Diversifying Selection of the Tumor-Growth Promoter Angiogenin in Primate Evolution

Jianzhi Zhang¹ and Helene F. Rosenberg

Laboratory of Host Defenses, National Institute of Allergy and Infectious Diseases, National Institutes of Health

Diversifying selection drives the rapid differentiation of gene sequences and is one of the main forces behind adaptive evolution. Most genes known to be shaped by diversifying selection are those involved in host-pathogen or male-female interactions characterized as molecular “arms races.” Here we report the unexpected detection of diversifying selection in the evolution of a tumor-growth promoter, angiogenin (ANG). A comparison among 11 primate species demonstrates that ANG has a significantly higher rate of nucleotide substitution at nonsynonymous sites than at synonymous sites, a hallmark of positive selection acting at the molecular level. Furthermore, we observed significant charge diversity at the molecular surface, suggesting the presence of selective pressures in the microenvironment of ANG, including its binding molecules. A population survey of ANG in chimpanzees, however, reveals no polymorphism, which may have resulted from a recent selective sweep of a charge-altering substitution in chimpanzee evolution. Functional assays of recombinant ANGs from the human and owl monkey indicate that primate ANGs retain angiogenic activity despite rapid evolution. Our study, together with findings of similar selection in the primate breast cancer suppressor gene, BRCA1, reveals an intriguing phenomenon of unusual selective pressures on, and adaptive evolution of, cancer-related genes in primate evolution.

Introduction

Although cases of positive Darwinian selection and adaptive evolution are well documented at the morphological and physiological levels (Darwin 1859), they are found only infrequently at the molecular level (Hughes 1999), reflecting the conservative nature of the evolutionary histories of most genes (Kimura 1983; Nei 1987). There are two types of positive selection pressures that describe events at the molecular level: directional selection and diversifying selection. Directional selection promotes a substantial change in gene function, often resulting in a new biochemical activity. As but one example, a novel antipathogen activity emerged in a duplicated ribonuclease (RNase) gene of primates, as a result of a sudden and dramatic increase in the number of encoded arginine residues (Zhang, Rosenberg, and Nei 1998). In contrast, diversifying selection accelerates the alteration of gene sequences of different alleles, species, or paralogous genes in a nondirectional manner, which often widens the spectrum of the ligands or substrates that can be recognized, but rarely changes the overall protein function. Almost all genes and gene families known to be affected by diversifying selection belong to one of two groups. The first group includes host-defense genes and their counterparts in pathogens, such as the immunoglobulin genes (Tanaka and Nei 1989), major histocompatibility complex (MHC) genes (Hughes and Nei 1988), and the human influenza virus hemagglutinin gene (Fitch et al. 1991). Sequence variation in pathogen genes is presumably favored by diversifying selection because new mutations may confer resistance to immune surveillance. At the same time, diversity in host-defense genes may increase the number and variety of pathogens recognized by the host and thus may also be selectively favored. Genes involved in prey-predator interactions, such as the conotoxin genes of marine gastropods (Duda and Palumbi 1999) and venom gland phospholipase genes of snakes (Nakashima et al. 1995), may also be included in this group. The second group of genes evolving under diversifying selection includes those involved in sexual reproduction, such as the protamines of primates (Rooney and Zhang 1999; Wyckoff, Wang, and Wu 2000), accessory gland proteins of flies (Tsaur, Ting, and Wu 2001), and gamete recognition proteins of abalone and sea urchin (Lee, Ota, and Vacquier 1995; Metz and Palumbi 1996). The selective pressures are presumed to relate in some fashion to issues of male-female conflict or compatibility, although the exact nature of these selective pressures is not well understood.

Unexpectedly, a recent study revealed what appears to be diversifying selection acting on a tumor suppressor gene, BRCA1, in the evolution of humans and chimpanzees (Huttley et al. 2000). In this report, we describe yet another case of diversifying selection in a cancer-related gene, angiogenin (ANG). ANG, also known as RNase 5, is a member of the RNase A superfamily (reviewed in Riordan 1996). ANG was originally isolated from human tumor cell-conditioned medium based on its ability to stimulate the formation of new blood vessels (Fett et al. 1985). This activity, as opposed to tumorigenesis, is a more relevant physiologic function from an evolutionary perspective. However, ANG expression is elevated in various tumors, and its activity is related to cancer progression (Shimoyama et al. 1996; Montero et al. 1998; Miyake et al. 1999; Shimoyama et al. 1999; Shimoyama and Kaminishi 2000), with its antagonists having the ability to inhibit cancer growth in vivo (Olson et al. 1994, 1995; Piccoli et al. 1998). We
begin by examining the phylogenetic relationships of the paralogous ANG genes of various mammalian species and then provide evidence for diversifying selection of the single-copy ANG gene in primates.

