# Domain and Genomic Sequence Analysis of Bdellin-KL, a Leech-Derived Trypsin-Plasmin Inhibitor[1]

**Young Hye Kim, Jung Gyoon Choi, Gyun Min Lee, and Ke Won Kang[2]**

*Bioresource Laboratory, Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejon, 305-701, Korea*

Received April 20, 2001; accepted June 25, 2001

**Bdellin-KL is a trypsin-plasmin inhibitor from *Hirudo nipponia*, whose N-terminal sequence was identified as a non-classical Kazal-type. A cDNA clone encoding the inhibitor was isolated by reverse transcription-PCR and 5′ rapid amplification of cDNA ends. The cDNA showed an open reading frame of 155 amino acids comprising one signal peptide and two separated domains. The C-terminal domain consists of distinct internal repeats, including HHEE and HHDD. The bdellin-KL sequence, from the constructed genomic library of Korean leech, was determined for the 2109 bases comprising the open reading frame and flanking regions (3′ and 5′). The promoter region contains potential regulatory sequence motifs, including TATA, CAAT, and GC boxes. To characterize the properties of each domain, an N-terminal fragment was prepared by limited proteolysis of the intact protein. The inhibitory activity of the region was as potent as that of the intact protein. This suggests that the compact domain plays an important part in the inhibitory action of bdellin-KL. The C-terminal domain was revealed to have binding affinity to ions such as $Ca^{2+}$, $Zn^{2+}$, $Fe^{3+}$, and $Fe^{2+}$ without an influence on the inhibitory activity. This study demonstrates that bdellin-KL may be a novel bifunctional protein with two distinct domains.**

**Key words: bdellin, *Hirudo nipponia*, leech genomic library, plasmin inhibitor, trypsin inhibitor.**

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

Leeches are known to possess various protease inhibitors whose biological functions are not clearly understood, but which have been shown to act as effective inhibitors of serine proteases. Recently, a trypsin-plasmin inhibitor with potent activity within a wide temperature and pH range was purified from *Hirudo nipponia* (*1*). It was named as bdellin-KL due to its amino acid composition and N-terminal sequence, which are similar to bdellin B-3 from a European leech that has also been partially sequenced (*2*). Bdellin-KL has disulfide bridges and a reactive site in the N-terminal region comprising about 40 amino acids, and has highly charged amino acids in the remaining region. Therefore, we expect that the inhibitor has a unique structure consisting of an N-terminal compact domain with a reactive site, and a C-terminal domain with an interesting sequence composed of His, Glu, Asp, and Gly. Large concentrations of bdellins are present in the salivary glands as well as other organs of leech, with high levels in the reproductive organ (*3, 4*). Moreover, the concentrations of bdellin along with several inhibitors increase remarkably within a few weeks after feeding blood (*5*). However, the entire primary structure and biological function of the protein are not yet revealed. Details of the primary structure and characteristics of bdellin-KL are essential for understanding the physiological function of this protein *in vivo*.

In this study, we establish the complete primary structure, including the C-terminal region with a highly charged repetitive sequence, of bdellin-KL by cDNA cloning, and describe the genomic sequence from the constructed library of Korean leech. We also analyze the biochemical properties of each of the domains and evaluate the application of the protein to a biologically active peptide.

## MATERIALS AND METHODS

*Materials*—Leeches collected from local ponds were kept in laboratory water tanks and fed on porcine blood (*6*). Porcine pancreatic trypsin, porcine blood plasmin, *N*-benzoyl-L-arginine *p*-nitroanilide (BAPNA), and endoproteinase Asp-N (sequencing grade) were purchased from Sigma. IPTG, X-gal, and PCR molecular markers were from Promega. The RNA ladder, all restriction endonucleases and T4 DNA ligase were from New England Biolabs. Hybond™-N, [$\alpha$-$^{35}$S]dATP, [$\alpha$-$^{32}$P]dCTP, and [$\gamma$-$^{32}$P]ATP were purchased from Amersham. Indo-1 and Phen Green were obtained from Molecular Probes. *Escherichia coli* strain ER1647 was used for plating libraries, amplification, titering, and screening. *E. coli* strain BM25 was used for automatic sub-

[1] This work was supported by grants from the Korea Ministry of Science and Technology and the Korea Science and Engineering Foundation. The nucleotide sequence data reported in this paper have been submitted to the GenBank sequence data bank and are available under accession numbers AF223972 and AF272790.

[2] To whom correspondence should be addressed. Tel: +82-42-869-2612, Fax: +82-42-869-2610, E-mail: kwkang@mail.kaist.ac.kr

Abbreviations: RACE: rapid amplification of cDNA ends; BAPNA: *N*-benzoyl-L-arginine *p*-nitroanilide; RP-HPLC: reversed phase high performance liquid chromatography; MALDI-MS: matrix-assisted laser desorption ionization-mass spectrometry; *tsp*: transcription start point.

© 2001 by The Japanese Biochemical Society.

cloning and DH5α was used as a host organism for transformation.

*Preparation of mRNA and cDNA Synthesis*—Total leech RNA was extracted by the acid guanidinium thiocyanate/phenol/chloroform method (*7*). The poly $(A)^+$ mRNA fraction was prepared using a polyA tract mRNA isolation kit (Promega). The first strand cDNA was synthesized from the mRNA using an *Eco*RI-*Not*I-$(dT)_{18}$ primer. For 5′ RACE, cDNA was synthesized using a Marathon™ cDNA amplification kit (CLONTECH).

*PCR Amplification and cDNA Cloning*—For amplification of the cDNA encoding the N-terminal part of bdellin-KL, a degenerate sense primer [PR1: 5′-GARTGYGTNTGYACNAARGA-3′] and a degenerate antisense primer [PR2: 5′-CCYTCRCANGCRTGYTCRTG-3′] were designed on the basis of the amino acid sequence corresponding to amino acid residues 21–27 and 54–60, respectively. Thirty cycles of PCR using TaKaRa Ex Taq polymerase were performed as follows: 1 min at 94°C, 30 s at 49°C, and 30 s at 72°C. To identify the 3′ end of the cDNA sequence, a nested sense primer [PR3: 5′-AAGACGGTGTCACCTACGACAAC-3′] was used together with a primer containing an *Eco*RI-*Not*I site [PR4: 5′-AACTGGAAGAATTCGCGGCCG-3′]. Then 5′ RACE was performed to clone the remainder of the cDNA using an antisense primer [PR5: 5′-TCCACAAATCCTTCGCAGGCATGTTCGT-3′] and an adapter primer AP1 (CLONTECH). The products of each PCR were subcloned into pT7Blue(R)T vector (Novagen) and sequenced by the dideoxy chain termination method (*8*) using a Sequenase version 2.0 kit (USB) or on Automated DNA Sequencer (Perkin Elmer ABI PRISM 377).

