# Regression of Multicomponent Sticking Probabilities Using a Genetic Algorithm

**Ian J. Laurenzi***

*Department of Chemical Engineering, Lehigh University, Bethlehem, Pennsylvania 18015*

**John D. Bartels**

*Department of Mathematics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260*

**Scott L. Diamond**

*Department of Chemical Engineering, Institute for Medicine and Engineering, University of Pennsylvania, Philadelphia, Pennsylvania 19104*

A genetic algorithm (GA) was developed for the purpose of regressing composition-dependent aggregation kernels from time series of experimentally measured component or size distributions. The GA evolves initially random populations of kernel models in accordance with the principles of microevolution. To test the robustness of the GA, functionally diverse kernels—including one describing the shear-mediated aggregation of blood cells—were constructed. The stochastic time evolution of their corresponding aggregation processes were then simulated under physiological conditions via Monte Carlo. The GA successfully regressed the kernels underlying these "gold standard" datasets—where we employ the term in the sense of a "trusted reference"—from these simulation results, reproducing the multicomponent kernels to a maximum relative deviation of less than 9% over their entire composition ranges. Finally, ramifications of these cases pertinent to experimental design are considered, including the effects of extreme initial population ratios for multicomponent aggregation experiments with extreme population ratios.



## Introduction

The receptor-mediated aggregation of blood platelets and neutrophils plays an important role in cardiovascular disease.[1−4] Aggregates contribute to myocardial infarction, ischemia, thrombosis,[5−7] the progression of unstable angina,[8] and complications associated with extracorporeal circulation during surgery.[9]

Motivated by these medical issues, the aggregation of these blood cells has garnered considerable interest among biophysicists over the past twenty years. Among the more popular approaches to the study of the dynamics of this process is the "population balance" method,[10−15] wherein the time evolution of the aggregation in a closed system with fixed volume $V$ is specified by an equation such as

$$\frac{\partial c(u,v)}{\partial t} = \frac{1}{2}\int_0^v \int_0^u K(u',v',u-u',v-v')\, c(u',v')\, c(u-u',v-v')\, \mathrm{d}u'\, \mathrm{d}v' - c(u,v)\int_0^\infty K(u,v,u',v')\, c(u',v')\, \mathrm{d}u'\, \mathrm{d}v' \quad (1)$$

Here, $c(u,v)\,\mathrm{d}u\,\mathrm{d}v$ is the concentration of aggregates with a platelet volume content [$\mu$m$^3$] between $u$ and $u + \mathrm{d}u$ and neutrophil volume content [$\mu$m$^3$] between $v$ and $v + \mathrm{d}v$.[16−19] The aggregation kernel $K(u,v,u',v')$ is the intrinsic rate at which particles of composition $(u,v)$ and $(u',v')$ aggregate. By specification of the functional form of the kernel and the initial composition distribution $c(u,v,t=0)$, one may specify the time evolution of the entire aggregation process.

When particles aggregate in a linear shear field ($v_x = \gamma_S y$), the aggregation kernel possesses the form[20,21]

$$K(u,v,u',v') = \epsilon\frac{\gamma_S}{\pi}((u+v)^{1/3} + (u'+v')^{1/3})^3 \quad (2)$$

where $\gamma_S$ is the shear rate [s$^{-1}$] and $\epsilon$ is the "efficiency" or "sticking probability" for the interaction. When the aggregating particles are on the order of micrometers in radius, this parameter is a function of both the hydrodynamic and electrostatic forces between colliding aggregates.[21−29] However, because adhesion between platelets and neutrophils is mediated by surface-bound macromolecular receptors,[30−39] their aggregation is a function of the kinetics of formation and dissociation of those bonds.[28,29,40−43] Therefore, since each cell line features a unique set of surface-bound adhesion molecules, the sticking probability must feature a composition dependence, i.e.,

$$\epsilon(u,v,u',v') = \epsilon(y,y',\gamma_S) \quad (3)$$

where $y = u/(u+v)$ and $y' = u'/(u'+v')$.

In 1964, Swift and Friedlander published a method of determining the sticking probability for the single-component aggregation of latex beads.[21] Insofar as the aggregation process was homotypic, the sticking probability $\epsilon$ could be related directly to the particle consumption and computed directly from time series measurements of the particle size distribution. However, this elegant approach has not been amenable to the regression of multicomponent biological kernels on account of the explicit composition dependence of the kernels. In the case of the heterotypic aggregation of platelets and neutrophils in linear shear flow, the sticking probability must be a function of at least two variables ($y_1$ and $y_1'$) and three parametric constants: the sticking probability for the interactions of platelets with neutrophils $\epsilon_{plat-neut}$, the sticking probability for homotypic interactions of neutrophils $\epsilon_{neut}$, and the sticking probability for homotypic interactions of platelets $\epsilon_{plat}$. Thus, to obtain the functional dependence of the heterotypic sticking probability

* To whom correspondence should be addressed. Tel.: (610) 758-6835. Fax: (610) 758-5057. E-mail: ian.laurenzi@lehigh.edu.

10.1021/ie051159t CCC: $33.50 
Published on Web 04/25/2006

$\epsilon(y,y')$ from experimental data, the PBE itself must be solved and inverted. However, this is not a trivial task: the infinite set of nonlinear partial-differential integral equations represented by eq 1 rarely admits analytical solution, and Monte Carlo (MC) simulations of the stochastic time evolution of the aggregation process are required.[17,18,44] Therefore, a more general regression algorithm is necessary to specify the functional dependence of the heterotypic sticking probability $\epsilon(y,y')$ from experimental data.

In this work, we employ the genetic algorithm (GA) to solve the "inverse problem" of deconvoluting the functional form of the sticking probability from timeseries measurements of component size distributions of platelet−neutrophil aggregates. The genetic algorithm (GA) is a method of minimizing or maximizing functions using the principles of biological evolution and has been used for optimizations as diverse as kinetic model fitting[45−47] and chemical plant design.[48,49]

## Methods

**Generation of "Gold Standard" Datasets.** Under different experimental, physiological, or pathological conditions, the sticking probabilities between aggregates of platelets and neutrophils may differ. However, our interest lies in the development of an algorithm to regress these probabilities without restricting our scope to a specific experimental system. Hence, we performed MC simulations of two qualitatively dissimilar models of $\epsilon(y_1,y_2)$. The resulting sets of time series of composition distributions (our trusted reference datasets or gold standards) were employed as proxy experiments, akin to the composition distributions of platelet−neutrophil[4] and platelet−LS174T[50] aggregates measured via flow-cytometry.

