The EMBO Journal vol.7 no.13 pp.4151-4156, 1988

The primary structure of human ribonuclease/angiogenin inhibitor (RAI) discloses a novel highly diversified protein superfamily with a common repetitive module

Rainer Schneider, Elisabeth Schneider-Scherzer, Martin Thurnher, Bernard Auer and Manfred Schweiger

Institute of Biochemistry, University of Innsbruck, Innrain 52a, 6020 Innsbruck, Austria

Communicated by M.Schweiger

Immunological screening of a λgt11 library, constructed from HeLa mRNA, yielded several ribonuclease/angiogenin inhibitor (RAI) cDNA clones containing 900-bp inserts. Northern blot analysis revealed that the length of the RAI mRNA is ~1.9 kb. Construction and screening of a eukaryotic cDNA expression library (HeLa) containing preferentially complete cDNA inserts led to the isolation of a full length clone. The complete nucleotide sequence was determined. The C-terminal amino acid sequence deduced from the cDNA is identical to the peptide sequence obtained from a CNBr fragment of RAI, confirming the identity of the clone. The deduced primary structure of RAI consists of eight homologous tandem repeats with remarkable periodicity of leucine and cysteine residues. Each repeat is derived from the duplication of a leucine-rich 28-amino-acid module. This prototype module is closely related to a repetitive 24-amino-acid motif of unclear function, previously found in proteins involved in important biological processes such as blood coagulation, embryonic development, cell morphogenesis and signal transduction. Although homologous, the RAI modules show distinct differences in length and amino acid composition to the modules of this group of proteins, demonstrating their high potential of variability, necessary for adaptation to very diverse roles. Based on our results we propose that these repetitive modules are a common structural feature of a novel protein superfamily whose members exert their function by highly specific protein-protein interactions.

Key words: α_2-glycoprotein/protein-binding modules/proteoglycan/ribonuclease inhibitor/von Willebrand factor-binding

Introduction

Inhibitors of neutral ribonuclease have long been known to be present in mammalian tissues (Roth, 1956, 1967; Traub *et al.*, 1966). The human placental ribonuclease inhibitor, an acidic protein with M_r 51 000 (Blackburn *et al.*, 1977), is presently used in molecular biology to stabilize preparations of polyribosomes and mRNA.

RNase activity is inhibited by the formation of a tight oxidation sensitive 1:1 complex with the inhibitor. *In vivo* >95% of the ribonuclease is complexed with inhibitor. The inhibitor/RNase ratio is elevated in proliferating tissues (Shortman, 1962; Kraft and Shortman, 1970) and diminished during catabolic metabolism (Quirin-Stricker *et al.*, 1968). Hence, this equilibrium seems to play a central role in the regulation of mRNA turnover and protein biosynthesis. Interestingly, the amounts of inhibitor and neutral RNase are affected by hormonal stimuli, e.g. by steroid hormones (Kershner and Meyer, 1976) and growth hormone (Brewer *et al.*, 1969). Changes in the inhibitor/RNase ratio are supposed to play a role in Alzheimer's disease (Sajdel-Sulkowska and Marotta, 1984; Majocha *et al.*, 1987), muscular dystrophy (Little and Meyer, 1970), hyperthyreosis (Greif and Eich, 1977), during cellular repair (Ferencz *et al.*, 1973; Schweiger *et al.*, 1988) and aging (Kraft and Shortman, 1970; Sarkar, 1970).

It was recently reported that human placental ribonuclease inhibitor abolishes both angiogenic and ribonucleolytic activities of angiogenin (Shapiro and Vallee, 1987), a blood-vessel-inducing protein from HT-29 human colon adenocarcinoma cells (Fett *et al.*, 1985). Since the inhibitor binds more tightly to angiogenin than to pancreatic RNase (Shapiro and Vallee, 1987), ribonuclease/angiogenin inhibitor (RAI) seems to be a more appropriate name.

The ribonucleolytic and the angiogenic activities of angiogenin are inter-related (Shapiro and Vallee, 1987). However, the distribution of angiogenin mRNA among various tissues is not consistent with their angiogenic status (Weiner *et al.*, 1987). These findings imply that RAI is a regulatory factor in angiogenesis. Since angiogenesis is a crucial step not only in the vascular development of the fetus and in the neovascularization that accompanies wound healing, but also in tumour development, *in vivo* regulation of angiogenin by RAI might be of pharmacologic and therapeutic interest.

In view of the aspects discussed above the protein-protein interaction between RAI and its substrates and hence the primary structure of RAI are of general interest. This prompted us to isolate a cDNA clone for RAI to elucidate its primary structure and thereby to provide means for studying the *in vivo* regulation of mRNA turnover and angiogenesis. Here we report the isolation and characterization of the RAI cDNA. The analysis of the deduced primary structure reveals that RAI consists entirely of short, tandemly repeated modules. Homologous modules of unclear function have also been found in other proteins (Takahashi *et al.*, 1985; Hashimoto *et al.*, 1988). Our results help to understand the common function of the repetitive modules of this novel protein superfamily.

