RAPID COMMUNICATION

A Single Genetic Origin for the Common Prothrombotic G20210A Polymorphism in the Prothrombin Gene

By Ariella Zivelin, Nurit Rosenberg, Shlomit Faier, Nurit Kornbrot, Hava Peretz, Christine Mannhalter, Marie Helene Horellou, and Uri Seligsohn

The polymorphism G20210A in the 3' untranslated region of the prothrombin gene is associated with an increased level of factor II activity and confers a twofold to fivefold increase in the risk for venous thromboembolism. Among Caucasian populations, the prevalence of factor II G20210A heterozygotes is 1% to 6%, whereas in non-Caucasian populations it is very rare or absent. The aim of the present study was to discern whether factor II G20210A originated from a single or recurrent mutational events. Allele frequencies of four dimorphisms spanning 16 of 21 kb of the factor II gene were determined in 133 unrelated Caucasian subjects of Jewish, Austrian, and French origins who bore factor II G20210A (10 homozygotes and 123 heterozygotes) and 110 Caucasian controls. Remarkable differences in the allele frequencies for each dimorphism were observed between the study groups (P = .0007 or less), indicating strong linkage disequilibrium and suggesting a founder effect. Indeed, a founder haplotype was present in 68% of 20210A mutant alleles and only in 34% of 20210G normal alleles (P < .0001). These data strongly support a single origin for factor II G20210A that probably occurred after the divergence of Africans from non-Africans and of Caucasians from Mongoloid subpopulations.

© 1998 by The American Society of Hematology.

IN RECENT YEARS, there has been a growing interest in finding genetic polymorphisms that are associated with an increased risk of venous thromboembolism (VTE). Resistance to activated protein C (APC) caused predominantly by a substitution of G by A at nt 1691 of the factor V gene is currently the most commonly observed prothrombotic polymorphism.1-4 It is found in patients with idiopathic VTE at a frequency of 20% to 60%, depending on the mode of their selection.5-7 In Caucasians, the allele frequency of factor V G1691A varies from 1% to 8%,8,10 whereas among Africans, Chinese, Japanese, and native North and South Americans, this polymorphism is absent.8 We recently showed that factor V G1691A originated from a single mutational event that took place about 21,000 to 34,000 years ago after the evolutionary divergence of non-Africans from Africans and of Caucasians from Mongoloids.9

Another prothrombotic polymorphism in the factor II (prothrombin) gene was recently shown to be associated with an increased risk of VTE.11 This polymorphism, a G to A nucleotide transition at nt 20210 in the 3' untranslated region of the factor II gene, is located at or near the cleavage site of the mRNA precursor to which poly A is added.12 The factor II gene is localized on chromosome 11 near the centromere.13 spans approximately 21,000 bp, and is composed of 14 exons and 13 introns.12 Patients bearing the factor II G20210A polymorphism have an increased plasma factor II level for which no explanation has so far been provided.

Four independent studies from The Netherlands, England, and Sweden showed that the factor II polymorphism was associated with a twofold to fivefold increase in risk of VTE.11,14-16 These studies included a total number of 1,293 unselected patients with idiopathic VTE, of whom 73 (5.6%; range, 5.0% to 7.1%) were heterozygotes for the factor II 20210A allele. In contrast, only 31 (1.8%) of 1,428 matched controls (range, 1.2% to 2.6%) were heterozygous for the polymorphism, predicting an allele frequency of 0.009 in North and Western European populations. Additional data recently compiled by Rosendaal et al17 disclosed that the prevalence of carriers of factor II G20210A in healthy Northern Europeans was similar, ie, 1.7%, whereas in Southern Europeans the prevalence was nearly twice (3%).17

In contrast to these figures, factor II G20210A was found in only 1 of 441 African Americans18 and was absent among 231 Amerindians from Brazil19-20 and 210 Japanese subjects.19,21 Taken together, these data suggest that, similarly to factor V G1691A, factor II G20210A is confined to Caucasian populations.

In this study, we addressed the question of whether factor II G20210A occurred in a single founder or resulted from recurrent mutational events.

