Volume 287, number 1,2, 146–148 FEBS 10006 August 1991

ADONIS 0014579391007160

Evolutionary origin of numerous kringles in human and simian apolipoprotein(a)

Kazuho Ikeo[1,2], Kei Takahashi[3] and Takashi Gojobori[1]

[1]*DNA Research Center, National Institute of Genetics, Mishima 411, Japan,* [2]*The Graduate University for Advanced Studies, Mishima 411, Japan* and [3]*Department of Physiology, Shimane Medical University, Izumo 693, Japan*

Received 15 May 1991

Human apolipoprotein(a) has a great size heterogeneity and consists of 38 kringle domains in the amino terminal and a serine protease domain in the carboxyl terminal. All but one kringle of apolipoprotein(a) are homologous to the fourth kringle of plasminogen. However, the 38th kringle resembles the fifth kringle of plasminogen and it seems to have been deleted in simian species. The phylogenetic trees suggest that an ancestral apolipoprotein(a) may have started with a duplicate of a plasminogen type protein. It also implies that deletion of the three kringles in the amino terminus followed, and that one of the remaining two kringles was duplicated in both human and simian species and the other was processed by a deletion in simian species after species separation. Thus, the number of kringles in other mammals not yet studied may vary considerably from species to species.

Molecular evolution; Kringle; Blood coagulation; Serine protease

1. INTRODUCTION

A group of serine proteases is involved in the cascade of blood coagulation and fibrinolysis [1]. Each of these active serine proteases consists of two polypeptide chains, A and B. The number of kringles varies with the type of protease, but the A-chain has 'kringle' domains each containing approximately 80 amino acids. Kringles are a characteristic secondary structure formed by three pairs of intrachain disulfide bonds. Such a structure was first found in human prothrombin [2]. The B-chain contains a protease domain that resembles trypsin.

Urinary urokinase (plasminogen activator) and coagulation factor XII contain one kringle each [3]. Tissue-type plasminogen activator and prothrombin have two each [4,5]. The growth factor found in hepatocytes has four kringles [6], and plasminogen has five [7]. It is of particular interest to know how such a mosaic protein emerged from its ancestor during evolution [8].

Recently, 38 kringles have been found in human apolipoprotein(a) (apo(a)) [9]. Ten kringles, at least, are in the sequence of rhesus macaque [10]; each of the kringles is very similar to its human counterpart. More kringles in rhesus macaques may be found by further analysis of DNA. A large number of kringles in apo(a) seem to be derived from the kringles of a plasminogen-type protein, since they are homologous to the fourth kringle of plasminogen (plgen 4).

Correspondence address: T. Gojobori, DNA Research Center, National Institute of Genetics, Mishima 411, Japan.

To elucidate the evolutionary origin of these homologous domains, we constructed a phylogenetic tree for the kringles of serine protease and apo(a) in various organisms.

2. MATERIALS AND METHODS

We constructed phylogenetic trees by the use of all available sequences of amino acids and nucleotides. The 50 amino acid sequences from PIR (version 25) and the 45 nucleotide sequences from GenBank (version 64) were compiled from various serine proteases. Homology alignment with maximum match was first performed. From comparisons of each pair of sequences, substitution numbers were calculated. For the correction of multiple substitutions at a site, Kimura's equation was used; $d_a = -\log_e(1 - p - 0.5p^2)$ [11], where d_a and p are the number of amino acid substitutions per site and the proportion of different amino acids, respectively. A phylogenetic tree was constructed from amino acid sequences by use of d_a values by unweighted pair-grouping (UPG) method, assuming a constant substitution rate over time [12]. In the case of a tree constructed from nucleotide sequences, the total number of nucleotide substitutions for each pair of kringles was calculated by the six-parameter method [11,13].

3. RESULTS AND DISCUSSION

Fig. 1 shows a phylogenetic tree of all kringles that was constructed by the UPG method, using amino acid sequences. The majority of kringles could be separated into the three major groups indicated by I, II and III. By this classification, all kringles of apo(a) except the 38th (apo 38) belong to group I. The second and third kringles (plgen 2 and plgen 3) of plasminogen and the second kringle (HGF 2) of hepatocyte growth factor are in group II. Apo 38, plgen 1, and plgen 5 in group III

Published by Elsevier Science Publishers B.V.

