Two novel human members of an emerging mammalian gene family related to mono-adp-ribosylating bacterial toxins

Published in:
Genomics

DOI:
10.1006/geno.1996.4520

Citation for published version (APA):
Two Novel Human Members of an Emerging Mammalian Gene Family Related to Mono-ADP-Ribosylating Bacterial Toxins

FRIEDRICH KOCH-NOLTE,*1 FRIEDRICH HAAG,* RICKMER BRAREN,* MAREN KÜHL,* JAN HOOVERS,† SRIRAM BALASUBRAMANIAN,‡ FERNANDO BAZAN,‡ AND HEINZ-GÜNTER THIELE*

* Department of Immunology, University Hospital, D-20246 Hamburg, Federal Republic of Germany; †Department of Human Genetics, Academic Medical Center, Amsterdam, the Netherlands; and ‡DNAX Research Institute of Molecular & Cellular Biology, Palo Alto, California

Received May 14, 1996; accepted November 18, 1996

Mono-ADP-ribosylation is one of the posttranslational protein modifications regulating cellular metabolism, e.g., nitrogen fixation, in prokaryotes. Several bacterial toxins mono-ADP-ribosylate and inactivate specific proteins in their animal hosts. Recently, two mammalian GPI-anchored cell surface enzymes with similar activities were cloned (designated ART1 and ART2). We have now identified six related expressed sequence tags (ESTs) in the public database and cloned the two novel human genes from which these are derived (designated ART3 and ART4). The deduced amino acid sequences of the predicted gene products show 28% sequence identity to one another and 32–41% identity vs the muscle and T cell enzymes. They contain signal peptide sequences characteristic of GPI anchorage. Southern Zoo blot analyses suggest the presence of related genes in other mammalian species. By PCR screening of somatic cell hybrids and by in situ hybridization, we have mapped the two genes to human chromosomes 4p14–p15.1 and 12q13.2–q13.3. Northern blot analyses show that these genes are specifically expressed in testis and spleen, respectively. Comparison of genomic and cDNA sequences reveals a conserved exon/intron structure, with an unusually large exon encoding the predicted mature membrane proteins. Secondary structure prediction analyses indicate conserved motifs and amino acid residues consistent with a common ancestry of this emerging mammalian enzyme family and bacterial mono(ADP-ribosyl)transferases. It is possible that the four human gene family members identified so far represent the "tip of an iceberg," i.e., a larger family of enzymes that influences the function of target proteins via mono-ADP-ribosylation.

INTRODUCTION

Mono-ADP-ribosylation is a posttranslational protein modification that involves the transfer of the ADP-ribose moiety from NAD+ to a specific amino acid in a target protein while the nicotinamide moiety is released (Jacobson and Jacobson, 1989). Mono-ADP-ribosylation was originally discovered as the mechanism by which diphtheria toxin blocks protein synthesis (Honjo et al., 1968). A number of other potent bacterial toxins, including cholera, pertussis, and Escherichia coli heat labile enterotoxins, interfere with signal transduction in human host cells by ADP-ribosylating regulatory G-proteins (Aktories, 1991; Moss and Vaughan, 1990). Since cellular protein functions are profoundly affected by ADP-ribosylation, these toxins have found wide application in cell and molecular biology, e.g., as recombinant toxins for the directed killing of specific cells (diphtheria and pseudomonas toxin) and to identify proteins involved in signal transduction (cholera and pertussis toxins).

The suspicion that mono-ADP-ribosylation might also be used as a mechanism to regulate endogenous protein functions was first corroborated in photosynthetic bacteria. Fixation of nitrogen is essential for all living organisms. In Rhodospirillum rubrum and Azospirillum braziliense this process is regulated by reversible ADP-ribosylation of the key enzyme dinitrogenase reductase (Ludden, 1994).

