Molecular cloning of two novel types of peptidylarginine deiminase cDNAs from retinoic acid-treated culture of a newborn rat keratinocyte cell line

Akihito Ishigami[a,*], Masashi Kuramoto[b], Michiyuki Yamada[b], Kazutada Watanabe[c], Tatsuo Senshu[a]

[a] *Department of Cell Chemistry, Tokyo Metropolitan Institute of Gerontology, 35-2 Sakae-cho, Itabashi-ku, Tokyo 173-0015, Japan*
[b] *Graduate School of Integrated Science, Yokohama City University, 3-2 Seto, Kanazawa-ku, Yokohama 236-0027, Japan*
[c] *Department of Experimental Biology, Tokyo Metropolitan Institute of Gerontology, 35-2 Sakae-cho, Itabashi-ku, Tokyo 173-0015, Japan*

Received 26 May 1998

Abstract Peptidylarginine deiminases (PADs) are a group of enzymes which convert protein arginine residues to citrulline residues. Using rat muscle PAD cDNA as a probe, we obtained two novel cDNAs, PAD-R11 and PAD-R4, from immortalized rat keratinocytes treated with all-*trans* retinoic acid. Comparison of the deduced amino acid sequences with those of muscle and hair follicle enzymes showed high conservation in the C-terminal region. Recombinant proteins encoded by both PAD-R11 and PAD-R4 showed the enzyme activities. That of PAD-R11 showed a characteristic feature of the enzyme found in the epidermis.

Key words: Peptidylarginine deiminase; All-*trans* retinoic acid; Keratinocyte differentiation

1. Introduction

The process of normal epidermal differentiation is characterized by a series of morphologic changes as keratinocytes progress from the germinative basal layer through the spinous and granular layers to the outer cornified layer. The climax of the process is the transition from the granular cells to the cornified cells, during which a number of proteins are subjected to various post-translational modifications. We recently reported the presence of deiminated forms of keratins and filaggrin, which is a kind of keratin-associated protein, in the cornified layer of mammalian epidermis [1,2]. This led us to suggest a possible role of 'protein deimination' during the cornification of epidermis. The 'protein deimination' is catalyzed by 'peptidylarginine deiminase' (PAD; EC 3.5.3.15) which converts arginine residues to citrulline residues [3–11]. There are at least three types of PADs in mammalian tissues, i.e. 'type I' or 'epidermal type', 'type II' or 'muscle type' and 'type III' or 'hair follicle type', all of which have a definite requirement for calcium ion. These enzymes differ in their relative activities towards synthetic substrates [8], elution profiles from anion exchange columns [9] or cross-reactivities with a polyclonal antibody to PAD type II [8,9]. The complete amino acid sequences of PAD type II and PAD type III have been deduced from their relevant cDNA sequences [12–14]. The cDNA cloning of the remaining type I enzyme is crucially important to explore not only its functional role but also its regulation during the terminal stages of epidermal differentiation. We found that all-*trans* retinoic acid (RA) induced multiple PADs in a newborn rat keratinocyte cell line, one of which was regarded as PAD type I from its chromatographic profile and relative activities towards synthetic substrates [15]. We constructed a cDNA library from the RA-treated cells as a possible source of PAD type I cDNA clones. Initial screening was conducted using PAD type II cDNA as a hybridization probe, and representative clones containing novel sequences were characterized by 5′-terminal extension using the 5′-RACE method and also by the enzymatic properties of the encoded recombinant proteins. Here we present a cDNA clone encoding catalytically active PAD type I. Moreover, we obtained another cDNA clone encoding a novel type of PAD, which was tentatively termed type IV.

2. Materials and methods

2.1. Construction and cloning of the cDNA library

A newborn rat keratinocyte cell line (kindly provided by Dr. Baden, Cutaneous Biology Research Center, Massachusetts General Hospital, Charlestown, MA, USA) was incubated with 5 μM RA for 4 days under the conditions described [15]. Poly(A)$^+$ RNA was extracted using a FastTrack mRNA isolation kit (Invitrogen). An oligo(dT)-primed cDNA library was constructed in a λgt11 vector using a complete rapid cloning system (Amersham) according to the supplier's instruction. About 5×10^5 independent clones were subjected to initial screening by plaque hybridization using a ^{32}P-labeled 2.3 kb *Eco*RI fragment of pKS-PAD [15] covering the entire length of the coding sequence of rat PAD type II. Hybridization was performed at 42°C in a solution containing 5×SSPE, 50% deionized formamide, 5×Denhardt's solution, 1% sodium dodecyl sulfate (SDS) and 10% dextran sulfate. Membranes were washed twice with 2×SSPE containing 0.1% SDS at room temperature and twice with 1×SSPE containing 0.1% SDS at 65°C, successively. Positive clones were selected, plaque-purified, and subcloned into a pBluescript SK$^-$ phagemid vector (Stratagene) for restriction mapping. Sequencing was performed by the dideoxynucleotide chain termination method using a Thermo Sequenase fluorescent labelled primer cycle sequencing kit (Amersham) and an automated DNA sequencing system (Shimadzu, model 1000L). Clones λPAD-R11 and λPAD-R4 were chosen for further characterization.

