Vol. 171, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 31, 1990 Pages 7-13

THE GENE FOR HUMAN ACIDIC FIBROBLAST GROWTH FACTOR ENCODES TWO UPSTREAM EXONS ALTERNATIVELY SPLICED TO THE FIRST CODING EXON

Gregg Crumley, Craig A. Dionne and Michael Jaye

Rorer Central Research
680 Allendale Road
King of Prussia, PA 19406

Received June 27, 1990

Summary: We have isolated two cDNA clones encoding human acidic fibroblast growth factor (aFGF) which represent the utilization of alternative upstream exons in aFGF mRNA. Isolation and sequence analysis of genomic clones spanning the first coding exon and each of the upstream sequences confirms that the divergent 5' sequences are separate exons, spliced alternatively to the first coding exon 34 nucleotides upstream of the initiator AUG codon. Restriction mapping of the genomic clones provides a minimum size estimate of 45 kilobase pairs for the aFGF locus.

Acidic fibroblast growth factor (aFGF; also known as heparin-binding growth factor 1) is one member of an expanding family of heparin-binding polypeptide growth factors including basic fibroblast growth factor (bFGF; HBGF 2) and the translated products of several recently identified genes (for review see ref. 1). Members of the family share significant amino acid sequence identity and an identical intron/exon structure within the coding regions of the genes (2-4). The FGF-related proteins include human keratinocyte growth factor (KGF) (5), the putative translation product of the murine int-2 oncogene (6), and the products of human oncogenes FGF-5 (7) and hstl/K-fgf (8,9) as well as the putative product of human FGF-6 (10). The FGF gene products promote new blood vessel growth (angiogenesis), stimulate mitogenesis in target cells of mesodermal and neuroectodermal origin (1) and induce mesoderm formation in the early amphibian embryo (12-14).

We previously reported the nucleotide sequence and deduced amino acid sequence of cDNA clones encoding human aFGF (15). Further characterization of human aFGF transcription by additional cDNA cloning and sequence analysis revealed the utilization of multiple polyadenylation sites, which explained in part the heterogeneity often observed on RNA transfer blots (16). Much of the major 4.4 kilobase (kb) mRNA is encoded within the 3' untranslated region and extends to a

0006-291X/90 $1.50

strong distal polyadenylation site, whereas less abundant RNA species are processed at less distal sites. The structure of aFGF mRNA at the 5' end, however, has not been completely elucidated and a definitive start site for transcription has not been identified. In this paper, we report evidence from cDNA cloning for the utilization of two alternative 5' exons in the synthesis and processing of human aFGF mRNA, each of which is spliced to the first coding exon 34 nucleotides (nt) upstream of the initiator AUG codon. The corresponding regions of human genomic DNA are shown to diverge from the cDNA sequences at authentic splice donor and acceptor sites.

MATERIALS AND METHODS

DNA Hybridization Probes: The cDNA insert from one of the original human aFGF clones, pDH15 (15), was utilized as probe (Probe 1) for the isolation of additional cDNA clones representing the 5' region. Probe 1 was also used to isolate a human genomic clone spanning the first coding exon. The cDNA probe was radiolabelled by nick-translation with [α-^{32}P]dATP and [α-^{32}P]dCTP as described (17).

A probe for detecting genomic sequences specific to Clone 7 (λGC7) upstream exon (Probe 2) was developed by synthesizing the following two oligonucleotides (oligos): a sense strand oligo with the sequence 5'CTTCCCCTGGGACAGCACTGAGCGAGTGTGGAGAGAGGTAC3' and an anti-sense oligo with the sequence 5'CTAAAGAGCTTGTAGGCCGAGGGCTGTACCTCTCTCCACACT3' The two oligos, which are complementary through the 17 nt at their 3' termini, were annealed by heating to 65°C for 5 min. and cooling for 1 hr. The annealed oligos (40 pmol) were filled in by incubation with 1 unit of E. coli DNA polymerase I (Klenow fragment) (17), creating a 66 base pair (bp) double-stranded DNA probe. For detection of genomic sequences specific to the Clone 36 (λMJ36) upstream exon a 30 nt oligo with the sequence 5'CCAACAGCCTTCGCTCCAGGGGAATCAGGG3' was synthesized (Probe 3). Probes 2 and 3 were radiolabelled using [γ-^{32}P]ATP and T4 polynucleotide kinase as described (17).

