Analysis of angiogenesis using *in vitro* experiments and stochastic growth models

Antti Niemistö,[1,2] Valerie Dunmire,[2] Olli Yli-Harja,[1] Wei Zhang,[2] and Ilya Shmulevich[3,2]
[1]*Institute of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland*
[2]*Department of Pathology, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Unit 85, Houston, Texas 77030, USA*
[3]*Institute for Systems Biology, 1441 N 34th St., Seattle, Washington 98103, USA*
(Received 29 November 2004; published 16 December 2005)

The global properties of vascular networks grown with an *in vitro* angiogenesis assay are compared quantitatively, using automated image analysis, with the global properties of networks obtained with discrete, stochastic growth models. The model classes that are investigated are invasion percolation and diffusion limited aggregation. By matching global properties to experimental data, one can infer which model classes and parameters are most reflective of angiogenesis in experimental cells. This sheds light on large-scale emergent properties of angiogenesis from a systems perspective. It is found that invasion percolation is better than diffusion limited aggregation at matching experimental data. We also present evidence that the distribution of the lengths of real tubule complexes follows a power law.

DOI: 10.1103/PhysRevE.72.062902 PACS number(s): 87.10.+e, 61.43.Hv

Angiogenesis, the formation of new capillary blood vessels, has become an important area of scientific research. It plays an important role in embryonic development, tissue repair and wound healing, tumor growth, and various vascular diseases [1–3]. Although several angiogenic and antiangiogenic factors have been discovered and some of the molecular mechanisms involved in angiogenesis have been identified [3,4], it is not fully known how these effects work together to determine the structural properties of vascular networks.

Angiogenesis involves interactions among angiogenic factors, inhibitors, and regulators at molecular, cellular, and tissue levels [5]. To understand such a complicated process, one must combine the experimental approach with mathematical modeling. The experimental approach enables us to investigate isolated factors and simple interactions, while through mathematical modeling we can understand the large-scale emergent properties of angiogenesis from a systems perspective.

Various mathematical models of angiogenesis have been developed [6–9]. Several of these models use partial differential equations to examine in space and time the distribution of variables such as endothelial cell density, capillary tip and branch density, and angiogenic factor concentration. The models have had some success in capturing the structure and morphology of capillary networks and even in examining strategies of antiangiogenesis.

In contrast to these continuous, deterministic models, discrete, stochastic models have also been considered [10–13]. Many of these models utilize experimental data to derive some of the model parameters, and have been able to generate realistic capillary network structures. On the other hand, some stochastic models do not model angiogenesis at a molecular, cellular, or tissue level, but rather attempt to capture the branching morphology of vascular network formation at a higher level of abstraction. These include invasion percolation (IP) as well as diffusion limited aggregation (DLA) models [14,15]. These models can be used to study the organizing principles of vascular networks in a well-defined framework. By matching global properties, such as distributions of the lengths of tubule complexes, to experimental data, using automated image analysis, we can infer which model classes and parameters are most reflective of angiogenesis. It is thus prudent to strive to extract higher level information or knowledge about vascular networks from currently available measurement data.

Just as the ensemble approach has been used to gain insight into the behavior of genetic regulatory networks [16], it can also be very useful, in conjunction with experimental measurements, for achieving a greater understanding of the dynamical process of angiogenesis. This is the approach taken by us here. The advantage of using such stochastic models as IP and DLA is that they enable us to study dominant coarse-scale properties of vascular networks without having to consider fine-scale quantitative details. Furthermore, due to a number of latent factors that may not be captured in a model, yet present in the environment of the vascular network and affecting the formation of local structures, stochastic models may be more appropriate than deterministic models. Another advantage is that these models are computationally much simpler than the models that are based on modeling various fine-scale details of neovascular growth.

In this work, we quantitatively compare the properties of real vascular networks with the properties of networks obtained with the IP and DLA models, by focusing on the distributions of the lengths of tubule complexes. Both models have several parameters that control their behavior. The simulations are made with different sets of parameters to obtain networks with different properties. The real vascular networks are obtained with the TCS Cellworks Angiokit (Buckingham, UK), a commercially available *in vitro* angiogenesis assay [17]. The quantitative properties of these networks are obtained with ANGIOQUANT, an image analysis tool that we have made freely available (see Ref. [26]) [18]. We also present evidence that the distribution of the lengths of real tubule complexes follows a power law.

