Functional and Structural Profiling of the Human Thrombopoietin Gene Promoter

Corinna Dördelmann, Ralph Telgmann, Eva Brand, Claudia Hagedorn, Bianca Schröer, Sandra Hasenkamp, Peter Baumgart, Peter Kleine-Katthöfer, Martin Paul, and Stefan-Martin Brand-Herrmann

From the Leibniz-Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, University of Münster, Münster D-48149, University Hospital Münster, Department of Internal Medicine, Nephrology and Hypertension, D-48149, Münster, Clemenshospital GmbH Münster, Internal Medicine I, D-48145, Münster, Charité, University Medicine Berlin, Campus Benjamin Franklin, Institute of Clinical Pharmacology and Toxicology, D10117 Berlin, Germany

Human thrombopoietin (TPO) is involved in cardiovascular disease as it regulates megakaryocyte development and enhances platelet adhesion/aggregation. The THPO promoter structure is still controversial. By reverse transcription-PCR, we confirm that THPO transcription is cell-line dependently initiated at two alternative promoters, which we newly designated P1a and P1. We subsequently electrophoretically scanned and resequenced these portions in 95 and 57 patients with cardiovascular disease, respectively, and identified seven variants (−1450/del58bp, C-920T [rs2855306], A-622G, C-413T [rs885838], C+4-5A, G+115A, and C+135T). After subcloning of 1032 bp of THPO P1 in pGL3-basic vector, five molecular haplotypes (MolHaps 1–5) were observed: [A−622-C−413-C+4-5G+115], wild type (wt), [A−622-T−413-C+4-5G+115], [G−622-T−413-C+4-5G+115], [A−622-C+4-5G−413-A+5G+115], [A−622-C+4-5C+4-5A+115], and analyzed in reporter gene assays in HEK293T and HepG2 cells. MolHaps 2, 4, and 5 were significantly more active than wt (all p values ≤0.01) in HEK293T cells, MolHap3 exerted a substantial loss of promoter activity (p < 0.0001 in HEK293T and p < 0.01 in HepG2, compared with wt). Electrophoretic mobility shift assays revealed that A−622G and C−413T individually differed from MolHaps in their DNA-protein interaction patterns. Supershift and chromatin immunoprecipitation assays identified CCAAT/enhancer-binding protein α as the binding protein exclusively for the −622A allelic portion.

The human glycoprotein hormone thrombopoietin (TPO), also known as myeloproliferative leukemia virus ligand (c-Mpl), is an important growth factor for the megakaryocyte lineage, controlling its proliferation and differentiation (1, 2). TPO affects platelet function by stimulating platelet aggregation via phosphoinositide-3 kinase (3), and it affects mature platelets by reducing the threshold levels of ADP, collagen, or thrombin necessary for aggregation (4, 5) and stimulates platelet adhesion (6). In TPO-deficient mice platelet counts are reduced by ~90% compared with wild type mice (wt) (7). Circulating TPO levels are strongly influenced by inflammatory processes, including the presence of coronary atherosclerosis (8). Although predominantly expressed in liver cells, TPO is also generated by kidney cells and bone marrow stroma as well as smooth muscle cells (9). Recently, a genome-wide linkage analysis in a large Asian kindred proposed TPO as a candidate for platelet count variation (10). The THPO locus has been mapped to chromosome 3q26.3-q27. Three groups independently reported that THPO comprises 6 exons and 5 introns spanning more than 6.2 kilobases (kb) (11–13), whereas Chang et al. (14) proposed that it consists of 7 exons and 6 introns spanning 8 kb. Thus, the 5′-regulatory region consists of two alternative promoters which we designated P1a and the intronic promoter P1 with multiple transcription initiation sites, with 10% of all transcripts synthesized from P1a and 90% originating from P1 (15, 16). Given the controversy with respect to the THPO structure, we analyzed the two alternative promoter regions and identified 7 exons instead of 6 and propose an alternative nomenclature for THPO architecture. Furthermore, we systematically scanned the promoter region P1 in 190 chromosomes of high risk myocardial infarction (MI) patients for genetic variation/single-nucleotide polymorphisms (SNPs) and resequenced this promoter portion in an additional set of 114 chromosomes of patients with cardiovascular disease (CVD) including MI and essential hypertension. We then performed molecular func-

---

*This work has been supported by grants from the German Ministry for Education and Science (BMBF; to M. P.) in the National Genome Research Network for Cardiovascular Disease and to E. B., S.-M. B.-H., and R. T. in Bio-Profile-Project BMBF 0313040C. This study was also supported by European Union Project Network of Excellence Grant FP6-2005-LIFESCI-HEALTH-6, Integrating Genomics, Clinical Research and Care in Hypertension, InGenious HyperCare proposal 037093 (to E. B. and S.-M. B.-H. of the European Union Project Network of Excellence Grant FP6-2005-LIFESCI-HEALTH-6, Integrating Genomics, Clinical Research and Care in Hypertension, InGenious HyperCare proposal 037093 (to E. B. and S.-M. B.-H.) and a Charité-University Medicine Berlin, Campus Benjamin Franklin, Institute of Clinical Pharmacology and Toxicology, D10117 Berlin, Germany.

