Polymorphisms in the osteopontin promoter affect its transcriptional activity

Francesca Giacopelli,1 Renato Marciano,1 Angela Pistorio,2 Paolo Catarsi,3 Silvana Canini,4 Gerard Karsenty,5 and Roberto Ravazzolo1,6

1Laboratory of Molecular Genetics and 2Epidemiology and Statistics Unit, 3Laboratory of Nephrology, and 4Laboratory of Chemical-Clinical Analysis, G. Gaslini Institute, Genova, Italy; 5Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; and 6Department of Pediatrics and Center of Excellence for Biomedical Research, University of Genova, Genova, Italy

Submitted 11 June 2004; accepted in final form 7 October 2004

Giacopelli, Francesca, Renato Marciano, Angela Pistorio, Paolo Catarsi, Silvana Canini, Gerard Karsenty, and Roberto Ravazzolo. Polymorphisms in the osteopontin promoter affect its transcriptional activity. Physiol Genomics 20: 87–96, 2004. First published October 12, 2004; doi:10.1152/physiolgenomics.00138.2004.—Understanding the molecular mechanisms that underlie regulation of transcription of the human osteopontin encoding gene (OPN) may help to clarify several processes, such as fibrotic evolution of organ damage, tumorigenesis and metastasis, and immune response, in which OPN overexpression is observed. With the aim to evaluate variants with functional effect on transcription, we have analyzed the promoter region and focused our investigation on three common variants present in the first 500 bp upstream of the transcription start site. Transfection of constructs carrying the four most frequent haplotypes relative to variants at −66, −156, and −443 fused to the luciferase reporter gene in a panel of different cell lines showed that one haplotype conferred a significantly reduced level of reporter gene expression in all tested cell lines. We describe that the −66 polymorphism modifies the binding affinity for the SP1/SP3 transcription factors, the −156 polymorphism is included in a yet uncharacterized RUNX2 binding site, and the −443 polymorphism causes differential binding of an unknown factor. The finding of differential effects of various combination of variants in haplotypes may contribute to explain data of association studies reported in several already published articles. Future association studies using haplotypes instead of single OPN variants will allow to achieve more accurate results referable to differential expression of OPN in several common diseases, in which OPN is considered a candidate susceptibility gene.

haplotype; gene regulation; Sp1; RUNX2

OSTEOPONTIN (OPN) is a phosphorylated acidic glycoprotein that displays several functions in different physiological and pathological processes, including bone remodeling, cell-mediated immunity, maintenance or reconfiguration of tissue integrity during inflammatory processes, coronary restenosis, and tumor cell metastasis. The protein contains an Arg-Gly-Asp (RGD) motif, is found both as an immobilized extracellular matrix molecule in mineralized tissues and as a cytokine in body fluids, is able to interact with various receptors, including αvβ3 and other integrins, and it may also be a ligand for certain variant forms of CD44. These interactions induce cellular signaling pathways, allowing OPN to mediate cell-matrix and cell-cell interactions (6, 40).

Overexpression of OPN has been described in several conditions in which basic inflammatory processes are activated, such as arthritis (24, 27, 43), myocardial remodeling after infarction (37), kidney interstitial fibrosis after obstructive urethropy and other renal insults (42), and wound healing (21, 29).

The OPN encoding gene [secreted phosphoprotein 1 (SPP1), MIM #166490] is mapped on human chromosome 4q21-q25, together with other members of the so-called SIBLING family of proteins, bone sialoprotein and dentin matrix protein-1, which share some structural characteristics (10). The OPN gene gives rise to different mRNA transcripts because of alternative spliced isoforms (30), and, moreover, numerous OPN isoforms are also due to posttranslational processing, such as phosphorylation and glycosylation and proteolytic cleavage by thrombin and matrix metalloproteinases (1, 33, 41).

