[Frontiers in Bioscience 10, 1768-1778, May 1, 2005]

ORGANIZATION AND EXPRESSION OF THE HUMAN SERPIN GENE CLUSTER AT 14q32.1

Mark D. Marsden and R. E. K. Fournier

Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, WA 98109

TABLE OF CONTENTS

1. ABSTRACT

The serpins are a superfamily of gene sequences that have been conserved through evolution. These genes encode protein products that perform a variety of functions *in vivo*, and their regulation differs among different cell types. About one-third of the serpin genes in the human genome are located at 14q32.1, and the serpin genes in this ~370 kb region are organized into discrete proximal, central, and distal subclusters of four, three, and four genes each. In this report we discuss the genomic organization of the 14q32.1 serpin gene cluster, and we summarize what is known about the regulation of each serpin gene in this region. An approach for studying locus-wide regulation of chromosomal serpin genes *in situ* is also described. Using this approach, specific mutations in the proximal serpin subcluster were prepared by homologous recombination. These mutant alleles define a serpin locus control region that regulates gene activity and chromatin structure of the entire proximal subcluster. Prospects for further analyses of this complex genomic domain are discussed.

2. THE SERPIN GENE SUPERFAMILY

The serpins comprise a superfamily of genes that have been identified in animals, plants, and viruses. Some serpins function as *ser*ine *p*rotease *in*hibitors, from which the name of the superfamily derives, but others perform a variety of diverse physiological functions *in vivo*. Phylogenetic analyses suggest that the serpins can be divided into sixteen distinct clades, plus a group of highly diverged orphan sequences (reviewed in (1)). This classification forms the basis of a systematic nomenclature for serpin genes and gene products. There are more than 35 different serpin gene sequences dispersed throughout the human genome (2), but most of these genes are located in two discrete clusters at 14q32.1 and 18q21.3. Each of these clusters contains 11 serpin genes, and they occupy 390 kb and 520 kb of genomic DNA, respectively. The 14q32.1 and 18q21.3 clusters contain only serpin genes within them, and the compositions and structures of these loci are conserved in mammalian genomes.

2.1. Serpin Genes at 14q32.1

The serpin genes at 14q32.1 are arranged in two different transcriptional orientations: Kallistatin-like (KAL-like), α1-antichymotrypsin (ACT/SerpinA3), Protein C inhibitor (PCI/SerpinA5), and Kallistatin (KAL/SerpinA4) are all transcribed proximal (centromeric) to distal (telomeric), while Vaspin (OL-64/SerpinA12), Centerin (CEN/GCET1/SerpinA9), Antiproteinase-like 2 (APL-2/SerpinA11), α1-antitrypsin (α1AT/SerpinA1), α1-antitrypsin-related (ATR/SerpinA2), Corticosteroid binding globulin (CBG/SerpinA6), and Protein Z inhibitor (ZPI/SerpinA10) are orientated distal to proximal (Figure 1). All 11 genes have significant sequence homology with each other, and most of them share a gene structure that is similar to the consensus, which consists of five exons, including an untranslated first exon and a large second exon. However, the α1AT gene has a different structure

Figure 1. The serpin gene cluster at 14q32.1. The genomic organization of the 11 serpin genes in the 14q32.1 cluster is shown, from distal (left) to proximal (right) on the chromosome. The distal subcluster includes Kallistatin-like (KAL-like), α1-antichymotrypsin (ACT/SerpinA3), Protein C inhibitor (PCI/SerpinA5), and Kallistatin (KAL/SerpinA4); these genes are all transcribed proximal to distal. The central subcluster includes Vaspin (OL-64/SerpinA12), Centerin (CEN/GCET1/SerpinA9), and Antiproteinase-like 2 (APL-2/SerpinA11). The proximal subcluster includes α1-antitrypsin (α1AT/SerpinA1), α1-antitrypsin-related (ATR/SerpinA2), Corticosteroid binding globulin (CBG/SerpinA6), and Protein Z inhibitor (ZPI/SerpinA10). Genes in the central and proiximal subclusters are all transcribed distal to proximal.

composed of alternate exons IA, IB, and IC in addition to exons II, III, IV, and V. The α1AT gene also has dual promoters that are differentially utilized in different cell types, and the α1AT primary transcript is subject to a complex pattern of differential splicing in macrophages (3-5). In contrast, the ATR pseudogene, which is highly homologous to α1AT and located ~10 kb downstream, lacks exon I and its associated promoter. This suggests that recombination between α1AT and ATR sequences may have generated the atypical structures of these two adjacent serpin genes.

The 14q32.1 serpin gene cluster is bounded by the Goosecoid (GSC) and KIAA1622 genes at the telomeric and centromeric extremes of the domain, respectively. The expression patterns and structures of these non-serpin genes are completely different from those within the serpin domain. In particular, GSC is expressed during gastrulation and encodes a homeodomain transcription factor with a DNA binding specificity identical to that of the anterior morphogen, bicoid, in *Drosophila*. KIAA1622 is expressed in brain, ovary, and testis, and it encodes a protein with HEAT-like tandem repeats found in a number of cytoplasmic proteins.

Despite genetic similarities, the serpins at 14q32.1 are expressed in a wide variety of different tissues, including liver, macrophages, and brain, and they have a diverse set of functions *in vivo*. For example, the protein products of the PCI, KAL, and ZPI genes are all important in the blood coagulation cascade, CBG transports hormones, CEN has a role in selective maturation of B cells, and both α1AT and ACT are acute phase reactants that limit the effects of proteases at sites of infection and inflammation.

Candidate sequences for important regulatory elements within the serpin locus at 14q32.1 have been identified by several means, including the proximity of sequences to promoter regions and the locations of non-transcribed, disease-associated mutations in patients. More recently, the availability of maps containing chromatin landmarks, such as DNase I-hypersensitive sites (DHSs) and matrix attachment regions (MARs), have provided signposts to putative regulatory elements. As the serpin locus is conserved among mammalian genomes, sequence comparisons between syntenic loci in humans, mice, and rats will allow further candidate regulatory elements to be identified.

Functional regulatory elements have been mapped within the serpin locus by two approaches- first, using an isolation approach, where candidate elements are cloned and assayed in heterologous tests, and, second, using an *in situ* strategy where putative regulatory elements are manipulated directly on the chromosome, and the effects assayed in context.

Here, we describe the serpin cluster at 14q32.1 on a gene-by-gene basis. Several of these genes have been identified only recently, while the expression patterns and functions of other genes have been studied in detail. For example, α1AT and ACT are the most well-studied genes of the serpin cluster, as they are among the most highly expressed genes in human liver; indeed, the α1AT protein makes up ~1% of total serum protein (6). Moreover, both α1AT and ACT are medically important loci, as mutations in the ACT gene are implicated in Alzheimer's disease, and changes in α1AT are associated with chronic emphysema.

2.2. Newly Discovered Serpin Genes

The availability of human genome primary sequence has allowed cDNA sequences to be mapped to their chromosomal locations, confirming the presence of *in silico* predicted genes within the 14q32.1 serpin cluster (7). For example, a cDNA isolated from a library using a signal peptide trap for transmembrane and secreted proteins maps to the KAL-like gene (8), and mRNA from this gene is detected in human liver (unpublished). Similarly, distinct cDNAs isolated from human melanoma and human fetal libraries map to the Vaspin and APL-2 genes, respectively (9). The APL-2 mRNA is also detected in the human hepatoma cell line, HepG2 (unpublished).

