

# **Dynamic Modeling of the Human Coagulation Cascade using Reduced Order Effective Kinetic Models**

Adithya Sagar and Jeffrey D. Varner\*

School of Chemical and Biomolecular Engineering  
Cornell University, Ithaca NY 14853

**Running Title:** Modeling blood coagulation

**To be submitted:** *Processes*

\*Corresponding author:

Jeffrey D. Varner,

Associate Professor, School of Chemical and Biomolecular Engineering,  
244 Olin Hall, Cornell University, Ithaca NY, 14853

Email: [jdv27@cornell.edu](mailto:jdv27@cornell.edu)

Phone: (607) 255 - 4258

Fax: (607) 255 - 9166

## **Abstract**

**Keywords:** Coagulation, Mathematical modeling, Systems Biology

## 1 Introduction

2 Developing mathematical models of biochemical networks is a significant facet of systems  
3 biology. Modeling approaches differ in their degree of detail, and the choice of approach  
4 is often determined by prior system knowledge [1]. The ability of ordinary differential  
5 equations (ODEs) to capture dynamics has made it one of the most common tools for  
6 studying signaling systems. However, ODEs come with difficult (and sometimes impossi-  
7 ble) parameter identification problems. For example, Gadkar et al., showed that even with  
8 near-perfect information and high frequency sampling, it was often impossible to identify  
9 (to within a specified uncertainty) all the parameters in typical signal transduction models  
10 [2]. This result highlights the perhaps under-appreciated role that experimental design  
11 could play in generating the best training and validation data sets for model identification  
12 [3]. ODE models also require significant mechanistic knowledge of the underlying biol-  
13 ogy. However, Bailey suggested more than a decade ago, that achieving qualitative or  
14 even quantitative understanding of biological systems should not require complete struc-  
15 tural and parametric knowledge [4]. Since Bailey's complex biology with no parameters  
16 hypothesis, Sethna and coworkers showed that model behavior and predictive ability are  
17 typically sensitive to only a few parameter combinations, a characteristic seemingly uni-  
18 versal to multi-parameter signaling models referred to as *sloppiness* [5]. Alternatively,  
19 data-driven approaches [6] or logical model formulations [7] are emerging paradigms that  
20 constrain model structure (and hence the identifiability challenge) by the availability of  
21 training and validation data. These are interesting techniques because the data naturally  
22 informs model structure, which can be expanded as more data becomes available. Thus,  
23 while there are many possible strategies to simulate biochemical networks, ODE models  
24 remain a common choice despite significant identification challenges.

25 The human coagulation system is an archetype proteolytic cascade regulated by both  
26 positive and negative feedback loops. The biology of coagulation is well studied [8–11].

Coagulation is mediated by a family proteases, called factors and a key group of blood cells, called platelets. The activation of coagulation factors and platelets, which are normally inactive in the circulation, requires a trigger event. Trigger events such as vessel injury expose procoagulant materials like collagen, fibronectin, von Willebrand factor (vWF) or tissue factor (TF). These materials drive platelet localization and activation, coagulation factor activation and ultimately clot formation [12]. Once coagulation has been activated, two converging pathways activate the master protease thrombin. The extrinsic cascade is generally believed to be the main mechanism of thrombinogenesis in the blood [8, 9, 13]. Thrombin generation consists of three phases, initiation, propagation and termination [14, 15]. Initially thrombin is produced upon cleavage of prothrombin by fluid phase activated factor X (FXa) [16]. Picomolar amounts of thrombin can then activate the cofactors factors V and VIII (fV and fVIII) and platelets, resulting in the formation of the *tenase* and *prothrombinase* complexes on the surface of activated platelets. Tenase amplifies the early coagulation signal by activating FXa, while prothrombinase, a complex of FVa-FXa, directly converts prothrombin to thrombin. There are several control points in the cascade that promote or inhibit thrombin formation. Tissue Factor Pathway Inhibitor (TFPI) downregulates FXa formation catalyzed by TF/FVIIa, while antithrombin III (ATIII) neutralizes several of the serine proteases generated during coagulation. Thrombin itself also inadvertently plays a role in its own inhibition; thrombin, through interaction with thrombomodulin and endothelial cell protein C receptor (EPCR), converts protein C to activated protein C (APC) which attenuates the coagulation response by the proteolytic cleavage of fV/FVa and fVIII/FVIIIa. Thus, termination occurs after prothrombin is consumed or activated thrombin is neutralized by inhibitors such as activated protein C (APC) or antithrombin III (ATIII).

In this study, we constructed a reduced order effective model of thrombin generation in normal and hemophilic blood. We used this model to understand how physiological

53 coagulation was altered by prothrombin levels, the activated protein C pathway and the  
54 level of ATIII. Previous models of coagulation have been formulated as systems of nonlin-  
55 ear ordinary differential equations, often using mass-action kinetics to describe the rates  
56 of biochemical conversions [REFHERE]. While this approach has proven highly effective,  
57 it does require significant experimental data for model identification and validation. For  
58 example, a previous coagulation model from our laboratory, which described 193 pro-  
59 teins and protein complexes interconnected by 301 biochemical interactions, contained  
60 467 unknown parameters (301 kinetic parameters and 166 initial conditions) [REFHERE].  
61 The key innovation of our approach is the seamless integration of simple effective rules  
62 encoding complex regulatory motifs with traditional kinetic pathway modeling. This inte-  
63 gration allows the description of complex regulatory interactions, such as time-dependent  
64 allosteric regulation of enzyme activity, in the absence of specific mechanistic informa-  
65 tion. The regulatory rules are easy to understand, easy to formulate and do not rely on  
66 overarching theoretical abstractions or restrictive assumptions. We tested our approach  
67 by modeling the time evolution of the human coagulation cascade.

## Results

**Formulation of reduced order coagulation models.** We developed a reduced order coagulation model to test our hybrid kinetic modeling approach (Fig. 1). A trigger event initiates thrombin formation (FIIa) from prothrombin (fII) through a lumped initiation step. This step loosely represents the initial fluid phase activation of thrombin by activated FXa. Once activated, thrombin catalyzes its own formation (amplification step), and inhibition via the conversion of protein C to activated protein C (APC). APC and tissue factor pathway inhibitor (TFPI) inhibit initiation, while antithrombin III (ATIII) inhibits amplification. All initiation and inhibition processes, as well as the dependence of amplification upon other coagulation factors, was approximated using our rule-based approach (Fig. 2). Individual regulatory contributions to the activity of pathway enzymes were integrated into control coefficients using an integration rule. These control coefficients then modified the rates of model processes at each time step. Hill-like transfer functions  $0 \leq f(\mathcal{Z}) \leq 1$  quantified the influence of components upon a target process. Components were either individual inhibitor or activator levels or some function of levels, e.g., the product of factor levels. In this study,  $\mathcal{Z}$  corresponded to the abundance of individual inhibitors or activators, with the exception of the dependence of amplification upon specific coagulation factors (modeled as the product of factors). When a process was potentially sensitive to multiple inputs, logical integration rules were used to select which transfer functions influenced the process at any given time. In our proof of concept model, we used a winner takes all strategy; the maximum or minimum transfer function was selected at any given time step. However, other integration rules are certainly possible. Taken together, while the reduced order coagulation model encodes significant biological complexity, it is highly compact, consisting of only five differential equations. Thus, it will serve as an excellent proof of principle example to study the reduction of a highly complex human subsystem.

**Identification of model parameters using particle swarm optimization.** A critical challenge for any dynamic model is the estimation of kinetic parameters. We estimated kinetic and control parameters simultaneously from *in vitro* coagulation data sets with and without the protein C pathway. The residual between model simulations and experimental measurements was minimized using particle swarm optimization (PSO). A population of particles ( $N = 20$ ) was initialized with randomized kinetic and control parameters and allowed to search for parameter vectors that minimized the residual. However, not all parameters were varied simultaneously. We partitioned the parameter estimation problem into two subproblems based upon the biological organization of the training data; (i) estimation of parameters associated with thrombin formation in the absence of the protein C pathway and (ii) estimation of parameters associated with the protein C pathway. Only those parameters associated with each subproblem were varied during the optimization procedure for that subproblem e.g., thrombin parameters were *not* varied during the protein C subproblem. The PSO procedure was run for 20 generations for each subproblem, where each generation was 1200 iterations. The best particle from each generation was used to generate the particle population for the next generation. We rotated the subproblems, starting with subproblem 1 in the first generation.