OPN knockout mice have provided several pieces of information about OPN function. The first report of targeted inactivation of OPN described impaired wound healing (21). Several other articles followed, describing reduced macrophage infiltration and interstitial fibrosis in the kidney in mouse models of renal fibrosis (26, 29); reduced loss of bone mineral in ovarectomized null mice (45); resistance to progression in an experimental mouse model of multiple sclerosis (4, 18); impaired type-1 immunity to viral and bacterial infection and granulomatous response (3); abnormal tissue remodeling after myocardial infarction (38); and increased arterial calcification (34). On the other hand, in OPN-overexpressing transgenic mice, OPN plays an important role in the development of vascular medial thickening without injury and in neointima formation after arterial injury in vivo (16).

Genetic variations in the OPN gene have been described, and associations of these polymorphisms with lupus erythematosus, multiple sclerosis, urolithiasis, primary biliary cirrhosis, and autoimmune lymphoproliferative syndrome (ALPS) have been reported (5, 11, 19, 23, 44). In previous work we have described a polymorphic variant in the first intron of human OPN (12) that, although devoid of functional significance, might be a marker for another or other linked functional sites.

Here we present results about the functional characterization of other polymorphisms in a 474-bp region upstream of the OPN transcription start site, estimation of linkage disequilibrium (LD) between all polymorphic sites, and haplotype frequencies.

MATERIALS AND METHODS

Analysis of polymorphisms in OPN regulatory regions. Genomic DNA was extracted by standard procedure from peripheral blood of unrelated individuals of the same geographical origin (Italian), collected after informed consent.
Allele T/H11002 shift assay are underlined. sequences of oligonucleotides used in gel polymorphism of 500 bp of the OPN promoter: location of R 11 EX reverse

\[ \text{OPN8 reverse} \]
\[ \text{OPN6 reverse} \]
\[ \text{OP2 reverse} \]
\[ \text{C reverse} \]
\[ \text{B reverse} \]
\[ \text{A reverse} \]
\[ \text{OP1 reverse} \]

previously described (12). transcription start site (arrow, are shown (not in scale) relative to the tran-

moter.

\[ \text{OP1 forward} \]
\[ \text{OP1 reverse} \]
\[ \text{A forward} \]
\[ \text{A reverse} \]
\[ \text{B forward} \]
\[ \text{B reverse} \]
\[ \text{C forward} \]
\[ \text{C reverse} \]

532 Forward

\[ \text{OP2 reverse} \]
\[ \text{OPN} \text{ forward} \]
\[ \text{OPN} \text{ 6 forward} \]
\[ \text{OP2 reverse} \]
\[ \text{OP1 forward} \]

Primers for PCR amplification and sequencing of regulatory region of the OPN gene Table 1.

The entire regulatory region from −2236 to +1200 was amplified by PCR with paired primers and sequenced with both forward and reverse primers; oligonucleotides used for PCR amplification and sequencing and PCR annealing temperature are reported in Table 1. PCR was performed using 50 ng DNA as a template under the following conditions: 94°C for 7 min, then 30 cycles of 94°C for 55 s, annealing temperature for 55 s and 72°C for 90 s, final extension at 72°C for 10 min. After affinity membrane purification, the PCR products were subjected to cycle sequencing with the respective forward and reverse primers. Sequence analysis was performed using an automated ABI PRISM 3100 DNA sequencer.

The intronic polymorphism at 245 (TG/TGTG) was analyzed as previously described (12).

Fig. 1. Location of single nucleotide polymorphisms (SNPs) in the human osteopontin (OPN) gene regulatory region from −2236 to +1200 and sequence of the OPN promoter. A: SNPs and their respective location are shown (not in scale) relative to the transcription start site (arrow, +1). B: sequence of 500 bp of the OPN promoter: location of polymorphism −443 C/T, G/GG −156, and T/G −66 is shown with nucleotide in bold; sequences of oligonucleotides used in gel shift assay are underlined.