MATERIALS AND METHODS

Study groups and collection of samples. To determine the prevalence of the polymorphism in different Jewish ethnic groups, subjects with no history of venous thrombosis were recruited from consecutive admissions to the hospital, outpatient clinics, hospital personnel, and army personnel. The ethnic origin was determined by the country of origin of the 4 grandparents of the subjects. DNA samples that turned out to be homozygous normal (G20210) served as controls for analysis of polymorphisms and haplotypes (see below). To define the origin of factor II G20210A, we analyzed DNA samples of previously diagnosed homozygotes from Israel, Austria, and France and heterozygotes of Jewish origin. The collection of the samples was approved by the Human Subject Ethics Committee of the Sheba Medical Center.

Detection of factor II gene 20210A allele. Factor II G20210A was detected by a slight modification of the originally described method.11 The 345-bp DNA fragment amplified by polymerase chain reaction (PCR) was simultaneously digested by HindIII for detection of the...
polymorphism and by Msp I (5 U/15 µL reaction) for detection of a constant 141-bp fragment indicating that digestion took place (Fig 1).

Analysis of intragenic polymorphic markers in the factor II gene.

Four additional diallelic polymorphisms in the factor II gene were analyzed. The first was a new unpublished polymorphic marker in intron D characterized by direct sequencing with SequiTherm Excell DNA sequencing kit (Epicentre Technologies, Madison, WI). The second was a previously described polymorphism in intron E. The third and fourth were new unpublished polymorphisms in exon 10 and intron M, respectively. They were characterized by us after noting variations in the described sequence of the factor II gene. All 4 polymorphisms were detected by PCR amplification and allele-specific restriction analyses (Fig 2). The site of the polymorphisms, the size of the PCR products, the restriction enzymes used, and forward (F) and reverse (R) primers were as follows: (1) intron D-nt T3728C, 480 bp, Nde I, F5’TCTGCTGGCCCTAATGATGATGCATGG3 and R5’TACCCAGACCTTCAGCACTTAC3; (2) intron E-nt C4125G, 418 bp, Mbo II, F5’AA TACGATCCCCCGCACCGC3 and R5’TGGTACCTGTCGCCCCACT3; (3) Exon 10-nt G8845A, 86 bp, Aci I, F5‘ACCGCCGGCCACTGCTCTG3 and R5’CGGGAGTCTTGCCAA TGCGCACC3; and (4) intron M-nt G19911A, 271 bp, Mnl I, F5’ATGGCTGAAGTGGAGACATGG3 and R5’GTTAGATGCCATTTTCCATCCCC3. The PCR products were generated in 20 µL reaction mixtures that contained 100 to 200 ng of genomic DNA,
250 nmol/L of each primer, 200 nmol/L of each dNTP, 0.125 U of Taq polymerase (Super Nova; Laboratory Products Int, Kent, UK), 1.5 mmol/L MgCl2, and 1X PCR buffer. The reactions were subjected to 30 cycles of 45 seconds of denaturation at 94°C, 45 seconds of annealing at 65°C, and 45 seconds of extension at 72°C. The PCR products (10 µL) were directly subjected for digestion with 5 U of restriction enzymes (New England Biolabs, Beverly, MA) according to the manufacturer’s instructions. The digested PCR products were analyzed on 4% Metaphor agarose (FMC Bioproducts, Rockland, ME) for exon 10 and intron M and on 3% agarose for introns D and E.

To analyze the linkage between intron M polymorphism at nt 19911 and the prothrombotic polymorphism at nt 20210, a 345-bp PCR product containing both sites was amplified as described by Poort et al. This amplified segment was simultaneously digested by EcoRI cleaved the amplified product 19 bp downstream from its 5’ end, yielding a 322-bp fragment. When A was present at nt 19911, EcoRI cleaved the amplified product 19 bp downstream from its 5’ end, yielding a 326-bp fragment. Simultaneous digestion by HindIII and EcoRI enabled detection of a 19911A-20210A haplotype that was represented by a 303-bp fragment (345 bp less the cleaved 23-bp and 19-bp fragments) and a 19911G-20210G haplotype that was represented by the 345-bp fragment resistant to both HindIII and EcoRI (Fig 3A).

