Volume 16 Number 20 1988 Nucleic Acids Research

A highly conserved, 5′ untranslated, inverted repeat sequence is ineffective in translational control of the α1(I) collagen gene

Paul Bornstein*, John McKay, Sreelekha Devarayalu and Susan C.Cook

Departments of Biochemistry and Medicine, University of Washington, Seattle, WA 98195, USA

Received June 2, 1988; Revised and Accepted September 19, 1988

ABSTRACT

An inverted repeat sequence, extending from the 5′ untranslated region of the first exon through the translation initiation codon, is highly conserved in the α1(I), α2(I) and α1(III) collagen genes of mammals and birds. It has been suggested that this sequence functions in translational control of collagen gene expression. When the upstream axis of the dyad of symmetry was deleted, the efficiency of translation of transcripts from a human α1(I) collagen-bovine growth hormone fusion gene was unchanged in either transiently or stably transfected cells. Furthermore, mRNA levels were not affected when the same deletion was transferred to a collagen-human growth hormone fusion gene in which the collagen sequence retained the first intron. Examination of human α1(I) DNA, extending from the start of transcription to the start of translation, by the DNAse I protection procedure revealed evidence for protein binding to a sequence just upstream of the inverted repeat sequence but not to the inverted repeat itself. Our studies therefore indicate that this highly conserved DNA sequence does not function generally in translational or transcriptional control of type I procollagen synthesis.

INTRODUCTION

The three collagen genes that code for types I and III collagens are characterized by a highly conserved nucleotide sequence in the untranslated region of the first exon (1-7). Thus, in the 120 nucleotides that separate the transcription and translation start sites in the human α1(I) gene only 9 substitutions, deletions or insertions occur when this sequence is compared with the corresponding mouse α1(I) sequence (3, 5). Comparison between the mouse and chicken α2(I) genes in this region also shows a high degree of conservation of sequence (2). A constant feature in these genes is the presence of a conserved inverted repeat sequence which contains the open reading frame ATG in its 3′ axis. A second ATG is consistently found 14 nucleotides upstream in the 5′ portion of the inverted repeat and a third ATG is placed 40-60 bases upstream of the second. Both upstream ATG codons are followed by in-frame stop codons creating minicistrons.

This high degree of conservation of sequence in a noncoding region of three related genes, and in three species, has led to suggestions that the 5′ untranslated portion of the first exon of interstitial collagen genes may function in regulation of gene expression. In particular, Yamada et al. (1) have proposed that the conserved palindrome could form a stem-loop structure in mRNA and cite, in support of this suggestion, that conservation of sequence among

genes and species is greatest in bases placed in the stem of the putative stem-loop structure and less in the loop. Various functions have been proposed for such a hairpin mRNA structure, including modulation of translational efficiency (1-3, 8). Translational control could result as a consequence of secondary structure in the mRNA (9, 10) or may be mediated by RNA-binding proteins. Efficient translation of collagen mRNAs may be particularly sensitive to such modulating influences since the open reading frame AUG initiator codon does not exist in an optimal sequence context, as defined by the work of Kozak (11), in any of the collagen genes that have been examined. Additional, indirect support for a regulatory function for the conserved inverted repeat sequence in the α1(I), α2(I) and α1(III) genes comes from the finding that the related cartilage collagen gene, α1(II), which is undoubtedly subject to different tissue-specific control, lacks such a sequence (12, 13).

We have examined the function of the conserved inverted repeat sequence by creating a deletion in the 5′ untranslated region of the human α1(I) gene that effectively destroys the dyad of symmetry but preserves the open reading frame ATG in this gene. No evidence for a role of the conserved sequence in either translational or transcriptional control was obtained when collagen promoter-bovine growth hormone (bGH) constructions were transfected, transiently or stably, into a variety of cells. In addition, constructions which included the first intron of the collagen gene, fused to a human growth hormone minigene, were transcribed with equal efficiency, whether the inverted repeat was intact or not, as judged by hGH mRNA levels in transfected cells.

MATERIAL AND METHODS

Plasmid constructions

pCol(IR)bGH represents a human α(I) collagen-bGH fusion gene in which an in-frame ligation was made in the DNA coding for the signal peptide sequence. (IR) indicates that the conserved inverted repeat sequence bridging the untranslated and translated portions of the first exon, was intact. A 2.25 kb collagen genomic sequence containing 804 bp of 5′ flanking sequence, the first exon and 1.2 kb of the first intron was cloned into pUC19 to produce pUC19E (14). The resulting plasmid was linearized at a unique BstXI site at +237, (15 bp downstream of the exon-intron junction) and digested with Bal 31 nuclease at 18° for 2 min. The DNA was blunt-ended with the Klenow fragment of DNA polymerase, dephosphorylated, circularized by ligation to an oligonucleotide containing a BamHI site, and cloned as a population of plasmids. The resulting DNA was digested with BamHI and AccI, the latter enzyme cleaving within the conserved inverted repeat sequence at +98, (see Fig. 1), and a population of DNA fragments with a size of 95 ±5 bp was isolated. pUC19E (in which the AccI site in the polylinker had been deleted) was then cleaved with AccI and BamHI and the population of AccI-BamHI fragments cloned into this vector. Clones were screened for distance between the AccI and BamHI sites and several were sequenced. A clone was chosen which, when subjected

to subsequent manipulations, would provide an in-reading-frame fusion with bGH. The resulting clone was cleaved with BamHI, the site blunt-ended with Mung bean nuclease, and the DNA cleaved with EcoRI. A 1.8 kb AhaII - EcoRI fragment of bGH (15) was subjected to appropriate trimming, and ligated into the collagen vector to yield pCol(IR)bGH (Fig 1A).

pCol(IR)*bGH represents a derivative of pCol(IR)bGH in which a portion of the conserved inverted repeat was deleted. pCol(IR)bGH was cleaved at the unique AccI site and digested for brief periods of time with Bal 31 nuclease. The DNA was filled in with Klenow and ligated. Clones were screened by restriction enzyme mapping and DNA sequencing. A clone was selected in which a new SmaI site had been created as a consequence of deletion of 25 bp (85-109; see Fig. 1B). The nucleotide sequence of pCol(IR)*bGH was then verified by DNA sequencing.