Table 1. Primers for PCR amplification and sequencing of regulatory region of the OPN gene

<table>
<thead>
<tr>
<th>Name of Primer</th>
<th>Primer Sequence</th>
<th>Annealing Temperature, °C</th>
<th>Amplified Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP1 forward</td>
<td>5’TGGGACTAGAAGGTTTCCAATAGGAG 3’</td>
<td>60</td>
<td>508 bp</td>
</tr>
<tr>
<td>OP1 reverse</td>
<td>5’TCAATTATGTAGGAGGCTGCCG 3’</td>
<td>60</td>
<td>542 bp</td>
</tr>
<tr>
<td>A forward</td>
<td>5’GTGGGACAGAAGTAAACTGAC 3’</td>
<td>60</td>
<td>555 bp</td>
</tr>
<tr>
<td>A reverse</td>
<td>5’CAGTGTGGAATGAGATTTGAC 3’</td>
<td>60</td>
<td>498 bp</td>
</tr>
<tr>
<td>B forward</td>
<td>5’CAGAATTTCCTTGGACGACAAAGCGG 3’</td>
<td>60</td>
<td>607 bp</td>
</tr>
<tr>
<td>B reverse</td>
<td>5’GGAAATGTCGCTGGCAATTTAGGACCG 3’</td>
<td>55</td>
<td>648 bp</td>
</tr>
<tr>
<td>C forward</td>
<td>5’ATTAGGCAAGTTCCGAGTGGAGG 3’</td>
<td>55</td>
<td>358 bp</td>
</tr>
<tr>
<td>C reverse</td>
<td>5’GGTAGCCGAGGAGTGCTCGG 3’</td>
<td>55</td>
<td>626 bp</td>
</tr>
<tr>
<td>532 Forward</td>
<td>5’TGTGGGAACTTCACTTGG 3’</td>
<td>55</td>
<td>626 bp</td>
</tr>
<tr>
<td>OP2 reverse</td>
<td>5’TGTGACCCATCGTGGGCTTGGT 3’</td>
<td>55</td>
<td>607 bp</td>
</tr>
<tr>
<td>OPN1 forward</td>
<td>5’TAAATTTCGAGGCTGTTGGG 3’</td>
<td>55</td>
<td>648 bp</td>
</tr>
<tr>
<td>OPN6 reverse</td>
<td>5’GCCTACAGAGTGTCACTACAG 3’</td>
<td>55</td>
<td>358 bp</td>
</tr>
<tr>
<td>OPN5 forward</td>
<td>5’ ACCGGAAACGAGCCAGGATCTA 3’</td>
<td>55</td>
<td>626 bp</td>
</tr>
<tr>
<td>OPN8 reverse</td>
<td>5’TGTGGGAACTTCACTTGG 3’</td>
<td>55</td>
<td>607 bp</td>
</tr>
<tr>
<td>F1309 forward</td>
<td>5’TGTGGGAACTTCACTTGG 3’</td>
<td>55</td>
<td>607 bp</td>
</tr>
<tr>
<td>R 11 EX reverse</td>
<td>5’TGTGGGAACTTCACTTGG 3’</td>
<td>55</td>
<td>607 bp</td>
</tr>
</tbody>
</table>

Oligonucleotides Used for Gel Shift Assay

<table>
<thead>
<tr>
<th>Allele</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T −66</td>
<td>5’CTCATGAGAAGTAAACTCCTGACGCTCCCTTGGTGTTGGT’3’</td>
</tr>
<tr>
<td>G −66</td>
<td>5’CTCATGAGAAGTAAACTCCTGACGCTCCCTTGGTGTTGGT’3’</td>
</tr>
<tr>
<td>G −156</td>
<td>5’TGTGACCCATCGTGGGCTTGGT 3’</td>
</tr>
<tr>
<td>G −156</td>
<td>5’TGTGACCCATCGTGGGCTTGGT 3’</td>
</tr>
<tr>
<td>C −443</td>
<td>5’AAGGACAGAGGAGGAGTTCCGAGCCTTGGGAGG 3’</td>
</tr>
<tr>
<td>T −443</td>
<td>5’AAGGACAGAGGAGGAGTTCCGAGCCTTGGGAGG 3’</td>
</tr>
</tbody>
</table>

OPN, osteopontin. *Upper strand is reported, with bold letters indicating nucleotides that differ in the two alleles of each variant

Haplotype reconstruction and statistical analysis. Pairwise LD was measured by D’ using the Arlequin software ver. 2.000 (32), by which the Hardy-Weinberg equilibrium was also tested. Haplotypes definition and frequency estimation were obtained by using the computer program PHASE 2.02 (35, 36). Estimates were obtained using 1,000 iterations, 10 thinning intervals, and 1,000 burn-in cycles.