Search for new polymorphic markers in the promoter region of the factor II gene. DNA amplification of nucleotides −1074 to −354 of the promotor was performed with the primers 5’AGAAAATGGG-CCTCCCAGGAA TTAAGG3’ and 5’GTGTGAGGACTAAAGG-AGATTAGG3’ under the same conditions for PCR as described above, except for 60 seconds of extension. Three restriction enzymes were chosen for cleavage of the amplified 721-bp segment to smaller fragments suitable for single-stranded conformation polymorphism (SSCP) analysis. The enzymes used and sizes of the obtained fragments in basepairs were: (1) RsrII, 206 and 515; (2) MboII, 395 and 326; and (3) PvuII, 432 and 289. SSCP was performed as described and the radiolabeled PCR fragments were analyzed on MDE gels (FMC Bioproducts) prepared according to the manufacturer’s instructions, with and without 10% glycerol.

PCR fragments of the promotor region from a 20210A homozygote and a 20210G normal homozygote were cloned into a TA cloning kit (Invitrogen, San Diego, CA) by the procedure recommended by the manufacturer. DNA was purified with Wizard Plus Miniprep DNA purification kit (Promega, Madison, WI) and sequenced by the core facilities of the Weizmann Institute (Rehovot, Israel).

RESULTS AND DISCUSSION

Prevalence of factor II G20210A polymorphism in Jewish ethnic groups. A total of 1,670 Jewish subjects belonging to different ethnic groups were examined for the prevalence of FII G20210A polymorphism. As shown in Table 1 the prevalence of heterozygotes for the prothrombotic factor II polymorphism was 6.7% in Ashkenazi (European) Jews and 5.5% in Sephardic Jews from North Africa. This is the highest prevalence so far observed among studied populations that may have stemmed from genetic drift due to isolation, periods of extinction, and times of extensive expansions of Ashkenazi and Sephardic Jews. In Middle Eastern Jews originating in Iraq, Iran, and Yemen, prevalences of 1% to 4% were observed, resembling the prevalences reported in European populations. Interestingly,

---

**Fig 3.** Determination of the linkage between nt 19911 and nt 20210 by simultaneous HindIII and EcoRI digestion of a 345-bp amplified fragment spanning both sites. (A) Schematic illustration of 4 possible alleles and specific fragment sizes in basepairs after cleavage at the indicated sites (X). (B) Electrophoretic separation of the following observed genotypes: 3 normal homozygotes for 20210G (lanes 1 through 3) who were GG, GA, and AA for the 19911 polymorphism, respectively; 2 homozygotes for 20210A (lanes 4 and 5) who were AA and AG for the 19911 polymorphism, respectively; and 2 heterozygotes for G20210A who were AA (lane 6) and AG (lane 7) for the 19911 polymorphism. The additional band in lane 7 between the 345- and 303-bp bands represents a heteroduplex (H.D.). This was proved by the following two mixing experiments. (1) When 345-bp PCR fragments from a homozygote for 20210G and 19911G and a homozygote for 20210A and 19911A were mixed and digested with HindIII and EcoRI, an uncut 345-bp fragment and a 303-bp fragment were obtained (lane 8). (2) When DNA samples from the same 2 individuals were mixed and then amplified by PCR, followed by digestion and electrophoresis, a pattern representing heteroduplex bands was obtained (lane 9) that is identical to the one shown in lane 7. Numbers alongside the gel indicate the size (in basepairs) of the bands.
in Ethiopian Jews, who share African genes, none of 177 subjects examined (354 alleles) bore the 20210A genotype.

Frequency of polymorphic markers in the factor II gene in controls. Table 2 and Fig 2 illustrate the identification of the 3 new polymorphic markers and of a previously published marker in intron E.21 All four di-allelic markers are located 5’ to nucleotide 20210 and cover 16 of 21 kb of the factor II gene. All polymorphisms are probably silent changes, because three of them are located in introns D, E, and M, respectively, and one is a change of codon 368 from CCG to CCA, both encoding for Proline.

