Discovering Novel Cancer Therapies: A Computational Modeling and Search Approach

Arthur W. Mahoney, Brian G. Smith, Nicholas S. Flann, and Gregory J. Podgorski

Abstract— **Solid tumors must recruit new blood vessels for growth and maintenance. Discovering drugs that block this tumor-induced development of new blood vessels (angiogenesis) is an important approach in cancer treatment. However, the complexity of angiogenesis and the difficulty in implementing and evaluating rationally-designed treatments prevent the discovery of effective new therapies. This paper presents a massively parallel computational search-based approach for the discovery of novel potential cancer treatments using a high fidelity simulation of angiogenesis. Discovering new therapies is viewed as multi-objective combinatorial optimization over two competing objectives: minimizing the cost of developing the intervention while minimizing the oxygen provided to the cancer tumor by angiogenesis. Results show the effectiveness of the search process in finding interventions that are currently in use and more interestingly, discovering some new approaches that are counter intuitive yet effective.**

I. Introduction

CANCER is the leading cause of death for Americans between the ages of 40 and 74 [6]. With survival rates at approximately 50%, new, more effective therapies are urgently needed. However, given the complexity of cancer development, discovering interventions through purely laboratory-based experimental techniques is woefully slow. This paper introduces a massively parallel computational search-based technique that shows promise in rapidly discovering potential therapeutic interventions.

Many kinds of cancers involve solid tumors that must recruit new blood vessels from host tissues in order to grow and remain viable. These new blood vessels sprout from existing vessels towards the tumor and provide needed nutrition and oxygen to enable the tumor to grow rapidly. Blocking this process of tumor-induced angiogenesis (the formation of new blood vessels) has been an important approach in cancer treatment [2]. This paper introduces a computational approach to search for novel intervention strategies that disrupt angiogenesis induced by solid tumors. Using the work of Bauer et al. [1] as a springboard, a high fidelity simulation model of angiogenesis has been developed that is sufficiently abstract to be computational feasible yet sufficiently detailed to identify specific medically relevant intervention targets. Running this model simulates the early stages of angiogenesis, when new blood vessels grow towards the tumor, form loops, that allow blood to flow, and secrete oxygen to support the tumor.

The simulation system integrates: (a) a cellular Potts model (CPM) that captures mechanisms of endothelial cell growth, cell adhesion, extracellular matrix (ECM) fiber adhesion and degradation, and tip cell chemotaxis and haptotaxis [5]; (b) a continuous model of vascular endothelial growth factor (VegF) secretion from the tumor, diffusion through the stroma (host tissue), and endothelial cell uptake and activation; (c) a flow model that estimates blood flow through the irregular network of vessels that emerge during angiogenesis; and (d) a continuous model of oxygen secretion from vessel loops, diffusion through the stroma and uptake by the tumor. This model captures behaviors such as vessel branching, loop formation (anastomosis), progression and termination of tip movement, and activation and growth of new vessels. All these complex behaviors emerge from interactions among the simpler, biologically-relevant component mechanisms of the model.

The state of the art in angiogenesis blocking treatments is reviewed in [11]. Current drugs disrupt angiogenesis through interference with the VegF system, which diffuses from the tumor cells to the existing vascular, triggering the formation of new vessel growth and guiding the newly growing sprouts towards the tumor. These treatments either bind to VegF slowing its diffusion through the supporting tissue and interfering with its binding to receptors, or bind to the VegF receptor on the endothelial cells, preventing activation and growth. While demonstrating effectiveness at slowing the progression of some tumors, these drugs target only one component of angiogenesis. To identify more effective medical interventions, it is worthwhile to consider disrupting other components of angiogenesis and particularly, combinations of mechanisms. However, the complexity of angiogenesis presents many component mechanisms that could be disrupted in multiple ways, presenting a large combinatorial space of possible therapies that is infeasible to search using laboratory-based biological methods. By utilizing a high fidelity *simulation* of angiogenesis, this large combinatorial space can be efficiently searched and new potential therapies discovered. A further advantage of this computational approach is the discovery of non-intuitive, currently unavailable, but potentially very effective cancer treatments. A computational search-based approach can suggest which disruptions to try first to laboratory-based experimenters, thereby focusing research on finding ways of clinically implementing disruptions to those mechanisms that are most likely to be effective.

Arthor Mahoney, Brian Smith and Nicholas Flann are with Computer Science Department, Utah State University. Gregory Podgorski is in the Biology Department and Center for Integrated Bio Systems, Utah State University (email: nick.flann@usu.edu). This work was supported in part by a grant of high performance computing (HPC) resources from the Arctic Region Supercomputing Center at the University of Alaska Fairbanks as part of the Department of Defense High Performance Computing Modernization Program.

The rest of this paper is organized as follows: First the Monte Carlo-based search engine is described in Section II, followed by the Cellular Potts-based angiogenesis simulation system in Section II-B. Section III describes potentially effective cancer therapies found by the search engine. Finally, the findings of this work are discussed in Section IV.

II. Search Method

The search engine uses an extension of the COMPUCELL [5] tissue simulation system that takes as input a vector $\vec{P}$ of 23 parameters, runs a simulation of angiogenesis and outputs a measure of oxygen absorbed by a cancer tumor. Based on values in the literature and simulation studies of standard blood vessel morphologies, each of the 23 parameters has been assigned a normative value, which represents the untreated condition [1]. The normative parameter vector, $\vec{P_N}$, contains the normative values for each parameter and is given in Table II. The units of each parameter is given in [1]. Running the simulation with the normative parameter vector produces the expected progression of blood vessel growth in tissues adjacent to the tumor without treatment intervention.