Descriptive statistics were reported either in text or in figures in terms of medians and first and third quartiles for quantitative data. As the number of observations was quite low, all nonparametric tests were used to perform statistical analysis. Mann-Whitney U test was used to compare parameters in the two groups of data (for example: luciferase activity B haplotype transfected with Runx2 vs. control vector lacking the Runx2 cDNA). The one-way nonparametric anal-
analysis of variance (Kruskal-Wallis test) was used to compare parameters in more than two groups of data, and the Newman-Keuls test was applied for multiple comparisons to explore post-hoc differences between pairs of experiments; whenever the sample size of the experiments was not the same, the Dunn test was performed. All statistical tests were two-sided; a P value of less than 0.05 was considered statistically significant; the statistical package “Statistica for Windows” (release 6, StatSoft) was used to perform all the analyses.

Plasmid construction. Fragments from −1206 to +42 were prepared by PCR amplification from genomic DNA (sense primer 1206 OPN and antisense primer 2304 R as in Ref. 12) of homoyzogous individuals for the four most common haplotypes and cloned in pCR II Topo TA vector (Invitrogen). These plasmids were then digested with HindIII, which cuts at −471 in the SSP1 promoter and in multiple cloning site of pCR II Topo vector. Fragments obtained were cloned in the pGL3 Basic Vector (Promega), giving rise to A, B, C, and D plasmids which contain fragments of different alleles of the human OPN gene from −474 to +42. Site-specific mutagenesis was performed according to a PCR-based procedure (13) to obtain E and F plasmid constructs, using C and D plasmids as templates.

The pCMV-Oct2 plasmid, containing the full-length mouse Runx2 cDNA (9), and pCMV-SPl plasmid, containing the Sp1 cDNA (46) under the control of the cytomegalovirus (CMV) promoter, were kindly provided by Drs. P. Ducy and F. Ramirez, respectively. All the plasmids were carefully verified by restriction analysis and complete DNA sequencing.

Cell culture and transfection. All cell lines were maintained in Dulbecco’s modified Eagle’s medium (DMEM) except for HeLa, which was maintained in minimum essential medium, supplemented with 10% (vol/vol) fetal bovine serum (GIBCO-BRL), 2 mM l-glutamine, 100 U/ml penicillin, and 100 μg/ml streptomycin at 37°C in a humidified atmosphere with 5% CO₂. Nonessential amino acids (Euroclone) were included for HEK293 and HeLa cells. Human proximal tubular epithelial cells (HK2) were cultured in DMEM/F-12 containing 5 μg/ml insulin, 5 μg/ml transferrin, 5 ng/ml selenium, 5 ng/ml hydrocortisone, 5 μg/ml 3,5,3’-triiodothyronine, 5 μg/ml prostaglandin E₁, 10 ng/ml epidermal growth factor, 5% (vol/vol) fetal bovine serum, 2 mM l-glutamine, and 100 U/ml penicillin and streptomycin.

All cell lines were available in the laboratory and originally obtained from the American Type Culture Collection.

Transfections were performed using the polyethylenimine (PEI) cationic polymer, as previously described (25), except for HeLa cells which were transfected using Lipofectamine 2000 reagent (Invitrogen) according to manufacturer’s instructions.

Cells were plated into six-well dishes at 50–60% confluence and transfected at an estimated 80–90% confluence after 24 h. The amount of DNA used for transfection was 2.1 μg for each haplotype plasmid, 3.0 μg for each expression vector, and 0.1 μg for Renilla luciferase control construct pRL-CMV (Promega).

Cell lysates were prepared with the Luciferase Assay System Kit (Promega), and luciferase activity was determined with a luminometer (Turner BioSystems).