The two models employed the shear kernel defined by eqs 2 and 3 with sticking probabilities featuring the following composition dependence:

$$\epsilon(y,y') = yy'C_1 + (y'(1-y) + y(1-y'))C_2 + (1-y)(1-y')C_3 \quad (4)$$

This model has a microphysical interpretation as a "single-bond" model, whereby $y$ acts as the probability that there is a single platelet in a particle of composition $(u,v)$, in accord with the symmetry property of all aggregation kernels, $\epsilon(y,y') = \epsilon(y',y)$, regardless of the values of $C_1$, $C_2$, and $C_3$.

The first gold standard model for which data were simulated (E1) was designed to emulate the behavior of true heterotypic aggregation of platelets and neutrophils in shear at $\gamma_S = 335$ $s^{-1}$. At this shear rate, $C_1$—the homotypic sticking probability for platelets $\epsilon_{plat}$—and $C_3$—the homotypic sticking probability $\epsilon_{neut}$ for neutrophils—are 0.05 and 0.31, respectively.[15,51,52] The value of the heterotypic sticking probability between platelet and neutrophil singlets $\epsilon_{plat-neut}$ was estimated to be 0.5.

The second gold standard model (E2) was chosen to have significantly different features than the first: parameters of 0.5, 0.1, and 0.6 were selected for $C_1$, $C_2$, and $C_3$, respectively, to favor homotypic aggregation of the two "cells" over heterotypic aggregation. Moreover, the homotypic parameters were specifically chosen to be similar in magnitude.

**Monte Carlo Simulation.** To obtain time series of the composition distributions of cellular aggregates that behave according to models E1 and E2, we conducted stochastic simulations of their aggregation processes. We employed the stochastic simulation algorithm (SSA) of Laurenzi and coworkers, which is based on Gillespie's direct method.[17,18,44,53,54] This method simulates the stochastic time evolution of the aggregation process using the kernels directly, avoiding the solution of a population balance equation (PBE) such as eq 1. The fundamental axiom of the approach is

$$\frac{K(u,v,u',v')}{V}\delta t + o(\delta t) = \Pr(\text{two specific particles of compositions } (u,v) \text{ and } (u',v') \text{ will aggregate within the next time interval } \delta t) \quad (5)$$

If there is more than one particle of each aggregation species as defined by composition, then it may be shown that[53,55]

$$\frac{K(u,v,u',v')}{V}h(X(u,v),X(u',v'))\delta t + o(\delta t) = \Pr(\text{any two particles of compositions } (u,v) \text{ and } (u',v') \; (u',v') \text{ will aggregate within the next time interval } \delta t) \quad (6)$$

where $h$ is the total number of ways that those particles can aggregate

$$h(X(u,v),X(u',v')) = \begin{cases} X(u,v)\,X(u',v') & (u,v) \neq (u',v') \\ \binom{X(u,v)}{2} & (u,v) = (u',v') \end{cases} \quad (7)$$

and $X(u,v)$ is the total number of particles with composition $(u,v)$.

By judicious use of these axioms, one can construct probability densities for the time until the next event (quiescence time) $\tau$ and the specific event to come, such as the interaction of a platelet of size $u$ with a neutrophil of size $v$.[17,18,44] By repetitive Monte Carlo selections of quiescence times and events from these probability densities and intermediating "updates" of species population and reaction rates concomitant with population change (cf. eqs 6 and 7), the entire time evolution of any aggregation process may be simulated. We note in particular that the derivation of the algorithm is exact, involving no numerical approximations, and remarkably fast due to innovations of data structures for aggregating systems.[18,44] The simulations return species populations $X(u,v)$ over a desired range of time points, which may be related to concentrations by division by $V$.

All simulations were performed using volumetrically homogeneous platelets (7.68 [$\mu m^3$]) and neutrophils (290.1 [$\mu m^3$]) at populations of 36 000 and 1000 cells, respectively. To approximate experimental error, redundant simulations were not performed, and the results of individual simulations of the models were employed as sets of proxy experimental results for GA regression.

The results of these simulations tend to be statistically indistinguishable from analytical solutions of eq 1 unless (a) the kernel grows geometrically in size or (b) the number of monomers of a component is too small. In the simulations of our model systems, the number of particles was sufficiently large enough to reproduce bulk behaviors. Simulations were performed on a 1.2 GHz Dell desktop computer running RedHat Linux and typically required only a few seconds of CPU time.

**Genetic Algorithm.** The first GA is attributed to John Holland,[56] who devised electronic entities whose "evolution" was guided by their conformity to a fitness measure. By correlating these quantities to mathematical statements and formulas, he was able to use the concept of natural selection to optimize functions. Since the early 1980s, GAs have been used extensively in physics, economics, chemistry, and their counterparts in engineering for optimization and design.[57]

The task of fitting these parameters to experimental or simulated experiments is nontrivial. Inasmuch as the free parameters characterize a kinetic process, a model should be able to predict a set of experimental results at multiple time points. If the derivatives of such an objective functions were known, the traditional second derivative test could be employed to minimize the difference between model and experiment. However, owing to the noise of both the results of simulations of models and the noise of experimental results, such derivatives cannot be determined in this case. Because GAs do not require derivatives of objective functions, they are well suited to the regression of such noisy data to stochastic simulation results. We then investigated whether a simple GA could evolve a kernel to fit new simulation results to the VERs resulting from simulations of models E1 and E2. To evaluate this, we defined a model as the ordered triple ($C_1$, $C_2$, $C_3$). We also supplement the model with the constraint that $C_1$, $C_2$, and $C_3$ are limited to values between 0 and 1, guaranteeing that $\epsilon$, $\epsilon_{neut}$, $\epsilon_{plat-neut}$, and $\epsilon_{plat}$ are all positive and less than one. By defining the model in this way, we also establish the first step of the GA.