*Northern Blot Analysis*—Total RNA (20 μg) was separated in a 1.0% agarose gel containing 2.2 M formaldehyde and then transferred to a Hybond™-N membrane. A cDNA probe specific to bdellin-KL was labeled with [α-$^{32}$P]dCTP using the Prime-a-Gene™ labeling system (Promega). Hybridization was performed at 68°C for 2 h in a QuikHyb hybridization solution (Stratagene) containing 100 $\mu g \cdot ml^{-1}$ sonicated salmon sperm DNA. The membrane was washed twice in 2× SSC containing 0.1% SDS at room temperature, and once in 0.2× SSC containing 0.1% SDS at 60°C

*Isolation of Genomic DNA and Construction of a Leech Genomic Library*—Leeches (5 g) were frozen in liquid nitrogen and ground with a prechilled mortar. After treatment in 90 ml lysis buffer containing RNaseA and collagenase at 50°C for 2 h, proteinaseK was added and the sample was incubated for 12 h. Purification of genomic DNA was performed a using Genomic-tip™ kit (QIAGEN). The purified genomic DNA (50 μg) was partially digested with *Sau*3AI at 37°C for 30 min and a 7–15 kb region of DNA was eluted from a 0.5% agarose gel. The size-fractionated DNA (2 μg) was ligated to λBlueSTAR vector with T4 ligase overnight at 4°C. Packaging of recombinant phages was performed with a PhageMaker® system (Novagen). Recombinant phages were amplified once on agar plates using *E. coli* strain ER1647. The procedures for growth, purification of phage particles, and extraction of phage DNA were performed according to the manufacturer's instructions.

*Screening of the Genomic Library and Isolation of Bdellin-KL Genomic Clones*—The bacteriophage plaques (5 × $10^4$ plaques/150 mm plate) were lifted onto a Hybond™-N membrane. Hybridization was carried out at 42°C for 24 h using the same radioactive probe used for the Northern blot analysis in 30% formamide, 5× Denhardt's solution, 5× SSPE, and 100 μg/ml salmon sperm DNA. The filters were washed three times in 2× SSC containing 0.1% SDS at 20°C and twice in 0.2× SSC containing 0.1% SDS at 68°C, and autoradiographed. Positive plaques were selected and subcloned into BM25.8 according to the manufacturer's instructions.

*Southern Blot Analysis*—Genomic DNA (10 μg) was digested with restriction enzymes, electrophoresed in a 0.7% agarose gel, and transferred to a Hybond™-N membrane. Hybridization and membrane washing were performed as for the Northern blot analysis.

*Identification of tsp*—To identify a *tsp*, primer extension analysis was performed as described by Boorstein and Craig (*9*). For DNA sequencing and reverse transcription, primers complementary to nucleotides 99–120 [5′-AAGTCTTCAACTTCTACTCACCC-3′] and codons 51–73 [5′-CTGTATCTGCGTTGATGGCGACC-3′] were synthesized.

*Purification of the N-Terminal Fragment by Limited Proteolysis*—Bdellin-KL was prepared as described previously (*1*). The intact protein was treated with endoproteinase Asp-N in 100 mM Tris-Cl (pH 8.0) at 37°C for 15 h. The digested peptides were separated by RP-HPLC using a Vydac218TP $C_{18}$ column with a gradient from 0.1% TFA in water to 0.1% TFA in 80% acetonitrile. The purified peptides were sequenced at the N-terminus by the automated Edman degradation method on an Applied Biosystem 476A protein sequencer. The molecular mass of the N-terminal fragment was determined by COMPACT matrix-assisted laser desorption ionization-mass spectrometry (MALDI-MS II, KRATOS Analytical) (*10*).

*$K_i$ Determination of the N-Terminal Fragment*—Varying concentrations of the N-terminal fragment were incubated with either porcine pancreatic trypsin (90 nM) or porcine blood plasmin (3 μM) for 15 min. After adding 2.4 mM BAPNA, the residual enzyme activity was determined by measuring the $A_{410}$ of the reaction at 30 s intervals. The inhibition constants ($K_i$) of the N-terminal peptide to each enzyme were estimated by tight-binding analysis as described by Stone and Hofsteenge (*11*). Active site titrations of trypsin and plasmin were performed using *p*-nitrophenyl *p*′-guanidinobenzoate (*p*-NPGB) HCl (*12*).

*pH and Temperature Effects on the Inhibitory Activity of the N-Terminal Fragment*—After incubation of the N-terminal fragment under varying pH and temperature conditions for 40 min, the solution was diluted 500-fold with 100 mM Tris-Cl (pH 8.0), and then the trypsin inhibitory activity was determined by comparison with controls. The incubation buffers used were 50 mM glycine-HCl (pH 2.0), 20 mM HEPES (pH 7.0), and 50 mM glycine-NaOH (pH 12.0).

*Ion-Binding Assay*—The bdellin-KL was mixed with divalent ions, and fluorescence spectra were obtained on a fluorometer using first indo-1 as $Ca^{2+}$ and $Zn^{2+}$ probes (*13*, *14*) and Phen Green as $Fe^{2+}$ and $Fe^{3+}$ probes. After mixing 0.1 μM of bdellin-KL or its N-terminal fragment with a fifteen-fold excess of each ion, 1 μl of 1 mM probe was added and the change in fluorescence was measured. A mixture of each ion and probe without peptide, and separate mixtures of peptide, each ion, 3 μM EGTA, and the probe were used as controls. Excitation was performed for indo-1 at 355 nm and for Phen Green at 492 nm.