The reduced order coagulation model captured the role of initial prothrombin abundance, and the decay of the thrombin signal following from ATIII activity (Fig. 3). However, we systematically under-predicted the thrombin peak and the strength of ATIII inhibition in this training data set. On the other hand, with fixed thrombin parameters, we captured peak thrombin values and the decay of the thrombin signal (at least for the 150% fII case) in the presence of both ATIII and the protein C pathway (Fig. 4). Lastly, we were unable to capture global differences in initiation time *across* separate data sets with a single ensemble of model parameters. These differences likely result from normal experimental variability. For example different thrombin generation experiments within our training data

(at the same physiological factor levels) have significantly different initiation times (supplemental results). However, this also highlights a potential shortcoming of the initiation module within the model. To capture the variability in initiation time *across* training data sets, we included a constant time-delay parameter ( $T_D$ ) for each data group. The delay parameter was constant within a data set, but allowed to vary *across* training data sets. Introduction of the delay parameter allowed the model to simulate multiple training data sets using a single ensemble of model parameters. Taken together, the model identification results suggested that our kinetic-rules based approach could reproduce a panel of thrombin generation data sets conducted at physiological factor and inhibitors concentrations. However, it was unclear whether the reduced order model could predict new data, without updating the model parameters.

**Validation of the reduced order coagulation model.** We tested the predictive power of the reduced order coagulation model with validation data sets not used during model training. Two validation data sets were used, thrombin generation for various prothrombin and ATIII concentrations with the protein C pathway, and thrombin generation in normal versus hemophilic plasma in the presence of the protein C pathway. Lastly, we compared the qualitative output of the model to rFVIIa addition in the presence of hemophilia. The hemophilia case was an especially difficult test as it was taken from a different study which used a plasma-based *in vitro* assay involving platelets instead of phospholipid vesicles (PCPS). All kinetic and control parameters were fixed for the validation simulations. The only globally adjustable parameter  $T_D$ , was fixed within each validation data set but allowed to vary between data sets. The reduced order model predicted the thrombin generation profile for ratios of prothrombin and ATIII in the absence of the protein C pathway (Fig. 5). Simulations near the physiological range ( $fII, ATIII$ ) = (100%, 100%) or (125%, 75%) tracked the measured thrombin values (Fig. 5B and C). On the other hand, predictions for factors levels outside of the physiological range ( $fII, ATIII$ ) = (50%, 150%)



or (150%, 50%), while qualitatively consistent with measured thrombin values, did show significant deviation from the measurements (Fig. 5A and D). Likewise, simulations of thrombin generation in normal versus hemophilia (missing both fVIII and fIX) were consistent with measured thrombin values (Fig. 6). We modeled the dependence of thrombin amplification on factor levels using a product rule ( $\mathcal{Z} = fV \times fX \times fVIII \times fIX$ ), which was then integrated using a `min` integration rule into the control variable governing amplification. Thus, in the absence of fVIII or fIX, the amplification control variable evaluated to zero, and the only thrombin produced was from initiation (Fig. 6B). However, the decay of the thrombin signal was underpredicted in the normal case (Fig. 6A), while the activated thrombin level was overpredicted in hemophilia simulations, although thrombin generation was far less than normal (Fig. 6B). Taken together, the reduced order model performed well in the physiological range of factors, even with unmodeled components such as platelet activation in the hemophilia data set. Thus, our kinetic-rules framework predicted the output of a physiologically important cascades such as coagulation, despite significant unmodeled components.

The model ensemble predicted a direct correlation between thrombin generation and rFVIIa addition in hemophilia (Fig. 7). In the current model, we cannot distinguish between different initiation sources, e.g., TF/FVIIa and rFVIIa, as we have only a single lumped initiation source (trigger). Thus, we simulated the addition of rFVIIa in hemophilia by removing fVIII and fIX from the model, and modulating the initial level of trigger. Simulations with a baseline level of trigger were consistent with the previous hemophilia simulations, where the only thrombin being produced was from initiation (Fig. 7,  $1 \times$  trigger). As we increased the strength of the trigger event, the thrombin peak time and the maximum value of thrombin increased (Fig. 7,  $50 \times$  trigger). However, as the trigger strength increased, the thrombin generated quickly decayed. For example, for the highest trigger strength simulated ( $50 \times$  trigger), 95% of the thrombin was gone by 20 min after initiation.

We performed flux analysis to understand how the reduced order coagulation model balanced initiation, amplification and inhibition of thrombin formation for normal coagulation and hemophilia. Analysis of the reaction flux through the reduced order network for thrombin generation in normal, hemophilia and rFVIIa-treated hemophilia identified three distinct operational modes (Fig. 8). The reduced order network includes four lumped reactions, initiation, amplification, thrombin-induced APC generation and total thrombin inhibition (including both APC and ATIII action). Directly after the addition of a trigger (e.g., TF/FVIIa or rFVIIa), the lumped initiation flux was the largest for all three cases. However, within a few minutes enough thrombin was generated by the initiation mechanism to induce the amplification stage. During amplification, thrombin catalyzes its own formation and inhibition by generating activated protein C (APC), a potent inhibitor of the coagulation cascade. For normal coagulation, amplification and thrombin inhibition are the dominant reactions by 6 min after initiation (Fig. 8, left). After 10 min, the dominant reaction has shifted to thrombin inhibition (both ATIII and APC action). In the current proof-of-principle model, APC inhibits upstream of amplification thus APC activity will slow the rate of thrombin formation directly from the initiation trigger. On the other hand, ATIII inhibits thrombin directly, as well as upstream coagulation factors. In hemophilia (missing both fVIII and fIX), the amplification reaction does not occur and the only thrombin produced is from initiation (Fig. 8, center). Initiation is quickly inhibited by APC, and the thrombin level stabilizes and eventually decays because of ATIII activity. Lastly, when 50×rFVIIa is used to induce thrombin formation in hemophilia (absence of fVIII/fIX), initiation mechanisms dominate for up to 6 min following initiation (Fig. 8, right). However, similar to hemophilia alone, without amplification the thrombin signal is quickly extinguished by the combined action of ATIII and APC generated by thrombin.

**Sensitivity analysis of the reduced order coagulation model.** We conducted a global sensitivity analysis to estimate which parameters controlled the performance of the reduced order model. We calculated the sensitivity of the time to maximum thrombin (peak time) and the thrombin exposure (area under the thrombin curve) for different levels of prothrombin, and protein C (Fig. 9). Globally, 41% of the parameters shifted in importance between the (fII,PC) = (50%, 0%) and (150%,100%) cases for the peak thrombin time (Fig. 9A). The majority of these shifts were involved increased prothrombin and the protein C pathway, while only 5% were directly associated with increased prothrombin alone. The rate constant for thrombin amplification was the most important parameter controlling the peak thrombin time. While this parameter was differentially important for different prothrombin levels, and in the presence or absence of the activated protein C pathway, it's sensitivity was consistently above all other parameters in the model. The saturation constant governing thrombin amplification was the second most important parameter, followed by the initiation control gain parameter. Other important parameters influencing the thrombin peak time included the control gain for activated protein C formation, as well as the rate constant controlling ATIII inhibition of thrombin activity. On the other hand, only 27% of the model parameters were differentially sensitive between the (%fII,%PC) (50%, 0%) and (150%,100%) cases for the thrombin exposure (Fig. 9B). Of these parameters, all of the shifts were associated the interplay between thrombin formation and the protein C pathway. The rate constant controlling ATIII inhibition was the most important parameter controlling the thrombin exposure. While this parameter was less important in the presence of protein C for 150% prothrombin levels, it was significantly above all other parameters. Similar to the peak time, for 150% prothrombin, the control gain for activated protein C formation was differentially important along with the rate constant controlling amplification. However, the influence of the amplification parameter was much less in the thrombin exposure case versus peak time.