An effective therapy disrupts the progression of blood vessels, thereby reducing the oxygen provided to the tumor and arresting its growth. Each potential therapy is represented as a vector of deviations $\Delta\vec{P}$ to the normative parameter values $\vec{P_N}$, where $\vec{P} = \vec{P_N} + \Delta\vec{P}$. The search for effective therapies can be viewed as a combinatorial optimization over the space of parameter deviations, with fitness of each potential therapy being evaluated using two competing minimization objective functions: (a) the estimated cost of medically implementing those changes in the patient, represented as $Cost(\Delta\vec{P})$, and (b) the estimated oxygen provided to the tumor by the simulated blood vessels formed when the model is run under the changed parameter values, represented as $O_2(\Delta\vec{P})$. The cost objective function used to calculate the cost fitness of each therapy is described in Section II-A.1, and the oxygen objective that calculates oxygen fitness is described in Section II-B.

Optimization occurs in the deviation space, X, defined as a subset of the $\mathbb{R}^{23}$ vector space, where each $\Delta\vec{P} \in X$ contains 23 parameter deviations (dimensions) bound by ranges. The deviation range for each parameter is limited to what changes are physically valid in the model and may be medically feasible to change now or in the future. Table II gives details of each parameter, which are summarized below, grouped by sub-system:

$$\vec{P} = (\underbrace{p_1 \ldots p_4}_{\text{growth}}, \underbrace{p_5 \ldots p_{13}}_{\text{adhesion}}, \underbrace{p_{14}, p_{15}}_{\text{forces}}, \underbrace{p_{16}, p_{17}, p_{18}}_{\text{ECM}}, \underbrace{p_{19} \ldots p_{23}}_{\text{VegF}}) \quad (1)$$

The subsystems are integrated into the cellular Potts model (CPM) and include endothelial cell growth, cell adhesion, the forces of chemotaxis and haptotaxis, ECM adhesion and degradation, and vasculoendothelial growth factor (VegF) secretion from the tumor, diffusion through the host tissue, and endothelial cell uptake and activation.

A. Improving Monte Carlo Search

Algorithm 1 The improving Monte Carlo search algorithm searches the combinatorial space of deviation vector sampled from X, where the number of active (non-zero) parameters in each sample is bound between 1 and k. $ImprovingMonteCarlo(k, n)$ returns a set, L^*, containing near Pareto optimal solution vectors found during the search.

```
ImprovingMonteCarlo(k, n)
L* ← ∅
S_k ← RandomSample(k, n)
for i = k downto 0 do
    for each ΔP_j ∈ S_i do
        ΔP_j.cost ← Cost(ΔP_j)
        ΔP_j.o2 ← O2(ΔP_j)
    end for
    S_i ← εParetoOptimal(S_i)
    L* ← εParetoOptimal(S_i ∪ L*)
    S_{i-1} ← ReduceResample(S_i, n)
end for
return L*
```

The search engine uses a naive, improving Monte Carlo search algorithm to discover novel and potentially counter-intuitive therapies in X that require the least medical cost for the largest average decrease in oxygen supplied to the tumor. For this study on practical medical therapies, solution vectors were limited to those with only 3, 2 or 1 parameters deviated from nominal The improving Monte Carlo search is an effective approach to this optimization problem because it maintains a set of promising $\Delta\vec{P}$ vectors while monotonically decreasing their estimated cost of clinical implementation.

Let S be a population of n parameter deviation vectors sampled from X. For $1 \leq j \leq n$, each deviation vector $\Delta\vec{P}_j \in S$ is defined as $\Delta\vec{P}_j = (\delta p_1, \ldots, \delta p_i, \ldots, \delta p_{23})$. If the deviation $\delta p_i \neq 0$ then the i^{th} parameter is said to be "active". $A(\Delta\vec{P})$ denotes the number of active parameters in the deviation vector $\Delta\vec{P}$.

The improving Monte Carlo search takes as input the maximum number of active (non-zero) parameter deviations k, then searches over the combinatorial space $\Delta\vec{P} \in X$, where $1 \leq A(\Delta\vec{P}) \leq k$. The algorithm maintains a population of promising deviation vectors S_i, where $1 \leq i \leq k$. The initial population S_k is seeded using $RandomSample(k, n)$ that uniformly samples the set of deviation vectors with exactly k active parameters. The search engine operates in a "top-down" fashion by systematically reducing the number of active parameters i in each population S_i until each vector has no active parameters remaining. During the search, the best solutions discovered so far are stored in the set L^*, which is returned by the search engine.

The search algorithm proceeds in an iterative fashion alternatively generating randomized potential solutions then selecting the most promising. The algorithm is described in

pseudo-code in Algorithm 1. At the i^{th} iteration, first the estimated cost and the amount of oxygen supplied to the tumor are evaluated and stored for each $\Delta\vec{P} \in S_i$ using $Cost()$ and $O_2()$. Selection is applied to the population S_i using $\varepsilon ParetoOptimal(S_i)$ that returns a subset of S_i that lie on or near the Pareto optimal frontier of the solution set S_i. The best-so-far solution set L^* is then updated with these near-optimal solutions from S_i. The next population of solutions S_{i-1} is generated by the $ReduceResample(S_i, n)$ function, which returns a new set of n solutions by simplifying and re-sampling each $\Delta\vec{P} \in S_i$. To simplify $\Delta\vec{P}$, one of the active parameters is randomly selected and its deviation set to 0, thus rendering that parameter inactive. Then the remaining active parameters are re-sampled by adding uniform noise to their values, while keeping the parameter deviations within the parameter range constraints. This set of perturbed and reduced deviation vectors then becomes S_{i-1} for the next iteration.