We observed whether the ions modulate the inhibitory activity of the protein. After mixing 50 μl of 30 μM bdellin-

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

KL with 50 μl of each ion (at 100 μM), 50 μl of 1 μM enzyme, such as trypsin or plasmin, was added. The absorbance change at 405 nm was monitored after the addition of 100 μl of 1 mM BAPNA. The percentage of inhibition (*%I*) was calculated as $\%I = (1-V_i/V_o) \times 100$, where $V_i$ and $V_o$ are the velocities (abs./min) with and without inhibitor.

## RESULTS

*Cloning and Sequencing of Bdellin-KL cDNA*—From the first strand cDNA, a 120-bp DNA fragment corresponding to amino acid residues 21–60 of bdellin-KL was amplified by PCR with the degenerate primers PR1 and PR2. Based on the DNA sequence, the 3′ end of cDNA was amplified with primers PR3 and PR4. After 5′- RACE with primers PR5 and AP1, the cDNA sequence was completed by combining the overlapping sequences from each PCR product (Fig. 1A) (GenBank accession No. AF223972).

There is an open reading frame of 465 nucleotides encoding a protein of 155 amino acids. The 18 residues from the

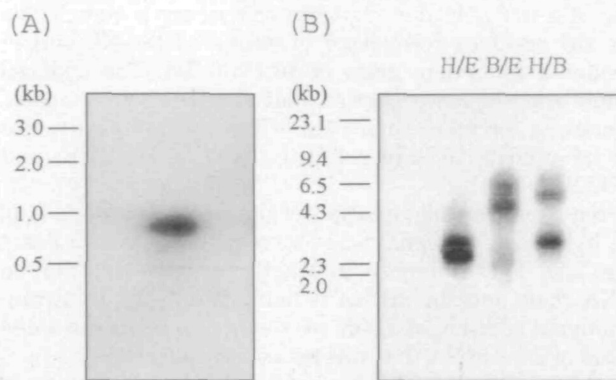


Fig. 2. **Northern and Southern blot analysis of bdellin-KL.** (A) Northern blot analysis. Twenty micrograms of total leech RNA was separated in a 1.0% agarose/formaldehyde gel, blotted onto a Hybond™-N membrane, and hybridized with a probe, [α-$^{32}$P]dCTP-labeled bdellin-KL cDNA. The size marker is an RNA ladder from New England Biolabs. (B) Southern blot analysis. Ten micrograms of genomic DNA was digested with *Hin*dIII/*Eco*RI (H/E), *Bam*HI/*Eco*RI (B/E), or *Hin*dIII/*Bam*HI (H/B), electrophoresed in a 0.7% agarose gel, and blotted onto the Hybond™-N membrane. Hybridization was performed with the bdellin-KL cDNA probe. The size marker is a *Hin*dIII digest of lambda DNA.

(A)

```
ACT TGA ATC AAC ATG AAG CTT TTG TTC GCT TTG GCC GTT TTT GGT GCT   48
                 M   K   L   L   F   A   L   A   V   F   G   A   12
CTG GTC GCC ATC AAC GCA GAT ACA GAA TGC GTC TGC ACC AAG GAA CTG   96
 L   V   A   I   N   A   D   T   E   C   V   C   T   K   E   L   28
CTG AGA GTT TGC GGA GAA GAC GGT GTC ACC TAC GAC AAC TCA TGT CTG  144
 L   R   V   C   G   E   D   G   V   T   Y   D   N   S   C   L   44
GCT ACG TGC CAT GGA ACG TCA GTT GCC CAC GAA CAT GCC TGC GAA GGA  192
 A   T   C   H   G   T   S   V   A   H   E   H   A   C   E   G   60
TTT GTG GAG CAC CAC GAA GAT GAA CAT CAC GAG GGC GAA GAG CAC AAG  240
 F   V   E   H   H   E   D   E   H   H   E   G   E   E   H   K   76
GAG GAA GGT CAC GAA GGT CAT GAT GAT CAT CAT GAC GAT GGT CAC GAG  288
 E   E   G   H   E   G   H   D   D   H   H   D   D   G   H   E   92
GAG CAC CAC GAG GGA GAA GAG CAC AAA GAT GAG CAT CAC GAG GAG GGT  336
 E   H   H   E   G   E   E   H   K   D   E   H   H   E   E   G  108
CAT GAT GAT CAT CAC GAG GAG GGT CAC GAT GAT CAC CAT GAC GAC GAA  384
 H   D   D   H   H   E   E   G   H   D   D   H   H   D   D   E  124
CAC AAG GAA GAT CAC CAT GAC GAC GAA CAC AAG GAA GAT GAT CAT CAT  432
 H   K   E   D   H   H   D   D   E   H   K   E   D   D   H   H  140
GAC GAC GAA CAC AAG GAT GAT GAT CAT CAC GAG GAG CAC CAC GAT TAG  480
 D   D   E   H   K   D   D   D   H   H   E   E   H   H   D stop 155
TCG GTG TCT TGG GCA ACA GAC CGG TCC TTT TTC TTG TTG CCA TAG AAC  528
AGT TAA CCG GTT CAA TAG CTT AGA GTA TTT TTT TCC TTA CTT TAT ATT  576
TTT TTA TTT CGG AAT CAT TCT GCT GAT GAT GTA AAT CTC AAT CGA ATG  624
TAT CCT GTT GCA GAG ACG AAT GAC GTA AAC ATG TTG TTG ACA TGA AAA  672
TAT GAA GTT TTT CAA ATG TCA AAA AAA AAA AAA AAA AAA              711
```

(B)

```
                   10         20       30         40         50        60
NH2-MKLLFALAVFGALVAINADTECVCTKELLRVCGEDGVTYDNSCLATCHGTSVAHEHACEGFVE-

 64  [H H] [E D E] [H H] [E] G [E E]
 75  [H K] [E E] G [H] [E] G [H] [D D]
 86  [H H] [D D] G [H] [E E]
 94  [H H] [E] G [E E] [H K] [D E]
104  [H H] [E E] G [H] [D D]
112  [H H] [E E] G [H] [D D]
120  [H H] [D D E] [H K] [E D]
129  [H H] [D D E] [H K] [E D D]
139  [H H] [D D E] [H K] [D D D]
149  [H H] [E E] [H H] [D]-COOH
```

Fig. 1. **Nucleotide and deduced amino acid sequences of the bdellin-KL cDNA.** (A) The nucleotide sequence of the cDNA encoding bdellin-KL and its deduced amino acid sequence (GenBank accession No. AF223972). There is an open reading frame of 465 nucleotides encoding a protein of 155 amino acids. (B) Rearrangement of the deduced amino acid sequence of bdellin-KL. The signal peptide is underlined and the arrow denotes the reactive site. The disulfide bridge pattern, shown as dotted lines, was elucidated for bdellin B-3 (*4*). Positively charged amino acids are shown in shaded boxes and negatively charged amino acids in open boxes.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

first Met are a leader sequence containing a signal. The last 137 residues correspond to native bdellin-KL with a calculated molecular mass of 16,137.0 Da. The deduced amino acid sequence showed that the C-terminal region, accounting for more than 65% of the total sequence, has unique repeats including HHEE, HHDD, HHDDE, and HKEDD (Fig. 1B).

