Vol. 136, No. 1, 1986 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 14, 1986 Pages 396-403

STRUCTURE AND ANALYSIS OF THE BOVINE ATRIAL NATRIURETIC PEPTIDE PRECURSOR GENE

George P. Vlasuk*, Judith Miller,
Gerard H. Bencen and John A. Lewicki**

California Biotechnology, Inc.
2450 Bayshore Frontage Road
Mountain View, CA 94043

Received March 7, 1986

Abstract: The isolation and sequence analysis of the gene encoding the bovine atrial natriuretic peptide (ANP) precursor is described. The bovine-ANP coding sequences are located on three exons which are interrupted by two intervening sequences. Comparison of the bovine, human, rat and mouse ANP gene sequences reveals a common organization of introns and exons and a high degree of sequence homology in the 5'-flanking and coding regions. Examination of the pre-proANP amino acid sequence derived from the bovine gene with those from rat, mouse and human, indicates a high degree of sequence homology in both the amino-terminal and biologically-active carboxy-terminal ANP region. The latter region in the bovine sequence resembles its human counterpart except for a carboxy-terminal Arg-Arg dipeptide.

It is now well established that the atria of the mammalian heart secretes polypeptides which have potent natriuretic, diuretic and vasorelaxant activities (1, 2). These hormones termed atrial natriuretic peptides (ANP), have been proposed to play a major role in maintaining cardiovascular homeostasis (1, 2). Elucidating the steps involved in the biosynthesis of ANP has been aided through the analysis of cloned gene sequences for human (3, 4, 5), rat (6, 7) and mouse (3) ANP. Through these studies it has been established that ANP comprises the carboxy-terminal portion of a larger precursor molecule (pre-ProANP). Though a circulating ANP form of 28 amino acids has been recently identified (8, 9), the mechanisms involved in

*Present Address: Merck, Sharp and Dohme Research Labs.
**To whom correspondence should be addressed.

0006-291X/86 $1.50

the regulation of synthesis and processing of pre-proANP to yield this peptide are unknown.

Various investigators have utilized in vitro assays employing cells and tissues of bovine origin, to investigate the mechanisms responsible for the biological actions of ANP (10-13). It has been suggested that there may be variability in the physiological action of ANP between different species (14). Therefore, it is important that the structure of the bovine form of ANP is known to correlate the results obtained from in vitro assays with ANP's derived from other species.

We now describe the structure and nucleotide sequence of the bovine-ANP gene and corresponding pre-proANP amino acid sequence.

Materials and Methods

Genomic DNA cloning - A bovine genomic library in bacteriophage λ Charon 28 (15) (a gift of Dr. F.M. Rottman) was screened with a nick-translated human-ANP cDNA (J.C. Fiddes, unpublished results) as described (16). DNA from one hybridizing phage was purified and submitted to restriction enzyme mapping using Southern blot analysis (17). A single 1.76 kb PstI-EcoRI fragment was subcloned into PUC 9 (18) and designated pBGPE-1.

DNA sequence analysis - Plasmid pBGPE-1 was submitted to limited restriction enzyme mapping to define suitable fragments for cloning into bacteriophage M13 vectors mp8 and mp9 (19). M13 subclones were sequenced on both strands with the dideoxy-chain termination method (20), using the 17-bp universal primer.

Results and Discussion

Analysis of the bovine-ANP gene structure - Using a human-ANP cDNA as probe, we isolated a single phage containing the bovine-ANP gene from a genomic library in bacteriophage λ. We subcloned the complete hybridizing region on a 1.76 kb PstI-EcoRI fragment and determined its nucleotide sequence (Fig. 1).