Materials and Methods

DNA Samples

The genomic DNAs of the gorilla (Gorilla gorilla), orangutan (Pongo pygmaeus), baboon (Papio hamadryas), rhesus monkey (Macaca mulatta), green monkey (Cercopithecus aethiops), talapoin monkey (Miopithecus talapoin), owl monkey (Aotus trivirgatus), squirrel monkey (Saimiri sciureus), and tamarin (Saguinus oedipus) were either purchased or extracted from cells of various cell lines. DNA samples of 35 humans (Homo sapiens) were from Coriell Cell Repositories and Dr. David McDermott of NIH, and those of 13 chimpanzees (Pan troglodytes) were from the San Diego Zoo, Yerkes Regional Primate Research Center, and Coriell Cell Repositories.

Genotyping

The ANG coding region was amplified from genomic DNAs of 11 primate species by PCR with primers ANG-5 (5'-GTGTGTGGAAGAGATGGTGATG-GGC-3') and ANG-3 (5'-AGCACCTTGACCAAGGGGCCTG GTTA-3'). The first four codons of the ANG gene were encoded in the ANG-5 primer and were excluded from subsequent analysis. The PCR products were cloned into a PCR II TA cloning vector (Invitrogen, San Diego, Calif.) and sequenced from both directions by the dideoxy automatic sequencer. The polymorphic alleles were obtained by PCR of 35 humans (8 Chinese, 5 Native Americans, 10 Europeans, and 12 African Americans) and 13 chimpanzees (5 obtained from San Diego Zoo, 1 from Coriell, and 7 from Yerkes), using the high fidelity Taq (Life Technologies, Rockville, Md.) with primers 253 (5'-GGTATGTCTTAATGGCCTCAGG-3') and ANG-3. The PCR products were subject to direct sequencing, as described in Zhang and Rosenberg (2000).

Evolutionary Analysis

Sequences of nonprimate ANGs were obtained from GenBank (accession numbers follow genus/species classification—rabbit Oryctolagus cuniculus: P31347; cattle Bos taurus: P10152; calf2: P080929; pig Sus scrofa: P31346; mouse Mus musculus: NMJ007447; mouse2: NMJ007449; mouse3: U72672; chicken Gallus gallus: X61193). Two ANG pseudogenes and an EST from the mouse (reviewed in Strydom 1998) were not included in the analysis. Phylogenetic trees were reconstructed by the neighbor-joining method (Saitou and Nei 1987) with 1,000 bootstrap replications (Felsenstein 1985). The MEGA2 program (Kumar et al. 2000) was used for phylogenetic analysis. The numbers of synonymous (dS) and nonsynonymous (dN) nucleotide substitutions per site between ANG sequences were computed as described in Zhang, Rosenberg, and Nei (1998), with an estimated transition-transversion ratio (Kimura 1980) of 1.2. The numbers of synonymous and nonsynonymous substitutions per site for each tree branch were estimated from the pairwise distances, using the least-squares method with a given tree topology (Zhang, Rosenberg, and Nei 1998). The ancestral sequences were inferred by the distance-based Bayesian method (Zhang and Nei 1997). Because the species concerned are closely related, the accuracy of the ancestral inference is greater than 96% for all nodes. The potential numbers of synonymous (S), nonsynonymous (N), conservative nonsynonymous (C), and radical nonsynonymous (R) sites as well as the observed substitutions (s, n, c, r) at these sites for each branch were estimated (Zhang, Rosenberg, and Nei 1998; Zhang 2000). Binomial tests were used to test the substitution rate difference at various types of sites. For instance, sN is compared with the binomial distribution B(s + N, N/[S + N]) to test if the nonsynonymous substitution rate is significantly greater than the synonymous rate for the entire tree, where s and n are sums of s and n over all tree branches. In the present case, the binomial test is more appropriate and more conservative than the test described in Zhang, Kumar, and Nei (1997) because multiple substitutions at individual sites can be taken into account. Following Zhang, Kumar, and Nei (1997), Fisher's test was used to test the difference in n/s among different branches (i.e., the episodic evolution hypothesis; Messier and Stewart 1997). DnaSP (Rozas and Rozas 1999) was used for estimating the nucleotide diversity π and its sampling error (Nei 1987) and for performing the Tajima's (1989) and Fu and Li's (1993) tests of neutrality. The age of the most recent common ancestor of a sample of alleles was estimated following the method of Fu and Li (1996).