2.2. 5′-Rapid amplification of cDNA ends (5′-RACE)

The 5′-terminal sequence was determined using a rapid amplification of cDNA ends (RACE) system (Gibco BRL) according to the supplier's instruction. The antisense strand synthesis was directed with 1 μg of poly(A)$^+$ RNA and primed with 5′-ACCTCGGAGAGGGGCCGTGTGT-3′ (R11 GSP1) or 5′-GTTCATCTTGGCCTTCG-

*Corresponding author. Fax: (81) (3) 3579-4776.
E-mail: ishigami@tmig.or.jp

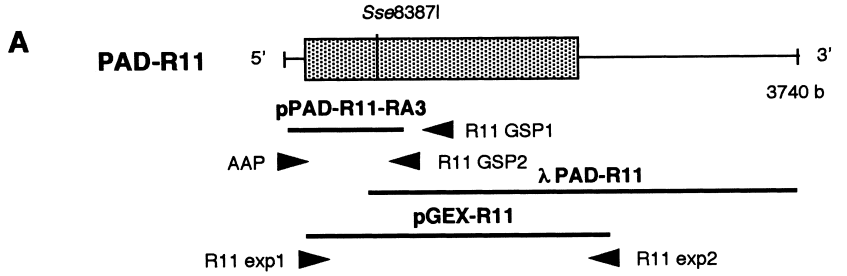
Abbreviations: PAD, peptidylarginine deiminase; RA, all-*trans* retinoic acid; PCR, polymerase chain reaction; RACE, rapid amplification of cDNA ends; GST, glutathione *S*-transferase; BAEE, benzoyl-L-arginine ethyl ester; Bz-L-Arg, benzoyl-L-arginine

0014-5793/98/$19.00
PII: S0014-5793(98)00893-X

GCACCT-3′ (R4 GSP1) for the extension of λPAD-R11 or λPAD-R4, respectively, followed by the terminal addition of the oligo-C tail and the sense strand synthesis. A polymerase chain reaction (PCR) was conducted using Takara Ex Taq polymerase (Takara Biomedicals) and primed with 5′-GGCCACGCGTCGACTAGTACGGGIIG-GGIIGGGIIG-3′ (AAP) and 5′-AGTCTTCTGGAATCCGGCAAA-3′ (R11 GSP2) or 5′-ACCTTGTCGTCC TCGAAGTCCATT-3′ (R4 GSP2) for the amplification of PAD-R11-RA3 or PAD-R4-RA6, respectively. The PCR products were cloned into a pGEM-T Easy vector (Promega) for sequencing.

2.3. Construction of plasmids for expression of recombinant proteins

A full length cDNA clone (pPAD-R11) was constructed by inserting the *Sac*II/*Sse*8387I fragment of pPAD-R11-RA3 into the *Sac*II/*Sse*8387I site of λPAD-R11. Another full length cDNA clone (pPAD-R4) was constructed by inserting the *Sac*II/*Nar*I fragment of pPAD-