**Gene Pool.** The GA requires a gene pool—a population of model kernels that will evolve to maximize the fit between the experiment and the simulation. The first step of the GA is the construction of a population of model extrema

$$\begin{aligned} s_1 &= (C_{1,1},C_{2,1},C_{3,1}) \\ s_2 &= (C_{1,2},C_{2,2},C_{3,2}) \\ &\vdots \\ s_n &= (C_{1,n},C_{2,n},C_{3,n}) \end{aligned} \tag{8}$$

randomly generated on the possible ranges of $C_1$, $C_2$, and $C_3$. In our GA, we used a population of 200 models. We denote $C_{1,i}$, $C_{2,i}$, and $C_{3,i}$ as the *genes* of model $i$ and the set $(C_{1,i},C_{2,i},C_{3,i})$ as its *chromosome*.

**Evaluation of Fitness.** Keeping in mind the physical interpretations of $C_1$, $C_2$, and $C_3$ discussed previously, we sought to fit them to the simulated models E1 and E2 via an objective function, to evaluate, the *fitness* of each model. We defined the fitness the $i$th model, $f_i$, as a measure of the time-averaged percentage of correctly allocated mass

$$f_i = \frac{1}{4}\sum_{k=1}^{4}\frac{\sum_m\sum_n (m+n)\,\min(X_{expt}(m,n,t_k),X_{s_i}(m,n,t_k))}{\sum_m\sum_n (m+n)X_{expt}(m,n,t_k)} \tag{9}$$

In eq 9, $X_{expt}(m,n,t_k)$ is the simulated composition distribution at time $t_k$ for either model E1 or E2 and $X_{s_i}(m,n,t_k)$ is the corresponding distribution for model $s_i$. Like the time-dependent composition distributions for models E1 and E2, $X_{s_i}(m,n,t)$ is generated via MC simulation of model $s_i$ with the same initial conditions as models E1 and E2.

Models with negative fitness were discarded at the beginning of the procedure, and new models were generated until the gene pool was set. In the results of models E1 and E2, measurements were taken at four time points at equal intervals. Hence, exact models from the "gene pool" will allocate the components almost identically to experimental composition distributions at all times, resulting in a fitness of one. By contrast, models that are nonpredictive or predictive only at large or small $t$ will have fitness values close to zero.

**Selection and Crossover.** Once the population was defined and each model evaluated, the GA selected pairs of models according to their fitness and employed "crossover" to generate new models. Like the selection of the event to come in the SSA, both selection and crossover were treated *probabilistically*. Here, the selection criterion followed from the event selection method of the SSA:

$$\frac{f_k}{\sum_{i=1}^{n} f_i} = \Pr(\text{the } k\text{th parent will procreate in the coming reproductive cycle}) \tag{10}$$

Equation 10 balances the relative probabilities of the coming event and favors "more fit" models for reproduction. The "parent" models were then chosen like the reaction events to come in the MC simulation algorithm for aggregation

$$\begin{aligned} \sum_{i=1}^{k-1} f_i \le r_1\sum_{i=1}^{n} f_i < \sum_{i=1}^{k} f_i \quad &\text{parent 1 selection} \\ \sum_{i=1,i\neq k}^{l-1} f_i \le r_2\sum_{i=1,i\neq k}^{n} f_i < \sum_{i=1,i\neq k}^{l} f_i \quad &\text{parent 2 selection} \end{aligned} \tag{11}$$

where $r_1$ and $r_2$ are random numbers on (0,1].

Upon selection of parent models, a new "child" model is created from them by crossover. In the context of the GA, crossover has a significantly different meaning than in nature. Upon fusion of two haploid cells, typically from different parents, the resulting (diploid) child cell possesses half of each parent's alleles. Our GA cannot produce child models in exactly this way, since each model is "haploid". To effect a mixing of parental information to produce a child model, we needed to ensure that different genes were not mixed. That is, none of one parent's $C_2$ genes should mix with the child's $C_1$ gene. Our solution was the following: if models $j$ and $k$ are the chosen parents, the crossover mechanism of our GA was defined by the following formula

$$s_{child} = \begin{cases} (rC_{1,j} + (1-r)C_{1,k},C_{2,k},C_{3,k}) & 0 \le r < {}^1/_3 \\ (C_{1,j},(r-{}^1/_3)C_{2,j} + ({}^4/_3 - r)C_{2,k},C_{3,k}) & {}^1/_3 \le r < {}^2/_3 \\ (C_{1,j},C_{2,j},(r-{}^2/_3)C_{3,j} + ({}^5/_3 - r)C_{3,k}) & {}^2/_3 \le r < 1 \end{cases} \tag{12}$$

The genes are distributed probabilistically from the parents in a way similar to the biological process. Moreover, this crossover formula does not allow accidental generation of illegal values for the parameters $C_1$, $C_2$, and $C_3$.

**Mutation.** To assist in the escape from local minima of the objective function, a mutation step was added to the procedure of generating new models from their parents. In nature, mutation most certainly has helped organisms adapt to their surroundings: If there is a mutation in a gene of a living organism (i.e., a change in its sequence such as ...GATC... from ...GTTC...), there may be consequences for the structure and function of the protein encoded by that gene. An abundance of literature testifies to the significance of single nucleotide polymorphisms (SNPs) to the function of proteins from which mutated genes are translated. If the floating point numbers we designated as genes are represented bitwise, then if a bit at the upper end is changed, there will be a change in the first significant digit of the corresponding floating point value. Likewise, if the last bit is changed, then the entire number becomes negative if the

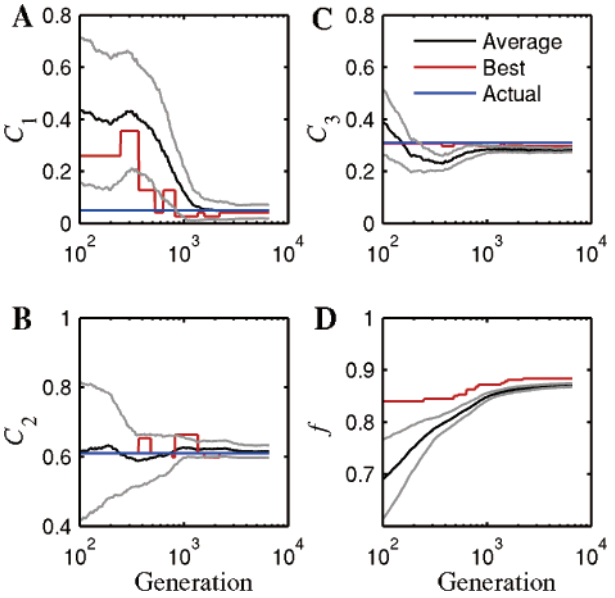


**Figure 1.** Evolution of the "$C_1$ gene" ($\epsilon_{plat}$), "$C_2$ gene" ($\epsilon_{plat-neut}$), "$C_3$ gene" ($\epsilon_{neut}$), and fitness $f$ as the GA evolves a population of aggregation models for the heterotypic aggregation of human blood platelets and neutrophils to model E1. The values corresponding to the best (most fit) model and the population mean are tracked with each generation. Gray lines are the population average ± SD.



computer stores numbers using twos complements. However, intermediate bits may be of little to no significance, i.e., changing the floating point value in the fourth significant digit.