## Discussion

Controlling hemorrhage has been a driving force to understand the molecular basis of blood coagulation. Replacement therapy improves bleeding times for most hemophilia patients. However, repeated fVIII/fIX administration can lead to neutralizing inhibitors. Alternatively, rFVIIa has been used as a universal hemostatic agent to initiate clotting with and without fVIII or fIX inhibitors [17]. Despite its wide adoption, the mechanism and the utility of rFVIIa in trauma-associated hemorrhage remains controversial [? ]. In this study, we formulated a model of the human coagulation cascade and studied thrombin formation in normal and rFVIIa treated and untreated hemophilic plasma. The model described 193 proteins or protein complexes connected by 301 interactions. The coagulation network architecture was based on literature and a previous mathematical model from our laboratory [? ]. Simulations of thrombin dynamics were independently validated using blood drawn from patients with coronary artery disease (reported here) and previous TF/FVIIa and FVIIa *in-vitro* studies [10]. While the human coagulation cascade is perhaps an ideal model system to develop network analysis tools, not all the model parameters were identifiable given the current training data. Instead, an *ensemble* of models (N = 437) was estimated using nine cell-based coagulation training sets [? ]. We used this family of consistent coagulation models in all simulations instead of a single best-fit but uncertain model. Additionally, using sensitivity analysis on the ensemble of models, we identified structurally sensitive components of the coagulation architecture as a function of condition. We also demonstrated that the ensemble robustly constrained model predictions of independent validation sets, despite having many poorly constrained parameters. Thus, the model ensemble displayed *sloppy* behavior similar to that observed by Sethna and coworkers for other signal transduction networks [? ]. By studying a family of models, we perhaps partially addressed the uncertainty stemming from the many poorly characterized model parameters. However, many other factors could influence our results. For

example, missing structural interactions or biophysical factors such as blood flow could play a large role. Thus, while the results presented here may be a valuable first step, more studies are required.

Malfunctions in coagulation can have serious or potentially even fatal consequences. For example, aggressive clotting is involved with Coronary Artery Diseases (CADs), which collectively account for 38% of all deaths in North America [18]. Coagulation management during surgery can also be challenging, particularly because of the increasing clinical use of antithrombotic drugs [19]. Insufficient coagulation due to genetic disorders such as hemophilia can also result in recurrent bleeding. The coagulation factors VIII (fVIII) and IX (fIX) are deficient in Hemophilia A and B, respectively [20–22]. People with mild hemophilia have 5-40% of the normal clotting factor levels while severe hemophiliacs have <1% [22]. Hemophilia can be controlled with regular infusions of the deficient clotting factors. However, clotting factor replacement sometimes leads to the formation of *in-vivo* fVIII and fIX inhibitors [23]. Activated Prothrombin Complex Concentrates (aPCCs) from pooled plasma have been used to treat hemophilia. However, aPCCs can have safety issues [24–26]. Alternatively, recombinant factor VIIa (rFVIIa) has been used to treat bleeding disorders [17, 27] including hemophilia with and without factor VIII/IX inhibitors [28–31]. However, rFVIIa requires frequent administration (every 2-3 hr) because of its short half-life in the circulation. Many questions also remain about its mechanism [16, 32–35] and effective dose range [23]. Thus, despite its wide adoption, the utility of rFVIIa in trauma-associated hemorrhage remains controversial [? ].

While the performance of the proof of principle reduced order model was impressive, there are several questions that should be explored further. [FINISH]

## Materials and Methods

**Formulation and solution of the model equations.** We used ordinary differential equations (ODEs) to model the time evolution of proteins ( $x_i$ ) in our reduced order coagulation model:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \quad i = 1, 2, \dots, \mathcal{M} \quad (1)$$

where  $\mathcal{R}$  denotes the number of reactions,  $\mathcal{M}$  denotes the number of protein species in the model. The quantity  $r_j(\mathbf{x}, \epsilon, \mathbf{k})$  denotes the rate of reaction  $j$ . Typically, reaction  $j$  is a non-linear function of biochemical species abundance, as well as unknown kinetic parameters  $\mathbf{k}$  ( $\mathcal{K} \times 1$ ). The quantity  $\sigma_{ij}$  denotes the stoichiometric coefficient for species  $i$  in reaction  $j$ . If  $\sigma_{ij} > 0$ , metabolite  $i$  is produced by reaction  $j$ . Conversely, if  $\sigma_{ij} < 0$ , metabolite  $i$  is consumed by reaction  $j$ , while  $\sigma_{ij} = 0$  indicates metabolite  $i$  is not connected with reaction  $j$ . Lastly,  $\lambda_i$  denotes the scaled enzyme degradation constant. The system material balances were subject to the initial conditions  $\mathbf{x}(t_o) = \mathbf{x}_o$ .

Each reaction rate was written as the product of two terms, a kinetic term ( $\bar{r}_j$ ) and a regulatory term ( $v_j$ ):

$$r_j(\mathbf{x}, \epsilon, \mathbf{k}) = \bar{r}_j v_j \quad (2)$$

We used multiple saturation kinetics to model the reaction term  $\bar{r}_j$ :

$$\bar{r}_j = k_j^{max} \epsilon_i \left( \prod_{s \in m_j^-} \frac{x_s}{K_{js} + x_s} \right) \quad (3)$$

where  $k_j^{max}$  denotes the maximum rate for reaction  $j$ ,  $\epsilon_i$  denotes the scaled enzyme activity which catalyzes reaction  $j$ , and  $K_{js}$  denotes the saturation constant for species  $s$  in reaction  $j$ . The product in Eqn. (3) was carried out over the set of *reactants* for reaction  $j$  (denoted as  $m_j^-$ ).

289 The control term  $v_j$  depended upon the combination of factors which influenced the  
 290 activity of enzyme  $i$ . For each enzyme, we used a rule-based approach to select from  
 291 competing control factors (Fig. 2). If an enzyme was activated by  $m$  metabolites, we  
 292 modeled this activation as:

$$v_j = \max(f_{1j}(\mathcal{Z}), \dots, f_{mj}(\mathcal{Z})) \quad (4)$$

293 where  $0 \leq f_{ij}(\mathcal{Z}) \leq 1$  was a regulatory transfer function that calculated the influence of  
 294 metabolite  $i$  on the activity of enzyme  $j$ . Conversely, if enzyme activity was inhibited by a  
 295  $m$  metabolites, we modeling this inhibition as:

$$v_j = 1 - \max(f_{1j}(\mathcal{Z}), \dots, f_{mj}(\mathcal{Z})) \quad (5)$$

296 Lastly, if an enzyme had both  $m$  activating and  $n$  inhibitory factors, we modeled the regu-  
 297 latory term as:

$$v_j = \min(u_j, d_j) \quad (6)$$

298 where:

$$u_j = \max_{j^+}(f_{1j}(\mathcal{Z}), \dots, f_{mj}(\mathcal{Z})) \quad (7)$$

$$d_j = 1 - \max_{j^-}(f_{1j}(\mathcal{Z}), \dots, f_{nj}(\mathcal{Z})) \quad (8)$$

299 The quantities  $j^+$  and  $j^-$  denoted the sets of activating and inhibitory factors for enzyme  $j$ .  
 300 If a process has no modifying factors, we set  $v_j = 1$ . There are many possible functional  
 301 forms for  $0 \leq f_{ij}(\mathcal{Z}) \leq 1$ . However, in this study, each individual transfer function took the  
 302 form:

$$f_i(\mathbf{x}) = \frac{\kappa_{ij}^\eta \mathcal{Z}_j^\eta}{1 + \kappa_{ij}^\eta \mathcal{Z}_j^\eta} \quad (9)$$

where  $\mathcal{Z}_j$  denotes the abundance of the  $j$  factor (e.g., metabolite abundance), and  $\kappa_{ij}$  and  $\eta$  are control parameters. The  $\kappa_{ij}$  parameter was species gain parameter, while  $\eta$  was a cooperativity parameter (similar to a Hill coefficient). Applying the general framework to the reduced coagulation network resulted in five ordinary differential equations:

$$\frac{dx_1}{dt} = -(r_{init}v_{init} + r_{amp}v_{amp}) \quad (10)$$

$$\frac{dx_2}{dt} = r_{amp}v_{amp} + r_{init}v_{init} - (r_{inh,ATIII}v_{inh,ATIII} + r_{inh,APC}v_{inh,APC}) \quad (11)$$

$$\frac{dx_3}{dt} = -r_{apc}v_{apc} \quad (12)$$

$$\frac{dx_4}{dt} = r_{apc}v_{apc} \quad (13)$$

$$\frac{dx_5}{dt} = -r_{inh,ATIII}v_{inh,ATIII} \quad (14)$$

where  $\mathbf{x} = (fII, FIIa, PC, APC, ATIII)^T$ . The model equations were encoded using the Python programming language and solved using the ODEINT routine of the SciPy module [36].