B

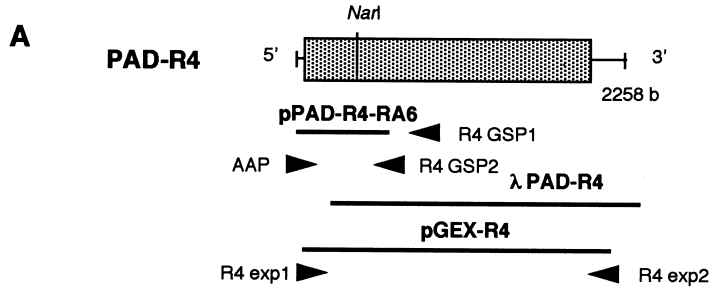
```
1    ATCAGAACGCACGCTCCTACCCGGCAGCCGGGTGTCCCCAAGAAGAGCTGGGGATCCGGGGCACAGCTGAGGCTTGTGGCTGCTGGGTACCCAAGCTGACCCGATCAGGAGACTGGAAGC
                                                                                                                  *
121  TGGGCCCAGGATGGCCTCCCAGCGAGCTGTGCAGCTGTCTCTGAGAAAGCCTGCTCATGCCGTGTGTGTGATAGGTGTGGGGACACTGGTGGATGTGTACAGTGATGTGCCCAAAGGCGC
 1           M  A  S  Q  R  A  V  Q  L  S  L  R  K  P  A  H  A  V  C  V  I  G  V  G  T  L  V  D  V  Y  S  D  V  P  K  G  A

241  CAAGACCTTTGAGGTGTCCGGGAGCTCTGAGGTGAAGATCTACATGGTCTACGATCCCTCAAGAGTGGCAGAGCCCGCAGGCCCGGCCCACTGGCCCCTGGATGCCAATGTGGATGTGGT
 38   K  T  F  E  V  S  G  S  S  E  V  K  I  Y  M  V  Y  D  P  S  R  V  A  E  P  A  G  P  A  H  W  P  L  D  A  N  V  D  V  V

361  GGTGGTGGCGGACACAGTCAGTAAGAACTTACACGACCTCAAGGTAAAGGTATCATACTTCAAGTCCCAGGACACTGCCGCCCTGGCCCACAGCCTTCTCTACCTAACCGCTGTTGATGT
 78   V  V  A  D  T  V  S  K  N  L  H  D  L  K  V  K  V  S  Y  F  K  S  Q  D  T  A  A  L  A  H  S  L  L  Y  L  T  A  V  D  V

481  GTCCCTTGATGTCGACACAGGCCGCACAGGCAAGGTGAAGAAAGGCTCTGGTGACAAGAAAAACTGGCGCTGGGGCTCCGGGGGCTCCGGGGCTATCCTGTTGGTGAACTGTGACAGAGA
118   S  L  D  V  D  T  G  R  T  G  K  V  K  K  G  S  G  D  K  K  N  W  R  W  G  S  G  G  S  G  A  I  L  L  V  N  C  D  R  D

601  TGTGCATGGGTCCAGAGAGGACCTGCATGATAGCCATTTGAGGTCACTGGAGGACCTGCAGGACATGTCCCCAATGGTGTTGAGCTGTGACGGCCCTGATGATCTTTTTAAGAGCCACAA
158   V  H  G  S  R  E  D  L  H  D  S  H  L  R  S  L  E  D  L  Q  D  M  S  P  M  V  L  S  C  D  G  P  D  D  L  F  K  S  H  K

721  GCCGGTCTTGAAGGCATCTTTGCCGGATTCCAGAAGACTGGGGGTCTTTTGTGCCCGGGGTGGGACCTCCCTCTCCAACTACAAACAGGTGCTGGGGCCCCGTCATAACTCCTATGAGGT
198   P  V  L  K  A  S  L  P  D  S  R  R  L  G  V  F  C  A  R  G  G  T  S  L  S  N  Y  K  Q  V  L  G  P  R  H  N  S  Y  E  V

841  GGAAAGGCATCCTGGGGAGAGGACATTCAGTTCTACGTGGAAGGACTCGCCTTCCCTGATGCCAGTTTCTCAGGGCTGCTGTCTCTCAGTGTCAGCCTGGTAGACACACGGCCCCTCTC
238   E  R  H  P  G  E  R  D  I  Q  F  Y  V  E  G  L  A  F  P  D  A  S  F  S  G  L  L  S  L  S  V  S  L  V  D  T  R  P  L  S

961  CGAGGTGTTGGTCTTCACGGACAGCGTGACCTTCCGAGTGGCCCCCTGGATCATGACCCCCAACACCCAGCCGCCCCTGGAACTGTACGTGTGCAGCGTGACAGATATGCATGGCAGCAA
278   E  V  L  V  F  T  D  S  V  T  F  R  V  A  P  W  I  M  T  P  N  T  Q  P  P  L  E  L  Y  V  C  S  V  T  D  M  H  G  S  N

1081 CGACAAGTTCTTGGAGGACATGTCCCACCTGGCCACGAAGGCCAACTGCAAACTAGTGGTCTGTCCCCGGCTAGAGAATCGCAATGATCGATGGATCCAGGATGAGATGGAGTTTGGCTA
318   D  K  F  L  E  D  M  S  H  L  A  T  K  A  N  C  K  L  V  V  C  P  R  L  E  N  R  N  D  R  W  I  Q  D  E  M  E  F  G  Y

1201 CATTGATGCCCCTCACAAATCGTTCCCTGTGGTCTTTGATTCCCCCGAAACAGAGGCTTGAGGGACTTCGCCCTTAAGAAGATCCTGGGTCCTGACTTTGGATATGTGACCCGGGAGAT
358   I  D  A  P  H  K  S  F  P  V  V  F  D  S  P  R  N  R  G  L  R  D  F  A  L  K  K  I  L  G  P  D  F  G  Y  V  T  R  E  I

1321 CCAGTTCGCAGGTGCCTCTGGCCTGGATTCCTTCGGCAACCTGGATGTGAGCCCGCCAGTGAGGGTGGGCAACACAGAGTACCCCCTCGGCAGAATCCTCATCGGTGGCAACTTCCCTAA
398   Q  F  A  G  A  S  G  L  D  S  F  G  N  L  D  V  S  P  P  V  R  V  G  N  T  E  Y  P  L  G  R  I  L  I  G  G  N  F  P  K

1441 GTCCAGCGGGCGGCGGATGGCCAGGGTGGTGCGTGACTTCCTGCATGCCCAGCAGGTGCAGGCTCCGGTGGAACTCTACTCTGACTGGCTCTCCGTGGGCCACGTGGATGAGTTTCTGAG
438   S  S  G  R  R  M  A  R  V  V  R  D  F  L  H  A  Q  Q  V  Q  A  P  V  E  L  Y  S  D  W  L  S  V  G  H  V  D  E  F  L  S

1561 CTTCGTGCCCACCTCTGACCAAAAGGGCTTCCGGTTACTCCTGGCCAGCCCCAGCGCCTGCCTCCAGCTGTTCCAGGAGAAGAAGGAGGAGGGCTACGGGGAGGCCAAGCAGTTTGATGG
478   F  V  P  T  S  D  Q  K  G  F  R  L  L  L  A  S  P  S  A  C  L  Q  L  F  Q  E  K  K  E  E  G  Y  G  E  A  K  Q  F  D  G

1681 CCTCAAGCATAAGACAAAGAGAAGCATCAATGACATCTTAGCTGACAGGCACCTCAGGAGGGACAGTGCACATGTGCAGAAATGCATCGACTGGAACCGTGAGGTGCTAAAACAGGAGCT
518   L  K  H  K  T  K  R  S  I  N  D  I  L  A  D  R  H  L  R  R  D  S  A  H  V  Q  K  C  I  D  W  N  R  E  V  L  K  Q  E  L

1801 GGGCCTCTCCGAGAGCGACATCGTGGATATTCCACAGCTCTTCTTCCTGAAGGGAGCCTACGCTGAAGCCTTCTTTCCTGACATGGTCAATATGGTGGTCCTCGGCAAGTACCTGGGCAT
558   G  L  S  E  S  D  I  V  D  I  P  Q  L  F  F  L  K  G  A  Y  A  E  A  F  F  P  D  M  V  N  M  V  V  L  G  K  Y  L  G  I

1921 CCCCAAGCCCTTCGGGCCCATCATCAATGGCCGCTGCTGCCTGGAAGAGAAGGTGCGCTCCCTGCTGGAGCCGCTGGGTCTACGCTGCGTCTTTATCGACGACTTTCTCTTTTACCACCA
598   P  K  P  F  G  P  I  I  N  G  R  C  C  L  E  E  K  V  R  S  L  L  E  P  L  G  L  R  C  V  F  I  D  D  F  L  F  Y  H  Q

2041 ACTGCTTGGAGAGATCCATTGTGGCACTAATGTGCGGAGGAAACCTTTTGCCTTTAGGTGGTGGAACTCCGTGCCCTGAGCCTGCCCGACCACCACCCTCGCTGCCCTGCTGATGAGAAC
638   L  L  G  E  I  H  C  G  T  N  V  R  R  K  P  F  A  F  R  W  W  N  S  V  P

2161 CCCAGAGCAAGGGCAAGGGCCACGAACCAGCTTCCACCCTCCTGGAGCCTGAAAGCGGACTGACTACTGTGGTCAGCTGGGCTCCCGGAGGACAACACACTTCTTCGTTTCTGGAGTCGC
2281 CTGTCCTGCCCAGGAGACTATTGACACTTTTCCTTCACCCTTTCCCAGCCAGAGCCCCCAGAGGGTCAAGGTCGGAAGCTAAGCCTGTAGGCCACGCCCTACTTTGAGCTGTTTCCCCCC
2401 TCCCTTCTGGACACCTCTCTGATGCTGTTTCAGAAATGCCCACCATGCCAGGACCCTGGGGCACGAACACCATAGTTTCGTATGACTCCTGGATTGCGCAAGATGAAAAGGAAGATTTGG
2521 GGCGTCTTTTTTTCCCAGAGGACCTCAGGCCATGGCCCCCTGGCACCCTTCAGCTCAACGTAGGCAAACAAAGGCAGGAAATCCCAGTCCTTGAAAGCGGGCAGGCACAAAAGCCCCTCT
2641 AATGCCAAACTCCATCCAGTCCCACCTCTGCAGAATGCTTGTTACCTGTTGGGGCAGCCTTTAATTACGAGCCCTGGTGCTTCCTGTTGGAAGATGCTTCGGGGCATAGGGAGGCCCTCG
2761 GAAGTCTGCAAGCGTGGGTGGGGGTGGGGAAGGTGGGCAGGAAGGTCGCACCAGGAAGGCAGAGAGGTTGCCCATCTCGGGGTGGGGGAACCTCGAAGCCTGGAACTCTTGGCTGGGTG
2881 TTTGGGGCCCCTGAACCTGACGGCTGTTGGTCCAGATGTGAAGGGGATGGGGCCCAAGGTATTAGAAGATGGTGGGGTAGGCCATTAAGTTACATCTGGAGGTGAGGCTTAATTTCCTCT
3001 AATAGTCTCTAATTACTCCCTAACCTGCTGCAGGCAGTGGGAGGGAGGGCCCACACTCCCCCAGCAGCCTGGGACCCTACACAACACTTGTACAGCCCCAGTAAACCCCATGGCAGGCC
3121 TGTGTCCCCCCTCCCACAGTAGGACAGACAGCTGGCCTGGAGGCTAATAGCATTGGGTGAGGTCACAGAGACATGTGTCGGCAAGCAGGCAGGCCAACCGGCTGCAGGGCACGGCTGGCA
3241 GAGGCTTCCTCTGGTCAAGGAGCCACAACAGGCTGCTGGATTCTGTCTTTGTATTTTGGTAGCTTCTGGTGCCACCCCTGACTTGAGGCAGCTGCTGTCACTTGGTAACCACTGGACACT
3361 TAGTTAGCAAAGGCTTCCTTGGAGCTGTTAGGATAGGCATGGCCAGGTAGACATTGGCTCTTCCCCCTTCCAGAAGGTTCTGTGGAACTGGAACCCCCATTTGTAGGCTCCCCACACCAA
3481 CACCCTCCTCTATGATCATGGCCTGGCTGAGGTTAAGCCATGGAGAAGGAGAAGAAGAAGAAGATAAAGGGGGATGAGAAGGTTGGGGGCGGGGCCTGGCCTCTGAAGACTCAGGAGACT
3601 GACTACAGAGCCTCGGCAAGCATGGGAGCAAGTCTTCTGCGCCACATGAGGTGGGCTCTGGAGTCGCTACTCAGACCGTCTGGGGGCCTGGCTGACTGGCCACCCTCTCCCTCATCAATT
3721 AAAAACATCGTTGTGCATTC  poly A (21)
```

