Vol. 188, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
October 15, 1992 Pages 57-63

STRUCTURE AND CHROMOSOMAL LOCATION OF THE HUMAN GRANULIN GENE

Vijay Bhandari and Andrew Bateman

Endocrine Laboratory, Royal Victoria Hospital/McGill University, Department of Medicine, Montreal, Quebec H3A 1A1, Canada

Received August 19, 1992

SUMMARY. Granulins are a family of cysteine rich polypeptides some of which have growth modulatory activity. We showed previously that the granulins are encoded within the same precursor consisting of seven granulin domains arranged in tandem. Here we report the chromosomal location and structural organization of the protein coding region of the granulin gene. The granulin gene was assigned to chromosome 17 using DNA from human-hamster somatic cell hybrids. The protein-coding region of the granulin gene was shown to comprise 12 exons covering about 3700 bp. Each tandem granulin repeat is encoded by two non-equivalent exons, a configuration unique to the granulins that would permit the formation of hybrid granulin-like proteins by alternate splicing.

Elicted human inflammatory cells (1), rat bone marrow (1), rat kidney preparations (2), and teleost spleen (3) contain novel growth regulatory peptides that have been called granulins (1) or epithelins (2). Myelogenous leukemic cell lines such as HL60 and U937 express abundant granulin mRNA (4), as do all epithelial cell lines we have examined, and most rat tissues except skeletal muscle (unpublished). Expression in fibroblastic cell lines such as NIH 3T3 is low or absent. The wide spread occurence of granulin mRNA in cells from the hemic system, and in epithelia implies important functions in hematopoietic and epithelial cell function.

Granulins have a very characteristic peptide structure. A highly conserved 12-cysteine backbone defines a concensus sequence; $CX_{5-6}CX_5CCX_8CCX_6CCXDXXHCCPX_4CX_{5-6}C$, which is repeated seven times in the granulin precursor (4). All four reported human granulin-like peptides are encoded in this precursor (4). Southern analysis shows that the granulin gene exists as a single copy in the human genome (4). Here we present the chromosomal location of the granulin gene and the genomic organisation of the protein-coding region of the granulin gene.

0006-291X/92 $4.00

MATERIALS AND METHODS

Chromosomal mapping.
Granulin-specific primers [forward primer,5'GGAAGTATGGCTGCTGC 3', corresponding to nucleotides 681 to 697 of the cDNA (4); reverse primer 5'GGATCAGGTCACACACA 3', nucleotides 753 to 769 (primer pair A) and foward primer 5'CTGTGTGTGACCTGATC 3', nucleotides 752 to 768; reverse primer 5'CTGGATGTGGTT(T/C)TC(A/G)CA(G/A)CA 3', nucleotides 940 to 959 (primer pair B)] were used in the polymerase chain reaction (PCR) with DNA from human-hamster somatic cell hybrids (BIOSMAP , BIOS Corp.,CT) as template. The reactions were subjected to 30 cycles of denaturation at 94°C for 1 min, annealing at 60°C for 1 min, and elongation at 72°C for 1 min. The amplified products were analysed by Southern blotting using granulin-specific probes (a 21-mer 5'TGCTCCGATCACCTGCACTGC 3' corresponding to nucleotides 718 to 738, and a 1890-bp EcoRI/SacI fragment (4) for PCR products from primer pairs A and B respectively).
Gene isolation and sequencing.
A human genomic library (Clontech) was probed with a 182 bp PCR derived genomic fragment (4) corresponding to part of granulin A. Conditions of hybridization and washes were as previously described (4). One positively hybridizing clone, HGL1, was obtained. The insert was removed with BamHI, SalI and XhoI, and the resulting fragments subcloned into Bluescript KSII-. Subclones containing the granulin gene were identified by Southern hybridzation with a full length granulin cDNA. Nucleotide sequence was obtained by double stranded dideoxy sequencing using T3, T7 or custom-synthesized 17-mer primers (Sheldon Biotechnology Centre, McGill University) with Sequenase and in some cases by automated sequencing (Nucleic Acids Facility, Columbia University). HGL1 did not contain the 3'-end of the protein coding region of the granulin gene. The human genomic library was rescreened with a 394 bp fragment corresponding to nucleotides 1706 to 2099 of the 3'-end of the granulin precursor cDNA. One positive clone, HGL23, was identified. The nucleotide sequence of the 3'-end of the granulin gene was then obtained in the same manner as described above.

