Journal of Bioinformatics and Computational Biology

A HYBRID-SYSTEM MODEL OF THE COAGULATION CASCADE: SIMULATION, SENSITIVITY, AND VALIDATION

JOSEPH G. MAKIN*

Department of Electrical Engineering and Computer Science, U.C. Berkeley/ International Computer Science Institute
Berkeley, CA 94720, USA
makin@icsi.berkeley.edu

SRINI NARAYANAN

International Computer Science Institute/ Div. of Cognitive Science & Inst. of Cognitive and Brain Sciences, U.C. Berkeley
1947 Center Street, Ste. 600
Berkeley, CA 94704, USA
snarayan@icsi.berkeley.edu

Received (Day Month Year)
Revised (Day Month Year)
Accepted (Day Month Year)

The process of human blood clotting involves a complex interaction of continuous-time/continuous-state processes and discrete-event/discrete-state phenomena, where the former comprise the various chemical rate equations and the latter comprise both threshold-limited behaviors and binary states (presence/absence of a chemical). Whereas previous blood-clotting models used only continuous dynamics and perforce addressed only portions of the coagulation cascade, we capture both continuous and discrete aspects by modeling it as a hybrid dynamical system. The model was implemented as a hybrid Petri net, a graphical modeling language that extends ordinary Petri nets to cover continuous quantities and continuous-time flows. The primary focus is simulation: (1) fidelity to the clinical data in terms of clotting-factor concentrations and elapsed time; (2) reproduction of known clotting pathologies; and (3) fine-grained predictions which may be used to refine clinical understanding of blood clotting. Next we examine sensitivity to rate-constant perturbation. Finally, we propose a method for titrating between reliance on the model and on prior clinical knowledge. For simplicity, we confine these last two analyses to a critical purely-continuous subsystem of the model.

Keywords: coagulation; hybrid systems; model validation

1. Introduction

The process of blood clotting in mammals is complicated, involving the interaction of more than a dozen coagulation factors as well as a number of proteins from

*Currently at the Center for Integrative Neuroscience, University of California, San Francisco.

the kinin-kallikrein system and protein inhibitors. Attempts to model coagulation mathematically therefore usually focus on a smaller subset of interactions, perhaps one of the so-called pathways or just a portion of one of them[20,10,4,11,23,5,22]. Such models generally consist of a set of coupled, usually nonlinear, differential equations governing the time evolution of protein concentrations. Although on one level of analysis all of the processes in blood clotting comprise discrete events (e.g., cleavage and formation of chemical bonds), at the scale of interest it is concentrations that matter, and these exhibit continuous dynamics.

Continuous systems are nevertheless inadequate to model the entire coagulation cascade, because (1) certain events are better modeled as "switched"[8] (e.g., reactions that take place if and only if calcium is present, of zinc ions exceed a threshold[24,25]); (2) the continuous reactions of the cascade take place on very different time scales, and replacing fast ones with switches avoids the numerical difficulties in simulating all of them at once; and (3) the current state of clinical knowledge is not sufficient to provide differential equations for all the interactions of the cascade, yet qualitative knowledge can nevertheless often be incorporated in the form of discrete events toggled (possibly) by thresholds (e.g., the thrombin concentration that is required for appreciable conversion of factor XIII into its activated form).

We therefore model the coagulation cascade as a hybrid system (HS), i.e. one consisting of interacting continuous and discrete dynamics. (See Ref. 14 for a thorough introduction.) We require the model to capture faithfully, and over a wide range of parameter settings, the major qualitative aspects of human blood clotting, but also that it be perspicuous and easily modifiable, to accommodate the biologist and clinician. In light of these constraints, the model was implemented using hybrid Petri nets (HPNs), a graphical modeling formalism for modeling hybrid systems. Classical Petri nets are a well known computational formalism for the modeling of discrete-event dynamical systems, with constructs for sequential and concurrent process execution, for resource consumption and production, and for inhibition. HPNs extend classical Petri nets from the domain of purely discrete phenomena to the domain of hybrid dynamics, by supplementing the traditional discrete-event architecture with continuously varying events and states. (See Chapter 3 of Ref. 15 for a formal and detailed description of HPNs. Our current implementation is based on the Visual Object Net++ platform, a dedicated HPN modeling and simulation environment.[6])

We show that the model can reproduce certain features from the clinical literature (time-to-clotting and the time course of thrombin, the most important coagulant) of (1) normal clotting, as well as of the coagulation disorders (2) hæmophilia A and (3) factor-V Leiden. More generally, the model provides a platform for *in silico* investigation of blood clotting, which is not presently fully understood. This includes the relevance of certain reactions, the nature of various clotting pathologies, and the effects of pharmacological interventions on those disorders. More broadly still, the model demonstrates the utility of using hybrid systems, and in particular

hybrid Petri nets, to model cascade-like biological processes where both discrete and continuous dynamics play a role. In virtue of its ability to incorporate both types of dynamics, the model is able to support robust analysis and prediction in cases where parts of a complex process may be known precisely (with differential equations) while other aspects may have qualitative descriptions only (through punctuated phase changes, discrete transitions, and threshold behaviors). This ability to reason effectively with representations of multiple granularities addresses a central requirement in modeling complex biological processes.

2. Model Implementation

Since the complete HPN is extremely large, it is omitted here, but the full model appears in Chapter 4 of Ref. 15, along with the blood factors and their initial concentrations ("ICs"; Table 2.1). Yet in spite of its size, the overall model comprises multiple instances of only five basic modules (along with a few other places and transitions), with different parameters for different pathways, factors, and enzymes. This motivates the use of an *object-oriented* HPN modeling language; that is, one with constructs for information hiding and object reuse. Two of the basic modules have purely continuous dynamics (blood factor activation and factor-factor binding); the other three have hybrid dynamics (the initiation of blood clotting in the intrinsic clotting pathway, the formation of a fibrin clot, and switched activation).