Fig. 1. Sequencing of rat PAD-R11. A: Schematic representation of the sequencing strategy. B: Nucleotide and deduced amino acid sequences of PAD-R11. Nucleotide and amino acid numbers are shown on the left. An asterisk denotes the position of the in-frame stop codon located upstream of the putative initiation codon. A consensus sequence for the polyadenylation signal is underlined. The sequence data will appear in the DDBJ/EMBL/GenBank Nucleotide Sequence Data Libraries under accession number AB010998.

R4-RA6 into the *Sac*II/*Nar*I site of λPAD-R4. A region containing the entire coding sequence of pPAD-R11 (nucleotides 131–2223) was amplified by PCR using 5′-CCGCTCGAGATTGAATTCCATATG-GCCTCCCAGCGAGCTGTGCA-3′ (R11 exp1) containing the underlined *Eco*RI site as a sense primer, and 5′-CCGGAATTC-GCGGCCGCTCAGTCCGCTTTCAGGCTCCAGG-3′ (R11 exp2) containing the underlined *Not*I site as an antisense primer. In a similar manner, a region containing the entire coding sequence of pPAD-R4 (nucleotides 17–2155) was amplified using 5′-ACGCGTCGACTC-GAATTCCATATGGCCCAGGGTGCGGTGATC-3′ (R4 exp1) containing the underlined *Eco*RI site and 5′-CCGGAATTC-GCGGCCGCAGATGTGGGTGAAGCAAAGCAAC-3′ (R4 exp2) containing the underlined *Not*I site as sense and antisense primers, respectively. The amplified DNA fragments were subcloned into a pGEX-6P vector (Pharmacia) containing a glutathione *S*-transferase (GST) sequence and its nucleotide sequence was confirmed by DNA sequencing. Plasmids pGEX-R11 and pGEX-R4 obtained were used to transform JM105 cells. The transformants were grown overnight and treated for 4 h with 0.1 mM isopropyl-1-thio-β-D-galactopyranoside at 20°C. The cells were then disrupted by sonication, and the recombinant proteins in the soluble fraction were adsorbed to Glutathione Sepharose 4B (Pharmacia). PAD moieties were released using PreScission Protease (Pharmacia) according to the supplier's instruction. PAD activities were determined by measuring the conversion of benzoyl-L-arginine ethyl ester (BAEE) or benzoyl-L-arginine (Bz-L-Arg) to the corresponding derivatives of citrulline [8], and the specific activity was determined as described [15].

2.4. Northern blot hybridization

The poly(A)$^+$ RNA was electrophoresed in a 0.8% agarose-formaldehyde gel and transferred to nylon membrane. cDNA fragments, nucleotides 3109–3523, 2023–2227 and 2409–2837 of PAD-R11, PAD-R4, and rat PAD type II cDNA, respectively, were amplified by PCR using appropriate primers for the preparation of ^{32}P-labeled probes. Hybridization was performed under the condition used for the plaque hybridization. The membrane was finally washed at 65°C in 1×SSPE containing 0.1% SDS. The signals were detected using a BASS 2500 image analyzer (Fuji Photo Film Co., Ltd.).



