A Polymorphism in the CTGF Promoter Region Associated with Systemic Sclerosis

Carmen Fonseca, M.D., Ph.D., Gisela E. Lindahl, Ph.D., Markella Ponticos, Ph.D., Piersante Sestini, M.D., Elisabetta A. Renzoni, M.D., Alan M. Holmes, Ph.D., Paolo Spagnolo, M.D., Panagiotis Pantelidis, Ph.D., Patrizia Leoni, Ph.D., Neil McHugh, M.D., Carmel J. Stock, B.Sc., Xu Shi-Wen, Ph.D., Christopher P. Denton, M.D., Ph.D., Carol M. Black, M.D., Kenneth I. Welsh, Ph.D., Roland M. du Bois, M.D., and David J. Abraham, Ph.D.

From the Centre for Rheumatology, Royal Free and University College Medical School, London (C.F., G.E.L., M.P., A.M.H., P.L., C.J.S., X.S.-W., C.P.D., C.M.B., D.J.A.); the Institute of Respiratory Diseases, University of Siena, Siena, Italy (P. Sestini); the Clinical Genomics Group, National Heart and Lung Institute, Imperial College London, London (E.A.R., P. Spagnolo, P.P., K.I.W., R.M.B.); and the Royal National Hospital for Rheumatic Diseases, Centre for Rheumatology, Bath, United Kingdom (N.M.). Address reprint requests to Dr. Abraham at the Department of Rheumatology, University College London, Hampstead Campus, Rowland Hill St., London NW3 2PF, United Kingdom, or at d.abraham@medsch.ucl.ac.uk.

Drs. Fonseca and Lindahl contributed equally to this article.


ABSTRACT

BACKGROUND
Systemic sclerosis (scleroderma) is a life-threatening autoimmune disease that is characterized by the presence of specific autoantibodies and fibrosis of the skin and major internal organs.

METHODS
We genotyped a polymorphism (G−945C) in the promoter of the connective-tissue growth factor (CTGF) gene in 1000 subjects in two groups: group 1, consisting of 200 patients with systemic sclerosis and 188 control subjects; and group 2, consisting of 300 patients with systemic sclerosis and 312 control subjects. The combined groups represented an estimated 10% of patients with systemic sclerosis in the United Kingdom. We tested the effect of the polymorphism on the transcription of CTGF.

RESULTS
The GG genotype was significantly more common in patients with systemic sclerosis than in control subjects in both groups, with an odds ratio for the combined group of 2.2 (95% confidence interval [CI], 1.5 to 3.2; P<0.001 for trend). Analysis of the combined group of patients with systemic sclerosis showed a significant association between homozygosity for the G allele and the presence of anti–topoisomerase I antibodies (odds ratio, 3.3; 95% CI, 2.0 to 5.6; P<0.001) and fibrosing alveolitis (odds ratio, 3.1; 95% CI, 1.9 to 5.0; P<0.001). We observed that the substitution of cytosine for guanine created a binding site of the transcriptional regulators Sp1 and Sp3. The C allele has high affinity for Sp3 and is associated with severely reduced transcriptional activity. A chromatin immunoprecipitation assay showed a marked shift in the ratio of Sp1 to Sp3 binding at this region, demonstrating functional relevance in vivo.

CONCLUSIONS
The G−945C substitution represses CTGF transcription, and the −945G allele is significantly associated with susceptibility to systemic sclerosis.
SYSTEMIC SCLEROSIS (SCLERODERMA) IS A clinically heterogeneous disorder of the connective tissue characterized by immune activation, vascular damage, and fibrosis. Major organ-based complications involving the lungs, heart, kidneys, and gastrointestinal tract determine mortality and morbidity. Moreover, the presence of highly specific and mutually exclusive autoantibodies is associated with distinct clinical phenotypes.

Clinical and experimental data suggest that the pathogenesis of systemic sclerosis is multifactorial, involving both genetic and environmental factors. Genetic susceptibility to systemic sclerosis is complex, involving several genes and chromosomal loci with small and probably additive individual effects.