Blood-factor activation. Fig. 1A depicts one of the basic aspects of the coagulation cascade (see Fig. 3), the enzyme-induced transformation of a blood factor (which may be either a serine protease or a glycoprotein) from its inactive (zymogen) form to its active configuration. The places IN_1, IN_2 and OUT represent the concentration (nM) of various blood factors: IN_1 is the zymogen and OUT is its activated form; IN_2 is the catalyzing enzyme; IN_1:IN_2 is an intermediate macromolecule. The places labeled with k's are the rate constants of classic enzyme kinetics: on-, off-, and catalytic rates. This reaction can also be written as a set of differential equations; square brackets are used to indicate concentrations:

$$\frac{d[\mathrm{IN}_1]}{dt} = k_{\mathrm{off}}[\mathrm{IN}_1{:}\mathrm{IN}_2] - k_{\mathrm{on}}[\mathrm{IN}_1][\mathrm{IN}_2] \tag{1}$$

$$\begin{aligned}\frac{d[\mathrm{IN}_2]}{dt} =& k_{\mathrm{off}}[\mathrm{IN}_1{:}\mathrm{IN}_2] - k_{\mathrm{on}}[\mathrm{IN}_1][\mathrm{IN}_2] \\ &+ k_{\mathrm{cat}}[\mathrm{IN}_1{:}\mathrm{IN}_2]\end{aligned} \tag{2}$$

$$\begin{aligned}\frac{d[\mathrm{IN}_1{:}\mathrm{IN}_2]}{dt} =& k_{\mathrm{on}}[\mathrm{IN}_1][\mathrm{IN}_2] - k_{\mathrm{off}}[\mathrm{IN}_1{:}\mathrm{IN}_2] \\ &- k_{\mathrm{cat}}[\mathrm{IN}_1{:}\mathrm{IN}_2]\end{aligned} \tag{3}$$

$$\frac{d[\mathrm{OUT}]}{dt} = k_{\mathrm{cat}}[\mathrm{IN}_1{:}\mathrm{IN}_2]. \tag{4}$$

Of course, since the variables IN_1, IN_2, and OUT participate in other reactions, Eqs. 1, 2, and 4 do not completely define the dynamics of any of these variables; the complete governing equations may contain additional additive terms from other reactions. That is why these three places have shaded outer rings in Fig. 1A: it indicates that they are "published," i.e. available to interact with other objects and hence other reactions.

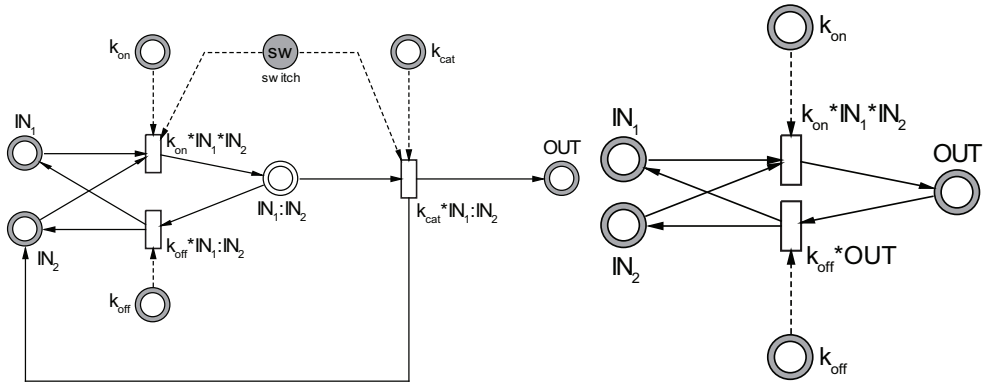


Fig. 1. Continuous time/state Petri net modules. (A) The activation module, which models the activation of a zymogen (IN_1) into its active configuration (OUT) by a catalyst (IN_2). In versions with the discrete switch *sw*, the module is hybrid, and models the switching on/off of a reaction by the presence/absence of a variable. (B) The binding module, which models the (reversible) binding of two components, IN_1 and IN_2, into the macromolecule OUT.

Fig. 1A also includes a discrete switch, *sw*, that can turn off the binding and catalytic reactions. Most instantiations of this module in the model lack the switch and are therefore purely continuous, but several of the activation reactions of the clotting cascade do require free calcium ions. In the corresponding hybrid module, the reactions will take place if and only if the *sw* place holds a token.

Factor-factor binding. The second recurring reaction is the binding of two blood factors (Fig. 1B). As in the activation reaction (Fig. 1A), the rates constants k_i are connected to transitions via "test arcs" (dashed lines), reflecting the fact that these quantities are unchanged by the reaction. The governing equations are simply:

$$\frac{d[\mathrm{IN}_1]}{dt} = k_{\mathrm{off}}[\mathrm{OUT}] - k_{\mathrm{on}}[\mathrm{IN}_1][\mathrm{IN}_2] \tag{5}$$

$$\frac{d[\mathrm{IN}_2]}{dt} = k_{\mathrm{off}}[\mathrm{OUT}] - k_{\mathrm{on}}[\mathrm{IN}_1][\mathrm{IN}_2] \tag{6}$$

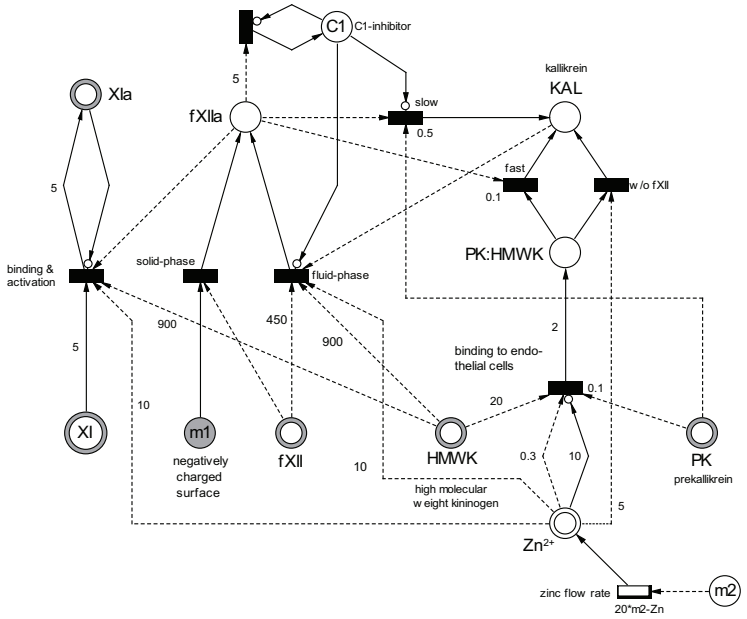
$$\frac{d[\mathrm{OUT}]}{dt} = k_{\mathrm{on}}[\mathrm{IN}_1][\mathrm{IN}_2] - k_{\mathrm{off}}[\mathrm{OUT}] \tag{7}$$

Both of the continuous modules appears in many instantiations throughout the model, differing from each other only in their rate constants and their interconnections with the rest of the network. The differential equations that they model were drawn from Ref. 4, and are summarized in Table 5.1 of Ref. 15.