The structure of pCol(IR)hGH has been described (16). pCol(IR)hGH contains 2.25kb of collagen sequence including 804 bp of 5′ flanking sequence, the first exon and 1.2kb of the first intron fused, in an intron, to a fragment of the human growth hormone (hGH) gene containing the last 2 exons of the gene as well as the termination and poly(A) addition signals. The deletion created in pCol(IR)bGH was transferred to pCol(IR)hGH by partial digestion of pCol(IR)hGH with XbaI followed by ligation of a 0.4 kb XbaI-XbaI fragment containing the deletion. The correct clone, pCol(IR)*hGH, was identified by restriction enzyme mapping, including identification of a new SmaI site. The intronic sequences in pCol(IR)hGH and pCol(IR)*hGH that were located between two SstII sites at bases 292 and 1440 were deleted by SstII digestion followed by religation to produce pCol(IR)hGHΔ292-1440 and pCol(IR)*hGHΔ292-1440.

Gene transfer into cells

Baby hamster kidney (BHK), chicken tendon fibroblasts (CTF), NIH 3T3, Swiss 3T3 and BalbC 3T3 cells were grown in monolayer culture in 60 mm dishes. Transient transfection, using the calcium phosphate precipitation technique, was performed as previously described (14, 17). In some experiments the efficiency of transfection was monitored by cotransfection of a plasmid containing the β-galactosidase gene driven by the RSV LTR. Cell lysates were assayed for β-galactosidase activity by fluorescence using 4-methylumbelliferyl β-D-galactoside as a substrate. Stably transfected cell lines were isolated by selection for hygromycin resistance. Cells were cotransfected with the hygromycin resistance gene under control of the SV40 basal promoter and enhancer and grown in 0.4 mg/ml hygromycin. Stably transfected clones appeared after 2-3 weeks.

Radioimmunoassay for bovine growth hormone

bGH and rabbit antiserum to ovine growth hormone (R-anti-ovGH) were obtained from the National Hormone and Pituitary Program of the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases. I^{125}bGH was prepared by combining 5 μg bGH, 0.5 mCi NaI^{125} (Amersham), solid phase lactoperoxidase and glucose oxidase (Enzymobeads, Bio-Rad)

and β-D-glucose in phosphate buffer, pH 7.2, for 60 min at room temperature. Following centrifugation, the supernatant was applied to a 0.7 x 18 cm column of Sephadex G-25 Fine (Pharmacia) equilibrated in 50 mM barbital buffer, 1% BSA, pH 8.6. The labelled protein was collected in the void volume, tested against antiserum, and frozen in aliquots. R-anti-ovGH cross reacts strongly with bovine growth hormone and was used at a final dilution of 1:50,000 to yield ~ 50% binding of I^{125}bGH in the absence of added ligand. Nonspecific binding was ~ 5%.

Samples of cell culture medium were collected and frozen. Varying aliquots of samples and bGH standards (0.2 - 25 ng per tube) were assayed in a final volume of 500 µl, consisting of 50 mM barbital buffer, pH 8.6, 1% BSA, 10 mM sodium azide, 8.6 mM sodium chloride, 0.94 mM EDTA, 20,000 cpm I^{125}bGH, 1:50,000 R-anti-ovGH, and 0.4% rabbit serum. After overnight incubation at room temperature, goat anti-rabbit IgG (Miles) was added to a final dilution of 1:160, and PEG 8000 (Sigma) to a final concentration of 6.5%. After 2-3 h incubation at room temperature the assay tubes were centrifuged, the supernatants were aspirated, and the radioactivity in the dry pellet was counted. A standard displacement curve of % radioactivity bound versus concentrations of standard bGH was generated; sample values were read from the curve and corrected for volume and cell number.

Analysis for mRNA

Analysis for bGH mRNA was initially performed by quantitative Northern analysis using a nick-translated genomic bGH probe. Considerable size heterogeneity of bGH mRNA (in CTF) was detected which could not be attributed to repetitive sequences in intronic regions of the probe. We believe that the initial transcript of the transfected bGH gene may not be completely spliced in CTF. More reliable measurements were obtained by solution hybridization with a riboprobe. A 169 bp SmaI-PvuII fragment, representing most of the last exon of the bGH gene (15) was cloned into pGEM-1 and ^{32}P-labeled antisense RNA was synthesized with T7 RNA polymerase in the presence of [^{32}P]UTP (18). RNA sequences that were capable of hybridizing to the bGH riboprobe were detected by an RNAse protection assay (18, 19). An adaptation of this assay for hGH mRNA has been described (16).

DNA Sequence Analysis

A 487 bp PvuII fragment of pCol(IR)*bGH that included the deletion in the conserved inverted repeat sequence was cloned into the filamentous phage vector, M13mp19, and clones were recovered which carried the fragment in both orientations. These clones were sequenced by the chain termination method of Sanger et al. (20).