The bovine-ANP gene is comprised of three coding exons interrupted by two intervening sequences of 98 and 525-bp. This organization is similar to the other mammalian ANP genomic sequences (3, 4, 6). The positions of the intron-exon junctions

```
                                                                    50                                                                  100
CTGCAGCTGA GGGTCCTGGG GGTTGTCGGG GCTGCTCAAG GCAGAGGGGC TGTGACAAGC AGGCTGGACT GATAACTTTA AAAGGGCATC TTCTGCTGCT
PstI
                                                                   150                                                                  200
TCCTCACTCA GCTGCTTTAT CACTGCAAGT GACAGAATGG GGAGGGTTCC GTCCCTCTCC CGGACGAGCT CCCAGAGAGC CAGGGGGCTA TAAAAAGAGG

                                                                   250                                                                  300
AGGCTCAGGG CAGCTGGGAG ACAGAGACGG ACAAAGGCCA ACAGCAAAAG GCCAAAGAGG ACAGGGAGGA GGCAGCAAGC ACCAGACCGA CCATTCCTTG

                                                                   350
 ACCGACGCCAGC ATG GGC TCC TCC GCC ATC ACC GTG AGC TTC CTC CTC TTT CTG GCA TTT CAG CTC CCA GGG CAA ACA GGA GCA
              MET Gly Ser Ser Ala Ile Thr Val Ser Phe Leu Leu Phe Leu Ala Phe Gln Leu Pro Gly Gln Thr Gly Ala

                    400                                                                 450
AAT CCC GTG TAT GGC TCT GTG TCC AAT GCA GAC CTG ATG GAT TTC AAG gtagggcc agggaacggc gatggtctgg ggctgagggg gtt
Asn Pro Val Tyr Gly Ser Val Ser Asn Ala Asp Leu MET Asp Phe Lys

                              500                                                                   550
gtgacat tgtgccaggc gagcgagacc tctccctttc cctgttttcc ttttgtaaag AAT TTG CTG GAC CGT TTG GAG GAC AAG ATG CCT TTA
                                                                                  Asn Leu Leu Asp Arg Leu Glu Asp Lys MET Pro Leu

                                                  600                                                                           650
GAA GAT GAG GCT GTG CCC TCA CAA GTA CTA AGT GAG CAG AAT GAA GAA GCT GGG GCC CCT CTC AGC CCC CTT TCA GAG ATG CCT
Glu Asp Glu Ala Val Pro Ser Gln Val Leu Ser Glu Gln Asn Glu Glu Ala Gly Ala Pro Leu Ser Pro Leu Ser Glu MET Pro

                                                                  700
CCC TGG ATG GGG GAG GTC AAC CCA GCC CAG AGA GAG GGG GGC GTC CTC GGG CGG GGC CCC TGG GAA TCC TCC GAT AGA TCT GCC
Pro Trp MET Gly Glu Val Asn Pro Ala Gln Arg Glu Gly Gly Val Leu Gly Arg Gly Pro Trp Glu Ser Ser Asp Arg Ser Ala

                    750                                                                 800
CTC CTG AAG AGC AAG CTG AGG GCA CTG CTC ACT GCC CCT CGG AGC CTG CGG AGG TCC AGC TGC TTC GGG GGA AGG ATG GAC AGG
Leu Leu Lys Ser Lys Leu Arg Ala Leu Leu Thr Ala Pro Arg Ser Leu Arg Arg Ser Ser Cys Phe Gly Gly Arg MET Asp Arg

                                                  850                                                                           900
ATT GGA GCC CAG AGT GGA TTG GGC TGC AAC AGC TTC CGG gtaagaggacctg agaatggaaa tgggatgggg aggaaggaaa ttgtggcttc
Ile Gly Ala Gln Ser Gly Leu Gly Cys Asn Ser Phe Arg

                                                  950                                                                  1000
attgaagttc aaaccttgtg aaagaacatc gccagggaat gccttcagta  ggaaaggggac agcatagaag caaccccttt gaaatttctg ccccaacttg

                                                 1050                                                                  1100
gcagggagga gggtgtgctc tgagtctcag gacaatgata ccaacctagc  tacagttttc tgagagaatg ctaagaaaaa aagactttac tgccacgagc

                                                 1150                                                                  1200
actggggact taaattgttc atggggccaa ataacctgtg ctttgctgat  tggtagtttg tgtcctttgc agaatcatca gatcccaaag gattgaaatt

                                                 1250                                                                  1300
gagcaggact gactttacta gttcctaatg ggcaatttgt ttaccagttt  atagaagtca gagggtcatc aggctggagt ggaggctggt gggaagggag

                                                 1350                                                                  1400
cacagtctga tgaagctggc ttttccagtg gagtcaggtc accaaaccaa acatgtctct gctctcttgtag TAT CGA AGA TAA TGGCCA GGGAGGAAAA
                                                                                  Tyr Arg Arg  .

                                                 1450                                                                  1500
GGCAGGCCAG GCCCTGGGCA GTCTTCAAGA GAATCCCCTG GGGTCTCTCA  CTCAACTTTG TCGCATCTGG TTGCCATCAA GTTGAGCTGT GACCGAGCAT

                                                 1550                                                                  1600
TCAAGCATCA GCTTCTTGTC AACATTTCTC ACATTTTATG CTAAATGTAG  ACAAAGTGAT TTAACTGTGG CCTTCTCCAC CTCTCCCACC CATGTGTTAA

                                                 1650                                                                  1700
GTTTTAATCA CCTGTTACCA ACATCAGTTT GAAAATGAAT AAACTTCAGC  ACCATGGACA GAAGCAGTAG GCTCGGGTTG GTGTGATTTC TTTCATTTCC

                                                 1750                  1769
GGAAGGGAGT TCAGCCTGAT ACTCCTTGTC ATTTTACCTT TTGTTGGAGA GAAGAATTC
                                                                      EcoRI
```

