TTG ATT  436
 D   P   E   I   I   K   S   P   S   D   P   K   Q   Y   R   Y   I   K   L   Q   N   G   L   Q   A   L   L   I   124
TCA GAT CTA AGT AAT GTG GAG GGT AAA ACA GGA AAT GCA ACA GAT GAA GAG GAA GAG GAA GAA GAG GAA GAG GAG GAG GGG GAA  520
 S   D   L   S   N   V   E   G   K   T   G   N   A   T   D   E   E   E   E   E   E   E   E   E   E   E   G   E   152
GAG GAA GAG GAA GAG GAA GAA GAC GAT GAC GAT GAT GAT GAT GAA GAC TCT GGA GCT GAG ATA CAG GAT GAT GAT GAG GAA GGC  604
 E   E   E   E   E   E   E   D   D   D   D   D   D   D   E   D   S   G   A   E   I   Q   D   D   D   E   E   G   180
TTT GAC GAT GAA GAG GAA TTT GAT GAT GAT GAA CAT GAT GAT GAT GAT CTT GAT AAT GAG GAA AAT GAA CTG GAA GAA CTG GAA  688
 F   D   D   E   E   E   F   D   D   D   E   H   D   D   D   D   L   D   N   E   E   N   E   L   E   E   L   E   208
GAG CGG GTG GAA GCC AGA AAG AAA ACC ACT GAG AAA CAG TCT GCA GCG GCT CTG TGT GTT GGA GTT GGG AGC TTT GCT GAT CCA  772
 E   R   V   E   A   R   K   K   T   T   E   K   Q   S   A   A   A   L   C   V   G   V   G   S   F   A   D   P   236
GAT GAC CTG CCT GGG CTG GCA CAC TTT TTG GAG CAC ATG GTA TTC ATG GGT AGT TTG AAA TAC CCA GAT GAG AAT GGA TTT GAT  856
 D   D   L   P   G   L   A   H   F   L   E   H   M   V   F   M   G   S   L   K   Y   P   D   E   N   G   F   D   264
GCC TTC CTC AAG AAA CAT GGA GGT AGT GAT AAT GCC TCA ACC GAT TGT GAA CGC ACA GTC TTT CAG TTT GAT GTC CAG AGA AAG  940
 A   F   L   K   K   H   G   G   S   D   N   A   S   T   D   C   E   R   T   V   F   Q   F   D   V   Q   R   K   292
TAT TTC AAG GAA GCC CTG GAT AGA TGG GCC CAG TTC TTC ATC CAT CCA CTG ATG ATC AGA GAT GCA ATT GAC CGA GAG GTG GAA 1024
 Y   F   K   E   A   L   D   R   W   A   Q   F   F   I   H   P   L   M   I   R   D   A   I   D   R   E   V   E   320
GCT GTG GAC AGT GAG TAT CAG CTT GCA AGA CCT TCT GAT GCA AAC AGA AAA GAA ATG TTG TTT GGA AGT CTT GCT AGA CCT GGA 1108
 A   V   D   S   E   Y   Q   L   A   R   P   S   D   A   N   R   K   E   M   L   F   G   S   L   A   R   P   G   348
CAT CCT ATG GGG AAG TTT TTT TGG GGA AAT GCT GAG ACA CTC AAG CAT GAA CCA AAG AAG AAT AAT ATT GAT ACA CAT GCC AGA 1192
 H   P   M   G   K   F   F   W   G   N   A   E   T   L   K   H   E   P   K   K   N   N   I   D   T   H   A   R   376
CTG AGA GAG TTC TGG ATG CGT TAC TAC TCT GCT CAT TAC ATG ACC TTA GTA GTC CAG TCC AAA GAA ACC TTG GAC ACT TTG GAA 1276
 L   R   E   F   W   M   R   Y   Y   S   A   H   Y   M   T   L   V   V   Q   S   K   E   T   L   D   T   L   E   404
AAG TGG GTG ACG GAG ATC TTC TCT CAG ATA CCA AAC AAT GGG CTA CCT AAA CCA AAC TTT AGC CAT TTA ACG GAT CCT TTT GAC 1360
 K   W   V   T   E   I   F   S   Q   I   P   N   N   G   L   P   K   P   N   F   S   H   L   T   D   P   F   D   432
ACA CCA GCA TTT AAC AAA CTT TAT AGA GTT GTT CCA ATC AGA AAA ATC CAT GCT CTG ACC ATC ACG TGG GCC CTT CCT CCG CAG 1444
 T   P   A   F   N   K   L   Y   R   V   V   P   I   R   K   I   H   A   L   T   I   T   W   A   L   P   P   Q   460
CAG CAA CAC TAC AGG GTG AAG CCA CTT CAT TAT ATC TCC TGG CTT GTG GGG CAT GAA GGC AAA GGC AGC ATC CTT TCT TAC CTT 1528
 Q   Q   H   Y   R   V   K   P   L   H   Y   I   S   W   L   V   G   H   E   G   K   G   S   I   L   S   Y   L   488
AGA AAA AAG TGC TGG GCT CTT GCG TTG TTT GGT GGA AAT GGT GAG ACA GGT TTT GAG CAA AAT TCT ACC TAC TCA GTG TTC AGC 1612
 R   K   K   C   W   A   L   A   L   F   G   G   N   G   E   T   G   F   E   Q   N   S   T   Y   S   V   F   S   516
ATT TCT ATC ACT CTG ACT GAT GAG GGG TAT GAG CAT TTT TAT GAG GTT GCT CAC ACT GTC TTT CAG TAT TTA AAA ATG CTG CAG 1696