DNase I footprinting

HeLa nuclear extracts were prepared as described by Dignam et al. (21) and Wildeman et al. (22) and dialyzed against 20 mM Hepes, pH 7.9, 20 mM KCl, 1 mM $MgCl_2$, 0.5 mM DTT and 17% glycerol. Nuclear extracts of mouse L cells, prepared by minor modifications of the method described by Piette et al. (23), were dialyzed against the same buffer. pCol(IR)bGH

was cleaved at a unique XmaIII site at base -83, 55 bases upstream of the TATA motif, and the linear plasmid was ^{32}P-5'-end labeled on the sense strand with T4 polynucleotide kinase. The DNA was digested with PvuII and a ^{32}P-labeled 340 bp XmaIII-PvuII fragment was resolved on a 5% acrylamide gel and electroeluted. DNase I footprinting was performed by modifications of the methods of Wildeman et al. (22) and Augereau and Chambon (24) as described by Bornstein et al. (14).

RESULTS

The structure of pCol(IR)bGH and pCol(IR)*bGH

In order to determine whether the highly conserved inverted repeat sequence, present in the first exon of all interstitial collagen genes with the exception of α1(II), was involved in regulation of gene expression, we initially constructed a fusion gene in which the bGH structural gene was under control of the collagen promoter. Since the collagen translation initiation codon was present in the conserved sequence (see Fig. 1B), and since we wished to preserve this relation in order to test a possible function of the inverted repeat in translational control of expression, we fused the two sequences in the DNA coding for the signal peptide. The strategy used to construct pCol(IR)bGH is described in Fig. 1A and in Materials and Methods. An in-frame fusion was devised which preserved the translated amino acid sequence, Gly-Ala, found in the bGH gene immediately preceding the NH_2-terminal Phe residue (15), as well as the sequence for the collagen signal peptide. As expected, the collagen-bGH fusion gene, when transfected into a variety of cells, was transcribed and translated, and immunologically detectable bGH could be assayed in the culture medium of these cells (see below). A similar transient gene expression system, employing immunological detection of hGH, was developed independently by Selden et al. (25).

In creating a deletion in the conserved inverted repeat sequence, we were constrained by two considerations. We wished to preserve the functional translation initiation codon that existed in the stem of the putative stem-loop structure (see Fig. 1B) and we wished to retain the upstream ATG followed by the in-frame TAA stop codon which could contribute to translational control. We took advantage of the presence of a unique AccI site within the conserved inverted repeat sequence (Fig. 1B) to achieve a 25 bp deletion by brief (30 sec) Bal 31 nuclease digestion. Clones were screened for the presence of a new SmaI site which would be expected to form if bases 85-109 were deleted. The deletion in the clone designated pCol(IR)*bGH was confirmed by DNA sequencing. This deletion should effectively prevent the formation of a stem-loop structure in the transcribed mRNA.

The conserved inverted repeat sequence does not function in control of translation

In order to assess the consequences of the deletion in pCol(IR)*bGH, the intact and deleted plasmids were transfected into a variety of cells, including 3T3, BHK and CTF, and bGH was measured immunochemically. Both transiently and stably transfected cells were

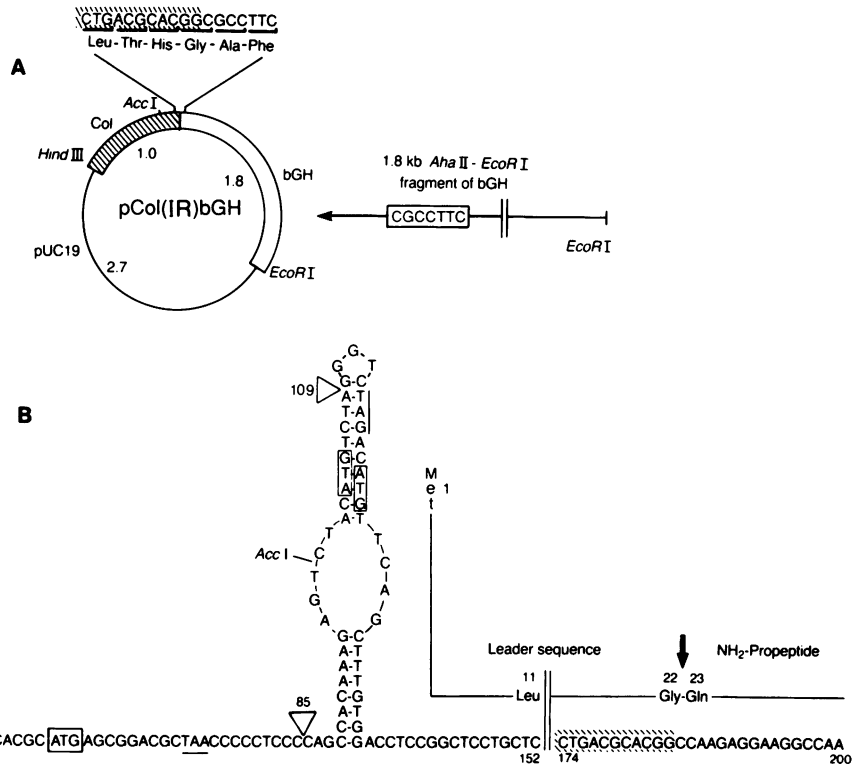

Figure 1. A) Structure of pCol(IR)-bGH. The pUC19-based plasmid contains a collagen-bGH fusion gene in which an in-frame fusion has been effected in the DNA coding for the signal peptide sequence. The collagen-derived sequence (hatched) consists of 804 bp of 5′ flanking sequence and 184 bp of the 1st exon. Fusion with the AhaII-EcoRI fragment of bGH recreates the Gly codon. Cleavage with signal peptidase would be expected at either the Gly-Ala or Ala-Phe bonds (the NH_2-terminal residue of mature bGH is Phe). B) Sequence of the top (sense) strand of the α1(I) collagen gene. The sequence is written as it might exist in mRNA if the interrupted inverted repeat were to form a stem-loop structure. The two upstream translation start sites (boxed) are followed by in-frame stop codons (underlined). The third, open-reading-frame ATG codes for the NH_2-terminal residue, Met, in procollagen. The leader or signal sequence is foreshortened as indicated by the parallel vertical lines. Cleavage of preprocollagen by signal peptidase occurs at the Gly-Gln bond (residues 22-23) as indicated by the solid arrow. The hatched sequence (bases 174-184) represents the 5′ portion of the collagen-bGH fusion as shown in A. The open arrowheads at bases 85 and 109 indicate the deletion used to construct pCol(IR)*bGH.