1 To whom correspondence should be addressed: Leibniz-Institute for Arteriosclerosis Research, Department of Molecular Genetics of Cardiovascular Disease, Domagkstrasse 3, D-48149 Münster. Tel.: 49-251-83-52996; Fax: 49-251-83-52997; E-mail: brandher@uni-muenster.de.

2 © 2008 by The American Society for Biochemistry and Molecular Biology, Inc. Printed in the U.S.A.

3 The abbreviations used are: TPO, thrombopoietin; THPO, thrombopoietin gene; CVD, cardiovascular disease; wt, wild type; MI, myocardial infarction; SNP, single nucleotide polymorphism; MolHap, molecular haplotype; ECTIM, Etude Cas-Témoins de l’Infarctus du Myocarde; MolProMD, Molecular Functional Profiling for Mechanism Detection; EMSA, electrophoretic mobility shift assay; TFBS, transcription factor binding site; ChIP, chromatin immunoprecipitation; C/EBP, CCAAT/enhancer-binding protein; HEK, human embryonic kidney; PMA, phorbol 12-myristate 13-acetate; RT, room temperature; NF-1, nuclear factor 1.
tional analyses of the identified promoter variations in single-allelic assays as well as for the molecular haplotypes (MolHaps), determined by individual subcloning procedures in the 57 CVD patients. Subsequently, in-depths reporter gene and band shift assays were performed in two different cell lines under three different stimulatory regimes.

EXPERIMENTAL PROCEDURES

Study Populations—The current investigation was based on the Etude Cas-Témoins de l’Infarctus du Myocarde (ECTIM) and Münster Molecular Functional Profiling for Mechanism Detection (MolProMD) studies. ECTIM is a study of patients with MI ($n = 1000$) from several regions covered by World Health Organization MONICA (MONItoring trends and determinants in Cardiovascular disease) registers and controls ($n = 1000$) representative of each geographic area. The ECTIM Study design was originally described in 1992 (17). Genetic informed consent was obtained from all study subjects. Ninety-five high risk MI patients (positive family history for CVD, MI $\leq 55$ years of age) from the ECTIM study were selected for a preliminary scan of the THPO promoter region. An additional sample of 57 CVD patients recruited in Münster hospitals (ongoing recruitment) as well as 35 healthy controls was used for molecular haplotyping and further functional profiling studies. The Münster MolProMD study is a running study of patients with CVD (MI, essential hypertension) and families, aimed at studying molecular genetic mechanism of CVD. The study was approved by the ethics committee of the Medical Faculty, Westfälische Wilhelms-University of Münster, and written informed consent was obtained from all study subjects.

Scanning of the THPO Promoter P1 for Genetic Variants—Genomic DNA was prepared from white blood cells using either phenol-chloroform extraction or a salting out method (18). From the published sequences of the THPO gene (accession number D32046), 15 overlapping fragments $<300$ bp in length from 95 MI patients (190 alleles) of the ECTIM study were amplified to cover 2084 bp of the upstream region. Specific amplification protocols can be obtained at GeneCanvas INSERM website.

Single-strand Conformation Polymorphism analysis was performed as previously described (19). DNA from patients presenting different migration patterns in the polyacrylamide gels were then sequenced twice (both DNA strands with sense and antisense primers) using an automated sequencing device (ABI Prism 377, PerkinElmer Life Sciences).

Resequencing of the THPO Promoter P1 and Identification of Molecular Haplotypes (MolHaps)—To identify THPO promoter MolHaps, 1032 bp of the promoter region P1 of genomic DNA from 57 CVD patients from the Münster study were amplified and subcloned into vector pCRII-TOPO (Invitrogen) and sequenced twice as described above.

Cell Culture—Human embryonic kidney (HEK) cell line 293T was maintained in Dulbecco’s modified Eagle’s medium (Sigma) with 10% bovine calf serum (iron-supplemented; Cell Concepts, Umkirch, Germany), penicillin (100 units/ml), streptomycin (100 ng/ml), and $\gamma$-glutamine (2 mmol/ml). Human hepatocellular carcinoma cells (HepG2) were maintained in the same medium without iron supplementation. For stimulation experiments, cells were incubated with 0.5 mmol/liter 8-Br-cAMP (Biolog, Bremen, Germany) or $10^{-8}$ mol/liter PMA (Fluka, Seelze, Germany) for 24 h.

Isolation of Total RNA and Generation of cDNA—Total RNA was isolated with the RNeasy Mini kit (Qiagen, Hilden, Germany) following the manufacturer’s protocol. Of the total RNA, 5 µg were used for the generation of cDNA, performed with the First Strand cDNA Synthesis kit (Fermentas, St. Leon-Rot, Germany).