 I   S   I   T   L   T   D   E   G   Y   E   H   F   Y   E   V   A   H   T   V   F   Q   Y   L   K   M   L   Q   544
AAA CTA GGT CCG GAG AAA AGA GTT TTT GAA GAG ATT CAG AAA ATT GAG GAT AAT GAA TTT CAT TAC CAA GAA CAG ACA GAT CCA 1780
 K   L   G   P   E   K   R   V   F   E   E   I   Q   K   I   E   D   N   E   F   H   Y   Q   E   Q   T   D   P   572
GTT GAG TAC GTG GAG AAC ATG TGT GAG AAC ATG CAG CTG TAC CCA CGT CAG GAC TTT CTC ACT GGA GAT CAG CTT CTG TTT GAG 1864
 V   E   Y   V   E   N   M   C   E   N   M   Q   L   Y   P   R   Q   D   F   L   T   G   D   Q   L   L   F   E   600
TAC AAG CCA GAA GTC ATT GCC GAG GCC CTT AAT CAG CTA GTT CCT CAA AAG GCA AAT CTC GTG CTT CTG TCT GGT GCT AAC GAG 1948
 Y   K   P   E   V   I   A   E   A   L   N   Q   L   V   P   Q   K   A   N   L   V   L   L   S   G   A   N   E   628
GGA AGA TGT GAC CTC AAG GAG AAG TGG TTT GGG ACT CAG TAC AGC ATA GAA GAT ATT GAA AAC TCT TGG ACT GAA CTG TGG AAG 2032
 G   R   C   D   L   K   E   K   W   F   G   T   Q   Y   S   I   E   D   I   E   N   S   W   T   E   L   W   K   656
AGC AAT TTT GAC TTA AAT TCA GAC CTT CAT CTA CCA GCT GAA AAC AAA TAC ATA GCC ACG GAC TTT ACA TTG AAG GCT TTT GAT 2116
 S   N   F   D   L   N   S   D   L   H   L   P   A   E   N   K   Y   I   A   T   D   F   T   L   K   A   F   D   684
TGC CCT GAA ACT GAA TAC CCT GCT AAA ATT GTG AAT ACC CCA CAA GGT TGC CTG TGG TAT AAG AAA GAC AAC AAA TTC AAA ATC 2200
 C   P   E   T   E   Y   P   A   K   I   V   N   T   P   Q   G   C   L   W   Y   K   K   D   N   K   F   K   I   712
CCC AAA GCC TAT ATA CGA TTT CAT CTG ATC TCA CCT TTG ATA CAG AAA TCT GCA GCA AAT GTT GTC CTC TTC GAC ATC TTC GTC 2284
 P   K   A   Y   I   R   F   H   L   I   S   P   L   I   Q   K   S   A   A   N   V   V   L   F   D   I   F   V   740
AAC ATT CTT ACT CAC AAC CTT GCT GAG CCA GCC TAT GAA GCC GAT GTG GCC CAG CTG GAG TAT AAA CTG GTA GCC GGA GAG CAC 2368
 N   I   L   T   H   N   L   A   E   P   A   Y   E   A   D   V   A   Q   L   E   Y   K   L   V   A   G   E   H   768
GGT TTA ATC ATC CGA GTG AAG GGA TTC AAC CAC AAA CTA CCT CTG CTC TTC CAG CTC ATC ATT GAC TAC CTG ACT GAG TTC AGC 2452
 G   L   I   I   R   V   K   G   F   N   H   K   L   P   L   L   F   Q   L   I   I   D   Y   L   T   E   F   S   796
TCC ACG CCC GCT GTC TTC ACC ATG ATA ACT GAG CAG CTG AAG AAG ACC TAC TTC AAC ATC CTC ATC AAG CCT GAG ACT CTG GCC 2536
 S   T   P   A   V   F   T   M   I   T   E   Q   L   K   K   T   Y   F   N   I   L   I   K   P   E   T   L   A   824
AAA GAT GTG CGG CTT TTA ATT CTG GAA TAT TCC CGC TGG TCT ATG ATT GAC AAG TAC CGA GCT TTG ATG GAT GGC CTT TCC CTT 2620
 K   D   V   R   L   L   I   L   E   Y   S   R   W   S   M   I   D   K   Y   R   A   L   M   D   G   L   S   L   852
GAG TCT CTG CTA AAC TTT GTC AAG GAT TTT AAA TCC CAA CTC TTT GTT GAG GGC CTG GTG CAA GGG AAT GTC ACC AGC ACG GAA 2704
 E   S   L   L   N   F   V   K   D   F   K   S   Q   L   F   V   E   G   L   V   Q   G   N   V   T   S   T   E   880
TCG ATG GAT TTC CTA AGA TAC GTT GTT GAC AAG CTG AAC TTC GTG CCC CTG GAG CGG GAG ATG CCC GTG CAG TTC CAG GTG GTG 2788
 S   M   D   F   L   R   Y   V   V   D   K   L   N   F   V   P   L   E   R   E   M   P   V   Q   F   Q   V   V   908
GAG CTG CCG AGC GGC CAC CAC CTG TGC AAA GTC AGA GCC CTG AAT AAG GGG GAC GCC AAC TCT GAA GTC ACT GTG TAC TAC CAG 2872
 E   L   P   S   G   H   H   L   C   K   V   R   A   L   N   K   G   D   A   N   S   E   V   T   V   Y   Y   Q   936
TCA GGT ACA AGG AGT CTG AGA GAA TAC ACG CTC ATG GAG CTG CTT GTG ATG CAC ATG GAG GAG CCC TGC TTT GAC TTC CTT CGA 2956