1) Evaluating the cost of each deviation vector: Given a deviation vector defined as $\Delta\vec{P} = (\delta p_1, \ldots, \delta p_i, \ldots, \delta p_{23})$. Let the range of parameter i be $R_i = H_i - L_i$ (where H_i and L_i are the allowable high and low parameter value respectively) then the cost of a deviation vector is estimated by summing the weighted individual errors:

$$Cost(\Delta\vec{P}) = \sum_{1}^{23} \omega_i * (\frac{\delta p_i}{R_i})^2 \tag{2}$$

$$\omega_i = \begin{cases} 2.0 & \text{if cost of } p_i \text{ is low} \\ 4.0 & \text{if cost of } p_i \text{ is medium} \\ 8.0 & \text{if cost of } p_i \text{ is high} \end{cases}$$

The relative costs of changing parameters, ω_i is determined by reviewing the literature on known medical intervention strategies [11] and considering the complexity of the biological sub-system to be manipulated. The medical cost is estimated based on whether specific treatments are currently available (low cost), have the potential to become available in the foreseeable future (medium cost), or are predicted to be difficult to develop with current technologies (high cost). For example, changing the diffusion properties (via the ligand) or binding efficiency of VegF is already clinically viable and is therefore given low cost, while ways to change the adhesive or elastic properties of specific cell types is poorly understood and therefore given a high cost.

2) Parallel evaluation of oxygen supplied to the tumor for each deviation vector: The COMPUCELL model simulates angiogenesis beginning with the secretion of VegF by the tumor. This induces the budding and growth of new blood vessels that sometimes form loops, enabling blood flow and oxygen delivery. The stochastic nature of the simulation generates different morphologies based on an initial random seed tied to each individual simulation. Due to the differing morphologies, the search engine evaluates 256 random initializations for a given $\Delta\vec{P}$. The metric $O_2(\Delta\vec{P})$ is the average O_2 absorbed by the tumor in each run.

For each simulation run, the domain is first initialized as illustrated in Figure 1 with initial blood vessel at the top, the tumor on the bottom and the stroma randomized with ECM fibers and stromal cells. The parameters used by COMPUCELL are set from $\vec{P} \leftarrow \vec{P_N} + \Delta\vec{P}$ and it is run for 1,500 Monte Carlo Steps (MCS). The O_2 score of this run is calculated by accumulating the O_2 absorbed by the tumor at each MCS. This number of MCS steps is sufficient to simulate angiogenesis within the space modeled between the existing blood vessel and the tumor.

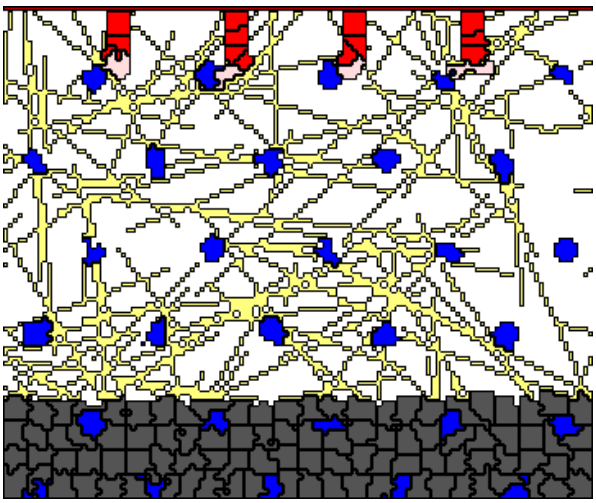


Fig. 1. The initial conditions of the angiogenesis simulation of the cell type domain $\sigma[,]$. The colors used are given in Table I

The COMPUCELL simulations performed while evaluating $O_2(\Delta\vec{P}_j)$ during each iteration are independent from each other. This implies that the improving Monte Carlo search is an "embarrassingly parallel algorithm" [3], [13]. Provided enough processors, the Compucell simulations required for individual $O_2()$ evaluations, as well as multiple $O_2()$ evaluations themselves, can be performed concurrently and executed in nearly the same amount of time as a single simulation. On a modern processor, one simulation can be completed in 5 to 7 minutes. Because the Monte Carlo search requires the execution of several thousand simulations, the algorithm cannot be executed in a reasonable amount of time without the use of a large set of processors. The improving Monte Carlo search used in this study was implemented using the common master/slave paradigm and was executed on a 2,312-processor AMD Opteron™ cluster maintained by the Arctic Region Supercomputing Center.

B. Angiogenesis Model

The CPM facilitates the combination of simpler sub-models into a larger model that is biologically accurate and computationally feasible. One advantage of the CPM is that multicellular pattern formation is an emergent property of local interactions between simple sub-cellular components. Another advantage is that all cellular mechanisms including adhesion, growth, death, haptotaxis and chemotaxis are easily and realistically implemented within the same architecture. The cell types used in this work interact with each other through the process of energy minimization to

TABLE I

TYPES USED IN SIMULATION

Cell Types		
Symbol	Description	Color in Figures
τ_m	Medium	White
τ_{ev}	Existing Vasculature	Dark red
τ_{tu}	Tumor	Black
τ_s	Endothelial Stalk	Red
τ_t	Endothelial Tip	Pink
τ_{st}	Stromal Cell	Blue
τ_e	ECM Fiber	Yellow

create emergent morphologies that simulate angiogenesis and anastomosis.

Initial Conditions: In our simulation, the angiogenesis domain is modeled as a two dimensional section illustrated in Figure 1 bordered by the existing vasculature on the top and the tumor, as a layer of 4 cells, on the bottom. The distance from the existing vasculature to the surface of the tumor wall is 165 μm. The area between the tumor and the vasculature is a mixture of medium, ECM fibers and stromal cells. Four sprouts are modeled with each new sprout vessel initially consisting of two cells. One cell is an endothelial tip cell, and the other is an endothelial-stalk cell which touches the existing vasculature.

TABLE II

THE PARAMETERS GOVERNING THE MODEL IN THIS WORK.