analyzed. In transiently transfected cells care was taken to use plasmid preparations with equal extents of supercoiling since the expression of transfected DNA may depend on the topology of the DNA (26). In CTF the efficiency of transfection was monitored by cotransfection of a plasmid containing the β-galactosidase gene driven by an RSV LTR. The results are presented in Tables 1 and 2. There was no reproducible, significant difference in the levels of bGH

Table 1. Secreted bGH Levels in Transiently Transfected Cells.

EXPERIMENT	CELL LINE	PLASMID	SECRETED bGH ($ng/10^6$ cells)
I	BHK	pCol(IR)bGH	519; 637
	"	pCol(IR)*bGH	426; 496
II	CTF	pCol(IR)bGH	42
	"	pCol(IR)*bGH	43
III	BHK	pCol(IR)bGH	53; 56; 66; 70; 119
	"	pCol(IR)*bGH	76; 84; 119; 151
IV	BHK	pCol(IR)bGH	100; 112
	"	pCol(IR)*bGH	82; 98; 120; 275

Cells were transfected with equal quantities of plasmid DNA in each experiment. In experiments I and II secreted bGH was expressed per 60 mm dish and in experiments III and IV per 10^6 cells. Each value represents a separate transfection. Culture medium was collected over a period of 48 hr.

secreted by transiently (Table 1) or stably (Table 2) transfected cells when transfected with either pCol(IR)bGH or pCol(IR)*bGH. Such differences as exist in stably transfected cells (eg. Experiment IV, Table 2) could result from the integration of different numbers of bGH genes in the several isolated clones. Other experiments, in which a clone with a smaller deletion

Table 2. Secreted bGH Levels in Stably Transfected Cells

EXPERIMENT	CELL LINE	PLASMID	bGH ($ng/10^6$ cells)
I	BHK (P)	pCol(IR)bGH	110; 145
	"	pCol(IR)*bGH	183
II	Balb C 3T3(C)	pCol(IR)bGH	28
	"	pCol(IR)*bGH	45
III	NIH 3T3 (P)	pCol(IR)bGH	132
	"	pCol(IR)*bGH	414
IV	NIH 3T3 (C)	pCol(IR)bGH	59; 151
	"	pCol(IR)*bGH	135; 227; 625
V	Swiss 3T3 (C)	pCol(IR)bGH	1519
	"	pCol(IR)*bGH	1341

Stably transfected cells were examined either as populations (P) or as isolated clones (C). Each value represents a separate transfection. In experiment IV different clones were examined. Culture medium was collected over a period of 48 hr.

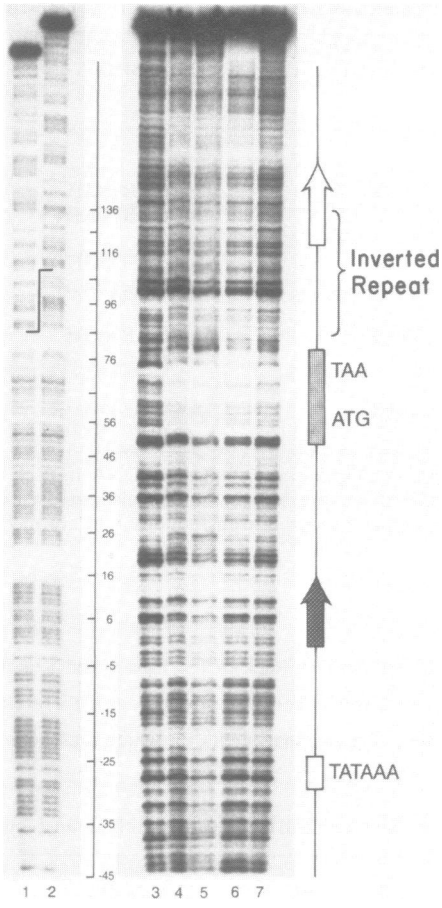

Figure 2. DNase I cleavage pattern of the sense strand of the XmaIII-PvuII fragment of pCol(IR)bGH in the presence of nuclear extracts. Base numbers between lanes 2 and 3 are relative to base 1 as the start of transcription (filled arrow). The start of translation(open arrow), the extent of the inverted repeat and the location of the TATAAA motif are shown. Lane 1, A and G sequence ladder of pCol(IR)*bGH; lane 2, A and G sequence ladder of pCol(IR)bGH. The bracket between lanes 1 and 2 indicates the extent of the deletion in pCol(IR)*bGH which encompasses the 5' half of the inverted repeat sequence; lane 3, cleavage with 5 ng DNase I in the absence of nuclear extract (naked DNA); lane 4, 30 μg L cell extract, 100 ng DNAse I; lane 5, 20 μg L cell extract, 50 ng DNase I; lane 6, 60 μg HeLa extract, 200 ng DNase I; lane 7, 40 μg HeLa extract, 100 ng DNAse I. Both extracts provide a zone of partial protection from bases 50 to 80 which corresponds to the position of the first ATG codon and in-frame stop codon (TAA) as shown on the right.

(bases 97-100) was transfected into BHK and CTF, again showed no significant differences in secreted bGH (data not shown).