The expression of the gene encoding connective-tissue growth factor (CTGF), also known as CCN2, has been found to be greatly up-regulated in gene-expression–profiling studies of skin-biopsy specimens and of fibroblasts cultured from skin and lung tissue (unpublished data) from patients with systemic sclerosis. Moreover, CTGF has been associated with key biologic functions, including fibroblast proliferation, the production of extracellular matrix, and the formation of adhesion and granulation tissue, processes that are relevant to systemic sclerosis. Indirect genetic evidence of the involvement of the CTGF gene in the pathogenesis of systemic sclerosis is provided in a recent study of monozygotic and dizygotic twins that showed a concordant increase in the expression of CTGF in monozygotic twins who were discordant for systemic sclerosis. In addition, genomewide screening of specimens obtained from Choctaw Indians of North America, who have a high incidence of systemic sclerosis, identified one region in particular, 6q23–27, to which a large number of genes, including CTGF, map. Taking all of these observations into account, we hypothesized that variations in the promoter region of CTGF confer susceptibility to systemic sclerosis.

METHODS

STUDY PATIENTS

We enrolled 500 white patients (426 women and 74 men) with systemic sclerosis, who fulfilled the preliminary criteria of the American College of Rheumatology for the disease, and 500 healthy white unrelated control subjects. Of these participants, 200 patients with systemic sclerosis and 188 control subjects were included in the initial screening group (group 1), and 300 patients with systemic sclerosis and 312 control subjects were assessed in the second group (group 2) (Table 1). All patients with systemic sclerosis were undergoing treatment at the Royal Free Hospital or the Royal Brompton Hospital in London or the Royal National Hospital for Rheumatic Diseases in Bath, and complete clinical and serologic information was available. The control samples were collected from healthy volunteers from the above-mentioned three hospitals and were unrelated to the patients with systemic sclerosis. Data on race and ethnic background were recorded from the patients’ medical records. For the control subjects, this information was self-reported as part of an open-ended questionnaire completed during the enrollment interview.

Systemic sclerosis was classified as either a limited subtype or a diffuse subtype according to the extent of the skin involved, as proposed by LeRoy et al. The presence of fibrosing alveolitis, scleroderma renal crisis, and isolated pulmonary

| Table 1. Clinical and Serologic Characteristics of Patients with Systemic Sclerosis. |
|-----------------|-----------------|-----------------|-----------------|
| Variable        | Total (N = 500) | Group 1 (N = 200) | Group 2 (N = 300) |
|                 | number (percent) | number (percent) | number (percent) |
| **Sex**         |                 |                 |                 |
| Female          | 426 (85.2)      | 166 (83.0)      | 260 (86.7)      |
| Male            | 74 (14.8)       | 34 (17.0)       | 40 (13.3)       |
| **Skin involvement** |               |                 |                 |
| Limited         | 364 (72.8)      | 137 (68.5)      | 227 (75.7)      |
| Diffuse         | 136 (27.2)      | 63 (31.5)       | 73 (24.3)       |
| **Organ involvement** |             |                 |                 |
| Fibrosing alveolitis | 207 (41.4)     | 87 (43.5)       | 120 (40.0)      |
| Isolated pulmonary arterial hypertension | 44 (8.8) | 22 (11.0) | 22 (7.3) |
| Scleroderma renal crisis | 52 (10.4) | 36 (18.0) | 16 (5.3) |
| No internal-organ involvement* | 224 (44.8) | 72 (36.0) | 152 (50.7) |
| **Antibodies**  |                 |                 |                 |
| Anti–topoisomerase I | 155 (31.0)   | 54 (27.0)       | 101 (33.7)      |
| Anticentromere  | 166 (33.2)      | 58 (29.0)       | 108 (36.0)      |
| No anti–topoisomerase I or anticentromere | 179 (35.8) | 88 (44.0) | 91 (30.3) |

* Patients in this category did not have fibrosing alveolitis, isolated pulmonary arterial hypertension, or scleroderma renal crisis.
arterial hypertension was defined according to internationally accepted criteria (see Methods in the Supplementary Appendix, available with the full text of this article at www.nejm.org). These clinically defined groups were not exclusive and consisted of 207 patients with fibrosing alveolitis, 44 with isolated pulmonary arterial hypertension, and 52 with scleroderma renal crisis (24 patients with fibrosing alveolitis with scleroderma renal crisis and 3 had isolated pulmonary arterial hypertension with scleroderma renal crisis); 224 patients had none of these complications. (The latter group was defined as having systemic sclerosis with no internal-organ involvement.)