Diagnostic PCR—Exon-specific oligonucleotide primers were designed for exon 1a (E1aS, 5’-atgacaacctgagaaag-3’), exon 1 (E1S, 5’-tcgctgcctgctgactttg-3’), exon 2 (E2AS, 5’-caggagattcggagga-3’), and exon 3 (E3AS, 5’-tcctcagtaaactctg-3’). Amplifications were confirmed by automated sequencing as described above.

Transient Transfections—HEK293T cells were transfected with the reporter gene vectors pGL3-basic (Promega, Mannheim, Germany), generating the reporter vectors pGL3-P1 containing the −413C and −622A allele (positions from the first nucleotide of exon 1; accession number U17071) as template. Sequence localization refers to the first nucleotide of exon 1; the sense primer was located at position −902, 5’-upstream of exon 1 (5’-AGAGAGCCT-GAGAAAGTTCTGTG-3’), and the antisense primer was located within exon 1 (5’-ACTACCCGGGTGGAAGGG-3’). Amplifications were subcloned into the luciferase reporter gene vector pGL3-basic (Promega, Mannheim, Germany), generating the reporter vectors pGL3-P1 containing the −413C and −622A allele, pGL3-P1 containing the T allele, and pGL3-P1 containing the G allele of A-622G. The presence and correctness of the sequences were confirmed by automated sequencing as described above.

Terminal Deoxynucleotidyltransferase Labeling and Annealing—All oligonucleotides for electrophoretic mobility shift assay (EMSA) (supplemental Table 1) were synthesized at a minimum coupling efficiency of $>98.5\%$ and purified twice by high performance liquid chromatography (IBA, Gottingen, Germany). Single-stranded DNA was labeled using the biotin 3’-end DNA labeling kit (Thermo Fischer, Bonn, Germany) with Biotin-11-dUTP, Biotin-16-dUTP (Roche Applied Science), or $\gamma$-ATP following the manufacturer’s protocol. To generate double-stranded labeled oligonucleotides for EMSA, 2 pmol each of sense and antisense oligonucleotides were mixed
in 100 mmol/liter NaCl, heated at 95 °C for 5 min, and then allowed to cool down to RT overnight.

**Extraction of Nuclear Proteins**—Nuclear proteins were harvested by a modified procedure of the protocol published by Schreiber et al. (20). Intactness of nuclear protein extracts was ascertained by SDS-PAGE and Coomassie Blue staining.

**EMSA—EMSAAs** were performed using the LightShift chemiluminescent EMSA kit (Thermo Fischer). For each reaction 5 μg of nuclear protein extract were mixed with 500 ng of presheared poly(dI-dC)/poly(dA-dT) (Amersham Biosciences) as nonspecific competitor plus, where applicable, a 200-fold molar excess of unlabeled oligonucleotide (8–10 pmol/reaction) as specific competitor and incubated with binding buffer (20 mmol/liter MgCl₂, 240 mmol/liter KCl, 40 mmol/liter HEPES/KOH pH 7.9, 5 mmol/liter spermidine (Sigma), 20% Ficoll) for 5 min at RT. DNA concentration in lanes with no specific competitor was adjusted with the according amount of poly(dI-dC)/poly(dA-dT). After the addition of 40–50 fmol of the labeled THPO probe bearing one allele, reactions were left for 20 min at RT or 45 min on ice. For supershift reactions, 0.2 μg of the chosen antibody was added and incubated for an additional 15 min at RT. Gels were run as a 6% native PAGE in 0.5× or 0.25× Tris-buffered EDTA, blotted, and visualized using the Chemiluminescent Nucleic Acid Detection Module (Thermo Fischer).

**Chromatin Immunoprecipitation (ChIP) Assay and Real-time PCR—ChIP was performed as described previously (21, 22).** Approximately 10⁸ cells were fixed by adding formaldehyde to a final concentration of 1% and incubated by modest shaking for 30 min at RT. Thereafter, cells were washed twice with cold phosphate-buffered saline (Sigma). The pellet was resuspended and lysed, and nuclei were isolated and sonicated until the chromatin had an average length of ~500–1500 bp. After centrifugation, the supernatant was incubated with 5 μg of antibody against CCAAT/enhancer binding protein α (C/EBPα; #2295, Cell Signaling, Frankfurt, Germany), C/EBPβ (#3082, Cell Signaling), and C/EBPβ (#2318, Cell Signaling) overnight at 4 °C for immunoprecipitation. The following day 10 μl of magnetic-G beads (Invitrogen) were added and further incubated at 4 °C for 1–3 h. After appropriate washing, the antibody-transcription factor-DNA complex was eluted from the beads, subsequently formaldehyde cross-links were reversed, and proteins were digested with proteinase K at 67 °C overnight. DNA was extracted by phenol/chloroform/isooamy alcohol, washed with 75% ethanol, resuspended in water, digested with the restriction enzyme TsoI (Fermentas) to shorten the DNA fragments, and used for PCR and real-time PCR. Primer for the promoter variant A-622G were 5′-GGA-CAGAGACTGTGGGGAG-3′ (sense) and 5′-GACCTCC-CGGCAATAAC-3′ (antisense). Real-time PCR was performed using the Platinum® SYBR® Green qPCR SuperMix-UDG with ROX (Invitrogen) and a 7500 ABI-prism sequence detection system (Applied Biosystems). Values were normalized against glyceraldehyde-3-phosphate dehydrogenase. Semiquantitative PCR was followed by densitometric analyses using NIH Image J 1.40 (rsb.info.nih.gov/ij/download.html).

