ERV3 LTR's did not appear to contain a long-direct repeat. However, the LTR's of two HTLV variants (*9, 10*) and an isolate of the related BLV (*11*) also lack these repeats and are certainly expressed. Four regions within the ERV3 LTR's displayed partial homology with the consensus sequence GTGG$^{AAA}_{TTT}$G that is present in many viral enhancers and is essential for the activity of the SV40 enhancers (*2*). These regions, beginning at LTR nucleotides 455, 505, and 563 in U5 and at nucleotide 597 within the putative primer binding site, each shared six out of eight nucleotide identities with the consensus sequence. The significance of these sequences is unclear, since retroviral enhancers characterized thus far are located in the U3 region of the LTR.

The LTR sequences of HTLV variants I and II are unrelated to each other except for the promoter sequences and a 21–base pair sequence repeated three times in the U3 region of each genome at comparable positions. These repeats are imputed to be equivalent to the enhancer sequences of other retroviruses (*9*). Similarly, the ERV3 LTR's contained a short repeated sequence (underlined in Fig. 2). This sequence, GAAAAACAAG, is also found in the U3 region of both BaEV (*12*) and HTLV-II (*9*) as a single copy. Further experiments are necessary to define whether these or other ERV3 LTR sequences function in enhancement of proviral expression.

Analysis of the nucleotide sequence of the ERV3 LTR's revealed a region of close nucleotide homology with the BaEV LTR that was responsible for their hybridization. Within a stretch of 41 nucleotides beginning at nucleotide 373 in the ERV3 3′ LTR there was only one nucleotide change with respect to the BaEV LTR (*12*) (two nucleotides were different in the ERV3 5′ LTR). This sequence was found in the U3 region of both ERV3 and BaEV LTR's but farther upstream in BaEV at U3 nucleotide 102. A human provirus characterized earlier, ERV1 (*13*), also contains this sequence but in the U5 region. We observed no other homology between the LTR's of ERV3 and those of other sequenced retroviruses, including HTLV (*9, 10*).

Because of the mechanism of retroviral replication, the two ERV3 LTR's were presumably identical when the provirus first integrated into the ancestral human genome (*1*). The two LTR sequences are not presently 100 percent homologous (Fig. 2). The 5′ LTR is 593 nucleotides long, whereas the 3′ LTR is only 590 nucleotides long. There are also 52 nucleotide substitutions spaced throughout the two LTR's, corresponding to an 8.8 percent divergence.

The possible use of $tRNA^{Arg}$ as a primer for replication of ERV3 suggests a separate lineage for this provirus. Because the ERV1 provirus is missing the 5′ LTR, the primer binding site cannot be identified. While this precludes classification of ERV3 and ERV1 in the same retroviral lineage by this criterion, comparisons of the DNA sequences from regions of the *gag* and *pol* genes indicate that these two proviruses are more closely related to each other than either is to BaEV or M-MuLV (Moloney murine leukemia virus) (*4*). It will be interesting to determine whether other BaEV LTR-hybridizing clones isolated from human DNA (*14*) also share this primer binding site and the 41-nucleotide homology.

In conclusion, the ERV3 LTR's contain sequences homologous to known transcriptional regulatory elements. Without knowing whether the ERV3 LTR's are, or were, used for expression, we were unable to assess the relevance of the nucleotide changes between them. It is unlikely that all the changes would have occurred within only one LTR even if the other was selected for by providing a necessary function, since every change would not be expected to be deleterious. Thus, the divergence of the two ERV3 LTR's by 8.8 percent does not necessarily preclude function by either of them. These sequences may therefore be capable of directing the expression of either ERV3 or nearby host genes. Consequently, the ERV3 provirus is a suitable probe with which to look for expression of endogenous retroviruses in both normal and tumor tissues of humans.