Patients with systemic sclerosis were also classified into three nonoverlapping autoantibody subgroups: 155 patients had anti–topoisomerase I antibodies, 166 had anticentromere antibodies, and 179 patients had neither of these antibodies. (More detailed information about the group without either anti–topoisomerase I or anticentromere antibodies is provided in Table 1 of the Supplementary Appendix.) Autoantibody specificities were determined by standard methods that have been described previously. The frequency and type of organ involvement in both groups are shown in Table 1.

Written informed consent was obtained from all subjects, and the study was approved by the ethics committee at the Royal Free Hospital.

Identification of Sequence Variants

We compared sequences of the CTGF promoter region available in the National Center for Biotechnology Information’s GenBank database (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=nucleotide) and directly sequenced approximately 1 kb of sequence upstream from the transcription start site from 10 subjects.

The sequences across polymorphic sites (30 bp) were analyzed for putative nuclear-factor–binding sites by a search of transcription-factor–binding sites (TFBIND at http://tfbind.ims.u-tokyo.ac.jp). Several single-nucleotide polymorphisms (SNPs) of low frequency were detected and are now all deposited in GenBank’s dbSNP database. A common SNP, which was described previously as located at position −743 bp from the transcription start site, has since been included in the dbSNP database at position −945 from the ATG start codon (rs6918698). Because of this SNP’s high frequency and location in a putative transcriptional element, we further investigated it in our study, and we adhere to the official reference of its location at −945 in this article.

Genotyping

Genotyping was performed with the use of a polymerase-chain-reaction (PCR) assay with sequence-specific primers, as detailed in the Supplementary Appendix. The genotypes of 25 random samples were confirmed by direct sequencing by MGW Biotech.

Cell Culture

We isolated lung fibroblasts as described previously. The cells were cultured under standard conditions in Dulbecco’s modified Eagle’s medium containing 10% fetal-calf serum. For transient transfection experiments, the cells were placed in 0.2% fetal-calf serum for 10 hours and then treated with fresh 0.2% fetal-calf serum with or without 1 U of thrombin per milliliter for 16 hours. For the electrophoretic mobility-shift assay (EMSA) and chromatin immunoprecipitation assay, cells were harvested after 24 hours in 0.2% fetal-calf serum. For all experiments except the chromatin immunoprecipitation assay, we used the same preparation of cells, at passage 4 to 7, from one patient with systemic sclerosis. For the chromatin immunoprecipitation assay, cell preparations from 13 subjects were genotyped and four homozygotes were selected: two control lines, genotypes CC and GG; and two systemic sclerosis lines, genotypes CC and GG.

Plasmid Constructs and Promoter-Activity Assays

To generate the parent CTGF promoter–reporter construct, we cloned a 4.6-kb fragment of the CTGF gene from a P1-derived artificial chromosome clone (provided by Andrew Mungal, Sanger Institute, Cambridge, United Kingdom; GenBank accession number, AL354866) into the luciferase (Luc) reporter vector pGL3 (Promega) (see the Supplementary Appendix). This construct, which contains 4.5 kb of CTGF promoter sequence, was denoted as p-4.5–CTGF–Luc. We amplified a 1.6-kb fragment spanning the polymorphism from nine unrelated subjects who had previously been genotyped as homozygotes. After restriction enzyme digestion, which generates Pmel- and SacII-compatible sites at the ends, the resulting 1.5-kb fragments were cloned into the parent construct p-4.5–
CTGF–Luc. The 10 constructs were transiently transfected into human lung fibroblasts overnight with the use of Fugene 6 (Roche). After treatment, cells were lysed with the use of Reporter Gene Assay Lysis Buffer (Roche), and luciferase was measured with the use of luciferase reagents (Promega) in a chemiluminometer (Mithras LB 940, Berthold Technologies). A thymidine kinase promoter–renilla luciferase reporter plasmid (pRL-TK, Promega) was used as a transfection efficiency control. Triplicate wells were tested, and experiments were repeated at least twice. The parent construct, p-4.5–CTGF–Luc, was subjected to site-directed mutagenesis with the use of the QuikChange II XL kit (Stratagene) to make the substitution of guanine for cytosine at position −945. Primers used were forward ATTTATGCGCCACTCTCCCTCTTTGCTTGGCC and reverse GGCAAGGACACAGGGAGAGTGGCCATCAAT. The two constructs were used in transient cotransfection assays, together with pCMV-Sp3 (provided by Guntram Suske), or the empty vector control pCMV, as described above.