**Estimation of model parameters from experimental data.** Model parameters were estimated by minimizing the difference between simulations and experimental thrombin measurements (squared residual):

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left( \frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2 \quad (15)$$

where  $\hat{x}_j(\tau)$  denotes the measured value of species  $j$  at time  $\tau$ ,  $x_j(\tau, \mathbf{k})$  denotes the simulated value for species  $j$  at time  $\tau$ , and  $\omega_j(\tau)$  denotes the experimental measurement variance for species  $j$  at time  $\tau$ . The outer summation is respect to time, while the inner summation is with respect to state. We minimized the model residual using Particle swarm optimization (PSO) [37]. PSO uses a *swarming* metaheuristic to explore



parameter spaces. A strength of PSO is its ability to find the global minimum, even in the presence of potentially many local minima, by communicating the local error landscape experienced by each particle collectively to the swarm. Thus, PSO acts both as a local and a global search algorithm. For each iteration, particles in the swarm compute their local error by evaluating the model equations using their specific parameter vector realization. From each of these local points, a globally best error is identified. Both the local and global error are then used to update the parameter estimates of each particle using the rules:

$$\Delta_i = \theta_1 \Delta_i + \theta_2 \mathbf{r}_1 (\mathcal{L}_i - \mathbf{k}_i) + \theta_3 \mathbf{r}_2 (\mathcal{G} - \mathbf{k}_i) \quad (16)$$

$$\mathbf{k}_i = \mathbf{k}_i + \Delta_i \quad (17)$$

where  $(\theta_1, \theta_2, \theta_3)$  are adjustable parameters,  $\mathcal{L}_i$  denotes local best solution found by particle  $i$ , and  $\mathcal{G}$  denotes the best solution found over the entire population of particles. The quantities  $r_1$  and  $r_2$  denote uniform random vectors with the same dimension as the number of unknown model parameters ( $\mathcal{K} \times 1$ ). In this study, we used  $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05564, 0.02886)$ . The quality of parameter estimates was measured using two criteria, goodness of fit (model residual) and angle between the estimated parameter vector  $\mathbf{k}_j$  and the best parameter set  $\mathbf{k}^*$ :

$$\alpha_j = \cos^{-1} \left( \frac{\mathbf{k}_j \cdot \mathbf{k}^*}{\|\mathbf{k}_j\| \|\mathbf{k}^*\|} \right) \quad (18)$$

If the candidate parameter set  $\mathbf{k}_j$  were perfect, the residual between the model and synthetic data and the angle between  $\mathbf{k}_j$  and the true parameter set  $\mathbf{k}^*$  would be equal to zero. The particle swarm optimization routine was implemented in the Python programming language. All plots were made using the Matplotlib module of Python [38].

## 337 **Acknowledgements**

338 This study was supported by an award from the Army Research Office (ARO #59155-LS).

## References

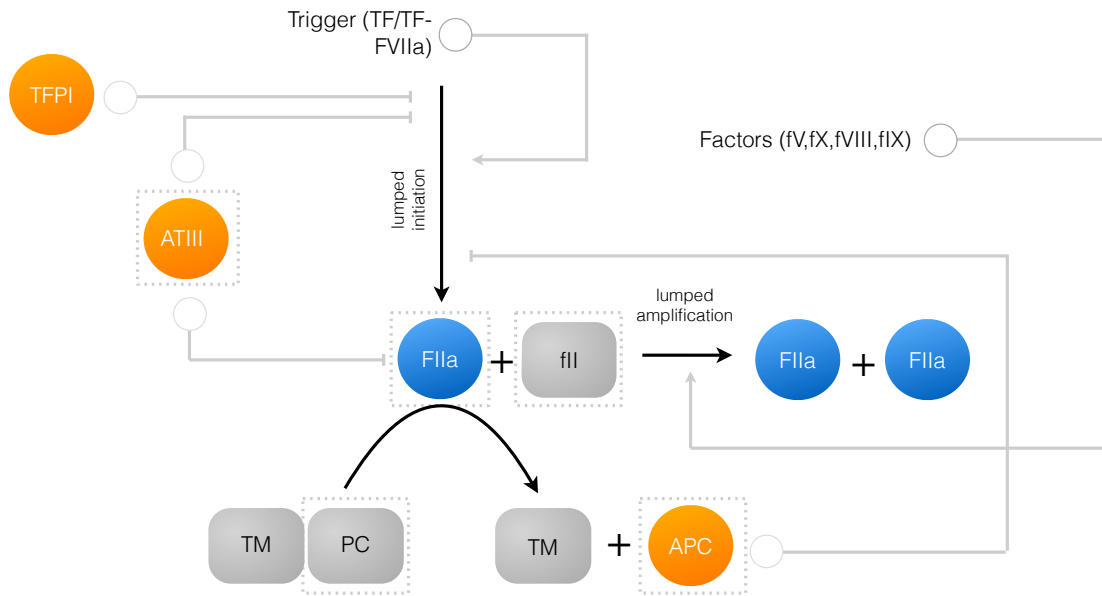
1. Kholodenko B, Yaffe MB, Kolch W (2012) Computational approaches for analyzing information flow in biological networks. *Sci Signal* 5: re1.
2. Gadkar KG, Varner J, Doyle FJ (2005) Model identification of signal transduction networks from data using a state regulator problem. *Syst Biol (Stevenage)* 2: 17–30.
3. Apgar JF, Witmer DK, White FM, Tidor B (2010) Sloppy models, parameter uncertainty, and the role of experimental design. *Mol Biosyst* 6: 1890-900.
4. Bailey JE (2001) Complex biology with no parameters. *Nat Biotechnol* 19: 503-4.
5. Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compression underlies emergent theories and predictive models. *Science* 342: 604-7.
6. Cirit M, Haugh JM (2012) Data-driven modelling of receptor tyrosine kinase signalling networks quantifies receptor-specific potencies of pi3k- and ras-dependent erk activation. *Biochem J* 441: 77-85.
7. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* 7: e1001099.
8. Roberts H, Monroe D, Oliver J, Chang J, Hoffman M (1998) Newer concepts of blood coagulation. *Haemophilia* 4: 331-334.
9. Mann K (1999) Biochemistry and physiology of blood coagulation. *Thromb Haemost* 82: 165-174.
10. Butenas S, Mann K (2002) Blood coagulation. *Biochemistry (Moscow)* 67: 3-12.
11. Schenone M, Furie BC, Furie B (2004) The blood coagulation cascade. *Curr Opin Hematol* 11: 272 - 277.
12. Giesen P, Rauch U, Bohrmann B, Kling D, Roque M, et al. (1999) Blood-borne tissue factor: another view of thrombosis. *Proc Natl Acad Sci USA* 96: 2311 - 2315.

13. Mann K, Nesheim M, Church W, Haley P, Krishnaswamy S (1990) Surface-dependent reactions of vitamin k-dependent enzyme complexes. *Blood* 76: 1-16.
14. Goldhaber SZ, Colman RW, Clowes AW, editors (2006) Hemostasis and Thrombosis: Basic Principles and Clinical Practice. Lippincott Williams and Wilkins.
15. Brummel KE, Paradis SG, Butenas S, Mann KG (2002) Thrombin functions during tissue factor-induced blood coagulation. *Blood* 100: 148-152.
16. Butenas S, Brummel KE, Branda RF, Paradis SG, Mann KG (2002) Mechanism of factor VIIa-dependent coagulation in hemophilia blood. *Blood* 99: 923-930.
17. Hedner U (2008) Factor viia and its potential therapeutic use in bleeding-associated pathologies. *Thromb Haemost* 100: 557–562.
18. GKHansson (2005) Inflammation, Atherosclerosis and Coronary Artery Disease. *N Engl J Med* 352: 1685 - 1695.
19. Tanaka KA, Key NS, Levy JH (2009) Blood coagulation: hemostasis and thrombin regulation. *Anesth Analg* 108: 1433-46.
20. Tuddenham E, Cooper D (1994) The molecular genetics of haemostasis and its inherited disorders., volume 25 of *Oxford monographs in medical genetics*. Oxford University Press.
21. Mannucci MP, Tuddenham EGD (2001) The hemophilias - from royal genes to gene therapy. *N Engl J Med* 344: 1773 - 1780.
22. Mitchell J, Phillott A (2008) Haemophilia and inhibitors 1: diagnosis and treatment. *Nursing Times* 104: 26–27.
23. Tomokiyo K, Nakatomi Y, Araki T, Teshima K, Nakano H, et al. (2003) A novel therapeutic approach combining human plasma-derived factors viia and x for haemophiliacs with inhibitors: evidence of a higher thrombin generation rate in vitro and more sustained haemostatic activity in vivo than obtained with factor viia alone. *Vox Sanguinis* 85: 290-299.