 S   G   T   R   S   L   R   E   Y   T   L   M   E   L   L   V   M   H   M   E   E   P   C   F   D   F   L   R   964
ACC AAG CAG ACC CTT GGG TAC CAC GTC TAC CCG ACC TGT AGG AAC ACA TCT GGC ATT CTA GGA TTT TCT GTA ACT GTT GGG ACT 3040
 T   K   Q   T   L   G   Y   H   V   Y   P   T   C   R   N   T   S   G   I   L   G   F   S   V   T   V   G   T   992
CAG GCA ACC AAA TAC AAC TCT GAA ACT GTT GAC AAG AAG ATA GAA GAA TTT CTT TCC AGC TTT GAG GAG AAG ATT GAG AAC CTC 3124
 Q   A   T   K   Y   N   S   E   T   V   D   K   K   I   E   E   F   L   S   S   F   E   E   K   I   E   N   L  1020
ACA GAG GAT GCA TTC AAT ACT CAG GTC ACA GCT CTG ATC AAG TTG AAA GAG TGT GAG GAC ACC CAC CTC GGG GAG GAG GTG GAC 3208
 T   E   D   A   F   N   T   Q   V   T   A   L   I   K   L   K   E   C   E   D   T   H   L   G   E   E   V   D  1048
AGG AAC TGG AAT GAA GTG GTG ACA CAG CAG TAT CTC TTT GAC CGC CTT GCC CAT GAG ATT GAA GCG CTG AAG TCT TTC TCA AAG 3292
 R   N   W   N   E   V   V   T   Q   Q   Y   L   F   D   R   L   A   H   E   I   E   A   L   K   S   F   S   K  1076
TCA GAC CTG GTC AGT TGG TTC AAG GCT CAC AGA GGA CCT GGG AGT AAA ATG CTC AGC GTT CAT GTT GTT GGA TAT GGG AAG TAT 3376
 S   D   L   V   S   W   F   K   A   H   R   G   P   G   S   K   M   L   S   V   H   V   V   G   Y   G   K   Y  1104
GAA CTG GAA GAG GAC GGC GCT CCC GTT TGT GAG GAT CCG AAT TCT CGT GAA GGG ATG CAG CTA ATC TAC CTG CCA CCC TCT CCT 3460
 E   L   E   E   D   G   A   P   V   C   E   D   P   N   S   R   E   G   M   Q   L   I   Y   L   P   P   S   P  1132
CTC CTG GCA GAG TCC ACC ACC CCC ATT ACT GAC ATC AGG GCC TTC ACA GCG ACG CTC AGC CTC TTC CCC TAC CAT AAG ATA GTC 3544
 L   L   A   E   S   T   T   P   I   T   D   I   R   A   F   T   A   T   L   S   L   F   P   Y   H   K   I   V  1160
AAA taa actgcagtctcactggcctgaagcaacttgt                                                                          3581
 K   *                                                                                                           1161
```

FIG. 1. Nucleotide and deduced amino acid sequences of NRD convertase. The putative signal peptide is indicated in italic type and the two possible cleavage sites are shown in boldface type. The boldface and underlined amino acid sequences correspond to trypsin or endolysine C peptide fragments obtained by digestion of the purified protein. The C-terminal peptide used to design the nucleotide probe is double underlined. Putative glycosylation sites are underlined; the acidic stretch is underlined with dots. The zinc-binding site is boldface and double underlined and the conserved Cys residue in IDEs is located at position 959 (boldface type underlined). The putative polyadenylylation signal is shown double underlined.

against purified rat testis NRD convertase (13). The plates were incubated at 42°C for 3 h, transferred to 37°C for an additional hour, and then overlaid with a nitrocellulose filter previously saturated with 10 mM isopropyl β-D-thiogalactoside. The plates and filters were then incubated for 3 h at 37°C. The filters were then removed, washed briefly in TBST (10 mM Tris·HCl, pH 8/150 mM NaCl/0.05% Tween 20), and stored in a fresh solution of TBST overnight. Positive phage plaques were detected using an alkaline phosphatase second antibody system as described using a 1:2000 dilution of primary antibody (22). Insert DNA was excised from all positive clones, isolated by immuno- or oligonucleotide screening, according to the manufacturer's recommendations. cDNA inserts were characterized by restriction mapping and Southern blot analysis.