Param.	$\vec{P_N}$	Description
ϕ_t	1.0	τ_t growth rate
α_t	10^{-4}	τ_t VegF activation level
ϕ_s	1.0	τ_s growth rate
α_s	10^{-4}	τ_s VegF activation level
ε_t	2.0	Elasticity of τ_t
ε_s	2.0	Elasticity of τ_s
$J_{t,t}$	16	Adhesion between τ_t and τ_t
$J_{t,s}$	0	Adhesion between τ_t and τ_s
$J_{s,s}$	1	Adhesion between τ_s and τ_s
$J_{t,e}$	2	Adhesion between τ_t and τ_e
$J_{s,e}$	2	Adhesion between τ_s and τ_e
$J_{t,st}$	10	Adhesion between τ_t and τ_{st}
$J_{s,st}$	10	Adhesion between τ_s and τ_{st}
μ_c	35000	VegF chemotactic force on τ_t
μ_h	2000	ECM haptotactic force on τ_t
k_t	0.1	ECM degradation by τ_t
k_{t0}	0.05	ECM degradation adjacent to τ_t
η	0.25	ECM to fluid threshold
λ_V	10^{-5}	VegF degradation rate
D_V	0.045	VegF diffusion rate
S_V	0.035	VegF secretion rate
β_V	10^{-4}	VegF binding and absorbtion rate
γ	1.0	VegF binding efficiency

Cellular Potts Model: The Cellular Potts Model [4] (CPM) is used to simulate the tissues and integrate the mechanisms during angiogenesis. The CPM represents the tissue domain as a two dimensional array of lattice sites or pixels, each $\approx 2\mu m$ square. In all the simulations reported here, the tissue array is 100 by 120 pixels. Each cell within the tissue is represented as a set of contiguous pixels. Cell-cell contacts occur through adjacent pixels which belong to different cells. As the simulation is run, cells form new contacts and move with restrictions in size and in shape. All cell rearrangement is driven by a process of stochastic energy minimization.

The energy of a specific CPM tissue configuration is described by a Hamiltonian equation over the two dimensional domain s, which comprises a set of pixels z, each with a designated cell id σ_z:

$$H_s = \sum_{z \in s, z' \in n(z)} J_{\tau_{\sigma_z}, \tau_{\sigma_{z'}}} + \sum_{\sigma \in s} \varepsilon_{\tau_\sigma} (a_\sigma - A_\sigma)^2 \qquad (3)$$

The first term implements differential adhesion based on the type of each cell, denoted by τ_{σ_z}, by estimating the total surface energy between all contacting cells σ_z and $\sigma_{z'}$. This is done by summing $J_{\tau_{\sigma_z}, \tau_{\sigma_{z'}}}$ over all adjacent pixels z and z' where $\sigma_z \neq \sigma_{z'}$. There are seven types used in the angiogenesis simulation, five types of cells, the medium and the ECM listed in Table I. The $J_{\tau,\tau'}$ values are given in Table II.

The second term in Equation 3 implements an area constraint on cells where a_σ is the actual area of a cell σ and A_σ is σ's target area. The elasticity of a cell (the ease with which the cell can change its area) is controlled by the parameter ε_τ. Table II gives the elasticity values for the endothelial tip and stalk cells. The target area of a cell is dependent on the cell-type. For those cell types which do not grow (tumor τ_{tu} and stromal τ_{st}) the target area is fixed. For those cell types that grow (endothelial tip τ_t and stalk τ_s) the target area A_σ is incremented each model iteration. When a growing cell's current target area reaches twice its original target area, it is split into two daughter cells and the growing process repeats. In this work, all cell divisions are symmetric and the cleavage angle is chosen randomly. In angiogenesis, growth of endothelial cells is dependent on the presence of VegF, and is controlled by the threshold value α_τ given in Table II.

Low energy cell arrangements are determined by repeatedly copying the state of one pixel $\sigma[x, y]$ at x, y to an adjacent pixel at x', y' for pixels belonging to different cells. Let s be the configuration before the copy and s' be the configuration after the copy, then $\Delta H_{s,s'}$ is defined as $H_{s'} - H_s$. Then, if $\Delta H_{s,s'} < 0$, the state change is always accepted, and if $\Delta H_{s,s'} = 0$, the state change is accepted with probability 0.5. Otherwise the state change is accepted with probability $e^{-\frac{\Delta H_{s,s'}}{T}}$, where T is the temperature, representing the agitation of the cells [4].

In angiogenesis, the endothelial tip cell responds to haptotaxis [1] based on the local concentration of ECM in the tissue, and chemotaxis [7] based on the local concentration of VegF. To integrate these mechanisms into the CPM, two additional terms are added to the energy change function $\Delta H_{s,s'}$:

$$\Delta H_{s,s'} = (H_{s'} - H_s) + k_H(E[x', y'] - E[x, y]) + \mu_\sigma(V[x', y'] - V[x, y]) \qquad (4)$$

The level of VegF at pixel x, y is defined as $V[x, y]$ with the strength of the chemotactic force controlled by parameter μ_σ given in Table II; the level of ECM at pixel x, y is defined as $E[x, y]$ with the strength of the haptotactic force controlled by parameter k_H also given in Table II. How $V[x, y]$ and

$E[x, y]$ are initialized and calculated is described in following sections.

To run the simulation forward from the initial conditions, a pixel and its neighbor are randomly selected and a pixel state copy is considered as described above. One iteration of the simulation is termed an MCS (for Monte Carlo step) and comprises repeatedly choosing a random pixel n times, where n is 16 times the number of the pixels in the domain, which is 100×120 in the simulation model described here. The remaining subsections describe each sub-system in detail.