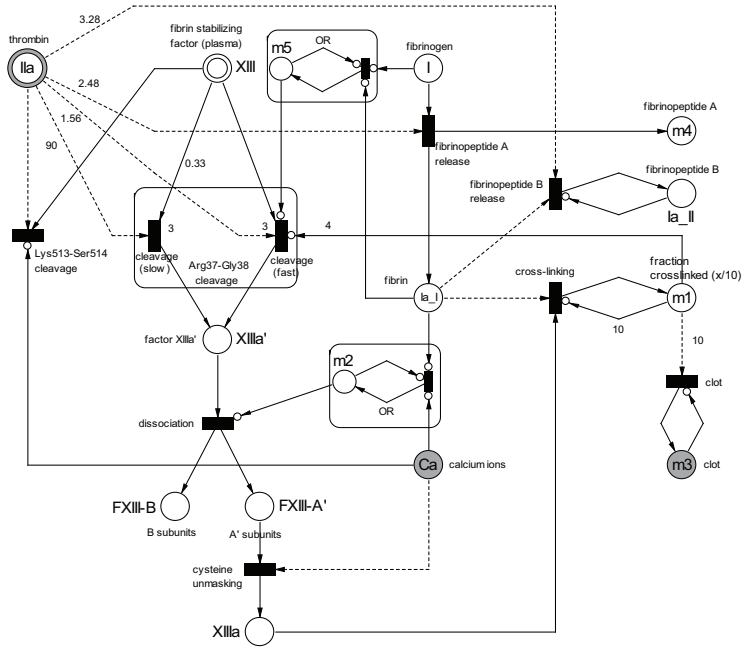
The intrinsic pathway. The INIT-INTRINSIC module, shown in Fig. 2A, models the initiation of blood clotting via the intrinsic pathway. Details of this process were drawn from Ref. 25 and Ref. 24; see also Table 1; we outline the main features here. The pathway begins with the exposure of a negatively charged surface and rising zinc concentrations, and ultimately activates factor XIa, through which it interacts with the rest of the cascade (see Fig. 3). Zinc-ion time evolution is modeled by a first-order differential equation (exponential growth up to an asymptote); various threshold concentrations trigger various intermediate reactions. The presence of the negatively charged surface, on the other hand, acts as a binary switch on solid-phase XII activation.

Notice that factor XIIa can activate kallikrein through either a "slow" or "fast" transition, where the former corresponds to activation of free prekallikrein and the latter to activation of prekallikrein bound to the surface of HMWK. The speeds of these reactions, fast and slow, are modeled by assigning appropriate time delays to the discrete transitions. Negative and positive feedback loops are also modeled: Activated kallikrein enables the fluid-phase activation of factor XII, which in turn activates more kallikrein. Conversely, a sufficient concentration factor XIIa inhibits its own activation via the serpin C1-inhibitor; sufficient quantities of factor XIa directly inhibit its own further activation by factor XIIa; and high zinc concentrations shits of the binding of high molecular weight kininogen to prekallikrein.

The fibrin module. Fig. 2B shows the FIBRIN module, which models the final portion of the blood clotting pathway: essentially the feedforward path from thrombin (IIa) to the formation of a clot. (Thrombin also takes part in numerous feedback reactions, which are modeled with the continuous modules lately discussed.) Data for this module were drawn from the clinical literature[19,7,27,12,2,18]; see Table 1 for details. Key features are the thrombin-thresholded intermediate reactions, modeled with test arcs; the fast and slow parallel pathways for cleavage of the Arg37-Gly38 bond of factor XIII (the fast cleavage requires lower levels of thrombin but additionally the presence of fibrin or fibrinogen); and the toggling of reactions by the binary presence/absence of calcium ions. Notice in particular that in addition to enabling certain reactions, calcium ions block the cleavage of the Lys513-Ser514 bond of factor XIII and therefore its inactivation. Negative feedback appears in the self-inhibiting production of fibrinopeptide B, and in the inhibitory effect of cross-linking of FXIIIa* and the fibrin polymers on the promoter action of fibrin and fibrinogen on factor XIII activation. The pathway terminates when cross-linking fraction reaches unity.



(A)



(B)

Fig. 2. Hybrid Petri net modules. (A) The initiation of the intrinsic pathway. (B) The final stages of blood clotting: the activation of factors I and XIII, and the formation of a clot. The semantics of hybrid Petri nets can be found in Chapter 3 of Ref. 15.

Table 1. Parameter values and types used in the HPN modules of Fig. 2A,B .

Parameter	Type	Description	Value
$[Zn^{2+}]$	arcweight	thresh. for HMWK:PK binding	0.3μM[25]
$[Zn^{2+}]$	arcweight	thresh. for PK activation on cells	5μM[24]
$[Zn^{2+}]$	arcweight	thresh. inhibition of HMWK:PK binding	10μM[25,24]
$[Zn^{2+}]$	arcweight	thresh. for FXII fluid-phase activation	10μM[25]
$[Zn^{2+}]$	arcweight	thresh. for FXI:HMWK binding	10μM[25]
Zinc flow rate	cont. transition	rate of zinc ion accumulation	$[\dot{Zn}] = 20 - [Zn]$[a]
[IIa]	arcweight	thresh. for fast cleavage of Arg37-Gly38	1.56nM[2b]
[IIa]	arcweight	thresh. for slow cleavage of Arg37-Gly38	90nM[2b]
[IIa]	arcweight	thresh. for release of fibrinopeptide A	2.48nM[2b]
[IIa]	arcweight	thresh. for release of fibrinopeptide B	3.28nM[2b]
[XIII]	arcweight/delay	fXIII cleaved/unit time: fast	1nM/3s[2b]
[XIII]	arcweight/delay	fXIII cleaved/unit time: slow	0.33nM/3s[2b]
cross-linking	arcweight	% cross-linking at which promoter effect of fibrin on Arg37-Gly38 cleavage is inhib.	40%[19]

[a] Extrapolated from intrinsic-pathway time data
[b] Interpolated

The complete cascade The portions of the coagulation cascade, Fig. 3, not modeled by the these two HPNs (Fig. 2) can be completely described in terms of the two continuous modules, factor-factor binding (Fig. 1B) and factor activation (Fig. 1A), and the latter's switched cousin (Fig. 1A). These include portions of the cascade not included in Fig. 3 for clarity, mostly the bindings of proteins to lipid substrates. The full model is depicted in Chapter 4 of Ref. 15.

Conceptually, the cascade is initiated through both the intrinsic and extrinsic pathways, the latter through the release of tissue factor (TF) and its binding to (lipid-bound) factor VII. Factor VII may be activated by factor Xa, but this requires completion of either the intrinsic or extrinsic pathways, both of which require the activation of factor VII itself. Thus it is generally accepted[4] that some very small quantity of factor VIIa (0.1 nM) exists in the blood stream prior to vascular injury. This does not imply spontaneous activation of the coagulation system since without tissue factor the rest of the pathway cannot proceed.

