# Analysis of the Human Alternative Complement Pathway Architecture Identified Globally Fragile and Robust Components.

Adithya Sagar and Jeffrey D. Varner[*]

School of Chemical and Biomolecular Engineering

Cornell University, Ithaca NY 14853

**Running Title:** Analysis of the alternative complement cascade

**To be submitted:** PLoS ONE

[*]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

Phone: (607) 255 - 4258

Fax: (607) 255 - 9166

# Abstract

**Keywords:** Complement activation, Systems biology

# Introduction

Complement is the central subsystem of innate immunity activated as part of host defense. Complement destroys invading microorganisms and eliminates immune complexes **?**. Complement can also be activated during trauma and is correlated with negative clinical outcomes [REFME]. Complement, discovered more than a century ago [REFME], consists of more than 30 plasma and cell surface proteins and is organized into three pathways, the classical, lectin binding and alternative pathways **?**. In this study, we explored only the alternative activation pathway. The alternative pathway is considered to be the primary amplification mechanism for complement activation [REFME]. The alternative pathway consists of complement proteins numbered C3-C9, plasma factors B, D, H and I and several regulatory proteins restraining activation. C3 undergoes spontaneous hydrolysis in plasma to form a reactive protein intermediate $C3(H_2O)$. $C3(H_2O)$ is almost immediately inactivated by the regulatory proteins Factor H and I, however, before inactivation $C3(H_2O)$ can form a complex with Factor B thereby protecting itself against Factor H/I mediated inactivation. The $C3(H_2O)$-Factor B complex is cleaved by Factor D to form the fluid phase C3-convertase, denoted as C3Bb. C3Bb cleaves a 9-kDa fragment from the amino terminus of the $\alpha-$chain of free C3 to form two fragments, C3b and C3a. The C3b fragment tags foreign surfaces for destruction by the Membrane Attack Complex (MAC). Surface bound C3b binds circulating Factor B, forming a complex which is cleaved by Factor D to produce the surface C3-convertase, denoted by C3bBb. Fluid-phase and surface C3-converatase amplifies the complement response; C3Bb and C3bBb cleave C3 to form C3b and C3a. Binding of an addition C3b by C3bBb yields the C5-converatase, denoted as C3bBbC3b. The C5-convertase cleaves circulating C5 to form C5a and C5b fragments; C5b anchors itself on hydrophobic regions of the invading surface. C5b binding is the first step in the formation of the MAC; surface bound C5b sequentially binds C6 and C7 forming the C5b67 complex. C8 binds C5b67 via the $\beta-$chain of C5b. Binding of C8 triggers the polymerization of C9 proteins and the insertion of the poly-C9 structure into the membrane of the invading pathogen. Insertion of the poly-C9 structure forms a

pore that disrupts membrane integrity resulting in an influx of ions and water into the cell leading to cell death and lysis **?**.

In this study, we developed and analyzed a mechanistic mathematical model of the alternative complement pathway. The objective of our analysis was to better understand the fragility and robustness of the components making up C3-converatase amplification loop. The central computational challenge was the identification of model parameters.

# Results

**Identification of an ensemble of alternative complement models.** We assembled a canonical model of the molecular architecture of the alternative complement pathway (Fig. Z and given in detail in Table Z) from primary literature. The model, which described 53 molecular species interconnected by 104 interactions, was encoded as a system of non-linear ordinary differential equations. The rate of each of the 104 interactions contained in the model was modeled using elementary mass action kinetics; thus, each rate equation had a single rate constant. A master set of model parameters (rate constants and initial conditions) was taken from literature, or adjusted to make model simulations *qualitatively* consistent with observations. The master parameter set was used to generate synthetic data from which we identified a low-correlation ensemble of model parameters using Particle Swarm Optimization (PSO).