Results

Antibodies, peptide sequencing and oligonucleotides

Antibodies were raised in rabbits using highly purified RAI from human placenta. In Western blot analysis of various human tissue homogenates these antibodies detected only a single polypeptide chain of M_r 51 000 which corresponds to the M_r of human RAI (Blackburn *et al.*, 1977).

CNBr-peptide: M E D R L Q A L E K D K P S L R V I S

OLR1: 5'- G G $\frac{T}{C}$ T T $\frac{A}{G}$ T C $\frac{T}{C}$ T T $\frac{T}{C}$ T C - 3'

Fig. 1. Amino acid sequence of a CNBr peptide from RAI. OLR1 is an oligonucleotide constructed according to the pentapeptide underlined in the CNBr peptide.

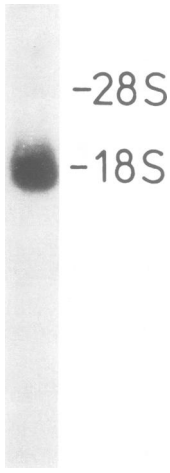

Fig. 2. Northern blotting: 20 μg of total cytoplasmic RNA from human monocytes were subjected to denaturing agarose gel electrophoresis and blotted to a nylon membrane. A human RAI-specific cDNA probe was used for hybridization.

The N terminus of RAI is blocked, therefore, RAI was CNBr digested and the resulting fragments were separated by HPLC. A prominent fragment was sequenced by automated gas-phase sequencing; the amino acid sequence is shown in Figure 1. The underlined pentapeptide provides minimum codon degeneracy and was therefore chosen for the design of an oligonucleotide (OLR1). OLR1 (Figure 1) had to be prepared as an equal mixture of 16 related sequences to cover the remaining degeneracies.

Isolation of cDNA clones for human RAI

RAI is of low abundancy in human cells; in placenta RAI represents 0.06% of the soluble protein (Blackburn, 1979). Fast-proliferating cells are supposed to contain a higher concentration of RAI to guarantee an elevated level of protein synthesis. Therefore we chose HeLa cells as a source of mRNA for the cDNA cloning procedure. Construction of a λgt11 library (Huynh *et al.*, 1985) from 10 μg of HeLa mRNA yielded 1.2×10^6 independent recombinants.

Screening of this library with antibodies against RAI yielded several identical cDNA clones containing a 900-bp insert. Western blot analysis of induced lysogens of these clones showed that the antibodies reacted specifically with a fusion protein of M_r 140 000. Additionally, OLR1 hybridized specifically with the cDNA inserts (data not shown). These independent lines of evidence confirmed the identity of the cDNA clones.