CATHERINE D. O'CONNELL
MAURICE COHEN
Laboratory of Molecular Virology and Carcinogenesis, Litton Bionetics Incorporated–Basic Research Program, National Cancer Institute, Frederick Cancer Research Facility, Frederick, Maryland 21701

References and Notes

1. H. M. Temin, *Cell* **27**, 1 (1981).
2. Y. Gluzman and T. Shenk, *Enchancers and Eukaryotic Gene Expression* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983).
3. W. S. Hayward, B. G. Neel, S. M. Astrin, *Nature (London)* **290**, 475 (1981).
4. C. O'Connell, S. O'Brien, W. G. Nash, M. Cohen, *Virology* **138**, 225 (1984).
5. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
6. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
7. D. H. Gauss and M. Sprinzl, *Nucl. Acids Res.* **11**, 1 (1983).
8. H. R. Chen and W. C. Barker, *ibid.* **12**, 1767 (1984).
9. K. Shimotohno, D. W. Golde, M. Miwa, T. Sugimura, I. S. Y. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 1079 (1984).
10. M. Seiki, S. Hattori, M. Yoshida, *ibid.* **80**, 3618 (1983).
11. D. Couez, J. Deschamps, R. Kettmann, R. M. Stephens, R. V. Gilden, A. Burny, *J. Virol.* **49**, 615 (1984).
12. T. Tamura, M. Noda, T. Takano, *Nucl. Acids Res.* **9**, 6615 (1981).
13. T. I. Bonner, C. O'Connell, M. Cohen, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4709 (1982).
14. M. Noda, M. Kurihara, T. Takano, *Nucl. Acids Res.* **10**, 2865 (1982).
15. We thank E. Brownell for helpful discussions and J. Clarke for technical assistance. Sponsored by the National Cancer Institute, Department of Health and Human Services, under contract NO1-CO-23909 with Litton Bionetics, Inc.

21 May 1984; accepted 2 August 1984

Nucleotide Sequences of the Human and Mouse Atrial Natriuretic Factor Genes

Abstract. *Mouse and human atrial natriuretic factor (ANF) genes have been cloned and their nucleotide sequences determined. Each ANF gene consists of three coding blocks separated by two intervening sequences. The 5′ flanking sequences and those encoding proANF are highly conserved between the two species, while the intervening sequences and 3′ untranslated regions are not. The conserved sequences 5′ of the gene may play an important role in the regulation of ANF gene expression.*

Atrial natriuretic factor (ANF), a potent vasoactive peptide synthesized in mammalian atria, is thought to play a key role in cardiovascular homeostasis (*1–6*). Analysis of cloned DNA sequences complementary to AFN messenger RNA's (mRNA's) has defined a precursor molecule (preproANF) from which this cardiac hormone is derived (*7–11*). The ANF gene is actively transcribed; preproANF mRNA comprises 1 to 3 percent of atrial mRNA (*7, 12*). To study further the transcription, processing, and regulated expression of this cardiac hormone, we have cloned the genes encoding murine and human ANF and determined their nucleotide sequences.

The single ANF gene in rodents and humans (*7, 10*) hybridizes to a cloned rat ANF complementary DNA (cDNA) probe (*7*). Procedures for cloning single-copy eukaryotic genes from bacteriophage libraries are well established (*13*). Five bacteriophage clones contain-

ing ANF gene sequences were isolated from 500,000 phage plaques derived from a human genomic library. Southern blot analysis of the DNA from these five bacteriophages defined a 3.3-kilobase (kb) Bam HI fragment that hybridized to the rat ANF cDNA probe. This fragment was subcloned into pBR322 for subsequent nucleotide sequence characterization (Fig. 1).

The murine ANF gene was isolated from size-fractionated (*14*) Bam HI–digested mouse DNA (strain B10.LP). A 10-kb fragment encoding ANF was selected by Southern blot hybridization (*15*) and cloned into pBR322. Coding segments of the ANF gene were identified with a rat cDNA probe or an oligonucleotide derived from the 3′ end of the human cDNA sequence (Fig. 1). This synthetic oligonucleotide was used because a restriction fragment from the 3′ end of the rat mRNA did not hybridize to the human gene (*12*). Nucleotide sequences (Fig. 2) of the human and mouse ANF genes were determined by the dideoxy chain termination procedure (*16*). Locations of intervening sequences were defined by comparison of genomic and cDNA sequences.