**DNA-BINDING ASSAYS**

For EMSA, we prepared nuclear extracts from human lung fibroblasts as described previously and used these extracts in binding reactions with 32P-labeled double-stranded oligonucleotides spanning the polymorphic site: top-strand C allele oligonucleotide CATTGATGCGCCACTCTCCCTCTTTGCTTGGCC and top-strand G allele oligonucleotide CATTGATGCGCCACTCTCGCTTTGCTTGGCC. We used standard conditions recommended by the kit manufacturer (Promega) and separated the reactions on a 4% polyacrylamide gel at 4°C and 160 V. Consensus oligonucleotides for Sp1 and EGR1 were purchased from Promega, and the following antibodies were used in supershift assays: Sp1 (sc-59X), Sp3 (sc-644X), and goat and rabbit IgG (Santa Cruz). A chromatin immunoprecipitation assay was performed with the use of a commercially available kit (Upstate), according to the manufacturer’s instructions (see the Supplementary Appendix).

**STATISTICAL ANALYSIS**

The genotype and allele frequencies were determined by direct counting. The genotype frequencies of the groups of patients and control subjects (in group 1, in group 2, and in the combined group, without consideration of serologic and clinical subgroups) were in Hardy–Weinberg equilibrium. The polymorphism was analyzed as an individual biallelic marker at a single locus. We used chi-square analysis, the Cochran–Armitage chi-square test for trend, and logistic-regression models to compare the distribution of the G–945C genotype among the groups of patients and control subjects.

Polytomous (multinomial) logistic-regression analyses were performed with the 500 control subjects as the base level and two possible outcomes of disease (with or without the specified feature) or more than two in the case of the antibodies, which were mutually exclusive. Polytomous logistic regression was also used in a case–case analysis to investigate the association between mutually exclusive serologic groups, with the group that was negative for both anti–topoisomerase I and anticientromere antibodies used as the control group. Since clinical features of the disease were not necessarily mutually exclusive, we analyzed their association with the G–945C genotype in a case–case approach with the use of logistic-regression models in which the genotype distribution in patients who presented with each clinical feature was compared with that in patients without that feature. Results are expressed as odds ratios and 95% confidence intervals. Statistical calculations were performed with the use of Stata statistical software (version 9).

The significance of the difference between genotype groups in promoter–reporter regions and chromatin immunoprecipitation assays was estimated by means of a two-tailed Student’s t-test (assuming equal variance) with the use of Excel (Microsoft).
pathogenesis of systemic sclerosis,22-24 made this SNP our prime candidate.

We screened our initial group of patients and control subjects for this polymorphism, and a significant difference in genotype distribution was found between the 200 patients with systemic sclerosis and the 188 control subjects (P=0.03), with an excess of the GG genotype in the disease group (in 27.5% of the patients vs. 18.0% of the control subjects, P=0.03). The association between the GG genotype and the diagnosis of systemic sclerosis was subsequently confirmed in an independent set (group 2), consisting of 300 patients with systemic sclerosis and 312 control subjects (difference in genotype distribution, P=0.001; proportion of G homozygotes, 32.3% in the group with systemic sclerosis vs. 19.9% in the control group; P<0.001) (Table 2). When the combined group of 500 patients with systemic sclerosis was compared with the combined group of 500 control subjects, the difference in the frequency of G homozygotes reached strong statistical significance (30.4% vs. 19.2%; odds ratio, 2.2; P<0.001 for trend) (Table 2). The genotype distribution of the G−945C polymorphism in all analyzed subgroups (group 1 and group 2) is shown in Table 2 of the Supplementary Appendix.

A significant trend toward an increase in the G allele copy number was observed in 155 patients with anti–topoisomerase I antibodies (odds ratio for CG, 1.3; odds ratio for GG, 3.3; P<0.001 for trend) and in 166 patients with anticentromere antibodies (odds ratio for CG, 1.7; odds ratio for GG, 2.7; P<0.001 for trend) but not in 179 patients with autoantibodies other than anti–topoisomerase I or anticentromere antibodies (odds ratio for CG, 1.1; odds ratio for GG, 1.3; P=0.35 for trend) (Table 2, and Table 4 of the Supplementary Appendix).

