Hereditary Thrombophilia as a Model for Multigenic Disease

Edwin G. Bovill,¹ Sandra J. Hasstedt,² Mark F. Leppert,² and George L. Long³

Departments of Pathology¹ and Biochemistry², University of Vermont, Burlington, VT, and
Department of Human Genetics³, University of Utah, Salt Lake City, UT, USA

Introduction

Nearly 150 years ago, Virchow postulated that thrombosis was caused by changes in the flow of blood, the vessel wall, or the composition of blood. This concept created the foundation for subsequent investigation of hereditary and acquired hypercoagulable states. This review will focus on an example of the use of modern genetic epidemiologic analysis to evaluate the multigenic pathogenesis of the syndrome of juvenile thrombophilia.

Juvenile thrombophilia has been observed clinically since the time of Virchow and is characterized by venous thrombosis onset at a young age, recurrent thrombosis, and a positive family history for thrombosis. The pathogenesis of juvenile thrombophilia remained obscure until the Egeberg observation, in 1965, of a four generation family with juvenile thrombophilia associated with a heterozygous antithrombin deficiency subsequently identified as antithrombin Oslo (G to A in the triplet coding for Ala 404).¹² The association of a hereditary deficiency of antithrombin III with thrombosis appeared to support the hypothesis, first put forward by Astrup in 1958, of a thrombohemorrhagic balance.³ He postulated that there is a carefully controlled balance between clot formation and dissolution and that changes in conditions, such as Virchow’s widely encompassing triad, could tip the balance toward thrombus formation.

The importance of the thrombohemorrhagic balance in hypercoagulable states has been born out of two lines of investigation: evidence supporting the tonic activation of the hemostatic mechanism and the subsequent description of additional families with antithrombin deficiency and other genetically abnormal hemostatic proteins associated with inherited thrombophilia. Assessing the activation of the hemostatic mechanism in vivo is achieved by a variety of measures, including assays for activation peptides generated by coagulation enzyme activity. Activation peptides, such as prothrombin fragment 1+2, are measurable in normal individuals, due to tonic hemostatic activity and appear elevated in certain families with juvenile thrombophilia.⁴

In the past 25 years since Egeberg’s description of antithrombin deficiency, a number of seemingly monogenic, autosomal dominant, variably penetrant hereditary disorders have been well established as risk factors for venous thromboembolic disease. These disorders include protein C deficiency, protein S deficiency, antithrombin III deficiency, the presence of the factor V Leiden mutation, and the recently reported G20210A prothrombin polymorphism.⁵⁶ These hereditary thrombophilic syndromes exhibit considerable variability in the severity of their clinical manifestations. A severe, life-threatening risk for thrombosis is conferred by homozygous protein C or protein S deficiency, which if left untreated, leads to death.⁷⁻⁸ Homozygous antithrombin III deficiency has not been reported but is also likely to be a lethal condition. Only a moderate risk for thrombosis is conferred by the homozygous state for factor V Leiden or the G20210A polymorphism.⁹¹⁰ In contrast to homozygotes, the assessment of risk in heterozygotes, with these single gene disorders, has been complicated by variable clinical expression in family members with identical genotypes.¹¹ Consideration of environmental interactions has not elucidated the variability of clinical expression. Consequently, it has been postulated that more than one genetic risk factor may co-segregate with a consequent cumulative or synergistic effect on thrombotic risk.¹²

A number of co-segregating risk factors have been described in the past few years. Probably the best characterized interactions are between the common factor V Leiden mutation, present in 3% to 6% of the Caucasian population,¹³¹⁴ and the less common deficiencies of protein C, protein S, and antithrombin III. The factor V Leiden mutation does not, by itself, confer increased risk of thrombosis. The high prevalence of the mutation, however, creates ample opportunity for interaction with other risk factors when present.