Nucleotide sequence analysis showed that the ANF gene contains many features of a typical eukaryotic gene. The genes contain a TATAA box (A, adenine; T, thymine), two intervening sequences bounded by splicing signals (GT-AG; G, guanine), and an AATAAA polyadenylation addition signal. Initiation of ANF mRNA transcription probably begins 27 base pairs (bp) after the TATAA box (cap site; Figs. 1 and 2). PreproANF is encoded in three separate segments (Figs. 2 and 3). Coding block I encodes the hydrophobic leader segment and the first 20 amino acid residues of proANF. Coding block II encodes the remainder of the proANF except for one amino acid of human ANF and three amino acids of mouse ANF. These carboxyl-terminal amino acids are encoded by coding block III.

Rat ANF contains two amino acids (Arg-Arg) more than human ANF (*3, 6*). Comparison of nucleotide sequences from rat and human cDNA clones (*7, 11*) suggests that this difference is due to a single nucleotide change (C to T; C, cytosine) between these genes, defining a termination codon in humans rather than the arginine codon in rodents.

Because ANF mRNA makes up 1 to 3 percent of the total polyadenylated [poly(A)$^+$] RNA of the rat atria (*7, 12*), mutations in this gene that adversely affect production of functional ANF would presumably be deleterious and lost during evolution. In contrast, sequences that are important for the regulated expression and production of ANF should be conserved in rodent and human genomes. We compared the nucleotide sequences of the ANF genes of these two species to identify the conserved regions (Fig. 2 and Table 1).

Table 1. Percent homology between different segments of mouse and human ANF genes. Homology was calculated from the nucleotide sequences in Fig. 2. The percent amino acid homology is reported for coding blocks I and II. A 25 percent homology was assigned to the second intervening sequence (IVS2) because neither dot matrix comparison (*19*) nor nucleic acid hybridization (*13*) revealed sequence similarities.

Gene segment	Homology (%)	
	Nucleotide	Amino acid
5′ Flank	79	
5′ UT*	61	
Coding block I	76	63
IVS1	56	
Coding block II	83	83
IVS2	(25)	
Coding block III	60	

*U, uracil.

The 5′ flanking sequences (1 to 576 bp) of the human and mouse ANF genes are highly conserved (Table 1). Because these sequences may play an important role in directing ANF transcription, they were screened for previously characterized regulatory sequences. Enhancer sequences that regulate expression of closely linked genes have been found near a number of viral and eukaryotic genes (*17*). A consensus enhancer sequence has not yet been defined; however, portions of the viral SV40 enhancer sequence are shared with the enhancers of other genes. An 11-bp sequence (residues 201 to 211) in the human and the mouse ANF genes is identical to the simian virus 40 (SV40) enhancer at 9 of 11 and at 8 of 11 bp, respectively. These sequences may play a role in causing the high level of atrial-specific transcription of the ANF gene. Assays for regulatory sequences should help define a role for these conserved 5′ flanking sequences.

The 5′ untranslated portions of ANF mRNA (residues 481 to 577) and those encoding the hydrophobic leader segment of preproANF are poorly conserved between mouse and human sequences (Fig. 2 and Table 1). While the role of 5′ untranslated portions of mRNA's is unclear, the hydrophobic leader segment of a nascent peptide is thought to be important for transport across the endoplasmic reticulum (*18*). Specific molecular interactions may not be required for this passage because human and mouse sequences are only 53 percent identical (Table 1) yet function