From Northern blot analysis of the total leech RNA, only one hybridization signal was observed at about 900 bases (Fig. 2A). The difference in length between the 711 bp cDNA clone and the mRNA is likely due to the post-transcriptional addition of a poly (A) tail or to a 5′-untranslated region of the mRNA that was not copied into cDNA.

*Genomic Sequence Analysis of Bdellin-KL*—Screening of the constructed genomic library revealed a clone encoding bdellin-KL, and its sequence was determined up to 2,109 bases (Fig. 3) (GenBank accession No. AF272790). The clone contains an open reading frame (ORF) composed of 3 exons and 2 introns spanning about 1 kb. The exon-intron organization was determined by comparing the nucleotide sequence with the cDNA sequence. The sequence of each exon-intron junction completely matched the consensus sequence for splicing acceptor/donor sites (GT/AG rule) (*15*).

The transcription start point (*tsp*) in bdellin-KL gene was determined by primer extension analysis (Fig. 4). Two extension products, 121 and 122 nt in length, were observed, corresponding to T and G located 59 and 60 bp upstream of the start codon (ATG), respectively. The +1 position was

```
 -572                             AAGCCGGATAATAATAATAATAATAATAATAA
 -540 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA
 -480 TAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAA
 -420 TAATAATAATAATAATAATAATAATACCCAAGTCACTCATCAAATTTTATTTTTTCAAGC
 -360 CAAAGTTTTTTTTCCAAATTATCTCGGCTGTTTTGAAGTTGCAAAAAACGTTCAAGCGAC
                         GATA-1
 -300 ATTTTCGAACATTAACATTAACAAAGGCTCCATCCAGCTTCAGTTAATTTACCAGCGGTT
                                    CAAT
 -240 GGTCGCTAATCTCGGCGAAAAGGAAAGCAAATCCACAAGCTTACTTGTATTTTTGTGAAA
                 SP-1                AP-3
 -180 AATGACAAGAAACTTAAATTTAGAAAAAAACAACCCTTAGGATTGAAGACAGATGGTGCA
                     Pit-1                              E box
 -120 TTTTAGACTCACGGCATCAATCAAACCAAGAGCTAAGACCGGTTCTGATTAACTGAACAC
                       CAAT                          AP-1
  -60 GGAGCAGTCTAGACGAGGTACCACGGCGAGCTTATATAAAGACACTCGTGATAAGTCGTG
      ↓                      SP-1        TATA         GATA-1
    1 GTAGGGTTATCTCTTTTTCCTCTAGTTGCGTTTTCTCTGCGGATAAACACTTGAAACAAC
            GATA-1                                  Pit-1
   61 ATGAAGCTTTTGTTCGCTTTGGCCGTTTTTGGTGCTCTGGGTGAGTAGAAGTTGAAGACT
                                                   AP-1
  121 TTTTTTTTGTCGGTAAAATGATTGAAGACAAATGCTTAATTTCAGGGAATATTTTAAAGC
  181 ATATCTCTGGTTTTTTTGTGTTTGTAACAATAAAATTATAATTTATCCGAAGTTTCGGTT
       GATA-1
  241 ACTGTCAAGTTACAGATGGTAATTTGAGAATGTCAGCAAGACAAAATTCAATAATTTCCT
                  E box
  301 CCAGATTTAGTTACAAAATACAACTACTGAATATCTTAAATATTTTATCTTATATATAAA
                                    GATA-1, Pit-1   GATA-1
  361 AAAAAATCGAATCACTTATCATTTGGCAAATTTGAATTTTACAGTCGCCATCAACGCAGA
                     GATA-1
  421 TACAGAATGCGTCTGCACCAAGGAACTGCTGAGAGTTTGCGGAGAAGACGGTGTCACCTA
  481 CGACAACTCATGTCTGGCTACGTGCCAGTCAGTAAATCTATTTTATTACTTAATGAACTA
  541 ATTGATTAGTATTGTTTGTTACTATTATTATTATTATTATTATTATTATTATTATTAATA
  601 GTATTATTAGTTACTATTATCACAGAGATACGAGATTAGATGCATTAGATTACATAAGGA
  661 AATAACTTCAAACACCAAAACTAAAATAAACATCAAAACCATCTTTCACAGTGGAACGTC
  721 AGTTGCCCACGAACATGCCTGCGAAGGATTTGTGGAGCACCACGAAGATGAACATCACGA
  781 GGGCGAAGAGCACAAGGAGGAAGGTCACGAAGGTCATGATGATCATCATGACGATGGTCA
  841 CGAGGAGCACCACGAGGGAGAAGAGCACAAAGATGAGCATCACGAGGAGGGTCATGATGA
  901 TCATCACGAGGAGGGTCACCGATGATCACCATGACGACGAACACAAGGAAGATCACCATG
  961 ACGACGAACACAAGGAAGATGATCATCATGACGACGAACACAAGGATGATGATCATCACG
 1021 AGGAGCACCACGATTAGTCGGTGTCTTGGGCAACAGACCGAGTCCTTTTTCTTGTTGCCG
 1081 TAGAACAGTTAACCGGTTCAATAGCTTAGAGTATTTTTTTCCTTACTTTATATTTTTTTA
 1141 TTTCGGAATCATTCTGCTGATGATGTAAATCTCAATCGAATGTATCCTGTTGCAGAGACG
 1201 AATGACGTAAACATGTTGTTGACATGAAAATATGAAGTTTTTCAAATGTCAGAGCTAGAA
 1261 ATAAAGTCTAACAAATAATCAAACTACCCCCTCTTAAGAGTAAGGCTAACGATAGCTACG
 1321 ACTTCAGCTTAACCAAAAAACTGTTAAGAATCCGCTGCCAAGCTACCTTTACTTTTTTTT
 1381 TTATTTTTTTATTTATTTTATTAGACACAATTGAAGGGCCCAATTGGACCACTAACACAA
 1441 TTGCAAGAAAAAAAAAAACCAAAAAACATCGACCGGCGTGTACAACAGGGGCAGCATATA
 1501 GATTAAAGTGCATTTAAATAATGGGTTAAGGCAAGGG
```