First-order sensitivity coefficients were calculated using the ODE15s routine of Matlab (The Mathworks, Natick MA) over a family of random parameter sets ($N = 100$) constructed by perturbing each nominal parameter by up $\pm 50\%$. The ODE15s solutions were used as high-order approximations of the correct senstivity solutiuon. Overall State Sensitivity Coefficients (OSSC) calculated using the ODE15s solutions were compared with sensitvity coefficients calculated using four additional methods over the same random parameter family (see Methods); Low-order Foward Finite Difference, FFD; low-order Backward Difference Formulation, BDF; high-order Interpolating Collocation, Collocation and a high-order Approximate Spectral Expansion, ASE. Complement mechanisms were rank-

ordered based upon their OSSC values generated by each method. The Spearman Rank Correlation was used to quantify the similarity between each of the rank-ordered lists and the ordering produced by the ODE15s method. Fixed-step methods produced results that were not consistent with ODE15s and required several iterations to correctly identify the proper perurbation size. The conclusions drawn from fixed-step methods were also found to be sensitive to model stiffness; both the classification of mechanisms as fragile and robust and the identifiablity of model parameters were effected by method choice.

# Discussion

# Materials and Methods

**Formulation and solution of model equations.** Mass balance equations were written around each protein or protein complex in the complement model yielding the set of nonlinear ordinary differential equations (vector-form):

$$\frac{d\mathbf{x}}{dt} = \mathbf{S}\mathbf{r}\left(\mathbf{x}, \mathbf{k}\right) \qquad \mathbf{x}\left(t_o\right) = \mathbf{x}_o \tag{1}$$

where $\mathbf{x}$ denotes the abundance vector (54×1), and $\mathbf{r}\left(\mathbf{x}, \mathbf{k}\right)$ denotes the vector of reaction rates (104×1). The matrix $\mathbf{S}$, which denotes the stoichiometric matrix (53×104), describes the biological connectivity of the model. Each row in $\mathbf{S}$ describes a particular protein or protein complex while each column describes the stoichiometry associated with a specific interaction in the network. The $(i, j)$ element of $\mathbf{S}$, denoted by $\sigma_{ij}$, describes how protein $i$ is connected to rate process $j$. If $\sigma_{ij} < 0$, then protein $i$ is consumed in $r_j$, conversely, if $\sigma_{ij} > 0$ protein $i$ is produced by $r_j$ and if $\sigma_{ij} = 0$ there is no connection between protein $i$ and rate process $j$. We have assumed mass action kinetics for each interaction in Table [RXNTABLE]. For the general reaction $q$:

$$\sum_{j \in \{\mathbf{R}_q\}} \sigma_{jq} x_j \rightarrow \sum_{k \in \{\mathbf{P}_q\}} \sigma_{kq} x_k \tag{2}$$

is given by:

$$r_q\left(\mathbf{x}, k_q\right) = k_q \prod_{j \in \{\mathbf{R}_q\}} x_j^{-\sigma_{jq}} \tag{3}$$

where $\{\mathbf{R}_q\}$ denotes the set of reactants for reaction $q$, $\{\mathbf{P}_q\}$ denotes the product set for reaction $q$, $k_q$ denotes the rate constant governing the qth reaction and $\sigma_{jq}, \sigma_{kq}$ denote stoichiometric coefficients (elements of the matrix $\mathbf{S}$). We have treated every rate as non-negative; all reversible reactions were split into two irreversible reaction steps. Thus, every element of the reaction rate vector $\mathbf{r}\left(\mathbf{x}, \mathbf{k}\right)$ takes the form shown in Eq. (**??**).