Sequencing of the NRD Convertase cDNA. After analysis of the inserts, suitable restriction fragments were cloned in both orientations in M13mp18 and tg131 (M13 phage derivatives). The sequence was determined by the dideoxynucleotide chain-termination method (23) using T7 polymerase (Pharmacia). Both strands were entirely sequenced.

Computer Analysis. Computer sequence analysis was performed using the Base Informatique sur les Séquences de Biomolécules pour les Chercheurs Européens (BISANCE,

Paris, ref. 24) and DNA Strider (Commissariàt á l'Energie Atomique, Saclay, France). Homology searches were performed using the FASTA algorithm with a KTPL value set to 2 (25) and the GenPro databank (version 79). Sequences presenting an optimized score >100 were retained. Initial multiple alignments in Fig. 4 were carried out with the CLUSTAL program with a KTPL value set to 1 (26); these were then refined visually to minimize the number of gaps.

Northern Blot. Total RNA was prepared from rat testis and brain and electrophoresed in a 1% agarose/formaldehyde gel (21). RNA markers were purchased from GIBCO/BRL. The filter was then hybridized with a 960-bp *Pst* I restriction fragment of clone 22, labeled by oligonucleotide random priming (Boehringer Mannheim) using [α-^{32}P]dCTP (Amersham; ref. 21).

***In Situ* Hybridization.** Male Wistar rats (180–200 g) were perfused with freshly prepared 4% (wt/vol) paraformaldehyde in 0.1 M sodium phosphate-buffered saline (pH 7.4). Testis were dissected and immersed in paraformaldehyde overnight at 4°C, embedded in paraffin blocks, and sectioned at a 7.5-μm thickness. The sense and antisense probes used were complementary RNAs obtained by *in vitro* transcription, using T3 and T7 promotors, of a 960-bp fragment of rat NRD convertase cDNA in pBS SK$^+$ plasmid (Stratagene). Probes were labeled with ^{35}S-labeled UTP (1409 Ci/mmol; 1 Ci = 37 GBq; Amersham) to a specific activity of 1–4 × 10^8 cpm/μg and separated from unincorporated nucleotides on G_{50} spin columns (Pharmacia). *In situ* hybridization methods are as described (27). Slides were dipped in liquid emulsion (K5, Ilford) and developed after a 3-day exposure.

RESULTS

A total of 3 × 10^6 clones from a rat testis cDNA library were screened. After analysis, 32 of the 40 selected clones appeared to correspond to NRD convertase cDNA and in addition represented the only class of clones found by both oligonucleotide and immunoscreening. The entire sequence is shown in Fig. 1. It can be divided into a 5′ untranslated region of 64 nt, an open reading frame of 1161 codons, and a 3′ untranslated region of 34 nt. Interestingly, the putative polyadenylylation signal appears to contain the last sense codon and the TAA stop codon (Fig. 1). The polyadenylylation site in clone 22, the only full-length cDNA obtained, is located 23 bases downstream from the stop codon. However, a number of clones showed an alternative polyadenylylation site 11 bases downstream of the first site. Two in-phase Met codons are present at positions 65 and 209, corresponding to residues 1 and 49, respectively. However, both the alignment of the N-terminal domains of homologous proteins (data not shown), namely, the mitochondrial matrix-processing peptidases (MPPs), and the absence of a putative signal peptide downstream from the second Met support the start of the open reading frame at position 65.

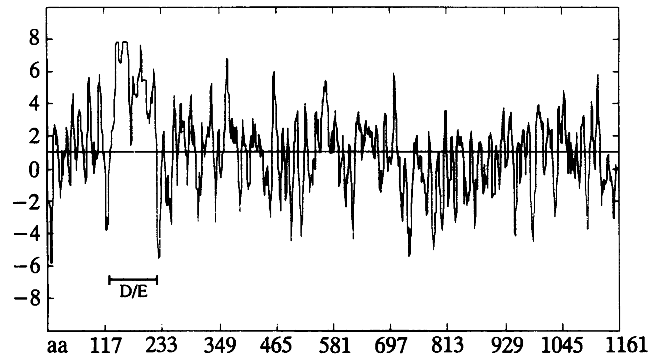

FIG. 2. Kyte–Doolittle diagram of NRD convertase. The acidic stretch (D/E) is indicated by a bar.

The deduced amino acid sequence encodes a 1161-residue protein with a calculated molecular mass of 133 kDa, consistent with the molecular mass of the largest form of the purified enzyme (13). All six oligopeptide sequences (Fig. 1) obtained from tryptic or endolysine C digestion of the purified enzyme could be found in the deduced protein sequence. Analysis of the N-terminal part of the amino acid sequence reveals a potential signal peptide whose cleavage site is predicted to occur after Cys-19 or Ala-21 (28). The corresponding hydrophobic stretch clearly appears in the Kyte–Doolittle hydropathy diagram (ref. 29; Fig. 2), whereas no hydrophobic transmembrane domain could be detected in the NRD convertase sequence.