To obtain a full length cDNA clone for RAI we constructed a eukaryotic expression cDNA library containing

```
   1  AGTGCCTTAGATTCCAGCGAGCTACGAAGCAATCCTGGCCCAGCCGAGCTTGCTTCCCCAAATCCCGTAATCCTTGACCTTATTCCCCCAAAGAAGCGGCCTCCCGGGAAGGAGCGCCCT

 121  GGCGGAGAAGACTCGAACGGCTCCCACAGCCGGGCGTTGGGGGTAAAGGCATGAAGAACTCTTGACTGACAGAAACGGAGGGTGTGTCCAAAGTTTTGAGGACGGCCGAGCGGCGCTCCA

 241  AAACCCGTCCTCACAGCCTCGCCCCGTTCGCCTCAGCTACAACAAATCATCGTCAACCTGTTCCACCTTCTCCAGTCTGGTAGCAAAAAGGGGTGTCTCAGGCCACTCTTCACCTCCACC

 361  ATGAGCCTGGACATCCAGAGCCTGGACATCCAGTGTGAGGAGCTGAGCGACGCTAGATGGGCCGAGCTCCTCCCTCTGCTCCAGCAGTGCCAAGTGGTCAGGCTGGACGACTGTGGCCTC
          M  S  L  D  I  Q  S  L  D  I  Q  C  E  E  L  S  D  A  R  W  A  E  L  L  P  L  L  Q  Q  C  Q  V  V  R  L  D  D  C  G  L
                                                                                                                              40

 481  ACGGAAGCACGGTGCAAGGACATCAGCTCTGCACTTCGAGTCAACCCTGCACTGGCAGAGCTCAACCTGCGCAGCAACGAGCTGGGCGATGTCGGCGTGCATTGCGTGCTCCAGGGCCTG
          T  E  A  R  C  K  D  I  S  S  A  L  R  V  N  P  A  L  A  E  L  N  L  R  S  N  E  L  G  D  V  G  V  H  C  V  L  Q  G  L
                                                                                                                              80

 601  CAGACCCCCTCCTGCAAGATCCAGAAGCTGAGCCTCCAGAACTGCTGCCTGACGGGGGCCGGCTGCGGGGTCCTGTCCAGCACACTACGCACCCTGCCCACCCTGCAGGAGCTGCACCTC
          Q  T  P  S  C  K  I  Q  K  L  S  L  Q  N  C  C  L  T  G  A  G  C  G  V  L  S  S  T  L  R  T  L  P  T  L  Q  E  L  H  L
                                                                                                                             120

 721  AGCGACAACCTCTTGGGGGATGCGGGCCTGCAGCTGCTCTGCGAAGGACTCCTGGACCCCCAGTGCCGCCTGGAAAAGCTGCAGCTGGAGTATTGCAGCCTCTCGGCTGCCAGCTGCGAG
          S  D  N  L  L  G  D  A  G  L  Q  L  L  C  E  G  L  L  D  P  Q  C  R  L  E  K  L  Q  L  E  Y  C  S  L  S  A  A  S  C  E
                                                                                                                             160

 841  CCCCTGGCCTCCGTGCTCAGGGCCAAGCCGGACTTCAAGGAGCTCACGGTTAGCAACAACGACATCAATGAGGCTGGCGTCCGTGTGCTGTGCCAGGGCCTGAAGGACTCCCCCTGCCAG
          P  L  A  S  V  L  R  A  K  P  D  F  K  E  L  T  V  S  N  N  D  I  N  E  A  G  V  R  V  L  C  Q  G  L  K  D  S  P  C  Q
                                                                                                                             200

 961  CTGGAGGCGCTCAAGCTGGAGAGCTGCGGTGTGACATCAGACAACTGCCGGGACCTGTGCGGCATTGTGGCCTCCAAGGCCTCGCTGCGGGAGCTGGCCCTGGGCAGCAACAAGCTGGGT
          L  E  A  L  K  L  E  S  C  G  V  T  S  D  N  C  R  D  L  C  G  I  V  A  S  K  A  S  L  R  E  L  A  L  G  S  N  K  L  G
                                                                                                                             240

1081  GATGTGGGCATGGCGGAGCTGTGCCCAGGGCTGCTCCACCCCAGCTCCAGGCTCAGGACCCTGTGGATCTGGGAGTGTGGCATCACTGCCAAGGGCTGCGGGGATCTGTGCCGTGTCCTC
          D  V  G  M  A  E  L  C  P  G  L  L  H  P  S  S  R  L  R  T  L  W  I  W  E  C  G  I  T  A  K  G  C  G  D  L  C  R  V  L
                                                                                                                             280

1201  AGGGCCAAGGAGAGCCTGAAGGAGCTCAGCCTGGCCGGCAACGAGCTGGGGGATGAGGGTGCCCGACTGCTGTGTGAGACCCTGCTGGAACCTGGCTGCCAGCTGGAGTCGCTGTGGGTG
          R  A  K  E  S  L  K  E  L  S  L  A  G  N  E  L  G  D  E  G  A  R  L  L  C  E  T  L  L  E  P  G  C  Q  L  E  S  L  W  V
                                                                                                                             320

1321  AAGTCCTGCAGCTTCACAGCCGCCTGCTGCTCCCACTTCAGCTCAGTGCTGGCCCAGAACAGGTTTCTCCTGGAGCTACAGATAAGCAACAACAGGCTGGAGGATGCGGGCGTGCGGGAG
          K  S  C  S  F  T  A  A  C  C  S  H  F  S  S  V  L  A  Q  N  R  F  L  L  E  L  Q  I  S  N  N  R  L  E  D  A  G  V  R  E
                                                                                                                             360

1441  CTGTGCCAGGGCCTGGGCCAGCCTGGCTCTGTGCTGCGGGTGCTCTGGTTGGCCGACTGCGATGTGAGTGACAGCAGCTGCAGCAGCCTCGCCGCAACCCTGTTGGCCAACCACAGCCTG
          L  C  Q  G  L  G  Q  P  G  S  V  L  R  V  L  W  L  A  D  C  D  V  S  D  S  S  C  S  S  L  A  A  T  L  L  A  N  H  S  L
                                                                                                                             400

1561  CGTGAGCTGGACCTCAGCAACAACTGCCTGGGGGACGCGGGCATCCTGCAGCTGGTGGAGAGCGTATCCGAGCCGGGCTGCCTCCTGGAGCAGCTGGTCCTGTACGACATTTACTGGTCT
          R  E  L  D  L  S  N  N  C  L  G  D  A  G  I  L  Q  L  V  E  S  V  S  E  P  G  C  L  L  E  Q  L  V  L  Y  D  I  Y  W  S
                                                                                                                             440

1681  GAGGAGATGGAGGACCGGCTGCAGGCCCTGGAGAAGGACAAGCCATCCCTGAGGGTCATCTCCTGAAGCTCTTCCTGCTGCTGCTCTCCCTGGACGACCGGCCTCGAGGCAACCCTGGGG
          E  E  M  E  D  R  L  Q  A  L  E  K  D  K  P  S  L  R  V  I  S  *
                                                                          461

1801  CCCACCAGCCCCTGCCATGCTCTCACCCTGCATATCCTAGGTTTGAAGAGAAACGCTCAGATCCGCTTATTTCTGCCAGTATATTTTGGACACTTTATAATCATTAAAGCACTTTCTTGG

1921  CAAAAAAAAAAAAAAAAAAAAAAAAAAAA
```