B

```
   1 TCCCTGCTGCCCGAGGATGGCCCAGGGTGCGGTGATCCATGTGGCCCCTGAGGAGCCCACCCATGCCGTGTGTGTGGTGGGCACGGCGACCCCGCTGGATATCCGTGGCTCTGCTCCAAG
   1               M  A  Q  G  A  V  I  H  V  A  P  E  E  P  T  H  A  V  C  V  V  G  T  A  T  P  L  D  I  R  G  S  A  P  R

 121 GGGCTCCACGTCCTTCAGCATCACAGCCTCCCCAGAAGTTGTTGTAGATGTCATCCATGGTCCTCCATCCAAGAAGAGTACCACGGGGGGCCTCCAAATGGCCTCTGGACCCCAAGCTGGA
  36  G  S  T  S  F  S  I  T  A  S  P  E  V  V  V  D  V  I  H  G  P  P  S  K  K  S  T  T  G  A  S  K  W  P  L  D  P  K  L  E

 241 GGTGACCCTACAGATGAAAGCAGCCAGCAGTAGGATAGATGACCAGAAGGTTCGAATTTCCTATTACGGACCCAAGACCTCATCAACCCAAGCCCTGCTCTACCTCACCGGGGTGGAACT
  76  V  T  L  Q  M  K  A  A  S  S  R  I  D  D  Q  K  V  R  I  S  Y  Y  G  P  K  T  S  S  T  Q  A  L  L  Y  L  T  G  V  E  L

 361 CTCCCTGAGTGCGGATGTCACCCGCACTGGCAAAGCGAAGCCAGCCCCAGCTGGGAAGGATCAGAGCACCTGGACCTGGGGCCCGGACGGTCACGGCGCCATCCTGTTGGTGAACTGTGA
 116  S  L  S  A  D  V  T  R  T  G  K  A  K  P  A  P  A  G  K  D  Q  S  T  W  T  W  G  P  D  G  H  G  A  I  L  L  V  N  C  D

 481 CAAGGAGGATCCCAAGTCCTCCGGAATGGACTTCGAGGACGACAAGGTCTTAGACAACAAAGACCTGCAGGACATGTCACCAATGACCCTAAGCACGAAGACTCCCAAAGACTTCTTTGA
 156  K  E  D  P  K  S  S  G  M  D  F  E  D  D  K  V  L  D  N  K  D  L  Q  D  M  S  P  M  T  L  S  T  K  T  P  K  D  F  F  D

 601 CAAGTATCAGTTGGTGCTGCAGGTGCCGAAGGCCAAGATGAACAAAGTGAGAGTCTTCCGGGCCACACGGGGCAAACTGCCATCCAGATACAAGGTGGTCCTAGGACCACAGCAGTTCTC
 196  K  Y  Q  L  V  L  Q  V  P  K  A  K  M  N  K  V  R  V  F  R  A  T  R  G  K  L  P  S  R  Y  K  V  V  L  G  P  Q  Q  F  S

 721 CCACCGCCTAGAGCTGCTGGGCGGCCAGCACAGCACGGACTTCTATGTGGAAGGCCTTGCTTTCCCTGACGCCGACTTCAAGGGGCTCATTCCCCTAACGATCTCCCTGTTGGACAAATC
 236  H  R  L  E  L  L  G  G  Q  H  S  T  D  F  Y  V  E  G  L  A  F  P  D  A  D  F  K  G  L  I  P  L  T  I  S  L  L  D  K  S

 841 TAACCCGGAGCTCCCTGAAGCCCTGGTGTTCCAAGACACCGTGATGTTCCGTGTGGCCCCCTGGATCATGACGCCCAACACTCAGCCCCCCCAGGAGGTATACGTGTGCAGGTTTTCCGA
 276  N  P  E  L  P  E  A  L  V  F  Q  D  T  V  M  F  R  V  A  P  W  I  M  T  P  N  T  Q  P  P  Q  E  V  Y  V  C  R  F  S  D

 961 CAATGAAGACTTCCTAAAGTCACTGGCTACTTTCACCAAGAAAGCCAAGTGCAAACTGACTGTGTGCCCAGAGGAGGAAAATCAAGATGACCAGTGGATGCAGGATGAAATGGAGATTGG
 316  N  E  D  F  L  K  S  L  A  T  F  T  K  K  A  K  C  K  L  T  V  C  P  E  E  E  N  Q  D  D  Q  W  M  Q  D  E  M  E  I  G

1081 CTACATACAGGCCCCCCACAAGACGCTGCCTGTGGTCTTTGACTCCCCGAGGGACAGAGGCCTGAAGGATTTCCCTGTCAAACGAGTCATGGGTCCAAATTTTGGCTATGTGACCCGAGG
 356  Y  I  Q  A  P  H  K  T  L  P  V  V  F  D  S  P  R  D  R  G  L  K  D  F  P  V  K  R  V  M  G  P  N  F  G  Y  V  T  R  G

1201 GCTCTATAGAGCAGAGGTCACTGGGCTGGATGCCTTTGGGAACCTGGAGGTGAGTCCCCCAGTCACTGTCAGAGGGAAGGAATACCCACTGGGCAGGATTCTCATCGGGAGCAGCGGTTA
 396  L  Y  R  A  E  V  T  G  L  D  A  F  G  N  L  E  V  S  P  P  V  T  V  R  G  K  E  Y  P  L  G  R  I  L  I  G  S  S  G  Y

1321 TTCCAGCAGTGAGAGCCGGGACATGCACCAGATCCTGCAGGACTTCCTGGGTGCCCAGCAGGTGCAGGCCCCTGTGAGACTCTTCTCCGATTGGCTCTTTGTGGGTCACGTGGATGAGTT
 436  S  S  S  E  S  R  D  M  H  Q  I  L  Q  D  F  L  G  A  Q  Q  V  Q  A  P  V  R  L  F  S  D  W  L  F  V  G  H  V  D  E  F

1441 CTTGAGCTTCGTCCCGGCACGGGGCAAGCAGGGCTTTCGGCTGCTACTGTCCAGCCCCAGAGCTTGCTACCAGATGTTCCAGGAGCTGCAGACCGAGGGCCACGGGGAGGCATCACTGTT
 476  L  S  F  V  P  A  R  G  K  Q  G  F  R  L  L  L  S  S  P  R  A  C  Y  Q  M  F  Q  E  L  Q  T  E  G  H  G  E  A  S  L  F

1561 CGAAGGACTCAAGAGGAAAAGGCAGACAATCAGTGACATTCTGTCCAGCCAGAAACTAAGAGACCAGAATGCCTACGTGGAGAGCTGTATTGACTGGAACCGTGAGGTGCTGAAGCGAGA
 516  E  G  L  K  R  K  R  Q  T  I  S  D  I  L  S  S  Q  K  L  R  D  Q  N  A  Y  V  E  S  C  I  D  W  N  R  E  V  L  K  R  E

1681 GTTGGGCCTGACGGAGGGCGACATCATTGATATACCCCAGCTCTTCAGGATTGTGGGAAACTCCAGAGGGAACCCGAAGGCTGAGGCCTTCTTCCCAAACATGGTGAACATGCTGGTACT
 556  L  G  L  T  E  G  D  I  I  D  I  P  Q  L  F  R  I  V  G  N  S  R  G  N  P  K  A  E  A  F  F  P  N  M  V  N  M  L  V  L

1801 GGGCAAGCACCTGGGCATCCCCAAGCCCTTTGGGCCCATCATCAACGGCCGCTGCTGCCTGGAGGAGAAGGTGTGCTCCCTGCTGGAGCCGCTGGGTCTGCACTGCACCTTCATCAACGA
 596  G  K  H  L  G  I  P  K  P  F  G  P  I  I  N  G  R  C  C  L  E  E  K  V  C  S  L  L  E  P  L  G  L  H  C  T  F  I  N  D

1921 CTTCTACAGCTATCACATGTACCACGGAGAGGTTCACTGTGGCACCAACGTGCGCAGGAAGCCCTTCGCCTTCAAGTGGTGGCACATGGTTCCCTGAGTCACCGCTCCTGGCAGCCTCCC
 636  F  Y  S  Y  H  M  Y  H  G  E  V  H  C  G  T  N  V  R  R  K  P  F  A  F  K  W  W  H  M  V  P

2041 TCGAGGACTGCCCTGGCTCCCTGCAGGGAGATACAGTGAGAGATCTGGGGCCTTGGGGCTTCCCTAGAGGTGGCTACCCCTCCCTGCTGTGAGTTGCTTTGCTTCACCCACATCTGGCTC
2161 CTGTCCCACCTCTGTATGTCCTAAAATGGTCCTTGCATAGCAAACTCAGTCCTGCTTAGAGATGCTGCAATAAAGATGTGGTTTTGTATCTTTATTTT poly A (23)
```