Due to these considerations, and in particular the twos complement issue, we choose to randomly generate a new gene on [0,1) if the GA dictates that mutation will occur. Like the desicion for the crossover point, the decision to mutate is made by MC. For each of the genes of a new child model, a random number $r \in (0,1]$ was chosen. If this number $r$ is greater than 0.95, that gene will be replaced as discussed. By choosing the tolerance 0.95, mutation was ensured to be both rare and sufficient for the generation of models that can escape from local maxima in fitness.

**Insertion.** Once the new model was generated by crossover of the parents and possibly mutated, its fitness was evaluated by stochastic simulation. If the fitness was nonnegative, the child model was added to the population. A fixed population size was maintained by displacing the least fit member of the population at each new generation, weeding out less fit models over successive generations.

## Results and Discussion

In Figure 1 we present typical results for the regression of model E1 from its simulation results. At the beginning of the GA regression, the model population average (MPA) values of $C_1$, $C_2$, and $C_3$ are all approximately 0.5 due to the fact that the initial genes were uniformly randomly generated on (0, 1]. With each generation or creation of a new model, the model populations became more and more homogeneous: in Figure 1A−D, the gray lines representing the means ± standard deviation (SD) monotonically approach the mean with each generation.

The convergence of the mean values of the parameters, however, is not monotonic. After 200 generations, the initial population was completely replaced in accordance with the population constraint of the GA but the best model had not been replaced. The evolution of a more fit model occurred between

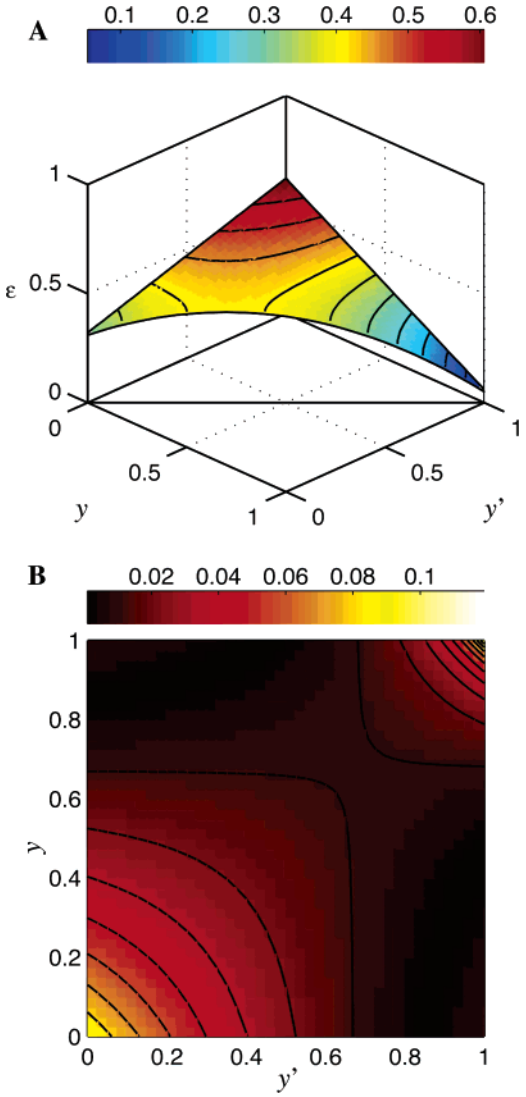


**Figure 2.** Convergence of the GA to model E1. (A) Gold standard model E1. The optimal values found were $\hat{C}_1 = 0.044 \pm 0.026$, $\hat{C}_2 = 0.615 \pm 0.017$, and $\hat{C}_3 = 0.283 \pm 0.026$ (population mean ± SD), which fall at $(y,y')$ coordinates of (0,0), (0,1), and (1,1), respectively. The actual values of the parameters are $C_1 = 0.05$, $C_2 = 0.61$, and $C_3 = 0.31$. The surface reflects about the line $y = y'$; values are shown for $y' < y$. (B) Fractional difference between the GA-regressed and true sticking probabilities $|\epsilon_{E1} - \hat{\epsilon}_{E1}|/\epsilon_{E1}$. The relatively large deviation for $C_3$ is a result of a large platelet/neutrophil ratio of 36:1 in all model simulations that minimizes the effects of homotypic neutrophilic aggregation, diminishing the sensitivity to $C_3$.

200 and 300 generations, but it actually produced a less accurate value of $C_1$. This escape from a local minimum then allowed the population to relax to a new minimum, and after the passing of 300 generations, the convergence was fairly monotonic.

The GA succeeded in reproducing all three constants for model E1. After 3000 generations, the population appeared to reach steady state, with optimal point estimates of $\hat{C}_1 = 0.044 \pm 0.026$, $\hat{C}_2 = 0.615 \pm 0.017$, and $\hat{C}_3 = 0.283 \pm 0.026$ (population mean ± SD). Of the GA estimates of the three parameters, those for $\hat{C}_1$ and $\hat{C}_2$ were the most accurate. The estimate of $C_3$—although less precise—was sufficiently accurate; the true value of $C_3$ for model E1 was 0.31. However, if one compares the functions for the sticking probability for the regressed and E1 models, the convergence is remarkably good (Figure 2). The worst case relative deviation $|\epsilon_{E1} - \hat{\epsilon}_{E1}|/\epsilon_{E1}$ corresponds to a percentage difference between $\epsilon_{E1}(0,1) = C_3$ and $\hat{\epsilon}_{E1}(0,1) = \hat{C}_3$ of only 8.7%. Generally, the relative deviation