Recombinant ANGs and the Angiogenic Activity Assay

The ANG genes of the human and owl monkey were subcloned into the bacterial expression vector pFLAG CTS (Kodak, New Haven, Conn.) and were verified by sequencing. The vector adds the octapeptide DYKDDDDK (FLAG) to the recombinant protein, which facilitates its purification and detection with M2 anti-FLAG monoclonal antibody. Recombinant proteins were isolated, purified, and quantified as described in Rosenberg and Dyer (1995). The angiogenic activity of the recombinant ANGs was tested using the chicken embryo chorioallantoic membrane (CAM) assay (Fett et al. 1985; Gho, Kleinman, and Sosne 1999). That is, salt-free aqueous solution containing recombinant ANGs was loaded onto a Thermonox disc (Nunc, Naperville, Ill.). The sample was dried under sterile air before being applied to the CAM of a chicken embryo aged 10 days. After 72-h incubation at 37°C, positive (appearance of a spokewheel pattern) or negative responses were assessed under a microscope. Each sample was tested in at least two independent experiments. The numbers of positive and negative responses for each sample from
multiple experiments were combined, and a G-test was used to test the difference in angiogenic activity between samples. Water was used as the negative control.

**Results**

**Phylogeny of Paralogous ANG Genes**

Three copies of ANG genes have been reported for the mouse and two for the cattle (Nobile, Vallee, and Shapiro 1996; Strydom, Bond, and Vallee 1997; Fu et al. 1999), whereas one ANG gene is known in every other species surveyed. With the chicken ANG as an outgroup, we reconstructed the phylogenetic tree of the currently available mammalian ANG sequences (fig. 1A). This tree strongly suggests that the cattle ANG2 gene was generated through an ancient gene duplication predating the diversification of mammalian orders (primates, artiodactyls, lagomorphs, and rodents) about 100 MYA, and the bootstrap support for this branching pattern is 97%. The tree also suggests that the three mouse genes were the result of recent gene duplications in rodents (fig. 1A). However, the bootstrap support for the clustering of all three mouse ANG genes is low (57%), and further evidence is needed to corroborate this result. No orthologs of the cattle ANG2 are known in other organisms. Our search of the draft human genome sequence also resulted in only one ANG gene. These results suggest that the ancient ANG2 might have been lost in most mammalian lineages, except in cattle. It remains to be seen whether ANG2 exists in other artiodactyls.

**Positive Selection of Primate ANG Gene**

We identified one ANG gene in each of the 11 higher primate species examined, and a phylogeny of these 11 ANG sequences (fig. 1B) is in general agreement with the known phylogeny of the species (fig. 2), indicating that these genes are orthologous. Rapid evolution of ANG is evident from the high sequence variation shown in the alignment of the primate sequences (fig. 3). The number of nonsynonymous substitutions per site \( (d_{ns}) \) is greater than that of synonymous substitutions per site \( (d_s) \) for 50 of the 55 pairs among the 11 sequences (fig. 4), suggesting that rapid substitution at nonsynonymous sites is a general feature of ANG in higher primates. To test whether the rapid evolution is because of positive Darwinian selection, we analyzed the ANG sequences based on the phylogeny of the 11 primates. In this analysis, we excluded the signal peptide region, because it does not affect the function of the mature protein, as well as the six structural cysteines, one catalytic lysine, and two catalytic histidines (see fig. 3) that are absolutely necessary for the RNase activity of ANG and are not subject to positive selection. The potential numbers of synonymous \( (S) \), nonsynonymous \( (N) \), conservative nonsynonymous \( (C) \), and radical nonsynonymous \( (R) \) sites of ANG are estimated to be 91.7, 247.3, 141.3, and 106.1, respectively, whereas the total numbers of substitutions at these sites over the whole tree are \( \sigma_s = 31.7 \), \( \sigma_n = 129.3 \), \( \sigma_C = 64.0 \), and \( \sigma_R = 65.3 \), respectively. The signal peptide, six structural cysteines, and three catalytic residues (fig. 3) are excluded from the analysis. The phylogenetic position of the talapoin monkey has been controversial (Van der Kuyl et al. 1995; Van der Kuyl, Dekker, and Goudsmit 2000), with some genes (including our own unpublished data of RNases 1 and 8) supporting its close relationship with the genus Cercocebus (e.g., green monkey), but a mitochondrial sequence data (which may potentially be discredited as nonorthologous nuclear insertions) supporting it as a basal lineage in the subfamily Cercoceboidea. Our results are virtually identical under the two trees, though the latter tree is presented here because it is strongly supported by the ANG data (99% bootstrap, fig. 1B).