The concept that analogous mechanisms operate to regulate endogenous metabolism in higher vertebrates has long aroused the curiosity of biochemists and cellular biologists, and ample evidence has been put forth to support this hypothesis (Jacobson and Jacobson, 1989; Moss and Vaughan, 1990). The responsible enzymes, however, eluded molecular cloning until the groups of J. Moss and M. Shimoyama succeeded in purifying proteins with enzyme activities akin to those of bacterial mono(ADP-ribosyl)transferases from rabbit skeletal muscle and chicken bone marrow cells (Tsuchiya et al., 1994; Zolkiewska et al., 1992). Homology searches revealed significant sequence similarity of these enzymes to one other previously known eukaryotic protein: the heat-labile enterotoxin (Jacobson and Jacobson, 1989). Recently, two mammalian GPI-anchored cell surface enzymes with similar activities were cloned (designated ART1 and ART2). We have now identified six related expressed sequence tags (ESTs) in the public database and cloned the two novel human genes from which these are derived (designated ART3 and ART4). The deduced amino acid sequences of the predicted gene products show 28% sequence identity to one another and 32–41% identity vs the muscle and T cell enzymes. They contain signal peptide sequences characteristic of GPI anchorage. Southern Zoo blot analyses suggest the presence of related genes in other mammalian species. By PCR screening of somatic cell hybrids and by in situ hybridization, we have mapped the two genes to human chromosomes 4p14–p15.1 and 12q13.2–q13.3. Northern blot analyses show that these genes are specifically expressed in testis and spleen, respectively. Comparison of genomic and cDNA sequences reveals a conserved exon/intron structure, with an unusually large exon encoding the predicted mature membrane proteins. Secondary structure prediction analyses indicate conserved motifs and amino acid residues consistent with a common ancestry of this emerging mammalian enzyme family and bacterial mono(ADP-ribosyl)transferases. It is possible that the four human gene family members identified so far represent the "tip of an iceberg," i.e., a larger family of enzymes that influences the function of target proteins via mono-ADP-ribosylation.

© 1997 Academic Press

Sequence data from this article have been deposited with the GenBank/EMBL Data Libraries under Accession Nos. X98526 and X98527.

1 To whom correspondence and reprint requests should be addressed at Department of Immunology, Medical Clinic, University Hospital, D-20246 Hamburg, FRG. Telephone: 49/40/4717 3612. Fax: 49/40/4717 4243. E-mail: nolte@uke.uni-hamburg.de.
T-cell alloantigen RT6 (Koch et al., 1990). RT6 proteins from rat and mouse, indeed, have recently been shown to possess the enzyme activities predicted by the structural homology (Haag et al., 1995; Koch-Nolte et al., 1996; Maehama et al., 1995; Takada et al., 1994).

Human homologues of the genes for the muscle enzyme and RT6 have been identified (these are here designated ART1 and ART2, respectively) (Haag et al., 1994; Okazaki et al., 1994). Intriguingly, while the human ART2 gene evidently has been inactivated by premature stop codons (Haag et al., 1994), the human ART1 gene appears to be functional, and its gene product exhibits properties similar to those of the rabbit homologue (Okazaki et al., 1994; Zolkiewska et al., 1992). Here we describe the identification, molecular cloning, and chromosomal mapping of two novel human gene family members (designated ART3 and ART4).

Our results suggest that these encode GPI-anchored membrane proteins operating in human testis and spleen.

**MATERIALS AND METHODS**

Identification of ART1/ART2-related ESTs. The database of expressed sequence tags (dbEST) was screened for ART1- and ART2-homologous sequences using the tBLASTn program with the mail server blast@ncbi.nlm.nih.gov (Altschul et al., 1990). Sequences of candidate RT6-homologous genes were analyzed in detail by visual inspection and on a Macintosh personal computer with the DNAStar and MacMolly programs to judge the significance of observed sequence identities to ART1 and ART2. Clones for ESTs from the Merck/Washington Univ. project were requested from and kindly provided by G. Lennon, IMAGE consortium, Lawrence Livermore Lab. (lnl@ncbi.nlm.nih.gov).

PCR primers and PCR. Primers derived from ART3 and ART4 ESTs were as follows: ART3: B14, TTT TAC CTC ACA AGA GCC CTG and B34, TTT AAT GTC AAC TCC AAG GCA; ART4: L01, GGC CAA TTC CTC TCC ACA TCC and L32, ATA GGG AGG GAT CAA GAC TTC). PCR amplifications were carried out in 20-µl reaction volumes on purified genomic DNA (100–500 ng), cDNA (10–50 ng), P1 or PAC1 DNA (10–50 ng), and plasmid DNA (0.1–1 ng) in 1× PCR buffer (low salt buffer; Stratagene) with a combination of 2.5 U Taq and 0.5 U Pfu polymerase (Taq plus; Stratagene) for 25–30 cycles (20 s, 95°C; 20 s, 60°C; 60 s, 72°C).