We further compared the association between the G−945C genotype and antibody status in a case–case analysis. There was a significant increase in the G allele copy number in samples that were positive for anti–topoisomerase I antibodies (P<0.005 by the Cochran–Armitage test) and for those that were positive for anticentromere antibodies (P<0.05) as compared with samples that were negative for both; no significant difference was found between samples that were positive for anti–topoisomerase I antibodies and those that were positive for anticentromere antibodies (P=0.24 by the Cochran–Armitage test).
The association between anti–topoisomerase I antibodies and fibrosing alveolitis in systemic sclerosis is well established\(^{25}\) and raises the question of whether the strong association between this polymorphism and anti–topoisomerase I antibodies also reflects an association with fibrosing alveolitis. Since clinical phenotypes of the disease were not necessarily mutually exclusive, we performed a logistic-regression analysis to compare the G−945C genotype between patients with and those without each clinical feature using a case–case approach. This analysis revealed a significant association between the G allele and the presence of fibrosing alveolitis when compared with patients without fibrosing alveolitis (odds ratio for CG, 1.3; odds ratio for GG, 1.8; \(P<0.05\)). Moreover, this association remained significant after adjustment for sex and autoantibody group (odds ratio for CG, 1.6; odds ratio for GG, 2.0; \(P<0.05\)).

When the association between the G allele copy number and the presence or absence of fibrosing alveolitis was analyzed separately in each autoantibody group (Fig. 1A), a significant association was observed in the group that was positive for anti–topoisomerase I antibodies (\(P<0.05\) by the Cochran–Armitage test), and a nonsignificant trend was observed in the small group of patients who were positive for ant centromere antibodies (\(P=0.11\)). However, no trend was found in the group without anti–topoisomerase I or ant centromere antibodies (\(P=0.48\)).

In a polytomous logistic-regression model with healthy control subjects as the base category and separate outcomes for patients with different autoantibodies and with or without fibrosing alveolitis, the combination of anti–topoisomerase I antibodies and fibrosing alveolitis was strongly associated with an increase in the G allele copy number (odds ratio for CG, 1.6; odds ratio for GG, 4.6; \(P<0.001\)). Similar results were ob-

![Figure 1](image:Figure.png)

**Figure 1.** The G−945C Polymorphism in CTGF as a Risk Factor for Fibrosing Alveolitis in Systemic Sclerosis.

Panel A shows the prevalence of fibrosing alveolitis according to the −945 genotype in CTGF in patients with systemic sclerosis who have either anti–topoisomerase I antibodies or ant centromere antibodies or lack both. T bars represent the standard error of proportions obtained with the use of generalized linear models. Panel B shows CTGF promoter–reporter constructs used in transfection assays, indicating the polymorphic site in relation to cloning sites and functional sites in the proximal promoter.\(^{23,26}\) Panel C shows the relative luciferase expression in human lung fibroblasts transfected with reporter constructs containing a promoter sequence cloned from 10 unrelated subjects, five G alleles and five C alleles. A thymidine kinase promoter–renilla luciferase vector was used as a transfection efficiency control. Shown are the means (+SE) of pooled data from three observations (triplicate cultures) for five constructs at basal conditions and after stimulation with 1 U of thrombin per milliliter; data are representative of two experiments. \(P<0.001\) for the G alleles as compared with the C alleles at basal conditions (luciferase activity increased by a factor of 3.3) and after thrombin stimulation (increased by a factor of 2.9). SMAD denotes similar to mothers against decapentaplegic homologue, BCE-1 basal control element 1, Luc promoter–luciferase, and RLU relative luciferase units.
served for anticentromere-positive systemic sclerosis either with fibrosing alveolitis (odds ratio for CG, 1.6; odds ratio for GG, 2.4; \( P=0.002 \)) or without fibrosing alveolitis (odds ratio for CG, 3.6; odds ratio for GG, 10.1; \( P=0.01 \)) but not in the groups without anti-topoisomerase I or antcentromere antibodies, regardless of the presence or absence of fibrosis. By contrast, after adjustment for autoantibody group, no significant association was found in patients with systemic sclerosis between the G allele copy number and the presence or absence of diffuse skin involvement, pulmonary hypertension, or renal crisis (Tables 3 and 4 of the Supplementary Appendix).