Secretion and Uptake of VegF: The recruitment of vessels by tumor cells is accomplished through the secretion of VegF, and the resulting VegF diffusion gradient. The endothelial tip cell exerts a chemotactic force in response to the VegF gradient, and thus induces vessel growth towards the tumor. The level of VegF at location x, y is defined as $V[x, y]$ and controlled by the following reaction-diffusion equation:

$$\frac{\partial V[x,y]}{\partial t} = D_V \nabla^2 V[x,y] - \lambda_2 V[x,y] + S_V(x,y) - B(x,y,V[x,y]) \tag{5}$$

There are four parameters that govern the VegF system. The coefficient of diffusion of VegF $D_V > 0$ is assumed to be homogeneous throughout the simulation domain. The degradation of VegF is also considered constant, at $\lambda_V > 0$. The two functions $S_V(x, y)$ and $B(x, y, V[x, y])$ describe the secretion and the absorption respectively of VegF in the domain. $S_V(x, y)$ describes the secretion of VegF from tumor cells positioned on the bottom border of the domain, while $B(x, y, V[x, y])$ describes the binding and uptake of VegF by the endothelial cells.

$$S_V(x,y) = \begin{cases} \phi & \text{if } \sigma[x,y].\tau = \tau_{tu} \\ 0 & \text{otherwise} \end{cases}$$

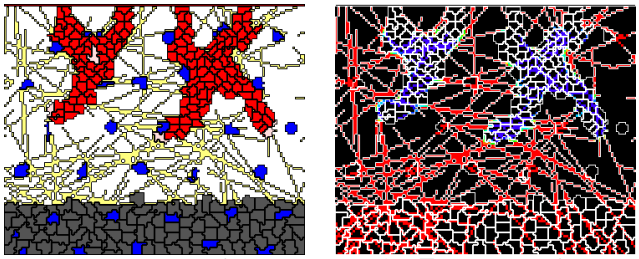
The secretion of VegF is at a constant rate ϕ from the tumor cells, positioned along the bottom side of the domain.

$$B(x,y,V) = \begin{cases} \beta & \text{if } \beta \leq V \text{ and } \sigma[x,y].\tau = \tau_t \text{ or } \tau_s \\ V & \text{if } 0 \leq V \leq \beta \text{ and } \sigma[x,y].\tau = \tau_t \text{ or } \tau_s \\ 0 & \text{otherwise} \end{cases}$$

The binding and uptake of V by the endothelial cells is defined in $B(x, y, V)$ and is limited to a maximum rate of $\beta > 0$ over the external surface of the endothelial cells. This is realistic since the capacity to bind and uptake VegF will saturate to a rate-limit.

Stroma: The endothelial tip cells exert haptotactic force in response to local concentrations of ECM in the stroma. This force can cause the vessel paths to be diverted towards other vessels resulting in anastomosis. The stroma is modeled using a combination of a discrete and a continuous model. The discrete model is over the $\sigma[,]$ domain and represents stromal cells τ_{st} and the ECM fibers denoted by τ_e. The continuous model contains the concentration level of ECM fiber proteins and is stored in the $E[,]$ domain.

Degradation of ECM fibers: The secretion of proteolytic enzymes by endothelial tip cells degrades the level of ECM protein when the endothelial cells are near or in contact with the ECM [12]. To model this process, the level of ECM at a point x, y is assumed to decay exponentially at those locations occupied by an endothelial tip cell or directly adjacent to that cell. This effect is illustrated in Figure 2(b). The rate of decay of ECM proteins is higher when the tip cell is over the fiber than next to it. The behavior is captured in the following equations:



(a) $\sigma[,]$ (MCS=600) (b) $E[,]$ (MCS=600)

Fig. 2. $\sigma[,]$ and $E[,]$ domains where the τ_t cells create vessel paths exhibiting anastomosis. The τ_t cells are pink, and the τ_s cells are red. The degradation of τ_e fibers is evident in the E lattice, where red indicates high concentrations of the ECM protein and blue indicates low concentrations. The vessel paths also exhibit haptotactic influence from the ECM.

$$E[x,y] = E[x,y] * \begin{cases} e^{k_i} & \text{if } \sigma[x,y].\tau = \tau_t \\ e^{k_n} & \text{if } \sigma[x',y'].\tau = \tau_t \text{ and } x', y' \text{ adjacent to } x, y \\ 1.0 & \text{otherwise} \end{cases}$$

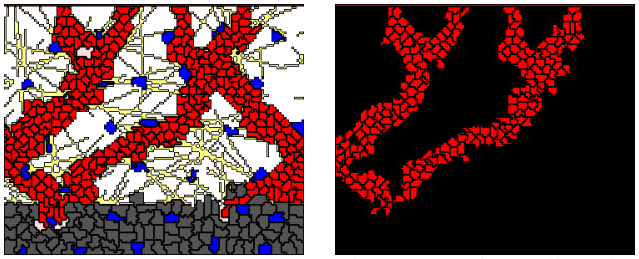
The physical effects of the ECM fibers are modeled within the CPM, where all the fibers are denoted $\sigma[x, y].\tau = \tau_e$ over the domain. τ_e is inelastic and initially arranged in random lines representing fiber bundles, see Figure 1. By integrating the ECM fibers into the CPM domain, the sprout's morphological development is affected by the growing sprouts pulling along the surfaces of the fiber bundles and being turned by the fiber bundle obstructions.

Initially, the ECM fibers are allocated a level of ECM protein $E[x, y]$ uniformly, then decayed in the presence of endothelial tip cells. To model the removal of the physical fibers once the ECM protein has been sufficiently degraded, the fibers become interstitial fluid (medium) when $E[x, y]$ drops below a fixed threshold η (given in Table II). This change in type is implemented by changing $\sigma[x, y].\tau$ from τ_e to τ_m.

Emergent Properties

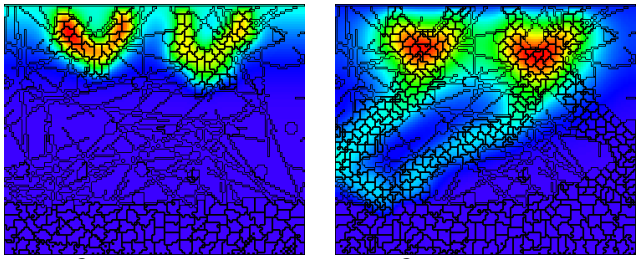
New blood vessel shape, structure and networks emerge from the complex interplay between the mechanisms of differential cell adhesion, VegF activated growth, chemotaxis, haptotaxis, ECM degradation by the tip cell and the secretion, diffusion and uptake of VegF.