```
C                  R.+.L.NG+..LL+SD.........................T+K.S+AAL.V.+GS.+DP.++.GL+HF.EHM+F+G+.KYP.EN.+..+L..H.GS.NA.T....T.+
rNRD          111-RYIKLQNGLQALLISDLSNVEGKTGNAT(D/E)RVEARKK-TTEKQSAAALCVGVGSFADPDDLPGLAHFLEHMVFMGSLKYPDENGFDAFLKKHGGSDNASTDCERTVF-285
rIDE           65-RGLELANGIKVLLISD-----------------------PTTDK-SSAALDVHIGSLSDPPNIPGLSHFCEHMLFLGTKKYPKENEYSQFLSEHAGSSNAFTSGEHTNY-149
hIDE           65-RGLELANGIKVLLMSD-----------------------PTTDK-SSAALDVHIGSLSDPPNIAGLSHFCEHMLFLGTKKYPKENEYSQFLSEHAGSSNAFTSGEHTNY-149
dIDE           38-RGLQLENGLKVLLISD-----------------------PNTDV-SAAALSVQVGHMSDPTNLPGLAHFCEHMLFLGTEKYPHENGYTTYLSQSGGSSNAATYPLMTKY-122
E.c.proIII     45-QAIRLDNGMVVLLVSD-----------------------PQAVK-SLSALVVPVGSLEDPEAYQGLAHYLEHMSLMGSKKYPQADSLAEYLKMHGGSHNASTAPYRTAF-129
------------------------------------------------------------------------------------------------------------------------------
K.p.pqqF        6-RTVTLPGGLQATLVHQ-----------------------PQADR-AAALARVAAGSHHEPSRFPGLAHLLEHLLFYGGERYQDDDRLMGWVQRQGGSVNATTLARHSAF- 90
B.s.PP          1-.............................................GISHFLEHMFFKGTSTKSAR-EIAESFDRIGGQVNAFTSKEYTCY- 44
rαMPP          67-KVTTLDNGLRVA--SQ-----------------------NKFGQFCTLGILINSGSRYEAKYLSGIAHFLEKLAFSSTARFDSKDEILLTLEKHGGICDCQTSRDTTMY-150
C                  +.+.L.NG+.....S+.........................+...+.+.+.+..GS..+P....G++H+.E++.F.G+.++...+.....+....G..NA.T....T.+

C                 .F+V.......ALDR AQFF..PL......+RE.+AV+SE.......D..R...+......P.H...KF....GN..TL...P.....ID....L..F...YYS+..M..
rNRD              QFDVQRKYFKEALDRWAQFFIHPLMIRDAIDREVEAVDSEYQLARPSDANRKEMLFGSLARPGHPMGKFFW--GNAETLKHEPKK-NNIDTHARLREFWMRYYSAHYMTL-392
rIDE              YFDVSHEHLEGALDRFAQFFLCPLFDASCKDREVNAVDSEHEKNVMNDAWRLFQLEKATGNPKHPFSKFGT--GNKYTLETRPNQ-EGIDVREELLKFHSTYYSSNLMAI-256
hIDE              YFDVSHEHLEGALDRFAQFFLCPLFDESCKDREVNAVDSEHEKNVMNDAWRLFQLEKATGNPKHPFSKFGT--GNKYTLETRPNQ-EGIDVRQELLKFHSAYYSSNLMAV-256
dIDE              HFHVAPDKLDGALDRFAQFFIAPLFTPSATEREINAVNSEHEKNLPSDLWRIKQVNRHLAKPDHAYSKFGS--GNKTTLSEIPKS-KNIDVRDELLKFHKQWYSANIMCL-229
E.c.proIII        YLEVENDALPGAVDRLADAIAEPLLDKKYAERERNAVNAELTMARTRDGMRMAQVSAETINPAHPGSKFSG--GNLETLSDKP----GNPVQQALKDFHEKYYSANLMKA-234
------------------------------------------------------------------------------------------------------------------------------
K.p.pqqF          FFEVAADALADGVARLQEMLQAPLLLREDIQREVAVIDAEYRLIQQHEPSRREAAVRHAAS--APAAFRRFQVGSADALAG-----DLAALQAALGDFHRTHYVARRMQL-193
B.s.PP            YAKVLDEHANYALDVLADMFFHSTFDENELKKEKNVVYEEIKMY---E--DAPDDIVHDLL--SKATYGNHSLGYPILGT---EETLASFNGDSLRQYMHDYYTPDRVVI-144
rαMPP             AVSADSKGLDTVVGLLADVVLHPRLTDEEIEMTRMAVQFELEDL---NMRPDPEPLLTEMI--HEAAFRENTVGLHRFC---PVENIGKIDREVLHSYLKNYYTPDRMVL-252
C                 ...V.......+...A+....P.......++E...V+.E.......+.................+......G...................L..+....Y+...+.+
```

FIG. 3. Multiple alignment of the conserved domain of NRD convertase and related proteins. rNRD, rat NRD convertase; rIDE, hIDE, and dIDE, rat, human, and *Drosophila* IDE, respectively; *E.c.*proIII, protease III from *E. coli*; *K.p.*pqqF, *Klebsiella pneumoniae* open reading frame F of the pyrroloquinoline quinone operon; *B.s.*PP, *Bacillus subtilis* processing protease; rαMPP, rat MPP α subunit. Consensus sequences of NRD convertase with IDEs and *E. coli* protease III or of all the sequences are indicated above and below the alignment, respectively. Strictly conserved positions are in boldface type, whereas plain text indicates only one exception to the consensus. +, Conserved positions according to the following rules: (V = L = I = M), (D = E = N = Q), (T = A = S), (K = R = H), (F = Y), (C), (G), (P), (W). The two short anchor regions of the acidic stretch (D/E) are underlined with dots.