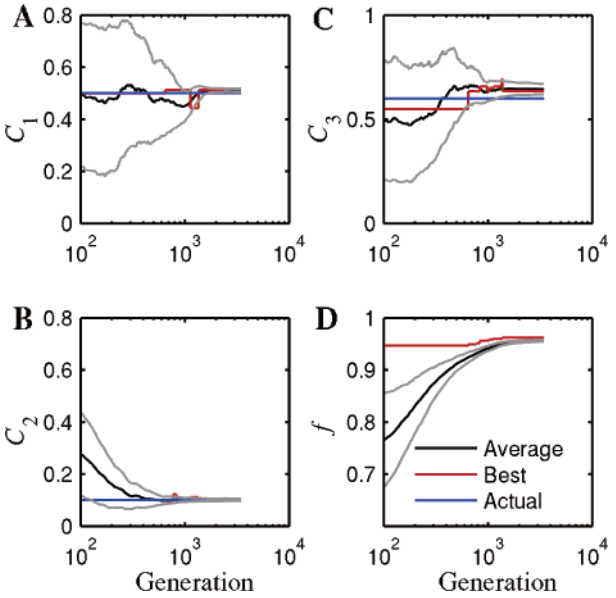


**Figure 3.** Evolution of the "$C_1$ gene" ($\epsilon_{plat}$), "$C_2$ gene" ($\epsilon_{plat-neut}$), "$C_3$ gene" ($\epsilon_{neut}$), and fitness $f$ as the GA evolves a population of aggregation models for the heterotypic aggregation of human blood platelets and neutrophils to model E2. The values corresponding to the best (most fit) model and the population mean are tracked with each generation. Gray lines are the population average ± SD.

between the true and regressed sticking probabilities over all $y$ and $y'$ is less than 3%.

Further study revealed the lower precision of $\hat{C}_3$ to be a consequence of (a) the heterotypic nature of the aggregation process described by eq 7 and (b) the small initial population of neutrophils. At the beginning of both the GA and gold standard model simulations, the 36:1 ratio of platelets to neutrophils causes most neutrophils to aggregate heterotypically, strictly out of population considerations (cf. eqs 6 and 7).[17] Moreover, the high platelet population and higher number of platelet combinations for small $t$ guarantee significant homotypic platelet aggregation even in the presence of neutrophils. Later, as the aggregation process proceeds and homotypic platelet aggregates aggregate with heterotypic aggregates out of entropic considerations, all aggregations tend to be heterotypic. Consequently, the heterotypic sticking probability $C_2$ and homotypic platelet sticking probability $C_1$ are the most important parameters in the process; the final results are not as sensitive to $C_3$. Hence, the GA does not resolve $C_3$ to the precision of the other parameters.

Results of a typical regression of model E2 are presented in Figure 3. $C_1$ and $C_2$ were reproduced very well and with high precision, whereas the GA estimate for $C_3$ is not as precise. After 3439 generations, the GA resolved $\hat{C}_1 = 0.512 \pm 0.006$, $\hat{C}_2 = 0.102 \pm 0.004$, and $\hat{C}_3 = 0.644 \pm 0.014$ (population mean ± SD), although the convergence is more or less achieved after 1500 generations. As with the regression of model E1, small differences between the parameters used to generate the "model time series" and the parameters of the population average model are insignificant over the entire range of aggregate compositions $y$ and $y'$: the relative deviation $|\epsilon_{E2} - \hat{\epsilon}_{E2}|/\epsilon_{E2}$ is at most 7% (Figure 4).

In as much as the form of model E2 is has a different type of curvature (i.e., an inverted saddle point with respect to model E1) and different magnitudes and ratios of the homotypic extrema, we conclude that extreme ratios of the initial cellular population are a potential cause of precision loss when regressing $\epsilon(y,y')$ from experimental composition distributions via the

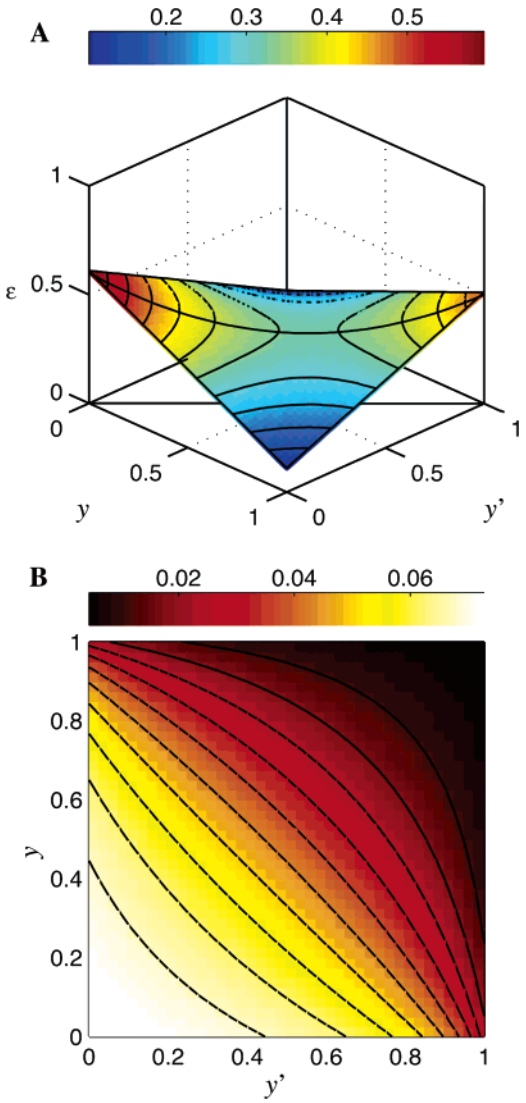


**Figure 4.** Convergence of the GA to model E2. (A) Model E1. The optimal values found were $\hat{C}_1 = 0.512 \pm 0.006$, $\hat{C}_2 = 0.102 \pm 0.004$, and $\hat{C}_3 = 0.644 \pm 0.014$ (population mean ± SD), which fall at $(y,y')$ coordinates of (0,0), (0,1), and (1,1), respectively. The actual values of the parameters are $C_1 = 0.5$, $C_2 = 0.1$, and $C_3 = 0.6$. (B) Fractional difference between the GA-regressed and true sticking probabilities $|\epsilon_{E2} - \hat{\epsilon}_{E2}|/\epsilon_{E2}$. As with the results for model E1, the relatively large deviation for $C_3$ is a result of a large platelet/neutrophil ratio of 36:1 in all model simulations that minimizes the effects of homotypic neutrophilic aggregation, diminishing the sensitivity to $C_3$.