Angiogenesis: The VegF chemical field is used as a growth signal to the stalk cells. When the VegF level is above the threshold α_s, then the cells grow. This method of growth, coupled with the uptake of VegF, the adhesion of cell types, and the chemotactic influence of the VegF gradient, produces the emergent property seen in tumor induced blood vessel growth. The tip cells climb the VegF gradient using the chemotactic force. The tip and stalk cells uptake a portion of the VegF that is contained in their lattice pixels. The stalk



(a) Cells (b) 0_2 secreting cells only

Fig. 3. (a) Typical angiogenesis morphologies form an inverted tree-like structure with some tip cells being dominated by other sprout cells. (b) Only endothelial cells in loops that have sufficient blood flow secrete O_2.



O_2 (MCS=300) O_2 (MCS=1430)

Fig. 4. O_2 values over the domain, where high concentrations of oxygen are red and low concentrations are blue. O_2 is secreted from vessels that form loops, diffuses through the stroma, and is absorbed by the tumor cells.

cells grow and divide in the presence of VegF, and adhere to the tip cells. This creates a vessel growth pattern that follows the movement pattern of the tip cells. In effect, the tip cells pull the stalk cells through the VegF gradient. The stalk cells grow directly behind the tip cells, but do not continue to grow as the cells get further from the tip cell. This is due to the low levels of VegF left in the growth path due to accumulated uptake. Since the ECM is only degraded by the tip cells, the stalk cells are often bounded by the remaining ECM fibers as they grow, forming regular width curving vessels under nominal conditions.

Anastomosis: The emergent property of anastomosis (loop formation) is critical to the usefulness of the simulations in evaluating the therapeutic effectiveness, since oxygen is only secreted when blood flows through vessel loops. Anastomosis arises from the inclusion of complex stroma in the model, which causes the separate sprouts to bend and collide. The path of tip cells is a combination of many factors including the haptotactic force provided by the ECM fibers, the chemotactic force of the VegF gradient, the adhesive force of stromal cells encountered and the density of ECM fiber bundles which provide obstructions.

Endothelial Tip Cell Growth Termination: When two sprouts join together it is common for one of the endothelial tip cells to continue growing, while the other becomes dormant. This behavior has been explicitly programmed into other models [9], [8], but in this system the behavior naturally emerges from interactions of sub-components of the model. When two sprouts join together the endothelial tip cell that is closer to the higher concentrations of the VegF gradient often continues, while the other stalls. It is common for the leading tip cell to continue growing towards the tumor through the chemotactic force, and thus consume the VegF available to both sprouts. The other endothelial tip cell most often terminates progression towards the tumor due to lack of a VegF gradient. The dominant endothelial-tip cell pattern recursively emerges to create an inverse tree-like vessel structure. This structure is consistent with biological observations and is illustrated in Figure 3(a).

Oxygen

The level of O_2 at location x, y is defined as $O_2[x, y]$ and controlled by the following reaction-diffusion equation:

$$\frac{\partial O_2[x,y]}{\partial t} = D_2 \nabla^2 O[x,y] - \lambda_2 O_2[x,y] + S(x,y) - U(x,y,O_2[x,y]) \quad (6)$$

There are four parameters that govern the O_2 system. The coefficient of diffusion $D_2 > 0$ is assumed to be homogeneous throughout the simulation domain. The degradation of O_2 is also considered constant, at $\lambda_2 > 0$. $S(x, y)$ describes the secretion of O_2 from endothelial cells contained within loop structures in the resulting morphology (see below for an explanation), while $U(x, y, O_2[x, y])$ describes the absorbtion of O_2 by the tumor cells.

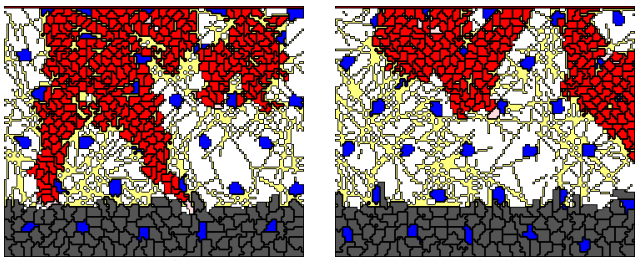
$$S(x,y) = \begin{cases} \alpha Q_{(i,j)} & \text{if } \sigma[x,y].\tau = \tau_t \text{ or } \tau_s \text{ and } \sigma[x,y] \text{ is secreting} \\ 0 & \text{otherwise} \end{cases}$$

The secretion of O_2 is at a constant rate α from the surface of all endothelial cells that are within loops formed by anastomosis. It is realistic to assume that only those sprouts that successfully link with other sprouts and form loops will be able to secrete O_2 because they will be capable of maturing and carrying a flow of blood through the formed vessel. To determine which endothelial cells secrete O_2 during the angiogenesis simulation, a pressure diffusion model is calculated at the cell level to identify loops as contiguous paths along contacting endothelial cells that join the original blood vessel at distinct locations. The flow (pressure difference) between cells is used to classify a cell as secreting, thereby enabling the correct implementation of $S(x, y)$ when the O_2 diffusion equation given in Equation **??** is solved.

$$U(x,y,O_2) = \begin{cases} \upsilon & \text{if } \upsilon \leq O_2 \text{ and } \sigma[x,y].\tau = \tau_{tu} \\ O_2 & \text{if } 0 \leq O_2 \leq \upsilon \text{ and } \sigma[x,y].\tau = \tau_{tu} \\ 0 & \text{otherwise} \end{cases}$$

The absorbtion of O_2 by the tumor is implemented in $U(x, y, O_2)$ and is limited to a maximum rate of $\upsilon > 0$ over the external surface of the tumor cells. This is realistic since the capacity to absorb oxygen will saturate to a rate-limit based on the maximum reaction rate on the membranes of the tumor cells.