Since bGH mRNA levels were not measured in the above experiments, it was possible that differences in translational efficiency of mRNAs existed but that these differences were

Table 3. Secreted bGH and mRNA levels in transiently and stably transfected CTF

Experiment	Plasmid	bGH (ng/10^6 cells)	mRNA (units/μg RNA)		Translational Efficiency (ng bGH/unit mRNA)	
			Northern Analysis	Solution Hybridization	1	2
I	pCol(IR)bGH	81	201		0.46	
	"	54	216		0.25	
	pCol(IR)*bGH	27	119		0.23	
	"	34	162		0.21	
II	pCol(IR)bGH	88	263	170	0.33	0.52
	"	241	277	190	0.87	1.27
	pCol(IR)*bGH	248	491	320	0.51	0.78
	"	381	458	430	0.83	0.89
III	pCol(IR)bGH	306	574	650	0.53	0.47
	"	41	227	180	0.18	0.23
	pCol(IR)*bGH	271	480	620	0.56	0.44
	"	130	281	320	0.46	0.41
IV	pCol(IR)bGH	54		271		0.20
	"	72		355		0.20
	pCol(IR)*bGH	100		535		0.19
	"	175		770		0.23

In Experiments I-III cells were transiently transfected with 2 μg of either pCol(IR)bGH or pCol(IR)*bGH, 1 μg of RSV-βGal and 3μg of pUC19 per 60 mm dish. bGH and mRNA levels were corrected for efficiency of transfection as determined by β-galactosidase levels in cell lysates. In Experiment IV stably transfected cells were analyzed as mixed populations of hygromycin-resistant cells. mRNA levels were determined either by densitometric scanning of Northern blots and expressed as arbitrary units per μg RNA or as radioactivity per μg RNA in a solution hybridization procedure (see Materials and Methods). Translational efficiency was determined in Column 1 by expressing bGH levels as a function of densitometric mRNA units in Northern analyses or in column 2 as a function of mRNA determined by solution hybridization.

compensated, coincidentally, by reciprocal changes in mRNA levels. We therefore performed a series of experiments in CTF in which both secreted bGH and mRNA levels were measured. Translational efficiency was calculated by expressing secreted bGH levels as a function of mRNA levels, determined either by quantitative Northern analysis or by a solution hybridization procedure. The results of these experiments, in which duplicate plates were analyzed, indicated that bGH mRNA was translated with the same degree of efficiency whether the conserved inverted repeat sequence was present in the 5′ untranslated region or not (Table 3).

Protein binding to DNA in the 5′ untranslated region of the α1(I) gene

Although analysis of bGH levels in transfected cells did not provide support for a role of the conserved inverted repeat sequence in gene regulation we examined binding of proteins to this and adjacent sequences in the 5′ untranslated region of the gene. DNase I protection footprinting of the region from the XmaIII site at -84 to the Pvu II site beyond the translation initation codon at +120 provided evidence for protein binding to DNA in the sequence from +50 to +80 (Fig. 2). Nuclear extracts from HeLa and L cells provided a roughly equal degree of partial protection. The zone of protection does not encompass the conserved inverted repeat sequence but does involve an equally conserved sequence which contains an upstream ATG followed by an in-frame stop codon (Fig. 2). As shown in Figure 2, the deletion that distinguishes pCol(IR)*bGH from pCol(IR)bGH involves the 5′ region of the inverted repeat, just downstream from the zone of DNaseI protection.

Lack of evidence for a role of the conserved inverted repeat sequence in transcriptional control

The experiments described in Table 3 do not support a role for the inverted repeat sequence in transcriptional control since bGH mRNA levels, expressed per μg RNA, did not change as a consequence of the deletion present in pCol(IR)*bGH. However, recent experiments indicate that a negatively acting element in the first intron of the α1(I) collagen gene functions optimally only in the presence of an intact 5′ flanking sequence (Bornstein, P. and McKay, J., manuscript in preparation.). It therefore occurred to us that the conserved inverted repeat sequence might not function as a regulatory element unless collagen intronic sequences were also introduced into the test plasmid. We therefore transferred the deletion in pCol(IR)*bGH to pCol(IR)hGH creating pCol(IR)*hGH (see Materials and Methods). The collagen-hGH fusion gene differs from the collagen-bGH construct in that the former contains the collagen intronic sequence present in pUC19E (14, 16) and substitutes a hGH minigene for the bGH gene. A protein product would not be expected following transfection of pCol(IR)hGH into cells but mRNA that hybridizes with an antisense riboprobe made to the 5th (last) exon in the hGH gene can be detected (16; see below).

CTF were transfected transiently with pCol(IR)hGH or pCol(IR)*hGH and RNA was assayed for hGH mRNA by a solution hybridization procedure. There was no difference in the level of specific hGH mRNA in cells transfected with pCol(IR)*hGH in comparison with pCol(IR)hGH (Figure 3, lanes 2 and 3; Figure 4). Furthermore, mRNA levels in cells trans-

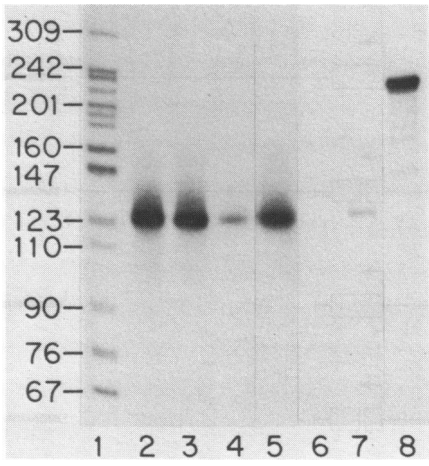

Figure 3. Autoradiogram of hGH mRNA analyzed by solution hybridization. RNA from CTF, transiently transfected with collagen-hGH plasmids, was hybridized with a ^{32}P-labeled antisense hGH riboprobe. After RNAse digestion, ethanol-precipitable material was analyzed by electrophoresis in a 6% acrylamide, 50% urea gel. Lane 1, Msp 1 digest of pBR322 end-labeled with ^{32}P and T4 kinase; lane 2, pCol(IR)hGH; lane 3, pCol(IR)*hGH; lane 4, pCol(IR)hGHΔ292-1440; lane 5, pCol(IR)*hGHΔ292-1440; lane 6 RNA from transgenic mice carrying the hGH gene; lane 7 RNA from cells transfected with a CAT gene; lane 8, hGH riboprobe. The relative hGH mRNA levels in this experiment, corrected for efficiency of transfection, are: lane 2, 1.0; lane 3, 1.1; lane 4, 0.28; lane 5, 0.31.