Factor X activation marks the meeting point of the extrinsic and intrinsic pathways, after which the cascade proceeds through a series of feedforward and feedback bindings and activations (which include inhibitions) to the most important blood protein, activated factor II (called thrombin, or IIa). Thrombin then initiates fibrin production along the lines described above.

3. Simulations

We choose to examine thrombin because it is the most important enzyme product of the coagulation cascade: it participates in far more reactions than any of the

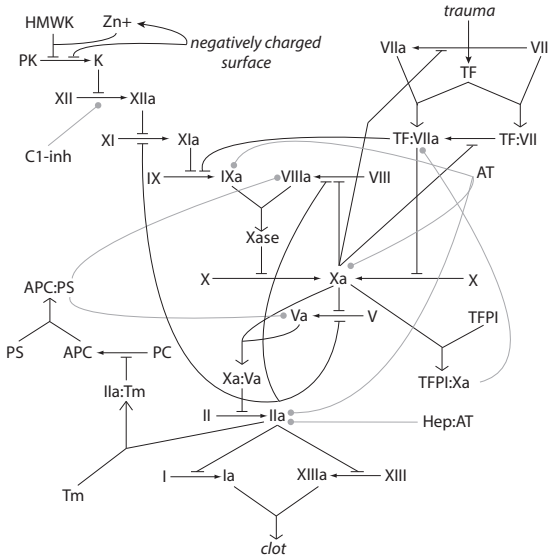


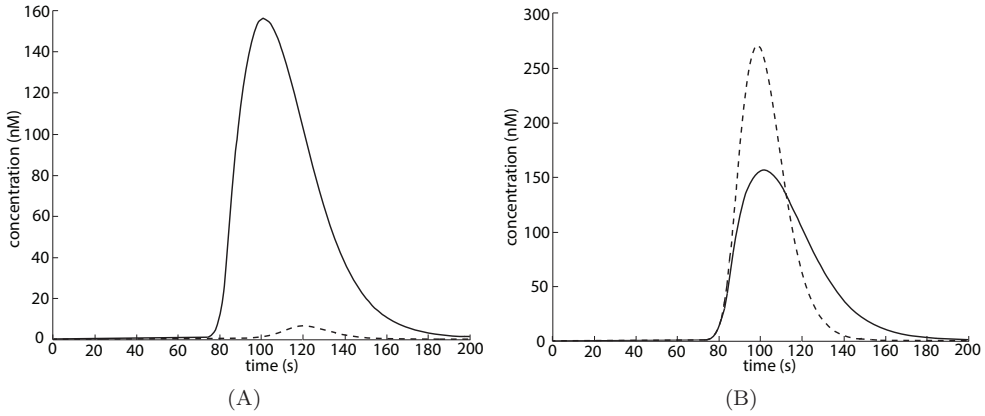
Fig. 3. The coagulation cascade. Converging arrows represent bindings; flat-ended lines terminating on triangle-headed arrows label activations; and gray, circle-headed lines denote inhibitory reactions, which may be either bindings or (de)activations. To minimize clutter, lipid bindings and other minor reactions have been omitted.

other factors, including both feedforward and feedback regulation, and is essential for normal blood clotting. Time to thrombin activation is consequently one of the major parameters measured in clinical tests. To evaluate the baseline performance of our model, we set the ICs as shown in Table 2.1 of Ref. 15 and compared the time course of thrombin production in our computational simulation with results reported in the clinical literature. Fig. 4A shows time-evolution of the thrombin concentration produced under normal conditions (solid line) upon triggering of the clotting cascade.

The subsystem of ODEs from extrinsic pathway to thrombin production had been partially validated against clinical data.[4] That comparison was appropriate because they both lack the contact pathway, the latter because clinical tests block factor-XII activation with corn-trypsin inhibitor (CTI). We therefore compare the effects of including the intrinsic pathway, Fig. 2A, on the thrombin profiles of that work.[4] The peak thrombin concentrations in the respective models is 160 nM (Fig. 4A) vs. 40 nM (Fig. 4 of Ref. 4). This is congruent with the empirically observed difference in peak thrombin of 162 vs. 74 nM when CTI is present (i.e., no intrinsic-pathway activation).[26] However, it should be noted that this was found at low initializing concentrations of tissue factor—0.5 pM vs. the 5 pM used in our model; and that at higher concentrations, the effect of the contact pathway is largely abolished.[26] (Fig. 4A) also shows that time to peak thrombin is 105 s—the same as in the pure ODE model.[4] This is consistent with the empirical finding that in-

creased negatively-charged contact surface, while affecting total thrombin amount, has no significant effect on lag time.[26]

Hæmophilia A, a common hæmoraphic disorder, results from deficiency of coagulation factor VIII. We therefore simulate it by setting the initial value of the continuous-HPN place VIII to 0.035 nM, which corresponds to the borderline between mild and moderate. Thrombin concentration peaks later (at 120 seconds) and much lower (at just over 6 nM); Fig. 4A, dashed line. It is more difficult to compare this directly with clinical results, since thrombin curves in hæmophiliacs are invariably measured with CTI (no intrinsic pathway). Nevertheless, the relative size and shapes of thrombin profiles for healthy and hæmophiliac patients (e.g., Fig. 1 of Ref. 30) match our data rather strikingly, both showing a reduction in peak thrombin by about 95%, with the latter peak occurring when healthy concentrations have fallen to 75% of their maximum. Likewise, the module of Fig. 2B yields maximum cross-linking of 40%, which is consistent with the finding that clots in severe hæmophiliac reach only about 55% of healthy densities.[3]



(A) (B)

Fig. 4. Simulations of the time course of thrombin concentrations (factor IIa) in nM. (A) Moderate-to-seve hæmophilia (dashed) and normal clotting (solid). (B) factor-V Leiden (dashed) and normal clotting (solid).

Factor-V Leiden is a coagulation disorder characterized by a condition called activated protein C resistance (APCR), in which a genetic mutation in the factor-V gene renders the resulting factor-V protein resistant to inactivation by activated protein C (APC). Factor V is a procoagulant, so the consequence of its slower rate of inactivation is generally a thrombophilic (propensity to clot) state. More sophisticated ways of simulating factor-V Leiden are available within the present model—we demonstrate one such elsewhere in the context of control schemes[16]—but for now, we choose simply to remove the module by which APC inactivates factor V (see Section 2). As expected, the result (Fig. 4B) is an increase in the

amount of thrombin produced. Again direct comparison with the clinical literature is impossible due to use of CTI, but two points of agreement can be noted: (1) our model predicts a 1.7-fold increase in peak thrombin, and a 2.6-fold increase has been reported; both model and data[28] predict no change in time to peak thrombin (see Fig. 6A).