Fig. 3. Nucleotide sequence of RAI cDNA and the deduced sequence of RAI protein. The presumable open reading frame spans nucleotides 361–1743, the amino acid sequence (in one letter code) is shown and numbered on the right side, the stop codon is marked with a star and the sequence identical to that from the CNBr fragment is underlined. The ATG triplet, the stop codon and the polyadenylation site are all underlined.

preferentially complete cDNA inserts (Okayama and Berg, 1983). Starting from 5 μg of HeLa mRNA the cloning procedure resulted in 800 000 independent recombinants. Screening of this library by hybridization with the 900-bp insert yielded several clones with inserts varying from 0.8 to 2.1 kbp. The clone with the longest insert (OBR 2.1) was chosen for further characterization and sequence analysis.

Northern blot analysis

Northern blot analysis of total RNA extracted from human monocytes showed a single hybridizing species of ~1.9 kb (Figure 2), which is large enough to code for a protein of M_r 51 000.

cDNA sequencing and deduced amino acid sequence of RAI

Taking into account the vector sequences of the 2.1-kbp fragment, OBR 2.1 contains ~1.9 kbp of RAI-specific sequences. A remarkable feature of the cDNA is the frequent occurrence of recognition sites for restriction endonucleases *Pst*I and *Pvu*II (each occurring eight times), of which four occur in an overlapping position. The complete sequence was established both by sequencing a set of subclones and by primer extension with appropriate oligonucleotides (data not shown).

The sequence of the cDNA and the inferred amino acid sequence are shown in Figure 3. The cDNA consists of 1947 nucleotides including the complete coding region, 5′- and 3′-flanking regions and a poly(A) tail of 26 residues. Only one open reading frame adequate to code for RAI is present. The reading frame codes for 461 amino acids corresponding to a M_r of 49 911. In the same reading frame three stop codons precede the AUG codon in position 361–363, which therefore represents the only possible initiation site for translation. The 3′-flanking region comprises 200 nucleotides; 155 nucleotides downstream from the A of the UGA stop codon a consensus termination and polyadenylation sequence (ATTAAA) is present. The predicted M_r of ~50 000 is consistent with the behaviour of the protein in polyacrylamide gels under denaturing conditions.

The C-terminal amino acid sequence deduced from the cDNA is identical to the peptide sequence obtained from a CNBr fragment of RAI, confirming the identity of the cDNA clone.

The GC content of the cDNA clone, especially of the coding sequence, is remarkably high (65%), with a bias against CpG doublets, which is usual in vertebrate genomes. The high GC content results from very selective codon usage. In the wobble position nucleotides G and C are highly preferred to A or T (e.g. the codons CUC and CUG for leucine occur 84 times, whereas the codons CUU and CUA are present only three times).

The deduced amino acid composition is in close agreement with that obtained by amino acid analysis of the protein