Centerin is transcribed in B cells within the germinal centers of secondary lymphoid tissue, where naive B cells undergo selective maturation (10). Four alternatively spliced mRNAs have been described (11).

Centerin is up-regulated in response to CD40 signaling *in vitro*, suggesting that the gene has a role in the selective maturation of naive B cells to memory and plasma lymphocytes.

2.3. ZPI (SerpinA10), PCI (SerpinA5), KAL (SerpinA4), CBG (SerpinA6), and ATR (SerpinA2)

ZPI is expressed in liver but not in heart, brain, placenta, lung, skeletal muscle, kidney, or pancreas (12). ZPI inhibits the blood coagulant factor Xa; however, inhibition is enhanced more than 1000-fold through an additional interaction with protein Z in the presence of calcium ions and procoagulant phospholipids (13).

PCI is expressed mainly in the liver, but it is also detected locally in kidney, prostate, testis, seminal vesicles, and epididymal glands (14). The PCI gene product inhibits proteases of the blood coagulation cascade, such as activated protein C, thrombin, factor Xa, factor XIa, kallikreins, and the thrombin/thrombomodulin complex (15-17). PCI also affects proteases important in fertilization, including the prostate-specific antigen and the sperm protease, acrosin (18).

The PCI promoter lacks canonical TATA and CAAT sequences (19). A 329 bp minimal promoter is sufficient to drive transgene expression in HepG2 cells, and this region contains an Sp-1 consensus binding site at -294 to –302 bp (20). In gel shift assays, antibodies against Sp-1 recognize protein complexes bound to the -269 to –307 bp region, and deletion of this sequence results in a 50-fold reduction in transgene activity. An enhancer located between -329 to -385 bp contains an AP-2 consensus binding site; deletion of this region causes an ~10-fold reduction in transgene activity.

KAL and CBG both have TATA and CAAT boxes in their promoter regions- KAL at –31 bp and –55 bp (21), and CBG at –28 bp and –73 bp (22), respectively. The KAL promoter contains consensus binding sites for AP-2 at 127 bp, and AP-1 at –330 bp, in addition to a 34 bp GC-rich region at -183 bp and a cAMP-response element (CRE) at –269 bp (21). The CBG promoter contains a putative nuclear factor IL-6 (NF-IL-6) binding element at –215 bp that is conserved with a known NF-IL-6 binding site present in the rat CBG promoter (23). In HepG2 cells treated with IL-6, CBG is downregulated, while other serpin genes, such as α1AT and ACT, are upregulated in a dose-dependent manner (24).

Both KAL and CBG are expressed in liver, kidney, and testes. KAL is also expressed in heart, smooth muscle, and endothelia (25). The KAL gene product inhibits the Kallikrein protease, but it may also have multiple roles in vascular function such as vasodilation (26) and angiogenesis (27). The CBG protein is not a serine protease inhibitor; rather, it acts as a plasma hormone carrier that binds cortisol.

There is no evidence that ATR is expressed, but there are no termination codons within ATR coding sequences, and all of the RNA splice sites are conserved in the genomic sequence. Thus, ATR is likely a recently evolved pseudogene.

2.4. ACT (SerpinA3)

ACT is expressed in liver, brain, lung, and cornea, and the gene product inhibits chymotrypsin-type proteases, including Cathepsin G (28). Expression of ACT is enhanced in the affected brain regions of patients with Alzheimer's disease, where the protein co-localizes with filamentous β-amyloid deposits. ACT is believed to affect both the formation and destabilization of β-amyloid fibrils in these areas (29, 30).

There is a single ACT transcription start site in both liver and brain, with a TATA consensus box at -29 bp (31). As ACT and α1AT are both acute phase reactants, the concentrations of their gene products increase rapidly in response to humoral signals mediated by cytokines. Oncostatin-M (Os-M), IL-1, and TNFα up-regulate ACT in hepatocytes (32). In primary human astrocytes, ACT is up-regulated by Os-M, IL-1β, and TNFα, and the synthetic glucocorticoid, dexamethasone, enhances the response of ACT to each of these cytokines (33). STAT1 (signal transducer and activator of transcription 1) and STAT3 binding sites located at –87 to –95 bp and –117 to -125 bp, respectively, mediate the Os-M response of ACT transgenes. Mutations in either site cause an ~5-fold reduction in the Os-M response of reporter constructs, while mutations in both sites result in ablation of the Os-M response (33).

Reporter genes linked to sequences between –13,227 and –12,814 bp respond to IL-1 and TNFα in astrocytes (34). This 413 bp region contains three elements- an NFκβ binding site (-13,202 to –13,213 bp), a consensus binding site for AP-1 (-12,979 to –12,985 bp), and a consensus binding site for NFκβ (-12,820 to –12,831 bp). Mutation and deletion analyses have shown that all three elements are required by transgenes for a full response to IL-1 and TNF, although mutations in the NFκβ and AP-1 elements have the greatest effect. This regulatory region is conserved in the rat and mouse (unpublished).

2.5. α1AT (SerpinA1)

α1AT is the most thoroughly studied gene in the 14q32.1 serpin locus, and it provides an example of how the two complementary approaches for examining regulatory elements can be employed- an isolation approach, where putative elements are cloned and assayed as transgenes, and an *in situ* strategy, where putative regulatory elements are examined in their normal chromosomal contexts.

As mentioned above, α1AT is atypical of the serpin genes at 14q32.1 in that transcription initiates from two distinct regions ~2 kb apart; a proximal promoter that is active in liver, intestine, and lung, and a distal promoter that is active in monocytes, macrophages, and cornea. The presence of two promoters indicates that gene expression is modulated by alternative mechanisms in different tissues. In the following discussion, we define 0 bp of the α1AT gene as the hepatic transcription start site.