RESULTS AND DISCUSSION

Southern blotting of the PCR-amplification products from a panel of hamster-human chromosome hybrid DNAs using two independent PCR-primer pairs revealed a single product of predicted size with human genomic DNA and with DNA from the somatic cell hybrid 811, which contains human chromosomes 8, 17 and 18 (Figure 1, Table 1). No PCR product was obtained with hamster genomic DNA or with the control (no DNA added). Localization of the granulin gene to chromosomes 8 or 18 can be excluded because of the absence of hybridization in hybrids 803, 909, 967 and 1006 which contain chromosome 8, or hybrids 324, 734, 750 and 867 which contain chromosome 18 (Table 1). Therefore the granulin precursor gene must be on chromosome 17.

The protein coding region of the granulin gene spans about 3700 bp, with 11 introns (Figure 2, Table 2). Postulated splice sites were determined by comparison with the published cDNA sequence (4) and conform to splice site concensus sequences (5,6). The introns are small, varying from 79 to 479 bp, with an average of 168 bp. Each 12-

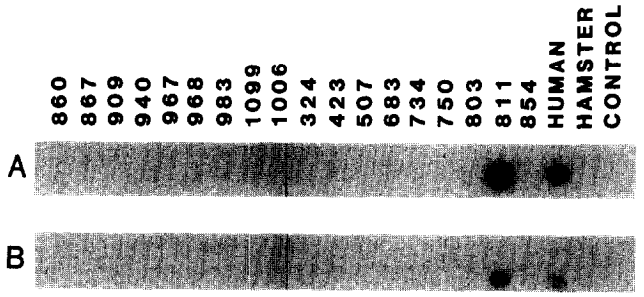

Figure 1. Southern Hybridization of PCR products. Conditions for the PCR are described in Materials and Methods. The amplified products were electrophoresed on a 1.5% agarose gel, transferred to nylon membranes (Zeta-Probe, Bio-Rad) by capillary blotting and hybridized with granulin-specific probes. Conditions for the hybridizations and washes were essentially as described by the manufacturer (Bio-Rad). (A) PCR with primer pair A and probed with a 21-mer oligonucleotide. (B) PCR with primer pair B and probed with a 1890-bp EcoRI/SacI fragment (4).

cysteine granulin domain is encoded by two exons. Phase 0 exonic boundaries, where the splice site occurs between two codons (7), bisect the translated granulin protein domains between the sixth and seventh cysteine. The position of these junctions is rigidly conserved in all seven granulin domains. The exonic boundaries in the intervening regions between granulin domains are all phase 1 (i.e. they occur after the first nucleotide of the codon), and their position relative to the granulin domains is variable.

Table 1. Segregation of Human Chromosomes in Human-Hamster Somatic Cell Hybrids

Cell Line	Human Chromosome Content																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y
324	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
423	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
507	-	-	+	-	+	-	-	-	-	-	-	+	-	+	-	-	-	-	-	40%	-	25%	-	+
683	-	-	-	-	+	-	-	-	-	-	+	45%	-	+	-	-	-	-	+	-	+	+	-	-
734	-	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
750	-	-	-	-	D	-	-	-	-	-	-	+	+	+	-	-	-	+	-	-	-	-	-	-
803	-	-	-	+	+	-	-	+	-	-	15%	-	-	-	-	-	-	-	-	-	-	+	+	-
811	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-
854	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
860	-	-	15%	-	+	+	-	-	-	15%	-	-	-	-	-	-	-	-	45%	-	+	-	-	-
867	60%	-	-	-	+	-	-	-	-	-	-	-	+	+	-	-	-	+	+	-	-	-	-	-
909	-	-	-	-	D	+	-	+	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	-
940	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-
967	-	-	-	-	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-
968	-	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-
983	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1006	-	-	-	55%	+	-	+	+	-	-	-	-	+	-	+	-	-	-	+	-	+	-	-	-
1099	+	-	-	-	D	-	-	-	-	-	-	-	+	-	-	-	-	-	+	-	+	+	-	-
Hamster	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Human	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

D - Deletion at p15.1 - 15.2
Percentage numbers are the percent of the cell population containing the noted chromosome.