GA. With this caveat, we note that the ultimate results of the GA are remarkably robust, as illustrated in Figures 2B and 4B.

To improve experimental designs for the measurement of $\epsilon(y,y')$, we suggest that experiments be run with varying concentrations of neutrophils. This may be achieved, for example, by isolation of neutrophils from venous blood via ficoll hypaque[58] and resuspending them with varying amounts of platelet-rich plasma prior to aggregation. Each diluted sample may then be aggregated, yielding composition distribution time series akin to the simulation results of models E1 and E2, which may then be used to calculate $\epsilon(y,y')$ via the GA. The constants regressed can then be compared via ANOVA or another method to systematically quantify the effect of the cell population ratio on the GA sensitivity.

Although we have not addressed the matter explicitly in this work, we note that the sticking probability may be a function of both the compositions of colliding aggregates as well as the shear rate of the suspending medium. However, depending on

the dynamics of tether bond formation and cell type, sticking probabilities may have a broad range of shear behaviors.[14,15,50] Therefore, although it is possible to modify the models and GA to accommodate the regression of experiments at multiple shear rates, we believe that it is more informative to perform regressions at each shear rate and then analyze the model constants for their shear dependence.

## Conclusions

We have developed a genetic algorithm to facilitate the kinetic characterization of heterotypic aggregation processes. Given a time series of cellular composition or size distributions such as may be measured via flow-cytometry or a Coulter counter, the GA regresses a model of the aggregation kernel for the process.

To obtain a diverse set of gold standard data for testing the GA, we developed unique models of aggregation of blood platelets and neutrophils and simulated these via the stochastic simulation algorithm of Laurenzi and co-workers.[17,18] The resulting time series of composition distributions were then processed via the GA to regress the original models. The GA was capable of fitting multicomponent kernels to experimental composition distributions with remarkable accuracy. Heterotypic kernels were reproduced very well over their entire composition range, with maximum relative deviations lower than 9% for both gold standard models.

However, we discovered in the process that the design of such aggregometry experiments is critical: the sensitivity of the results may be affected by the initial relative numbers of cells in an experiment. Should this ratio favor one of the cell lines too strongly, the composition dependence of the kernel for the other cell line may be reproduced less reliably. In our example of the heterotypic aggregation of neutrophils and platelets in venous blood, wherein the ratio of platelets to neutrophils is 36:1, we found that the relative error of the sticking probability returned by the GA could be as high as 9%. Hence, we suggest care in choosing the relative concentrations of blood cells prior to aggregometry and/or flow-cytometry.

We suggest that the GA described may be useful in the "kinetic fingerprinting" of human blood, whereby variations between individuals' heterotypic cellular sticking probabilities could be characterized as a diagnostic measure of immunological or thrombotic activity. Moreover, due to the general applicability of the stochastic simulation algorithm on which it is based and the potential to use a wide variety of other kernel models, we believe it may have considerable impact in the characterization of colloidal, atmospheric, and polymer systems in which multicomponent aggregation plays a central role.



## Acknowledgment

This work was supported by National Institutes of Health Grant HL 56621 and National American Heart Association Grant 96-6670. S.L.D. is an Established Investigator of the National American Heart Association.

## Notation

$c(u,v)\ \mathrm{d}u\ \mathrm{d}v$ = concentration of aggregates with a platelet volume content ($\mu$m$^3$) between $u$ and $u + \mathrm{d}u$ and neutrophil volume content ($\mu$m$^3$) between $v$ and $v + \mathrm{d}v$

$C_1$, $C_2$, $C_3$ = model constants

$\hat{C}_1$, $\hat{C}_2$, $\hat{C}_3$ = point estimates of model constants

$h(X(u,v),X(u',v'))$ = number of ways that $(u,v)$-mers and $(u',v')$-mers can aggregate

$K(u,v,u',v')$ = aggregation kernel ($\mu$m$^3$/s)

$t$ = time (s)

$r$, $r_1$, $r_2$ = uniform random numbers on (0,1]

$f_i$ = fitness of model $i$

$s_i$ = ordered set of parameters constituting model $i$

$u$, $u'$ = platelet volume content ($\mu$m$^3$) of an aggregate

$v$, $v'$ = neutrophil volume content ($\mu$m$^3$) of an aggregate

$V$ = batch volume

$X(u,v)$ = population of aggregates with platelet volume content ($\mu$m$^3$) $u$ and neutrophil volume content ($\mu$m$^3$) $v$

$y$, $y'$ = platelet volume fraction of an aggregate

$\epsilon$ = efficiency or sticking probability

$\epsilon_{plat}$ = homotypic sticking probability for platelets

$\epsilon_{neut}$ = homotypic sticking probability for neutrophils

$\epsilon_{plat-neut}$ = heterotypic sticking probability for interactions between platelet and neutrophil singlets

$\gamma_S$ = shear rate (s$^{-1}$)

$\tau$ = quiescence time

## Literature Cited

(1) Hamburger, S. A.; McEver, R. P. GMP-140 mediates adhesion of stimulated platelets to neutrophils. *Blood* **1990**, *75*, 550.

(2) Rinder, H. M.; Bonan, J. L.; Rinder, C. S.; Ault, K. A.; Smith, B. R. Activated and unactivated platelet adhesion to monocytes and neutrophils. *Blood* **1991**, *78*, 1760.

(3) Evangelista, V.; Manarini, S.; Rotondo, S.; Martelli, N.; Polischuk, R.; McGregor, J. L.; de-Gaetano, G.; Cerletti, C. Platelet/polymorphonuclear leukocyte interaction in dynamic conditions: evidence of adhesion cascade and cross talk between P-selectin and the beta 2 integrin CD11b/CD18. *Blood* **1996**, *88*, 4183

(4) Konstantopoulos, K.; Neelamegham, S.; Burns, A. R.; Hentzen, E.; Kansas, G. S.; Snapp, K. R.; Berg, E. L.; Hellums, J. D.; Smith, C. W.; McIntire, L. V.; Simon, S. I. Venous levels of shear support neutrophil-platelet adhesion and neutrophil aggregation in blood via P-selectin and beta(2)-integrin. *Circulation* **1998**, *98*, 873.

(5) Bednar, M.; Smith, B.; Pinto, A.; Mullane, K. M. Neutrophil depletion suppresses 111In-labeled platelet accumulation in infarcted myocardium. *J. Cardiovasc. Pharmacol.* **1985**, *7*, 906.