Southern and Northern blot analyses. Genomic DNAs were prepared, restriction digested, and subjected to Southern blot analysis essentially as described previously (Haag et al., 1994). Northern blots were purchased from Clontech. ART3- and ART4-exonic probes were generated by PCR amplification from respective cloned DNAs. The reaction products were purified on Chromaspin columns (Clontech) and radiolabeled to high specific activity (>10⁶ cpm/µg) by the random primer labeling procedure (Amersham) according to the manufacturer’s instructions. Hybridization and washing were performed under conditions of moderate stringency (16 h in 7% SDS, 0.5 M sodium phosphate, 1 mM EDTA at 62°C and 3 × 15 min in 1% SDS, 40 mM sodium phosphate, 100 mM NaCl at 62°C, respectively). Bound probe was detected by autoradiography using X-Omat AR films (Kodak) at –80°C for 8 h–7 days. In some cases, this was followed by high-stringency washings (2 × 15 min in 1% SDS, 40 mM NaPhosphate at 70°C) and further autoradiography. Blots were stripped and reprobed with a radiolabeled cDNA probe for glyceraldehyde 3-phosphate dehydrogenase (G3PDH)(Clontech).

Isolation, restriction mapping, and sequencing of genomic and cDNA clones. Human genomic PAC1 and P1 DNA libraries (Genomic Systems, St. Louis, MO) were screened with PCR with ART3- and ART4-specific primers. Purified P1 and PAC DNAs were subjected to restriction digestion with a panel of restriction enzymes, size fractionated by agarose gel electrophoresis, and subjected to Southern blot analyses with ART3- and ART4-specific probes. Suitable restriction fragments were subcloned into pBluescript (Stratagene) and

![FIG. 1. Southern blot analyses of ART3 and ART4 genes in the human and other species. Human genomic DNA was digested with BamHI, BglII, EcoRI, HindIII, or Sall (a and b, lanes 1–5, respectively) and subjected to Southern blot analysis with radiolabeled ART3 (a)- and ART4 (b)-specific exonic probes. The ART3 probe contains an internal HindIII site (accounting for the two bands in a, lane 4) but no EcoRI site; the ART4 probe does not contain any sites for the enzymes used. Genomic DNAs from four members of our laboratory staff (c and d, lanes 1–4), rhesus monkey (lane 5), lemur monkey (lane 6), C57BL 10SnJ mouse (lane 7), tupai (lane 8), and Lewis rat (lane 9) were digested with EcoRI and subjected to Southern blot analysis with radiolabeled ART3 (c)- and ART4 (d)-specific probes. M, “Drigest” marker (Pharmacia).](image-url)
FIG. 2. ESTs for ART3 and ART4 map to human chromosomes 4p14–p15.1 and 12q13.2–q13.3, respectively. Genomic DNAs from human/rodent somatic cell hybrids (NIGMS mapping panel 2, Coriell Institute, Camden, NJ) were screened by PCR with ART3 (a)- and ART4 (b)-specific primers. Reaction products were analyzed by agarose gel electrophoresis in the presence of ethidium bromide. The single or two human chromosomes retained by each of the cell lines are indicated above the lanes. Control DNAs were from human (H), mouse (m), and hamster (h). M, “Drigest” marker (Pharmacia). In situ hybridization (c and d) was performed with labeled ART3 (c)- and ART4 (d)-specific probes (derived from respective PAC1 and P1 genomic DNA clones) on human prometaphase chromosomes at the 550-band stage. The 95% confidence intervals calculated from 20 FLpter (fractional length of the whole chromosome relative to pter) values were 0.14–0.17 for ART3 and 0.34–0.36 for ART4. By comparing these measurements to chromosome ideograms in the same state of contraction, the ART3 gene is assigned to 4p14–p15.1, the ART4 gene to 12q13.2–q13.3.

RESULTS AND DISCUSSION

Identification and molecular cloning of ART3 and ART4. Homology searches of dbEST (the database of expressed sequence tags) (Boguski, 1995) revealed four human ESTs from the Merck/Washington University project and one human EST from a project at INSERM (Pawlak et al., 1995) with noticeable sequence similarities to the known mammalian skeletal muscle and T cell mono(ADP-ribosyl)transferases (here designated ART1 and ART2) (Koch et al., 1990; Koch-Nolte et al., 1996; Okazaki et al., 1994; Zolkiewska et al., 1992). Primers derived from these ESTs were used to clone subjected to further restriction mapping and sequence analyses. Human spleen and testis “marathon” cDNAs were purchased from Clontech and subjected to PCR amplification with ART3- and ART4-specific primers. Amplification products were cloned into the pCRll vector (Invitrogen). Sequences obtained from PCR products were confirmed by sequencing cDNAs obtained from two separate PCR amplifications. The ART3 and ART4 nucleotide sequences have been deposited in the EMBL database (Accession Nos. X95827 and X95826, respectively).