The frequencies of the more common alleles for the 4 markers ranged from 0.55 to 0.84 among 220 alleles of 110 normal controls (Table 2). This frequency distribution was considered adequate for linkage analysis. Interestingly, the frequency of 0.65 and 0.35 for the two complementary alleles in intron E was similar to the frequency observed in Japanese subjects.22

Determination of a founder effect. Table 2 compares the allele frequencies for the 4 intragenic markers in 110 normal homozygotes (20210G) that represent the different Jewish ethnic distribution in Israel, 123 unrelated heterozygotes (A20210G) from different Jewish ethnic groups, and 10 unrelated homozygotes (20210A), 6 from Israel and 4 non-Jewish subjects (2 from Austria and 2 from France). In homozygotes for the 20210A genotype, the frequency of one of the alleles for each marker was 0.95 to 1.00, whereas in normal homozygotes (20210G), the frequency was only 0.45 to 0.84 (see underlined alleles in Table 2). It was remarkable to note that in heterozygotes bearing one 20210A allele, the frequency of the depicted polymorphic alleles was very high (0.77 to 0.96). The differences in allele frequencies for each marker between 110 normals and 133 subjects bearing 20210A (123 heterozygotes and 10 homozygotes) were highly significant (P = .0007 or less). This degree of linkage disequilibrium strongly suggested that the prothrombotic polymorphism occurred only once on the haplotype 3728T-4125C-8845G-19911A.

We then calculated the frequency of the founder haplotype in normal alleles and in alleles bearing the factor II G20210A polymorphism. Because we were unable to amplify the marker in intron D in DNA samples from 2 20210A homozygotes, only 16 alleles of the homoyzgotes were analyzed. As can be seen in Table 3, only 28% of 220 alleles derived from normals bore the founder haplotype, which was in striking contrast to 64% of 123 20210A alleles derived from heterozygotes and 94% of 16 alleles of homozygotes (P < .0001). Similarly, the founder haplotype was present in 34% of the combined 343 20210G alleles (of normals and heterozygotes) and in 68% of the combined 139 20210A alleles (of heterozygotes and homozygotes). However, the linkage between the prothrombotic polymorphism and the other polymorphisms could not be established in 46% of the alleles in normals and in 36% of the alleles in heterozygotes. Similar values of undefinable alleles were observed among 20210G alleles derived from normals and heterozygotes (43%) and 20210A alleles combined from heterozygotes and homozygotes (32%).

To increase the number of definable alleles, we took advantage of the proximity of the dimorphic marker at nt 19911 and the prothrombotic polymorphism at nt 20210. A 345-bp PCR-amplified product containing both polymorphic sites and specific restriction analysis with HindIII and EcoNI enabled us to phase the linkage of these two sites in all alleles examined (Fig 3). Full linkage of 19911A and 20210A was observed in all 123 heterozygotes (A20210G). Similarly, with the exception of one allele that probably represents a recombination event (see below), 19 of 20 alleles of 20210A homozygotes had nt A at position 19911. Thus, of 143 alleles (123 + 20) bearing the prothrombotic polymorphism, 142 (>99%) shared the same marker in intron M, in contrast to its frequency of 0.45 in normals (Table 2).

Detection of a recombination in a homozygote for 20210A. In one 20210A homozygous Jewish proband, the genotype was 3728T-G4125C-8845G-G19911A, which was different from the founder haplotype at nts 4125 and 19911. Direct sequencing with specific restriction analysis with HindIII and EcoNI enabled us to phase the linkage of these two sites in all alleles examined (Fig 3). Full linkage of 19911A and 20210A was observed in all 123 heterozygotes (A20210G). Similarly, with the exception of one allele that probably represents a recombination event (see below), 19 of 20 alleles of 20210A homozygotes had nt A at position 19911. Thus, of 143 alleles (123 + 20) bearing the prothrombotic polymorphism, 142 (>99%) shared the same marker in intron M, in contrast to its frequency of 0.45 in normals (Table 2).