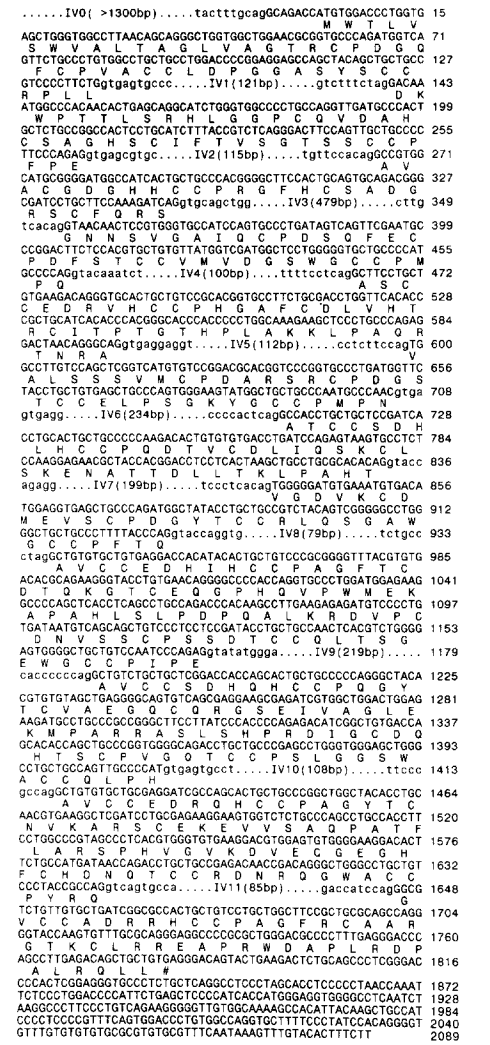

Figure 2. Nucleotide sequence of the human granulin gene. The nucleotide sequence is a composite of two overlapping clones, HGL1 and HGL23. HGL1 and HGL23 overlap by about 1400 bp. Sequences of exons, intron/exon boundaries and 3'-untranslated region are shown. The amino acid sequence is shown below the nucleotide sequence. The stop codon is indicated by #. The lengths of the introns are given in parentheses. IV refers to intervening sequence. The nucleotides are numbered according to the cDNA (4). Nucleotides 104, 126, 133, 500, 512, 644, 689, 1033 and 1567 were assigned based on the cDNA sequence. Several nucleotide differences were observed between the genomic sequence and the published cDNA sequence. These differences are due to reading and typographical errors in the cDNA sequence. The corrected nucleotides and their corresponding amino acids are given in the figure above. The errors were noted at nucleotides 1219 and 1221 (CGA to GGC; R to G), 1299 and 1301 (CGG to GGC; G to A), 1379 (A to T; Q to L), 1639 and 1640 (GC to CG; A to R), 1699 and 1700 (CG to GC; R to A), and in the 3'-untranslated region, nucleotide 1957 (A to T). Nucleotides A, G, C, C and A were omitted at positions 1968, 2051, 2054, 2056, 2063 and 2080, respectively.

Three types of exons, which we call α, ß and ßα can be recognised based on the peptide sequences that they encode and their phase class (Figures 2, 3). α-exons, encode a six cysteinyl peptide sequence with the cysteine motif $CX_{5-6}CX_5CCX_8CCX_4$,

Table 2. Size and splice junction sequences of exons and introns in the protein coding region of the human granulin gene. Exon sequences are given in uppercase and intron sequences in lowercase. Amino acids with interrupted codons were assigned to the exon containing two of the three codon nucleotides.