Fig. 2. Sequencing of rat PAD-R4. A: Schematic representation of the sequencing strategy. B: Nucleotide and deduced amino acid sequences of PAD-R4. Nucleotide and amino acid numbers are shown on the left. A consensus sequence for the polyadenylation signal is underlined. The sequence data will appear in the DDBJ/EMBL/GenBank Nucleotide Sequence Data Libraries under accession number AB010999.

3. Results

3.1. Cloning and sequencing of PAD cDNA clones

Nine positive clones were characterized by restriction mapping and partial sequencing. These clones were classified into three groups. Representative clones of these groups bearing the longest insert sizes were completely sequenced. They were termed λPAD-R5 (~0.9 kbp), λPAD-R11 (~3.2 kbp), and λPAD-R4 (~2.1 kbp). The other clones had overlapping restriction maps with one of these representative cDNAs. λPAD-R5 was not used for further characterization, since it consisted of 875 bp which exactly coincided with the 57 bp coding sequence and the contiguous 818 bp 3′-untranslated sequences of rat PAD type II cDNA. λPAD-R11 (3172 bp) and λPAD-R4 (2116 bp) contained sequences encoding 516 and 512 amino acid residues, respectively, both showing similarities to the coding sequence of rat PAD type II (Figs. 1A and 2A, respectively). Such similarities were not found in the adjacent 3′-untranslated sequences. Sequences missing from the 5′-end regions of the relevant mRNAs were estimated by the 5′-RACE method. This yielded pPAD-R11-RA3 (759 bp) for λPAD-R11 and pPAD-R4-RA6 (528 bp) for λPAD-R4. These combinations constructed cDNA sequences PAD-R11 (3740 bp) and PAD-R4 (2258 bp), respectively.

As Fig. 1B shows, PAD-R11 contains an open reading frame of 1989 nucleotides encoding a polypeptide of 662 amino acid residues with a calculated molecular mass of 73 856 Da. The putative initiation codon (nucleotides 131–133) is located within a sequence context favorable for Kozak's rule [16], and is preceded by an in-frame stop codon (nucleotides 68–70). A putative polyadenylation signal (AATTAAA) [17] is located at nucleotides 3717–3723. As Fig. 2B shows, PAD-R4 contains an open reading frame of 2001 nucleotides encoding a polypeptide of 666 amino acids, with a calculated molecular mass of 74 466 Da. The putative initiation codon (nucleotides 17–19) is located within a sequence context favorable for Kozak's rule. No in-frame stop codon is present in the estimated sequence. A putative polyadenylation signal (AATAAA) is located at nucleotides 2229–2234.