	NRD CONV.	hIDE	rIDE	dIDE	pro III
NRD CONV.		40 53	41 53	44 53	42 48
hIDE	31 47		97 99	59 74	43 75
rIDE	32 48	95 nd		60 74	42 75
dIDE	29 44	48 nd	47 nd		42 55
pro III	22 35	26 nd	27 nd	28 nd	

FIG. 4. This matrix of homology for NRD convertase (CONV.), human (h), rat (r), and *Drosophila* (d) IDEs, and protease (pro) III shows the percentages of identity (upper line) or similarity (lower line) considering either the entire sequence (left side of the diagonal) or the alignment presented in Fig. 5 (right side of the diagonal) (see Fig. 5 for similarity rules). nd, Not determined.

The protein sequence presents two major features: (*i*) a 71-residue acidic stretch (aa 139–209) rich in Asp and Glu (79%; Figs. 1 and 2) and (*ii*) the presence of the zinc binding motif, HXXEH, residues 244–248 (Fig. 1), that is part of a conserved motif present in several other zinc-metalloendopeptidases (see below; for review, see ref. 30).

A FASTA search in the GenPro databank (version 79) was performed using the NRD convertase amino acid sequence deleted of its acidic stretch (71 residues) and identified seven significantly homologous proteins: the human, rat, and *Drosophila* IDEs (17–19), protease III from *E. coli* (16), the open reading frame F of the pyrroloquinoline quinone operon of *Klebsellia pneumoniae* (pqqF; ref. 31), the rat MPP α subunit (32), and the *Bacillus subtilis* processing protease, a putative gene flanking the diaminopimelate operon (33). This search for related protein sequences did not select 12 additional species that could be classified in the M16 family, among which a majority exhibit the insulinase signature (30). A multiple alignment of NRD convertase with the selected protein sequences in the region containing the zinc binding motif (Fig. 3) shows a separation into two subgroups on the basis of their degree of similarity with NRD convertase. The first one contains IDEs and protease III (44 and 40% identity); the second one consists of two proteins, pqqF and *B. subtilis* processing protease (27 and 21% identity), whose functions are not as yet defined, and rat α-MPP (17% identity). The relative percentages of identity/similarity of the proteins belonging to the first group are presented in Fig. 4, considering either their entire sequence or the restricted alignment in Fig. 3. Interestingly, whereas the *Drosophila* IDE presents only 47–48% identity with its mammalian counterparts, the 3 IDEs share an equivalent identity with NRD convertase. Furthermore, protease III, which shows the weakest global homology with NRD convertase, appears

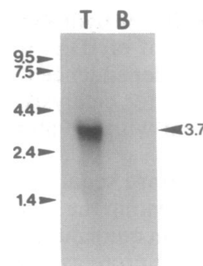

FIG. 5. Northern blot analysis of 26 μg of total testis (lane T) and brain (lane B) RNA. RNA markers, expressed in kilobases, are indicated by arrowheads on the left and the estimated size of NRD convertase mRNA is shown by an arrowhead on the right side.

to share the same homology than the IDEs in the zinc binding domain. The consensus sequence (see Fig. 3) reveals that the insulinase signature, G(X)$_9$GX(S/T/A)H(L/I/V/M/F/Y)-X(D/E)(H/R/K)(L/I/V/M/F/Y)X(L/I/V/M/F/Y)X(G/S/T)(G/S/T) (30), which is an extended zinc binding motif, is perfectly conserved in NRD convertase.

From the multiple alignment, it appears that the NRD convertase acidic stretch is flanked by two short anchors of 12 and 7 residues and, surprisingly, is inserted in the most conserved region between all the members of the family, $\approx$25 residues upstream the HXXEH zinc binding site.

A Northern blot analysis of both brain and testis total RNA is shown in Fig. 5. The NRD convertase mRNA was estimated to be 3.7 kb in testis, in good agreement with the length of clone 22, but remained undetectable in brain. *In situ* hybridization studies showed that NRD convertase mRNA is abundant in all round and elongating seminiferous tubules within adult testis. Two labeling patterns are observed: a very strong one that is principally concentrated all around the lumen of the tubules (Fig. 6*B*) and another one in which a large number of silver grains is uniformly distributed over the tubule (Fig. 6*A*). In contrast no signal could be observed with the control sense probe (Fig. 6*C*).