Library Screening: For the isolation of new cDNA clones, the radiolabelled cDNA probe (Probe 1) was hybridized to 1 x 10^6 plaques from a neonatal human brainstem library in λgt11 (18) on replica nitrocellulose filters and washed at high stringency under conditions previously reported (16).

Genomic clones were obtained by screening a library prepared by inserting Sau3A partial digests of human placental DNA (size-fractionated to ~ 20 kb) into BamHI-digested λEMBL3 arms (Promega Biotec). Approximately 1 x 10^6 recombinant phage were screened separately with Probes 1, 2 and 3 for detection of genomic clones spanning the 5' portion of the coding region, the λGC7-specific upstream sequence and the λMJ36-specific upstream sequence, respectively. Probes 1 and 2 were hybridized to genomic plaque lifts under the same high stringency conditions employed for cDNA screening. Probe 3 was hybridized to genomic DNA using conditions previously reported for oligo screening (16). Restriction maps of purified clones were compiled by a combination of single and double digestion with the restriction endonucleases BamHI, Bg1II, EcoRI, HindIII and SalI.

DNA Sequence Analysis: The cDNA inserts and cDNA-hybridizing restriction fragments from genomic clones were subcloned into pGEM1 (Promega Biotec) and M13 vectors and sequenced by dideoxy chain

termination (19) on either single-stranded templates or double-stranded templates denatured in alkali.

RESULTS AND DISCUSSION

Investigation of the 5' structure of aFGF mRNA was initiated by screening for human genomic clones spanning the 5' coding region and for cDNA clones extending further 5' than previous isolates (15,16). The original cDNA clone was used as probe (Probe 1) to screen both a neonatal human brainstem cDNA library in λgt11 (18) and a human genomic DNA library in the λEMBL3 vector (see MATERIALS AND METHODS). Restriction mapping of several new cDNA isolates indicated that two aFGF recombinants, designated λGC7 and λMJ36, span the entire coding region and extend into previously uncharacterized sequence at their 5' termini. Nucleotide sequence comparison of clones λGC7 and λMJ36 (Fig. 1) showed them to be identical through most of their length, except that they diverge at nt position -35 relative to the initiator ATG throughout the remainder of their 5' untranslated sequences. Clones λGC7 and λMJ36 have 100 bp and 358 bp, respectively, of 5' untranslated sequence. The comparisons shown in Fig. 1 are truncated within the coding region to emphasize the 5' divergence.

The initial genomic screen resulted in isolation of three non-overlapping genomic clones hybridizing to aFGF cDNA. Restriction analysis of the clones and genomic Southern blots (data not shown) confirmed the overall gene structure recently reported (3,4) showing

```
λMJ36  -358   atccccaa ggctaggagg ccaacctact aacaggtggg tgggtatggt gtgtggtttc
λMJ36  -300 actcagttct tctcatgggg tttctctgag ctccattcat accagaaagg gagcaggaga
λMJ36  -240 gagaggacaa gtggatccaa cagccttcgc tccaggggaa tcagggcatc gcctcctttt
λMJ36  -180 ctgggaggac actcccttct gatggtgaat gggaactccc ttcctcctgc agcagcctgc
λMJ36  -120 ctgcagctgt cctggtagaa cagtgtggac attgcagaag ctgtcactgc cccagaaaga
λGC7   -100                    cttcccctgg gacagcactg agcgagtgtg gagagaggta
                                          ↓
λMJ36   -60 aagcacccca gagccaaggc aaagagtctt gaaagcgcca caagcagcag ctgctgagcc
λGC7    -60 cagccctcgg cctacaagct ctttagtctt gaaagcgcca caagcagcag ctgctgagcc

λMJ36    +1 ATGGCTGAAG GGGAAATCAC CACCTTCACA GCCCTGACCG AGAAGTTTAA TCTGCCTCCA
λGC7     +1 ATGGCTGAAG GGGAAATCAC CACCTTCACA GCCCTGACCG AGAAGTTTAA TCTGCCTCCA

λMJ36   +61 GGGAATTACA AGAAGCCCAA ACTCCTCTAC TGTAGCAACG GGGGCCACTT CCTGAGGATCC
λGC7    +61 GGGAATTACA AGAAGCCCAA ACTCCTCTAC TGTAGCAACG GGGGCCACTT CCTGAGGATCC
```