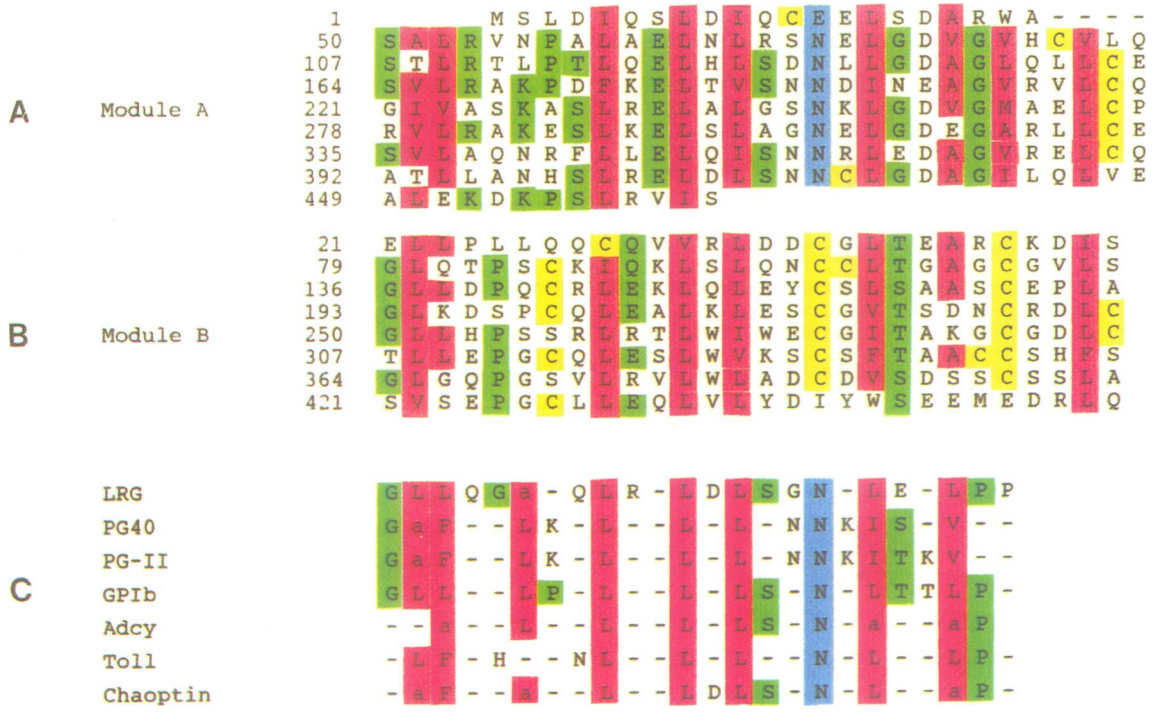

Fig. 4. Tandem repeats of RAI and consensus sequences of homologous proteins. (**A**) and (**B**) 57-residue tandem repeats of RAI; due to topological reasons the left (**A**) and right (**B**) halves of the tandem domains are each grouped and the two resulting groups are shown one beneath the other. Coloured boxing shows identities or conservative replacement within the groups of similar amino acids: IVMLFA, WYF, TS, RKH, NEQD, PG, C. Conserved hydrophobic positions are shown in red, all cysteines in yellow, asparagine in blue and other conserved positions in green. Positions are either marked, if at least four identical residues or six similar residues are found on one column. (**A**) and (**B**) The left (module A) and the right (module B) halves of the 57-residue tandem domains are correctly aligned one beneath the other in order to show that they are homologous to each other. This may clearly be seen by comparing the periodicity of the conserved leucine residues in these two module families. (**C**) Alignment of the consensus sequences of the tandem repeats of RAI-related proteins. LRG, human leucine-rich α_2-glycoprotein; PG40 and PG-II, proteoglycan core proteins PG40 of human fibroblasts and PG-II of bovine bone respectively; GPIb, human major platelet glycoprotein Ib; Adcy, yeast adenylatcyclase; Toll, *Drosophila* Toll gene product; chaoptin, *Drosophila*. The shown consensus sequences are defined by residues appearing with a frequency >40%, with 'a' indicating hydrophobic residues (IVMLFA). Coloured boxing means that an identical or similar residue is also conserved in one of the RAI module families.

(Blackburn *et al.*, 1977). RAI shows a remarkably high leucine (92 residues; 20%) and cysteine (32 residues; 7%) content. The surplus of acidic amino acids (63 residues) versus basic amino acids (39 residues) is consistent with the experimentally determined low pI value of 4.6 (Blackburn *et al.*, 1977).

Domain structure of RAI

Inspection of the deduced amino acid sequence of RAI revealed a most interesting feature. The primary structure of RAI consists of eight highly homologous tandemly repeated domains of 57 amino acids with no interspacing sequences (Figure 4A and B). Only one gap of four amino acids has to be introduced to optimize the sequence alignment. As is apparent from Figure 4A and B, there are >50% highly conserved positions preferentially occupied by leucine or hydrophobic amino acids (red) and cysteine (yellow). Almost all cysteine residues are clustered in the right half of the domains (Figure 4B) and >80% of all 92 leucine residues of RAI are located at highly conserved positions. Additionally there are 14 positions that show significant conservation concerning other amino acids or amino acid similarity groups (green and blue).

This highly regular domain structure of RAI indicates that this protein evolved by multiple gene duplications of an exact manner. Such duplications could result from unequal crossovers or transposon-like mechanisms (Patthy, 1985).

Intra-domain homologies

Further rigorous analysis of the predicted RAI domains revealed a significant intra-domain homology. The 57 amino acid domains can be divided into two clearly homologous modules of 28 and 29 amino acids (Figure 4A and B). The most conserved feature of these modules is the periodicity of leucine residues. Without introduction of any gaps or insertions eight out of nine invariant leucine positions are conserved in these modules, and the only variant position has conserved cysteine in place of hydrophobic residues. Likewise an invariant asparagine in one module coincides with an invariant cysteine in the other. As is apparent from Figure 4A and B, there are additionally numerous conserved substitutions at identical positions which further confirm this homology.

This finding implies that the first ancestral 57 amino acid domain of RAI has evolved by duplication of a leucine-rich peptide of 28 amino acids. Since the cysteine residues are predominantly in module B it seems that the stabilization of the protein structure by disulphide bonds evolved after the first duplication.

Secondary structure

A Chou and Fasman prediction of the secondary structure of RAI was performed (Chou and Fasman, 1978), and some regularity could be found; e.g. the conserved asparagine at position 17 of module A almost always occurs in a turn structure and this turn is predominantly flanked by helix or β-sheet structures. But the many disulphide bridges in RAI make it hard to predict a secondary structure according to the Chou and Fasman rules and the validity of this method is therefore low in such cases.