DISCUSSION

We report here the cloning and sequencing of a cDNA encoding a dibasic selective metalloendopeptidase that cleaves on the N terminus of Arg residues. The presence in the deduced amino acid sequence of the zinc-binding motif confirms the metalloenzyme character of NRD convertase. In addition to their sequence homology, IDEs and NRD convertase share several biochemical characteristics such as their capacity to cleave peptide bonds on the N terminus of a limited number of residues and their sensitivity to some sulfhydryl reagents (34). The alignment of NRD convertase and the three IDE sequences identified a unique Cys residue (Cys-959, Fig. 1) common to the four proteins (Cys-819 in rat IDE and human IDE; Cys-793 in *Drosophila* IDE). This amino acid is absent in protease III, which is not sensitive to sulfhydryl reagents (35). Site-directed mutagenesis and analysis of the mutant convertase would indicate whether this

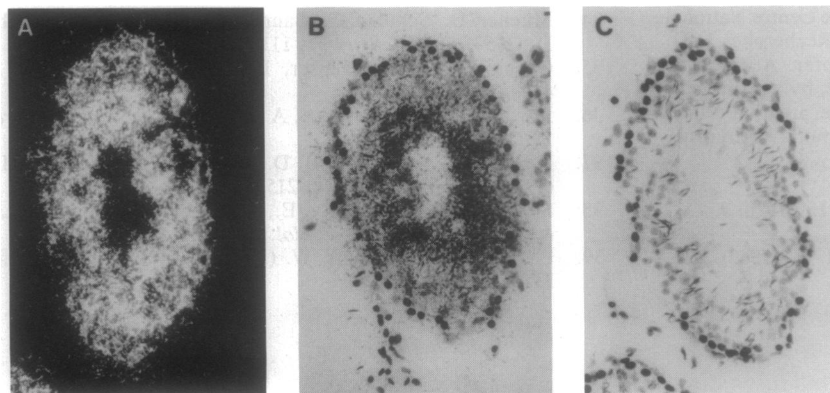

FIG. 6. Localization of NRD convertase mRNA in adult testis by *in situ* hybridization. Dark-field (*A*) and bright-field (*B* and *C*) cellular distribution of silver grains using ^{35}S-labeled antisense (*A* and *B*) and sense (negative control) (*C*) NRD convertase complementary RNA probes and nuclear emulsion coating. ($\times$400.)

residue is involved in either substrate binding or hydrolysis. The conserved Cys of the zinc-binding motif of IDEs was initially proposed to account for their thiol dependence (36) but was recently shown to be not essential for catalysis or inhibitor sensitivity (37), a result consistent with the above hypothesis.

A remarkable feature of the NRD convertase sequence is the presence of a 71-residue acidic segment within the region of high homology. Such stretches are found in quite different proteins such as nucleolins, calcium binding proteins, transcription factors, and even the carboxypeptidase Kex1 from *S. cerevisiae* but their function remains as yet undefined. For NRD convertase, the proximity of this highly charged region with the putative catalytic domain of the enzyme is compatible with a participation in substrate recognition and binding. Alternatively, this motif may play a role in the appropriate routing and targeting of the enzyme to a given subcellular compartment. No transmembrane domain could be predicted by hydropathy analysis of the NRD convertase sequence and a role of this acidic segment in the interaction with membranes cannot be excluded.

Northern blot analysis did not reveal any signal in brain extracts. This could be accounted for by the fact that RNA was prepared from whole brain leading to a dilution of specific NRD convertase mRNA. Indeed, *in situ* hybridization studies show an intense signal in restricted areas of the rat brain (data not shown).

The NRD convertase gene is highly expressed in seminiferous tubules of the adult rat. Endopeptidase mRNA is present at early stages of spermatogenesis, whereas no enzyme immunoreactivity is observed in spermatogonia and spermatocytes (13). A more intense signal is detected in the late stages of spermatogenesis possibly as a result of transcript accumulation throughout spermatogenesis or of a transcription rise or both. Further studies on isolated seminiferous tubule cell types will determine the corresponding stages of NRD convertase gene expression and their physiological significance.

Although its physiological substrates have not yet been identified, NRD convertase was shown to exhibit catalytic properties consistent with a role in processing dibasic sites in propeptides and proproteins (13–15). Furthermore, the presence at the N terminus of a 19- or 21-amino acid sequence compatible with the known properties of signal peptides may account for its sorting to the secretory pathways in the cells that produce the enzyme. In addition, despite its higher homology to IDEs and protease III, which exhibit an apparently lower specificity (34, 35, 38), NRD convertase is also distantly related to MPPs, which have been implicated in mitochondrial processing (39). We therefore suggest that the NRD convertase may constitute the prototype of a distinct class of processing enzymes.

We gratefully acknowledge M. Rivas and C. Carles for microsequencing tryptic fragments. This work was supported in part by funds from the Université Pierre et Marie Curie, the Centre National de la Recherche Scientifique (CNRS to Unité de Recherches 1682) and by program BioAvenir of Rhône-Poulenc-Rorer. A.R.P. was supported by Fondation de la Recherche Médicale, by a position of "chercheur associé" from the CNRS and by the program "Plan Science" from the European Economical Community (Contract SCI000407).