Figure 1 - Sequence of the bovine gene encoding the ANP-precursor. The complete 1,769-bp nucleotide sequence of the PstI-EcoRI fragments in pBGPE-1 is shown. Nucleotides are numbered from the PstI site. Introns are represented by lower case letters. The amino acid sequence is shown below the corresponding nucleotides. The potential TATA and polyadenylation (AATAAA) sequences are boxed. A putative cap site in the 5′-flanking region is circled.

were defined by comparing the bovine genomic-ANP sequence with other ANP gene sequences (3, 4, 6) and the homology with consensus splice sequences (21).

The first exon encodes the 5′-untranslated region, a 24 amino acid signal peptide and 16 amino acids following a putative

signal peptidase cleavage site which occurs in the same position as other pre-proANP sequences (3-7). The second exon encodes the remainder of the pre-proANP coding sequence including the biologically-active ANP region. The remaining carboxy-terminal -Tyr-Arg-Arg residues and 3'-untranslated region reside on the third exon.

Analysis of the 5' flanking region of the bovine-ANP gene revealed a perfect TATAAAA sequence located 118 bp upstream from the ATG translation initiation codon. This sequence is similar to the TATA box which has been found in the 5'-flanking regions of many eucaryotic genes (22). A putative cap site was assigned approximately 30 bp downstream of the TATAAAA sequence after comparison with similar regions in the rat (6) and human (3, 4) ANP-genomic sequences. The 3'-untranslated region residing on exon III, contains a single AATAAA polyadenylation signal (23) which is similar to that observed for the mouse and human-ANP sequences (3-5).

Comparison of the bovine-ANP genomic and pre-proANP sequence with rat, mouse and human sequences - The overall homologies in the nucleotide sequence of the bovine-ANP gene with the corresponding sequences from rat (6), mouse (3) and human (3, 4) (excluding the intervening sequences) are given in Table I. The 5'-flanking sequence of the four genes are homologous which may reflect the importance of this region in the regulation of pre-proANP gene expression. There is also a high degree of homology in the nucleotide sequence corresponding to the coding regions of both exons and a lesser degree in the 3'-untranslated region residing on exon III. The first intron of the bovine gene (98-bp) is shorter than the corresponding intron in the human, rat and mouse genes whereas the larger of the two introns (525-bp) is slightly

Table I

Percent homology between the bovine-ANP gene and those of human, mouse and rat

	HUMAN	RAT	MOUSE
5'-flanking region (1-312)	75.6%	70.5%	72.4%
Exon I (313-432)	80.8%	80.8%	79.2%
Exon II (531-857)	89.3%	82 %	81.3%
Exon III (1383-1653)	70.5%	53.1%	59 %

Homology was calculated from the nucleotide sequence shown in Figure 1 and those reported for the human (3, 4), rat (6) and mouse (3) ANP genomic sequences. Numbers in parenthesis indicate base positions in the bovine-ANP sequence shown in Fig. 1.

longer than the corresponding rat sequence (391-bp) but shorter than the mouse (527-bp) and human (1093-bp) sequences.