**Synonymous rate \( (S/N) = 0.346 \)**

**Nonsynonymous rate \( (2N/sN) = 0.523 \)**

**Conservative rate \( (2C/N) = 0.453 \)**

**Radical rate \( (2R/N) = 0.615 \)**
the synonymous mutations in ANG are not subject to selection. Therefore, the significantly greater rate of nonsynonymous substitution than that of synonymous substitution among ANG genes can be taken as an indication of positive selection promoting nonsynonymous substitutions. Analysis of complete sequences (including signal peptides) yielded similar results (data not shown). We also inferred the ancestral gene sequences for all interior nodes of the tree and counted the numbers of synonymous and nonsynonymous changes in each tree branch (fig. 2). This analysis gave essentially the same results as those described earlier. That is, the total number of synonymous substitutions over the entire tree is 0.346 per site, whereas that of nonsynonymous substitutions is 0.523 per site, with the latter being significantly greater than the former (P = 0.02, binomial test; fig. 2). In addition, there seem to be fluctuations in the n/s ratio among tree branches. For instance, the branch leading to the three-species group of baboon, rhesus monkey, and green monkey has an n/s ratio of 13/0, significantly higher than that (14/8) for the branches linking these three species (P < 0.02, Fisher's test). But tests of this kind may result in statistical artifacts, as the null hypothesis was specified after a great difference of n/s had been observed among branches (Zhang and Nei 2000). Additional information is necessary to confirm whether there is indeed a significant variation of n/s among lineages.

Positive Selection Favors Charge Diversity

To examine the nature of amino acid substitutions favored by selection, we separated nonsynonymous substitutions into two groups: those altering the amino acid charge are called radical substitutions, whereas those that leave charge unaltered are conservative. Earlier studies showed that for most mammalian genes, the rate of radical substitution is significantly lower than that of conservative substitution because of stronger purifying selection on the former (Zhang 2000). In the case of primate ANG, however, the opposite pattern is found. The number of radical substitutions per site (0.616) is greater than that of conservative substitutions per site (0.453) over the entire tree (P = 0.02, binomial test). Furthermore, the rate of radical nonsynonymous substitution exceeds that of synonymous substitution (P < 0.01, binomial test), whereas the rate of conservative substitution is not significantly different from the synonymous rate (P > 0.10). These comparisons strongly suggest that charge-altering amino acid substitutions have been favored by positive selection. Interestingly, substitutions that add both positive and negative charges are found in many branches of the tree (fig. 2), and the net charge of ANG has remained relatively constant throughout primate evolution. With the exclusion of the deepest branch of the tree (fig. 2) in which the evolutionary direction is unclear, there were a total of 21 ami-
no acid substitutions from neutral or negative toward positive charge and another 21 leading to negative charge. Apparently, rather than creating a directional shift in the net charge of the protein, here we observe a diversifying selection serving to promote the diversity of amino acid charge at individual positions. When mapped to the three-dimensional structure of the human ANG (Leonidas et al. 1999), these charge-altering substitutions are found at the protein surface (data not shown). As such, they may influence contacts of ANG with the molecules with which it interacts, including actin, angiostatin, elastase, heparin, plasminogen, RNAse inhibitor (RI), and a 170-kDa receptor on endothelial cells (Hu et al. 1993; Hu, Riordan, and Vallee 1997; Strydom 1998). In fact, 9 of the 24 RI-binding residues in ANG (Papageorgiou, Shapiro, and Acharya 1997) as well as 6 of the 21 residues of the actin-binding region (motif III) (Strydom 1998) show charge-altering substitutions (fig. 3). We also examined polarity and size profiles (see Zhang 2000) of the amino acid substitutions, but neither analysis yielded significant patterns.