Fig. 3. **Nucleotide sequence of the bdellin-KL gene (GenBank accession No. AF272790).** Nucleotide numbering begins at the transcription start point (arrow). The exon is boxed, and the start and stop codons are in bold face. The unusual ATA repeats are underlined. The potential CAAT boxes, TATA box, GC boxes, transcription factor binding sites, and the polyadenylation signal are shadowed.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

assigned to an intense band and corresponded to the G residue.

To define DNA regulatory elements that may mediate the expression of the bdellin-KL gene, the 5′-flanking region was analyzed (Fig. 3). A consensus TATA box sequence (TATAAA) was found at −26 bp from the *tsp*. CAAT sequences were found at −103 bp and −271 bp. Two putative Sp1 boxes were located 39 bp and 230 bp upstream as CACGGCG and CTCGGCG, respectively. From −395 bp to −565 bp, the sequence ATA was repeated. Two AP1 elements were found at positions −75 and +102. Some potential binding sites for specific transcription factors such as GATA-1, E box, and Pit-1 were found in the 5′-flanking region as well as the first intron of the ORF. The canonical polyadenylation signal (AATAAA) was found 222 bp downstream of the stop codon.

Southern blot analysis was performed after digestion of the genomic DNA with *Hin*dIII/*Bam*HI, *Bam*HI/*Eco*RI, and *Hin*dIII/*Eco*RI. This revealed that more than two copies of the bdellin-KL gene are present in the genome (Fig. 2B).

*The N-Terminal Domain as a Protease Inhibitor*—The N-terminal region comprising 48 amino acids (amino acid residues 19–66) was purified from RP-HPLC and identified by N-terminal amino acid sequencing and mass spectrometry. This region contains various amino acids with the reactive site $Lys^{26}$ and all disulfide bridges (*4*), while the remaining C-terminal region consists of several repetitive combinations of four amino acids (H, E, D, and G). Therefore, we specified the N-terminal region with the reactive site and disulfide bridges as an N-terminal domain and the remaining region as a C-terminal domain.

By tight-binding inhibition analysis (*11*), the inhibition constants ($K_i$) of the N-terminal fragment were estimated to be (3.58 ± 1.24) × $10^{-9}$ M (means ± SD) for trypsin and (8.64 ± 4.37) × $10^{-9}$ M for plasmin (Table I). In the case of intact native bdellin-KL, the inhibition constants ($K_i$) were (1.36 ± 0.42) × $10^{-9}$ M for trypsin and (1.56 ± 0.64) × $10^{-9}$ M for plasmin. These results demonstrate that protein has inhibitory activity in the N-terminal domain only.

The inhibitory activity of the N-terminal fragment was not strongly influenced by pH or temperature (Fig. 5). More than 90% of the activity remained up to 90°C at neutral pH and 50°C at extreme pH. After denaturation and *S*-pyridylethylation of the fragment, however, the inhibitory activity was lost. According to these results, the inhibitory activity of bdellin-KL appears to depend on the structure of the N-terminal domain being maintained by three disulfide bridges.

*The C-Terminal Domain as a Binding Region to Charged Molecule*—In the C-terminal part of bdellin-KL, charged amino acids (H, D, and E) are distributed regularly, as mentioned previously. This region is likely to be involved in molecular interactions (*1*, *2*). We investigated whether the protein binds to ions commonly present in the leech body ($Ca^{2+}$, $Zn^{2+}$, $Fe^{3+}$, and $Fe^{2+}$). We obtained fluorescence spectra of the protein after incubation with fifteenfold amounts of each ion (Fig. 6). Compared with the fluorescence spectrum of $Ca^{2+}$ alone, that of a mixture of bdellin-KL and $Ca^{2+}$ decreases markedly to the levels of a mixture to which EGTA is added. Also in the $Zn^{2+}$, $Fe^{3+}$, and $Fe^{2+}$-binding assays, the fluorescence spectra of the mixture of bdellin-KL and each ion showed a marked decrease compared to mixture of ion alone. These results indicate that bdellin-KL binds to these ions. Moreover, the fluorescence spectra of the N-terminal fragment with each ion showed patterns similar to those of each ion. This shows that the C-terminal domain plays a part in ion-binding.

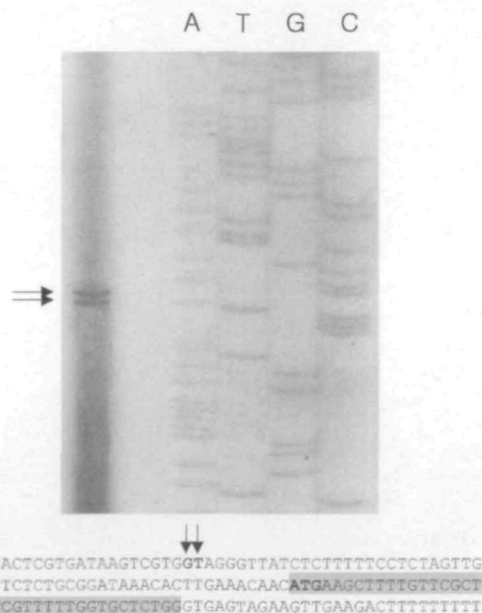


Fig. 4. **Primer extension analysis.** The arrows indicate *tsp* corresponding to the nt G and T in the sequence below. The +1 position was assigned to the more intense band corresponding to nt G. The start codon (ATG) is in bold face and the first exon is shadowed. The sequencing primer used is underlined.