The comparison of the derived amino acid sequence of the bovine-preproANP with those for rat, mouse and human is shown in Figure 2. The length of the bovine pre-proANP sequence (152 amino acids) is identical to that of rat and mouse and one amino acid longer than the human form. The biologically active ANP region which is located at the carboxy-terminal end of the pre-proANP molecule is extremely well conserved among the four species. The only difference among the four ANP peptides is at position 135 (Fig. 2) where a methionine is located in the bovine and human sequences and isoleucine in the rat and mouse. In addition, the carboxy-terminus of the ANP's differ, with bovine, rat and mouse terminating in the sequence Tyr-Arg-Arg and human in Tyr. This latter heterogenity at the carboxy-terminus presumably arises from a single base change in the codon for the first Arg (position 1372 in the bovine sequence, Fig. 1). However, the carboxy-terminal Arg-Arg dipeptide predicted from the gene sequence is not found on the circulating form of the

```
                                                    10
BOVINE  Met Gly Ser Ser Ala — — Ile Thr Val Ser Phe Leu Leu Phe Leu Ala Phe Gln Leu
HUMAN       Ser     Phe Ser Thr Thr                     Leu
MOUSE           .   Phe Ser —       .   Leu Gly     Phe     Val             Trp
RAT             .   Phe Ser —       .   Lys Gly     Phe                     Trp

        20                                          30
BOVINE  Pro Gly Gln Thr Gly Ala|Asn Pro Val Tyr Gly Ser Val Ser Asn Ala Asp Leu Met
HUMAN   Leu             Arg             .   Met     Asn Ala
MOUSE           His Ile                 .           Ser Ala             Thr
RAT             His Ile                 .           Ser Ala             Thr

            40                                          50
BOVINE  Asp Phe Lys Asn Leu Leu Asp Arg Leu Glu Asp Lys Met Pro Leu Glu Asp Glu Ala
HUMAN                               .   His     Glu                             Val
MOUSE                               .   His     Glu             Val             Val
RAT                                 .   His     Glu             Val             Val

                60                                          70
BOVINE  Val Pro Ser Gln Val Leu Ser Glu Gln Asn Glu Glu Ala Gly Ala Pro Leu Ser Pro
HUMAN           Pro                 .   Asp Pro                     Ala
MOUSE   Met     Pro     Ala         .           Thr                 Ala         Ser
RAT     Met     Pro     Ala         .           Thr Asp             Ala         Ser

                    80                                          90
BOVINE  Leu Ser Glu Met Pro Pro Trp Met Gly Glu Val Asn Pro Ala Gln Arg Glu Gly Gly
HUMAN       Pro     Val             .   Thr             Ser                 Asp
MOUSE       Pro     Val             .   Thr                     Pro Leu     Asp     Ser
RAT                 Val             .   Thr                     Ser         Asp

                100                                         110
BOVINE  Val Leu Gly Arg Gly Pro Trp Glu Ser Ser Asp Arg Ser Ala Leu Leu Lys Ser Lys
HUMAN   Ala                         .   Asp
MOUSE   Ala Ser Arg     Ser         .   Asp Pro
RAT     Ala                         .   Asp Pro

                            120                                         130
BOVINE  Leu Arg Ala Leu Leu Thr Ala Pro Arg Ser Leu Arg Arg Ser Ser Cys Phe Gly Gly
HUMAN                               .
MOUSE                       Ala Gly .
RAT                         Ala Gly .

                                140                                         150
BOVINE  Arg Met Asp Arg Ile Gly Ala Gln Ser Gly Leu Gly Cys Asn Ser Phe Arg Tyr Arg
HUMAN                               .                                           -
MOUSE       Ile                     .
RAT         Ile                     .

BOVINE  Arg
HUMAN    -
MOUSE
RAT
```

Figure 2 - Comparison of the pre-proANP amino acid sequence derived from the bovine gene shown in Fig. 1 with the human (3-5), rat (6, 7) and mouse (3) sequences. The human, rat and mouse sequences are shown only where they differ from the bovine sequence. The putative signal peptide region (....), signal peptidase cleavage site (11), and biologically active ANP region (__) are indicated.

ANP hormone (8, 9). Thus, removal of this dipeptide results in a common carboxy-terminus for all four ANP species.