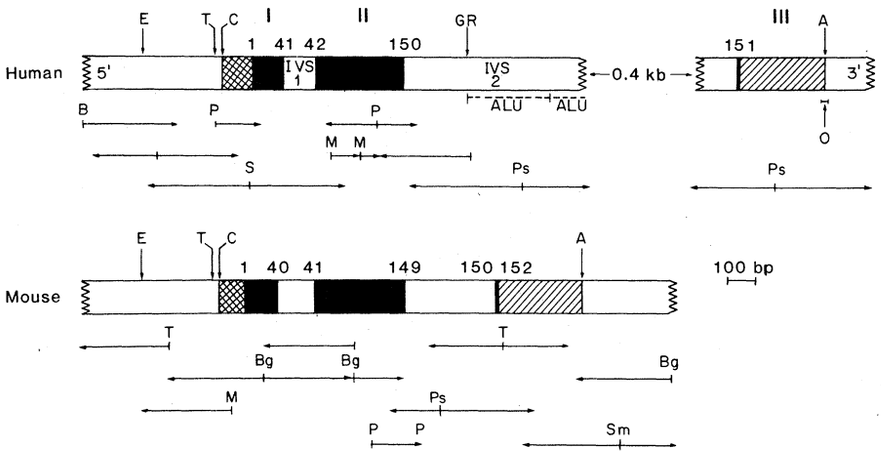

Fig. 1. Structure of the human and mouse ANF genes and strategy for determining their nucleotide sequences, showing the three coding blocks (I, II, and III) and two intervening sequences (IVS1 and IVS2). The segment of the ANF gene encoded in each coding block is indicated by the amino acid number. A putative enhancer (E), TATAA box (T), cap site (C), glucocorticoid receptor binding site (GR), and the polyadenylation addition sequence (A) defined by the nucleotide sequence are located on the gene map. The human gene contains two Alu repeat sequences (*22*). Fragments encoding the cloned ANF genes were digested with restriction enzymes [Bam HI (B), Sst I (S), Pvu II (P), Sau 3A (M), Pst I (Ps), Taq I (T), and Bgl II (Bg)], subcloned into appropriate M13 vectors (MP8, MP9, and MP11), and identified by hybridization to a rat ANF cDNA probe (*13*). The subclones encoding the 3′ end of the human gene were identified by the oligonucleotide (O) GAATAAACTTCAGCACCATGG derived from the human cDNA sequence (*10*). Three M13 subclones were derived after cleavage with Sau 3A, but lacked a Sau 3A site. Nucleotide sequences were determined by the dideoxy chain termination procedure.

Fig. 2. A comparison of the nucleotide sequences of the human and mouse atrial natriuretic factor genes. The nucleotide sequence of the human gene is drawn above the sequence of the mouse gene. The amino acid sequence of human preproANF is indicated. Only bases that differ between the mouse and human gene are shown for the mouse gene. A dot in any nucleotide position represents absence of a base from that species. Symbols indicate a potential enhancer sequence (dashed box), a TATAA sequence (jagged underline), and the probable site of initiation of mRNA transcription (circled). A potential glucocorticoid receptor binding site is boxed, and a polyadenylation signal is underlined.