Homology to other proteins

By reviewing recently published sequences with internal periodicity we found a strong homology to a group of proteins containing 24-amino-acid repeats of unknown function. This group includes so far: leucine-rich α_2-glycoprotein (LRG) with unknown function (Takahashi *et al.*, 1985), the α- and β-chains of human glycoprotein Ib (GPIb), which is the platelet receptor for von Willebrand factor (Lopez *et al.*, 1987, 1988; Titani *et al.*, 1987), proteoglycan PG40 of fibroblasts (Krusius and Ruoslahti, 1986), proteoglycan II of bovine bone (PG-II) (Day *et al.*, 1987), yeast adenylatcyclase (Kataoka *et al.*, 1985), the Toll gene product of *Drosophila*, which plays a central role in the establishment of the embryonic dorsal–ventral pattern (Hashimoto *et al.*, 1988), and *Drosophila* chaoptin, which is required for photoreceptor cell morphogenesis (Reinke *et al.*, 1988).

Figure 4 shows the alignment of the RAI modules with the consensus sequences of these proteins. As is apparent, the periodicity of conserved leucine residues can be aligned with that of the RAI modules without introduction of gaps or inserts (red). Additionally the highly conserved asparagines at position 17 of the motifs in Figure 4C (blue) coincide exactly with the asparagines at position 17 of module A of RAI. On the basis of amino acid similarity six additional positions show remarkable conservation, four of which are predominantly occupied by glycine or proline (green). Although the RAI modules and the consensus sequences of the proteins shown in Figure 4 are of different size (28 and 24 amino acids respectively), their close relationship is clearly evident because all of the conserved residues in the consensus sequences also appear in the RAI repeats.

Application of the program RELATE to the pair RAI/LRG resulted in a score of 11 SD units which demonstrates that the similarities between RAI and these proteins are statistically significant.

Searching protein databases with the best conserved core consensus sequence, LXXLXLXXN, yielded ~150 proteins which contained such a sequence motif, but only the examples in Figure 4C showed both tandemly arranged repeats and further extending homologies according to the 24-amino-acid consensus sequence. Due to the shortness and variability of the modules and since no extract structure–function relationship is evident, no reliable distinction between homology and chance similarity is possible if only one module is found. Therefore proteins which contain just one similar module are not considered to be members of this protein family, even if this module is more closely related to any of the repeats in Figure 4, than some of these repeats are to each other. For example, although a single 24-amino-acid segment of *Escherichia coli* traE protein (Frost *et al.*, 1984) gave an optimal score by matching with 11 highly conserved positions of the prototype repeat, it lacks the usual periodicity of hydrophilic residues between these positions. Thus as long as no structure–function relationship of the repeats is known, it remains uncertain whether traE belongs to the module family.

All of the relationships shown in Figure 4 concern glycosylated proteins. But although RAI also contains a potential N-glycosylation site at position 397, no carbohydrate moiety could be detected (data not shown) as already described (Blackburn *et al.*, 1977).

Discussion

In this paper we describe the cloning, sequencing and characterization of RAI cDNA; additionally we present a

comparative analysis of the deduced primary structure of RAI.

The deduced primary structure of RAI revealed some most interesting features. Almost the entire primary structure consists of highly homologous tandem repeats. Eight repeats of 57 amino acids can clearly be aligned just by introducing one small gap. The repeats show a remarkable periodicity of conserved leucine residues, and the 32 cysteine residues of RAI are mainly clustered in the right half of the 57 amino acid domains. Most of these cysteines are located at four highly conserved positions with 7, 9 and 6 amino acids respectively between them.

Further analysis of the RAI domains revealed that these domains themselves are derived from a duplication of an ancestral leucine-rich prototype module. The alignment of the two halves of the domains can be achieved without introduction of gaps or insertions. The most conserved features in this alignment are the positions of the leucines which show an identical periodicity in the two modules. Thus the primary structure of RAI finally consists of 16 closely related tandemly arranged modules, with an unusually high conservation of leucine periodicity.

These two modules of 28 and 29 amino acids show significant relationships to a group of proteins including integral membrane or membrane-associated proteins, proteoglycans, and a plasma protein. A 24-amino-acid tandem repeating unit of unclear function was reported for these proteins (Takahashi *et al.*, 1985; Patthy, 1987; Hashimoto *et al.*, 1988).

The first such protein to be identified was a human serum protein of unknown function, leucine-rich α_2-glycoprotein. Its eight tandemly arranged repeats were proposed to form an amphipathic structure and therefore to associate with membranes (Takahashi *et al.*, 1985), although the repeats do not show the appropriate hydrophobic–hydrophilic structure spacing (Eisenberg, 1984; Reinke *et al.*, 1988). However, in light of our results we propose protein–protein interaction to be the function of this structural motif and, as is discussed below, this assumption is consistent with the data available of the known members of this group of proteins.