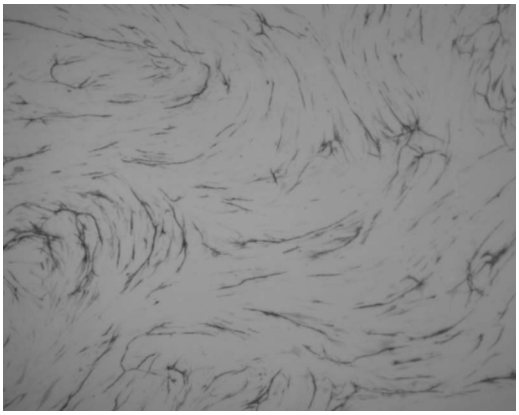

FIG. 1. A well treated with VEGF.

The TCS Cellworks Angiokit assay is based on coculturing endothelial cells with fibroblasts in wells arranged in a plate. Nine wells were treated with 2 ng/ml vascular endothelial growth factor (VEGF), a potent angiogenic stimulator. The effects of the growth factor can be seen with a light microscope, as the tubule complexes (connected vascular structures) are large. Each of the nine wells was imaged with a digital camera attached to a light microscope. Figure 1 shows an example of an obtained image. The total length of each tubule complex in each image was then obtained with ANGIOQUANT.

The invasion percolation growth model was introduced in 1983 by Wilkinson and Willemsen [19]. The model is based on a lattice of potential growth sites. It is initialized by assigning a random number for each potential growth site, and by occupying a single site at the center of the lattice. The initially occupied site can also be called a seed. Growth is thereafter possible only in the sites adjacent to an occupied site. In each time step, the site that has the lowest number assigned to it gets occupied. The IP model is thus a stochastic model that produces a single connected structure. The size of the structure can be controlled by the number of time steps. It is important to select the size of the lattice large enough with respect to the time steps, so as not to affect the shape of the structure that is grown.

We propose to use the IP model with multiple seeds that are positioned randomly as the model is initialized. This results in the growth of multiple structures, the sizes of which are not predefined by any of the parameters of the model. The structures can also merge to form a single structure. Thus the number of seeds is not necessarily the same as the number of structures that are present when the growth is halted.

We have also added an elongation factor $e \in [0,1)$ into the IP model. At each time step, the values assigned to the growth sites that are adjacent to an endpoint of the existing structure are multiplied by $1-e$. If the value of the elongation factor is greater than 0, growth in the endpoint neighborhoods is promoted, which results in more elongated structures. The value $e=0$ corresponds to no elongation.

The modified IP model was run with several different parameters sets. The size of the lattice was always 400×400 growth sites. The numbers of time steps N used were 1000, 2000, and 3000, the numbers of seeds n were 1, 5, 10, and 20, and the elongation factor e took the values 0, 0.2, 0.5, 0.8, and 0.9. The model was run with all combinations of these parameters, totalling 60 different sets of parameters. The simulation was repeated 334 times for each parameter set.

The diffusion limited aggregation growth model was introduced in 1981 by Witten and Sander [20]. The model is initialized by occupying a single site at the center of a lattice of potential growth sites. Growth particles are then released one by one from random sites that are at least a minimum distance from the center of the lattice. Each particle diffuses (walks randomly) until it hits a site adjacent to an occupied site, in which case it becomes part of the growing structure with a preset sticking probability p_s. If a particle touches the boundary of the lattice in its random walk, it is removed and another one is released.

The model can also include local amplification, whereby the particles are released in periods of N_p particle releases, and after each period, F additional particles are released at each growth site of that period. Thus after each period of N_p particle releases, local amplification increases the growth probability at the sites neighboring a site that was occupied within that period. Biologically, local amplification of growth factor levels can be achieved by autocrine release of growth factors [15]. This means that the growing structure itself is a source of growth factors. In the DLA model, elongated growth can be further promoted by using a higher sticking probability p_e in endpoint neighborhoods than in other potential growth sites.

We propose to use the DLA model with multiple seeds that are positioned randomly as the model is initialized. This results in the growth of multiple structures, which can merge to form a single structure. We also use separate launching and killing circles on a square lattice [21]. Each particle is released from a random site on the launching circle and it is removed if it hits the killing circle. Therefore the launching circle must have a smaller diameter than the killing circle.