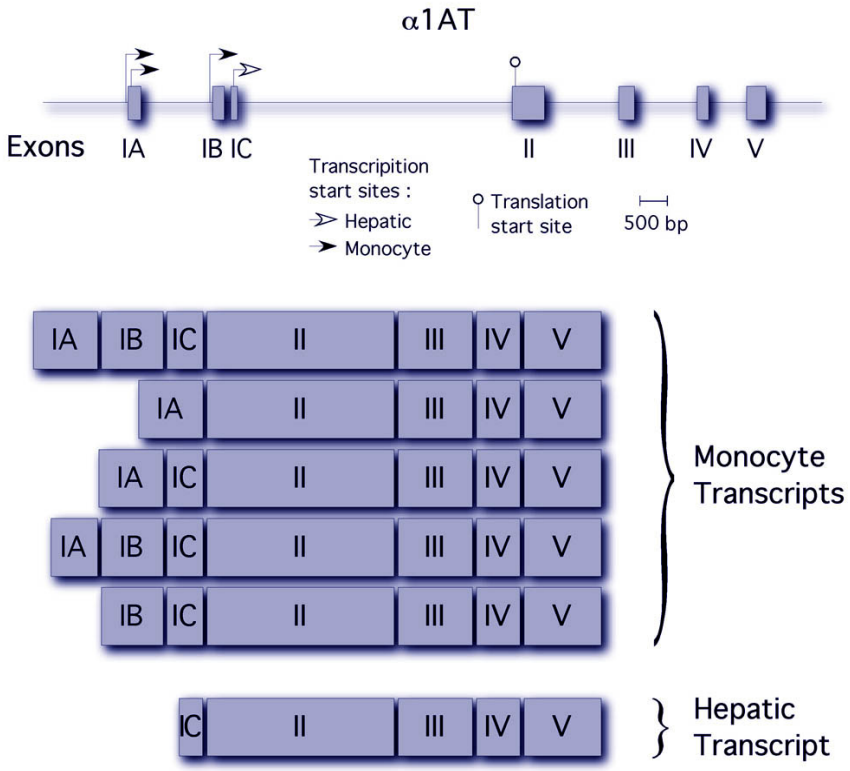

Figure 2. Promoter utilization and alternative splicing of human α1AT transcripts. Top: The α1AT gene. Transcription initiates at –2064, -2039 and –406 bp in monocytes. In hepatocytes, transcription initiates at 0 bp. The translation start site (ATG) is at position 5360 bp. Bottom: The structures of α1AT transcripts in monocytes and hepatocytes.

In monocytes, α1AT transcription initiates at three distinct sites, at -2064, -2028, and -406 bp. The three nuclear transcripts produce alternatively spliced mRNAs, which differ only in the presence or absence of the untranslated exons, IA, IB and IC (Figure 2) (3-5). Monocyte α_1AT production is further stimulated by circulating IL-6, lipopolysaccharide, IL-1β, and TNFα, which all cause an ~3-fold increase in α_1AT protein production (35-37). Interestingly, when the leukemic monocyte cell line, U937, is stimulated with IL-6, the proportion of α1AT transcripts lacking exon IB increases relative to other α1AT transcripts, suggesting that removal is important, and that exon IB may act as a translational repressor (38).

Corneal α1AT transcription initiates at -2039 bp within the α1AT distal promoter (39). Sequences between -2.0 and -3.4 kb are sufficient to target reporter gene expression to cornea and lymphocytes in transgenic mice (40). A shorter 489 bp promoter fragment is sufficient to drive transgene expression in corneal stromal cells. Transgene experiments show that the isolated promoter fragment contains two regions that are important for transgene expression in corneal stromal cells- enhancer elements between -2519 and -2359 bp, and repressive elements between -2139 and -2126 bp that respond to overexpression of Sp-1 (41). In cornea organ culture, α1AT expression responds to humoral stimuli such as IL-2 and IL-1β, with an ~3-fold increase in α1AT protein secretion (37).

Transcripts from the α1AT proximal promoter are not known to be subject to alternative splicing. A short DNA fragment between -37 and -137 bp linked to reporter genes is sufficient to drive cell-specific expression from the SV40 promoter in the human hepatoma cell line, Hep3B, but not in HeLa cells (42-44). Within this 100 bp DNA fragment are distinct binding sites for the transactivators HNF-1α (-73 to -66 bp) and HNF-4 (-119 to -108 bp). HNF-1α is the major transactivator of α1AT transcription in the liver, as site-specific mutagenesis of the HNF-1α binding site reduced reporter gene activity over 100-fold in hepatic cells (43). In contrast, mutation of the HNF-4 binding site reduced reporter activity only about 10-fold. These results were recapitulated in experiments where cosmid constructs containing these same HNF-1α and HNF-4 mutations upstream of the human α1AT gene were introduced into transgenic mice: disruption of the HNF-1α binding site showed an ~100-fold reduction in liver-specific expression of the human α1AT gene, but disruption of the HNF-4 binding site had only a small effect on human α1AT gene expression (45).

Two regions further upstream of α1AT seem to enhance transgene expression in expressing cells, although these effects are modest. Deletion of sequences between -

208 and -261 bp causes a 2-3-fold decrease in reporter gene expression, while sequences between -488 and -356 bp mediate an ~4-fold increase in reporter gene activity over an enhancerless SV40 promoter (43). Cloned sequences in this region are also important for expression in the human intestinal epithelium cell line, CaCo-2, but this effect is only ~2-fold (46). Elements between -661 and -991 bp also enhance gene expression in CaCo-2, but not HepG2, cells.

The mouse may not be an ideal model system with which to study human α1AT gene regulatory elements, as there are several murine α1AT isoforms, suggesting that murine and human α1AT physiology may be different. However, a transgene driven by the human α1AT promoter in transgenic mice (-348 to +15 bp) is sufficient to target reporter gene activity to the liver, and the correct developmental pattern of gene expression is observed (47). Interestingly, this short promoter fragment targets α1AT expression to the liver with a greater specificity than that of larger α1AT promoter fragments.

2.5.1. Humoral Regulatory Elements

Cytokines have an important role in the up-regulation of α1AT expression in response to physiological events such as inflammation. In HepG2 cells, Os-M and IL-6 stimulate an ~3-fold increase in α1AT protein secretion (48). An OsM-responsive element maps between +11,536 and +11,670 bp, and it is distinct from an adjacent IL-6-responsive element between +11,167 and +11,536 bp. The OsM- and IL-6-responsive elements function synergistically in reporter constructs, both with each other and with sequences in the α1AT proximal promoter. Cells transfected with reporter constructs containing the IL-6- and OsM-responsive elements and a minimal α1AT proximal promoter fragment (-275 to -35 bp) yield an ~5-fold increase in reporter activity when stimulated with OsM and IL-6. However, the same construct shows an ~3-fold response when the cells are stimulated with either IL-6 or OsM alone (49). The 5′ proximal elements are necessary for any response to IL-6 or OsM.

The sequence of the IL-6 responsive element indicates that it binds Oct-1 (50). The OsM-responsive element may bind STAT3, as competition assays with anti-STAT3 antibodies diminish STAT3 binding to oligonucleotides containing the element.

2.6. Pros and Cons of Transgene Experiments

Transgene experiments are based primarily on assaying the activities of cloned putative regulatory elements, most commonly in the context of heterologous reporter genes. Such experiments are useful in that many different putative regulatory elements can be assayed in a variety of different cell types both quickly and efficiently. In this way, the functions of these elements *in situ* may be inferred. Moreover, such experiments may be invaluable in cases where the functions of endogenous regulatory elements are redundant. However, more complex gene regulatory phenomena, such as those that require interactions between regulatory sequences, or that involve regulation over large genomic distances, may not be apparent using this general approach. Clearly, elements that are components of regulatory networks cannot be effectively studied in isolation.

Another important issue is transgene structure. In recent years it has become apparent that the regulation of chromatin structure is an important, perhaps central, aspect of gene control. Furthermore, this regulation may affect regions of the genome that are hundreds, perhaps thousands, of kilobases in length. Thus, even when transgenes are integrated into the genome, their structure is unlikely to recapitulate that of the corresponding chromosomal alleles, and their expression is often subject to position effects and/or other artifacts that affect tandem transgene arrays (51). A solution to these problems would be to study regulatory elements *in situ*; that is, to prepare specific mutant alleles of the locus of interest in its normal chromosomal context. As the behavior of the wild-type α1AT locus, both in terms of gene expression and chromatin structure, has been well-defined after chromosome transfer (52), this constitutes a useful model system for the analysis of regulatory elements *in situ*. The DT40 chromosome shuttle system provides an efficient means to construct and analyze regulatory mutations in the serpin locus at 14q32.1 (53, 54).