Exon	Amino Acids	5'donor (C,A)AG/gt(a,g)agt[a]	Intron	Length (bp)	3'acceptor (c,t)ag/G[a]	Type[b]
I	46	CTG/gtgagt	1	121	tag/G	0
II	42	GAG/gtgagc	2	115	cag/G	0
III	28	CAG/gtgcag	3	479	cag/G	I
IV	38	CAG/gtacaa	4	100	cag/G	0
V	45	CAG/gtgagg	5	112	cag/T	I
VI	37	AAC/gtgagt	6	199	cag/G	0
VII	42	CAG/gtacca	7	200	cag/T	I
VIII	33	CAG/gtacca	8	79	tag/G	0
IX	83	GAG/gtatat	9	219	cag/G	0
X	78	CAT/gtgagt	10	108	cag/G	0
XI	77	CAG/gtcagt	11	85	cag/G	0
XII	45					

[a] Consensus sequence from Mount (6).
[b] Intron type is according to Sharp (7). Type 0 indicates introns lying between two codons and type I indicates introns lying between the first and second nucleotides of a codon.

corresponding to the amino-terminal half of the granulins, and have phase 1 and 0 5' and 3' exonic boundaries respectively. ß-exons encode a six cysteinyl peptide sequence with a $X_2CCX_5CCX_5CX_{5-6}C$ array corresponding to the carboxyl terminus of a granulin domain, with phase 0 and 1 5' and 3' boundaries respectively. ßα exons encode a 12 cysteine peptide, $X_2CCX_5CCX_5CX_{5-6}CX_{25-30}CX_{5-6}CX_5CCX_8CCX_4$ comprising the carboxyl-terminal half of a granulin, an intervening spacer sequence, and then the amino-terminal half of a granulin domain, with phase 0 boundaries at both 5' and 3'

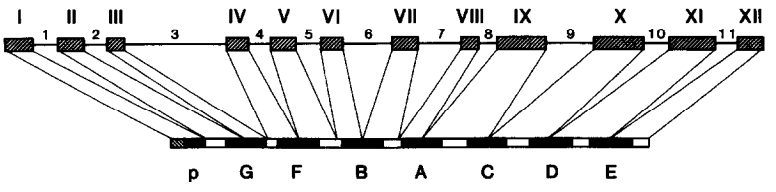

Figure 3. Structural organization of the human granulin gene. Exons are indicated by hatched boxes and introns by lines. Exons I, II, IV, VI and VIII are α-exons, exons III, V, VII and XII are ß-exons and exons IX, X and XI are ßα-exons (see text). The organization of the granulin precursor was determined from its cDNA (4). Closed boxes represent the granulin domains, open boxes represent the intervening spacer regions and the hatched box represents the signal sequence.

ends. The 5' end of the protein coding region of the granulin precursor consists of alternating α and ß exons. The pattern changes toward the 3' end of the gene where the ßα exons appear. The granulin gene therefore is constructed of three types of exons arranged in the following manner: α-α-ß-α-ß-α-ß-α-ßα-ßα-ßα-ß, where α encodes the amino terminus and ß the carboxy terminus of a granulin repeat.

In proteins constructed as tandem repeats, exon boundaries and protein domains are often congruent. Many examples could be cited, including the epidermal growth factor precursor and related proteins (8,9,10), and adhesion molecules of the selectin class (11,12,13). Thus the construction of the granulin repeat from two exons appears to go against a well established trend. The unusual exonic organisation of the granulin gene may allow the formation of hybrid granulin domains by alternate splicing. Single exon deletion of the ßα exons would generate the hypothetical granulin hybrids AC, CD, and DE without causing frame shifts in the nucleotide sequence. Other hybrids could be generated by the deletion of two or more exons provided identical phase junctions are brought together. Hence alternate splicing of the granulin gene may generate hybrid granulin-like proteins with potentially new biological properties. We have reported evidence for two granulin transcripts in some cell lines differing by about 200 bp which may be due to alternate splicing (4).