In situ hybridization. In situ hybridization was performed with labeled ART3 PAC DNA and ART4 P1 DNA probes on human prometaphase chromosomes (550-band stage) essentially as described previously (Hoovers et al., 1992). The 95% confidence intervals were calculated from 20 FLpter (fractional length of the whole chromosome relative to pter) values. The genes were assigned to specific chromosome bands by comparing these measurements to chromosome ideograms in the same state of contraction.

Amino acid sequence alignment and secondary structure prediction analyses. Multiple sequence alignments were performed with a weighted dynamic programming method (HSSP/MaxHom), and the generated multiple alignments were used as input for secondary structure predictions that were produced by profile-based neural network systems (PHDsec) (Rost and Sander, 1993; Sander and Schneider, 1994). These analyses were performed with the Predictprotein mail server of EMBL (Predictprotein@EMBL-Heidelberg.DE). Hydrophathy profiles were generated on a Macintosh with the MacMolly software (Softgen, Berlin) by using the Kyte–Doolittle algorithm and a window setting of 19 amino acid residues.
the corresponding genes from P1 and PAC1 genomic DNA libraries. Sequence analyses revealed that ESTs H12146, R35364, and T19112 were derived from one gene (here designated ART3) and that T70606 and R07880 were derived from a different gene (designated ART4). Recently, the single ART-related EST from Nature’s genome directory (Adams et al., 1995) also became accessible in the public database. Sequence comparison reveals that this EST (L49677) evidently also derives from ART3.

ART3 and ART4 are single-copy genes located on chromosomes 4p14–p15 and 12q13.2–q13.3, respectively. Our Southern blot analyses (Fig. 1) with exonic ART3- and ART4-specific probes revealed bands for each of the enzymes tested consistent with single-copy genes except for the two EcoRI bands detected with the ART3 probe (Fig. 1a, lane 3). Probing EcoRI-digested genomic DNAs from several members of our laboratory staff revealed that these bands reflect an allelic polymorphism in the gene rather than the presence of a second gene copy (Fig. 1c, lanes 1 and 2 vs lanes 3 and 4). Zoo blot analysis indicated the presence of ART3 and ART4 homologues in other primates and in tupaias (tree shrew) (Fig. 1c). A cross-hybridizing band was observed also in mouse DNA with ART4 but not with ART3 (Figs. 1c and 1d).

PCR screening of human/rodent somatic cell hybrids revealed that ART3 and ART4 genes are located...
on chromosomes 4 and 12, respectively (Fig. 2a). In situ hybridization revealed a regional localization to bands 4p14–p15.1 and 12q13.2–q13.3, respectively (Fig. 2b). Thus, neither of these genes is linked to ART2P or ART1 on chromosome 11q13 and 11p15, respectively (Koch-Nolte et al., 1993, 1996).

The deduced ART3 and ART4 amino acid sequences exhibit conserved structural motifs and amino acid residues similar to those of ADP-ribosylating toxins. Hydrophathy profiles of the ART3 and ART4 deduced amino acid sequences reveal hydrophobic N-terminal and C-terminal signal peptides characteristic of GPI-anchored membrane proteins (Fig. 3a). While the signal peptides of ART3 and ART4 show almost no recognizable amino acid sequence identities, the predicted native proteins show 28% identity. The gene family member with the highest sequence similarity to ART3 is ART2 (41% identity); the closest relative of ART4 is ART1 (39% identity). The ART3 sequence contains one potential Asn-X-Ser/Thr glycosylation sites; ART4 contains five. Four conserved cysteine residues probably form conserved disulfide bonds.

We obtained secondary structure predictions for ART3 and ART4 by a profile-based neural network system (Rost and Sander, 1993; Sander and Schneider, 1994) based on multiple alignments (Fig. 3b). While the N-terminal portion of these proteins is predicted to be mainly helical, the C-terminal region is dominated by \( \beta \)-sheets and a single prominent \( \alpha \)-helix. This pattern of \( \beta \)-sheets and \( \alpha \)-helix is quite similar to the secondary structure motifs of the catalytic domain in bacterial toxins of known 3D structure (Allured et al., 1985; Bennet et al., 1994; Choe et al., 1992; Domenighini et al., 1994; Sixma et al., 1991; Stein et al., 1994). Moreover, three motifs (designated \( \beta 1, \beta 3-\alpha 3 \), and \( \beta 6 \) in Fig. 3) show recognizable sequence similarities to the secondary structure units that line the active site crevice in bacterial toxins (Koch-Nolte et al., 1996). These include, in particular, two residues that have been shown to be essential for catalysis in bacterial toxins (arginine in \( \beta 1 \) and glutamic acid in \( \beta 6 \)) (see Domenighini et al., 1994, for review). Similarly, rabbit ART1 and rodent ART2 do not tolerate even conservative substitutions of these residues (Takada et al., 1995; and our unpublished observations).