The data obtained represent median values within first and third quartile of at least three independent experiments, each performed either in triplicate or in duplicate. Luciferase activity was calculated as normalized activity by comparison to a cotransfected Renilla luciferase control construct pRL-CMV (Promega). All data were also confirmed using at least two different plasmid preparations.

Preparation of nuclear extracts and electrophoretic mobility shift assays. Nuclear extracts were prepared as described previously (7) from U2OS, COS7, and HeLa cell lines, and protein content was determined using a Bradford protein assay kit (Bio-Rad). The probes used for electrophoretic mobility shift assay (EMSA) are listed in Table 1 (upper strand, nucleotide variant in bold). The probes were incubated with 5–10 μg of nuclear extract for 20 min at room temperature in the presence of 1× gel shift binding buffer containing 10 mM HEPES, pH 7.9, 0.1 mM EDTA, 50 mM KCl, 0.5 mM dithiothreitol, 10% glycerol, 1 μg of poly(dI-dC)/poly(dI-dC), and 1× protease inhibitor cocktail (Roche Molecular Biochemicals). Competition in EMSA was performed with either 50-, 150-, or 250-fold molar excess of unlabeled oligonucleotide. For the supershift assay, Sp1, Sp3 (Santa Cruz Biotech), and Runx2 antibodies (8) were included in the incubation mixture for 20 min in ice. Complexes were separated in a 5% nondenaturing polyacrylamide gel. After electrophoresis, the gel was dried on Whatman 3MM paper and subjected to autoradiography.

RESULTS

SNP identification and haplotype reconstruction. In a previous article, we described a polymorphism located at +245 with respect to the transcription start site in the first intron of the OPN gene, for which we did not find a direct functional role (12). To verify whether other variants, in linkage disequilibrium with the above polymorphism, might play some role on

### Table 2. Alleles at 5 loci from −443 to +2080 and respective frequencies

<table>
<thead>
<tr>
<th>Region</th>
<th>Position</th>
<th>Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoter</td>
<td>−443, C &gt; T</td>
<td>C(0.5)/T(0.5)</td>
</tr>
<tr>
<td>Promoter</td>
<td>−156, G &gt; GG</td>
<td>G(0.65)/GG(0.35)</td>
</tr>
<tr>
<td>Intron 1</td>
<td>+245, TG &gt; GTGG</td>
<td>TGG(0.78)/TGG(0.22)</td>
</tr>
<tr>
<td>Intron 2</td>
<td>+2080, C &gt; T</td>
<td>C(0.5)/T(0.5)</td>
</tr>
</tbody>
</table>

### Table 3. Linkage disequilibrium between pairs of variants

<table>
<thead>
<tr>
<th></th>
<th>−443</th>
<th>−156</th>
<th>−66</th>
<th>+245</th>
<th>+2080</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.85</td>
<td>0.86</td>
<td>0.81</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>−156</td>
<td>1.00</td>
<td>0.96</td>
<td>0.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>−66</td>
<td>1.00</td>
<td>1.00</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+245</td>
<td>1.00</td>
<td>1.00</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+2080</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

All values of linkage disequilibrium (D’) were statistically significant (P < 0.001)

### Table 4. Haplotype reconstruction and relative frequencies

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>00000</td>
<td>0.0018</td>
</tr>
<tr>
<td>00001</td>
<td>0.4157</td>
</tr>
<tr>
<td>00010</td>
<td>0.0029</td>
</tr>
<tr>
<td>00011</td>
<td>0.0144</td>
</tr>
<tr>
<td>00101</td>
<td>0.0006</td>
</tr>
<tr>
<td>01001</td>
<td>0.0099</td>
</tr>
<tr>
<td>01100</td>
<td>0.0211</td>
</tr>
<tr>
<td>10000</td>
<td>0.0966</td>
</tr>
<tr>
<td>10001</td>
<td>0.0504</td>
</tr>
<tr>
<td>10010</td>
<td>0.0052</td>
</tr>
<tr>
<td>10011</td>
<td>0.0003</td>
</tr>
<tr>
<td>11010</td>
<td>0.1175</td>
</tr>
<tr>
<td>11100</td>
<td>0.0016</td>
</tr>
<tr>
<td>11110</td>
<td>0.2597</td>
</tr>
</tbody>
</table>