3. THE ENDOGENOUS SERPIN LOCUS

3.1. The Wild-Type α1AT Allele

DNase I-hypersensitive sites (DHSs) often demarcate regions where proteins bind to DNA *in vivo*, and they are useful for mapping putative regulatory sequences for functional tests. There are 29 DHSs in the ~150 kb region around the α1AT gene in human HepG2 hepatoma cells, but only 7 of these DHSs are present in non-expressing HeLa cells (52). Several DHSs are coincident with sequences that have regulatory functions in cloned transgenes; for example, a DHS lies within the IL-6 and OsM-responsive element, and DHSs are found in both the proximal and distal α1AT promoters. Neighboring genes within the proximal serpin subcluster, CBG and ZPI, also display expression-associated DHSs in their promoter regions. Significantly, expression-associated DHSs within the serpin locus are formed *de novo* when wild-type human chromosome 14 is transferred from non-expressing to expressing cells, and serpin gene expression is activated (52). Thus, chromatin features of the α1AT locus serve as useful guides to search for putative regulatory elements; they also provide phenotypic readouts for specifically mutated chromosomal alleles.

3.2. The Chromosome Shuttle System

Precise changes can be made to specific chromosomal loci by homologous recombination. In most somatic cell types, the efficiency of homologous recombination between introduced DNA substrates and the corresponding chromosomal allele is so low as to be experimentally impractical. However, in transformed avian B-cells, such as the DT40 cell line, homologous recombination between chromosomal DNA and introduced DNA substrates can occur at efficiencies approaching 100% (55). Microcell hybrids created by transferring

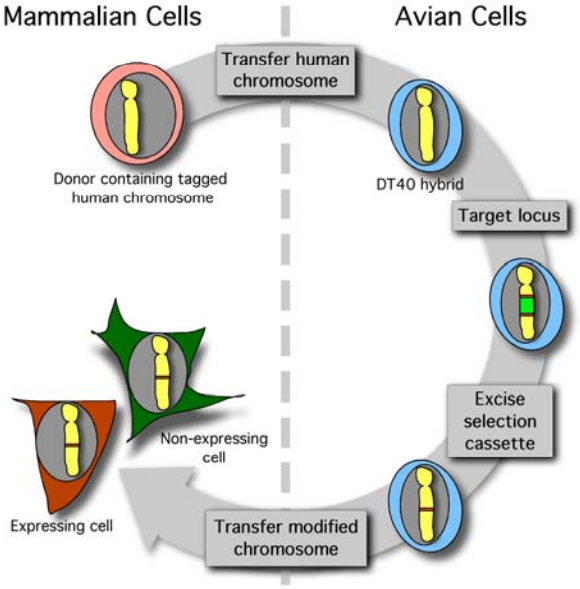

Figure 3. The DT40 chromosome shuttle scheme. A human chromosome is transferred from a mammalian donor cell line into the DT40 cell line by microcell fusion, creating a recombination-proficient chicken-human microcell hybrid. A specific locus on the human chromosome is targeted by transfecting the chicken-human microcell hybrid with a DNA substrate for homologous recombination. The substrate contains a selection cassette flanked by loxP sites and targets the human locus by homologous recombination. The selection cassette is excised from the modified human chromosome by CRE-mediated recombination. To assay the effects of the modifications, the modified human chromosome is transfered by microcell fusion to expressing and/or non-expressing mammalian cells.

single, specific human chromosomes to DT40 recipient cells provide a system in which human chromosomal loci can be modified efficiently by homologous recombination (Figure 3) (53, 54).

To make precise modifications in the human α1AT locus, human chromosome 14 was transferred to DT40 cells by microcell fusion (54), providing a permanent human/avian hybrid cell line that could be used in all future chromosome modification experiments. To create specific mutant alleles, the hybrid cells were transfected with substrates for homologous recombination, drug-resistant clones were isolated, and transfectants in which homologous recombination had occurred were identified (Figure 4). After removal of the selection cassette within the original recombination template, the modified chromosomes were then transferred to different mammalian cell types for functional tests.

3.3. First Generation Modification Experiments

There are six DHSs upstream of the α1AT hepatic transcription start site; two of these are in the proximal and distal promoters, respectively, and four expression-associated DHSs are clustered between -8 and -4 kb. These expression-associated DHSs are evolutionarily conserved, yet no function had been attributed previously to this ~4 kb region from transfection experiments. More recently, a series of cosmid transfections demonstrated that integrated α1AT transgenes containing ~14 kb of 5' flanking sequence were transcribed at wild-type levels at several ectopic insertion sites, while expression of α1AT in transfectants with only ~1.5 kb of 5′ flanking sequence was more variable (56). Nonetheless, both transgenes were expressed, indicating that upstream sequences were not required for α1AT expression, at least in this context. Furthermore, DHS formation within the transgene sequences resembled that of the expressing α1AT allele (56). From these experiments, one might conclude that the 4 kb region containing expression-associated DHSs upstream of α1AT was not necessary for gene expression. This conclusion is refuted by the results of chromosome modification experiments, as outlined below.

A mutant allele (Δ8.0) of α1AT was created in which an ~8 kb segment of endogenous sequence upstream of the α1AT proximal promoter (-8.37 to –0.32 kb) was deleted, as shown in Figure 4 (54). When this Δ8.0 mutant chromosome was transferred to hepatic cells by microcell fusion, human α1AT gene activation failed to occur (Figure 5). Moreover, activation of the CBG and ZPI genes, which are ~65 and ~93 kb downstream, also failed to occur in the Δ8.0 mutant chromosome. Furthermore, DHS mapping experiments indicated that expression-associated DHSs both upstream and downstream of the deletion failed to form on the Δ8.0 mutant chromosome. Thus, sequences in the -8.37 to –0.32 kb region are required both for α1AT, CBG, and ZPI gene activation and for locus-wide formation of expression-associated DHSs. These data define a serpin locus control region (LCR) that is necessary for gene activation and cell-specific chromatin remodeling of the entire proximal serpin subcluster, a region that is at least 150 kb in length. As noted above, the activity of the serpin LCR was not apparent in transgenesis experiments; the functions of this regulatory element were only apparent in its normal chromosomal context.