Finally, the objective function $O_2(\Delta P)$ used in the search engine can be defined. First the parameters that control the simulation are set $\vec{P} \leftarrow \vec{P_N} + \Delta\vec{P}$, the domain initialized as in Figure 1 and the simulation is run 1,500 Monte Carlo Steps



(a) Solution c (b) Solution j

Fig. 5. Example morphologies from solutions c and j given in Figure 7

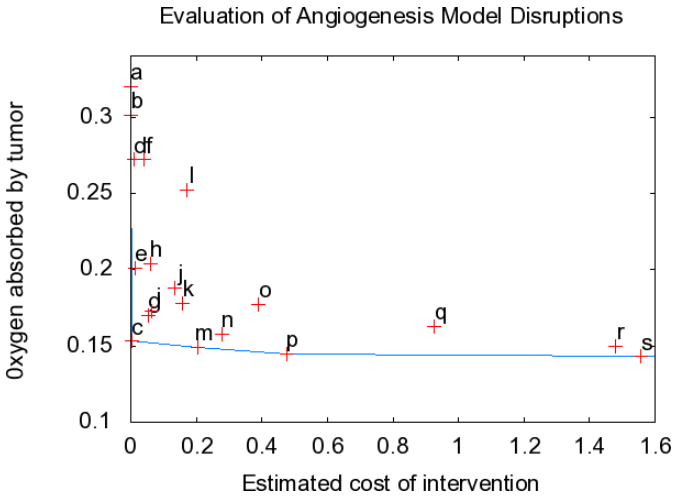


Fig. 6. Results: The Pareto optimal frontier and near optimal potential cancer therapies identified by the improving Monte Carlo search process. The best solutions provide the minimum 0_2 to the tumor (vertical axis) for the least estimated treatment cost (horizontal axis). Solution a represents the no treatment case at zero cost, but with the highest O_2. The point labels correspond to the solutions shown in Figure 7

(MCS). This number of steps allows us to model the early phase of angiogenesis in which tip cells sprout from existing vasculature to form new vessels that supply O_2 the micro tumor. Let $O_2[x, y]_t$ be the oxygen array at MCS iteration t, then the oxygen score for this run is:

$$O_2(\Delta) = \sum_{t=0}^{T} \sum_{x=1}^{X} \sum_{y=1}^{Y} U(x, y, O_2[x, y]_t)$$

Summary of Angiogenesis Simulation: A Cellular Potts model (CPM) is used to model tissue as a stochastic energy minimization process over an array of pixels. Cell signalling occurs through extracellular chemical gradients of VegF. VegF is secreted by the tumor cells and absorbed by the endothelial cells, thereby forming a diffusion gradient from the tumor to the vessel sprouts. The endothelial tip-cell responds to the VegF concentration by moving towards higher concentrations. The VegF also activates the endothelial sprout cells to grow. VegF is consumed by the endothelial cells, and eventually the concentration is low enough that the endothelial sprout cells become inactive. A vessel emerges that follows the chemotactic path of the endothelial tip cell.

The ECM fibers form obstructions to the endothelial cells. The tip cells adhere to the ECM fibers which causes a haptotactic force along contacting fibers. The tip cell degrades the ECM to interstitial fluid. The ECM diverts the vessel sprouts, causing them to grow towards other sprouts. As anastomosis occurs, one of the tip cells may become dormant while the other tip cell continues to progress towards higher concentrations of VegF. This competition between tip cells results in blood vessels exhibiting an inverted tree structure. This model only represents the early stages of angiogenesis, when tip cells arise from existing vessels and does not consider the formation of new tip cells along the newly formed vessels or subsequent tumor growth.

Oxygen begins to be secreted from the endothelial cells that form closed loops as the difference in blood pressure across the created loops increases above a threshold. The oxygen diffuses through the medium and reaches the tumor. The tumor cells consume the available oxygen, which is used as the principle metric for the search engine.

III. Results

The Monte Carlo improving search method was run on the AMD Opteron™ cluster at the Arctic Region Supercomputing Center. The oxygen score of each parameter vector was evaluated by running it on 256 processors in parallel, using different random initial ECM fiber and stomal cell configurations, and different random seeds to the stochastic CPM. Each angiogenesis simulation takes approximately five to seven minutes of cpu time. A total of 140 solutions were generated and evaluated for this preliminary study.

Out of these 140 solutions, the best 18 near-Pareto optimal solutions are given in Figure 6 as a Pareto-optimal frontier, and in detail in Figure 7. Reviewing the results, it is clear that the Monte Carlo improving search is capable of quickly discovering novel and potentially important cancer therapies. All the solutions presented reduce the oxygen to the tumor during early development, with the optimal solutions reducing oxygen by approximately 50%. This level of reduction is expected to significantly slow the growth of the tumor and has the potential to improve the survival rate of patients [10].

Reviewing the individual solutions in Figure 7 it is satisfying to see that the current clinically viable strategy of disrupting VegF [11] is discovered in solutions b and m, which both reduce the effectiveness of VegF uptake by the endothelial cells, and in solution g, which increases the degradation rate of VegF. Of more interest are the many counter-intuitive and novel solutions found. While it is possible to form a post hoc explanation of each solution's effectiveness, it is unlikely that these solutions would have been manually discovered. Consider a pair of effective solutions c and l, that work by modifying the threshold parameter that controls when the degraded ECM fibers become medium. Solution c works by reducing the threshold thereby increasing the resistance of the ECM fibers to proteolytic degradation. A typical morphology for solution c is illustrated in Figure 5(a) and shows clumped, stunted vessels intermixed with non-degraded ECM. Solution l works by increasing the threshold thereby increasing the susceptibility of the ECM to prote-