Angiogenic Activity of Recombinant ANGs

As mentioned earlier, diversifying selection often widens the spectrum of the recognizable ligands but rarely changes the main function of the protein. To examine if this is the case for ANG, we made recombinant ANG proteins from the human and owl monkey, a New World monkey, and compared their angiogenic activities, using the chicken embryo CAM assay (Table 1). We found that both ANGs are angiogenic ($P < 0.01$, $G$-test), with no statistically significant difference in activity ($P > 0.4$, $G$-test). This result suggests that, despite rapid evolution, primate ANG has maintained its ability to stimulate blood vessel formation. This finding provides further evidence suggesting that the positive selection acting on ANG is unlikely to be directional selection.

Intraspecific Polymorphisms at ANG

The rapid evolution of ANG between species prompted us to examine its intraspecific variations. We sequenced an 829-bp region that includes the entire coding region (441 bp) and an upstream noncoding region (388 bp) in humans and chimpanzees (fig. 5). Four single-nucleotide polymorphisms (one synonymous, one nonsynonymous, and two noncoding) are found among 35 humans of different ethnic backgrounds (fig. 6). The nucleotide diversity estimated ($0.00063 \pm 0.00012$) is within the normal range for human nuclear genes (Li and Sadler 1991; Cargill et al. 1999; Halushka et al. 1999). No departures from the Hardy-Weinberg equilibrium or neutrality are detected. In chimpanzees, however, an examination of the same region in 13 unrelated animals reveals no polymorphism. A randomization test indicates that the polymorphic level of ANG is significantly lower in chimpanzees than in humans ($P < 0.001$). As a comparison, polymorphisms were detected in a subsample of six chimpanzees (five from San Diego...
Zoo and one from Coriell) in the RNase 2 and RNase 3 loci (Zhang and Rosenberg 2000), which are loosely linked with the ANG gene on human chromosome 14. Because chimpanzees have substantially higher variation than humans for most nuclear genes (Kaessmann, Wiebe, and Paabo 1999; Kaessmann et al. 2001), ANG seems abnormal in this regard. Interestingly, whereas there was only one fixed synonymous substitution during human evolution since the human-chimpanzee separation, there was one fixed nonsynonymous substitution in chimpanzee evolution, which resulted in a charge-altering amino acid substitution from Asp to His at position 23 of the mature peptide of ANG (figs. 3, 5, and 6). It is possible that the low level of polymorphism at ANG of chimpanzees was because of a recent selective sweep of the His23 allele through the population that replaced the Asp23 allele and that insufficient time has elapsed to restore polymorphism.

Discussion

In this work we provide evidence demonstrating that ANG of higher primates evolved rapidly under positive selection. We have identified the selective pressure as promoting charge diversity at residues on the protein surface, including those sites involved in ANG-ANG binding–protein interactions, whereas our functional angiogenesis assay indicated that the selection did not change the basic function of ANG as an angiogenic factor. Taken together, these results lead us to conclude that the ANG genes of higher primates have been evolving under the pressures of diversifying positive selection.

There are a number of intriguing similarities in the evolutionary histories of ANG and those of the breast cancer suppressor gene, BRCA1 (Huttley et al. 2000). Similar to our results with ANG, comparisons of BRCA1 among humans, chimpanzees, and several other primate lineages yield  \( d_{SG}/d_S \) ratios either higher than or close to 1. This finding of positive selection operating in many lineages is suggestive of diversifying selection and is similar to that reported here for primate ANG genes. Similarly, the most variable sites in BRCA1 reside in a specific protein-protein interaction domain. These common features characterizing the positive selection of both BRCA1 and ANG raise the intriguing possibility of unusual selective pressures on specific cancer-related genes in primates. It is noteworthy that all these genes have normal physiological functions apart from their roles in cancer promotion or suppression, and their adverse effects may be related to a side effect of physiological function, random mutations, or alterations in expression. Because most tumors form after reproductive age, tumor promotion or suppression itself is not likely to be subject to natural selection, and the selective pressures acting on ANG are most likely related to another, more physiological role for this protein. Several authors have discussed a role for ANG during pregnancy in tissue vascularization of the developing embryo (Hayashi, Yanagihara, and Hata 2000; Koga et al. 2000; Malamitsi-Puchner et al. 2000). The diversifying selection on ANG may be related to physiologic issues of conflict of interests between mothers and fetuses, as suggested by Haig (1993) in explaining the origin of genetic imprinting. It is intriguing to consider the possibility that the natural selection that has improved ANG’s fitness for prereproductive, physiologic events yielded tumor promotion and cancer progression as undesirable, albeit postreproductive results.