3.2. Comparison of sequence similarity

The amino acid sequences of PAD-R11 and PAD-R4 are compared with those of rat PAD types II and III in Fig. 3. About 50% of amino acid residues in the C-terminal halves are identical, while only about 24% of amino acid residues are common in the N-terminal halves. The entire amino acid sequences of PAD-R11 and PAD-R4 showed 50–57% identity with that of rat PAD type II or type III when compared in pairs. The corresponding nucleotide sequences showed a somewhat higher identity (61–65%).

3.3. PAD activities of recombinant proteins

To construct GST expression plasmids, the coding regions of PAD-R11 and PAD-R4 were amplified and inserted into

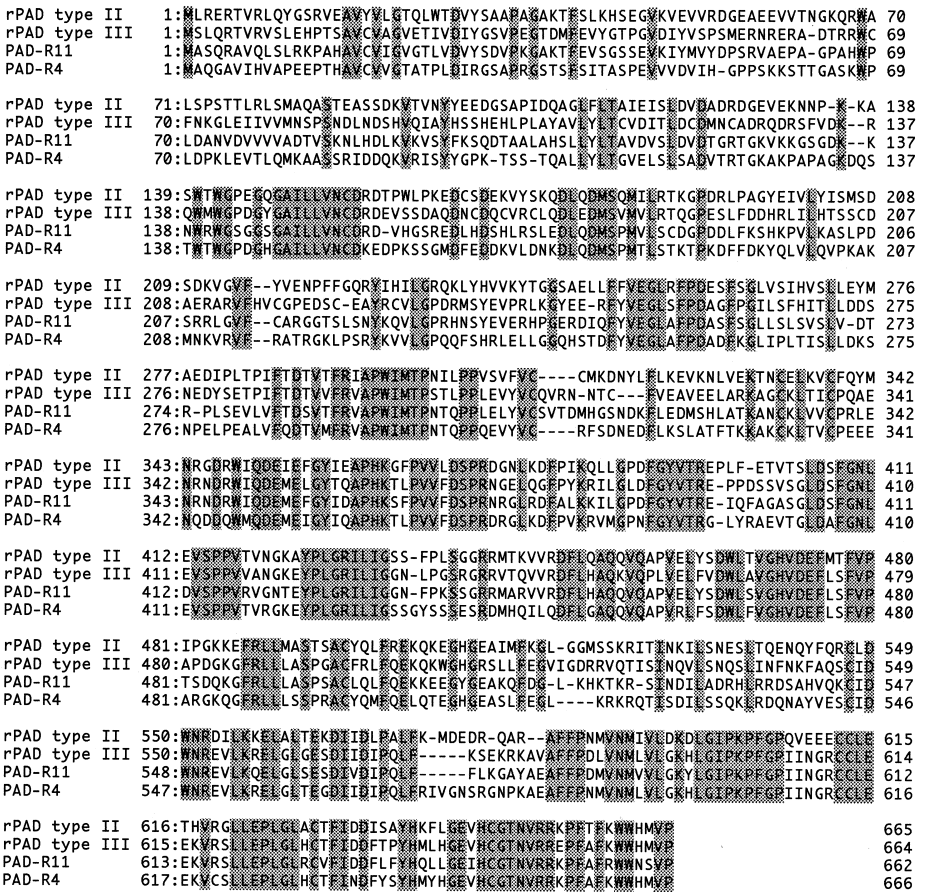


Fig. 3. Sequence comparison of PAD-R11 and PAD-R4 with rat PAD type II and type III. Alignment of putative amino acid sequences of PAD-R11 and PAD-R4 with rat PAD type II and type III. Amino acids conserved in all proteins are shadowed. Dashes indicate gaps inserted to maximize alignment.

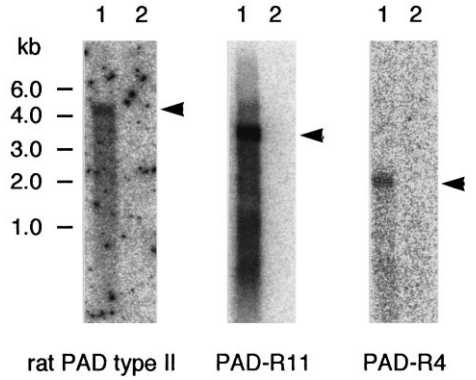


Fig. 4. Expression of the PAD type II, PAD-R11 and PAD-R4 genes in RA-treated cells. Poly(A)$^+$ RNA fractions were obtained for Northern blot hybridization with ^{32}P-labeled probes representing rat PAD type II (left), PAD-R11 (center) and PAD-R4 (right) as described in the text. Lane 1, poly(A)$^+$ RNA (1 μg) obtained from RA-treated cells; lane 2, poly(A)$^+$ RNA (1 μg).

the pGEX-6P vector. The PAD moieties of these recombinant proteins in bacteria were obtained by the affinity adsorption followed by the PreScission Protease cleavage. This resulted in about 50-fold purification of the recombinant protein encoded by PAD-R11 from the crude extract. The recombinant protein encoded by PAD-R4 was purified several-fold. The former showed specific activities of 2.1 to BAEE and 2.4 to Bz-L-Arg, in terms of units of the enzyme activity per mg protein. Those the latter were 6.3 and 1.1, respectively. No enzyme activities were detected in the extract of the bacteria transformed with the control pGEX-6P vector.

3.4. Analysis of mRNA expression

We performed Northern blot hybridization of poly(A)$^+$ RNA fractions to see if relevant mRNAs are specifically expressed in RA-treated cells (Fig. 4). Appropriate regions in the 3′-untranslated sequences were selected for probing to minimize cross-hybridization. The selected probes representing rat PAD type II cDNA, PAD-R11, and PAD-R4 yielded a single band of transcript ∼4.5 kbp, ∼3.8 kbp and ∼2.3 kbp, respectively. No significant hybridization signals were detected with the poly(A)$^+$ RNA fraction obtained from untreated cells.