fected with pCol(IR)hGHΔ292-1440 and pCol(IR)*hGHΔ292-1440 are indistinguishable (Figure 3, lanes 4 and 5; Figure 4). We therefore conclude, subject to the reservations raised in the Discussion, that the conserved inverted repeat sequence in the α1(I) collagen gene, and by inference in related genes, does not function in transcriptional or translational control of gene expression.

DISCUSSION

A number of studies have shown that synthetic rates for type I collagen do not always correlate with mRNA levels, suggesting an element of translational control (27-31), but the mechanisms responsible for such control have remained elusive. In 1979 Wiestner et al. (32), showed that peptide fragments representing the NH_2-terminal pro-domains of bovine types I and III collagens inhibited the synthesis of these collagens by fibroblasts in culture, but that type II collagen synthesis by chondrocytes was not affected. Subsequently it was shown that bovine type I collagen "propeptides" were capable of inhibiting, rather specifically, the translation of type I procollagen mRNA in an in vitro reticulocyte lysate translation system (33, 34). A synthetic peptide representing 22 amino acids in the COOH-terminal propeptide of the human α2(I) procollagen chain was also shown to inhibit selectively collagen and fibronectin

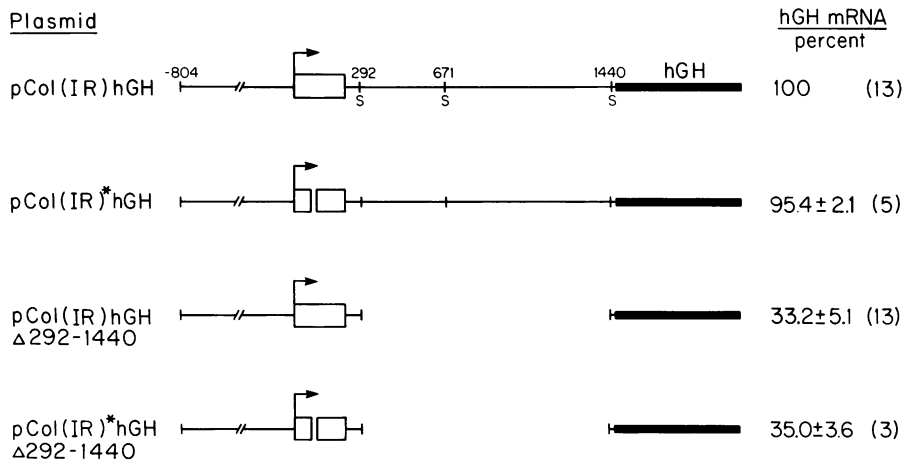

Figure 4. Diagrammatic representation of deletions in pCol(IR)hGH and pCol(IR)*hGH and tabulation of transcriptional activity of the resulting plasmids. The start of transcription is shown by a horizontal arrow and the first (collagen) exon by an open rectangle. The hGH minigene is represented by a solid bar. The collagen part of the first intron extends from nucleotide 222 to 1452 and contains 3 SstII cleavage sites (S) at nucleotides 292, 671 and 1440. The deletion in the 5′ untranslated part of the first exon is indicated by a vertical space in pCol(IR)*hGH. Deletion of much of the first intron was achieved by SstII digestion and religation and is indicated by Δ 292-1440. The transcriptional activity of each modified plasmid ± S.E.M., relative to pCol(IR)hGH which was set at 100, is tabulated. The number of independent determinations is given in parentheses.

synthesis by human fibroblasts in the absence of detectable changes in mRNA levels (35). On the other hand, evidence against specificity of translational inhibition by chick α1(I) NH_2-terminal propeptides (36) and evidence for specific pretranslational effects of both NH_2-terminal and COOH-terminal chick propeptides (37) has been presented.

None of the above effects has been directly attributed to a specific element in collagen genes. Nevertheless, the finding of a conserved inverted repeat sequence in the 5′ untranslated region of type I and III collagen genes, and its absence in the homologous type II gene, has led to speculation that this sequence might be involved in translational control of collagen synthesis (1, 3).

Several features of the 5′ untranslated region of type I and type III collagen genes do provide a basis for translational control. Kozak (10) has shown that highly stable stem loop structures (in mRNA) can inhibit translation of preproinsulin in COS cells transfected by several vectors. This was true whether the oligonucleotides used to create the hairpin structure were placed near the initiator ATG codon or upstream. On the other hand, an only moderately stable stem loop structure was ineffective in translational control (10). A role in posttranscriptional control has also been attributed to dyads of symmetry present in the 5′ untranslated regions of

the human T cell type III lymphotropic virus (38, 39), in the human heavy chain ferritin gene (40) and in the cytomegalovirus β gene (42). In these cases other factors, such as the ability of the sequence to serve as a site for an RNA binding protein, could influence the translatability of the mRNA.