4. Sensitivity Analysis

There are three distinct questions that can be addressed by analyzing the sensitivity of the model to its parameters. (1) The numerical values of these parameters were determined empirically, by chemical experiment, and are consequently subject to error; which errors have the biggest effect on model fidelity? (2) Suppose we wanted to change the course of coagulation by changing a reaction rate; how shall we achieve the biggest effect at minimum cost? (3) If a blood-clotting disease alters certain rate constants (as e.g. factor-V Leiden does), how dramatic will the effect be on the concentrations of blood proteins? For simplicity, we restrict the sensitivity analysis to a large portion of the cascade which *can* be described entirely by a set of ordinary differential equations; however, the sensitivity analysis can (in principle) and should be extended to the HPNs, as well.[9] The results must be interpreted accordingly.

We follow the technique employed on another model of blood clotting,[13] where a complete description can be found. Briefly, we derive (analytically, *contra* Ref. 13) the differential equation governing the time evolution of the sensitivities, $\sigma_{ij}(t) := \partial x_i/\partial k_j$, for all states x_i and parameters k_j, and then simulate it numerically in tandem with the ODEs for the state variables. A Euclidean norm is taken across time and (normalized) states to compute an *overall* sensitivity for each parameter. Since we are interested in the sensitivity of the system not only along the orthodox trajectory predicted by our model, but for nearby trajectories as well, we simulate the ODEs for a range of different parameter values: on each of 200 trials we independently draw each k_j from a uniform distribution between 50% to 150% of its nominal value and re-simulate the two systems of differential equations. The resulting overall sensitivity vectors $\hat{\sigma}$ are normalized at each trial into a $[0, 1]$ range, and averaged over trials.

Most sensitive parameters. Table 2 shows the 25 most sensitive rate constants, ranked according to average normalized sensitivity $\bar{\sigma}_j$. We note that 200 trials appears to be sufficient for the $\bar{\sigma}_j$ to converge.

How does this compare to the sensitivity analysis of Ref. 13, which models essentially the same portions of the cascade, but also includes equations for platelet binding? First of all, these platelet reactions account for about half their list of most "fragile" (sensitive) reactions (despite the relative rarity of platelet disorders in clinical observation). Second, the reactions involved in Xa:Va activation of thrombin are on average more sensitive in our model than the reactions for activation of factor

X by the TF:VIIa complex; whereas the converse is true in that work. In both, however, these reactions (and their related platelet reactions in Ref. 13) dominate the list.

Table 2. Averaged, normalized sensitivities (see text) for the 25 most influential rate constants, and their associated reactions.

Rank	Reaction	mean ± std	Rank	Reaction	mean ± std
1	Xa:Va:II ⇀ Xa:Va:mIIa	0.87 ± 0.17	15	XI + IIa ⇀ XI:IIa	0.46 ± 0.15
2	Xa:Va + II_L ⇀ Xa:Va:II	0.72 ± 0.20	16	V_L + $mIIa_L$ ⇀ V:mIIa	0.43 ± 0.13
3	TF:VIIa:X ⇀ TF:VIIa:Xa	0.71 ± 0.21	17	X_L + IXa:VIIIa ⇀ IXa:VIIIa:X	0.43 ± 0.13
4	TF + $VIIa_L$ ⇀ TF:VIIa	0.65 ± 0.20	18	$VIIIa_L$ + APC:PS ⇀	
5	Xa_L + Va_L ⇀ Xa:Va	0.65 ± 0.20		APC:PS:VIIIa	0.41 ± 0.13
6	Xa:Va + II_L ↽ Xa:Va:II	0.62 ± 0.19	19	IIa:Tm:PC ⇀ APC_L + IIa:Tm	0.40 ± 0.13
7	V:mIIa ⇀ Va_L + $mIIa_L$	0.56 ± 0.017	20	VIII:mIIa ⇀ $VIIIa_L$ + $mIIa_L$	0.39 ± 0.12
8	Xa_L + Va_L ↽ Xa:Va	0.53 ± 0.16	21	XI + IIa ↽ XI:IIa	0.38 ± 0.13
9	X_L + TF:VIIa ⇀ TF:VIIa:X	0.52 ± 0.18	22	V_L + $mIIa_L$ ↽ V:mIIa	0.38 ± 0.12
10	X_L + TF:VIIa ↽ TF:VIIa:X	0.51 ± 0.17	23	IX_L + XIa ⇀ XIa:IX	0.37 ± 0.11
11	XIa:IX ⇀ IXa_L + XIa	0.48 ± 0.15	24	IXa:VIIIa:X ⇀	
12	PC_L + IIa:Tm ⇀ IIa:Tm:PC	0.47 ± 0.14		Xa_L + IXa:VIIIa	0.36 ± 0.12
13	XI:IIa ⇀ XIa + IIa	0.46 ± 0.15	25	APC:PS:VIIIa ⇀	
14	IXa_L + $VIIIa_L$ ⇀ IXa:VIIIa	0.46 ± 0.014		$VIIIai_L$ + APC:PS	0.35 ± 0.10

Sensitivity and prothrombin time. One of the (two) most common clinical tests for a clotting problem is the prothrombin-time (PT) test, a measure of the time required for TF-initiated coagulation of blood plasma. Blood is drawn into a test tube containing citrate, which soaks up the free calcium ions and thus prevents the main clotting reactions; is centrifuged to separate out the blood cells; and is then activated by addition of an excess of calcium and tissue factor. Clotting time is measured from the addition of TF to the first visible signs of clot formation in the tube, i.e. when about 10 nM thrombin (factor IIa) has been produced[17]. The healthy range for this time is considered to be 12-15 seconds.