4. Discussion

Our principal aim was to obtain a cDNA clone encoding catalytically active PAD type I. We chose RA-treated immortalized rat keratinocytes as the source of cDNA, which were found to synthesize multiple PADs including PAD type I and type II [15]. We obtained three different groups of PAD cDNA clones based on the homology to rat PAD type II cDNA. One group of cDNA clones represented by λPAD-R5 appeared to be derived from PAD type II mRNA itself. Sequencing of the representative clones and 5′-terminal extension using the 5′-RACE method enabled us to construct two cDNA sequences, PAD-R11 and PAD-R4, that have not been reported previously. To distinguish which cDNA encodes PAD type I, we constructed bacterially expressed recombinant proteins to measure their relative activities towards the synthetic substrates. The purified recombinant protein encoded by PAD-R11 showed nearly equal activities towards BAEE and Bz-L-Arg. Since this is a unique feature of PAD type I among the three known types of PAD [8,9], PAD-R11 was identified to encode PAD type I. The purified recombinant protein encoded by PAD-R4 showed a higher activity towards BAEE than to Bz-L-Arg, resembling PAD types II and III. We tentatively call it type IV PAD cDNA. These three different PAD mRNAs were confirmed to be expressed in RA-treated cells but not in untreated cells by Northern blot analyses. Since RA is known to negatively modulate various terminal differentiation processes such as synthesis and processing of profilaggrin [18], and expression of keratins [19] and loricrin [20,21], it is not certain if PAD types II and IV are expressed in animal epidermis. Such questions should be answered by immunocytochemical approaches using specific antibodies and/or in situ hybridization analyses using relevant cDNA sequences.

Deiminated forms of keratins and filaggrin in the cornified layer were found as the reaction products of PAD type I in our previously studies [1,2]. Trichohyalin was shown to be the natural substrate for PAD type III and found in highly deiminated forms in the mature hardened layers of the inner root sheath of hair follicles and the medulla of hair [22,23]. Recently, we reported that vimentin was preferentially deiminated in mouse peritoneal macrophages undergoing calcium ionophore-induced apoptosis, although the PAD type involved was not identified precisely [24]. These data suggest that PADs may act on intermediate filament proteins and/or intermediate filament-associated proteins in cells undergoing degenerative processes. The relatively high sequence conservation in the C-terminal region suggests its possible involvement in the common functions such as the catalytic activity and the calcium binding activity. The N-terminal region might be involved in the selective recognition of target proteins in relevant tissues. The availability of multiple types of PAD cDNA sequences will open a fascinating avenue by which to explore the regulation of their expression, as well as the molecular characterization of PADs themselves.

Acknowledgements: This work was partly supported by Grant-in-Aid 07680703 from the Ministry of Education, Science, Sports and Culture, Japan.

References

[1] Senshu, T., Akiyama, K., Kan, S., Asaga, H., Ishigami, A. and Manabe, M. (1995) J. Invest. Dermatol. 105, 163–169.
[2] Senshu, T., Kan, S., Ogawa, H., Manabe, M. and Asaga, H. (1996) Biochem. Biophys. Res. Commun. 225, 712–719.
[3] Rogers, G.E., Harding, H.W.J. and Llewellyn-Smith, I.J. (1977) Biochim. Biophys. Acta 495, 159–175.
[4] Kubilus, J., Waitkus, R. and Baden, H.P. (1980) Biochim. Biophys. Acta 615, 246–251.
[5] Fujisaki, M. and Sugawara, K. (1981) J. Biochem. (Tokyo) 89, 257–263.
[6] Kubilus, J. and Barden, H.P. (1983) Biochim. Biophys. Acta 745, 285–291.
[7] Takahara, H., Oikawa, Y. and Sugawara, K. (1983) J. Biochem. (Tokyo) 94, 1945–1953.
[8] Watanabe, K., Akiyama, K., Hikichi, K., Ohtsuka, R., Okuyama, A. and Senshu, T. (1988) Biochim. Biophys. Acta 966, 375–383.
[9] Terakawa, H., Takahara, H. and Sugawara, K. (1991) J. Biochem. (Tokyo) 110, 661–666.
[10] Senshu, T., Akiyama, K., Nagata, S., Watanabe, K. and Hikichi, K. (1989) Endocrinology 124, 2666–2670.

[11] Takahara, H., Tsuchida, M., Kusubata, M., Akutsu, K., Tagami, S. and Sugawara, K. (1989) J. Biol. Chem. 264, 13361–13368.
[12] Watanabe, K. and Senshu, T. (1989) J. Biol. Chem. 264, 15255–15260.
[13] Rogers, G., Winter, B., McLaughlan, C., Powell, B. and Nesci, T. (1997) J. Invest. Dermatol. 108, 700–707.
[14] Nishijyo, T., Kawada, A., Kanno, T., Shiraiwa, M. and Takahara, H. (1997) J. Biochem. (Tokyo) 121, 868–875.
[15] Ishigami, A., Ohsawa, T., Watanabe, K. and Senshu, T. (1996) Biochem. Biophys. Res. Commun. 223, 299–303.
[16] Kozak, M. (1991) J. Biol. Chem. 266, 19867–19870.
[17] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.
[18] Asselineau, D., Dale, B.A. and Bernard, B.A. (1990) Differentiation 45, 221–229.
[19] Breitkreutz, D., Stark, H.J., Plein, P., Baur, M. and Fusenig, N.E. (1993) Differentiation 54, 201–217.
[20] Magnaldo, T., Bernerd, F., Asselineau, D. and Darmon, M. (1992) Differentiation 49, 39–46.
[21] Brown, L.J., Geesin, J.C., Rothnagel, J.A., Roop, D.R. and Gordon, J.S. (1994) J. Invest. Dermatol. 102, 886–890.
[22] Lee, S.-C., Kim, I.-G., Marekov, L.N., O'Keefe, E.J., Parry, D.A. and Steinert, P.M. (1993) J. Biol. Chem. 268, 12164–12176.
[23] Fietz, M.J., McLaughlan, C.J., Campbell, M.T. and Rogers, G.E. (1993) J. Cell Biol. 121, 855–865.
[24] Asaga, H., Yamada, M. and Senshu, T. (1998) Biochem. Biophys. Res. Commun. 243, 641–646.