Fig. 1: The Nucleotide Sequence of cDNA Clones λGC7 and λMJ36. The cDNA clones λGC7 and λMJ36 both agree with the previously reported sequence throughout the coding region; hence, the sequences shown are arbitrarily truncated in the coding region at a BamHI site (overlined). The two sequences are aligned to emphasize the divergence upstream of position -34. Numbers given are relative to the initiator ATG. The initiator ATG and in-frame upstream TGA codons are boxed, and upstream ATG codons in λMJ36 are underlined.

that the coding region of human aFGF is, like other FGF-related genes, interrupted by two introns. One of the genomic clones, designated λGC1, spans the most 5' of the three coding region exons and contains 10 kb of 5' flanking DNA as well as 2.8 kb of 3' flanking intron sequence (Fig. 2). In a preliminary characterization of this clone (2), we noted the presence of a potential splice acceptor site at nt position -35 upstream of the initiator ATG codon. Other groups have made the same observation (3,4), but the presence of upstream exon sequences has not previously been reported.

We hypothesized that the divergent 5' sequences observed in λGC7 and λMJ36 represent two alternative upstream exons because the point at which they diverge is exactly the same site where each diverges from λGC1 genomic sequence, the site already noted as a potential splice acceptor (2-4). To test the hypothesis we sought to isolate human genomic clones corresponding to each of the new 5' sequences. An initial hybridization of Probes 2 and 3 to transfer blots of λGC1 DNA indicated that neither of the upstream sequences lies within the 10 kb of 5' flanking DNA contained in λGC1. Hence, we screened the EMBL3 genomic library with Probes 2 and 3 to isolate clones containing the unique portions of λGC7 and λMJ36, respectively.

A single 14 kb genomic clone, λGC20, was obtained in the λGC7-specific screen and a 12.5 kb clone, λGC6, was isolated with the λMJ36 probe. Restriction enzyme analysis and DNA transfer blot hybridization of the new clones allowed us to establish a general physical map of the upstream region (Fig. 2). Since neither clone

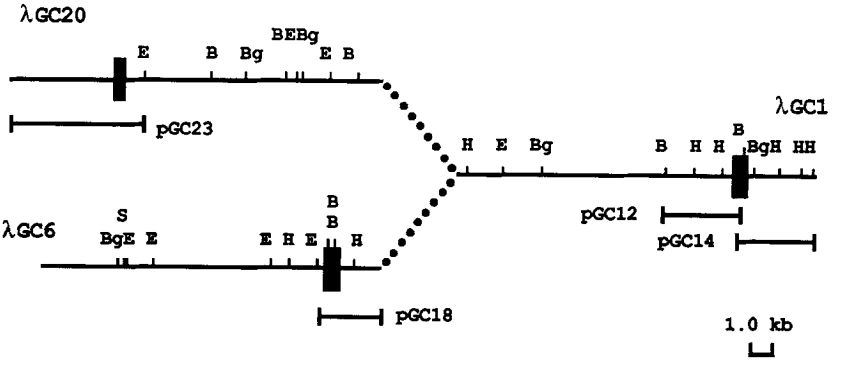

Fig. 2: Restriction Maps of the Genomic Clones λGC1, λGC20 and λGC6. The genomic clones were mapped by single and double-digestion with BamHI (B), BglII (Bg), EcoRI (E), HindIII (H) and SalI (S). λGC20 and λGC6 contain fragments hybridizing to unique upstream sequences in cDNA clones λGC7 and λMJ36, respectively, and λGC1 spans the 5' end of the coding region. The upstream clones are each shown in their general relationship to the coding region clone. The relative positions of upstream clones and the absolute distance has not been determined. cDNA-hybridizing fragments that were subcloned and sequenced are shown below the phage DNA maps. The scale in kb is shown.

overlapped with the other or with λGC1 (coding region clone), we were not able to assess the relative position of upstream clones or their total distance from the coding region. Restriction mapping of genomic clones places the λGC7-specific exon (in λGC20) at least 20 kb upstream of the first aFGF coding exon (Fig. 2). Likewise, the λMJ36-specific exon (in λGC6) is estimated to be at least 12 kb upstream of the first coding exon. Neither of these estimates, however, takes into consideration the relative placement of these two exons, i.e. that one must be upstream of the other. Placing λGC6, hypothetically, upstream of λGC20 provides a minimum size estimate of the human aFGF locus of 45 kb, including the estimated 19 kb already reported (3,4) for the region spanning the coding exons.