ART3 and ART4 have conserved exon/intron structures. Comparisons of genomic and cDNA sequences show that the N- and C-terminal signal peptides are encoded by separate exons in ART1–ART4, whereas the entire predicted processed polypeptide chains are encoded by a single unusually large exon. (As an exception to this rule, ART3 contains a large insertion of \( \geq 50 \) amino acids immediately preceding the GPI signal peptide that is encoded by the 3rd exon. This region includes an intriguing, very hydrophilic 10-residue repeat of unknown function.) It has been suggested that encoding of a functional polypeptide by a single large exon may reflect acquisition of this gene by horizontal gene transfer (Hawkins, 1988), an interesting possibility also for the vertebrate ART gene family, many of whose prokaryotic relatives, notably, are encoded by bacteriophages (Aktories, 1991).

ART3 and ART4 are specifically expressed in testis and spleen, respectively. Northern blot analyses revealed a prominent ART3-specific band of 1.7 kb and a weak band of 3.8 kb in testis RNA (Fig. 4a, lane 4). The 1.7-kb band was weakly detected also in intestine, spleen, heart, and skeletal muscle (Fig. 4a). Considering that two of the four ART3 ESTs originate from a human infant brain cDNA library and were evidently derived from processed transcripts, it is likely that ART3 is expressed also in the (infant) brain. As such, it is a candidate gene for mono(ADP-ribosyl)transferase activities that have been detected in the adult brain by biochemical means (Maehama et al., 1991; Schuman et al., 1994).

Probing the same blots with an ART4-specific probe revealed prominent bands of 1.4, 2.4, and 5.5 kb only in spleen RNA (Figs. 4b and 4c). The same bands were weakly detected also in intestine and ovary. Interestingly, spleen (T cells) is also the tissue in which ART2 is prominently expressed in rodents (Koch et al., 1990). It is conceivable that the ART4 gene product compensates for the universal loss of ART2 in the human species (Haag et al., 1994).
It is of note that the signal intensities of ART3- and ART4-specific bands were much lower than those for the housekeeping enzyme G3PDH (not shown). Possibly, this reflects a tight control of expression levels of these presumptive regulatory enzymes. A low transcription level of ART genes is also in accord with the fact that no ESTs for ART1 or ART2 have yet appeared in the public database, whereas 79% of the positionally cloned human genes are represented by exact DNA sequence matches with one or more ESTs in dbEST (Bassett et al., 1996).

Implications. The prediction that ART3 and ART4 are expressed as GPI-anchored membrane proteins raises the intriguing question of whether they have extra- or intracellular targets or both. If they target intracellular proteins as do their bacterial toxin cousins (Moss and Vaughan, 1990), how do they enter the cytoplasm? If they have extracellular targets as suggested for ART1 and ART2 (Maehama et al., 1995; Zolkiowska and Moss, 1993), how is access to the required substrate NAD⁺, a classic intracellular metabolite, assured? As cell surface proteins, the ART gene products will be accessible targets for inhibitors and antibodies. Moreover, it will be of interest to test whether recombiant forms of these intriguing enzymes may be tools for experimental interventions. Finally, it is not unlikely that ART1–ART4 represent the “tip of an iceberg,” i.e., a larger family of mammalian enzymes that influences the function of cellular proteins by mono-ADP-ribosylation.

Note added in proof. Results of our recent PCR and sequence analyses confirm that homologues of ART3 and ART4 exist also in Mus musculus.

ACKNOWLEDGMENTS

This work was supported in part by Grants Ko1114 and No310 from the Deutsche Forschungsgemeinschaft to F. K.-N. We thank G. Guellean, Creteil, France, for sharing data on testis EST T19112 before publication. We thank the IMAGE consortium at the Lawrence Livermore Lab. for generously providing EST clones. We thank R. Rümpler, Strasbourg, France, for providing tissue samples from lemur monkeys and D. von Holst, Bayreuth, FRG, for providing samples from tupaia (tree shrews). We thank Marja Jakobs for excellent technical assistance.

REFERENCES


Okazaki, I. J., Zolkiowska, A., Nightingale, M. S., and Moss, J.