The modified DLA model was run using the launching and killing circle sizes 400 and 420 growth sites, respectively. The number of released particles N was always 10 000. The numbers of seeds n were 1, 5, 10, 20, and 25. With the chosen lattice size and number of released particles, using a larger number of seeds is not possible, because that would cause the structures to outgrow the lattice. The sticking probabilities p_s and p_e both took the values 0.4, 0.8, and 1. The period length N_p took the values 10 and 100 and the local amplification factor F took the values 0, 1, and 2. The model was run with all combinations of these parameters, totalling 270 different sets of parameters. The simulation was repeated 208 times for each parameter set.

The tubule complex lengths obtained from the experimental angiogenesis data were accumulated to form a single data set that contains the length of 2529 tubule complexes. With both growth models, the obtained structures were skeletonized before their lengths were measured. This was done with the well-known thinning algorithm by Guo and Hall [22]. The structures were skeletonized, because skeletonization was also performed in the ANGIOQUANT software for the experimental angiogenesis data. Skeletonization ensures that

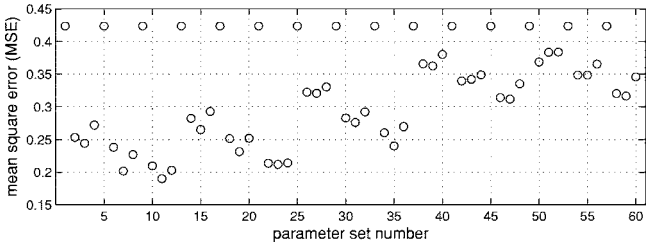

FIG. 2. The MSE values obtained with the IP model using different parameters.

the width of the structures is always one pixel (growth site), which makes measuring their lengths straightforward.

When multiple seeds are used, each simulation produces multiple connected structures. These structures are clearly not statistically independent. To ensure independence between the structures that we include in the study, we randomly select only one structure from each simulation. Thus with the IP model the data sets for each parameter set contain the lengths of 334 structures, and with the DLA model each data set contains the lengths of 208 structures.

Each data set was normalized by dividing each tubule complex length with the length of the longest tubule complex in the respective data set. This enables us to use the empirical (Kaplan-Meier) cumulative distribution function [23] to compare the tubule complex length distributions of the obtained simulated data sets with the tubule complex length distribution obtained from the experimental angiogenesis data.

In the following discussion, the word "model" refers to a model together with its parameters. The tubule length distributions are compared by calculating the mean square error (MSE) between each pair of empirical cumulative distribution functions. The procedure of picking the structures randomly from each of the simulated images along with the consecutive MSE calculation is repeated 1000 times. The model with the lowest mean over the 1000 MSE values is assumed to be the one providing the best fit to the experimental data. The mean of the MSE values alone does not tell us whether the difference between the distributions produced by different models is statistically significant. Therefore, we used the *t*-test to make this assessment.

The MSE values obtained with the IP model using different parameters are shown in Fig. 2. The MSE values above 0.4 are all equal and correspond to the IP model with one seed. We can conclude that if there is just one seed in the IP model, the fit to the experimental data is poor irrespective of

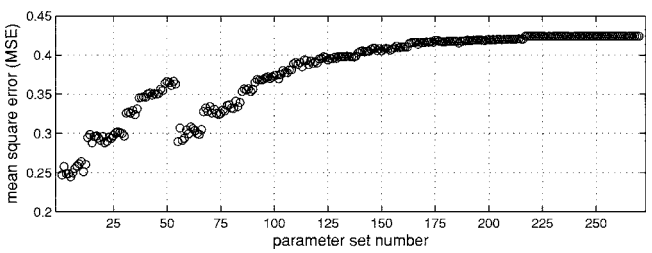

FIG. 3. The MSE values obtained with the DLA model using different parameters.

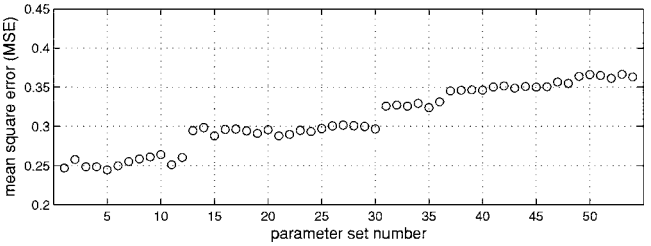

FIG. 4. The MSE values obtained with the DLA model using 25 seeds.

the other parameters. This is because in the case of one seed the size of the structure grown is always equal to N.