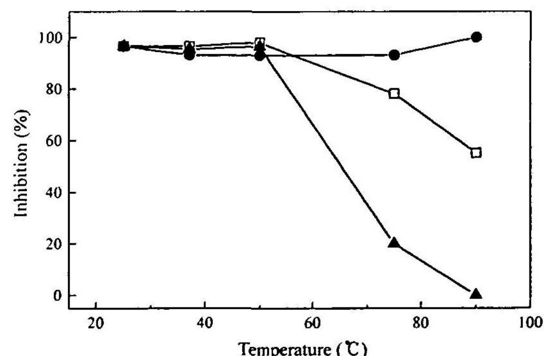


Fig. 5. **Temperature and pH effects on the inhibitory activity of the N-terminal fragment.** The trypsin inhibitory activity of the N-terminal fragment was measured after 40 min incubation at various temperatures and at three different pH: (□) 50 mM glycine-HCl (pH 2.0); (●) 20 mM HEPES (pH 7.0); (▲) 50 mM glycine-NaOH (pH 12.0). The percentage of inhibition (*%I*) was calculated as $\%I = (1 - A_i/A_o \times 100$, where $A_i$ and $A_o$ are the absorbance at 405 nm with and without inhibitor.

TABLE I. **Comparison of the inhibitory activities of the intact protein and the N-terminal fragment.**

| | Intact ptotein (Residues 19–155) | N-terminal fragment (Residues 19–66) |
|---|---|---|
| Trypsin $K_i$ (M) | (1.36 ± 0.42) × $10^{-9}$ | (3.58 ± 1.24) × $10^{-9}$ |
| Plasmin $K_i$ (M) | (1.56 ± 0.64) × $10^{-9}$ | (8.64 ± 4.37) × $10^{-9}$ |



Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

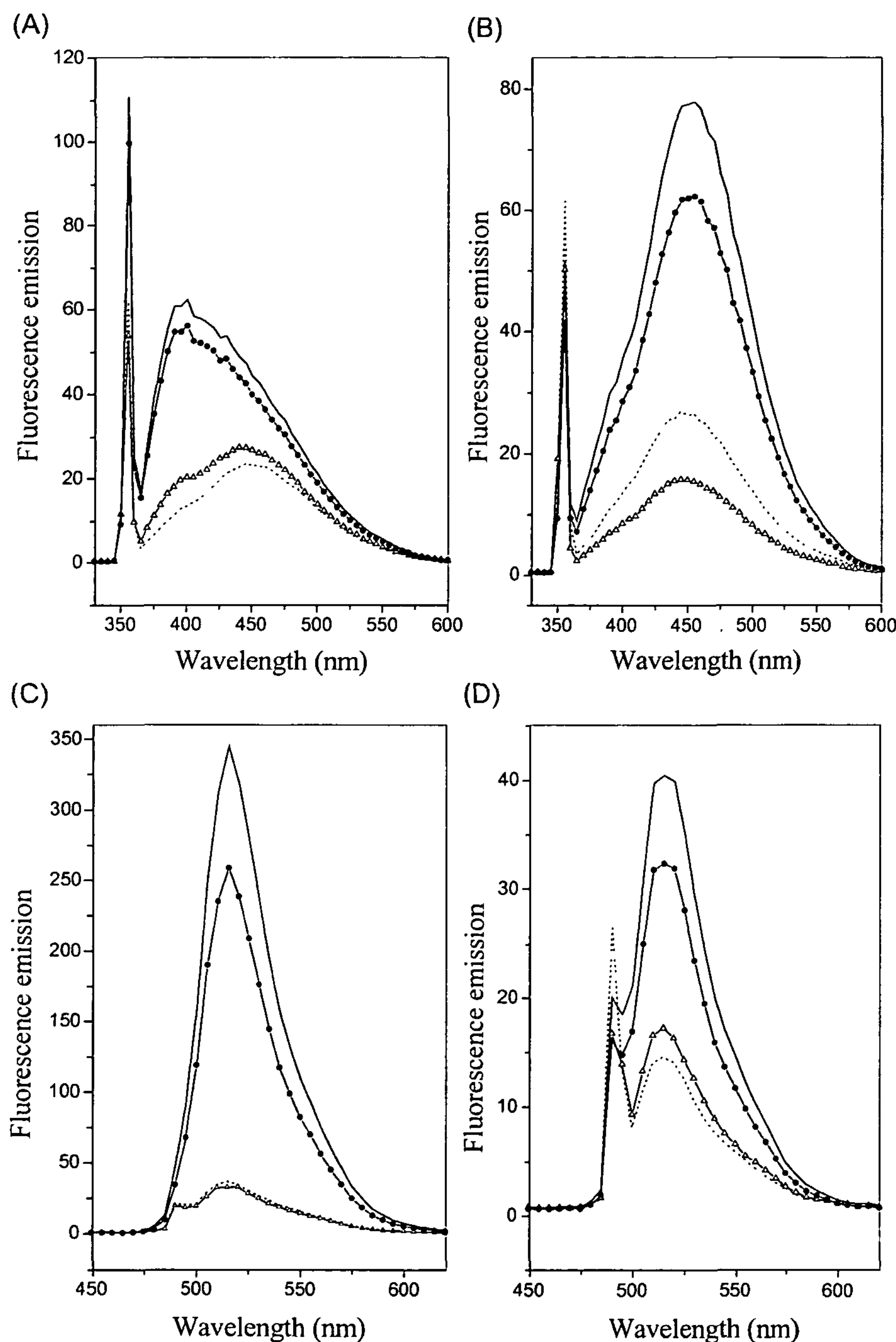


Fig. 6. **Ion-binding assay of bdellin-KL.** The fluorescence spectra of bdellin-KL after incubation with ions were obtained for (A) $Ca^{2+}$, (B) $Zn^{2+}$, (C) $Fe^{3+}$, and (D) $Fe^{2+}$. We used indo-1 for $Ca^{2+}$ and $Zn^{2+}$ probes and Phen Green for $Fe^{2+}$ and $Fe^{3+}$ probes. Following the mixing of 0.1 μM bdellin-KL or its N-terminal fragment with a fifteenfold excess of each ion, 1 μl of 1 mM probe was added and the change in fluorescence was measured. (—) ion alone; (△) bdellin-KL/ion mixture; (...) addition of 3 μM EGTA to the bdellin-KL/ion mixtures; (●) N-terminal fragment/ion mixtures. Excitation was at 355 nm for indo-1 and at 492 nm for Phen Green.