Subdeletions of the 8.0 kb region demonstrated that the serpin LCR was composed of muliple regulatory elements, and interactions between these elements, both in terms of gene activation and DHS formation, were complex (Figure 5). For example, the initial 8.0 kb region was divided into subsegments of 4.3 and 3.4 kb, corresponding to the distal and proximal halves of the original Δ8.0 deletion. The 3.4 kb (proximal) region did not contain DHSs, and cells containing the Δ3.4 mutant chromosome were wild-type with respect to both serpin gene activation and DHS formation. In contrast, the 4.3 kb (distal) region contained four DHSs, and the Δ4.3 mutant chromosome was refractory to α1AT, CBG, and ZPI gene activation, although not to the same extent as the original Δ8.0 mutation. Furthermore, DHS formation both upstream and downstream of the Δ4.3 deletion was abolished, although DHSs flanking the α1AT transcription unit were formed on the Δ4.3 (but not the Δ8.0) mutant chromosome (Figure 5). In a similar fashion, the 4.3 kb region was subdivided into distal (2.3 kb) and proximal (2.0 kb) halves, each containing two DHSs. The Δ2.0 mutant chromosome was

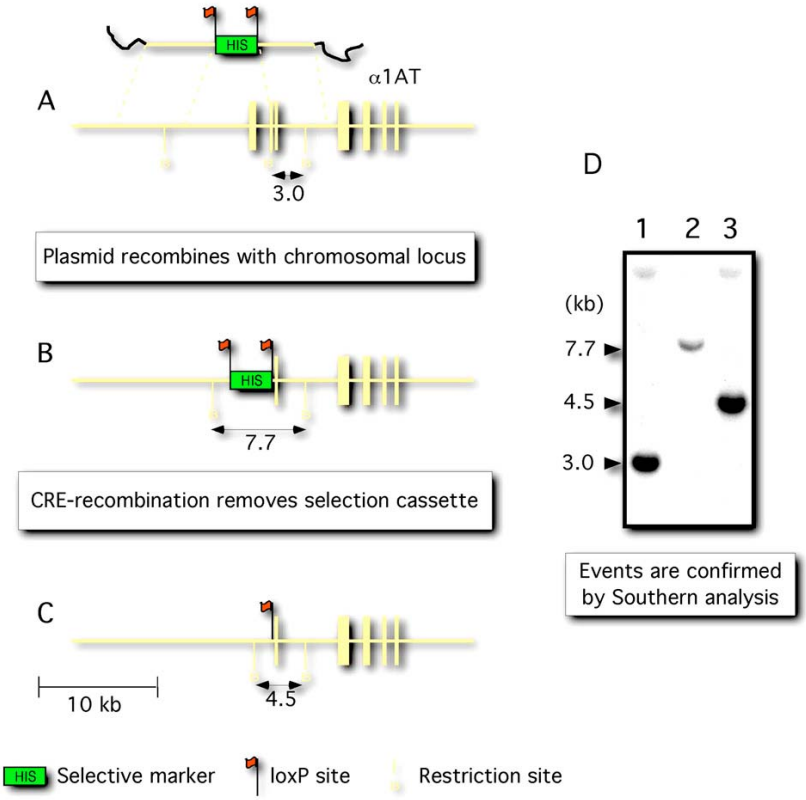

Figure 4. Generation of the Δ8.0 mutant chromosome by homologous recombination. A. The chromosomal (wild-type) α1AT allele and the homologous recombination template are shown. The recombination template contains a HisD selection cassette with *loxP* sites at each end, and the selection cassette is flanked by chromosomal α1AT DNA sequences just upstream and downstream of the 8 kb region to be deleted. B. Homologous recombination between the transfected DNA template and the chromosomal α1AT allele generates a mutant allele in which the 8 kb chromosomal DNA segment upstream of the α1AT gene has been replaced by the HisD selection cassette. C. Transient transfection of cells containing the modified allele shown in panel B with an expression plasmid encoding CRE recombinase results in site-specific recombination between *loxP* sites, resulting in precise excision of the HisD selection cassette. The resulting mutant allele is one in which the 8 kb DNA segment upstream of the α1AT gene has been replaced by a single 34 bp *loxP* site. D. Southern blot hybridization of genomic DNAs from cells containing wild-type (lane 1), homologously recombined (lane 2), and fully modified (lane 3) chromosomal α1AT alleles.

wild-type with respect to serpin gene activation and DHS formation, but the Δ2.3 chromosome displayed a mutant phenotype similar to that of the Δ4.3 deletion, although DHS formation in the far upstream region was different on the Δ4.3 and Δ2.3 chromosomes (Figure 5). These data provide clear evidence that the serpin LCR contains multiple regulatory elements, and these elements necessarily interact to encode the gene expression and chromatin structure phenotypes displayed by the wild-type serpin allele (54).

Given the profound effects that the serpin LCR has on both gene expression and chromatin structure in the proximal serpin subcluster, it is reasonable to ask why those effects were not apparent in transfection tests. We suggest that this reflects the operation of a hierarchy of regulatory controls, some of which affect the expression of individual serpin genes, and others that affect the activity and chromatin structure of the entire locus. For example, α1AT transgenes require only the cell-specific enhancer just upstream of the promoter for high-level, cell-specific expression, and other sequences, including the serpin LCR, have only modest effects on gene activity (42-44, 56). This is likely a consequence of the fact that α1AT transgenes in stable transfectants are necessarily integrated at transcriptionally permissive chromosomal sites, because they are invariably linked to a selected marker. In the normal chromosomal context, however, the α1AT gene cannot be expressed in the absence of the serpin LCR, presumably because the LCR is required for the formation of a transcriptionally competent chromatin state, a state that is marked by extensive DHS formation throughout the locus. According to this view, this permissive chromatin state is a prerequisite for activation of the α1AT enhancer by its cognate transactivators, HNF-1α and HNF-4. This suggests that a hierarchy of regulatory controls are involved, with the serpin LCR affecting the structure and activity of the entire locus, and gene-specific transcriptional activators acting in parallel and/or downstream. Experimental approaches like those described here should permit this complex regulatory circuitry to be resolved.