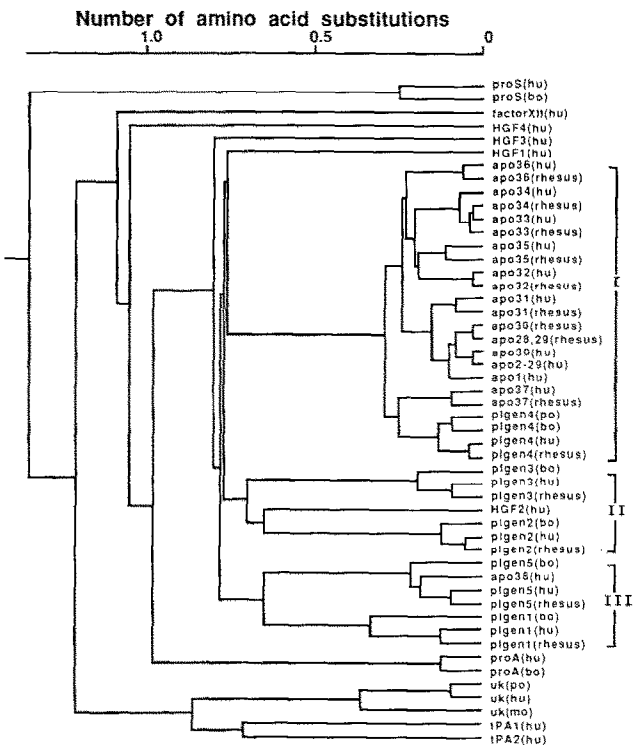

Fig. 1. Phylogenetic trees of kringle domains. The tree was constructed by analysis of amino acid sequences. Three major groups are indicated by I, II and III. Abbreviations: tPA, tissue-type plasminogen activator; uk, urokinase; plgen, plasminogen; apo, apolipoprotein(a); HGF, human hepatocyte growth factor; hu, human; bo, bovine; po, porcine; mo, mouse. The numbers represent the kringle numbers from the N-terminus of each protein. Sources: pro(hu) [2,5], factorXII(hu) [3], tPA(hu) [4], HGF(hu) [6], plgen(bo) [7], apo(hu) [9], apo(rhesus) [10], plgen(rhesus) [10], pro(bo) [17], uk(mo) [18], uk(hu) [19], uk(po) [20], plgen(hu) [21].