These ions modulate the inhibitory activity of the protein both before or after ion-binding. We conclude that the C-terminal domain serves as a binding region for ions without influencing the inhibitory activity (Fig. 7).

## DISCUSSION

Recently, a trypsin-plasmin inhibitor, bdellin-KL, was purified from *H. nipponia* (*1*). This protein was found to contain high compositions of four amino acids: H, E, D, and G. However, the entire amino acid sequence was not determined by peptide analysis due either to a lack of fragments or to excessive disintegration after proteolytic enzyme treatment. For further characterization, molecular cloning approaches were needed to determine the primary structure of the inhibitor.

A unique repeated sequence in the C-terminus of bdellin-KL was revealed by a combination of RT-PCR and 5′-RACE techniques on mRNA from *H. nipponia*. The C-terminal part is composed of repeated sequences, HHEE or HHDD. A common characteristic of the alignments is that two positively charged amino acids are followed by two or three negatively charged amino acids with an occasional insertion of Gly between the repeats. There was a little discrepancy between the deduced amino acid sequence and the sequence of native bdellin-KL. We confirmed that the clone



Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

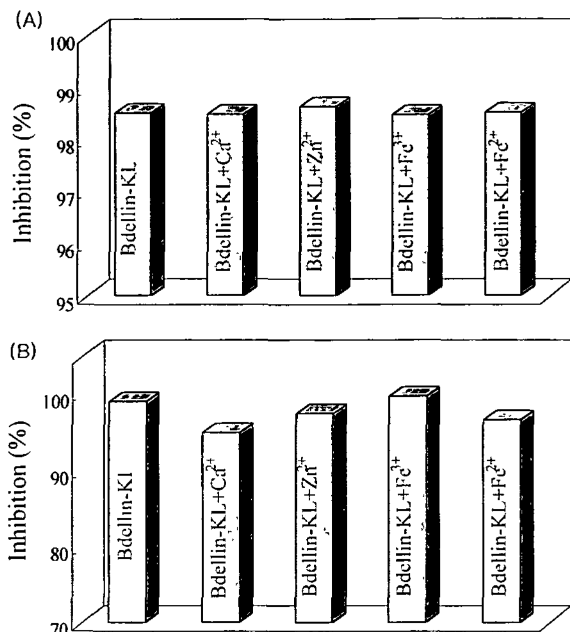


Fig. 7. **Effect of ion-binding on the inhibitory activity.** For trypsin (A) and plasmin (B), the inhibitory activities of bdellin-KL were measured after ion-binding. After mixing 50 µl of 30 µM bdellin-KL with 50 µl of 100 µM each ion, 50 µl of 1 µM enzyme was added. The absorbance change at 405 nm was monitored after adding 100 µl of 1 mM BAPNA. The percentage of inhibition (*%I*) was calculated as $\%I = (1-V_i/V_0 \times 100$, where $V_i$ and $V_0$ are the velocity (abs./min) with and without inhibitor.

obtained really encodes bdellin-KL by internal peptide sequencing and amino acid analysis. The identified internal sequence after trypsin treatment is DEHHEEGHDDHH-HEEGHD, which corresponds to amino acid residues 102–118. Therefore, the most probable explanation for the differences between the sequences would be errors in the sequence analysis of native bdellin-KL.

In the case of bdellin-KL, all six cysteines are in the N-terminal part and are involved in disulfide bonds (*4*). The N-terminal fragment was not digested by further incubation with endoproteinase Asp-N, even though two more cleavage sites are present. Moreover, the inhibitory activity of the region is as potent as that of the intact protein and very stable over wide ranges of pH and temperature. This result shows that the N-terminal domain constitutes a compact structure resistant to proteolysis and plays an important role in the inhibitory action of bdellin-KL.

Some proteins with similar internal repeats have been found. The histidine rich protein (HRP) from *Plasmodium falciparum* demonstrates the existence of a heme-binding site (HHAANA) (*16*). Histidine, glutamate, and aspartate are repeated in the histidine-rich $Ca^{2+}$-binding protein (HCP) from sarcoplasmic reticulum (*17*). The structural motif HD/EXXH is common in proteins that bind divalent metal ions with high affinity, especially $Zn^{2+}$ (*18*). Therefore, we expected that bdellin-KL would bind highly charged molecules such as heme, divalent ions, and other charged molecules. In practice, the C-terminal domain shows binding affinity for $Ca^{2+}$, $Zn^{2+}$, $Fe^{3+}$, and $Fe^{2+}$ without an influence on the inhibitory activity.

The bdellin-KL gene includes several potential binding sites for tissue-specific transcription factors such as E box, Pit-1, and GATA-1. The E box has been shown to be a B-cell and myocyte-specific transcription factor binding site (*19*), while the Pit-1 binding site is found in genes expressed specifically in the anterior pituitary gland (*20*). GATA-1 has been determined to play key roles in the regulation of a number of different erythroid-specific genes (*21*), and in the generation of hematopoietic cell types (*22*). Considering the hypothetical role of bdellin in blood preservation in leeches (*5*), and the similar sequences of heme binding sites in histidine rich proteins from *P. falcifarum* (*16*), the presence of GATA-1 consensus sequences support the possible function of bdellin-KL in blood preservation. Thus, the C-terminal part may sequestrate or release heme while the N-terminal part regulates protease activity. Bdellin might also be involved in sperm-egg interactions. The concentrations of bdellin are present in the reproductive organ, and bdellin strongly inhibits acrosin, which occurs in the acrosome of spermatozoa (*23*). GATA-1 has also been suggested to regulate genes during the earliest stage of spermatogenesis (*24*, *25*). There have been some reports that trypsin inhibitors might control such sperm functions as $Ca^{2+}$ influx in the acrosome reaction and the activation of acrosin and other serine proteases (*26*). Therefore, we would expect that the C-terminal part plays a part in interactions with charged molecules while the N-terminal part acts as a protease inhibitor during acrosome reaction.