We have recalculated the stability of the highly stable and moderately stable structures designed by Kozak (10) using the newly reported parameters of Freier et al. (42) and find them to be ΔG°_{37} = -31.3 and -21.9 Kcal/mole, respectively. In comparison, the naturally occurring structure in the human α1(I) gene, depicted in Figure 1B, was calculated to have a stability of ΔG°_{37} = -15.0 Kcal/mole. Similarly, the analogous structures found in the chick α1(I), α2(I) and α1(III) genes (1) and in the mouse α1(III) gene (4) were all found to be less stable than the moderately stable structure designed by Kozak (10). However, Rossi and de Crombrugghe (8) have recently provided evidence for the formation of an RNA dimer of the inverted repeat sequence in the mouse α2(I) gene by antisense, intermolecular, self-hybridization. The stability of such a structure would be considerably greater than that of an intramolecular stem-loop. Dimerization was favored by addition of a fibroblast ribosomal eluate. Rossi and de Crombrugghe (8) suggest, on the basis of unpublished experiments analogous to those reported here, but involving collagen-CAT constructs, that the inverted repeat sequence does exert translational control. We do not have a good explanation for these differences. It is possible that secondary structure in intrinsically unstable CAT mRNA may adversely affect mRNA half-life but be less of a factor in more stable growth hormone or collagen mRNA. It should also be noted that the relation of the promoter to the first exon in the collagen-bGH constructs described here more closely resembles those in the native collagen gene than is true of collagen-CAT constructions. Although the activation of a putative RNA binding protein could be under hormonal or other physiologic control or could be restricted to certain cell types, our data indicate that the conserved inverted repeat is not involved in general translational regulation.

A second feature of the 5′ untranslated region of types I and III collagen genes that could mediate translational control is the invariant presence of two ATG codons followed by in-frame stop codons upstream of the initiator ATG. Short upstream reading frames have been found in the mRNA of the yeast GCN4 gene (43), and in the cytomegalovirus β transcript (41) and have been thought to modulate gene expression. In the human, mouse and chicken α1(I) and α2(I) genes the proximal upstream ATG is separated by 14 bases from the initiator codon whereas the more 5′ ATG is located 36-40 bases further upstream in the α1(I) genes and about 60 bases upstream in the α2(I) genes. In accord with the scanning model for eukaryotic translational initiation, upstream translational initiation codons can severely depress initiation of translation from the authentic start of translation (44). In some studies, the presence of in-frame termination codons following these upstream initiation sites was found to modify this inhibitory effect (44, 45) although this did not appear to be the case for GCN4, the yeast transcriptional activator of amino acid biosynthetic genes (43). Peabody and Berg (46) have provided

additional evidence for translational initiation at internal AUGs and suggest that reinitiation can occur even if the termination codon in the 5′ reading frame is positioned a short distance downstream of the 3′ AUG.

The extent to which the several initiation codons conform to the most favorable sequence context, CC$\begin{matrix}A\\G\end{matrix}$CCATGG, will also modify the effects of multiple initiation codons. Purines at -3 and +4 (relative to ATG in positions +1 to +3) appear to be most crucial in this regard (11, 47). The presence of the authentic start codon in a strong sequence context and upstream start codons in weaker contexts would tend to reduce the negative influence of the latter codons. Examination of the sequence of the human α1(I) gene (Figure 1) reveals that both upstream ATG codons are followed by in-frame termination codons and that none of the three start sites exists in a strong sequence context. As a consequence, the potential for translational control by this region remains uncertain. In our experiments the deletion in pCol(IR)*bGH removes the middle ATG (Figure 1B) without affecting the translatability of the resulting transcript (Table 3).

We have considered the possibility that the conserved inverted repeat sequence may be involved in transcriptional control of collagen gene expression. Our preliminary experiments do not support a role in initiation of transcription since hGH mRNA levels were unchanged when the upstream arm of the dyad of symmetry was deleted (Figures 3 and 4) and DNAse I protection experiments using HeLa and L cell nuclear extracts fail to reveal evidence for binding of proteins to the inverted repeat sequence (Fig. 2). It is nevertheless possible that additional regions of the collagen gene, either upstream or downstream from the sequence placed in pCol(IR)hGH, will be required in order that a transcriptional function be manifest. Alternatively, the inverted repeat sequence may mediate a functional change in the rate of collagen gene transcription, such as that caused by a hormone or growth factor. Experiments to test these possibilities are currently in progress.

ACKNOWLEDGMENTS

This work was supported in part by Grants AM11248 and DEO2600 from the National Institutes of Health. We are greatly indebted to Richard Palmiter for suggesting the strategy for construction of pCol(IR)bGH, for providing bovine and human growth hormone plasmids, and for helpful advice throughout the course of this work. We also thank Helene Sage for a critical reading of the manuscript and Kathleen Doehring for secretarial assistance.

*To whom correspondence should be addressed.

REFERENCES

1. Yamada, Y., Mudryj, M., and de Crombrugghe, B. (1983) J. Biol. Chem. **258**, 14914-14919.
2. Schmidt, A., Yamada, Y., and de Crombrugghe, B. (1984) J. Biol. Chem. **259**, 7411-7415.