The amino-terminal signal sequence region lacks the homology seen in the remainder of the coding region of the pre-proANP molecules. However, this is not surprising since the overall structure of this region is important to its function in directing the ANP-precursor to the secretory pathway. The remainder of the amino-terminal segment of the ANP-precursor is

conserved among the four species. This homology indicates this region may have an important role in ANP function. However, to date the exact role of the amino-terminal portion of the ANP-precursor remains to be elucidated.

The determination of the bovine-ANP gene and corresponding amino acid sequence will be useful in studies comparing the biological activities of ANP's in model systems derived from different mammalian species.

Acknowledgement

We wish to thank Dr. Fritz Rottman for providing the bovine genomic library and Dr. John Fiddes for the helpful advice and critical reading of the manuscript. The expert secretarial support of Candice Remington and Mara Matsumura is appreciated.

References

1. Maack, T., Camargo, M.J.F., Kleinert, H.D., Laragh, J.H. and Atlas, S.A. (1985) Kidney Int. 607-615.
2. Needleman, P., Adams, S.P., Cole, B.R., Currie, M.G., Geller, D.M., Michener, M.L., Saper, C.B., Schwartz, D. and Standaert, D.G. (1985) Hypertension 7, 469-482.
3. Seidman, C.E., Bloch, K.D., Klein, K.A., Smith, J.A. and Seidman, J.G. (1984) Science 226, 1206-1209.
4. Greenberg, B.D., Bencen, G.H., Seilhamer, J.J., Lewicki, J.A. and Fiddes, J.C. 91984) Nature 312, 656-658.
5. Oikawa, S., Imai, M., Veno, A., Tanaka, S., Nogushi, T., Nakazato, H., Kangawa, K., Fukuda, A. and Matsuo, H. (1984) Nature 309, 724-726.
6. Argentin, S., Nemer, M., Drouin, J., Scott, G.K., Kennedy, B.P. and Dvaies, P.L. (1985) J. Biol. Chem. 260, 4568-4571.
7. Yamanaka, M., Greenberg, B., Johnson, L., Seilhamer, J., Brewer, M., Friedemann, T., Miller, J., Atlas, S., Laragh, J., Lewicki, J. and Fiddes, J. (1984) Nature 309, 719-722.
8. Schwartz, D., Geller, D.M., Manning, P.T., Siegel, N.R., Fok, K.F., Smith, C.E. and Needleman, P. (1985) Science 229, 397-400.
9. Thibault, G., Lazure, C., Schiffrin, E.L., Gutkowska, J., Chartier, L., Garcia, R., Seidah, N.G., Chretien, M. Genest, J. and Cantin, M. (1985) Biochem. Biophys. Res. Comm. 127, 433-442.
10. Schenk, D.B., Johnson, L.K., Schwartz, K., Sista, H., Scarborough, R.M. and Lewicki, J.A. (1985) Biochem. Biophys. Res. Comm. 127, 433-442.
11. Misono, K.S., Grammer, R.T., Rigby, J.W. and Inagami, T. (1985) Biochem. Biophys. Res. Comm. 130, 994-1001.
12. Hirose, S., Akiyama, F., Shinjo, M., Ohno, H. and Murakami, K. (1985) Biochem. Biophys. Res. Comm. 130, 574-579.
13. DeLean, A.K., Racz, J., Gutkowska, T-T., Nguyen, M., Cantin, M. and Genest, J. (1984) Endocrinol. 115, 1636-1638.
14. Wakitani, K., Oshima, T., Loewy, A.D., Holmberg, S.W., Cole, B.R., Adams, S.P., Fok, K.F., Currie, M.G. and Needleman, P.

(1985) Circ. Res. 56, 621-627.
15. Woychik, R.P., Camper, S.A., Lyons, R.H., Horowitz, S., Goodwin, E.C. and Rottman, F.M. (1982) Nucleic Acids Res. 10, 7197-7210.
16. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
17. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
18. Vieira, J. and Messing, J. (1982) Gene 19, 259-268.
19. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
21. Breathnach, R. and Chambon, P.A. (1981) Ann. Rev. Biochem. 50, 349-383.
22. Corden, J., Wasyik, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
23. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.