In order to establish that the new, alternative cDNA sequences can actually be derived from transcription of the genomic sequences

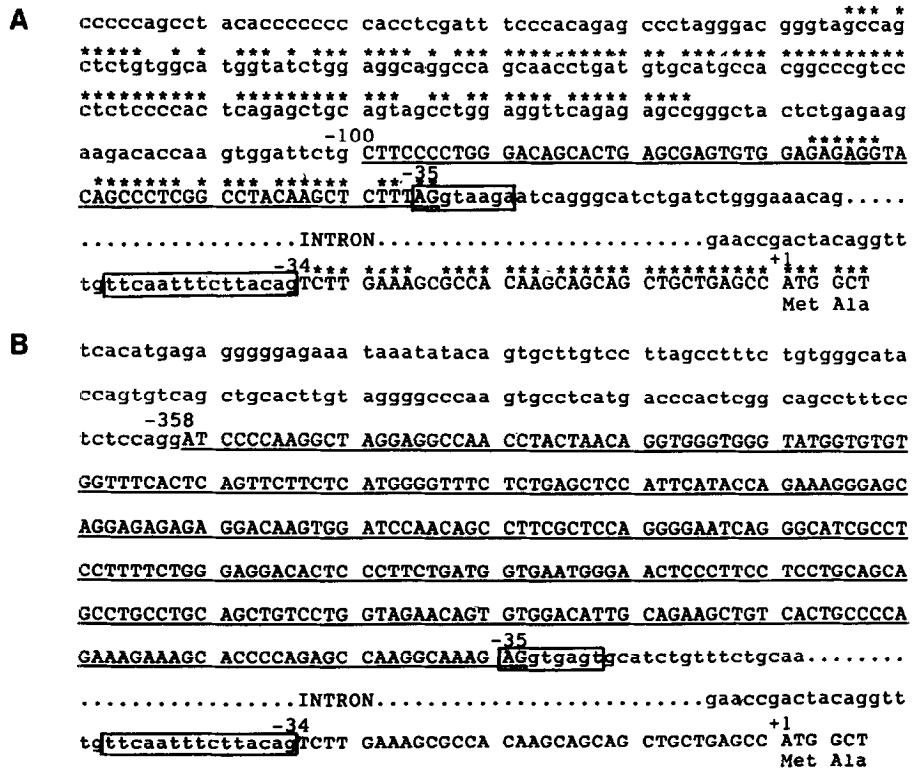

Fig. 3: Genomic Sequences of Clones λGC20 and λGC6. The sequences of λGC20 and λGC6 spanning the Clone 7 (Panel A) and Clone 36 (Panel B) exons, respectively, are shown in relationship to coding exon 1 and the boundaries of the intron between. Nucleotides found in both cDNA and genomic clones are shown in upper case letters. The upstream exon sequences are underlined. Numbers refer to the cDNA nt sequence relative to the initiator ATG codon. The putative splice donor and acceptor sequences are boxed. Positions of identity between a bovine aFGF cDNA clone (20) and the clone 7-specific genomic clone are marked with asterisks.

isolated, we subcloned restriction fragments from the genomic clones which hybridized to exon-specific probes. The subcloned fragments shown in Fig. 2 were sequenced. The nt sequence spanning exon regions of λGC20, λGC6 and λGC1 are shown in Fig. 3 in the context of the relationships revealed by the cDNA clones. In Fig. 3, one observes a canonical splice donor site (20) on the 3' side of each alternative upstream exon and a splice acceptor site on the 5' side of the first coding exon. These splice sites are located at precisely the correct positions to generate mRNA species corresponding to cDNA clones 7 and 36. Finally, a bovine aFGF cDNA reported recently (20) exhibits substantial sequence similarity to the λGC7 cDNA clone on both sides of the upstream splice junction and also in the 5' flanking λGC20 genomic sequence (see Fig. 3). Clearly, transcription of the human aFGF locus can initiate well upstream of the coding region.

The presence of 5' untranslated exons located far from the coding region raises anew the question of how aFGF transcription is initiated and controlled. There are now numerous examples in the literature of genes encoding one or more upstream exon. For example, a number of oncogenes encode alternative upstream exons that appear to regulate negatively the translational efficiency of the mRNA, wherein truncation of the leader exon can lead to cell transformation (for review see ref. 22). In this light, it is worth noting that the Clone 36 exon has four upstream ATG codons that could impede translation. In human and rat IGF-II mRNA, the use of alternative 5' exons appears to reflect tissue-specific transcription from two different promoters (23,24). The function of 5' exons in aFGF mRNA cannot be evaluated at present but future experiments will attempt to distinguish whether the upstream exons reflect a special role in aFGF regulation, such as tissue-specific use of alternative promoters or translational modulation. We have examined the genomic sequence for 1 to 2 kb upstream of each 5' exon, including coding exon 1, for promoter sequences (data not shown). No set of canonical promoter elements could be identified at any of these sites, but functional characterization of the three available genomic regions may ultimately reveal novel promoter elements. Alternatively, aFGF transcription may initiate at yet unrecognized sites in human genomic DNA.