Proteoglycans PG40 of human fibroblasts (10 repeats) and PG-II of bovine bone (10 repeats) are known to bind to collagen, even after removal of their carbohydrate moiety (Vogel *et al.*, 1984); thus a protein–protein interaction seems to be associated with these structures. The major platelet glycoprotein GPIb (seven repeats) (Lopez *et al.*, 1987; Titani *et al.*, 1987) and the Toll gene product of *Drosophila* (15 repeats) (Hashimoto *et al.*, 1988) are both integral membrane proteins with leucine-rich repeats in their extracellular domains. The von Willebrand factor binding domains of GPIb, which contain these repeats, are located on a tryptic fragment, called glycocalicin, which is freely water soluble (Titani *et al.*, 1987). This is in obvious agreement with the fact that RAI, which consists entirely of such repeats, is a soluble cytoplasmic protein (Blackburn *et al.*, 1977). Therefore membrane association seems unlikely as a function for these repeated structures. Nevertheless, based on fragmentation and extraction experiments, similar repetitive structures in yeast adenylatcyclase (26 repeats) and *Drosophila* chaoptin (41 repeats) were proposed to be responsible for tight membrane association (Kataoka *et al.*, 1985; Reinke *et al.*, 1988). But taking into account the extremely tight protein–protein interaction between RAI and angiogenin (K_I = 0.1 nM) (Shapiro and Vallee, 1987), one could speculate that the membrane association of these proteins is mimicked by very tight binding to integral membrane proteins. The high number of repeats found in these proteins might reflect an especially strong protein–protein interaction, achieved by multiple specific contact sites.

Therefore a plausible function of the found repeated structures is tight and specific protein binding. The RAI modules show that an adaptation to the binding of very different proteins might be accomplished by modulation of the length, number and amino acid composition of these repeats or by further structural variations like disulphide bonds and glycoslyation. Thus similar repetitive motifs will probably be found in many more proteins comprising a novel protein superfamily of general significance.

The finding of such repeats in RAI will be crucial for the elucidation of the structure–function relationship of this module family. RAI is the first member of this family with a well-studied and defined function. It is available in milligram quantities by potent purification procedures (Blackburn, 1979) and its substrates, especially RNase, number among the hitherto most extensively studied proteins. Therefore, the elucidation of the tertiary structure of RAI and RAI–substrate complexes should be feasible.

In conclusion the presented cDNA cloning and characterization of ribonuclease/angiogenin inhibitor provide not only potent tools for studying the regulation of angiogenesis and RNA turnover, but disclose a novel protein superfamily with a common highly versatile protein-binding module and will contribute to our understanding of molecular protein evolution.

Materials and methods

Preparation and screening of cDNA libraries

RNA was isolated from HeLa cells by phenol extraction of Mg^{2+} precipitated polysomes (Palmiter, 1974) and the poly(A) fraction was obtained by affinity chromatography on oligo(dT) cellulose (Aviv and Leder, 1972). The purified mRNA was used to prepare double-stranded cDNA (Gubler and Hoffmann, 1983), which was cloned into λgt11 (Huynh *et al.*, 1985). The DNAs of the positive phages were purified and after *Eco*RI digestion and Southern blotting reprobed with radiolabelled RAI-specific oligonucleotide (Maniatis *et al.*, 1982).

The purified HeLa mRNA was also used to construct a eukaryotic expression cDNA library (Okayama and Berg, 1983). Starting from 5 μg of mRNA this cloning procedure yielded 800 000 independent recombinants. *Bam*HI digestion of 12 of these recombinants showed an average insert size of 1.1 kbp. The largest insert (900 bp) of the positive λgt11 recombinants was labelled to high sp. act. with [^{32}P]dCTP (Feinberg and Vogelstein, 1983) and was used to screen 10^6 recombinants of this library by colony hybridization (Maniatis *et al.*, 1982).

Northern blot analysis

Total cytoplasmic RNA was prepared from human monocytes by a modified hot phenol method. RNA was size fractionated on a 1% agarose–formaldehyde gel and transferred to a positively charged nylon membrane (NYP, Hoefer Scientific). RAI-specific cDNA was labelled with [^{32}P]dCTP by random hexanucleotide priming, hybridized in 50% formamide by 42°C and washed at 65°C to a final stringency of 0.015 M NaCl.

DNA sequencing

cDNA inserts were subcloned into the Bluescript plasmid (Stratagene, San Diego, CA). The insert sequence was determined by the dideoxy chain termination method (Sanger *et al.*, 1977) using universal primers that hybridize to the vector DNA. In three cases the obtained sequence was extended using 17mer oligonucleotide primers.

Sequence comparisons

Internal homologies of RAI were found by visual inspection. Related proteins

were found with the FAST-P program (Lipman and Pearson, 1985) and evaluated with the RELATE program (Barker *et al.*, 1978) on the MIPS protein database at the Max Planck Institut in Martinsried/Munich.

Acknowledgements

We are grateful to H.Reinke for peptide sequencing, E.deGroot for oligonucleotide synthesis, F.Lottspeich and J.Kellermann for help in searching current protein databases and J.Willingale-Theune for carefully reading this manuscript.