The G20210A prothrombin polymorphism has a prevalence of 1% to 2% in the Caucasian population and, thus, may play a similar role to factor V Leiden. A number of small studies have documented an interaction of G20210A with other risk factors.¹⁵⁻¹⁷ A limited evaluation of individuals with antithrombin III, protein C, or protein S deficiency revealed a frequency of 7.9% for the G20210A polymorphism, as compared to a frequency of 0.7% for controls.¹⁸ The G20210A polymorphism was observed in only 1 of the 6 protein C-deficient patients.¹⁸

Correspondence to: Edwin G. Bovill, MD, Department of Pathology, University of Vermont, Burlington, VT 05405, Fax: 802-656-8892.
In the present state, the elucidation of risk factors for venous thromboembolic disease attests to the effectiveness of the analytical framework constructed from the molecular components of Virchow’s triad, analyzed in the context of the thrombohemorrhagic balance hypothesis. Two investigative strategies have been used to study thrombophilia: clinical case-control studies and genetic epidemiologic studies. The latter strategy has gained considerable utility, based on the remarkable advances in molecular biology over the past two decades. Modern techniques of genetic analysis of families offer important opportunities to identify cosegregation of risk factors with disease.\textsuperscript{19} The essence of the genetic epidemiologic strategy is the association of clinical disease with alleles of specific genes. It is achieved either by the direct sequencing of candidate genes or by demonstration of linkage to genetic markers.

Example of the Multigenic Pathogenesis of Thrombophilia

The evaluation of large families represents the most robust application of the genetic epidemiologic strategy and will be the focus of the following discussion and example. In major single-gene disorders, the inheritance of a single mutant allele confers a high probability for exhibiting the phenotype. The challenge in multigene disorders lies in the fact that families have generally been evaluated based on the presence of a variably penetrant mutant allele associated with thrombosis. This can leave the investigator with a daunting array of candidate genes for the possible modifying locus or loci.

An example of variable penetrance is demonstrated in Figure 1, which shows the observed inheritance pattern in a selected branch of a large, previously described\textsuperscript{11,20,21} protein C-deficient kindred of French Canadian extraction. The protein C deficiency is due to a mutation in the protein C gene, leading to a coding frame shift at His107Pro with premature termination of protein C translation and resultant type I protein C deficiency. It is evident from the figure that not all carriers of the His107Pro mutation experience thrombosis and that thrombotic episodes occur primarily, but not exclusively, in mutation carriers.

An initial problem posed in the search for a cosegregating modifying locus is whether there is, indeed, a second modifying locus or multiple modifying loci. This question was addressed by the use of two-locus segregation analysis in the 283 studied members of this large French Canadian family.\textsuperscript{21} A second thrombosis susceptibility gene, interacting with protein C deficiency, was inferred (p<0.00005), and environmental factors were excluded (p<0.0005). Penetrances estimated from the segregation analysis model showed that protein C deficiency interacting with the unknown susceptibility gene resulted in a penetrance, by 60 years of age, in 79% of men and 99% of women.

Previous phenotypic evaluation of a variety of candidate genes in this family was uninformative. Consequently, a gene mapping strategy was elected to localize the unknown modifying locus in the genome. Figure 2 outlines the phases of gene mapping as reviewed in Gelehrter et al.\textsuperscript{22} The first step in gene

### Table 1. Observed LOD Scores for Selected Candidate Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>LOD *</th>
<th>Chromosome</th>
<th>Marker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antithrombin III</td>
<td>-7.30</td>
<td>1</td>
<td>UT417</td>
</tr>
<tr>
<td>Factor V</td>
<td>-8.11</td>
<td>1</td>
<td>D1S2681</td>
</tr>
<tr>
<td>Protein S</td>
<td>-5.85</td>
<td>3</td>
<td>D3S1595</td>
</tr>
<tr>
<td>Prothrombin</td>
<td>-7.67</td>
<td>11</td>
<td>D11S1349</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>-6.38</td>
<td>4</td>
<td>FGA.PCR2</td>
</tr>
<tr>
<td>Tissue factor</td>
<td>-0.32</td>
<td>1</td>
<td>D1S497</td>
</tr>
<tr>
<td>Thrombomodulin</td>
<td>-1.71</td>
<td>20</td>
<td>D20S871</td>
</tr>
<tr>
<td>TPA</td>
<td>-2.90</td>
<td>8</td>
<td>PLAT.PCR2</td>
</tr>
<tr>
<td>PAI-I</td>
<td>-6.87</td>
<td>7</td>
<td>UT682</td>
</tr>
<tr>
<td>von Willebrand factor</td>
<td>-6.46</td>
<td>12</td>
<td>VWA1(TCTA)</td>
</tr>
<tr>
<td>MTHFR</td>
<td>-2.66</td>
<td>222/223</td>
<td>MTHFR</td>
</tr>
<tr>
<td>Interleukin 1</td>
<td>-5.71</td>
<td>1</td>
<td>222/223</td>
</tr>
</tbody>
</table>

*All LOD scores at θ\textsuperscript{**} = 0. LOD is the log likelihood of linkage. LOD scores of 3 or greater are considered statistically significant evidence for linkage. LOD scores below -2 exclude linkage.