In Silico Analyses of Putative Transcription Factor Binding Sites and CpG Islands—Portions of about 50 bp (~25 bp flanking either side of the polymorphic positions) were subjected to computer-aided analyses using the Alibaba2.1 net-based search tool (23). For the detection of CpG islands, 1188 bp of THPO promoter region P1a were analyzed using the “CpG island searcher” tool.

**RESULTS**

**Transcription of THPO Is Initiated at Exon 1a and Exon 1—**To analyze the promoter selection of THPO in HEK293T and HepG2, we designed sense primers within exon 1a (E1aS) and 1 (E1S) and antisense primers within exons 2 (E2AS) and 3 (E3AS) for reverse transcription-PCR (Fig. 1B). We detected THPO in both cell lines (Fig. 1B) and under all stimulatory regimes, without stimulation and with 10⁻⁸ mol/liter PMA and 0.5 mmol/liter 8-Br-cAMP. THPO transcription initiated from promoter P1a was much more prominent in the embryonic cell line HEK293T than in the adult liver cell line HepG2 except for cAMP-stimulated cells.

**Analyses of the THPO Promoter Region P1a—**We designed two pairs of primers flanking P1a and P1, directly sequenced, and subsequently subcloned genomic DNA of 57 MI patients to determine MolHaps of four promoter variants of promoter P1 (Fig. 2). The P1a region was amplified from positions −664 to −1653 (from the first nucleotide of exon 1; accession number U17071). As this area was GC-rich, we subsequently performed a CpG island analysis. CpG islands are defined as regions of DNA (∼200 bp) with a GC content >50% and a ratio of observed CpG to expected CpG of ≥0.6 (24). We analyzed 1188 bp of THPO promoter region P1a from positions 120601 to 121680 (accession number AC078797) and observed a GC content of 65.7% with an ObsCpG/ExpCpG ratio of 0.688, strongly indicating the presence of CpG islands.

**Computational Prediction of Putative Transcription Factor Binding Sites (TFBSs) in the THPO Promoter P1—**Analyses of A-622G, C-413T, C+5A, and G+115A with the net-based program Alibaba2.1 predicted several TFBSs, some of which were altered by the presence of different alleles (Fig. 3).

**Identification of Genetic Variants in the THPO Promoter and Determination of Molecular Haplotypes—**By the combined scanning approach using PCR-single-strand conformation polymorphism analysis in 95 MI patients from the ECTIM study and resequencing of the THPO promoter P1 in an additional set of 57 CVD patients from the MollProMD Study, we identified 6 SNPs and a 58-bp deletion variant (Fig. 1A). Three variants were located in the 5′-flanking (C-920T [rs2855306], A-622G, and C-413T [rs885838]) and three in the 5′-translated region (C+5A, G+115A, and C+135T). The 58-bp deletion identified in 4 CVD patients is located between positions −1450 and −1507 (−1450/del58bp; accession number U17071), 137 bp upstream of exon 1a. The fragment containing the A-622G site display sequence similarity to members of the C/EBP family, particularly C/EBPβ and C/EBPα, which belong to the basic leucine zipper protein (bZIP) transcription factor family. This SNP is also located in a binding site for Sp1, but the position of this binding site is displaced for 2 bp by the G allele. For the G allele, the sequence shows similarity to a member of
site for PEA3, a member of the Ets family of transcription factors, was identified by visual accounting and is only present for -622G. The sequence containing C-413T displays similarity for the consensus sequence of a member of the nuclear factor 1 (NF-1) family and the T allele represents a putative binding site for C/EBP. The consensus sequence for this transcription factor additionally shows sequence similarity to both alleles of C+5A and G+115A, which are both located in exon 1. The +5C allele is located in a binding site for the thyroid hormone receptor (T3R). Both alleles of G+115A show sequence similarity for members of the cAMP response element-binding protein (CREBP) family, more precisely CREB1, which has a leucine zipper structure (26). For the A allele, we also found a putative binding site for members of the PU.1 transcription factor family and GATA-binding protein 1, a zinc finger transcription factor. The members of the PU.1 family belong to a divergent subclass of the Ets transcription factor family (27).