In the present study, we have deduced the primary structure of bdellin-KL by molecular cloning and demonstrate that the protein is a novel bifunctional inhibitor with two distinct domains. While the N-terminal domain acts as a stable trypsin-plasmin inhibitor, the C-terminal domain serves as a binding region for charged molecules. In addition, the gene sequence analyses provide a clue to the structure-function relationship of bdellin, and may be helpful for understanding multifunctional proteins *in vivo*.

We would like to thank Dr. Seok Jin Hong, Dr. Dong Ryoung Kim, Chang Soo Kang, and Seong Won Cha for helpful discussions. We also thank Dr. Christian P. Sommerhoff and Prof. Hans Fritz (Ludwig-Maximilians-Universität München, Germany) for technical assistance with the nested PCR and for discussions.

REFERENCES

1. Kim, Y.H., Hong, S.J., Kim, D.R., Kim, S.I., Ha, K.S., and Kang, K.W. (1996) A new trypsin inhibitor from *Hirudo nipponia*. *Mol. Cells* **6**, 571–576
2. Fink, E., Rehm, H., Gippner, C., Bode, W., Eulitz, M., Machleidt, W., and Fritz, H. (1986) The primary structure of bdellin B-3 from the leech *Hirudo medicinalis*. Bdellin B-3 is a compact proteinase inhibitor of a 'non-classical' Kazal type. It is present in the leech in a high molecular mass form. *Biol. Chem. Hoppe-Seyler* **367**, 1235–1242
3. Fritz, H. and Krejci, K. (1976) Trypsin-plasmin inhibitors (bdellins) from leeches. *Methods Enzymol.* **45**, 797–806
4. Krejci, K. and Fritz, H. (1976) Structural homology of a trypsin-plasmin inhibitor from leeches (bdellin B-3) with secretory trypsin inhibitors from mammals. *FEBS Lett.* **64**, 152–155
5. Roters, F.J. and Zebe, E. (1992) Protease inhibitors in the alimentary tract of the medicinal leech *Hirudo medicinalis*: *in vivo* and *in vitro* studies. *J. Comp. Physiol. [B]* **162**, 85–92
6. Jung, H.I., Kim, S.I., Ha, K.S., Joe, C.O., and Kang, K.W. (1995) Isolation and characterization of guamerin, a new human leukocyte elastase inhibitor from *Hirudo nipponia*. *J. Biol. Chem.* **270**, 13879–13884



Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012

7. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156–159
8. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467
9. Boorstein, W.R. and Craig, E.A. (1989) Primer extension analysis. *Methods Enzymol.* **180**, 347–369
10. Beavis, R.C. and Chait, B.T. (1996) Matrix-assisted laser desorption ionization mass spectrometry of proteins. *Methods Enzymol.* **270**, 519–551
11. Stone, S.R. and Hofsteenge, J. (1986) Kinetics of the Inhibition of Thrombin by hirudin. *Biochemistry* **25**, 4622–4628
12. Chase Jr., T. and Shaw, E. (1970) Titration of trypsin, plasmin, and thrombin with $p$-Nitrophenyl $p'$-guanidinobenzoate HCl in *Methods in Enzymology* (Perlmann, G.E. and Lorand, L, eds.) Vol. 19, pp. 20–27, Academic Press, New York
13. Jefferson, J.R., Hunt, J.B., and Ginsburg, A. (1990) Characterization of indo-1 and quin-2 as spectroscopic probes for $Zn^{2+}$-protein interaction. *Anal. Biochem.* **187**, 328–336
14. Foehr, U.G., Heizmann, C.W., Engelkamp, D., Schaefer, B.W., and Cox, J.A. (1995) Purification and cation binding properties of the recombinant human s100 calcium-binding protein A3, an EF-hand motif protein with high affinity for zinc. *J. Biol. Chem.* **270**, 21056–21061
15. Senapathy, P., Shapito, M.B., and Harris, N.L. (1990) Splice junctions, branch point sites, and exons: Sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**, 252–278
16. Sullivan, D.J., Ayala, Y.M., and Goldberg, D.E. (1996) An unexpected 5′ untranslated intron in the *Plasmodium falciparum* genes for histidine-rich proteins II and III. *Mol. Biochem. Parasitol.* **83**, 247–251
17. Hofmann, S.L., Topham, M., Hsieh, C.L., and Francke, U. (1991) cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum protein, and localization of the gene to human chromosome 19 and mouse chromosome 7. *Genomics* **9**, 656–669
18. Arnold, F.H. and Hagmore, B.L. (1991) Engineered metal-binding proteins: purification to protein folding. *Science* **252**, 1796–1797
19. Murre, C., Voronova, A., and Baltimore, D. (1991) B-cell- and myocyte-specific E2-box-binding factors contain E12/E47-like subunits. *Mol. Cell. Biol.* **11**, 1156–1160
20. Anderson, B. and Rosenfeld, M.G. (1994) Pit-1 determines cell types during the development of the anterior pituitary gland. *J. Biol. Chem.* **269**, 29335–29338
21. Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.F., D'Agati, V., Orkin, S.H., and Constantini, F. (1991) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* **349**, 257–260
22. Shivadasami, R.A., Fujiwara, Y., McDevitt, M.A., and Orkin, S.H. (1997) A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. *EMBO J.* **16**, 3965–3973
23. Sawyer, R.T. (1986) *Leech Biology and Behavior*, Vol. II, pp. 473–505, CLARENDON Press, Oxford
24. Ito, E., Toki, T., Ishihara, H., Otani, H., Gu, L., Yokoyama, M., Engel, J.D., and Yamamoto, M. (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature* **362**, 466–468
25. Moroni, E., Cairns, L., Ottolenghi, S., Giglioni, B., Ashihara, E., Migliaccio, G., and Migliaccio, A.R. (1997) Expression in hematopoietic cells of GATA-1 transcripts from the alternative "testis" promoter during development and cell differentiation. *Biochem. Biophys. Res. Commun.* **231**, 299–304
26. Winnica, D.E., Novella, M.L., Dematteis, A., and Coronel, C.E. (2000) Trypsin/acrosin inhibitor activity of rat and guinea pig caltrin proteins. Structural and functional studies. *Biol. Reprod.* **63**, 42–48

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 30, 2012