The PT test can be simulated in our model by initializing clotting with a saturating amount of tissue factor (20 nM, vs. 0.005 nM in the simulations above). Yet this yields a clot time around 30 seconds. If we assume the *structure* of the model is correct, then this discrepancy must be caused by parameter errors—which is plausible, since literature values for rate constants vary by upwards of orders of magnitude.[21,4,13]

Can this discrepancy be eliminated by small changes in parameters? No explicit solution is possible, but the sensitivity analysis allows a form of gradient ascent: we compute the sensitivity of thrombin over the first p seconds to changes in the rate constants, i.e. $\partial[\mathrm{IIa}]/\partial k_j$, and then change only the most sensitive parameters, moving them in the direction of greater thrombin. True gradient ascent would require the recalculation of the thrombin sensitivity at every step, but this requires

prohibitively expensive numerical solution (see Ref. 15 for details), so instead we average:

$$s_j^{[IIa]} = k_j \left(\frac{1}{x_i^*} \frac{\partial [\mathrm{IIa}]}{\partial k_j} \right) \Bigg|_{t=13},$$

the thrombin sensitivity at the time of the clot (13 s), over 200 numerical simulations of the governing ODEs, each with randomly perturbed parameters, as above. Also note that although we have again neglected the discrete parts of the network, we can do so with confidence, since (1) the PT test ignores the intrinsic pathway, and (2) we can estimate clotting time from thrombin alone under the assumption of healthy functioning of fibrin and factor XIII.

How few parameters can be altered and still achieve the desired PT, given an upper limit on those parameter shifts? Two interesting features of the results should be noted. First, allowing parameter deviations of 60%—plausible, in light of the dispersion of literature values—enables the model to match clinical PT times with the shifting of just 12 (out of over 100) parameters. This demonstrates the disproportionate impact of a small number of parameters. Limiting deviation from nominal values to just 40% precipitously increases this minimum number of rate constants, up to about two-thirds of them. Conversely, if all parameters are allowed to vary, the minimum variance that will achieve the desired PT is 38%. In sum, the effects of parameter variation are highly nonlinear, at least at $t = 13$ seconds and in the range of deviations examined: a small number of parameters ($\sim 1/10$) has significant effects on PT; the weakest 1/3 have little effect at all.

5. Overcoming the limitations of an imperfect model

As we have emphasized, the parameters and even structure of the model are subject to error; and although sensitivity analysis gives some insight into the relative importance of variation in parameters and narrows the scope for further clinical investigation, it does not *per se* increase the usefulness of the imperfect model. The ability to use an imperfect model precisely to the extent that it is accurate is in fact a very general desideratum for models of complex biological processes. Here we briefly present a novel approach to this problem, and some preliminary results. (For a longer discussion, see Chapter 7 of Ref. 15.) It exploits the fact that, in addition to the qualitative information already incorporated into the model through discrete states, thresholds, and the like, even more coarse-grained information is available.

We treat the model as a map from the space of ICs and parameters into an output space of "coagulatory degree," whose ten dimensions are the average thrombin concentration during consecutive 30-second intervals of the first 300 seconds of a clotting event. To avoid the "curse of dimensionality," we assumed that a deviant IC of each zymogen is the result of some independent and rare event, making the chances of more than two such events in one individual rare enough to ignore. (The Monte-Carlo approach also allows, in principle, exploitation of empirical relationships between pathological parameters, in place of these agnostic pairs.) Therefore,

for every pair of proteins with nonzero ICs (15/80), we simulated the 25 different combinations of deficiencies of 2%, 26.5%, 51%, 75.5%, and 100% (no deficiency). The continuous subsystem of the model was used to map these input data into output space, where they were then projected onto the most varying dimensions—in this case just two—via a singular-value decomposition, and then clustered with k-means, initialized at $k = 3$. The number of clusters (k) was then increased and the data re-clustered until three "exemplars"—normal coagulation, severe hæmophilia A (factor VIII at 2% of normal pre-injury levels, all others at 100%), and antithrombin deficiency (AT at 26.5% of normal; all others at 100%)—were segregated. The clustering is thus influenced both by the blood-clotting model and by our prior knowledge of IC/disease pairs (a less *ad hoc* approach is discussed below).

Finally, we learn the structure in the input-output map by constructing a decision tree: The input data (sets of ICs) are recursively partitioned into the two subsets that are maximally informative about output-cluster identity. The probability distribution of ICs was assumed to be independent and uniform (we explore alternatives below), with splitting based on Gini's diversity index[1] until each datum had been assigned to its own leaf.

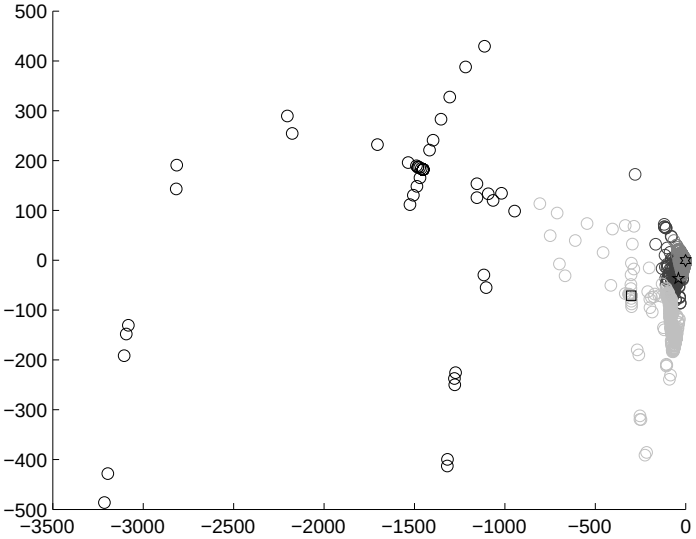


Fig. 5. An output space (units are arbitrary) in which hypo- (medium gray), hyper- (light gray), and normal (dark gray) coagulation are cleanly separated by k-means, with $k = 4$. The hexagram, square, and pentagram (resp.) are the known exemplars of the three classes. See text for an interpretation of the fourth class (black).

Results These results are preliminary, and are meant primarily to demonstrate the uses of the techniques just described. Using $k = 3$ clusters in the k-means algorithm pushed the hæmophiliac and normal vectors into the same cluster, which is obviously not acceptable. However, four clusters suffice to segregate all three exemplars; Fig. 5. For all members of the fourth cluster, thrombomodulin is at or below 25% —and in 90% of these cases, it is at 2%. We conclude that lack of thrombomodulin induces extreme hypercoagulation, in (literally) a class of its own.

The decision tree is shown in Fig. 6. Left branches correspond to concentrations being *less than* some amount. Since the input space was sampled only at pairwise deficiencies of ICs, only two left branches are ever required to reach a leaf. We note the following:

- The IC of protein C is most informative; *ceteris paribus*, it ought to be assayed first.
- Disease boundaries are perspicious: e.g., a patient with antithrombin deficiency no worse than 38% of normal levels (1317 nM) cannot have healthy clotting unless native concentrations of factor XI are also below about 87% (26 nM).
- Minimal adjustments (fewest number of protein concentrations to change) can be read off the tree: e.g., the hypercoagulatory consequences of protein-C deficiency can be rectified by decreasing native tissue-factor levels below 40% (0.002 nM) (as long as protein C concentrations are at least 38%).