Future studies on the evolution of ANG-binding molecules are likely to determine whether these molecules are also under diversifying selection and how they cope with the rapid evolution of ANG. RI is of particular interest in this regard because it is an antagonist of ANG and thus may play a regulatory role in one or more of the compatibility issues noted earlier. Consistent with this idea, RI is expressed in the placenta and other tissues, and our present work reveals charge-altering substitutions in nearly 40% of the RI-binding residues in primate ANGs.

Intraspecific polymorphisms of genes under diversifying selection vary greatly. For example, the accessory gland protein Acp26Aa of Drosophila mauritiana (Tsaur, Ting, and Wu 2001) and human MHC loci (Nei and Hughes 1991) have very high degrees of polymorphism, but human immunoglobulin (Li and Hood 1995; Sasso, Buckner, and Suzuki 1995), human protamine (Wyckoff, Wang, and Wu 2000), and abalone sperm lysin loci (Metz, Robles-Sikisaka, and Vacquier 1998) have quite low levels of variation. This dichotomy may reflect two different modes of diversifying selection, with one being an overdominant or frequency-dependent selection process that maintains polymorphism, and the other, a selection process that promotes interspecific changes by purging the intraspecific variations through selective sweeps. In the case of ANG, we detected a normal level of polymorphism in humans but no polymorphism in chimpanzees. Further analysis suggested the possibility of a selective sweep of a His23 allele replacing the Asp23 allele in chimpanzee evolution. Using the nucleotide differences between the human and chimpanzee ANG sequences (fig. 5) and assumptions of a generation time of 20 years, an effective population size of 30,000, and a divergence time of 6 Myr between humans and chimpanzees (Kaessmann et al. 2001), we estimated that the age of the most recent common ancestor of the 26 chimpanzee alleles examined here is \( T_{node} = 915,200 \) years, which is about half of the corresponding age (1.9 Myr) estimated from noncoding regions (Kaessmann et al. 2001). Although this comparison is consistent with the selective sweep hypothesis for the chimpanzee ANG gene, it has to be pointed out that \( T \) has a substantial stochastic variance. A larger sample of chimpanzees and inclusion of sequence from closely related bonobo species, together with an analysis of a larger region of gene sequence surrounding the ANG coding sequence, needs to be surveyed in order to time this potential selective sweep more precisely. Examination of pygmy chimpanzees may also help time the event. It is also possible that the selective sweep does not result from selection at ANG per se, but rather from a closely linked locus.
Although diversifying selection of ANG is detected in higher primates, it is unclear whether the same selection operates in other mammals. The evolutionary pattern of ANG may differ when multiple copies of ANG genes are present, as in the mouse and cattle. It is known that cattle ANG2 and mouse ANG3 are angiogenic (Strydom, Bond, and Vallee 1997; Fu et al. 1999), whereas mouse ANG2 is not (Nobile, Vallee, and Shapiro 1996). It will be interesting to compare the selective pressures on and evolutionary rates of the angiogenic and nonangiogenic ANG genes because such information will be useful toward improving our understanding of the physiology of these proteins and may have an impact on future studies relating to ANG and cancer research.

Acknowledgments

We thank David McDermott, Oliver Ryder, and Harold McClure for the human and chimpanzee DNAs, Hynda Kleinman for instructions on the chicken embryo CAM assay, Kimberly Dyer for assistance, and two anonymous referees for their helpful comments.

LITERATURE CITED


STEPHEN PALUMBI, reviewing editor

Accepted November 12, 2001