Transfection Experiments of the THPO Promoter P1—To investigate the identified genetic variants A-622G, C-413T, C+5A, and G+115A for their effect on promoter activity, DNA fragments corresponding to the five MolHaps were inserted into pGL3-basic (Promega). Transient transfection experiments were performed in HEK293T and HepG2 cells, because these cell lines expressed THPO endogenously. In HEK293T, the activity of MolHap4 was significantly increased under basic conditions \( (p < 0.0001) \) and under both stimulatory conditions, \( 10^{-8} \) mol/liter PMA \( (p = 0.0124) \) and 0.5 mmol/liter 8-Br-cAMP \( (p = 0.0136) \) (Fig. 4A), which was not observed in HepG2 cells under stimulatory conditions. Under basic conditions MolHap4 activity was increased compared with wt \( (p = 0.013) \) (Fig. 4B). MolHap5 also showed a significantly increased transcriptional activity compared with MolHap1 (wt) in HEK293T cells under all stimulatory regimes (basic conditions, \( p < 0.0001) \); PMA, \( p < 0.0001); cAMP, \( p < 0.0001) \; \text{Fig. 4A}), which was not observed in HepG2 cells under basic conditions. Stimulation with PMA and cAMP, however, resulted in a slightly increased transcriptional activity \( (p, p = 0.02) \; \text{cAMP, } p = 0.033) \) (Fig. 4B). Because MolHap4 occurred only once in the MolProMD and twice in the ECTIM sample and MolHap5 only once in MolProMD and ECTIM, they were not further studied. In HEK293T cells, under basic

the Ets family, c-Ets-1, which is involved in the transcriptional activation of different genes such as stromelysin 1, collagenase 1, or urokinase type plasminogen activator (25). A core binding
and stimulatory conditions MolHap2 was significantly more transcriptionally active than wt (basic conditions, \( p = 0.0036 \); PMA, \( p = 0.0014 \); cAMP, \( p = 0.0004 \)) (Fig. 4A). In HepG2 cells there was no significant difference in transcriptional activity of MolHap1 and 2 (Fig. 4B). Conversely, MolHap3 displayed a significant abrogation in promoter activity in both cell lines; 1) in HEK293T (basic conditions, \( p < 0.0001 \); PMA, \( p < 0.0001 \); cAMP, \( p = 0.0003 \)) (Fig. 4A), 2) in HepG2 (basic conditions, \( p < 0.0001 \); PMA, \( p = 0.001 \); cAMP, \( p = 0.034 \)) (Fig. 4B) compared with wt. C-920T did not show any allelic differences in promoter activity in transient transfection assays using HEK293T and HepG2 cells (data not shown) and was, therefore, not further investigated.

EMSAs for A-622G and C-413T with Nuclear Extracts from HEK293T and HepG2—Investigation of A-622G within the THPO promoter P1 by EMSA resulted in an altered DNA-protein binding pattern with nuclear proteins from HEK293T (Fig. 5A) and HepG2 (Fig. 5B). Incubation with nuclear proteins from HEK293T cells resulted in a specific interaction that was more pronounced for the G allele under basic conditions and...
PMA stimulation (Fig. 5A, arrow) and another faster migrating non-allele-specific band (Fig. 5A, single asterisk). Conversely, the interaction with HepG2 nuclear proteins showed an A allele-specific band (Fig. 5B, arrow), suggesting different expression patterns of transcription factors in the two different cell lines. Supershift assays with a C/EBPα antibody resulted in a specific shift of the A allele with nuclear proteins from HEK293T (Fig. 5C). This finding could be confirmed by ChIP assays, which identified C/EBPα as one of the binding factors at position −622 (Fig. 5D). Nuclear extracts from HepG2 cells did not lead to any specific supershift, indicating that C/EBPα does not bind in this cell line or is not present to a sufficient amount. Despite the in silico prediction of an Ets-binding site for the −622G allele, bandshift assays using antibodies against Ets-1/2 as well as antibodies against Sp1, C/EBPα and C/EBPβ failed to interact with nuclear proteins from HEK293T or HepG2 (data not shown).

EMSA for C-413T resulted in specific interaction with nuclear proteins from HEK293T and HepG2 for both alleles upon basic and stimulatory conditions (Fig. 6, A and B), whereas the binding of the T allele seemed to be much stronger with nuclear extracts from HEK293T. Competition with consensus sites for transcription factors Sp1, NF-1, and C/EBP resulted in a specific competition of both alleles and both cell lines with the NF-1 probe (Fig. 6, A and B). The Sp1 and the C/EBP consensus sites did not compete any DNA-protein interaction. These results indicate that NF-1, but neither SP1 nor a member of the C/EBP family, is involved in THPO promoter binding in this region.