\textsuperscript{**} θ = theta = genetic recombination frequency; a measure of genetic distance.\textsuperscript{22}

![Figure 1. Selected members of the large protein C-deficient kindred\textsuperscript{21} that demonstrate marked differences in penetrance of clinical disease in different wings of the family. Unaffected individuals, whose current age is not given, were not studied.](image-url)
mapping is genetic mapping, in which the meiotic recombination frequencies between loci are used to determine the distance between loci. The closer the loci become on a given chromosome, the more likely they will remain linked together during meiotic recombination. Thus, this phase of gene mapping is referred to as general linkage analysis. Positioning of an unknown modifying locus by general linkage analysis depends on the availability of a large number of polymorphic DNA markers whose relative position in the genome is known.

The demonstration of linkage also depends on the allelic heterogeneity of the markers, with greater heterogeneity improving discrimination at any locus. The Human Genome Project has mapped an enormous number of polymorphic markers, including variable number of tandem repeats (VNTRs) and microsatellite or short tandem repeats (STRs) within the genome. Analyzing the segregation of these markers within a disease pedigree can localize an unknown disease gene within 5 to 10 centiMorgans (5,000 to 10,000 kb). The measure of significance, indicating the presence of linkage, is the LOD score or “logarithm of the odds” ratio. The LOD score comprises the logarithm of the ratio of

\[
\text{Likelihood of observing pedigree data at varying degrees of recombination} \\
\text{Likelihood of observing pedigree data if there is random recombination}
\]

A LOD score of 3 is conventionally accepted as evidence of linkage (e.g., evidence for linkage is 1,000-fold greater than evidence against linkage). Conversely, a LOD score of –2 is accepted as evidence against linkage.

If linkage is observed, the second phase of gene mapping is physical mapping. As shown in Figure 2 and reviewed by Gelehrter, physical mapping utilizes an array of molecular biology and cytogenetic techniques to identify the physical location of the unknown gene on the chromosome. This review will focus on the utility of linkage analysis in the identification or exclusion of candidate genes.

The large protein C-deficient family described above affords an illustrative example of the power of using genetic epidemiologic strategies to search for modifying loci. The hemostatic mechanism is complex, and there are numerous candidate genes. Individual sequencing of the candidate genes is feasible but expensive and time consuming. The advantage of the evaluation of a single, large kindred is that the members of the family can be assumed to carry the same allelic variant of the candidate gene. By focusing on individual branches of the kindred exhibiting high prothrombotic penetrance, this approach is further strengthened, thus decreasing allelic heterogeneity. Figure 3 demonstrates an informative subset constructed from this French Canadian family, which shows high prothrombotic penetrance. This subset has an estimated maximum LOD score of 8 for demonstrating the unknown susceptibility gene, thus predicting a high likelihood of success.

The hemostasis genes listed in Table 1 were screened as possible candidates for the unknown susceptibility gene using STR markers and DNA from the informative subset shown in Figure 3 (n=31). Linkage was tested using likelihood analysis, and the results are summarized in Table 1. None of these candidates demonstrated linkage to disease, as reflected in the uniformly negative LOD scores.

Because of the importance of factor V Leiden and the G20210A prothrombin polymorphism as risk factors, further analysis was performed on both. Factor V Leiden had a prevalence of only 0.6% in the family and showed no relationship to disease. Poort demonstrated that individual carriers of the G20210A prothrombin polymorphism, not only have a 2.8-fold increased risk of thrombosis, they also have higher plasma prothrombin concentrations than noncarriers. In the original description of the G20210A variant, prothrombin plasma concentrations greater than 115 IU/\text{DL} increased the risk of thromboembolic disease two-fold. The pathophysiological significance of this latter observation remains to be determined.