The exonic organisation of a gene is intimately related to its evolutionary history (14). The pattern observed in the granulin gene may have arisen in several ways. The ancestral exon may have been of the ßα class which was then tandemly duplicated followed by the insertion of introns (15) to create the α and the ß exons. Alternatively, the α and ß exons may have arisen as separate units shuffled to form a primordial granulin, and then duplicated as an [exon-intron-exon] unit, the ßα exons arising from a deletion of an intervening intron (16) bringing together the phase 1 junctions of an α3' and a ß5' exon. Another possibility is that a single primordial exon duplicated and the two copies then evolved to form the α and ß exons encoding a granulin-like protein. Multiplication of this [exon-intron-exon] unit would generate a protein resembling the present granulin precursor. It will be interesting to determine whether granulin related exons occur in other proteins, and if so, whether they occur as α, ß, ßα, [α-intron-ß] pairs, or as combinations of these exonic units.

In conclusion, the seven tandem repeats that make up the granulin precursor mRNA are constructed from two exons. The construction of a tandem repeat from two regularly alternating exons is very unusual and has implications for the generation of novel granulins by alternate splicing, and in understanding the evolutionary origins of the 12 cysteine granulin repeat.

ACKNOWLEDGMENTS. We acknowledge valuable conversations with Dr. R.G.E. Palfree and Dr. H.P.J. Bennett and the support of Dr. S. Solomon, Director of the Endocrine Laboratory, Royal Victoria Hospital. The work was supported by a grant from the Medical Research Council of Canada (MRC) to A.B., who is the recipient of a Scholarship from the MRC. V.B. was supported by studentships from the Department of Medicine and Research Institute at the Royal Victoria Hospital and is the recipient of a studentship from the MRC.

REFERENCES

1. Bateman,A., Belcourt,D., Bennett,H.P.J., Lazure,C., and Solomon,S. (1990) Biochem. Biophys. Res. Commun. 173, 1161-1168
2. Shoyab,M., McDonald,V.L., Byles,C., Todaro,G., and Plowman,G.D. (1990) Proc. Natl. Acad. Sci. USA 87, 7912-7916
3. Belcourt,D., Okawara,Y., Fryer,J., and Bennett,H.P.J. (1991) 34th Annual Meeting of the Canadian Fed. Biol. Sci., Kingston, Ont. Abstract No. 343
4. Bhandari,V., Palfree,R.G.E., and Bateman,A. (1992) Proc. Natl. Acad. Sci. USA 89, 1715-1719
5. Breathnach,R., and Chambon,P. (1981) Annu. Rev. Biochem. 50, 349-383
6. Mount,S.M. (1982) Nucleic Acids Res. 10, 459-472
7. Sharp,P.A. (1981) Cell 23, 643-646
8. Sudhof,T.C., Russel,D.W., Goldstein,J.C., Brown,M.S., Sanchez-Pescador,R., and Bell,G.I. (1985) Science 228, 893-895
9. Delgadillo-Reynoso,M.G., Rollo,D.R., Hursh.,D.A., and Raff,R.A. (1989) J. Mol. Evol. 29, 314-327
10. Shimasaki,S., Koga,M., Esch,F., Cooksey,K., Mercado,M., Koba,A., Ueno,N., Ying, S-Y., Ling,N., and Guillemin,R. (1988) Proc. Natl. Acad. Sci USA 85, 4218-4222
11. Collins,T., Williams,A., Johnston,G.I., Kim,J., Eddy,R., Shows,T., Gimbrone,M.A., and Bevilacqua,M.P. (1991) J. Biol. Chem. 266, 2466-2473
12. Johnston,G.I., Bliss,G.A., Newman,P.J., and McEver,R.P. (1990) J. Biol. Chem. 265, 21381-21385
13. Ord,D.C., Ernst,T.J., Zhou,L-J., Rambaldi,A., Spertini,O., Griffin,J., and Tedder,T.F. (1990) J. Biol. Chem. 265, 7760-7767
14. Patthy,L. (1987) FEBS Lett. 214, 1-7
15. Rogers,J. (1985) Nature 315, 458-459
16. Blake,C. (1983) Nature 306, 535-537