(6) Bednar, M.; Smith, B.; Pinto, A.; Mullane, K. M. Nafazatrom-induced salvage of ischemic myocardium in anesthetized dogs is mediated through inhibition of neutrophil function. *Circ. Res.* **1985**, *57*, 131.

(7) Neumann, F. J.; Marx, N.; Gawaz, M.; Brand, K.; Ott, I.; Rokitta, C.; Sticherling, C.; Meinl, C.; May, A.; Schomig, A. Induction of cytokine expression in leukocytes by binding of thrombin-stimulated platelets. *Circulation* **1997**, *95*, 2387.

(8) Ott, I.; Neumann, F. J.; Gawaz, M.; Schmitt, M.; Schomig, A. Increased neutrophil-platelet adhesion in patients with unstable angina. *Circulation* **1996**, *94*, 1239.

(9) Rinder, C. S.; Bonan, J. L.; Rinder, H. M.; Matthew, J.; Hines, R.; Smith, B. R. Cardiopulmonary bypass induces leukocyte-platelet adhesion. *Blood* **1992**, *79*, 1201.

(10) Belval, T. K.; Hellums, J. D. Analysis of shear induced platelet aggregation with population balance mathematics. *Biophys. J.* **1986**, *50*, 479.

(11) Huang, P. Y.; Hellums J. D. Aggregation and disaggregation kinetics of human blood platelets. I. Development and validation of a population balance method. *Biophys. J.* **1993**, *65*, 334.

(12) Huang, P. Y.; Hellums J. D. Aggregation and disaggregation kinetics of human blood platelets. II. Shear-induced platelet aggregation. *Biophys. J.* **1993**, *65*, 344.

(13) Huang, P. Y.; Hellums, J. D. Aggregation and disaggregation kinetics of human blood platelets. III. Disaggregation under shear stress of platelet aggregation. *Biophys. J.* **1993**, *65*, 354.

(14) Neelamegham, S.; Taylor, A. D.; Hellums, J. D.; Dembo, M.; Smith, C. W.; Simon, S. I. Modeling the reversible kinetics of neutrophil aggregation under hydrodynamic shear. *Biophys. J.* **1997**, *72*, 1527.

(15) Taylor, A. D.; Neelamegham, S.; Hellums, J. D.; Smith, C. W.; Simon, S. I. Molecular dynamics of the transition from L-selectin to 2-integrin dependent neutrophil adhesion under defined hydrodynamic shear. *Biophys. J.* **1996**, *71*, 3488.

(16) Lushnikov, A. A. Evolution of coagulating systems III. Coagulating mixtures. *J. Colloid Interface Sci.* **1976**, *54*, 94.

(17) Laurenzi, I. J.; Diamond, S. L. Monte Carlo simulation of the heterotypic aggregation kinetics of platelets and neutrophils. *Biophys. J.* **1999**, *77*, 1733.

(18) Laurenzi, I. J.; Bartels, J. D.; Diamond, S. L. A General Algorithm for Exact Simulation of Multicomponent Aggregation Processes. *J. Comput. Phys*. **2002**, *177*, 418.

(19) Wright, D. L.; McGraw, R.; Rosner, D. E. Bivariate extension of the quadrature method of moments for modeling simultaneous coagulation and sintering of particle populations. *J. Colloid Interface Sci*. **2001**, *236*, 242.

(20) Smoluchowski, M.v. Vensuch einer mathematischen theorie der koagulation kinetic kolloider losungen. *Z. Phys. Chem*. **1917**, *92*, 129.

(21) Swift, D. L.; Friedlander, S. K. The coagulation of hydrosols by brownian motion and laminar shear flow. *J. Colloid Sci.* **1964**, *19*, 621.

(22) Batchelor, G. K.; Green, J. T. The hydrodynamic interaction of two small freely moving spheres in a linear flow field. *J. Fluid Mech.* **1972**, *56*, 375.

(23) Wakiya, S. Slow motion in shear flow of a doublet of two spheres in contact. *J. Phys. Soc. Jpn.* **1971**, *31*, 1581.

(24) Zeichner, G. R.; Schowalter, W. R. Use of trajectory analysis to study stability of colloidal dispersions in flow fields. *AIChE J*. **1977**, *23*, 243.

(25) Adler, P. M. Interaction of unequal spheres. I. Hydrodynamic interactions: colloidal forces. *J. Colloid Interface Sci.* **1981**, *84*, 461.

(26) Adler, P. M. Heterocoagulation in shear flow. *J. Colloid Interface Sci*. **1981**, *83*, 106.

(27) Van de Ven, T. M. *Colloidal Hydrodynamics*; Academic Press: London. 1989.

(28) Tandon, P.; Diamond, S. L. Hydrodynamic effects and receptor interactions of platelets and their aggregates in linear shear flow. *Biophys. J*. **1997**, *73*, 2819.

(29) Tandon, P.; Diamond, S. L. Kinetics of $\beta_2$-integrin and L-selectin bonding during neutrophil aggregation in shear flow. *Biophys. J*. **1998**, *75*, 3163.

(30) Staunton, D. E.; Dustin, M. L.; Springer, T. A. Functional cloning of ICAM-2, a cell adhesion ligand for LFA-1 homologous to ICAM-1. *Nature* **1989**, *339*, 61.

(31) Moore, K. L.; Stults, N. L.; Diaz, S.; Smith, D. L.; Cummings, R. D.; Varki, A.; McEver, R. P. Identification of a specific glycoprotein ligand for P-selectin (CD62) on myeloid cells. *J. Cell Biol.* **1992**, *118*, 445.

(32) Palabrica, T.; Lobb, R.; Furie, B. C.; Aronovitz, M.; Benjamin, C.; Hsu, Y. M.; Sajer, S. A.; Furie, B. Leukocyte accumulation promoting fibrin deposition is mediated in vivo by P-selectin on adherent platelets. *Nature.* **1992**, *359*, 848.

(33) Tu, L. L.; Chen, A. J.; Delahunty, M. D.; Moore, K. L.; Watson, S. R.; McEver, R. P.; Tedder, T. F. L-selectin binds to P-selectin glycoprotein ligand-1 on leukocytes. Interactions between the lectin, epidermal growth factor, and consensus repeat domains of the selectins determine ligand binding specificity. *J. Immunol*. **1996**, *157*, 3995.