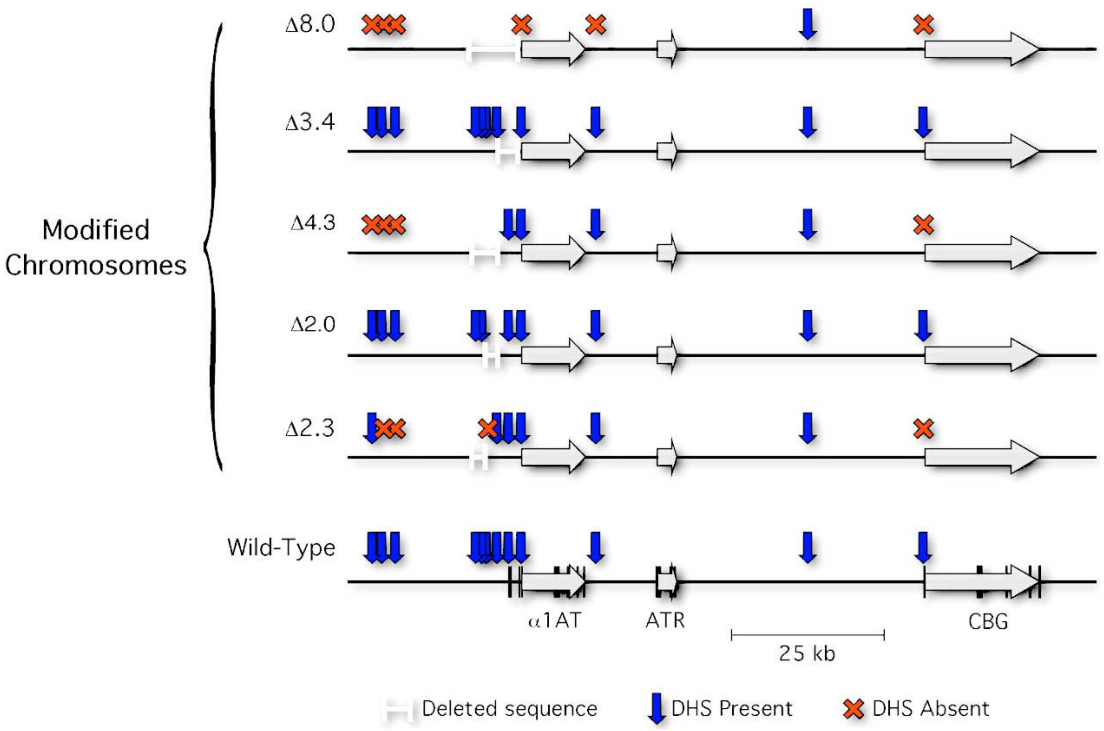

Figure 5. DHS formation on wild-type and mutant chromosomes. Twelve DHSs in the ~100 kb region from ~25 kb upstream of α1AT to the CBG promoter are shown. All of these sites were formed in hepatic cells that contained a wild-type copy of human chromosome 14 (bottom). In contrast, none of the eleven expression-associated DHSs in the region were formed on the Δ8.0 deletion chromosome; this mutant chromosome displayed only a single constitutive DHS between ATR and CBG. These data define a serpin locus control region (LCR) that resides within the deleted 8 kb region. The Δ4.3 and Δ2.3 subdeletions displayed partial mutant phenotypes with respect to DHS formation; the Δ3.4 and Δ2.0 deletions were wild-type. For a detailed discussion of these phenotypes, see the text.

4. CONCLUDING REMARKS AND PROSPECTS

As summarized in this report, we have learned much about the cell-specific regulation of individual genes in the 14q32.1 serpin cluster, and about the humoral control of serpin gene expression. However, our understanding of long-range regulation within the serpin locus is in its early stages. Clearly, the identification of a serpin locus control region in the proximal subcluster has been an important first step in identifying and analyzing locus-wide regulation of this region. The identification of the serpin LCR in chromosome modification experiments, but not in transfection tests, provides a convincing demonstration of the utility of the homologous recombination/chromosome transfer approach for studying the regulation of mammalian gene expression and chromatin structure.

The serpin LCR plays a dominant and dramatic role in activating expression of the proximal serpin subcluster, and it affects the activity and structure of a genomic region that is at least 150 kb. In this regard, its functions resemble those of locus control regions in the globin, immunoglobulin, and growth hormone loci. However, other activities of the serpin LCR appear unique. For example, genetic dissections of the serpin LCR clearly demonstrate that DHS formation in the proximal subcluster is not an all-or-none phenomenon; rather specific deletions of the serpin LCR affect the formation of discreet subsets of DHSs. These observations suggest that there exist subdomains within the proximal subcluster whose chromatin structure is regulated independently. This intriguing possibility can now be approached using a variety of experiment tests.

Much remains to be done. Although it appears at present that regulation of the proximal serpin subcluster is independent of the central and distal subclusters, little is known about regulation of the other two subclusters. Exploring this regulation will also require that we study the expression and organization of this region in a variety of cell types. Finally, it will be interesting to determine how aspects of serpin locus regulation have been conserved through evolution. Answering these questions is likely to make the serpin locus at 14q32.1 one of the paradigms of mammalian gene and chromatin control.

5. ACKNOWLEDGEMENTS

We thank Stephanie Namciu and Richard Friedman for critically reading the manuscript. Studies from the author's laboratory that are cited in this report were supported by grant GM26449 from the National Institute of General Medical Sciences.

6. REFERENCES

1. Silverman, G. A., P. I. Bird, R. W. Carrell, F. C. Church, P. B. Coughlin, P. G. Gettins, J. A. Irving, D. A. Lomas, C. J. Luke, R. W. Moyer, P. A. Pemberton, E. Remold-O'Donnell, G. S. Salvesen, J. Travis, and J. C. Whisstock: The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. Evolution, mechanism of inhibition, novel functions, and a revised nomenclature. *Journal of Biological Chemistry*. 276(36), 33293-6 (2001)

2. van Gent, D., P. Sharp, K. Morgan, and N. Kalsheker: Serpins: structure, function and molecular evolution. *International Journal of Biochemistry & Cell Biology*. 35(11), 1536-47 (2003)

3. Perlino, E., R. Cortese, and G. Ciliberto: The human alpha 1-antitrypsin gene is transcribed from two different promoters in macrophages and hepatocytes. *EMBO Journal*. 6(9), 2767-71 (1987)

4. Hafeez, W., G. Ciliberto, and D. H. Perlmutter: Constitutive and modulated expression of the human alpha 1 antitrypsin gene. Different transcriptional initiation sites used in three different cell types. *Journal of Clinical Investigation*. 89(4), 1214-22 (1992)

5. Rollini, P. and R. E. K. Fournier: Differential regulation of gene activity and chromatin structure within the human serpin gene cluster at 14q32.1 in macrophage microcell hybrids. *Nucleic Acids Research*. 28(8), 1767-77 (2000)

6. Gils, A. and P. J. Declerck: Structure-function relationships in serpins: current concepts and controversies. *Thrombosis & Haemostasis*. 80(4), 531-41 (1998)

7. Namciu, S. J., R. D. Friedman, M. D. Marsden, L. M. Sarausad, C. L. Jasoni, and R. E. K. Fournier: Sequence organization and matrix attachment regions of the human serine protease inhibitor gene cluster at 14q32.1. *Mammalian Genome*. 15, 162-178 (2004)

8. Clark, H. F., A. L. Gurney, E. Abaya, K. Baker, D. Baldwin, J. Brush, J. Chen, B. Chow, C. Chui, C. Crowley, B. Currell, B. Deuel, P. Dowd, D. Eaton, J. Foster, C. Grimaldi, Q. Gu, P. E. Hass, S. Heldens, A. Huang, H. S. Kim, L. Klimowski, Y. Jin, S. Johnson, J. Lee, L. Lewis, D. Liao, M. Mark, E. Robbie, C. Sanchez, J. Schoenfeld, S. Seshagiri, L. Simmons, J. Singh, V. Smith, J. Stinson, A. Vagts, R. Vandlen, C. Watanabe, D. Wieand, K. Woods, M. H. Xie, D. Yansura, S. Yi, G. Yu, J. Yuan, M. Zhang, Z. Zhang, A. Goddard, W. I. Wood, and P. Godowski: The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: a bioinformatics assessment. *Genome Research*. 13(10), 2265-70 (2003)