In Fig. 2 the indices 1–12, 13–24, 25–36, 37–48, and 49–60 correspond to the elongation factors $e=0$, $e=0.2$, $e=0.5$, e$=0.8$, and $e=0.9$, respectively. We immediately see that, other parameters being equal, the smaller the elongation factor, the better the fit. In particular, the best fit (the lowest MSE value) is obtained with $e=0$, a value which is equivalent to using the IP model without the elongation factor. This is an interesting result, since it implies that simpler models are preferred.

The indices 1–4, 5–8, and 9–12 correspond to the time step values $N=1000$, $N=2000$, and $N=3000$. Clearly, the best fit is obtained using 3000 time steps. The indices 9, 10, 11, and 12 correspond 1, 5, 10, and 20 seeds, respectively. The lowest MSE value is obtained with $n=10$. In summary, the best fit was obtained with the following set of parameters:

$$\{N = 3000,\ n = 10,\ e = 0\}.$$

The respective MSE value is 0.190.

Figure 2 also reveals that when n and e are kept constant and N is varied, the best fit is always obtained with $N=3000$. Similarly, when N and e are kept constant, the best fit is always obtained with $n=10$.

The MSE values obtained with the DLA model using different parameters are shown in Fig. 3. In the figure, the indices 1–54, 55–108, 109–162, 163–216, and 217–270 correspond to simulations with 25, 20, 10, 5, and 1 seeds, respectively. It can be clearly seen that the best fit (the lowest MSE value) is obtained with 25 seeds. Moreover, other parameters being equal, the higher the number of seeds, the better the fit. Also, as the number of seeds is increased, the effect of the other parameters on the fit increases. For example, with five or fewer seeds, the other parameters have almost no effect on the fit, whereas with 25 seeds the MSE ranges between 0.245 and 0.366 depending on the other parameters.

For the sake of improved visualization, the first 54 indices that correspond to simulations with 25 seeds are replotted in Fig. 4. The indices 1–18, 19–36, and 37–54 correspond to the local amplification factors $F=2$, $F=1$, and $F=0$. Clearly, $F=2$ always provides a better fit than the smaller values of the local amplification factor. This is in line with some of the results reported in earlier studies [15].

The indices 1–6, 7–12, and 13–18 correspond to the sticking probabilities $p_e=1$, $p_e=0.8$, and $p_e=0.4$, respectively.

Clearly, the fit with $p_e=1$ is somewhat better than with $p_e=0.8$ and much better than with $p_e=0.4$. Since the MSE values for indices 1–6 are all very close to each other, it can be concluded that the sticking probability p_s and the local amplification period N_p have only a small effect on the fit. Indices 1–2, 3–4, and 5–6 correspond to $p_s=1$, $p_s=0.8$, and $p_s=0.4$, respectively, and it can be seen that $p_s=0.4$ produces a better fit than $p_s=0.8$ or $p_s=1$. Indices 5 and 6 correspond to $N_p=10$ and $N_p=100$, respectively, and it can be seen that $N_p=10$ produces a slightly better fit than $N_p=100$. In summary, the best fit was obtained with the following set of parameters:

$$\{n=25,\ p_s=0.4,\ p_e=1,\ N_p=10,\ F=2\}.$$

The respective MSE value is 0.245.

We found that the IP model in its simplest form provides a better fit than the DLA model. With the parameters that provide the best fits, the MSE values are 0.190 and 0.245 for the IP and DLA models, respectively. The latter value is approximately 1.29 times as high as the former value, so the fit with the IP model is significantly better.

Finally, we have found that the tubule complex length distribution in the experimental angiogenesis data follows a power law. This may suggest that such a power law distribution provides optimal efficiency of blood transfer. However, more work needs to be carried out to determine why this distribution is optimal. One possible explanation for the power law behavior is that large tubule complexes have more potential branching points than small tubule complexes. Since tubule complexes can grow either by branching or at the endpoints of existing branches, large tubule complexes have more potential to grow than small tubule complexes. Thus the larger a tubule complex is, the faster it grows. This is akin to preferential attachment, which is one of the mechanisms known to give rise to power law degree distributions in networks [24]. The histogram of the tubule complex length data is plotted on a log-log scale in Fig. 5. Indeed, because the power law appears to be ubiquitous in many other contexts, such as metabolic networks of many organisms, social networks, the world wide web, the Internet, and others [25], similar underlying organizational principles may play a major role in angiogenesis.