**PROMOTER-ACTIVITY ANALYSIS**

We investigated whether the polymorphism at position −945 affects the transcriptional activity of the CTGF promoter in primary pulmonary fibroblasts. We chose these cells because the lung is the most strongly implicated organ in our study, and fibroblasts are major producers of CTGF in the lung.\(^{27}\) We cloned 1.6-kb DNA fragments amplified by PCR spanning the polymorphic site from patients with defined CC or GG genotypes (five in each group) into a 4.5-kb promoter–DNA sample before immunoprecipitation, diluted by a factor of 10 and used as a positive control; IP immunoprecipitation; and Ab antibody. In Panel D, the mean (+SE) ratios of Sp1 to Sp3 PCR product, measured by densitometry, are shown for pooled data from five independent experiments. \( P<0.01 \) for the GG genotype as compared with the CC genotype in control cells (ratio increased by a factor of 5.8), and \( P<0.001 \) for the GG genotype as compared with the CC genotype in cells from patients with systemic sclerosis (increased by a factor of 3.3). Panel E shows that Sp3 acts as a repressor at −945C. Luciferase expression from a G allele construct, p-4.5–CTGF–Luc (p-4.5G), and the corresponding C allele construct (p-4.5C), derived by site-directed mutagenesis from p-4.5–CTGF–Luc, is cotransfected with an expression vector for Sp3 (pSp3) or an empty expression vector (pEV). Data shown are means (+SE) of three independent experiments. The asterisk denotes \( P<0.01 \) for the difference between the effect of pSp3 and pEV on p-4.5C under basal conditions (a factor of 0.65). The dagger denotes \( P<0.001 \) for the difference between the effect of pSp3 and pEV on p-4.5C after thrombin stimulation (a factor of 0.47). Panel F shows a proposed model of interaction of the polymorphic site with the proximal promoter in the CTGF gene. The balance of complex components and transcriptional activity — for example, through acetylation (Ac) and phosphorylation (P) — is probably affected by the loss of Sp3 binding at position −945. Histone deacetylase (HDAC) and p300, a histone acetylase, may form parts of this complex together with Sp1, Sp3, SMAD protein, and other factors.

**FORMATION OF PROTEIN–DNA COMPLEX**

We next assessed the interactions between protein and DNA at this site with the use of fibroblast nuclear extracts and EMSA. A striking difference between the G allele and C allele was observed, showing strong binding of nuclear proteins to the C allele, with two distinct complexes formed, but with a lack of binding to the G allele probe (Fig. 2A). Since the C variant forms part of a TCCTCCC element (polymorphic base underlined), an atypical Sp1-like binding site reported in the collagen type Iα1 gene,\(^{28}\) we investigated whether the complexes contained such factors. Specificity of binding, tested by competition with unlabeled oligonucleotides, showed a loss of binding with the use of a C allele probe and an Sp1 consensus oligonucleotide, but not with the G allele probe and a consensus probe for another GC-rich binding factor, early growth response 1 (EGR1) (Fig. 2A). The use of specific supershift antibodies showed that Sp3 is the major factor binding to this site (C allele), with weak binding also of Sp1 (Fig. 2B).

We performed a chromatin immunoprecipita-
A SNP in the CTGF Promoter Associated with Systemic Sclerosis

![Image](A)

**A**

- C G
- None
- C G
- G
- Sp1 G
- EGR1

Free probe

---

**B**

- None
- Sp1
- Sp3
- Sp3
- GG systemic sclerosis
- GG systemic sclerosis

Free C-allele probe

---

**C**

<table>
<thead>
<tr>
<th>Genotype</th>
<th>IP</th>
<th>Input (10%)</th>
<th>Negative Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC</td>
<td>Sp1</td>
<td>Sp3</td>
<td>CC IgG</td>
</tr>
<tr>
<td>GG</td>
<td>Sp1</td>
<td>Sp3</td>
<td>GG IgG IgG IgG IP Ab</td>
</tr>
</tbody>
</table>

![Image](C)

**D**

![Image](D)

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Ratio of Sp1 to Sp2</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC control</td>
<td>0.01</td>
</tr>
<tr>
<td>GG control</td>
<td>0.33</td>
</tr>
<tr>
<td>CC systemic sclerosis</td>
<td>1.59</td>
</tr>
<tr>
<td>GG systemic sclerosis</td>
<td>2.35</td>
</tr>
</tbody>
</table>