Genotypes for the G20210A variant and the His107Pro mutation were available from 279 members of this French Canadian family. The prevalence of the G20210A variant was 13% in the kindred, which contrasts with the approximate 2% prevalence in the general Caucasian population. Thus, the G20210A variant appeared to be a promising candidate for a modifying locus. Relative risk analysis by the case-control method is not valid in a family study setting, so the Transmission Disequilibrium Test of Spielman and Ewens, as extended to pedigrees, was employed to test for excess transmission of G20210A or His107Pro to thrombosis cases. Using this test, a transmission of 0.5, specifying no effect, was obtained. Table 2 shows that, although the His107Pro mutation was overtransmitted to thrombosis cases in this family, the G20210A variant was not. Inspection of Figure 3 shows there was ample opportunity for interaction between the His107Pro mutation and the G20210A variant in a subset of the family with high prothrombotic penetrance.

Measured genotype analysis was used to evaluate the relationship between prothrombin plasma concentration and both the G20210A variant and thrombosis. The prothrombin concentration in carriers of G20210A was increased from 97±1% to
Table 2. Genetic Transmission* of G20210A and His107Pro Mutations by Thrombosis Diagnosis

<table>
<thead>
<tr>
<th>Thrombosis</th>
<th>G 20210A</th>
<th>His107Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verified</td>
<td>0.490 ± 0.130</td>
<td>0.837 ± 0.075 (p&lt; 0.001)</td>
</tr>
<tr>
<td>Verified + Suspected</td>
<td>0.505 ± 0.120</td>
<td>0.716 ± 0.076 (p&lt; 0.01)</td>
</tr>
</tbody>
</table>

*Transmission disequilibrium test results showing the estimate +/- standard error of tau, the probability of the transmission of mutations to different disease designations.

Figure 3. An informative subset chosen from the large French Canadian protein-C deficient kindred because of high prothrombotic penetrance. Clinical disease status, presence of the His107Pro mutation in protein C, and presence of the G20210A prothrombin polymorphism are displayed.

Conclusions

Thus, genetic epidemiologic analysis of a large thrombophilic kindred with well-documented protein C deficiency strongly supports the hypothesis that thromophilia is a multigenic disease. This example demonstrates the power of modern genetic epidemiologic analysis using the tools of molecular biology. Although the complete genome scan and general linkage study has not been completed, much has been learned from this family. Initial segregation analysis strongly supports the presence of a discreet second modifier locus interacting with the His107Pro protein C mutation, which increases the risk of venous thromboembolic disease. A wide range of possible candidate genes has been excluded by linkage analysis using STR markers adjacent to or within the candidate genes. This method of excluding the candidate genes is considerably more efficient than mutation analysis, which is expensive, time consuming, and may miss subtle alterations in areas, such as promoter regions outside the coding area of the gene. In this family, important modifier loci, such as factor V Leiden and the G20210A prothrombin polymorphism, were excluded both by linkage analysis of STRs close to the respective genes and by direct analysis for the specific mutations. In the case of factor V Leiden, the mutation was shown to be exceptionally rare in the family. Since the fami-
ily is of French Canadian ancestry, this may have been due to genetic drift. In contrast, the G20210A variant is highly prevalent in the family and yet does not confer additional risk. This runs counter to our expected result and raises interesting questions about the role of the interaction of this risk factor with the type I protein C deficiency observed in this family.

Humans are remarkably diverse in their biology and behavior. Although influenced by the environment, human diversity is largely determined by each individual’s genetic endowment. Unraveling the contribution of genes to common, complex multigenic disorders, such as atherosclerosis, diabetes, and hypertension, is exceptionally difficult, because of the sheer number of genes interacting in these polygenic disorders. Oligogenic diseases afford a unique opportunity to study simpler models of gene–gene interactions. This is especially the case when two discreet genes are involved, such as in some families with thrombophilia. As this review points out, modern genetic epidemiology, together with the tools of molecular biology, have created an unparalleled opportunity to sort out the pathogenesis of the thrombophilia as a model for other oligogenic, and more complex polygenic, diseases. In conclusion, identification and careful genetic and biochemical study of large kindreds with juvenile thrombophilia will not only pay off in better care for those affected with thrombophilia, but it will also help to improve our understanding of the far more common endemic polygenic diseases.

Acknowledgements

The authors gratefully express their appreciation to Julia Valliere, Bruce Scott, PhD, and Peter Cullas, PhD, for their help in the laboratory and in the field; Kenneth Bauer, MD, for years of collaboration and performance of prothrombin assays; and Mrs. Joanne Teel for manuscript preparation. This work was supported in part by the Public Health Service National Heart, Lung and Blood Institute, Grant R01 HL46703 “Surface dependent reactions in thrombosis and thrombolysis” and by NIH/NICHD, Grant HD17463.

References