ACKNOWLEDGMENTS

We would like to thank Dr. G. Ricca and Ms. Carol Herron for the genomic library and Ms. Robin McCormick for excellent typing of the manuscript.

REFERENCES

1. Burgess, W.H. and Maciag, T. (1989) Annu. Rev. Biochem. **58**, 575-606.
2. Crumley, G., Schlessinger, J. and Jaye, M. (1988) In NATO ASI Series C (M. Alexis and C. Sekeris, Eds.), Vol. 295, pp. 41-48. Kluwer Academic Publishers, Dorbrecht, The Netherlands.
3. Wang, W.P., Lehtoma, K., Varban, M.L., Krishnan, I. and Chiu, I.M. (1989) Mol. Cell. Biol. **9**, 2387-2395.
4. Mergia, A., Tischler, E., Graves, D., Tumolo, A., Miller, J., Gospodarowicz, D., Abraham, J.A., Shipley, G.D. and Fiddes, J.C. (1989) Biochem. Biophys. Res. Commun. **164**, 1121-1129.
5. Finch, P.W., Rubin J.S., Miki, T. Ron, D. and Aaronson, S.A. (1989) Science **245**, 752-755.
6. Dickson, C. and Peters, G. (1987) Nature **326**, 833.
7. Zhan, X., Bates, B., Hu, X. and Goldfarb, M. (1988) Mol. Cell. Biol. **8**, 3487-3495.
8. Taira, M., Yoshida, T., Miyagawa, K., Sakamoto, H., Terada, M. and Suzimura, T. (1987) Proc. Natl. Acad. Sci. USA **84**, 2980-2984.
9. Delli Bovi, P., Curatola, A.M., Kern, F.G. Greco, A., Ittman, M. and Basilico, C. (1987) Cell **50**, 729-737.
10. Marics I., Adelaide, J., Raybaud, F., Mattei, M.G., Coulier, F., Planche, J., de Lapeyriere, O., and Birnbaum, D. (1989) Oncogene **4**, 335-340.
11. Folkman, J. and Klagsbrun, M. (1987) Science **235**, 442-447.
12. Slack, J.M.W., Darlington B.G., Heath, J.K. and Godsave, S.F. (1987) Nature **326**, 197-200.
13. Kimelman, D., Abraham, J.A., Hosparanta, T., Palisi, T.M. and Kirschner, M.W. (1988) Science **242**, 1053-1058.
14. Paterno, G.D., Gillespie L.L., Dixon, M.S., Slack, J.M.W. and Heath, J.K. (1989) Development **106**, 79-83.
15. Jaye, M., Howk, R., Burgess, W., Ricca, G.A., Chiu, I.M., Ravera, M.W., O'Brien, S.J., Modi, W.S., Maciag, T. and Drohan, W.N. (1986) Science **233**, 541-544.
16. Crumley, G.R., Howk, R., Ravera, M.W. and Jaye, M. (1989) Gene **85**, 489-497.
17. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 112-123.
18. Kamholz, J., de Ferra, F., Puckett, C. and Lazzarini, R. (1986) Proc. Natl. Acad. Sci. USA **83**, 4962-4966.
19. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
20. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. **50**, 349-383.
21. Halley, C., Courtois, Y. and Laurent, M. (1988) Nucl. Acids Res. **16**, 10913.
22. Kozak, M. (1988) J. Cell. Biol. **107**, 1-7.
23. de Pagter-Holthuizen, P., Jansen, M. van Schaik, F.M., van der Kammen, R., Oosterwijk, C., van der Brande, J.L. and Sussenbach, J.S. (1987) FEBS Lett. **214**. 259-264.
24. Soares, M.B., Turken, A., Ishii, D., Mills, L. Episkopou, V., Cotter, S., Zeitlin, S. and Efstradiatis, A. (1986) J. Mol. Biol. **192**, 737-752.