---

**E**

<table>
<thead>
<tr>
<th>Luciferase Activity (RLU)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>p-4.5C pEV</td>
</tr>
<tr>
<td>p-4.5C pSp3</td>
</tr>
<tr>
<td>p-4.5G pEV</td>
</tr>
<tr>
<td>p-4.5G pSp3</td>
</tr>
</tbody>
</table>

**F**

Repression by Sp3 in C allele

Repression released in G allele
tion assay using pulmonary fibroblasts that were homozygous for either cytosine or guanine at position −945. Chromatin was immunoprecipitated with the use of specific anti-Sp1 and anti-Sp3 antibodies, and DNA that was extracted from five independent experiments was amplified by CTGF-promoter–specific PCR spanning the polymorphic site. Binding of Sp3 and Sp1 was demonstrated in both CC and GG homozygote cells (Fig. 2C), a finding that might be explained by the presence of flanking Sp-factor sites (which was confirmed by sequencing). However, there was a strong and significant shift in the ratio of Sp1 binding to Sp3 binding toward higher ratios in GG homozygote cells (Fig. 2D). This shift was caused by the combination of a moderate loss in Sp3 binding and a dramatic gain in Sp1 binding (Fig. 2C). The allelic difference was observed in cells from both patients with systemic sclerosis and control subjects, with cells from patients with systemic sclerosis having a generally higher ratio of Sp1 to Sp3 binding than cells from control subjects.

SITE-DIRECTED MUTAGENESIS AND Sp3 OVEREXPRESSIO

Unlike Sp1, Sp3 has been reported to act as a transcriptional repressor or as an activator that is weaker than Sp1 but is competing for the same site. This finding would be consistent with our finding that enhanced transcriptional activity of the G allele promoter constructs correlated with a lack of Sp3 binding to this site. To test this hypothesis further, and to avoid influences by other putative sequence variants in the amplified fragments from different subjects, we substituted the G at this position with a C in the parent p-4.5–CTGF–Luc construct by site-directed mutagenesis and used these two constructs in cotransfection assays with an expression vector for Sp3 or an empty vector as the control. As predicted, Sp3 overexpression resulted in a significant repression of the C allele construct but had no repressive effect on the G construct (and sometimes had a slightly enhancing effect) (Fig. 2E).

DISCUSSION

Our data clearly show an association between the G–945C polymorphism in CTGF and systemic sclerosis, which renders CTGF a susceptibility gene for this complex disease. It appears that this SNP is associated with defined subgroups of patients within the disease, which may contribute to its typical heterogeneity. In particular, there is a strong association between G-allele homozygotes and the presence of specific antibodies (anti–topoisomerase I and anticentromere) associated with systemic sclerosis, as well as with patients with lung fibrosis, in whom the GG genotype represents a risk that is also independent of the antibody profile. Thus, homozygosity for the G allele carries an increased risk of scleroderma and, in particular, of positivity for anti–topoisomerase I antibodies and an increased risk of interstitial lung fibrosis among patients with this disorder. The association with anticentromere-antibody positivity, which has been linked to vascular complications, including isolated pulmonary arterial hypertension, may suggest that a second mechanism involving CTGF overexpression is playing a role here. The effects of CTGF on cell proliferation or extracellular matrix production, if induced within certain vessels, could be plausible explanations for this association.

The genetics of autoimmune diseases in general is complex. The fact that major-histocompatibility-complex alleles confer a risk of such disorders has been known for more than 30 years, and more recent genetic studies have underscored a large number of novel loci and biologic pathways. Our study adds to a small number of recent studies linking polymorphisms of candidate genes more strongly with the presence of specific autoantibodies than with systemic sclerosis itself. The findings include genetic associations between tumor necrosis factor α and anticentromere antibodies, between interleukin-10 and anti–RNA polymerase and anti–U1-ribonucleoprotein autoantibodies, between endothelin receptor A and anti–RNA polymerase autoantibodies, and between protein tyrosine phosphatase nonreceptor type 22 (PTPN22) and anti–topoisomerase I and anticentromere antibodies.