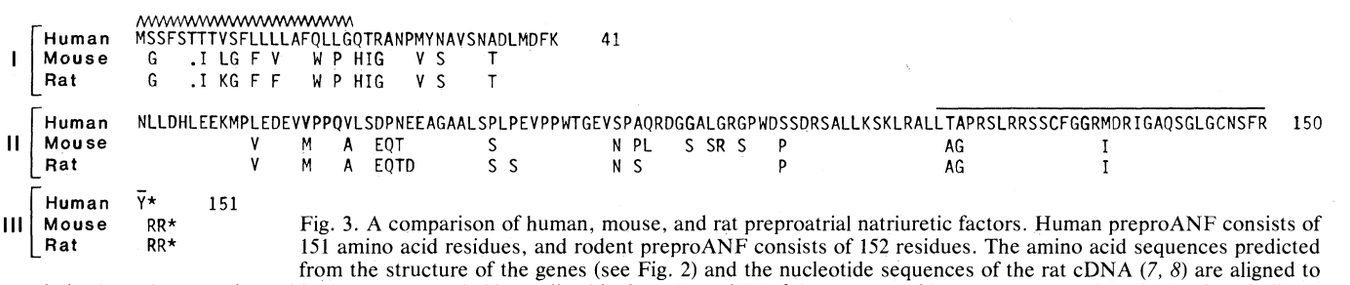

Fig. 3. A comparison of human, mouse, and rat preproatrial natriuretic factors. Human preproANF consists of 151 amino acid residues, and rodent preproANF consists of 152 residues. The amino acid sequences predicted from the structure of the genes (see Fig. 2) and the nucleotide sequences of the rat cDNA (*7*, *8*) are aligned to maximize homology. Amino acid sequences encoded by coding blocks I, II, and III of the mouse and human genes are shown. Numbers indicate the last amino acid encoded in each human coding block. The precise location of the intervening sequence in the rat is not known but probably occurs at the same relative location. The hydrophobic leader segment (residues 1 to 22) and ANF (residues 118 to 151) are defined by a jagged line and a straight line, respectively. ProANF consists of amino acid residues 22 to 151.

equally well. Maintaining hydrophobicity in the leader peptide appears to be the primary evolutionary constraint.

Sequences encoding both the mature ANF peptide and the proANF precursor are highly conserved, suggesting that these structures are critical to their functions. While a biological role of ANF may be regulation of salt and water homeostasis, a role for the proANF molecule is less well defined. Extensive sequence homology between human and mouse proANF implies a specific, necessary role. This region has been suggested to encode a separate hormone (cardiodilatin) (*11*); alternatively, these sequences may play a role in the processing and storage of ANF. Atrial granules are thought to be related to ANF production (*1–6*). Perhaps proANF sequences direct the biosynthesis of these granules.

A striking difference between human and mouse ANF is at the carboxyl terminus (coding block III, Fig. 3). The Arg-Arg sequence in rodent ANF is a putative peptidase cleavage signal similar to that in other prohormones. Some preparations of purified rat ANF lack this sequence (*6*), presumably as a result of proteolytic digestion. These data have two implications. First, the larger coding block in rodents suggests that ANF may have evolved from a longer peptide. Perhaps this ancestral proANF molecule, larger than either the rodent or human proANF, required proteolytic processing to form its carboxyl terminus. Subsequent evolution in humans may have introduced a stop codon at the end of the gene, removing the requirement for any proteolytic action. Second, persistence of the carboxyl tyrosine residue, despite rapid evolution of portions in the second intervening sequence (IVS2) and 3′ untranslated regions, suggests a critical role for this residue in ANF activity.

The IVS2 of the human gene is 800 bp longer than the corresponding IVS2 of the murine gene (Fig. 2). A dot matrix comparison (*19*) of these sequences suggests that there is no homology between the IVS2's of the two genes. Two tandem reiterated sequences present only in the human sequence are 80 percent identical to previously characterized Alu sequences.

A potential glucocorticoid receptor binding site is present in the IVS2 of the human gene (Fig. 2). The structure of the glucocorticoid receptor binding site is poorly understood; however, a consensus sequence has been proposed (*20*). Human ANF contains a sequence (residues 1291 to 1306) that is identical to the consensus glucocorticoid receptor binding sequence at 9 of 11 bp. The mouse gene lacks this glucocorticoid receptor binding site. An analogous situation is found in the human growth hormone gene, which contains a functional glucocorticoid receptor binding site that is absent in the murine gene (*21*). Whether or not the putative glucocorticoid receptor binding site can bind the glucocorticoid receptor remains to be determined. Further studies should identify the role of steroid hormones in regulation of ANF gene expression. Studies of the conserved nucleotide sequences should provide insight into the processes that regulate production of the gene and the mechanisms that control intravascular pressure and volume.

CHRISTINE E. SEIDMAN*
Cardiac Unit, Massachusetts General Hospital, Boston 02114
KENNETH D. BLOCH
KAREN A. KLEIN
Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115
JOHN A. SMITH
Department of Molecular Biology, Massachusetts General Hospital, and *Department of Pathology, Harvard Medical School*
J. G. SEIDMAN
Department of Genetics, Harvard Medical School