9. Strausberg, R. L., E. A. Feingold, L. H. Grouse, J. G. Derge, R. D. Klausner, F. S. Collins, L. Wagner, C. M. Shenmen, G. D. Schuler, S. F. Altschul, B. Zeeberg, K. H. Buetow, C. F. Schaefer, N. K. Bhat, R. F. Hopkins, H. Jordan, T. Moore, S. I. Max, J. Wang, F. Hsieh, L. Diatchenko, K. Marusina, A. A. Farmer, G. M. Rubin, L. Hong, M. Stapleton, M. B. Soares, M. F. Bonaldo, T. L. Casavant, T. E. Scheetz, M. J. Brownstein, T. B. Usdin, S. Toshiyuki, P. Carninci, C. Prange, S. S. Raha, N. A. Loquellano, G. J. Peters, R. D. Abramson, S. J. Mullahy, S. A. Bosak, P. J. McEwan, K. J. McKernan, J. A. Malek, P. H. Gunaratne, S. Richards, K. C. Worley, S. Hale, A. M. Garcia, L. J. Gay, S. W. Hulyk, D. K. Villalon, D. M. Muzny, E. J. Sodergren, X. Lu, R. A. Gibbs, J. Fahey, E. Helton, M. Ketteman, A. Madan, S. Rodrigues, A. Sanchez, M. Whiting, A. C. Young, Y. Shevchenko, G. G. Bouffard, R. W. Blakesley, J. W. Touchman, E. D. Green, M. C. Dickson, A. C. Rodriguez, J. Grimwood, J. Schmutz, R. M. Myers, Y. S. Butterfield, M. I. Krzywinski, U. Skalska, D. E. Smailus, A. Schnerch, J. E. Schein, S. J. Jones, M. A. Marra, and The Mammalian Gene Collection Program: Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences of the United States of America*. 99(26), 16899-903 (2002)

10. Frazer, J. K., D. G. Jackson, J. P. Gaillard, M. Lutter, Y. J. Liu, J. Banchereau, J. D. Capra, and V. Pascual: Identification of centerin: a novel human germinal center B cell-restricted serpin. *European Journal of Immunology*. 30(10), 3039-48 (2000)

11. Pan, Z., Y. Shen, C. Du, G. Zhou, A. Rosenwald, L. M. Staudt, T. C. Greiner, T. W. McKeithan, and W. C. Chan: Two newly characterized germinal center B-cell-associated genes, GCET1 and GCET2, have differential expression in normal and neoplastic B cells. *American Journal of Pathology*. 163(1), 135-44 (2003)

12. Han, X., Z. F. Huang, R. Fiehler, and G. J. Broze, Jr.: The protein Z-dependent protease inhibitor is a serpin. *Biochemistry*. 38(34), 11073-8 (1999)

13. Han, X., R. Fiehler, and G. J. Broze, Jr.: Characterization of the protein Z-dependent protease inhibitor. *Blood*. 96(9), 3049-55 (2000)

14. Laurell, M., A. Christensson, P. A. Abrahamsson, J. Stenflo, and H. Lilja: Protein C inhibitor in human body fluids. Seminal plasma is rich in inhibitor antigen deriving from cells throughout the male reproductive system. *Journal of Clinical Investigation*. 89(4), 1094-101 (1992)

15. Suzuki, K., J. Nishioka, H. Kusumoto, and S. Hashimoto: Mechanism of inhibition of activated protein C by protein C inhibitor. *Journal of Biochemistry*. 95(1), 187-95 (1984)

16. Meijers, J. C., D. H. Kanters, R. A. Vlooswijk, H. E. van Erp, M. Hessing, and B. N. Bouma: Inactivation of human plasma kallikrein and factor XIa by protein C inhibitor. *Biochemistry*. 27(12), 4231-7 (1988)

17. Rezaie, A. R., S. T. Cooper, F. C. Church, and C. T. Esmon: Protein C inhibitor is a potent inhibitor of the thrombin-thrombomodulin complex. *Journal of Biological Chemistry*. 270(43), 25336-9 (1995)

18. Hermans, J. M., R. Jones, and S. R. Stone: Rapid inhibition of the sperm protease acrosin by protein C inhibitor. *Biochemistry*. 33(18), 5440-4 (1994)

19. Meijers J. C. and D. W. Chung: Organization of the gene coding for human protein C inhibitor (plasminogen activator inhibitor-3). Assignment of the gene to chromosome 14. *Journal of Biological Chemistry*. 266(23), 15028-34 (1991)

20. Hayashi, T., M. Usui, J. Nishioka, Z. X. Zhang, and K. Suzuki: Regulation of the human protein C inhibitor gene expression in HepG2 cells: role of Sp1 and AP2. *Biochemical Journal*. 332(Pt 2), 573-82 (1998)

21. Chai, K. X., D. C. Ward, J. Chao, and L. Chao: Molecular cloning, sequence analysis, and chromosomal localization of the human protease inhibitor 4 (kallistatin) gene (PI4). *Genomics*. 23(2), 370-8 (1994)