Var = −443 | Var = −156 | Var = −66 | Var = +245 | Var = +2080
C = 0 | G = 0 | T = 0 | TG = 0 | C = 0
T = 1 | GG = 1 | G = 1 | GTGG = 1 | T = 1
expression or function of the OPN protein, we decided to investigate on polymorphisms located in the 5’ flanking region in search for variants which might have a functional role in transcriptional regulation.

To identify naturally occurring sequence variations within OPN regulatory sequences, a fragment flanking the 5’ end and including the first intron (−2100 to +1015) of the gene was amplified and sequenced in six Italian individuals. A total of nine variants were found: the genomic structure and locations of variant sites are shown in Fig. 1A; Fig. 1B shows the sequence between −462 and +3. Three of these SNPs (−245, +954, +1020) were located in the first intron: the first one (−245) was previously described by us, and the two others were already reported in the Japanese population (17). Six novel SNPs were found in the sequence upstream of the transcription start site and also confirmed in a recent Japanese
study (22). Five of the observed variants (−443, −156, −66, +245, and +2080) were further analyzed in 50 Italian individuals, and allele frequency was determined. Allele frequencies of all polymorphic variants are shown in Table 2.

Calculation of linkage disequilibrium between pairs of variants (D’ value) was done using the Arlequin software, resulting in highly significant values, as shown in Table 3. Haplotype reconstruction, obtained by the computer program PHASE, resulted in finding 14 haplotypes as shown in Table 4 with their respective frequencies. Considering the −443 C/T, −156 G/GG, and −66 T/G polymorphisms in the immediate 5′ flanking region, four haplotype combinations accounted for 96% of the total haplotype frequencies, the remaining 4% consisting of three haplotypes with frequency lower than 2%. The four most frequent haplotypes are indicated as A, B, C, and D, as shown in Fig. 2.

Transfection experiments. The location of three polymorphic variants in the immediate 500 bp at the 5′ end of the OPN gene prompted us to investigate on their effect in promoter activity; therefore, DNA fragments (from −474 to +42 relative to the transcription start site) corresponding to the four common haplotypes were inserted in pGL3 expression plasmids upstream of the luciferase reporter gene (Fig. 2). Transient transfection experiments were performed in a panel of cell lines, chosen because these are related to OPN expression and regulation: HeLa (epithelial, derived from human uterine cervical carcinoma), MCF7 (epithelial, derived from human breast adenocarcinoma), U2OS (osteoblast-like, derived from human osteogenic sarcoma), COS7 (fibroblast, derived from monkey kidney), HEK293 (fibroblast, derived from human embryonic kidney), and HK2 (human immortalized epithelial cells derived from normal proximal convoluted tubule).

Figure 3 shows that in all tested cell lines the A haplotype showed the lowest promoter activity, significantly different from the remaining three haplotypes, with a maximum of 4.87-fold with respect to B haplotype for the HEK293 cell line. Differences among B, C, and D haplotypes, more evident in some of the tested cell lines, did not reach statistical significance. Since the A haplotype differs from B, C, and D at the −66 position, to further confirm that this G/T variant has a major effect on promoter activity, we prepared two additional plasmids by mutagenesis of C and D constructs, carrying G in place of T at the −66 position (see Fig. 2). These plasmids, named E and F, were transfected in COS7 and U2OS cell lines. In both tested cell lines, luciferase activity of haplotypes A, E, and F did not show statistical difference, whereas E vs. C and F vs. D differed significantly (data not shown), consistent with a major role of the T/G variant at −66.

Transcription factor binding sites and variant alleles. An analysis of sequences corresponding to the three variant locations was carried out by using computer comparison with consensus binding sites for known transcription factors, both in the MacVector program. This analysis suggested putative binding site for the transcription factor around the −66 SNP, for CBF1/RUNX2 factor around the −156 small insertion and for the MYT1 zinc finger factor around the −443 SNP.