3. Chu, M. L., deWet, W., Bernard, M. and Ramirez, F. (1985) J. Biol. Chem. **260**, 2315-2320.
4. Liau, G., Mudryj, M., and de Crombrugghe, B. (1985) J. Biol. Chem. **260**, 3773-3777.
5. Harbers, K., Kuehn, K., Delius, H. and Jaenisch, R. (1984) Proc. Natl. Acad. Sci. USA **81**, 1504-1508.
6. Dickson, L. A., de Wet, W., Di Liberto, M., Weil, D. and Ramirez, F. (1985) Nucleic Acids Res. **13**, 3427-3438.
7. Finer, M. H., Aho, S., Gerstenfeld, L. C., Boedtker, H. and Doty, P. (1987) J. Biol. Chem. **262**, 13323-13332.
8. Rossi, P. and deCrombrugghe, B. (1987) Nucleic Acids Res. **15**, 8935-8956.
9. Pelletier, S. and Sonenberg, N. (1985) Mol. Cell. Biol. **5**, 3222-3230.
10. Kozak, M. (1986) Proc. Natl. Acad. Sci. USA **83**, 2850-2854.
11. Kozak, M. (1986) Cell 44, 283-292.
12. Kohno, K., Sullivan, M., and Yamada, Y. (1985) J. Biol. Chem. **260**, 4441-4447.
13. Nunez, A. M., Kohno, K., Martin, G. R., and Yamada, Y. (1986) Gene **44**, 11-16.
14. Bornstein, P., McKay, J., Morishima, J., Devarayalu, S., and Gelinas, R. E. (1987) Proc. Natl. Acad. Sci. USA **84**, 8869-8873.
15. Gordon, D. F., Quick, D. P., Erwin, C. R., Donelson, J. E. and Maurer, R. A. (1983) Mol. & Cellular Endocrin. **33**, 81-95.
16. Bornstein, P. and McKay, J. (1988) J. Biol. Chem. **263**, 1603-1606.
17. Gorman, C. (1985) In Glover, D. M. (ed.), DNA Cloning, A Practical Approach, IRL Press, Oxford, Vol. II, pp.143-190.
18. Melton, D. A., Krieg, P. A., Rebagliati, T., Maniatis, T., Zinn, K. and Green, M. R. (1984) Nucleic Acids Res. **12**, 7035-7056.
19. Myers, R. M., Larin, Z. and Maniatis, T. (1985) Science **230**, 1242-1246.
20. Sanger, F., Nicklen, S., and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
21. Dignam, J. D., Lebowitz, R. M., and Roeder, R. G. (1983) Nucleic Acids Res. **11**, 1475-1489.
22. Wildeman, A. G., Sassone-Corsi, P., Grundström, T., Zenke, M., and Chambon, P. (1984) EMBO J. **3**, 3129-3133.
23. Piette, J., Kryszki, M-H., and Yaniv, M. (1985) EMBO J. **4**, 2675-2685.
24. Augereau, P. and Chambon, P. (1986) EMBO J. **5**, 1791-1797.
25. Selden, R. F., Howie, K. B., Rowe, M. E., Goodman, H. M. and Moore, D. D. (1986) Mol. and Cell. Biol. **6**, 3173-3174.
26. Weintraub, H., Cheng, P. F. and Conrad, K. (1986) Cell **46**, 115-122.
27. Tolstoshev, P., Berg, R. A., Rennard, S. I., Bradley, K. H., Trapnell, B. C., and Crystal, R. G. (1981) J. Biol. Chem. **256**, 3135-3140.
28. Tolstoshev, P., Haber, R., Trapnell, B. C., and Crystal, R. G. (1981) J. Biol. Chem. **256**, 9672-9679.
29. Allebach, E. S., Boettiger, D., Pacifici, M., and Adams, S. L. (1985) Mol. Cell. Biol. **5**, 1002-1008.
30. Focht, R. J., and Adams, S. L. (1984) Mol. Cell. Biol. **4**, 1843-1852.
31. Stepp, M. A., Kindy, M. S., Franzblau, C. and Sonenshein, G. E. (1986) J. Biol. Chem. **261**, 6542-6547.
32. Wiestner, M., Krieg, T., Hörlein, D., and Müller, P. (1979) J. Biol. Chem. **254**, 7016-7023.
33. Paglia, L., Wilczek, J., Diaz de Leon, L., Martin, G. R., Hörlein, D., and Muller, P. (1979) Biochemistry **18**, 5030-5034.
34. Hörlein, D., McPherson, J., Goh, S. H. and Bornstein, P. (1981) Proc. Natl. Acad. Sci. USA **78**, 6163-6167.
35. Aycock, R. S., Raghow, R., Stricklin, G. P., Seyer, J. M., and Kang, A. H. (1986) J. Biol. Chem. **261**, 14355-14360.
36. Goldenberg, R. and Fine, R. E. (1985) Biochimica et Biophysica Acta **826**, 101-107.
37. Wu, C. H., Donovan, C. B., and Wu, G. Y. (1986) J. Biol. Chem. **261**, 10482-10484.
38. Rosen, C. A., Sodroski, J. G., Goh, W. C., Dayton, A. I., Lippke, J. and Haseltine, W. A. (1986) Nature **319**, 555-559.

39. Okamoto, T. and Wong-Staal, F. (1986) Cell **47**, 29-35.
40. Hentze, M. W., Rouault, T. A., Caughman, S. W., Dancis, A., Harford, J. B., and Klausner, R. D. (1987) Proc. Natl. Acad. Sci. USA **84**, 6730-6734.
41. Geballe, A. P., Spaete, R. R. and Mocarski, E. S. (1986) Cell **46**, 865-872.
42. Freier, S. M., Kierzek, R., Jaeger, J. A., Sugimoto, N., Caruthers, M. H., Neilson, T. and Turner, D. H. (1986) Proc. Natl. Acad. Sci. USA **83**, 9373-9377.
43. Mueller, P. P. and Hinnebusch, A. G. (1986) Cell **45**, 201-207.
44. Kozak, M. (1984) Nucl. Acids Res. **12**, 3873-3893.
45. Liu, C-C., Simonsen, C. S., and Levinson, A. D. (1984) Nature **309**, 82-85.
46. Peabody, D. S. and Berg, P. (1986) Mol. and Cell. Biol. **6**, 2695-2703.
47. Peabody, D. S. (1987) J. Biol. Chem. **262**, 11847-11851.