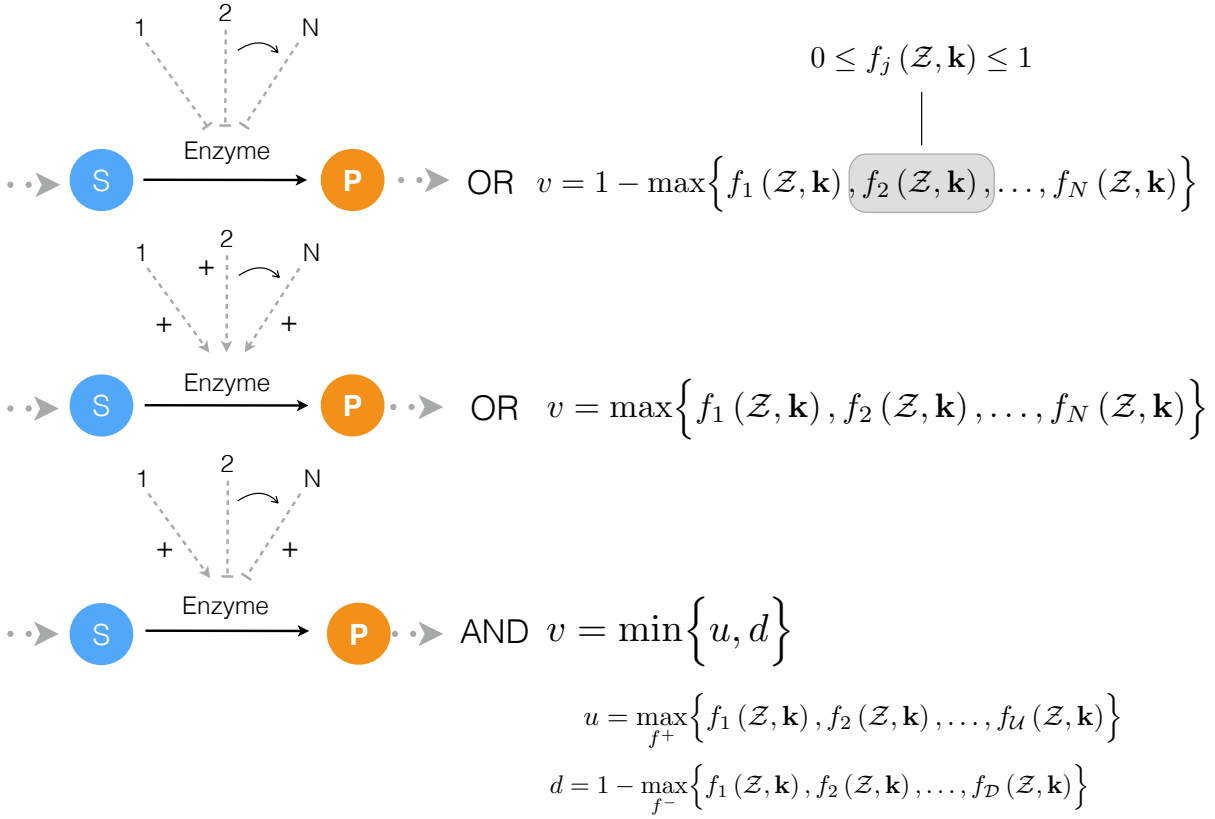
24. Chavin SI, Siegel DM, Rocco TA, Olson JP (1988) Acute myocardial infarction during treatment with an activated prothrombin complex concentrate in a patient with factor viii deficiency and a factor viii inhibitor. *Am J Med* 85: 245-249.
25. Allen GA, Hoffman M, Roberts HR, Monroe DM (2006) Manipulation of prothrombin concentration improves response to high-dose factor VIIa in a cell-based model of haemophilia. *Br J Haematology* 134: 314 - 319.
26. Schneiderman J, Rubin E, Nugent DJ, Young G (2007) Sequential therapy with activated prothrombin complex concentrates and recombinant FVIIa in patients with severe haemophilia and inhibitors: update of our previous experience. *Haemophilia* 13: 244-248.
27. Talbot M, Tien HC (2009) The use of recombinant factor viia in trauma patients. *J Am Acad Orthop Surg* 17: 477-81.
28. Lusher JM, Roberts HR, Davignon G, Joist JH, Smith H, et al. (1998) A randomized, double-blind comparison of two dosage levels of recombinant factor viia in the treatment of joint, muscle and mucocutaneous haemorrhages in persons with haemophilia a and b, with and without inhibitors. rfvia study group. *Haemophilia* 4: 790–798.
29. Shapiro AD, Gilchrist GS, Hoots WK, Cooper HA, Gastineau DA (1998) Prospective, randomised trial of two doses of rfvia (novoseven) in haemophilia patients with inhibitors undergoing surgery. *Thromb Haemost* 80: 773–778.
30. Shapiro AD (2000) Recombinant factor viia in the treatment of bleeding in hemophilic children with inhibitors. *Semin Thromb Hemost* 26: 413–419.
31. Shapiro AD (2008) Single-dose recombinant activated factor vii for the treatment of joint bleeds in hemophilia patients with inhibitors. *Clin Adv Hematol Oncol* 6: 579–586.
32. Komiyama Y, Pedersen AH, Kisiel W (1990) Proteolytic activation of human factors ix and x by recombinant human factor viia: Effects of calcium, phospholipids and tissue

factor. Biochemistry 29: 9418-9425.

33. Payne MA, Neuenschwander PF, Johnson AE, Morrissey JH (1996) Effect of soluble tissue factor on the kinetic mechanism of factor viia: Enhancement of p-guanidinobenzoate substrate hydrolysis. Biochemistry 35: 7100-7106.
34. Monroe DM, Hoffman M, Oliver JA, Roberts HR (1997) Platelet activity of high-dose factor viia is independent of tissue factor. Br J Haematol 99: 542-547.
35. Hoffman M, Monroe III DM, Roberts HR (1998) Activated factor VII activates factors IX and X on the surface of activated platelets: thoughts on the mechanism of action of high-dose activated factor VII. Blood Coag Fibrin 9: S61-S65.
36. Jones E, Oliphant T, Peterson P (2001–). SciPy: Open source scientific tools for Python. <http://www.scipy.org/>.
37. Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the International Conference on Neural Networks. pp. 1942 - 1948.
38. Hunter JD (2007) Matplotlib: A 2d graphics environment. Computing in Science and Engineering 9: 90 - 95.
39. Butenas S, van't Veer C, Mann KG (1999) "normal" thrombin generation. Blood 94: 2169-78.