Supershift assays with a Sp1 antibody as well as EMSA experiments with the SP1 binding domain instead of nuclear proteins did not result in any shift, indicating that SP1 is either not involved in DNA binding or not present at sufficiently high levels to being detected (data not shown). Supershift assays with anti-NF-1 antibody were also negative.
EMSAs for MolHaps1–3 with Nuclear Extracts from HEK293T and HepG2—Because −622G only occurred together with −413T, we investigated the interaction of a fragment that contained both SNPs in one strand with nuclear extracts from HEK293T and HepG2 by EMSA. For that purpose, we amplified 249-bp fragments by PCR that represented MolHaps1–3. All three labeled haplotypes competed either with the unlabeled MolHap probes or specifically with individual double-stranded 31-bp oligonucleotides with either −622A/G or −413T/C alleles. The addition of nuclear proteins from HEK293T resulted in the presence of two specific bands (Fig. 7A). One band (Fig. 7A, single asterisk) was specifically competed by the full-length homologous amplicon or by the oligonucleotide bearing the −413C or T allele but failed to interact with the oligonucleotide bearing the −622A or −622G allele. A slower migrating band appeared in probes bearing the major −622A allele (Fig. 7A, double asterisks) but was undetected in the probe with the G allele. This slower band was significantly more prominent when probed with MolHap2. The band was specifically competed by the cold probe alone but did not disappear upon competition with either SNP. This indicates that the MolHap sequence regions are specifically interacting with nuclear proteins and that this DNA-protein interaction depends on the presence of both variants in a strand. The same experiment with nuclear extracts from HepG2 showed a similar result (Fig. 7B). There was also a slow migrating-specific band that was competed with the full-length amplicon but not with either single SNP, and it did not appear in the presence of the −622G allele.

DISCUSSION

In the wake of non-hypothesis-driven genome-wide approaches (nk-profiling) (28, 29), which offer the opportunity of detection of genomic loci associated with complex genetic disease, specific and accurate molecular functional profiling of genetic variation in any gene region is warranted. In the present study we analyzed the transcriptional organization of the human THPO and propose a revised nomenclature for its molecular architecture, taking into account previous and partly contradictory reports. We confirm that THPO consists of seven exons and six introns and that transcription is initiated at two alternative promoters. The groups who reported that THPO consisted of 6 exons and 5 introns obtained their results by RNase protection and 5′-rapid amplification of cDNA ends assays; those experiments were performed with adult liver cell lines and revealed a single promoter region with multiple transcription initiation sites (12, 13). In the HepG2 cells used in our experiments, transcription was nearly exclusively started at the downstream promoter, leading to transcripts missing upstream exonic sequences, backing the above-mentioned finding, which led to the OMIM nomenclature for its molecular transcriptional organization. Taking into account the above-mentioned finding, which led to the OMIM nomenclature of exons 1–6 (accession number D32046). On the other hand, Chang et al. (14) reported that THPO contains seven exons and six introns. In embryonic HEK293T cells we demonstrated that transcription is also strongly driven by an upstream promoter region located 5′ of an additional, noncoding exon. Analogous to the nomenclature of the human prolactin gene (30, 31), we suggest designating this exon as 1a (alternative) and retaining the OMIM nomenclature of exon 1–6 for the remaining structure. The respective promoter regions, which we characterized in the present work, are consequently assigned promoter P1a, upstream of exon 1a, and promoter P1, upstream of exon 1.
P1 is the major, if not the only promoter utilized in HepG2 cells and probably responsible for THPO expression in the adult individual. This is in good agreement with results from Ghilardi et al. (15) reporting that in human liver 90% of the transcripts originate from P1. The intronic promoter is located between exons 1a and 1; thus, the region responsible for its transcriptional regulatory potency is clearly defined. By subcloning and resequencing patient DNA, we identified the common MolHaps1–3, \([A^{−622}, C^{−413}, C^{+5}, G^{+115}],\) wt, \([A^{−622}, T^{−413}, C^{+5}, G^{+115}],\) and \([G^{−622}, T^{−413}, C^{+5}, G^{−115}],\) respectively. As MolHaps4 and 5, even if functional, were observed only twice or once in 57 patients, respectively, we focused our analyses on MolHaps1–3, which were defined by allelic constellations representing the distal promoter variant sites A-622G and C-413T. Transient transfection of these haplotypic strands in the context of the full-length P1 revealed a strong transcriptional activity in both HEK293T and HepG2, both variants residing in functional promoter regions.

Furthermore, we were able to demonstrate that single genetic and molecular haplotypic variation in the THPO promoter P1 have distinct and profound functional consequences for transcription as well as DNA-protein interactions as identified by appropriate promoter analyses and bandshift assays. MolHap3 led to a significant loss in transcriptional activity in both HEK293T and HepG2, both variants residing in functional promoter regions.

To identify the molecular basis for this differential promoter usage, we first subcloned and sequenced 1188 bp of promoter P1a and exon 1a of 57 patients with CVD. The THPO promoter P1a is very GC-rich and contains several CpG islands, rendering it a possible target for methylation with activating or inactivating capacities. Because methylation events often represent long-lived regulatory procedures, it is conceivable that transcription from P1a might be repressed by methylation during ontogeny or due to differentiation, as observed in HepG2.