6. Discussion and Conclusions

The simulations of factor-V Leiden and hæmophilia A were consistent with clinical results, and to that extent vindicate both the present model and its parameters as well as the methodology, i.e. a HS approach. Of course, consistency is not tantamount to quantitative identity, but here the obstacle lies not with the model so much as the state of the clinical data. Most thrombin curves can be fit by an equation with a mere four parameters,[29] so these data vastly underdetermine the parameters of the model. Including clot time does not substantially decrease the number of free parameters.

We interpret our simulations, then, as having demonstrated a method for incorporating qualitative or otherwise discrete information into a single model, which model is *consistent* with clinical findings; but we do not claim the model to be error-free. On the other hand, the model also provides a platform for testing clinical hypotheses, quantitative or qualitative, about the coagulation cascade, since consistency with known data is a necessary (if not sufficient) criterion for the correctness of those hypotheses. This provides a method for model refinement, as well. Furthermore, inasmuch as this kind of model can indeed capture all the relevant details of coagulation, the analysis techniques demonstrated in this paper will apply equally well to a "final," error-free model.

We also showed that (restricting our view to the continuous subsystem) relative changes in protein concentrations are most sensitive to changes in the rate constants involving activation of thrombin by the Xa:Va complex and activation of Xa by the TF:VIIa complex. These sensitivities can be exploited to change the model in the most parsimonious way in order to match the results of the PT test: shifting a mere 12 rate constants by as little as 60% sufficed to achieve a reasonable 15 seconds for this test. On the other hand, it turns out that using these parameters in the original clotting simulations of this study vitiates their agreement with clinical results, suggesting that something more subtle than rate constant alterations is responsible for the discrepancy in PT-test results.

The clustering analysis weakly validates the model, but more importantly provides a method for incorporating prior information about the relationships between ICs and disease states. Here we simply increased the number of clusters until the exemplars were properly segregated, but the clustering function could be parameterized more generically, and then optimized by (e.g.) gradient descent of the mean-square-error of exemplar misclassification. (See Chapter 7 of Ref. 15 for a more technical and general exposition.)

The decision tree generalizes over the mapping from ICs to clusters by providing efficient *summaries*. This makes it possible to read off which deficiencies are sufficient to induce pathology. The tree also tells us what factors are most informative about *disease states* (cf. the sensitivity analysis); the first three are protein C, antithrombin, and thrombomodulin. It is interesting to note that all three are inhibitors. The generalizations provided by the tree are merely "summaries" since they contain no more information than that in the ICs with their corresponding cluster labels, though they do provide that information in a much more useful form. However, replacing the non-uniform priors over the ICs with their observed clinical distributions would go beyond summary, since the tree would then reflect a compromise between the information given by the model/clustering and the priors. Predictions from this tree might be quite useful in a clinical setting. In principle, these could be integrated with costs in a decision-theoretic framework.

Acknowledgments

Jerry Feldman provided useful feedback on a draft of this paper.

References

1. Breiman L, Friedman J, Stone CJ, Olshen R, *Classification and Regression Trees*, Chapman & Hall, New York, 1984.
2. Brummel K, Paradis S, Butenas S, Mann K, Thrombin functions during tissue factor-induced blood coagulation, *Blood* **100**(1):148–152, 2002.
3. Brummel KE, Branda R, Butenas S, Mann KG, Discordant Fibrin Formation in Hemophilia, *Thrombosis and Haemostasis* **7**(5):825–832, 2009.
4. Bungay S, Gentry P, Gentry R, A mathematical model of lipid mediated thrombin generation, *Mathematical Medicine and Biology* **20**:105–129, 2003.

5. Butenas S, Orfeo T, Gissel M, Brummel K, Mann K, The significance of circulating factor ixa in blood, *The Journal of Biological Chemistry* **279**(22):22875–22882, 2004.
6. Drath R, Description of hybrid systems by modified Petri nets, in *Modelling, Analysis, and Design of Hybrid Systems*, , Springer-Verlag, pp. 15–36, 2002.
7. Greenberg C, Miraglia C, Rickels F, Shuman M, Cleavage of blood coagulation factor xiii and fibrinogen by thrombin during in vitro clotting, *The Journal of Clinical Investigation* **75**:1463–1470, 1985.
8. Halkier T, *Mechanisms in blood coagulation, Fibrinolysis, and the Complement System*, Cambridge University Press, Cambridge, England, 1991, translation from Danish into English by Paul Woolley.
9. Hiskens IA, Pai M, Trajectory sensitivity analysis of hybrid systems, *IEEE Transactions on Circuits and Systems* **47**(2):204–220, 2000.
10. Kogan A, Kardakov D, Khanin M, Analysis of the activated partial thromboplastin time test using mathematical modeling, *Thrombosis Research* **101**:299–310, 2001.
11. Leipold R, Bozarth T, Racanelli A, Dicker I, Mathematical model of serine protease inhibition in the tissue factor pathway to thrombin, *The Journal ob Biological Chemistry* **270**(43):25383–25387, 1995.
12. Lewis K, Teller D, Fry J, Lasser G, Bishop P, Crosslinking kinetics of the human transglutaminase, factor xiii[a2], acting on fibrin gels and γ-chain peptides, *Biochemistry* **36**:995–1002, 1997.
13. Luan D, Zai M, Varner JD, Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies, *PLoS Computational Biology* **3**:1347–1359, 2007.
14. Lygeros J, Sastry S, Tomlin C, The art of hybrid systems, 2001, URL `http://robotics.eecs.berkeley.edu/~sastry/ee291e/book.pdf`, textbook on hybrid systems, forthcoming.
15. Makin JG, *A Computational Model of Human Blood Clotting: Simulation, Analysis, Control, and Validation*. PhD Thesis, EECS Department, University of California, Berkeley, 2008, URL `http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-165.html`.
16. Makin JG, Narayanan S, Real-time control of human coagulation, *IET Control Theory and Applications* **6**:2630–2643, 2012.
17. Mann K, Brummel K, Butenas S, What is all that thrombin for?, *Journal of Thrombosis and Hæmostasis* **1**:1504–1514, 2003.
18. Mosesson M, Fibrinogen functions and fibrin assembly, *Fibrinolysis & Proteolysis* **14**(2):182–186, 2000.
19. Muszbek L, Yee V, Hevessy Z, Blood coagulation factor xiii: structure and function, *Thrombosis Research* **94**:271–305, 1999.
20. Panteleev M, Zarnitsina V, Ataullakhanov F, Tissue factor pathway inhibitor - a possible mechanism of action, *European Journal of Biochemistry* **269**:2016–2031, 2002.
21. Panteleev MA, Ovanesov MV, Kireev DA, Shibeko AM, Sinauridze EI, Ananyeva NM, Butylin AA, Saenko EL, Ataullakhanov FI, Spatial propagation and localization of blood coagulation are regulated by intrinsic and protein c pathways, respectively, *Biophysical Journal* **90**:1489–1500, 2006.
22. Pohl B, Beringer C, Bomhard M, Keller F, The quick machine-a mathematical model for the extrinsic activation of coagulation, *Hæmostasis* **24**:325–337, 1994.
23. Qiao Y, Xu C, Zeng Y, Xu X, Zhao H, Xu H, The kinetic model and simulation of blood coagulation–the kinetic influence of activated protein c, *Medical Engineering and Physics* **26**:341–347, 2004.
24. Røjkjær R, Schmaier A, Activation of the plasma kallikrein/kinin system on endothe-