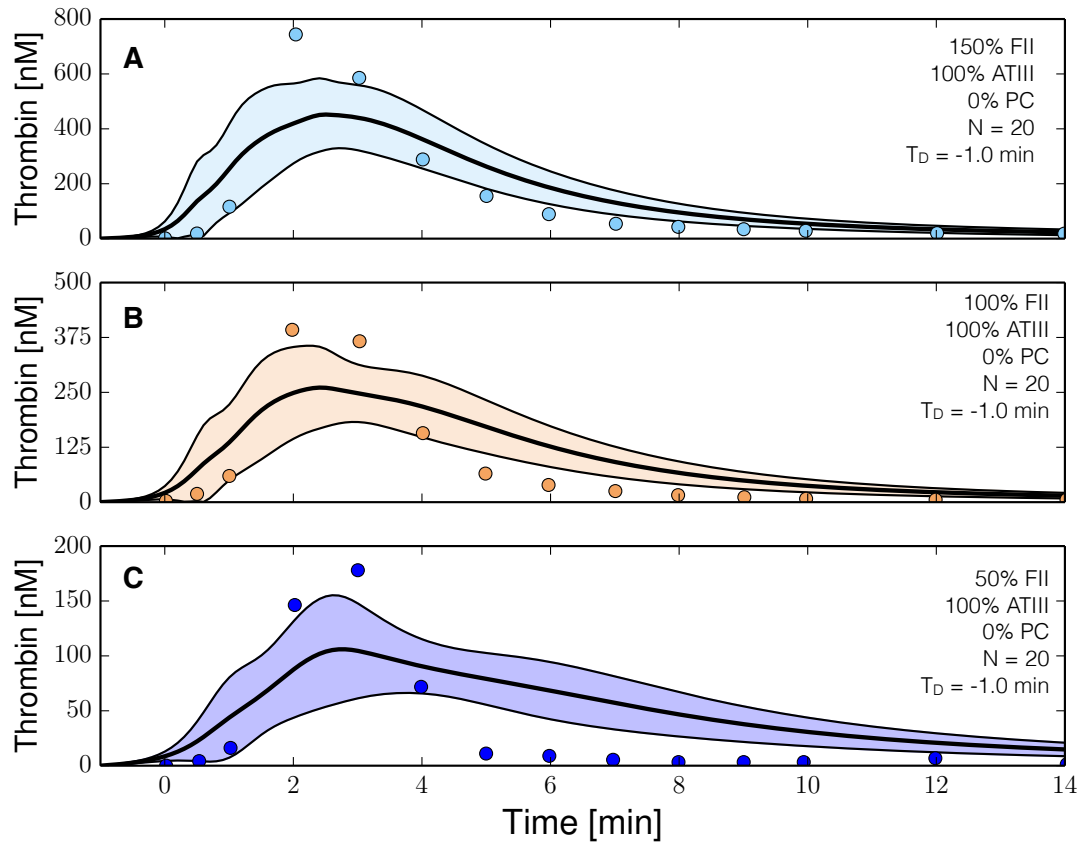


**Fig. 1:** Schematic of the connectivity of the reduced order coagulation model. A trigger compound e.g., TF/FVIIa initiates thrombin production (FIIa) from prothrombin (fII). Once activated, thrombin catalyzes its own activation (amplification step), as well as its own inhibition via the conversion of protein C to activated protein C (APC). APC and tissue factor pathway inhibitor (TFPI) inhibit initiation, while antithrombin III (ATIII) directly inhibits thrombin. All inhibition steps, and trigger-induced initiation were modeled using a rule-based approach. Likewise, the dependence of amplification on other coagulation factors was also modeled using a rule-based approach.

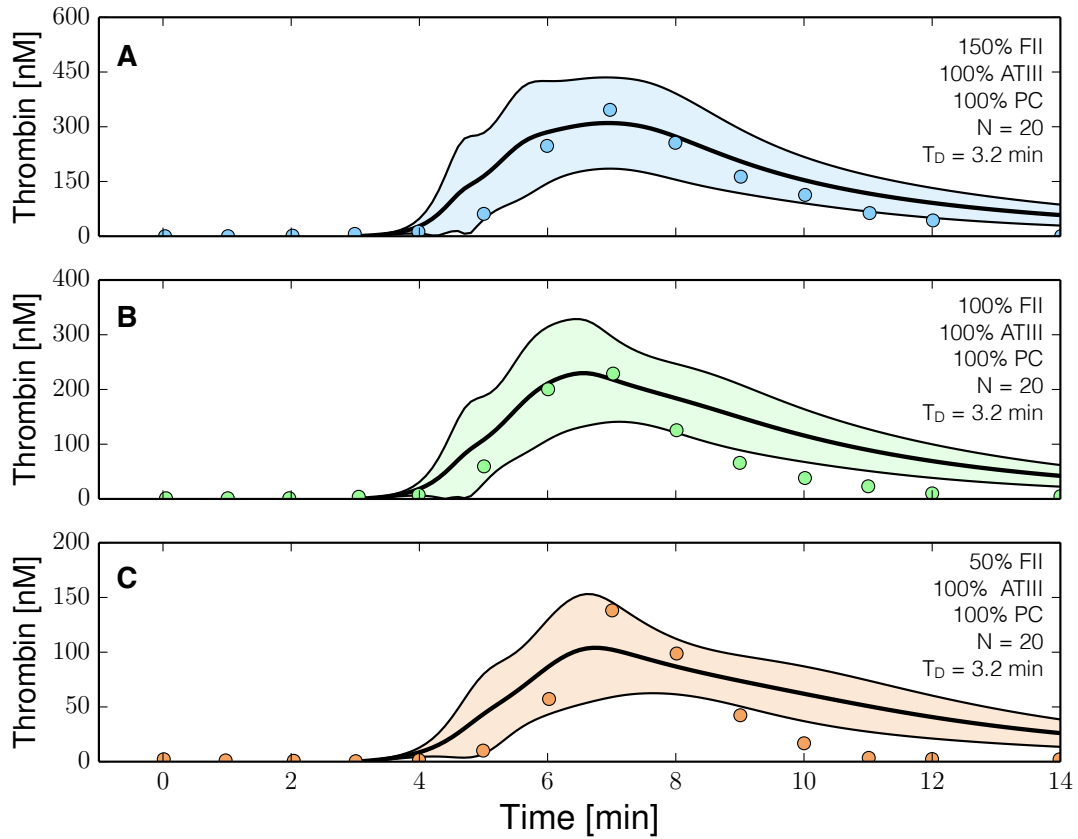


**Fig. 2:** Schematic of rule based effective control laws. Traditional enzyme kinetic expressions e.g., Michaelis–Menten or multiple saturation kinetics are multiplied by an enzyme activity control variable  $0 \leq v_j \leq 1$ . Control variables are functions of many possible regulatory factors encoded by arbitrary functions of the form  $0 \leq f_j(\mathcal{Z}) \leq 1$ . At each simulation time step, the  $v_j$  variables are calculated by evaluating integration rules such as the max or min of the set of factors  $f_1, \dots, f_n$  influencing the activity of enzyme  $E_j$ .