References

Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408–1412.
Barker,W.C., Ketcham,L.K. and Dayhoff,M.O. (1978) In Dayhoff,M.O. (ed.), *Atlas of Protein Sequence and Structure*. Natl. Biomed. Res. Found., Washington, DC, Vol. 5, Suppl. 3, pp. 359–362.
Blackburn,P. (1979) *J. Biol. Chem.*, **254**, 12484–12487.
Blackburn,P., Wilson,G. and Moore,S. (1977) *J. Biol. Chem.*, **252**, 5904–5910.
Brewer,E.N., Foster,L.B. and Sells,B.H. (1969) *J. Biol. Chem.*, **244**, 1389–1392.
Chou,P.Y. and Fasman,G.O. (1978) *Annu. Rev. Biochem.*, **47**, 251–276.
Day,A., McQuillan,C.I., Termine,J.D. and Young,M.R. (1987) *Biochem. J.*, **248**, 801–805.
Eisenberg,D. (1984) *Annu. Rev. Biochem.*, **53**, 595–623.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **123**, 6–13.
Ferencz,A., Hidvegi,E.J., Szabo,L.D. and Varteresz,V. (1973) *Radiat. Res.*, **55**, 304–317.
Fett,J.W., Strydom,D.J., Lobb,R.R., Alderman,E.M., Bethune,J.L., Riordan,J.F. and Vallee,B.L. (1985) *Biochemistry*, **24**, 5480–5486.
Frost,L.S., Paranchych,W. and Willetts,N.S. (1984) *J. Bacteriol.*, **160**, 395–401.
Greif,R.L. and Eich,E.F. (1977) *Metab. Clin. Exp.*, **26**, 851–856.
Gubler,U. and Hoffmann,B.J. (1983) *Gene*, **25**, 263–269.
Hashimoto,C., Hudson,K.L. and Anderson,K.V. (1988) *Cell*, **52**, 269–279.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning. A Practical Approach*. IRL Press, Oxford, Vol. 1, pp. 49–78.
Kataoka,T., Broek,D. and Wigler,M. (1985) *Cell*, **43**, 493–505.
Kershner,R.M. and Meyer,W.L. (1976) *Biochem. Biophys. Res. Commun.*, **70**, 513–517.
Kraft,N. and Shortman,K. (1980) *Biochim. Biophys. Acta*, **217**, 164–175.
Krusius,T. and Ruoslahti,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7683–7687.
Lipman,D. and Pearson,W. (1985) *Science*, **227**, 1435–1441.
Little,B.W. and Meyer,W.L. (1970) *Science*, **170**, 747–749.
Lopez,J.A., Chung,D.W., Fujikawa,K., Hagen,F.S., Papayannopoulou,T. and Roth,G.J. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5615–5619.
Lopez,J.A., Chung,D.W., Fujikawa,K., Hagen,F.S., Davie,E.W. and Roth,G.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2135–2139.
Majocha,R.E., Sajdel-Sulkowska,E.M., Rodenrys,A., Ventosa-Michelman, M. and Marotta,C.A. (1987) *Soc. Neurosci. Abstr.*, **13**, 819.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Okayama,H. and Berg,P. (1983) *Mol. Cell. Biol.*, **3**, 280–289.
Palmiter,R.D. (1974) *Biochemistry*, **13**, 3606–3615.
Patthy,L. (1985) *Cell*, **41**, 657–663.
Patthy,L. (1987) *J. Mol. Biol.*, **198**, 567–577.
Quirin-Stricker,C., Gross,M. and Mandel,P. (1968) *Biochim. Biophys. Acta*, **159**, 75–80.
Reinke,R., Krantz,D.E., Yen,D. and Zipursky,S.L. (1988) *Cell*, **52**, 291–301.
Roth,J.S. (1956) *Biochim. Biophys. Acta*, **21**, 34–43.
Roth,J.S. (1967) *Methods Cancer Res.*, **3**, 151–243.
Sajdel-Sulkowska,E.M. and Marotta,C.A. (1984) *Science*, **225**, 947–949.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sarkar,N.K. (1970) *Life Sci.*, **9**, 675–681.
Schweiger,M., Schneider,R., Hirsch-Kauffmann,M., Auer,B., Klocker,H., Scherzer,E., Thurnher,M., Herzog,H., Vosberg,J.P. and Wagner,E.F. (1988) In Kleinkauf,H., vanDöhren,H. and Jaenicke,L. (eds), *The Roots of Modern Biochemistry*. Walter de Gruyter & Co., Berlin, pp. 379–387.
Shapiro,R. and Vallee,B.L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2238–2241.
Shortman,K. (1962) *Biochim. Biophys. Acta*, **61**, 50–55.
Takahashi,N., Takahashi,Y. and Putnam,F.W. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1906–1910.
Titani,K., Takio,K., Handa,M. and Ruggeri,Z.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5610–5614.
Traub,P., Zillig,W., Millete,R.L. and Schweiger,M. (1966) *Hoppe-Seyler's Biol. Chem.*, **343**, 261–275.
Vogel,K.G., Paulsson,M. and Heinegard,D. (1984) *Biochem. J.*, **223**, 587–597.
Weiner,H.L., Weiner,L.H. and Swain,J.L. (1987) *Science*, **237**, 280–282.

Received on July 19, 1988; revised on September 22, 1988