lial cell membranes, *Immunopharmacology* **43**:109–114, 1999.
25. Shariat-Madar Z, Mahdi F, Schmaier A, Assembly and activation of the plasma kallikrein system: a new interpretation, *International Immunopharmacology* **2**:1841–1849, 2002.
26. Spronk HMH, Dielis AWJH, Panova-Noeva M, van Oerle R, Govers-Riemslag JWP, Hamulyák K, Falanga A, Cate HT, Monitoring thrombin generation: Is addition of corn trypsin inhibitor needed?, *Thrombosis and Haemostasis* **101**:1156–1162, 2009.
27. Standeven K, Ariëns R, Grant P, The molecular physiology and pathology of fibrin structure/function, *Blood Reviews* **19**:275–288, 2005.
28. van't Veer C, Kalafatis M, Bertina R, Simioni P, Mann K, Increased tissue factor-intiated prothrombin activitation as a result of the $\mathrm{Arg}^{506} \rightarrow$ Gln mutation in factor V^{LEIDEN}, *The Journal of Biological Chemistry* **272**(33):20721–20729, 1997.
29. Wagenvoord R, Hemker P, Hemker H, The limits of simulation of the clotting system, *Journal of Thrombosis and Hæmostatis* **4**:1331–1338, 2006.
30. Young G, Sø rensen B, Dargaud Y, Negrier C, Brummel-Ziedins K, Key NS, Thrombin generation and whole blood viscoelastic assays in the management of hemophilia: current state of art and future perspectives., *Blood* **121**(11):1944–50, 2013.

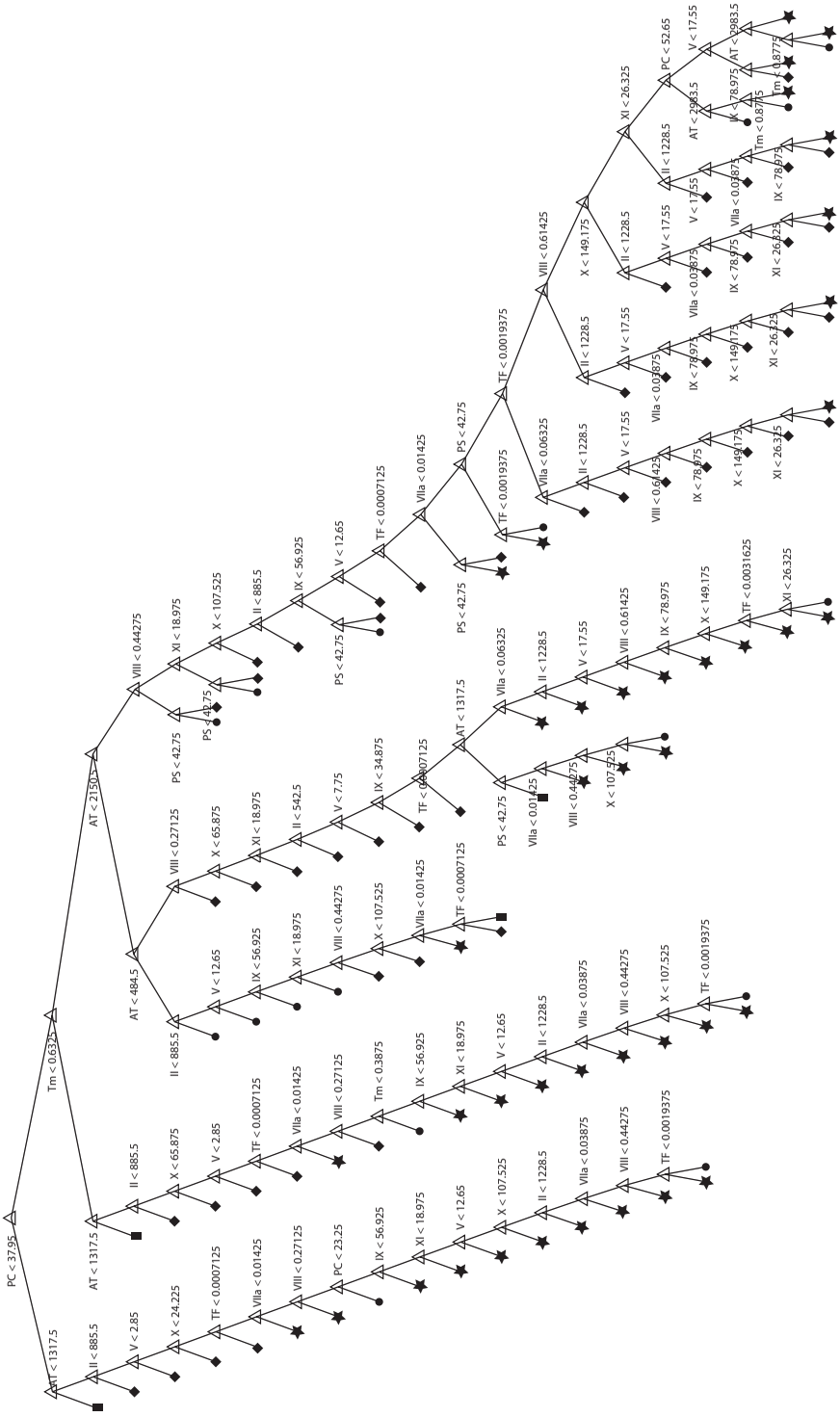


Fig. 6. The decision tree. The class labels are hypo- (stars), hyper- (squares), normal (diamonds) and extreme hyper- (circles) coagulation. Inequalities designate left branches.