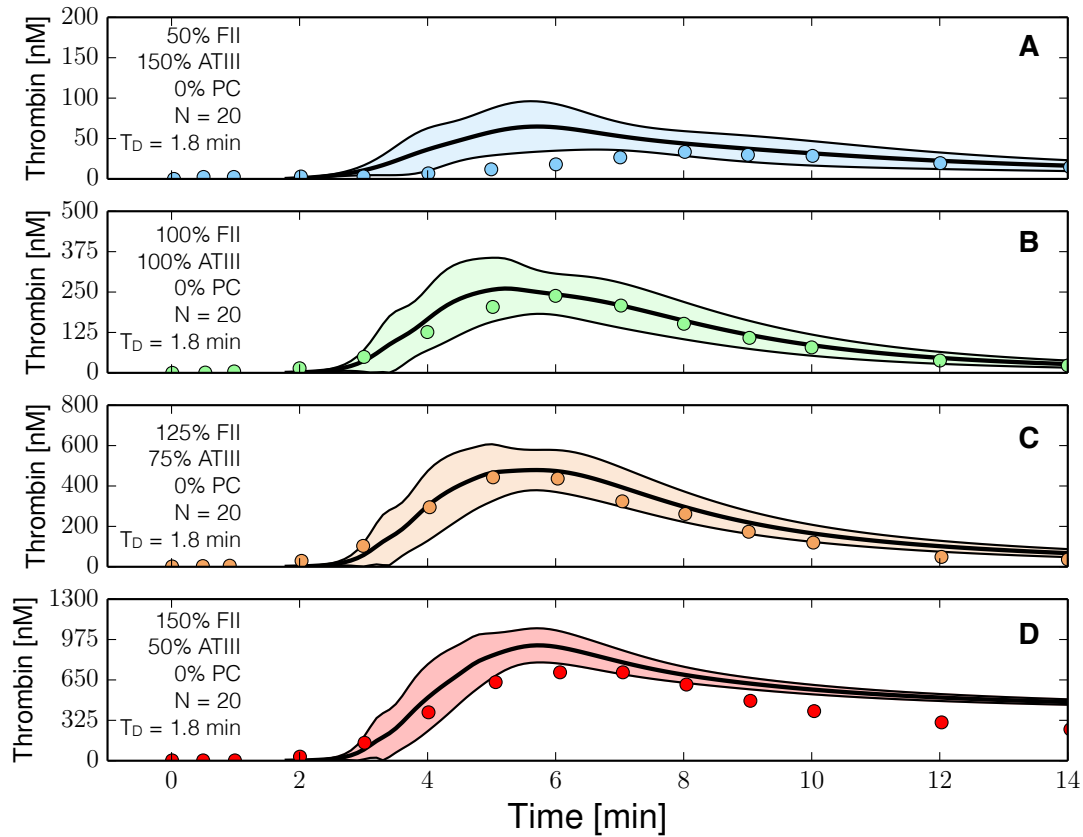




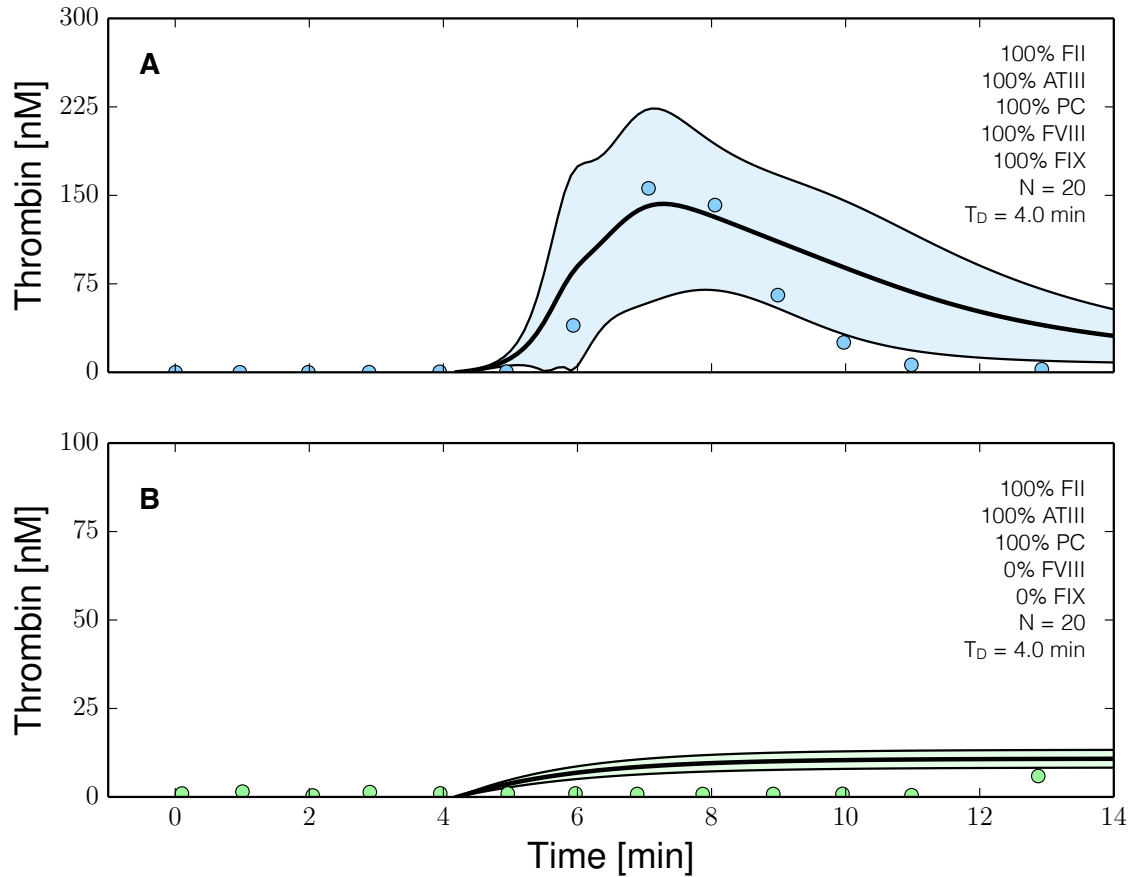
**Fig. 3:** Reduced order coagulation model training simulations. Reduced order coagulation model parameters were estimated using particle swarm optimization (PSO) with and without the protein C pathway as a function of prothrombin. Solid lines denote the simulated mean value of the thrombin profile for  $N = 20$  independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B,C) training results for 150%, 100% and 50% of physiological prothrombin levels in the absence of the protein C pathway. Thrombin generation was initiated using 5 pmol/L FVIIa-TF in the presence of 200  $\mu$ mol/L of phospholipid vesicles (PCPS). All factors and control proteins were at their physiological concentration unless others denoted. The experimental training data was reproduced from the study of Butenas et al. [39].



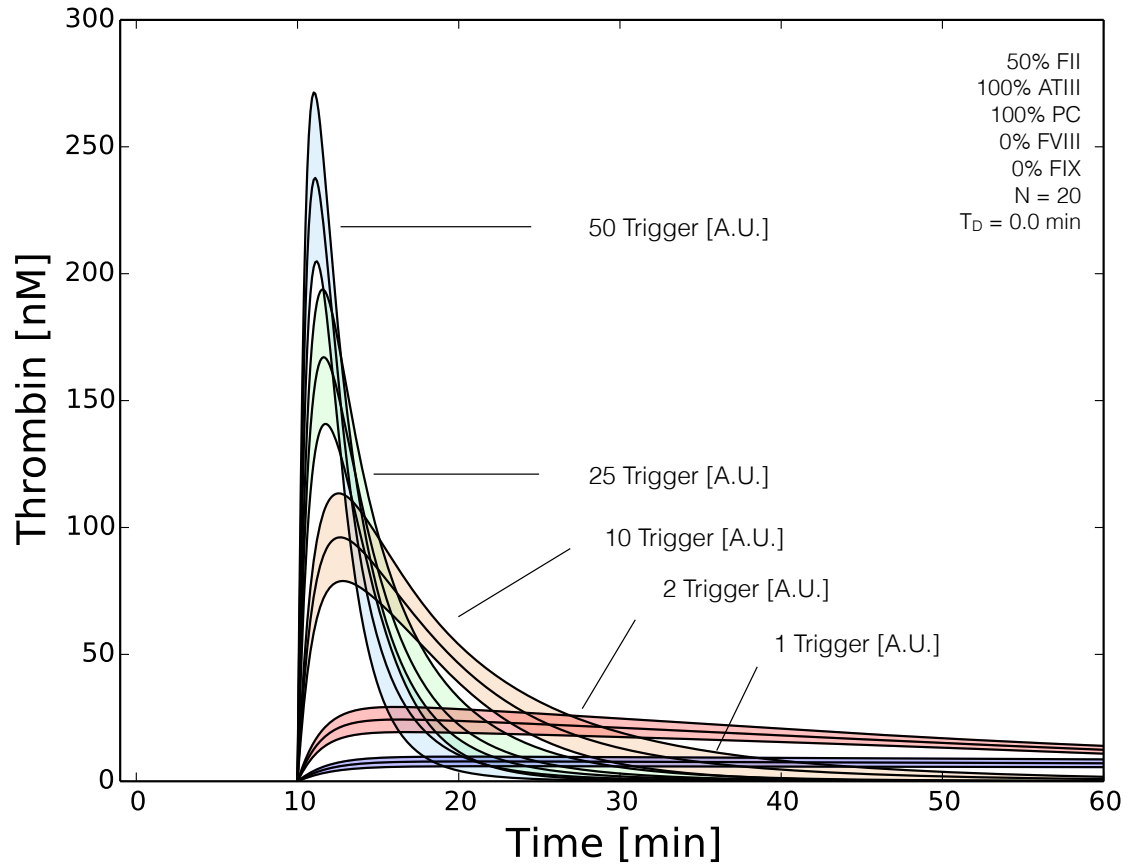
**Fig. 4:** Reduced order coagulation model training simulations. Reduced order coagulation model parameters were estimated using particle swarm optimization (PSO) with and without the protein C pathway as a function of prothrombin. Solid lines denote the simulated mean value of the thrombin profile for  $N = 20$  independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B,C) training results for 150%, 100% and 50% of physiological prothrombin levels in the presence of the protein C pathway. Only APC pathway parameters were allowed to vary in the simulations on the right. Thrombin generation was initiated using 5 pmol/L FVIIa-TF in the presence of 200  $\mu\text{mol/L}$  of phospholipid vesicles (PCPS). All factors and control proteins were at their physiological concentration unless others denoted. The experimental training data was reproduced from the study of Butenas et al. [39].