**Sensitivity and robustness analysis of the complement model population.** Sensitivity coefficients were calculated over the model ensemble (N = 500). First-order sensi-

tivity coefficients at time $t_q$:

$$s_{ij}(t_q) = \frac{\partial x_i}{\partial k_j}\bigg|_{t_q} \tag{4}$$

were computed by solving the kinetic-sensitivity equations **?**:

$$\begin{pmatrix} d\mathbf{x}/dt \\ d\mathbf{s}_j/dt \end{pmatrix} = \begin{bmatrix} \mathbf{S} \cdot \mathbf{r}(\mathbf{x}, \mathbf{k}) \\ \mathbf{A}(t)\,\mathbf{s}_j + \mathbf{b}_j(t) \end{bmatrix} \quad j = 1, 2, \ldots, P \tag{5}$$

subject to the initial condition $\mathbf{s}_j(t_0) = \mathbf{0}$. The quantity $j$ denotes the parameter index, $P$ denotes the number of parameters in the model, $\mathbf{A}$ denotes the Jacobian matrix, and $\mathbf{b}_j$ denotes the $j$th column of the matrix of first-derivatives of the mass balances with respect to the parameters. Sensitivity coefficients were calculated by repeatedly solving the extended kinetic-sensitivity system for forty parameters sets selected from the final 400 member ensemble. These sets were chosen to be comparable to the final 400 member ensemble on the basis of parametric coefficient of variation (CV); the sets selected for sensitivity analysis had a mean CV of $0.85 \pm 0.5$ and a mean correlation of approximately 0.6. Thus, there were diverse and uncorrelated. The Jacobian $\mathbf{A}$ and the $\mathbf{b}_j$ vector were calculated at each time step using their analytical expressions generated by UNIVERSAL.

The resulting sensitivity coefficients were scaled and time-averaged (Trapezoid rule):

$$\mathcal{N}_{ij} \equiv \frac{1}{T}\int_0^T dt \cdot |s_{ij}(t)| \tag{6}$$

where $T$ denotes the final simulation time. The time-averaged sensitivity coefficients were then organized into an array for each ensemble member:

$$\mathcal{N}^{(\epsilon)} = \begin{pmatrix} \mathcal{N}_{11}^{(\epsilon)} & \mathcal{N}_{12}^{(\epsilon)} & \cdots & \mathcal{N}_{1j}^{(\epsilon)} & \cdots & \mathcal{N}_{1P}^{(\epsilon)} \\ \mathcal{N}_{21}^{(\epsilon)} & \mathcal{N}_{22}^{(\epsilon)} & \cdots & \mathcal{N}_{2j}^{(\epsilon)} & \cdots & \mathcal{N}_{2P}^{(\epsilon)} \\ \vdots & \vdots & & \vdots & & \vdots \\ \mathcal{N}_{M1}^{(\epsilon)} & \mathcal{N}_{M2}^{(\epsilon)} & \cdots & \mathcal{N}_{Mj}^{(\epsilon)} & \cdots & \mathcal{N}_{MP}^{(\epsilon)} \end{pmatrix} \quad \epsilon = 1, 2, \ldots, N_\epsilon \tag{7}$$

where $\epsilon$ denotes the index of the ensemble member, $P$ denotes the number of parame-

ters, $N_\epsilon$ denotes the number of ensemble samples and $M$ denotes the number of model species. To estimate the relative fragility or robustness of species and reactions in the network, we decomposed the $\mathcal{N}^{(\epsilon)}$ matrix using Singular Value Decomposition (SVD):

$$\mathcal{N}^{(\epsilon)} = \mathbf{U}^{(\epsilon)}\Sigma^{(\epsilon)}\mathbf{V}^{T,(\epsilon)} \tag{8}$$

Coefficients of the left (right) singular vectors corresponding to largest $\theta \leq 15$ singular values of $\mathcal{N}^{(\epsilon)}$ were rank-ordered to estimate important species (reaction) combinations. Only coefficients with magnitude greater than a threshold ($\delta$ = 0.001) were considered. The fraction of the $\theta$ vectors in which a reaction or species index occurred was used to determine its importance (sensitivity ranking). The sensitivity ranking was compared between different conditions to understand how control in the network shifted as a function of perturbation or time.

# Acknowledgement