22. Underhill D. A. and G. L. Hammond: Organization of the human corticosteroid binding globulin gene and analysis of its 5'-flanking region. *Molecular Endocrinology*. 3(9), 1448-54 (1989)
23. Underhill D. A. and G. L. Hammond: *Cis*-regulatory elements within the proximal promoter of the rat gene encoding corticosteroid-binding globulin. *Gene*. 162(2), 205-11 (1995)
24. Emptoz-Bonneton, A., J. C. Crave, H. LeJeune, C. Brebant, and M. Pugeat: Corticosteroid-binding globulin synthesis regulation by cytokines and glucocorticoids in human hepatoblastoma-derived (HepG2) cells.[see comment]. *Journal of Clinical Endocrinology & Metabolism*. 82(11), 3758-62 (1997)
25. Wolf, W. C., R. A. Harley, D. Sluce, L. Chao, and J. Chao: Localization and expression of tissue kallikrein and kallistatin in human blood vessels. *Journal of Histochemistry & Cytochemistry*. 47(2), 221-8 (1999)
26. Chao, J. J. N. Stallone, Y. M. Liang, L. M. Chen, D. Z. Wang, and L. Chao: Kallistatin is a potent new vasodilator. *Journal of Clinical Investigation*. 100(1), 11-7 (1997)
27. Miao, R. Q., J. Agata, L. Chao, and J. Chao: Kallistatin is a new inhibitor of angiogenesis and tumor growth. *Blood*. 100(9), 3245-52 (2002)
28. Travis, J., J. Bowen, and R. Baugh: Human alpha-1-antichymotrypsin: interaction with chymotrypsin-like proteinases. *Biochemistry*. 17(26), 5651-6 (1978)
29. Ma, J., A. Yee, H. B. Brewer, Jr., S. Das, and H. Potter: Amyloid-associated proteins alpha 1-antichymotrypsin and apolipoprotein E promote assembly of Alzheimer beta-protein into filaments.[see comment]. *Nature*. 372(6501), 92-4 (1994)
30. Potter, H., I. M. Wefes, and L. N. Nilsson: The inflammation-induced pathological chaperones ACT and apo-E are necessary catalysts of Alzheimer amyloid formation. *Neurobiology of Aging*. 22(6), 923-30 (2001)
31. Hwang, S. R., B. Steineckert, A. Kohn, M. Palkovits, and V. Y. Hook: Molecular studies define the primary structure of alpha1-antichymotrypsin (ACT) protease inhibitor in Alzheimer's disease brains. Comparison of act in hippocampus and liver. *Journal of Biological Chemistry*. 274(3), 1821-7 (1999)
32. Baumann, H. and J. Gauldie: The acute phase response. *Immunology Today*. 15(2), 74-80 (1994)
33. Kordula, T., R. E. Rydel, E. F. Brigham, F. Horn, P. C. Heinrich, and J. Travis: Oncostatin M and the interleukin-6 and soluble interleukin-6 receptor complex regulate alpha1-antichymotrypsin expression in human cortical astrocytes. *Journal of Biological Chemistry*. 273(7), 4112-8 (1998)
34. Kordula, T., M. Bugno, R. E. Rydel, and J. Travis: Mechanism of interleukin-1- and tumor necrosis factor alpha-dependent regulation of the alpha 1-antichymotrypsin gene in human astrocytes. *Journal of Neuroscience*. 20(20), 7510-6 (2000)
35. Knoell, D. L., D. R. Ralston, K. R. Coulter, and M. D. Wewers: Alpha 1-antitrypsin and protease complexation is induced by lipopolysaccharide, interleukin-1beta, and tumor necrosis factor-alpha in monocytes. *American Journal of Respiratory & Critical Care Medicine*. 157(1), 246-55 (1998)
36. Barbey-Morel, C., J. A. Pierce, E. J. Campbell, and D. H. Perlmutter: Lipopolysaccharide modulates the expression of alpha 1 proteinase inhibitor and other serine proteinase inhibitors in human monocytes and macrophages. *Journal of Experimental Medicine*. 166(4), 1041-54 (1987)
37. Boskovic, G. and S. S. Twining: Local control of alpha1-proteinase inhibitor levels: regulation of alpha1-proteinase inhibitor in the human cornea by growth factors and cytokines. *Biochimica et Biophysica Acta*. 1403(1), 37-46 (1998)
38. Kalsheker, N. and T. Swanson: Exclusion of an exon in monocyte alpha-1-antitrypsin mRNA after stimulation of U937 cells by interleukin-6. *Biochemical & Biophysical Research Communications*. 172(3), 1116-21 (1990)
39. Li, Y. and B. Y. Yue: Involvement of Sp1 elements in the promoter activity of the alpha1-proteinase inhibitor gene. *International Journal of Molecular Medicine*. 1(2), 339-46 (1998)
40. Ueda, J., Y. Li, M. S. Goh, Y. Maruyama, J. Sugar, and B. Y. Yue: Reporter gene construct containing 1.4 kb alpha 1-proteinase inhibitor promoter confers expression in the cornea of transgenic mice. *Anatomical Record*. 266(1), 5-9 (2002)
41. Maruyama, Y., Y. Li, Y. Zhang, X. Wang, J. Sugar, and B. Y. Yue: Mapping of Sp1 regulation sites in the promoter of the human alpha1-proteinase inhibitor gene. *Journal of Cellular Biochemistry*. 85(3), 482-9 (2002)
42. Ciliberto, G., L. Dente, and R. Cortese: Cell-specific expression of a transfected human alpha 1-antitrypsin gene. *Cell*. 41(2), 531-40 (1985)
43. De Simone, V., G. Ciliberto, E. Hardon, G. Paonessa, F. Palla, L. Lundberg, and R. Cortese: *Cis*- and *trans*-acting elements responsible for the cell-specific expression of the human alpha 1-antitrypsin gene. *EMBO Journal*. 6(9), 2759-66 (1987)
44. Shen, R. F., M. Tripodi, R. Cortese, and E. F. Wagner: Tissue-specific regulation of human alpha 1-antitrypsin gene expression in transgenic mice. *Nucleic Acids Research*. 15(18), 7519-29 (1987)
45. Tripodi, M., C. Abbott, N. Vivian, R. Cortese, R. Lovell-Badge, and A. Tripodi: Disruption of the LF-A1 and LF-B1 binding sites in the human alpha-1-antitrypsin gene has a differential effect during development in transgenic mice. *EMBO Journal*. 10(11), 3177-82 (1991)
46. Hu, C. and D. H. Perlmutter: Regulation of alpha1-antitrypsin gene expression in human intestinal epithelial cell line CaCo-2 by HNF-1alpha and HNF-4. *American Journal of Physiology*. 276(5 Pt 1), G1181-94 (1999)
47. Yull, F. E., R. M. Wallace, and A. J. Clark: Restricted tissue-specific but correct developmental expression mediated by a short human alpha1AT promoter fragment in transgenic mice. *Transgenic Research*. 4(1), 70-4 (1995)
48. Sallenave, J. M., G. M. Tremblay, J. Gauldie, and C. D. Richards: Oncostatin M, but not interleukin-6 or leukemia inhibitory factor, stimulates expression of alpha1-proteinase inhibitor in A549 human alveolar epithelial cells. *Journal of Interferon & Cytokine Research*. 17(6), 337-46 (1997)
49. Morgan, K., P. Marsters, S. Morley, D. van Gent, A. Hejazi, M. Backx, E. R. Thorpe, and N. Kalsheker: Oncostatin M induced alpha1-antitrypsin (AAT) gene expression in Hep G2 cells is mediated by a 3' enhancer. *Biochemical Journal*. 365(Pt 2), 555-60 (2002)

50. Morgan, K., G. Scobie, P. Marsters, and N. A. Kalsheker: Mutation in an alpha1-antitrypsin enhancer results in an interleukin-6 deficient acute-phase response due to loss of cooperativity between transcription factors. *Biochimica et Biophysica Acta*. 1362(1), 67-76 (1997)
51. Eszterhas, S. K., E. E. Bouhassira, D. I. Martin, and S. Fiering: Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Molecular & Cellular Biology*. 22(2), 469-79 (2002)
52. P. Rollini and R. E. K. Fournier: Long-range chromatin reorganization of the human serpin gene cluster at 14q32.1 accompanies gene activation and extinction in microcell hybrids. *Genomics*. 56(1), 22-30 (1999)
53. E. S. Dieken, E. M. Epner, S. Fiering, R. E. K. Fournier, and M. Groudine: Efficient modification of human chromosomal alleles using recombination-proficient chicken/human microcell hybrids. *Nature Genetics*. 12(2), 174-82 (1996)
54. M. D. Marsden and R. E. K. Fournier: Chromosomal elements regulate gene activity and chromatin structure of the human serpin gene cluster at 14q32.1. *Molecular & Cellular Biology*. 23(10), 3516-26 (2003)
55. Buerstedde, J. M. and S. Takeda: Increased ratio of targeted to random integration after transfection of chicken B cell lines. *Cell*. 67(1), 179-88 (1991)
56. P. Rollini, L. Xu, and R. E. K. Fournier: Stable expression and cell-specific chromatin structure of human alpha1-antitrypsin cosmid transgenes in rat hepatoma cells. *Nucleic Acids Research*. 28(18), 3605-14 (2000)

Key Words: gene regulation, chromatin structure, serpins, locus control region, homologous recombination

Send correspodence to: Dr R. E. K. Fournier, Division of Basic Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue N., Seattle, WA 98109, Tel: 206-667-5217, Fax:206-667-6522, E-mail: kfournie@fhcrc.org

http://www.bioscience.org/current/vol10.htm