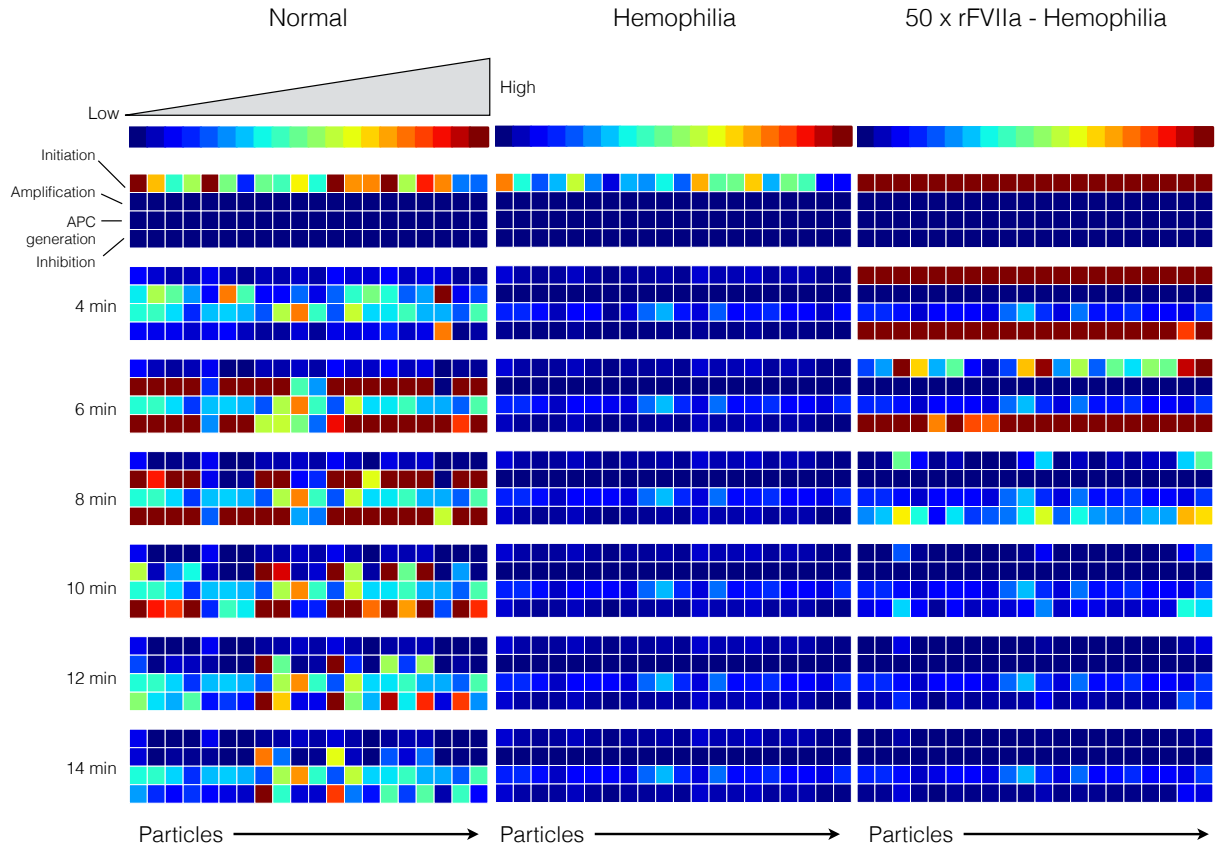
**Fig. 5:** Reduced order coagulation model predictions versus experimental data for normal coagulation. The reduced order coagulation model parameter estimates were tested against data not used during model training. Simulations of different levels of prothrombin and ATIII were compared with experimental data in the absence of the protein C pathway. Solid lines denote the simulated mean value of the thrombin profile for  $N = 20$  independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B,C,D) prediction results for (FII,ATIII): (50%,150%), (100%, 100%), (125%, 75%) and (150%, 50%) of physiological prothrombin and ATIII levels in the absence of the protein C pathway. Thrombin generation was initiated using 5 pmol/L FVIIa-TF in the presence of 200  $\mu$ mol/L of phospholipid vesicles (PCPS). All factors and control proteins were at their physiological concentration unless others denoted. The experimental validation data was reproduced from the study of Butenas et al. [39]



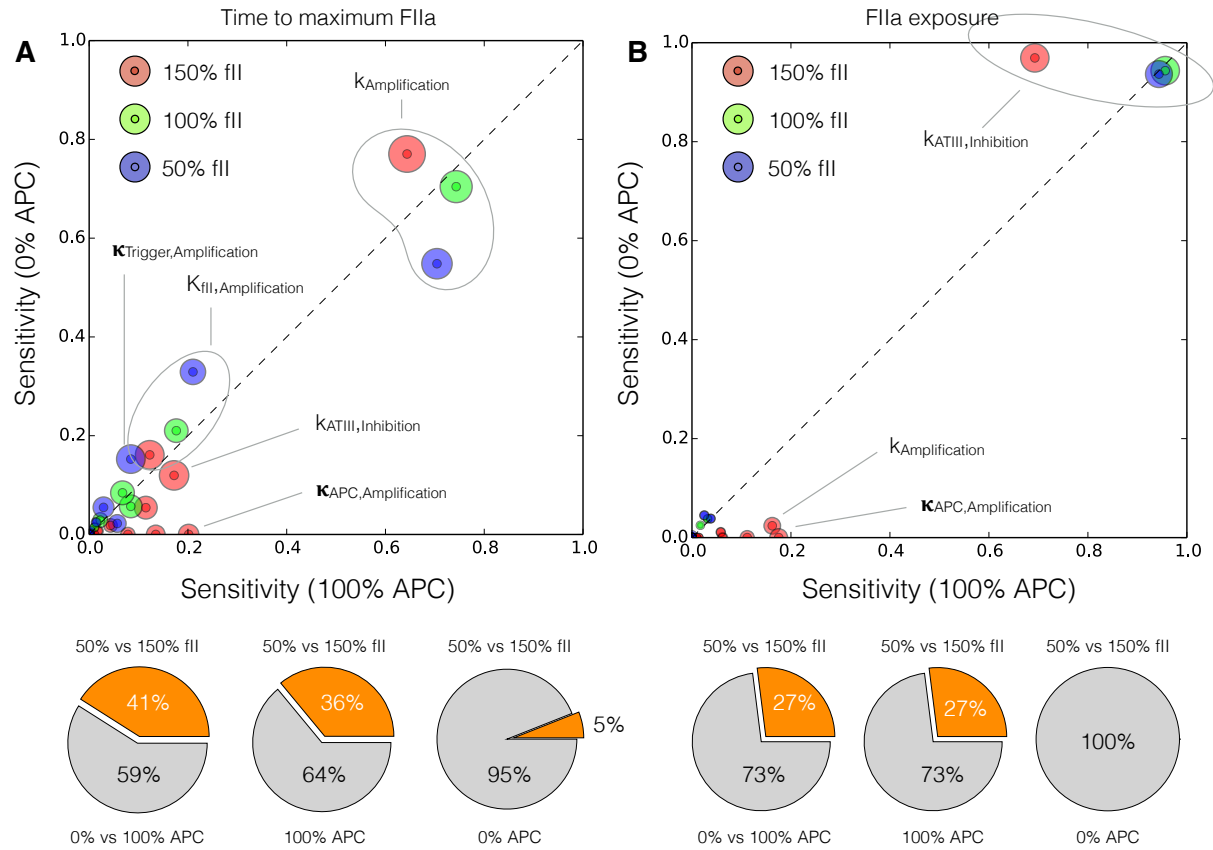
**Fig. 6:** Reduced order coagulation model predictions versus experimental data with and without FVIII and FIX. The reduced order coagulation model parameter estimates were tested against data not used during model training. Simulations of normal thrombin formation with ATIII and the protein C pathway were compared with thrombin formation in the absence of fVIII and fIX. Solid lines denote the simulated mean value of the thrombin profile for  $N = 20$  independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B) prediction results for normal thrombin generation and thrombin generation in hemophilia. All factors and control proteins were at their physiological concentration unless others noted. The experimental validation data was reproduced from the study of Allen et al. [25].



**Fig. 7:** Reduced order coagulation model predictions of rFVIIa administration. Simulations of thrombin formation in the presence of ATIII and the protein C pathway were conducted for a range of trigger values (1x - 50x) in the absence of fVIII and fIX. Solid lines denote the simulated mean value of the thrombin profile for  $N = 20$  independent particles. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). All factors and control proteins were at their physiological concentration unless others noted.



**Fig. 8:** Reaction flux distribution as a function of time for thrombin generation under normal (left), hemophilia (center) and rFVIIa treated hemophilia (right). Reaction flux was calculated for each particle at  $T = 0, 4, 6, 8, 10, 12, 14$  min after the initiation of coagulation. Reaction fluxes were calculated for each particle in the parameter ensemble ( $N = 20$ ). Blue colors denote low flux values while red colors denote high flux values.



**Fig. 9:** Global sensitivity analysis of the reduced order coagulation model with respect to the model parameters. A: Sensitivity analysis of the thrombin peak time for different prothrombin levels (150%,100% and 50% of the physiological value) as a function of activated protein C. B: Sensitivity analysis of the thrombin exposure for different prothrombin levels (150%,100% and 50% of the physiological value) as a function of activated protein C. Points denote the mean total sensitivity value, while the area around each point denotes the uncertainty in the sensitivity value. The gray dashed line denotes the 45° degree diagonal, if sensitivity values are equal for different conditions they will lie on the diagonal. Sensitivity values significantly above or below the diagonal indicate differentially important model parameters.