MolHap2 was significantly more transcriptionally active compared with the wt sequence in HEK293T cells. To investigate whether these obviously functional sequences interacted with transcription factors, we performed in silico analyses, which revealed that A-622G is located within a consensus binding site for CCAAT box/enhancer-binding proteins, a family of proteins with either activating or inhibiting (or both) nature (C/EBPα, -β, -γ, -δ, -ε, -ζ) (32). The physiological activity of C/EBP transcription factors within a physiological cellular sit-
THPO Functional Promoter Profiling

utation depends on the actual availability of these factors, their interaction with other partners, and their activation status. After having shown in EMSAs that specific DNA–protein interaction occurred, supershift experiments revealed that C/EBPβ specifically binds to the −622A allele in HEK293T cells irrespective of the stimulatory regime applied. We were able to confirm a specific C/EBPβ binding at this site by ChIP assays. Despite the in silico prediction of an Ets-binding site for the −622G allele, bandshift assays using antibodies against Ets-1/2 as well as antibodies against Sp1, C/EBPα, and C/EBPβ failed to interact with nuclear proteins from HEK293T or HepG2.

By comparing the C-413T variant with consensus sites of SP1, NF-1, and C/EBP, we observed that both alleles bind NF-1 in both cell lines. The −413T-containing promoter, a potential binding site for C/EBPα and NF-1, displayed a significantly enhanced promoter activity compared with the wt sequence in HEK293T cells, whereas in the MolHap3 constellation comprising −622G the promoter activity in both cell lines was substantially decreased. This suggests that −622G may have a key role as a repressor element. Moreover, with the use of EMSAs, we were able to demonstrate that nuclear proteins specifically interact with the region covering the A−622G and C−413T polymorphic sites and that this interaction was more prominent in the presence of the −413T allele in HEK293T cells and missing in the presence of −622G allele in both cell lines. This indicates a loss of protein binding, which may explain the decreased transcriptional activity of MolHap3 in transfection experiments. We did not see any differences between PMA and cAMP-stimulated ceils in protein binding. In the present analysis, we were able to demonstrate that protein binding patterns differed between promoter fractions containing single variation and MolHaps. Indeed, in the embryonic cell line HEK293T and for all three MolHaps we identified distinct and specific binding patterns, which were compatible with the entire sequence fragment as well as C−413T alone. This gives rise to the notion that some DNA–protein interactions were specific exclusively in the presence of both alleles simultaneously in the promoter strand. In the adult liver cell line HepG2 we could not identify this specific DNA–protein interaction, which indicates a different protein binding pattern in these two cell lines. This clearly differed from single interactions involving C/EBPβ, as the −622A-containing sequence was not able to compete the DNA–protein binding of MolHaps. In silico analyses might fail to accurately predict protein binding for MolHap constellations because they cannot integrate the impact of protein–protein interactions on the recruitment of factors to a transcriptional module. Spectrometric approaches, however, might provide a valuable tool to identify and quantify all proteins involved in promoter binding complexes and elucidate the kinetics of their assembly.

Interestingly, the THPO locus on chromosome 3q26-27 has recently been linked with bipolar disorders (29), and a significant association between THPO rs10513797 (intron 2) has been shown with type 1 diabetes mellitus and rheumatoid arthritis (see the Wellcome Trust Case Control Consortium; other disease phenotypes may emerge in future genome-wide studies). As these (mostly) tagging SNPs used in genome-wide studies are only markers in linkage disequilibrium with functional variants, it would be interesting for these consortia to test whether our functional MolHaps are associated with these latter disease phenotypes.

In conclusion, in the present analyses we functionally and structurally characterized the alternative promoter P1a and propose to designate the most distal exon as 1a. Within P1, we identified three common functional MolHaps, which substantially influenced its transcriptional activity. We provide strong evidence that both A−622G and C−413T are functional as they reside in an active region of THPO promoter P1. Further investigations should identify associations of MolHaps with TPO mRNA levels in appropriate cells extracted from typed individuals as well as differential gene expression profiles. Additionally, coexpression experiments and RNA-mediated interference approaches may help to elucidate the interaction of different transcription factors within the THPO promoter. With respect to haplotype-specific protein binding, protein trapping technologies might help to identify candidate proteins that act in concert to regulate THPO transcription.