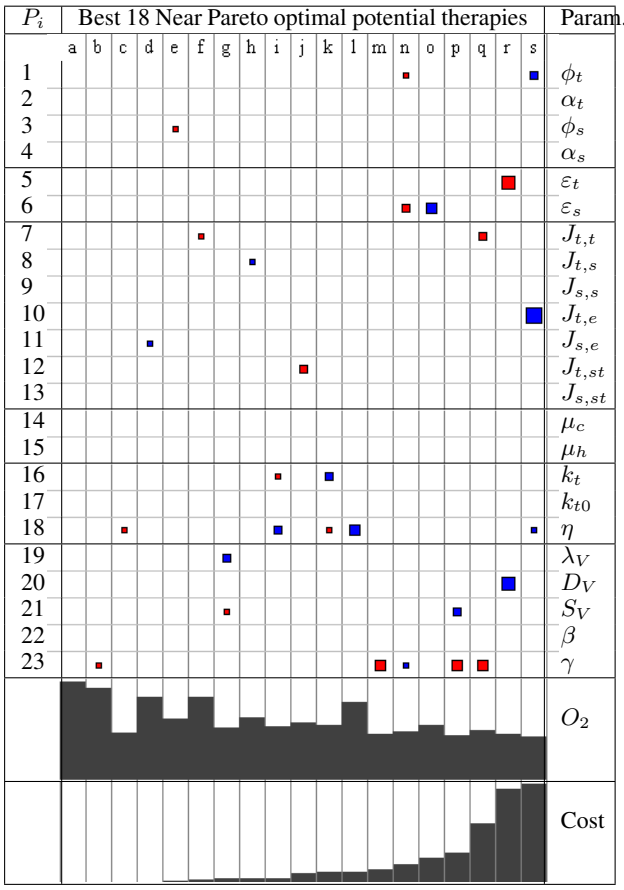


Fig. 7. Results: Near Pareto optimal model disruptions found by the search engine. Each column represents one solution with the letter designations (across the top) corresponding to those in Figure 6. The parameter disruptions for each solution is given in the column, with white meaning no change, red a reduction and blue an increase. Solution `a` represents the no treatment case, with no changes in the parameters. The size of the box represents the magnitude of the change. The average O_2 provided to the tumor and cost is given in the last two rows.

olytic degradation. Both changes in the threshold decrease the probability of the growing vessels forming loops. Or consider solution `j` that suggests reducing $J_{t,st}$, and thereby increasing the adhesive force between endothelial tip cells and stromal cells. A typical morphology for solution `j` is illustrated in Figure 5(b). This shows poorly formed thick vessels, with each of the 4 original tip cells adhering to a stromal cell that it encountered while moving up the VegF gradient. Solutions `c`, `j` and `l` are examples of the power of this method to discover novel potential cancer therapies that disrupt angiogenesis in unexpected ways.

IV. Conclusions

This report describes a massively parallel combinatorial search method for exploring the space of possible angiogenesis-blocking medical interventions for treating cancer. The method combines a high fidelity angiogenesis simulation system with a massively parallel improving Monte Carlo search process.

In a preliminary study of only 140 sample angiogenesis disruptions, many effective potential therapies were found. A few of the solutions were disruptions to the VegF system that are currently being deployed in clinical treatments [11]. Significantly, many of the solutions were novel and surprising, and suggest that this approach could help guide the development of powerful new cancer treatments.

References

[1] A. L. Bauer, T. L. Jackson, and Y. Jiang. A cell-based model exhibiting branching and anastomosis during tumor-induced angiogenesis. *Biophys. J.*, 92(9):3105–3121, May 2007.

[2] P. Carmeliet and R. K. Jain. Angiogenesis in cancer and other diseases. *Nature*, 407(6801):249–257, September 2000.

[3] K. Esselink, L. D. J. C. Loyens, and B. Smit. Parallel monte carlo simulations. *Phys. Rev. E*, 51(2):1560–1568, Feb 1995.

[4] F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended potts model. *Phys. Rev. Lett.*, 69:2013–2016, 1992.

[5] J. A. Izaguirre, R. Chaturvedi, C. Huang, T. Cickovski, J. Coffland, G. Thomas, G. Forgacs, M. Alber, G. Hentschel, S. A. Newman, and J. A. Glazier. Compucell, a multi-model framework for simulation of morphogenesis. *Bioinformatics*, 20(7):1129–1137, May 2004.

[6] A. Jemal, E. Ward, Y. Hao, and M. Thun. Trends in the leading causes of death in the united states, 1970-2002. *JAMA*, 294(10):1255–1259, 2005.

[7] J. Käfer, P. Hogeweg, and A. F. Marée. Moving forward moving backward: Directional sorting of chemotactic cells due to size and adhesion differences. *PLoS Computational Biology*, 2(6):e56+, June 2006.

[8] S. R. McDougall, A. R. Anderson, and M. A. Chaplain. Mathematical modelling of dynamic adaptive tumour-induced angiogenesis: clinical implications and therapeutic targeting strategies. *J Theor Biol*, 241(3):564–589, August 2006.

[9] S. R. McDougall, A. R. Anderson, M. A. Chaplain, and J. A. Sherratt. Mathematical modelling of flow through vascular networks: implications for tumour-induced angiogenesis and chemotherapy strategies. *Bull Math Biol*, 64(4):673–702, July 2002.

[10] S. Shinkaruk, M. Bayle, G. Laïn, and G. Déléris. Vascular endothelial cell growth factor (vegf), an emerging target for cancer chemotherapy. *Current medicinal chemistry. Anti-cancer agents*, 3(2):95–117, March 2003.

[11] F. Shojaei and N. Ferrara. Antiangiogenic therapy for cancer: An update. *The Cancer Journal*, 13(6):345–348, November 2007.

[12] S. Turner and J. A. Sherratt. Intercellular adhesion and cancer invasion: A discrete simulation using the extended potts model. *Journal of Theoretical Biology*, 216(1):85–100, May 2002.

[13] B. Wilkinson and A. Michael. *Parallel Programming: Techniques and Applications Using Networked Workstations and Parallel Computers*. Upper Saddle River, NJ: Prentice Hall, second edition, 2005.