Search for polymorphism in the promoter region of the factor II gene. Because the mechanism by which the change in nt 20210 at the 3’ untranslated region of the factor II gene is associated with an elevated plasma level of factor II is dubious,

---

**Table 1. Prevalence of Factor II G20210A Polymorphism in Jewish Ethnic Groups in Israel**

<table>
<thead>
<tr>
<th>Ethnic Group</th>
<th>Subjects (N)</th>
<th>Heterozygotes N</th>
<th>Allele Frequency</th>
<th>95% CI of Allele Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashkenazi</td>
<td>464</td>
<td>31</td>
<td>6.7</td>
<td>0.033</td>
</tr>
<tr>
<td>North African</td>
<td>273</td>
<td>15</td>
<td>5.5</td>
<td>0.028</td>
</tr>
<tr>
<td>Iraqi</td>
<td>247</td>
<td>10</td>
<td>4.0</td>
<td>0.020</td>
</tr>
<tr>
<td>Iranian</td>
<td>199</td>
<td>4</td>
<td>2.0</td>
<td>0.010</td>
</tr>
<tr>
<td>Yemenite</td>
<td>310</td>
<td>3</td>
<td>1.0</td>
<td>0.005</td>
</tr>
<tr>
<td>Ethiopian</td>
<td>177</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
</tr>
</tbody>
</table>

---

**Table 2. Allele Frequencies of 4 Dimorphic Markers of the Factor II Gene in Normals and Subjects Bearing G20210A**

<table>
<thead>
<tr>
<th>Polymorphism</th>
<th>Allele*</th>
<th>Normals (N = 110)</th>
<th>Heterozygotes (N = 123)</th>
<th>Homozygotes (N = 10)</th>
<th>χ² Analysis (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intron D</td>
<td>T</td>
<td>0.75</td>
<td>0.88</td>
<td>1.00†</td>
<td>.0007</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.25</td>
<td>0.12</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Intron E</td>
<td>C</td>
<td>0.65</td>
<td>0.67</td>
<td>0.95</td>
<td>≪.0001</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.35</td>
<td>0.13</td>
<td>0.05</td>
<td>≪.0001</td>
</tr>
<tr>
<td>Exon 10</td>
<td>G</td>
<td>0.84</td>
<td>0.96</td>
<td>1.00</td>
<td>≪.0001</td>
</tr>
<tr>
<td></td>
<td>A</td>
<td>0.16</td>
<td>0.04</td>
<td>0.00</td>
<td>≪.0001</td>
</tr>
<tr>
<td>Intron M</td>
<td>A</td>
<td>0.45</td>
<td>0.77</td>
<td>0.95</td>
<td>≪.0001</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>0.55</td>
<td>0.23</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

*The underlined nt are those present in the founder haplotype.
†Based only on 8 homozygotes.
a search for a linked sequence variation in the promotor region seemed warranted.

Poort et al. [1] found no sequence variations in approximately 400 bp upstream of the prothrombin gene transcription site in 28 homozygous normals (20210G) and 5 heterozygotes (G20210A). We extended the search for a possible alteration in sequence to nucleotides −350 to −1050 of the promotor region that contains regulatory elements. DNA samples from 8 normal homozygotes for 20210G and 2 homozygotes for 20210A were amplified by PCR and the 721-bp fragments were subjected to digestion by 3 restriction enzymes (Rsa I, Mbo II, and Pvu II), as described. SSCP analysis of the cleaved fragments showed no aberrant bands suggestive for polymorphisms (data not shown). The 721-bp fragments obtained from one homozygote (20210A) and one normal homozygote (20210G) were cloned into a TA vector and sequenced. Only one sequence variation, insertion of G at position 646, was found in both subjects. Thus, our data and the data of Poort et al. [1] failed to show a change in the −1,050-bp region of the promotor that might be linked with 20210A.

Similarly to factor V G1691A, our findings are consistent with a common Caucasian founder for the factor II G20210A prothrombotic polymorphism. However, unlike factor V G1691A, for which APC resistance is the assigned mechanism for VTE, the mechanisms by which factor II G20210A is linked with 20210A.

REFERENCES
18. Dilley A, Hooper WC, Austrin H, Lally C, Wenger NK, Evatt BL:
The prevalence of the prothrombin 20210 G → A variant in African Americans. Blood 90:652a, 1997 (abstr, suppl 1)


A Single Genetic Origin for the Common Prothrombotic G20210A Polymorphism in the Prothrombin Gene

Ariella Zivelin, Nurit Rosenberg, Shlomit Faier, Nurit Kornbrot, Hava Peretz, Christine Mannhalter, Marie Helene Horellou and Uri Seligsohn

Updated information and services can be found at:
http://www.bloodjournal.org/content/92/4/1119.full.html

Articles on similar topics can be found in the following Blood collections

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml