Received: 29 June 2020 | Revised: 30 September 2020 | Accepted: 12 October 2020

DOI: 10.1002/rth2.12456

BRIEF REPORT

# Carrier frequencies of antithrombin, protein C, and protein S deficiency variants estimated using a public database and expression experiments

Keiko Maruyama PhD | Koichi Kokame PhD

Department of Molecular Pathogenesis, National Cerebral and Cardiovascular Center, Suita, Japan

**Correspondence**
Koichi Kokame, Department of Molecular Pathogenesis, National Cerebral and Cardiovascular Center, 6-1 Kishibe-Shimmachi, Suita, Osaka 564-8565, Japan.
Email: kame@ncvc.go.jp

**Funding information**
This work was supported in part by the Ministry of Health, Labour, and Welfare of Japan; by Japan Society for the Promotion of Science (JSPS); and by the Takeda Science Foundation. KM was a Research Fellow of JSPS during this study.

**Handling Editor:** Neil Zakai

## Abstract

**Background:** Genetic deficiencies of antithrombin (AT), protein C (PC), and protein S (PS) are risk factors for venous thromboembolism. In the general population, the prevalence of heterozygous deficiency of AT, PC, and PS are reported as approximately 0.02%-0.2%, 0.2%-0.4%, and 0.03%-0.5%, respectively. The Exome Aggregation Consortium (ExAC) provides a public database containing reference data for over 60 000 exomes.

**Objective:** This study aimed to determine the frequency of AT, PC, and PS deficiencies using the ExAC database and transient expression experiments.

**Methods:** In total, 133, 157, and 221 variants of *SERPIN1* (encoding AT), *PROC* (PC), and *PROS1* (PS), respectively, were registered as missense and putative loss-of-function variants in the ExAC database. Variants with relatively high allele frequencies were selected and randomly sampled. Recombinant proteins were expressed in human embryo kidney 293 cells and their secretion and anticoagulant activities examined.

**Results and Conclusion:** We assessed 9 AT, 4 PC, and 14 PS variants with relatively high allele frequencies and randomly sampled 12 AT, 15 PC, and 19 PS missense variants. All 21 AT variants showed normal or mildly reduced secretion, and 6 showed reduced total activity (specific activity × antigen level). Of the 19 PC variants, 11 showed impaired total activity. All 33 PS variants showed normal or mildly reduced secretion, and 4 showed reduced total activity. Based on allele frequencies in the ExAC database, we calculated the frequencies of AT, PC, and PS genetic deficiency as 0.36%, 0.63%, and 0.39%, respectively.

KEYWORDS

antithrombin, genetic deficiency, protein C, protein S, venous thromboembolism

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2020 Research and Practice in Thrombosis and Haemostasis published by Wiley Periodicals LLC on behalf of International Society on Thrombosis and Haemostasis (ISTH).

**Essentials**

- Antithrombin (AT), protein C (PC), and protein S (PS) deficiencies are risks for thrombophilia.
- Exome Aggregation Consortium data and expression experiments were used to estimate deleterious allele frequencies.
- Secretion and activity of 21 AT, 19 PC, and 33 PS selected variants were assessed.
- Estimated carrier frequencies of the deleterious alleles were 0.36% (AT), 0.63% (PC), and 0.39% (PS).

## 1 | INTRODUCTION

Venous thromboembolism (VTE) is a common and serious disorder afflicting millions of individuals annually worldwide. VTE is a multifactorial disease in which acquired or inherited predispositions of thrombosis interact with various risk factors.[1] Deficiencies in natural anticoagulant proteins antithrombin (AT), protein C (PC), and protein S (PS) severely increase the risk of VTE. The serine protease inhibitor AT regulates blood coagulation and inactivates multiple coagulation factors, such as thrombin, activated factor XIa, activated factor Xa (FXa), activated factor IXa, and activated factor VIIa. In contrast, PC and PS are vitamin K–dependent plasma glycoproteins that exert their anticoagulant actions through the degradation of coagulation factors activated factor VIIIa, and activated factor Va. PS also serves in the efficient inhibition of FXa as a cofactor of tissue factor pathway inhibitor.

Although the pathophysiology and risk factors of VTE have been long studied, it remains difficult to estimate the precise prevalence of inherited AT, PC, and PS deficiencies. In patients with suspected hereditary VTE (juvenile and recurrent events), the prevalence of AT, PC, and PS deficiencies have been reported as 2%-5%, 5%-10%, and 5%-10%, respectively.[2] Among the general Caucasian population, the prevalence of AT, PC, and PS deficiencies are estimated to be 0.02%-0.2%, 0.2%-0.4%, and <0.5%, respectively.[2] The AT protein is encoded by *SERPINC1* (7 exons), PC by *PROC* (9 exons), and PS by *PROS1* (15 exons). Although more than 200 missense and nonsense variants have been registered in the Human Gene Mutation Database, their allele frequencies in the general population are unclear.

The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a wide variety of large-scale sequencing projects and to make summary data available to a wider scientific community. The data provided through ExAC included 60 706 unrelated individuals whose genomes were sequenced for various disease-specific and population genetic studies with frequencies provided for each variant.[3]

In the current study, ExAC data and expression experiments were used to estimate the allele frequencies of AT, PC, and PS genetic deficiencies that may cause thrombophilia.

## 2 | METHODS

### 2.1 | Sampling of ExAC data for nonsynonymous variants

In total, 133 *SERPIN1*, 157 *PROC*, and 221 *PROS1* missense and putative loss-of-function variants with allele counts have been registered in the ExAC database version 1.0.[3] The variants come from seven regional populations, including African, East Asian, European (non-Finnish), Finnish, Latino, South Asian, and Other. Variants with relatively high allele frequencies (≥0.1%) and allele counts ≥2 in each region were selected as a representative subgroup for subjection to expression experiments and included 9 *SERPIN1*, 4 *PROC*, and 14 *PROS1* variants. In addition, we randomly selected approximately 10% of the remaining variants of each gene, regardless of frequency and allele count. This included 12 *SERPIN1*, 15 *PROC*, and 19 *PROS1* variants.

### 2.2 | Mutagenesis and construction of expression vectors

The human AT expression vector ORF-NM_000488-pcDNA3 was generously provided by Dr Eriko Morishita (Kanazawa University). The human PC expression vector ORF-NM_000312-pCMV-SPORT6 was purchased from RIKEN BRC (Tsukuba, Japan). The human PS

**TABLE 1** The numbers of variants retrieved from ExAC data

| | | | Allele frequency | | Type | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Protein | Gene | Total | ≥0.1% and ≥2 counts | Less | Missense | In-frame insertion | In-frame deletion | Nonsense | Frame shift | Splicing defect | Start loss | Stop loss |
| AT (464 a.a.) | *SERPINC1* | 133 | 9 | 124 | 131 | 0 | 1 | 0 | 0 | 0 | 1 | 0 |
| PC (461 a.a.) | *PROC* | 157 | 4 | 153 | 147 | 1 | 2 | 4 | 1 | 1 | 0 | 1 |
| PS (676 a.a.) | *PROS1* | 221 | 14 | 207 | 206 | 0 | 0 | 4 | 5 | 6 | 0 | 0 |

Abbreviation: ExAC, Exome Aggregation Consortium.

expression vector was previously described.[4] In total, we constructed 73 variants using a site-directed mutagenesis technique and verified their sequences using an Applied Biosystems 3500xL genetic analyzer (Life Technologies, Rockville, MD, USA). Two clones of each variant were selected for expression experiments and are described below.

### 2.3 | Preparation of recombinant AT, PC, and PS

Human embryo kidney 293 cells were transiently transfected with each expression vector using Lipofectamine 3000 (Thermo Fisher Scientific, Waltham, MA, USA). For measuring expression levels in the culture supernatants, transfected cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM; Wako Pure Chemical, Osaka, Japan) supplemented with 10% fetal bovine serum (FBS) for 72 hours. For PC and PS expression, 10 μg/mL vitamin $K_1$ was added to the culture media. Culture supernatants were centrifuged to remove cell debris. For measuring AT and PC activity, the cells were cultured for 72 hours in OptiMEM I Reduced Serum Medium (Thermo Fisher Scientific). The culture supernatants were centrifuged to remove cell debris, concentrated using Amicon Ultra Centrifugal Filters (Sigma-Aldrich, St. Louis, MO, USA), and exchanged against reaction buffer (50 mmol/L Tris-HCl, 100 mmol/L NaCl, pH 7.4). For measuring PS activity, the cells were cultured in DMEM supplemented with 10% FBS for 72 hours, and His-tagged PS molecules in the supernatant were purified using Talon Metal Affinity Resin (Takara Bio, Mountain View, CA, USA). PS was eluted with elution buffer (20 mmol/L Tris-HCl, 300 mmol/L NaCl, and 250 mmol/L imidazole) and dialyzed against reaction buffer.

### 2.4 | Measurement of AT, PC, and PS expression levels

Antigen levels of AT and PC in culture media were measured using Matched-Pair Antibody Sets for ELISA (Affinity Biologicals, Ancaster, ON, Canada). Antigen levels of PS were measured using an in-house sandwich ELISA with a polyclonal rabbit antihuman PS primary antibody (Agilent Dako, Carpinteria, CA, USA) and horseradish peroxidase–labeled anti-human PS secondary antibody (Affinity Biologicals).[4] Although the protein S ELISA used can detect both C4b-bound and free (unbound) protein S, all recombinant protein S was free in the present study. ELISA standards were prepared using purified human AT, PC, and PS (Haematologic Technologies, Essex Junction, VT, USA). Results are shown as means of duplicate measurements relative to their respective recombinant wild-type (WT) protein.

### 2.5 | Measurement of activity of AT, PC, and PS

Anti-FXa activity of AT was determined using a Testzym S ATIII kit (Sekisui Diagnostics, Framingham, MA, USA) with a modified protocol. Briefly, 1.2 μL of test solution was mixed with 40 μL of FXa solution and 95 μL of reaction buffer, incubated for 3 minutes, and 20 μL of substrate solution then added. Absorbance was measured at 405 nm against a reference wavelength of 505 nm at 1-minute intervals for 10 minutes at 37°C. Anti-thrombin activity of AT was similarly determined using a Berichrom ATIII kit (Siemens Healthcare Diagnostics, Deerfield, IL, USA). Briefly, 2.5 μL of test solution was mixed with 20 μL of thrombin solution and 130 μL of reaction buffer, incubated for 30 seconds, and 25 μL of substrate solution then added. Absorbance was measured at 405 nm at 1-minute intervals for 10 minutes at 37°C. PC activity was determined using a Staclot Protein C kit (Diagnostica Stago, Parsippany, NJ, USA) according to the manufacturer's protocol. APC cofactor activity of PS was determined using a Staclot Protein S kit (Diagnostica Stago) using the manufacturer's protocols. Standard curves for all assays were generated using recombinant WT proteins of AT, PC, and PS. Results are shown as means of duplicate measurements relative to their respective WT proteins. Specific activities were determined by dividing total activity in culture supernatant by antigen level.

### 2.6 | In silico analysis

Three bioinformatics tools, PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/),[5] PROVEAN Protein (http://provean.jcvi.org/seq_submit.php),[6] and MutationTaster2 (http://www.mutationtaster.org)[7] were used for in silico analysis.

## 3 | RESULTS AND DISCUSSION

### 3.1 | Sampling of ExAC data for nonsynonymous variants

We retrieved 131 missense, 1 in-frame deletion, and 1 start loss variants of *SERPINC1*; 147 missense, 1 in-frame insertion, 2 in-frame deletion, 4 nonsense, 1 frameshift, 1 splicing defect, and 1 stop loss variant of *PROC*; and 206 missense, 4 nonsense, 5 frameshift, and 6 splicing defect variants of *PROS1* (Table 1). We then selected 21 AT, 19 PC, and 33 PS variants using the criteria described in the Methods section.

### 3.2 | Variant expression and activity

Culture-supernatant antigen levels of all AT variants were >50% compared to that of WT except for AT-D141V (Table 2), suggesting the variants had no severe expression defects. Anti-FXa–specific activity of AT variants R56C and E413K were <50% of WT and anti-thrombin–specific activity of R56C and A416S were <50% of WT. Both expression (supernatant antigen levels) and specific activities of the variants should be considered when evaluating their function in plasma. Therefore, we emphasized total activity (specific

**TABLE 2** Characteristics of AT (*SERPINC1*) variants sampled from ExAC data

| Protein | Transcript | Ag (%) | Anti-FXa sAct (%) | Anti-thrombin sAct (%) | Anti-FXa tAct (%) | Anti-thrombin tAct (%) | PolyPhen-2 | PROVEAN | MutationTaster2 | Reported to be causative |
|---|---|---|---|---|---|---|---|---|---|---|
| WT | … | 100 | 100 | 100 | 100 | 100 | … | … | … | … |
| V30E[a] | c.89T>A | 127.6 (± 22.5) | 97.9 (± 0.8) | 111.4 (± 4.2) | 124.9 | 142.1 | Possibly damaging | Neutral | Polymorphism | 10 |
| D38N | c.112G>A | 101.3 (± 10.9) | 92.7 (± 0.9) | 104.9 (± 3.7) | 93.9 | 106.3 | Probably damaging | Neutral | Disease causing | |
| I54V | c.160A>G | 105.4 (± 15.3) | 93.8 (± 0.1) | 96.5 (± 6.5) | 98.8 | 101.7 | Benign | Neutral | Disease causing | |
| R56C | c.166C>T | 56.7 (± 10.7) | 29.0 (± 1.3) | 8.8 (± 0.1) | 16.5 | 5.0 | Probably damaging | Deleterious | Disease causing | 11 |
| P73L[a] | c.218C>T | 103.4 (± 6.2) | 96.7 (± 3.8) | 56.3 (± 4.7) | 99.9 | 58.2 | Probably damaging | Deleterious | Disease causing | 12 |
| N105K | c.315T>A | 88.8 (± 3.4) | 104.8 (± 2.2) | 109.4 (± 6.4) | 93.1 | 97.2 | Benign | Neutral | Polymorphism | |
| D141V | c.422A>T | 45.3 (± 1.8) | 62.1 (± 3.3) | 54.8 (± 5.3) | 28.1 | 24.8 | Probably damaging | Deleterious | Disease causing | |
| T147A[a] | c.439A>G | 94.3 (± 4.9) | 100.7 (± 1.2) | 92.3 (± 6.2) | 95.0 | 87.0 | Possibly damaging | Deleterious | disease causing | 13 |
| R177C[a] | c.529C>T | 57.4 (± 2.6) | 105.3 (± 2.0) | 107.5 (± 0.0) | 60.5 | 61.7 | Probably damaging | Deleterious | Disease causing | |
| T185A[a] | c.553A>G | 102.4 (± 3.6) | 105.4 (± 2.0) | 103.8 (± 8.7) | 108.0 | 106.3 | Benign | Neutral | Polymorphism | |
| E212K | c.634G>A | 76.9 (± 0.4) | 92.2 (± 2.1) | 73.8 (± 1.2) | 71.0 | 56.8 | Benign | Deleterious | Disease causing | |
| N240S[a] | c.719A>G | 88.8 (± 4.5) | 105.4 (± 2.0) | 95.7 (± 2.7) | 93.6 | 85.0 | Benign | Neutral | Polymorphism | 14 |
| T250A | c.748A>G | 129.5 (± 7.0) | 84.0 (± 1.5) | 106.1 (± 7.2) | 108.8 | 137.4 | Benign | Neutral | Disease causing | |
| Y292F | c.875A>T | 103.5 (± 4.4) | 105.4 (± 2.4) | 106.6 (± 9.7) | 109.1 | 110.3 | Probably damaging | Neutral | Disease causing | |
| R293Q[a] | c.878G>A | 112.1 (± 17.5) | 106.0 (± 1.9) | 108.5 (± 7.5) | 118.9 | 121.7 | Possibly damaging | Neutral | Polymorphism | |
| P305H[a] | c.914C>A | 63.9 (± 15.1) | 105.3 (± 1.8) | 91.5 (± 4.1) | 67.2 | 58.4 | Probably damaging | Deleterious | Disease causing | |
| T312I | c.935C>T | 56.3 (± 8.5) | 94.9 (± 10.4) | 82.8 (± 3.4) | 53.5 | 46.7 | Probably damaging | Deleterious | Disease causing | |
| R356C | c.1066C>T | 78.1 (± 11.8) | 103.0 (± 0.6) | 97.1 (± 7.4) | 80.5 | 75.9 | Possibly damaging | Deleterious | Disease causing | |
| E413K | c.1237G>A | 76.9 (± 4.4) | 47.3 (± 0.5) | 115.4 (± 1.9) | 36.4 | 88.8 | Probably damaging | Deleterious | Disease causing | |
| A416S[a] | c.1246G>T | 99.1 (± 7.1) | 54.2 (± 6.5) | 13.9 (± 5.0) | 53.6 | 13.8 | Probably damaging | Deleterious | Disease causing | 15 |
| T418I | c.1253C>T | 72.3 (± 15.5) | 57.9 (± 3.8) | 79.0 (± 0.6) | 41.9 | 57.1 | Probably damaging | Deleterious | Disease causing | |

*Note:* All data are the means (± range) of duplicate measurements.

Abbreviations: AT, antithrombin; Ag, antigen; ExAC, Exome Aggregation Consortium; sAct, specific activity; tAct, total activity (Ag X sAct).

[a]Variants with relatively high allele frequencies in ExAC data (≥0.1% and ≥2 allele counts).

**TABLE 3** Characteristics of PC (*PROC*) variants sampled from ExAC data

| Protein | Transcript | Ag (%) | sAct (%) | tAct (%) | PolyPhen-2 | PROVEAN | MutationTaster2 | Reported to be causative |
|---|---|---|---|---|---|---|---|---|
| WT | ... | 100 | 100 | 100 | ... | ... | ... | ... |
| K70R | c.209A>G | 61.7 (± 19.4) | 96.0 (± 3.8) | 59.2 | Benign | Neutral | Polymorphism | |
| F181V | c.541T>G | 21.3 (± 1.1) | ND | ... | Benign | Deleterious | Disease causing | [16] |
| R189W[a] | c.565C>T | 73.3 (± 1.9) | 60.6 (± 0.7) | 44.4 | Benign | Neutral | Polymorphism | [17] |
| K193Q | c.577A>C | 137.0 (± 4.3) | 55.5 (± 0.0) | 76.1 | Possibly damaging | Neutral | Polymorphism | |
| K193del[a] | c.577_579delAAG | 151.6 (± 0.9) | 51.1 (± 0.6) | 77.5 | - | Deleterious | Polymorphism | [18] |
| L197P | c.590T>C | 61.3 (± 15.5) | 106.3 (± 3.1) | 65.1 | Possibly damaging | Neutral | Polymorphism | |
| E205_Q207dup | c.614_622dup AAGACCAAG | 141.1 (± 0.2) | 44.0 (± 1.7) | 62.1 | ... | Neutral | Polymorphism | |
| P210L | c.629C>T | 11.7 (± 2.8) | ND | ... | Possibly damaging | Neutral | Disease causing | [17] |
| R220Q | c.659G>A | 14.6 (± 6.6) | ND | ... | Benign | Neutral | Disease causing | [18] |
| A251V[a] | c.752C>T | 19.5 (± 0.6) | ND | ... | Probably damaging | Deleterious | Disease causing | [19] |
| D297N | c.889G>A | 47.4 (± 1.5) | 92.3 (± 3.0) | 43.7 | Probably damaging | Deleterious | Disease causing | |
| A309T | c.925G>A | 7.2 (± 1.4) | ND | ... | Possibly damaging | Neutral | Disease causing | [20] |
| E349D | c.1047G>C | 84.0 (± 1.4) | 98.1 (± 7.2) | 82.4 | Benign | Neutral | Polymorphism | |
| V368I | c.1102G>A | 41.4 (± 6.1) | 117.4 (± 15.9) | 48.6 | Benign | Neutral | Disease causing | |
| N371D[a] | c.1111A>G | 35.1 (± 11.2) | 102.1 (± 7.3) | 35.8 | Benign | Deleterious | Polymorphism | |
| G392R | c.1174G>A | 3.9 (± 0.1) | ND | ... | Probably damaging | Deleterious | Disease causing | [21] |
| M406I | c.1218G>A | 3.7 (± 0.5) | ND | ... | Probably damaging | Deleterious | Disease causing | [22] |
| V434I | c.1300G>A | 68.1 (± 2.6) | 101.3 (± 2.9) | 69.0 | Possibly damaging | Neutral | Polymorphism | |
| S458R | c.1374C>G | 71.3 (± 7.4) | 130.5 (± 7.4) | 93.0 | Benign | Neutral | Polymorphism | |

*Note:* All data are the means (± range) of duplicate measurements.

Abbreviations: Ag, antigen; ExAC, Exome Aggregation Consortium; ND, not detectable; PC, protein C; sAct, specific activity; tAct, total activity (Ag X sAct).

[a]Variants with relatively high allele frequencies in ExAC data (≥0.1% and ≥2 allele counts).

activity × antigen level) relative to WT AT. Anti-FXa total activities of R56C, D141V, E413K, and T418I were <50% of WT and anti-thrombin total activities of R56C, D141V, T312I, and A416S were <50% of WT (Table 2). Therefore, these six AT variants were considered loss-of-function alleles.

For PC, 10 variants (F181V, P210L, R220Q, A251V, D297N, A309T, V368I, N371D, G392R, and M406I) showed impaired secretion (<50% of WT; Table 3). The specific activity of E205_Q207dup was <50% of WT, and those of 7 variants (F181V, P210L, R220Q, A251V, A309T, G392R, and M406I) could not be measured because of severely impaired secretion. Therefore, the total activities of 11 variants (F181V, R189W, P210L, R220Q, A251V, D297N, A309T, V368I, N371D, G392R, and M406I) were considered to be <50% of WT, that is, loss-of-function.

Culture-supernatant antigen levels of all PS variants were >50% compared to that of WT PS (Table 4). The specific activities of D129G and F273L were <50% of WT. The total activities of four variants (E60K, V191D, F273L, and V510M) were <50% of WT and considered loss-of-function alleles.

## 3.3 | In silico prediction of variant impact

Three in silico tools were used to predict whether the variants could be pathogenic due to loss of function (Tables 2-4). Of the 73 variants examined, 32 showed unanimous predictions using all three tools, with 12 predicted as normal (benign, neutral, or polymorphism) and 20 as pathogenic (probably damaging, deleterious, or disease causing). The concordance rate between the expression experiments and in silico analyses was low. Frequencies of AT, PC, and PS deficiencies were estimated using the expression experiment results and ExAC data.

**TABLE 4** Characteristics of PS (*PROS1*) variants sampled from ExAC data

| Protein | Transcript | Ag (%) | sAct (%) | tAct (%) | PolyPhen-2 | PROVEAN | MutationTaster2 | Reported to be causative |
|---|---|---|---|---|---|---|---|---|
| WT | … | 100 | 100 | 100 | … | … | … | … |
| R40L[a] | c.119G>T | 270.5 (± 6.7) | 53.9 (± 2.8) | 145.8 | Probably damaging | Deleterious | Disease causing | 23 |
| E60K | c.178G>A | 66.0 (± 5.1) | 58.0 (± 0.5) | 38.3 | Probably damaging | Deleterious | Disease causing | |
| P76L[a] | c.227C>T | 91.9 (± 3.3) | 85.8 (± 1.7) | 78.9 | Probably damaging | Deleterious | Disease causing | 8 |
| D129G | c.386A>G | 107.6 (± 0.2) | 47.0 (± 2.8) | 50.5 | Benign | Neutral | Disease causing | 8 |
| T144N[a] | c.431C>A | 87.7 (± 5.7) | 77.9 (± 3.9) | 68.3 | Benign | Neutral | Polymorphism | 23 |
| Q150R | c.449A>G | 101.0 (± 1.4) | 93.4 (± 0.4) | 94.3 | Benign | Neutral | Disease causing | |
| N166H | c.496A>C | 93.0 (± 6.0) | 88.1 (± 2.4) | 82.0 | Probably damaging | Neutral | Polymorphism | |
| N168S[a] | c.503A>G | 109.2 (± 1.5) | 71.1 (± 0.8) | 77.6 | Possibly damaging | Deleterious | Disease causing | |
| D176H | c.526G>C | 99.0 (± 10.9) | 143.5 (± 3.4) | 142.0 | Possibly damaging | Neutral | Polymorphism | |
| P179R | c.536C>G | 89.6 (± 5.8) | 81.0 (± 2.1) | 72.6 | Possibly damaging | Deleterious | Disease causing | |
| V191D | c.572T>A | 84.1 (± 6.9) | 56.7 (± 3.3) | 47.7 | Benign | Neutral | Polymorphism | |
| R233K[a] | c.698G>A | 83.7 (± 2.7) | 80.7 (± 1.1) | 67.5 | Benign | Neutral | Polymorphism | 24 |
| F273L | c.819C>A | 79.2 (± 5.3) | 36.3 (± 1.4) | 28.8 | Benign | Neutral | Disease causing | 25 |
| E283D | c.849G>C | 91.8 (± 15.3) | 130.3 (± 0.4) | 119.6 | Possibly damaging | Neutral | Disease causing | |
| A307S | c.919G>T | 98.7 (± 9.5) | 88.9 (± 2.4) | 87.7 | Benign | Neutral | Polymorphism | |
| R316C | c.946C>T | 82.1 (± 10.1) | 62.9 (± 0.8) | 51.6 | Probably damaging | Deleterious | Disease causing | 26 |
| A341T[a] | c.1021G>A | 80.2 (± 12.4) | 81.0 (± 1.5) | 64.9 | Probably damaging | Deleterious | Disease causing | |
| I344F | c.1030A>T | 82.1 (± 2.0) | 104.4 (± 1.5) | 85.7 | Benign | Neutral | Polymorphism | |
| A348V | c.1043C>T | 81.4 (± 2.4) | 82.3 (± 0.8) | 67.0 | Possibly damaging | Neutral | Polymorphism | |
| N365K[a] | c.1095T>G | 65.2 (± 3.9) | 122.3 (± 0.0) | 79.8 | Probably damaging | Neutral | Disease causing | 9 |
| P410H[a] | c.1229C>A | 88.9 (± 7.8) | 81.7 (± 0.3) | 72.6 | Possibly damaging | Deleterious | Disease causing | 9 |
| P416L[a] | c.1247C>T | 98.9 (± 0.9) | 104.8 (± 0.3) | 103.6 | Benign | Neutral | Disease causing | |
| V425I | c.1273G>A | 94.5 (± 2.4) | 80.2 (± 6.2) | 75.8 | Benign | Neutral | Disease causing | |
| R432W | c.1294C>T | 95.4 (± 4.7) | 67.7 (± 1.8) | 64.6 | Probably damaging | Neutral | Disease causing | |
| E435G | c.1304A>G | 87.7 (± 5.8) | 88.2 (± 2.1) | 77.4 | Possibly damaging | Deleterious | Polymorphism | |
| P441L | c.1322C>T | 62.5 (± 6.0) | 88.2 (± 0.8) | 55.1 | Possibly damaging | Deleterious | Disease causing | |
| R445C | c.1333C>T | 109.8 (± 8.8) | 81.0 (± 2.9) | 88.9 | Probably damaging | Deleterious | Disease causing | |
| S501P[a] | c.1501T>C | 90.5 (± 6.0) | 80.8 (± 1.9) | 73.1 | Possibly damaging | Neutral | Disease causing | 27 |
| V510M[a] | c.1528G>A | 50.2 (± 3.3) | 99.4 (± 1.2) | 49.96 | Possibly damaging | Neutral | Disease causing | 28 |
| N530I | c.1589A>T | 78.3 (± 4.4) | 84.5 (± 0.7) | 66.1 | Possibly damaging | Neutral | Polymorphism | |
| T588A[a] | c.1762A>G | 97.4 (± 8.8) | 70.9 (± 0.9) | 69.0 | Possibly damaging | Neutral | Polymorphism | |
| D599H[a] | c.1795G>C | 107.4 (± 1.6) | 94.6 (± 0.1) | 101.6 | Probably damaging | Neutral | Polymorphism | 29 |
| T630I[a] | c.1889C>T | 81.8 (± 11.8) | 98.5 (± 1.9) | 80.6 | Probably damaging | Deleterious | Disease causing | 30 |

*Note:* All data are the means (± range) of duplicate measurements.

Abbreviations: Ag, antigen; ExAC, Exome Aggregation Consortium; PS, protein S; sAct, specific activity; tAct, total activity (Ag X sAct).

[a]Variants with relatively high allele frequencies in ExAC data (≥0.1% and ≥2 allele counts).

## 3.4 | Allele frequency estimation of AT, PC, and PS deficiencies

First, we estimated allele frequencies of AT deficiency. Of 133 AT variants in the ExAC data, 9 had relatively high allele frequencies of ≥1% with allele counts ≥2 (Table 1). Only one of the 9 frequent variants showed loss-of-function effects in the expression experiments (A416S) and had an allele frequency in the ExAC data of 0.0007172 (87 of 121 312 alleles). The 124 rare (nonfrequent) variants contained one start loss variant (c.3G>A) with an allele frequency of 0.0000083, which was predicted to be pathogenic. For the remaining 123 rare variants the mean allele frequency was 0.0000205. Of the randomly selected 12 variants, 5 (R56C, D141V, T312I, E413K, and T418I) showed loss-of-function effects

(Table 2). The total frequency of "loss-of-function alleles" was estimated as follows:

$$0.0007172 + 0.0000083 + [(0.0000205 \times 123) \times 5/12] = 0.0017761$$

Therefore, carriers of AT loss-of-function alleles in the general population were estimated as 35.5 of 10 000 individuals.

The frequency of PC deficiency was similarly estimated. Of 157 PC variants in the ExAC data, 4 had relatively high allele frequencies of ≥1% with allele counts ≥2. In the expression experiments, 3 of the frequent variants showed loss-of-function effects (R189W, A251V, and N371D) and had allele frequencies of 0.0007187, 0.0002246, and 0.0003802, respectively. The 153 rare variants included four nonsense (W83*, C101*, E274*, R348*), one frameshift (H149Pfs*13), one splicing defect (c.678+1G>A), and one stop loss (c.1384T>C) variants that were predicted to be pathogenic with a mean allele frequency of 0.0000142. The remaining 146 rare missense variants had a mean allele frequency of 0.0000219. Of the 15 randomly selected variants, 8 (F181V, P210L, R220Q, D297N, A309T, V368I, G392R, and M406I) showed loss-of-function effects (Table 3). The total frequency of "loss-of-function alleles" was estimated as follows:

$$0.0007187 + 0.0002246 + 0.0003802 + (0.0000142 \times 7) + [(0.0000219 \times 146) \times 8/15] = 0.0031282$$

Therefore, carriers of PC loss-of-function alleles in the general population were estimated as 62.5 of 10 000 individuals.

The frequency of PS deficiency was also estimated. Of 221 PS variants in the ExAC data, 14 had relatively high allele frequencies of ≥1% with allele counts ≥2. In the expression experiments, only one of the frequent variants showed loss-of-function effects (V510M) and had an allele frequency of 0.0010630. The 207 rare variants included 4 nonsense [L317*, Y485*, Y560* (c.1680T>A), Y560* (c.1680T>G)], 5 frameshift (C14Sfs*24, Y54Lfs*17, L45Yfs*42, N188Mfs*20, E437Gfs*3), and 6 splicing defect (c.77-1G>C, c.728-1G>A, c.728-1G>T, c.850-1G>A, c.1645-2A>G, c.1870+1G>T) variants with a mean allele frequency of 0.0000100. The remaining 192 rare missense variants had a mean allele frequency, 0.0000246. Of the 19 randomly selected variants, 3 (E60K, V191D, F273L) showed loss-of-function effects (Table 4). The total frequency of "loss-of-function alleles" was estimated as follows:

$$0.0010630 + (0.0000100 \times 15) + [(0.0000246 \times 192) \times 3/19] = 0.0019588$$

Therefore, carriers of PC loss-of-function variants in general were estimated as 39.1 of 10 000 individuals.

### 3.5 | Limitations

In the current study, we estimated the frequencies of AT, PC, and PS deficiencies using a simplified approach, which had several limitations. First, the ExAC database does not represent a general population and may cause some bias. Second, recombinant proteins expressed using cDNA vectors, and cell lines do not always properly reflect plasma proteins in humans. Some predicted normal functioning variants in our study have been reported in case reports as genetic causes of VTE (Tables 2-4). Third, expression experiment results may differ depending on materials and procedures used. In fact, our results of some variants differed from previous reports.[8,9] Fourth, the total activity threshold of 50% may not be appropriate to evaluate the variants. Since the cutoff value "50%" is for the total activity of each variant but not for the estimated plasma activity, the plasma activity of an individual with one "50%" mutant allele and one "100%" normal allele is theoretically 75% (25% + 50% activity in plasma). Fifth, sampling of 10% variants for expression experiments and extrapolating the results of the remaining 90% may cause some inaccuracy. Finally, other functional aspects that we did not examine, such as clearance from the circulation and binding affinity of partner proteins, may affect anticoagulant activities.

The present study highlighted the low concordance rate between expression experiments and in silico analyses, suggesting that clinical sequencing results may be difficult to interpret. There is growing interest in the clinical use of next-generation sequencing panels to identify inherited thrombophilia. How does clinical practice determine if a detected mutation is pathogenic? Neither phenotypic assays, gene sequencing, and phenotypic-genotyping correlations (interpretation of sequencing results) are complete. Moreover, the penetrance of these heritable thrombophilia seems to be far from 100%. Therefore, it is important to combine sequencing data with patient phenotypic assay data. In addition, it would be useful to obtain and examine sequencing and phenotypic assay data from the patient's family. These data can provide helpful information for a definitive diagnosis.

## 4 | CONCLUSION

Although our study had several limitations, our simplified approach was able to estimate the allele frequencies of AT, PC, and PS genetic deficiencies 0.36%, 0.63%, and 0.39%, respectively. In addition to this genetic basis, other factors such as penetrance, genetic modifiers, and environmental conditions may affect the onset of VTE.

### ACKNOWLEDGMENTS

We are grateful to Dr Eriko Morishita (Kanazawa University, Japan) for providing the human AT cDNA, Dr Masashi Akiyama for technical advice, and Ms Hiroko Ikejima for technical assistance.

### RELATIONSHIP DISCLOSURE

The authors have no conflicts of interest to disclose.

### AUTHOR CONTRIBUTIONS

KM performed the experiments and wrote the manuscript. KK conceived this study and wrote the manuscript. Both approved the final version of the manuscript.

## REFERENCES


1. Heit JA. Epidemiology of venous thromboembolism. Nat Rev Cardiol. 2015;12:464–74.
2. Coppola A, Tufano A, Cerbone AM, Di Minno G. Inherited thrombophilia: implications for prevention and treatment of venous thromboembolism. Semin Thromb Hemost. 2009;35:683–94.
3. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.
4. Maruyama K, Akiyama M, Miyata T, Kokame K. Protein S K196E mutation reduces its cofactor activity for APC but not for TFPI. Res Pract Thromb Haemost. 2018;2:751–6.
5. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.
6. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012;7:e46688.
7. Schwarz JM, Cooper DN, Schuelke M, Seelow D. Mutationtaster2: mutation prediction for the deep-sequencing age. Nat Methods. 2014;11:361–2.
8. Biguzzi E, Razzari C, Lane DA, Castaman G, Cappellari A, Bucciarelli P, et al. Molecular diversity and thrombotic risk in protein S deficiency: the PROSIT study. Hum Mutat. 2005;25:259–69.
9. Tang L, Jian XR, Hamasaki N, Guo T, Wang HF, Lu X, et al. Molecular basis of protein S deficiency in China. Am J Hematol. 2013;88:899–905.
10. Daly M, Bruce D, Perry DJ, Price J, Harper PL, O'Meara A, et al. Antithrombin Dublin (-3 Val → Glu): an N-terminal variant which has an aberrant signal peptidase cleavage site. FEBS Lett. 1990;273:87–90.
11. Perry DJ, Carrell RW. CpG dinucleotides are "hotspots" for mutation in the antithrombin III gene. Twelve variants identified using the polymerase chain reaction. Mol Biol Med. 1989;6:239–43.
12. Chang JY, Tran TH. Antithrombin III Basel. Identification of a Pro-Leu substitution in a hereditary abnormal antithrombin with impaired heparing cofactor activity. J Biol Chem. 1986;261:1174–6.
13. Picard V, Nowak-Göttl U, Biron-Andreani C, Fouassier M, Frere C, Goualt-Heilman M, et al. Molecular bases of antithrombin deficiency: twenty-two novel mutations in the antithrombin gene. Hum Mutat. 2006;27:600.
14. Zeng W, Tang L, Jian X-R, Li YQ, Guo T, Wang QY, et al. Genetic analysis should be included in clinical practice when screening for antithrombin deficiency. Thromb Haemost. 2015;113:262–71.
15. Perry DJ, Daly M, Harper PL, Tait RC, Price J, Walker ID, et al. Antithrombin Cambridge II, 384 Ala to Ser. Further evidence of the role of the reactive centre loop in the inhibitory function of the serpins. FEBS Lett. 1991;285:248–50.
16. Miyata T, Sakata T, Zheng YZ, Tsukamoto H, Umeyama H, Uchiyama S, et al. Genetic characterization of protein C deficiency in Japanese subjects using a rapid and nonradioactive method for single-stand conformational polymorphism analysis and a model building. Thromb Haemost. 1996;76:302–11.
17. Reitsma PH, Bernardi F, Doig RG, Gandrille S, Greengard JS, Ireland H, et al. Protein C deficiency: a database of mutations, 1995 update. On behalf of the Subcommittee on Plasma Coagulation Inhibitors of the Scientific and Standardization Committee of the ISTH. Thromb Haemost. 1995;73:876–89.
18. Miyata T, Sato Y, Ishikawa J, Okada H, Takeshita S, Sakata T, et al. Prevalence of genetic mutations in protein S, protein C and antithrombin genes in Japanese patients with deep vein thrombosis. Thromb Res. 2009;124:14–8.
19. Gandrille S, Aiach M. Identification of mutations in 90 of 121 consecutive symptomatic French patients with a type I protein C deficiency. Blood. 1995;86:2598–605.
20. Conard J, Horellou MH, Van Dreden P, Samama M, Reitsma PH, Poort S, et al. Homozygous protein C deficiency with late onset and recurrent coumarin-induced skin necrosis. Lancet. 1992;339:743–4.
21. Zheng YZ, Sakata T, Matsusue T, Umeyama H, Kato H, Miyata T. Six missense mutations associated with type I and type II protein C deficiency and implications obtained from molecular modelling. Blood Coagul Fibrinolysis. 1994;5:687–96.
22. Miyata T, Zheng YZ, Sakata T, Tsushima N, Kato H. Three missense mutations in the protein C heavy chain causing type I and type II protein C deficiency. Thromb Haemost. 1994;71:32–7.
23. Gandrille S, Borgel D, Eschwege-Gufflet V, Aillaud M, Dreyfus M, Matheron C, et al. Identification of 15 different candidate causal point mutations and three polymorphisms in 19 patients with protein S deficiency using a scanning method for the analysis of the protein S active gene. Blood. 1995;85:130–8.
24. Gandrille S, Borgel D, Sala N, Espinosa-Parrilla Y, Simmonds R, Rezende S, et al. Protein S deficiency: a database of mutations - Summary of the first update: for the Plasma Coagulation Inhibitors Subcommittee of the Scientific and Standardadization Committee of the International Society on Thrombosis and Haemostasis. Thromb Haemost. 2000;84:918.
25. Manderstedt E, Lind-Halldén C, Svensson P, Zöller B, Halldén C. Next-generation sequencing of 17 genes associated with venous thromboembolism reveals a deficit of non-synonymous variants in procoagulant genes. Thromb Haemost. 2019;119:1441–50.
26. Okada H, Yamazaki T, Takagi A, Murate T, Yamamoto K, Takamatsu J, et al. In vitro characterization of missense mutations associated with quantitative protein S deficiency. J Thromb Haemost. 2006;4:2003–9.
27. Duchemin J, Gandrille S, Borgel D, Feurgard P, Alhenc-Gelas M, Matheron C, et al. The Ser 460 to Pro substitution of the protein S alpha (PROS1) gene is a frequent mutation associated with free protein S (type IIa) deficiency. Blood. 1995;86:3436–43.
28. Borgel D, Duchemin J, Alhenc-Gelas M, Matheron C, Aiach M, Gandrille S. Molecular basis for protein S hereditary deficiency: genetic defects observed in 118 patients with type I and type IIa deficiencies. The French Network on Molecular Abnormalities Responsible for Protein C and Protein S Deficiencies. J Lab Clin Med. 1996;128:218–27.
29. Chan NCN, Cheng CK, Chan KCF, Wong CML, Lau KM, Kwong JHY, et al. Distinctive regional-specific PROS1 mutation spectrum in Southern China. J Thromb Thrombolysis. 2018;46:120–4.
30. Tsuda H, Urata M, Tsuda T, Wakiyama M, Iida H, Nakahara M, et al. Four missense mutations identified in the protein S gene of thrombosis patients with protein S deficiency: effects on secretion and anticoagulant activity of protein S. Thromb Res. 2002;105:233–9.


**How to cite this article:** Maruyama K, Kokame K. Carrier frequencies of antithrombin, protein C, and protein S deficiency variants estimated using a public database and expression experiments. *Res Pract Thromb Haemost*. 2021;5:179–186. https://doi.org/10.1002/rth2.12456