Although there is a well-recognized association between particular autoantibodies and the clinical profile in patients with systemic sclerosis, little is known about the cellular and molecular mechanisms leading to the production of these antinuclear antibodies or how they are linked to particular organ involvement. However, some recent reports have suggested mechanisms by which autoantibodies can cause tissue damage at a cellular level, providing the missing connection between antibody profile and organ involvement. Our study indicates that CTGF,
through its biologic effects (e.g., cell survival, differentiation, and proliferation\(^{26}\)) may be contributing to the pathogenesis of systemic sclerosis in a manner specific to either organs or cells. In this way, dysregulation of CTGF could directly contribute to the processes causing organ-specific damage associated with autoimmunity in systemic sclerosis. A possible alternative explanation for the associations we observed is that the effect of the CTGF polymorphism does not contribute to disease manifestations unless there is underlying tissue damage of a certain type (e.g., that associated with the presence of anti-topoisomerase I antibodies). Further detailed functional genetic studies will be required to address these issues. A third possibility is that the SNP is in linkage disequilibrium with another functional variant.

We have shown that the presence of a C allele at position −945 is critical for transcriptional suppression of the CTGF gene and that such suppression is mediated through Sp3 binding. This suppression would result in reduced CTGF production and potential protection against CTGF overexpression in disorders such as systemic sclerosis, in which a persistent wound-healing response appears to be an underlying pathogenic process. Such a functional explanation would be in agreement with the observation that the G allele is associated with systemic sclerosis, in which CTGF overexpression is generally observed. Thus, the functional experiments we describe not only unequivocally demonstrate differences in the transcriptional activity of the C and G alleles and differential DNA binding in vitro and in vivo but also identify a novel major repressor site in the CTGF promoter (the C allele).

Our chromatin immunoprecipitation assays showed that fibroblasts from patients with systemic sclerosis had an overall ratio of Sp1 to Sp3 binding that was higher than that of cells from control subjects. On the basis of previous reports that Sp1 often acts as a stronger activator than Sp3 and that Sp3 also often acts as a repressor (depending on the binding site and conditions),\(^{39}\) the increased shift in the ratio of Sp1 to Sp3 probably would alter the rate of transcription and govern higher activity. This hypothesis is consistent with an overall increase in the transcriptional activity of CTGF in fibroblasts from patients with systemic sclerosis, as we have shown previously,\(^{23}\) and with enhanced Sp1 binding activity.\(^{22,24}\)

From a genetic point of view, our chromatin immunoprecipitation assays showed that the ratios of Sp1 to Sp3 binding within each cell type were approximately three times as high for GG homozygote cells as for CC homozygotes, a finding that was consistent with higher transcriptional activity from the G alleles. The ratios of Sp1 to Sp3 binding in the fibroblasts from patients with systemic sclerosis with a CC genotype fell within the higher range in cells from control subjects with the GG genotype (Fig. 2C). This finding suggests that the effect of the polymorphism is superimposed on an underlying imbalance in the Sp1–Sp3 system in lung fibroblasts from patients with systemic sclerosis.\(^{22}\)

On the basis of our findings, we propose a model in which the absence of a strong Sp3 site within the context of the promoter, as in the G allele, could change the balance of the basic transcription complex toward enhanced transcription (e.g., through altered acetylation or phosphorylation) (Fig. 2F). From an evolutionary perspective, sequence comparison with the CTGF orthologues of other species suggests that the C allele, the more prevalent allele in humans, has evolved recently (see the figure in the Supplementary Appendix). Although this finding requires confirmation through more extensive study, it implies that there is strong positive selection and the possibility of advantages for this repressor site in humans.

In summary, our study shows the direct genetic association between CTGF and systemic sclerosis and identifies a CTGF risk genotype involving a polymorphism within a novel Sp3-dependent transcriptional repressor site. The G allele is strongly linked to transcriptional activation of CTGF and to the risk of systemic sclerosis, particularly in patients with anti–topoisomerase I autoantibodies and pulmonary fibrosis. Thus, our findings not only identify a novel regulatory mechanism governing CTGF promoter activity but also directly place CTGF on the genetic map of human disease. These data provide new insight into the pathogenesis of systemic sclerosis, including clues to the mechanisms leading to specific disease subtypes. Moreover, they may also be relevant to mechanisms underlying a wide range of other human disorders with a fibrotic component.

Supported by the Arthritis Research Campaign, the Raynaud’s and Scleroderma Association, the British Heart Foundation, the Asmarley Trust, the Scleroderma Society, and the Welton Foundation and by a grant (to Dr. Fonseca) from Instituto Nacional Enfermedades Respiratorias, Mexico.
REFERENCES


Copyright © 2007 Massachusetts Medical Society.