The nucleotide sequence of the human osteopontin promoter between −170 and −30, compared with the rat, mouse, and porcine sequence, is shown in Fig. 4A in which we observed that the positions of the SNPs −66 and −156 are located in a conserved region. Considering RUNX2 binding sites, which are very important for regulation of OPN expression in bone tissue, we observed that the −156 SNP generates a RUNX2 binding site very close to another RUNX2 binding site located 14 bp downstream, which is highly conserved between species, as shown in Fig. 4B.

We then designed double-stranded oligonucleotides reproducing the sequences corresponding to the above loci in both allelic forms (see Table 1) and performed EMSA using nuclear extracts from different cell lines. Figure 5 shows that the −82 to −47 oligonucleotides, including the −66 T or G alleles formed retarded complexes with U2OS nuclear extracts, which were also observed in comparable way with HEK293 nuclear extracts (not shown). These complexes were sequence specific since they could be competed by the respective unlabeled oligonucleotides. Differences in intensity were appreciable for the two alternative probes and cross-competitions highlighted that the oligonucleotide with the T allele contained a higher affinity binding site than the G allele. Addition of specific anti-SPI and anti-S3 antibodies in the EMSA caused appearance of a specific supershifted complex for SPI and disappearance of the specific complex for S3. A nonrelated control antibody did not af-
EMSA experiments were also performed using the −460 to −425 probes with −443 T or C alleles. This probe, predicted to contain a putative binding site for the MYT1 zinc finger factor (14), gave rise to retarded complexes that differed according to nuclear extracts used and, in part, according to T or C allele (not shown). Lack of specific antibodies prevented us from proving binding of a specific transcription factor.

Cotransfection experiments with SP1 and Runx2 expression vectors. SP1 overexpression in U2OS cotransfected with haplotypes A and B, which differ at the −66 position only, and a CMV-SP1 expression plasmid resulted in around 10-fold activation of promoter activity with a 1.24-fold higher activation of the B plasmid compared with A (average of two experiments performed in triplicate: 1.21–1.27), consistent with higher binding affinity of SP1 to the T allele at −66, located in the B haplotype.

To test whether the −156 polymorphism could underlie a different response to the RUNX2 transcription factor, cotransfections with reporter plasmids containing two promoter haplotypes differing at −156 (haplotypes B and D) and an expression vector carrying the Runx2 murine cDNA were performed in U2OS, MCF7, and HEK293 cell lines. When we tested the values obtained by comparing the two haplotypes in the three different cell lines, we obtained two types of information, as shown in Fig. 7: 1) U2OS and MCF7 cells displayed statistically significant response to RUNX2, whereas HEK293 were virtually unresponsive; 2) the haplotype with the GG allele

Fig. 5. Gel retardation assay with a probe including the −66 SNP. Gel retardation assay with a double-stranded oligonucleotide that reproduces the OPN promoter from −88 to −48 containing the G/T polymorphism, and nuclear extracts from U2OS cells: lanes 7 and 8, in the absence, and lanes 1–6 and 9–14, in the presence of competitor oligonucleotides at 50-, 100-, or 250-fold molar excess in the two allelic forms. Presence of antibody against SP1 generates a supershifted complex (lane 17), whereas antibody against SP3 shows disappearance of a complex as indicated with asterisk (lane 18). No effect observed in the absence of antibody or in the presence of a nonspecific antibody (lanes 15 and 16).
showed higher response than the one with the G allele, in both U2OS ($P = 0.0078$) and MCF7 cells ($P = 0.0001$).

**DISCUSSION**

Changes of expression of the OPN gene, mostly overexpression, have been described in a variety of conditions; therefore, investigating mechanisms of OPN transcriptional regulation can provide useful information on the molecular basis of expression variation.

Several polymorphisms in the human OPN encoding gene have been identified in different populations: some have been reported in the 5′ flanking region in a Japanese population (22), others in exons and introns and in the 3′ untranslated region (12, 17). Our work was aimed at verifying whether alleles at polymorphic sites in the 5′ flanking region were able to influence the level of promoter activity and characterizing their ability to bind transcription factors.