(34) Walcheck, B.; Moore, K. L.; McEver, R. P.; Kishimoto, T. K. Neutrophil−neutrophil interactions under hydrodynamic shear stress involve L-selectin and PSGL-1: a mechanism that amplifies initial leukocyte accumulation on P-selectin in vitro. *J. Clin. Invest.* **1996**, *98*, 1081.

(35) Spertini, O.; Cordey, A. S.; Monai, N.; Giuffre, L.; Schapira, M. P-selectin glycoprotein ligand-1 (PSGL-1) is a ligand for L-selectin on neutrophils, monocytes and CD341 hematopoietic progenitor cells. *J. Cell Biol*. **1996**, *135*, 523.

(36) Lynam, E.; Sklar, L. A.; Taylor, A. D.; Neelamegham, S.; Bruce, S.; Edwards, C.; Smith, W.; Simon, S. I. $\beta$ 2-Integrins mediate stable adhesion in collisional interactions between neutrophils and ICAM-1-expressing cells. *J. Leukocyte Biol*. **1998**, *64*, 622.

(37) Romo, G. M.; Dong, J. F.; Schade, A. J.; Gardiner, E. E.; Kansas, G. S.; Li, C. Q. McIntire, L. V.; Berndt, M. C.; Lopez, J. A. The glycoprotein Ib-IX-V complex is a platelet counterreceptor for P-selectin. *J. Exp. Med*. **1999**, *190*, 803.

(38) Simon, D. I.; Chen, Z. P.; Xu, H.; Li, C. Q.; Dong, J. F.; McIntire, L. V.; Ballantyne, C. M.; Zhang, L.; Furman, M. I.; Berndt, M. C.; Lopez, J. A. Platelet glycoprotein Ib alpha is a counter-receptor for the leukocyte integrin Mac-1 (CD11b/CD18). *J. Exp. Med*. **2000**, *192*, 193.

(39) Santoso, S.; Sachs, U. J.; Kroll, H.; Linder, M.; Ruf, A.; Preissner, K. T.; Chavakis, T. The junctional adhesion molecule 3 (JAM-3) on human platelets is a counterreceptor for the leukocyte integrin Mac-1. *J. Exp. Med*. **2000**, *196*, 679.

(40) Bell, G. I. A theoretical model for adhesion between cells mediated by multivalent ligands. *Cell Biophys.* **1979**, *1*, 133.

(41) Bell, G. I. Estimate of the sticking probability for cells in uniform shear flow with adhesion caused by specific bonds. *Cell Biophys.* **1981**, *3*, 289.

(42) Erickson, J.; Goldstein, B.; Holowka, D.; Baird, B. The effect of receptor density on the forward rate-constant for binding of ligands to cell-surface receptors. *Biophys. J.* **1987**, *52*, 657.

(43) Long, M.; Goldsmith, H. L.; Tees, D. F. J.; Zhu, C. Probabilistic modeling of shear-induced formation and breakage of doublets cross-linked by receptor−ligand bonds. *Biophys. J*. **1999**, *76*, 1112.

(44) Laurenzi, I. J.; Diamond, S. L. Kinetics of random aggregation-fragmentation processes with multiple components. *Phys. Rev. E*. **2003**. *67*, 051103.

(45) Yang, J.; Liu, Y.; Chang, J.; Wang, Y. N.; Bai, L.; Xu, Y. Y.; Xiang, H. W.; Li, Y. W.; Zhong, B. Detailed kinetics of Fischer−Tropsch synthesis on an industrial Fe−Mn catalyst. *Ind. Eng. Chem. Res*. **2003**, *42*, 5066.

(46) lliott, L.; Ingham, D. B.; Kyne, A. G.; Mera, N. S.; Pourkashanian, M.; Wilson, C. W. Genetic algorithms for optimisation of chemical kinetics reaction mechanisms. *Prog. Energy Combust. Sci.* **2004**, *30*, 297.

(47) Katare, S.; Caruthers, J. M.; Delgass, W. N.; Venkatasubramanian, V. An intelligent system for reaction kinetic modeling and catalyst design. *Ind. Eng. Chem. Res.* **2004**, *43*, 3484.

(48) Phimister, J. R.; Fraga, E. S.; Seider, W. D. Plantwide Controller Tuning Using a Multiobjective Genetic Algorithm. *Proceedings of the DYCOPS '98 Conference*, Corfu, Greece, 1998.

(49) Dietz, A.; Azzaro-Pantel, C.; Pibouleau, L.; Domenech, S. A framework for multiproduct batch plant design with environmental consideration: Application to protein production. *Ind. Eng. Chem. Res*. **2005**, *44*, 2191.

(50) McCarty, O. J. T.; Jadhav, S.; Burdick, M. M.; Bell, W. R.; Konstantopoulos, K. Fluid Shear Regulates the Kinetics and Molecular Mechanisms of Activation-Dependent Platelet Binding to Colon Carcinoma Cells. *Biophys. J.* **2002**, *83*, 836.

(51) Bell, D. N.; Spain, S.; Goldsmith, H. L. Adenosine diphosphate induced aggregation of human platelets in flow through tubes. I. Measurement of concentration and size of single platelets and aggregates. *Biophys. J*. **1989**, *56*, 817.

(52) Bell, D. N.; Spain, S.; Goldsmith, H. L. Adenosine diphosphate induced aggregation of human platelets in flow through tubes. II. Effect of shear rate, donor sex, and ADP concentration. *Biophys. J*. **1989**, *56*, 817.

(53) Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys*. **1976**, *22*, 403.

(54) Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*. **1977**, *81*, 2340.

(55) Gillespie, D. T. An exact method for numerically simulating the stochastic coalescence process in a cloud. *J. Atmos. Sci.* **1975**, *32*, 1977.

(56) Holland, J. H. *Adaptation in natural and artificial systems*; The University of Michigan Press: Ann Arbor, MI, 1975.

(57) Goldberg, D. E. *Genetic algorithms in search, optimization, and machine learning*; Addison-Wesley Publishing Co.: Reading, MA, 1989.

(58) English, D.; Andersen, B. R. Single-step separation of red blood cells. Granulocytes and mononuclear leukocytes on discontinuous density gradients of Ficoll-Hypaque. *J. Immunol. Methods* **1974**, *5*, 249.

*Received for review* October 17, 2005
*Revised manuscript received* February 15, 2006
*Accepted* February 23, 2006

IE051159T