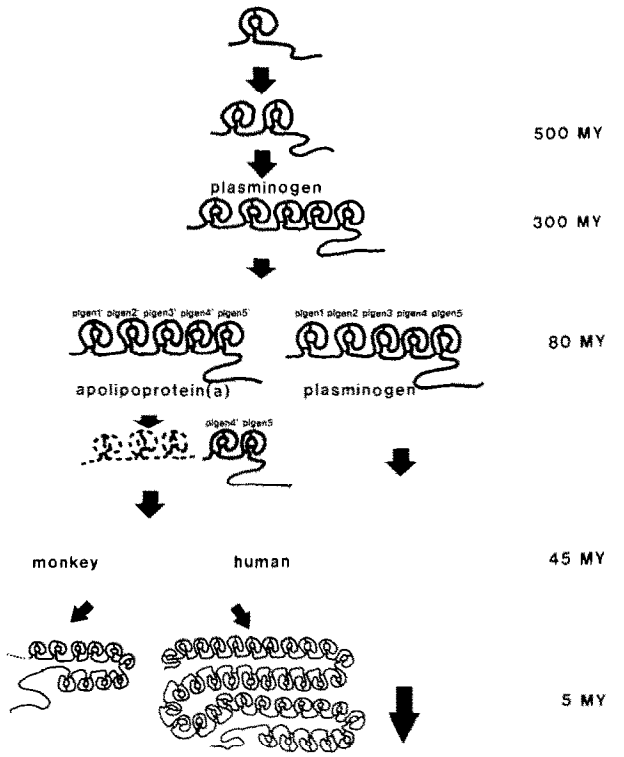

Fig. 2. Evolutionary change in apolipoprotein(a). The ancestral molecule of apo(a) is considered to have been a protein with one kringle and one serine protease domain. The kringle was continuously duplicated to form two types of kringles, plgen 4′ and plgen 5′, about 500 million years ago. An ancestral plasminogen-like protein with five kringles emerged by duplication of kringles about 300 million years ago. The entire sequence of the protein was duplicated; one formed apo(a) and the other formed plasminogen about 80 million years ago when mammals diverged from their ancestor. Then, plgen 1′–3′ were deleted in the human sequence. In simian species, plgen 5′ was also deleted. Both in human and simian species, plgen 4′ followed duplications about 2 to 3 million years ago after the two species were separated.

are differentially separated from the other 37 kringles of apo(a).

The rate of amino acid substitution was calculated to be 0.956×10^{-9} per amino acid site per year, assuming that the divergence of mammals occurred 80 million years ago. Using this rate, we estimated the time of the evolutionary events that must have taken place in the past.

The evolutionary development of kringle and apo(a) in human and simian species is summarized in Fig. 2. The following is based on an estimation of the divergence times for various branching points in the phylogenetic tree. An ancestral gene must have been of the plasminogen type with one kringle and one serine protease domain. About 500 million years ago, the kringle was duplicated into two domains, one similar to an ancestral domain in groups I and II, and the other similar to an ancestor in group III. The former could have been similar to plgen 4 (and plgen 4′) and the latter similar to plgen 5′; corresponding to apo 38). Some 300 million years ago, plgen 4′ was triplicated into three domains; plgen 1′, plgen 2′, and plgen 3′. Thus, the tree shows that all the five kringles have existed in the plasminogen since about 300 million years ago. Therefore, it is considered that the ancestral molecule of apo(a) was a plasminogen-type protein having five kringles.

About 80 million years ago, this plasminogen-like gene was duplicated into two genes; one formed apo(a) and the other became the present plasminogen. It is supported by a cytogenetical study that genes for human apo(a) and plasminogen are located very close to each other on chromosome 6 with at q2.6–2.7 and q2.7, respectively [14]. Accordingly, apo 38 was probably derived from plgen 5 about 80 million years ago. Separation of the ancestral gene of the 37 other kringles from plgen 4′ was at almost the same time as when apo 38 diverged from its ancestor. Then, a portion containing plgen 1′ to plgen 3′ was deleted in apo(a) of human and simian species. Thereafter, plgen 4′ underwent multiple duplications about 2 to 3 million years ago, giving rise to the 37 kringles in humans. In particular,

the 36th kringle acquired the ability to bind to apolipoprotein B-100 at an early stage of these duplications. In simian species, the process was similar but independent of the human homologue because the divergence between humans and rhesus macaque had occurred (about 45 million years ago) much earlier than the multiple duplication of plgen 4′ in humans. In particular, plgen 5′ was deleted while plgen 4′ (apo28 - apo37) was multiplied in rhesus macaque. A similar result was obtained by another tree that was constructed by the neighbour-joining method [15] (data not shown). Note that this tree-making method does not require assumption of a constant rate of amino acid substitution.

To clarify the evolutionary processes of kringles in human and simian species further, we constructed a tree by the use of nucleotide sequences. The tree is much more informative at the nucleotide sequence level than at the amino acid sequence level, because the identical sequences of amino acids such as (apo2-29) were different sequences as the corresponding sequences of nucleotides. However, this tree topology was essentially the same as the one in Fig. 1. The rate of nucleotide substitution was calculated to be 1.48×10^{-9} per site per year. Using this rate, the multiple duplications of plgen 4′ in human apo(a) were estimated to have occurred 2 to 3 million years ago. This is consistent with the results that were obtained from the tree of amino acid sequences.