1. Sossin, W. S., Fisher, J. M. & Scheller, R. H. (1990) *J. Cell Biol.* **110**, 1–12.
2. Schnabel, E., Mains, R. E. & Farquhar, M. G. (1989) *Mol. Endocrinol.* **3**, 1223–1235.
3. Lepage-Lezin, A., Joseph-Bravo, P., Devilliers, G., Benedetti, L., Launay, J.-M., Gomez, S. & Cohen, P. (1991) *J. Biol. Chem.* **266**, 1679–1688.
4. Julius, D., Brake, A., Blair, L., Kunisawa, R. & Thorner, J. (1984) *Cell* **37**, 1075–1089.
5. Mizuno, K., Nakamura, T., Ohshima, T., Tanaka, S. & Matsuo, H. (1988) *Biochem. Biophys. Res. Commun.* **156**, 246–254.
6. Fuller, R. S., Brake, A. J. & Thorner, J. (1989) *Science* **246**, 482–486.
7. Van de Ven, W. J. M., Van Duijnhoven, J. L. P. & Roebroek, A. J. M. (1993) *Crit. Rev. Oncogen.* **4**, 115–136.
8. Seidah, N. G. & Chrétien, M. (1992) *Trends Endocrinol. Metab.* **3**, 133–140.
9. Steiner, D. F., Smeekens, S. P., Ohagi, S. & Chan, S. J. (1992) *J. Biol. Chem.* **267**, 23435–23438.
10. Bloomquist, B. T., Eipper, B. & Mains, R. E. (1991) *Mol. Endocrinol.* **5**, 2014–2024.
11. Darby, N. & Smyth, D. (1990) *Biosci. Rep.* **10**, 1–13.
12. Rholam, M., Nicolas, P. & Cohen, P. (1986) *FEBS Lett.* **207**, 1–6.
13. Chesneau, V., Pierotti, A. R., Barré, N., Créminon, C., Tougard, C. & Cohen, P. (1994) *J. Biol. Chem.* **269**, 2056–2061.
14. Gluschankof, P., Gomez, S., Morel, A. & Cohen, P. (1987) *J. Biol. Chem.* **262**, 9615–9620.
15. Gomez, S., Gluschankof, P., Lepage, A. & Cohen, P. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5468–5472.
16. Finch, P. W., Wilson, R. E., Brown, K., Hickson, I. D. & Emmerson, P. T. (1986) *Nucleic Acids Res.* **14**, 7695–7703.
17. Affholter, J. A., Fried, V. A. & Roth, R. A. (1988) *Science* **242**, 1415–1418.
18. Kuo, W.-L., Gehn, B. D. & Rosner, M. R. (1990) *Mol. Endocrinol.* **4**, 1580–1591.
19. Baumeister, H., Müller, D., Rehbein, M. & Richter, D. (1993) *FEBS Lett.* **317**, 250–254.
20. Rawlings, N. D. & Barrett, A. J. (1993) *Biochem. J.* **290**, 205–218.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
22. Pierotti, A., Dong, K.-W., Glucksman, M. J., Orlowski, M. & Roberts, J. L. (1990) *Biochemistry* **29**, 10323–10329.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
24. Dessen, P., Fondrat, C., Valencien, C. & Mugnier, C. (1990) *Comp. Appl. Biosci.* **6**, 355–356.
25. Lipman, D. J. & Pearson, W. R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
26. Higgin, D. G. & Sharp, P. M. (1988) *Gene* **73**, 237–244.
27. Gaudoux, F., Boileau, G. & Crine, P. (1993) *J. Neurosci. Res.* **34**, 426–433.
28. von Heijne, G. (1986) *Nucleic Acids Res.* **14**, 4683–4690.
29. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
30. Rawlings, N. D. & Barrett, A. J. (1991) *Biochem. J.* **275**, 389–391.
31. Meulenberg, J. J. M., Sellink, E., Riegman, N. H. & Postma, P. W. (1992) *Mol. Gen. Genet.* **232**, 284–294.
32. Kleiber, J., Kalousek, F., Swaroop, M. & Rosenberg, L. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 7978–7982.
33. Chen, N.-Y., Jiang, S.-Q., Klein, D. A. & Paulus, H. (1993) *J. Biol. Chem.* **268**, 9448–9465.
34. Müeller, D., Schulze, C., Baumeister, H., Buck, F. & Richter, D. (1993) *Biochemistry* **31**, 11138–11143.
35. Cheng, Y.-S. E. & Zipser, D. (1979) *J. Biol. Chem.* **254**, 4698–4706.
36. Rawlings, N. D. & Barrett, A. J. (1991) *Biochem. Soc. Trans.* **19**, 289S.
37. Perlman, R. K., Gehm, B. D., Kuo, W.-L. & Rosner, M. R. (1993) *J. Biol. Chem.* **268**, 21538–21544.
38. Stentz, F. B., Kitabchi, A. E., Schilling, J. W., Schronk, L. R. & Seyer, J. M. (1989) *J. Biol. Chem.* **264**, 20275–20282.
39. Pfanner, N. & Neupert, W. (1990) *Annu. Rev. Biochem.* **59**, 331–353.