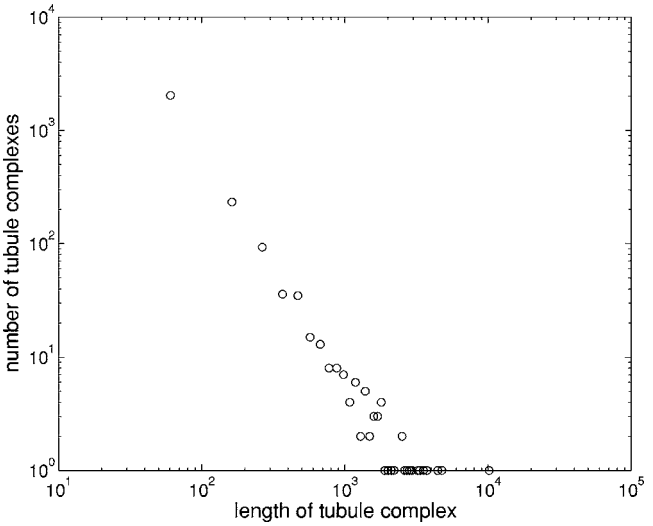

FIG. 5. The power law behavior of VEGF data.

We thank Dr. Ilya Gluhovsky for many useful discussions. The support of the Academy of Finland is gratefully acknowledged.

[1] F. Arnold and D. C. West, Pharmacol. Ther. **52**, 407 (1991).
[2] J. Folkman, Nat. Med. (N.Y.) **1**, 27 (1995).
[3] W. Risau, Nature (London) **386**, 671 (1997).
[4] N. Ferrara and K. Alitalo, Nat. Med. (N.Y.) **5**, 1359 (1999).
[5] J. Folkman, EXS **79**, 1 (1997).
[6] M. A. J. Chaplain, J. Neuro-Oncol. **50**, 37 (2000).
[7] *Vascular Morphogenesis: In Vivo, In Vitro, In Mente*, edited by C. D. Little, V. Mironov, and E. H. Sage (Birkhäuser, Boston, 1998).
[8] L. Olsen, J. A. Sherratt, P. K. Maini, and F. Arnold, IMA J. Math. Appl. Med. Biol. **14**, 261 (1997).
[9] M. E. Orme and M. A. J. Chaplain, IMA J. Math. Appl. Med. Biol. **14**, 189 (1997).
[10] F. Nekka, S. Kyriacos, C. Kerrigan, and L. Cartilier, Bull. Math. Biol. **58**, 409 (1996).
[11] M. J. Plank and B. D. Sleeman, IMA J. Math. Appl. Med. Biol. **20**, 135 (2003).
[12] C. L. Stokes and D. A. Lauffenburger, J. Theor. Biol. **152**, 377 (1991).
[13] S. Tong and F. Yuan, Microvasc. Res. **61**, 14 (2001).
[14] J. W. Baish and R. K. Jain, Cancer Res. **60**, 3683 (2000).
[15] Y. Gazit, D. A. Berk, M. Leunig, L. T. Baxter, and R. K. Jain, Phys. Rev. Lett. **75**, 2428 (1995).
[16] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York, 1993).
[17] E. T. Bishop, G. T. Bell, S. Bloor, I. J. Broom, N. F. K. Hendry, and D. N. Wheatley, Angiogenesis **3**, 335 (1999).
[18] A. Niemistö, V. Dunmire, O. Yli-Harja, W. Zhang, and I. Shmulevich, IEEE Trans. Med. Imaging **24**, 549 (2005).
[19] D. Wilkinson and J. F. Willemsen, J. Phys. A **16**, 3365 (1983).
[20] T. A. Witten and L. M. Sander, Phys. Rev. Lett. **47**, 1400 (1981).
[21] S. C. Ferreira, Jr., Eur. Phys. J. B **42**, 263 (2004).
[22] Z. Guo and R. W. Hall, Commun. ACM **32**, 359 (1989).
[23] D. R. Cox and D. Oakes, *Analysis of Survival Data* (Chapman and Hall, London, 1984).
[24] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[25] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[26] http://www.cs.tut.fi/sgn/csb/angioquant/