Tomlinson et al. [10] speculated that apo(a) emerged from its ancestor about 40 million years ago. However, our analysis shows that the ancestral molecule of apo(a) evolved from the plasminogen-like gene about 80 million years ago. It implies that apo(a) appeared at almost the same time that the mammalian divergence took place. If this is the case, apo(a) may be found in a certain variety of mammalian species. It is also possible that the number of kringles in the proteins homologous to apo(a) would be found to vary largely with the species if the DNA of apo(a) in other mammals was sequenced. The regions that were multiply-duplicated in apo(a) may be unstable because further duplications are likely to occur in this region. These duplications of plgen 4′ can be the cause of the size heterogeneity of apo(a). In fact, it is known that apo(a) has several isoforms in human and simian species [15].

Kringles are structurally autonomous units [16]. However, their consensus function is not well understood. Accumulation of sequence data from various proteins would provide us with a better understanding of kringles and an answer to the question of where they came from [15].

REFERENCES

[1] Neurath, H. (1984) Science 224, 350–357
[2] Walz, D.A., Hewett-Emmett, D. and Seegers, W.H. (1977) Proc. Natl. Acad. Sci. USA 74, 1969–1972.
[3] McMullen, B.A. and Fujikawa, K. (1985) J. Biol. Chem. 260, 5328–5341.
[4] Penica, D., Holmes, W.E., Kohr, W.J., Harkins, R.N., Vehar, G.A., Bennet, W.M., Yelventon, E., Seeburg, P.H. Heyneker, L. and Goeddel D.V. (1983) Nature 301, 214–221.
[5] Sandra, J.F.D., Ross, T., MacGillivary, A. and Davie, E.W. (1983) Biochemistry 22, 2087–2097.
[6] Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K. and Shimizu, S. (1989) Nature 342, 440–443.
[7] Schaller, J., Moser, P.W., Dannegger-Muller, G.A.K., Roselt, S.J., Kampfer, V. and Rikli, E.E. et al. (1985) Eur. J. Biochem 149, 267–278.
[8] Patthy, L. (1985) Cell 41, 657–663.
[9] McLean, J.W., Tomlinson, J.E., Kuang, W-J., Eaton, D.L., Ellson, Y.C., Fless, G.M., Scanu, A.M. and Lawn, R.M. (1987) Nature 330, 132–137.
[10] Tomlinson, J.E., McLean, J.W. and Lawn, R.M. (1989) J. Biol. Chem. 264, 5957–5965.
[11] Kimura, M. (1983) The Neutral Theory of Molecular Evolution (Cambridge University Press, Cambridge, U.K.
[12] Nei, M. (1975) Molecular Population Genetics and Evolution, Elsevier, Amsterdam.
[13] Gojobori, T., Ishii, K. and Nei, M. (1982) J. Molec. Evol. 18, 414–423.
[14] Frank, S.L., Klisak, I., Sparkes, R.S., Mohandas, T., Tomlinson, J.E., McLean, J.W., Lawn, R.M. and Lusis, A.J. (1988) Hum. Genet. 79, 352–356.
[15] Saitou, N. and Nei, M. (1987) Mol. Biol. Evol. 4, 406–425.
[16] Utermann, G. (1989) Science 246, 904–910.
[17] Trexler, M. and Patthy, L. (1983) Proc. Natl. Acad. Sci. USA 80, 2457–2461.
[18] MacGillivary, R.T.A. and Davie, E.W. (1984) Biochemistry 23, 1626–1634.
[19] Belin, D., Vassalli, J-D., Combepine, C., Godeau, F., Nagamine, Y., Reich, E., Kocher, H.P. and Duvoisin, R.M. (1985) Eur. J. Biochem. 148, 225–232.
[20] Verde, P., Stoppelli, M.P., Galeffi, P., Nocera, P.D. and Blasi, F. (1984) Proc. Natl. Acad. Sci. USA 81, 4727–4731.
[21] Nagamine, Y., Pearson, D., Altus, M.S. and Reich, E. (1984) Nucleic Acids Res. 12, 9525–9541.
[22] Forsgren, M., Raden, B., Israelsson, M., Larsson, K. and Haden, L.-O. (1987) FEBS Lett. 213, 254–260.