References and Notes

1. H. Sonnenberg *et al.*, *Can. J. Physiol. Pharmacol.* **60**, 1149 (1982).
2. M. G. Currie *et al.*, *Science* **221**, 71 (1983).
3. R. T. Grammer *et al.*, *Biochem. Biophys. Res. Commun.* **116**, 696 (1983).
4. N. C. Trippodo, A. A. MacPhee, F. E. Cole, *Hypertension* **5** (suppl. 1), I-81 (1983).
5. G. Thibault, R. Garcia, M. Canten, J. Genest, *ibid.*, p. I-75.
6. M. G. Currie *et al.*, *Science* **223**, 67 (1984).
7. C. E. Seidman *et al.*, *ibid.* **225**, 324 (1984).
8. M. Yamanaka *et al.*, *Nature (London)* **309**, 719 (1984).
9. M. Maki *et al.*, *ibid.*, p. 722.
10. S. Oikawa *et al.*, *ibid.*, p. 724.
11. K. Nakayama *et al.*, *ibid.* **310**, 699 (1984).
12. C. E. Seidman *et al.*, unpublished observations.
13. For a detailed description of these techniques see T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982).
14. K. McIntyre and J. G. Seidman, *Nature (London)* **308**, 551 (1984).
15. E. M. Southern, *J. Mol. Biol.* **8**, 503 (1975).
16. F. Sanger and S. Nicklen, A. R. Coulsen, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
17. G. Khoury and P. Gruss, *Cell* **33**, 313 (1983).
18. G. Blobel and B. Dobberstein, *J. Cell Biol.* **67**, 852 (1975).
19. P. A. Hieter, J. V. Maizel, P. Leder, *J. Biol. Chem.* **257**, 1516 (1982).
20. M. Karin *et al.*, *Nature (London)* **308**, 513 (1984).
21. D. D. Moore *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
22. C. M. Rubin *et al.*, *Nature (London)* **284**, 372 (1980).
23. We thank M. Stempowski for the preparation of the synthetic oligonucleotide and M. Swartz for providing computer facilities. Supported by NIH training grant HL-070208 (C.E.S.), a fellowship from the Leukemia Society of America (K.D.B.), a grant from Hoechst AG (West Germany) (J.A.S.), and a fellowship from the Mallinckrodt Foundation and NIH grant AI-18436 (J.G.S.).
* To whom correspondence should be addressed.

10 September 1984; accepted 24 October 1984

Synapsin I in Nerve Terminals: Selective Association with Small Synaptic Vesicles

Abstract. *Immunocytochemistry revealed that synapsin I is preferentially (and possibly exclusively) associated with small (40- to 60-nanometer) synaptic vesicles and not with large (greater than 60-nanometer) dense-core vesicles in bovine hypothalamus. These observations may explain why synapsin I is found exclusively in neurons, since small synaptic vesicles are specific to neurons whereas large dense-core vesicles in neurons may be considered the equivalent of secretory organelles in endocrine cells.*

Synapsin I is a major neuron-specific protein that is highly concentrated in nerve endings, where it is associated with the surface of synaptic vesicles (*1–4*). Synapsin I is phosphorylated at multiple sites both in vivo and in vitro by Ca^{2+}-calmodulin–dependent and adenosine 3′,5′-monophosphate–dependent protein kinases (*5*) and probably plays an important regulatory role in the function of synaptic vesicles (*3, 4*). Such a role is presumably related to some general aspect of synaptic vesicle function, since synapsin I is present in virtually all synaptic vesicles in most or all nerve endings (*2, 3*). On the other hand, the absence of synapsin I from nonneuronal cells and even from neuron-related cells, such as the chromaffin cells of the adrenal medulla (*6*), indicates that it is involved in some aspect of secretion specific to "bona fide" neurons.

Neurons can, by exocytosis, secrete at least two classes of substances: small nonpeptide molecules (classical neurotransmitters) and peptides (*7, 8*). Classical and peptide neurotransmitters are often released from the same neuron, even though their secretion seems to involve two distinct types of secretory organelles (*7–10*). It appears that typical synaptic vesicles, the small (40 to 60 nm in diameter) synaptic vesicles (SSV's), are storage sites only for classical neurotransmitters, the content of each vesicle representing a quantum of neurotrans-