REFERENCES

1. Kaushansky, K. (2005) J. Clin. Investig. 115, 3339–3347
2. Kaushansky, K. (1998) N. Engl. J. Med. 339, 746–754
3. Kojima, H., Shinagawa, A., Shimizu, S., Kanada, H., Hibi, M., Hirano, T., and Nagasawa, T. (2001) Exp. Hematol. 29, 616 – 622
4. Oda, A., Miyakawa, Y., Druker, B. J., Ozaki, K., Yabusaki, K., Shirasawa, Y., Handa, M., Kato, T., Miyazaki, H., Shimosaka, A., and Ikeda, Y. (1996) Blood 87, 4664–4670
5. Pasquet, J. M., Gross, B. S., Gratacap, M. P., Quek, L., Pasquet, S., Payrastre, B., van Willigen, G., Mounhoff, J. C., and Watson, S. P. (2000) Blood 95, 3429–3434
6. Van Os, E., Wu, Y. P., Pouteels, J. G., Jisseldijk, M. I., Sixma, J., Akkerman, J. W., De Groot, P. G., and Van Willigen, G. (2003) Br. J. Haematol. 121, 482–490
7. Mutone, M., Carpenter, D. A., and de Sauvage, F. J. (1998) Stem Cells 16, 1–6
8. Cotton, J. M., Hong, Y., Hawe, E., Mathur, A., Humphries, S. E., Brown, A. S., Martin, J. F., and Erusalimsky, J. D. (2003) Thromb. Haemostasis 89, 538–543
9. Sungaran, R., Markovic, B., and Chong, B. H. (1997) Blood 89, 101–107
10. Garner, C., Best, S., Menzel, S., Rooks, H., Tector, D. T., and Their, S. L. (2006) Eur. J. Hum. Genet. 14, 101–108
11. Soino, Y., Akahori, H., Seki, N., Horii, T., Ogami, K., Kato, T., Shimada, Y., Kawamura, K., and Miyazaki, H. (1994) FEBS Lett. 353, 57–61
12. Foster, D. C., Sprecher, C. A., Grant, F. J., Kramer, J. M., Kuijper, J. L., Holly, R. D., Whitmore, T. E., Heipel, M. D., Bell, L. A., and Ching, A. F. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 13023–13027
13. Kamura, T., Handa, H., Hamasaki, N., and Kitajima, S. (1997) J. Biol. Chem. 272, 11361–11368
14. Chang, M. S., McNinch, J., Basu, R., Shutter, J., Hsu, R. Y., Perkins, C., Mar, V., Suggs, S., Welcher, A. M., and Li, L. (1995) J. Biol. Chem. 270, 511–514
15. Ghilardi, N., Wienerst, A., and Skoda, R. C. (1998) Blood 92, 4023–4030
16. Wiester, A., Schlemper, R. J., van der Maas, A. P., and Skoda, R. C. (1998) Nat. Genet. 18, 49–52
17. Cambien, F., Poirier, O., Leret, L., Evans, A., Cambou, J. P., Arveiler, D., Luc, G., Bard, J. M., Bara, L., Ricard, S., Tiroet, L., Amouyel, P., Ahencel-Gelas, F., and Soubrerier, F. (1992) Nature 359, 641–644
18. Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988) Nucleic Acids Res. 16, 1215
19. Herrmann, S. M., Funke-Kaiser, H., Schmidt-Petersen, K., Nicola, V., Gautier-Bertrand, M., Evans, A., Kee, F., Arveiler, D., Morrison, C., Orzechowski, H. D., Elbaz, A., Amarenco, P., Cambien, F., and Paul, M. (2001) Arterioscler. Thromb. Vasc. Biol. 21, 1538–1543
20. Schreiber, E., Matthias, P., Muller, M. M., and Schaffner, W. (1989) Nu-
21. Boyd, K. E., Wells, J., Gutman, J., Bartley, S. M., and Farnham, P. J. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 13887–13892
22. Liu, S., Spinner, D. S., Schmidt, M. M., Danielsson, J. A., Wang, S., and Schmidt, J. (2000) J. Biol. Chem. 275, 41364–41368
23. Grabe, N. (2002) In Silico Biol. 2, 1–15
24. Gardiner-Garden, M., and Frommer, M. (1987) J. Mol. Biol. 196, 261–282
25. Grevin, D., Chen, J. H., Raes, M. B., Stehelin, D., Vandenbunder, B., and Desbiens, X. (1993) Int. J. Dev. Biol. 37, 519–529
26. Maekawa, T., Sakura, H., Kanei-Ishii, C., Sudo, T., Yoshimura, T., Fujisawa, J., Yoshida, M., and Ishii, S. (1989) EMBO J. 8, 2023–2028
27. Anderson, M. K., Sun, X., Miracle, A. L., Litman, G. W., and Rothenberg, E. V. (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 553–558
28. McPherson, R., Pertsemidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., Hinds, D. A., Pennacchio, L. A., Tybjaerg-Hansen, A., Folsom, A. R., Boerwinkle, E., Hobbs, H. H., and Cohen, J. C. (2007) Science. 316, 1488–1491
29. The Wellcome Trust Case Control Consortium (2007) Nature 447, 661–678
30. Gellersen, B., Kempf, R., Telgmann, R., and DiMattia, G. E. (1994) Mol. Endocrinol. 8, 356–373
31. DiMattia, G. E., Gellersen, B., Duckworth, M. L., and Friesen, H. G. (1990) J. Biol. Chem. 265, 16412–16421
32. Cao, Z., Umek, R. M., and McKnight, S. L. (1991) Genes Dev. 5, 1538–1552