Our analysis, focused on a 1-kb region upstream of the transcription start site, confirmed the presence of polymorphic variants, the same as those observed in the Japanese population. The reconstruction of haplotypes showed that four of them have frequencies above 10% in our population with a high degree of linkage disequilibrium inside this genomic region.

The functional characterization of three of the promoter polymorphic variants was carried out by preparation of expression plasmids in which the luciferase reporter gene was fused to the OPN promoter sequence in the 4 most frequent haplotypes found in our population. Transfection of these plasmids in a panel of different cell lines showed that the...
For the SNP at −156, our sequence analysis suggested that a binding site for RUNx2 factors, although not exactly matching the reported consensus sequence (9), might be created by the presence of two G nucleotides inside a long stretch of Ts, which, on the complementary strand, gives rise to the AAC-CAAA sequence. We therefore conducted an analysis, by both EMSA and cotransfection experiments with a Runx2 expression plasmid, which showed that an antibody to Runx2 was able to supershift a retarded complex and that Runx2 expression was able to activate the OPN promoter in a differential way, according to the allele. Both the G and the GG containing putative RUNT site are able to bind the factor with an apparently better ability for the GG allele, as shown by higher intensity of the retarded complex detectable when this becomes the only component (Fig. 6C), and higher intensity of the supershifted band. This was verified with nuclear extracts from two cell lines, U2OS and HEK293. Cotransfection experiments showed that expression of Runx2 activated the OPN promoter in U2OS and MCF7 cell lines, whereas no activation was observed in the HEK293 cell line, suggesting that, although a RUNT factor is present in nuclear extracts of HEK293, promoter activation depends on the presence of some additional factor absent in this last cell line. It is interesting to note that besides U2OS, an osteoblast-like cell line derived from osteosarcoma, activation took place in MCF7 cells, derived from human breast cancer, which is a tumor highly prone to bone metastasis. Another interesting observation comes from recent reports about variants in binding sites for the RUNX1 gene, which belongs to the RUNT family of transcription factors as well as RUNX2, found associated with common autoimmune diseases (2). Since the RUNX factors share their DNA binding sites and can heterodimerize, our finding raises various hypotheses on possible regulation of OPN related to susceptibility to disease in which the immune response has a relevant role.

The −443 SNP, at our EMSA, showed differential binding of a yet unidentified nuclear factor. Our sequence analysis suggests that this may be a putative binding site for the MYT1 zinc finger transcription factor, which, up to now, has been reported as involved in neurogenesis and pancreas development (14), or a related factor belonging to the same zinc finger subfamily. Although our results show differential binding of a nuclear factor to the sequence according to the polymorphic alleles, further work is required to clarify this point.

Our results clearly suggest that different haplotypes related to common variants in the promoter of OPN gene may influence the individual degree of gene expression, although other levels of regulation have to be considered especially in the case of this multifunctional protein. Moreover, individual variability in response to various stimuli, such as those triggered by inflammation, necrosis, growth factors, and hormones, is even more complicated, although variability in sequences that are targets of different factors is likely to play a role.

Study of genetic factors involved in susceptibility to complex traits relies on analysis of DNA variants at the susceptibility locus and, subsequently, on correlation of alleles with functional properties. The finding that haplotypes in the OPN encoding gene affect promoter activity will be of great help in the study of diseases for which osteopontin is a good candidate as a susceptibility gene.
ACKNOWLEDGMENTS

We thank Drs. Renata Bocciardi and Marco Musso for helpful discussion and review of the manuscript and Loredana Velo for secretarial assistance.

GRANTS

This work was supported by a Fondo per gli Investimenti della Ricerca di Base (FIRB) grant of the Italian Ministry of University to R. Ravazzolo.

REFERENCES

36. Tamura A, Shingai M, Aso N, Hazuku T, and Nasu M. Osteopontin is released from the heart into the coronary circulation in patients with


