Mathematical models of diffusion-constrained polymerase chain reactions: basis of high-throughput nucleic acid assays and simple self-organizing systems

John Aach, George M. Church*

Department of Genetics and Lipper Center for Computational Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, New Research Building, Rm 238, Boston, MA 02115, USA

Abstract

DNA templates amplified by polymerase chain reaction in thin polyacrylamide gels form diffusion-constrained amplicons called "polonies" (polymerase colonies) that have been used to phase DNA haplotypes over long distances, to analyse RNA splice variants, and to assay other phenomena of biological interest. We present two sets of mathematical models, one for single polony growth (SPGM) and one for two polony interaction (TPIM), that will be used to optimize polony technology. The models provide detailed predictions of polony yield, concentration profiles, growth of isolated polonies, and the interaction of neighboring polonies. The TPIM explains an experimental observation that nearby polonies deform against each other rather than interpenetrate, an effect important for optimizing polony protocols. However, the TPIM also predicts that polonies may invade each other with a complex geometry when sufficiently close. Polonies are also of interest as simple abiotic systems that exhibit lifelike properties of self-organization, growth, and development, and the models may also apply to biological phenomena involving propagation through tethering and diffusion. Our polony modeling software is available at our web site http://arep.med.harvard.edu/polony_models/.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Polymerase chain reaction; Diffusion; Polony; Self-organizing system

1. Introduction

High-throughput assays for nucleic acid (NA) sequences have become core technologies in molecular biology. While large-scale electrophoretic sequencing (IHGS, 2001; Venter et al., 2001) and expression level analysis using DNA microarrays (DeRisi et al., 1997; Lockhart et al., 1996) are perhaps the most widely recognized uses, many other assay technologies have been developed (Brenner et al., 2000; Velculescu et al., 1995), and NA assays have proved applicable to many other kinds of biological problems, e.g. identification of DNA–protein binding sites (Bulyk et al., 2001; Lee et al., 2002) and protein–protein interactions (Ito et al., 2001; Uetz et al., 2000), and phenotyping of mutants (Badarinarayana et al., 2001; Winzeler et al., 1999). However, regardless of the technology platform and

application (excepting single-molecule methods in development, Braslavsky et al., 2003), at some point in their protocols all of these assays require that millions of copies of thousands of specific NA molecules be collected as pure populations in discrete spatial locations. This step is always resource-intensive. For example, sequencing projects and expression assays like SAGE (Velculescu et al., 1995) separate and amplify fragments to be sequenced by cloning, while cDNA microarrays require the separate amplification of individual cDNA probe sequences and robotic spotting to distinct locations on a glass slide. To increase the efficiency of this step, and to form the basis of a new generation of high-throughput NA assays, we and other laboratories have worked out methods for using the polymerase chain reaction (PCR) to amplify dilute populations of linear NA sequences in thin polyacrylamide gels on microscope slides (Mitra and Church, 1999). The amplicon of each individual template molecule remains localized in the gel, forming a discrete

*Corresponding author. Tel.: +617-432-1278; fax: +617-432-6513.
E-mail address: cvallaro@arep.med.harvard.edu (G.M. Church).

"polony" (polymerase colony), and this retention of template location information yields a second advantage of polonies over other high-throughput NA methods. Moreover, we are developing methods to sequence all polonies on a slide in parallel and in situ (Mitra et al., 2003b), further increasing the efficiency and throughput of NA sequencing and, through it, other NA-based applications. Even with only preliminary versions of these protocols, polony technology has been applied successfully in several contexts, including long range haplotype phasing (Mitra et al., 2003a), comprehensive RNA splice variant analysis (Zhu et al., 2003), and competitive phenotyping of mutants (Merritt et al., 2003).

Ultimately, the cost, throughput, and success of polony technology will be heavily influenced by the degree to which polony production itself may be optimized. Many interrelated factors and trade-offs have been identified: (i) Generally speaking, the more polonies that can be packed onto a slide, the higher throughput and lower cost for polony assays. Therefore, it is advantageous to generate polonies that are as small as possible within the limits of detection. (ii) Polony size can be controlled by the length of the DNA templates, the polymer density of the gel, and the number of PCR cycles (Mitra and Church, 1999). However, the ability to detect and sequence polonies will depend on the yield of the PCR reaction. Therefore trade-offs between polony yield and size need to be characterized, as well as their relation to PCR cycles. (iii) Because DNA templates are randomly dispersed in the gel, a fraction of polonies will overlap by chance. For most applications these are not usable (however, see Mitra et al., 2003a). As random overlap is governed by Poisson statistics, it might appear that the fraction of wasted polonies is an irreducible function of polony diameter and not amenable to improvement. However, it has been observed experimentally that polonies that are very close together sometimes appear to deform against each other instead of interpenetrate (see Fig. 1b), a phenomenon that has been called polony "exclusion" (Mitra et al., 2003b). If this phenomenon can be understood and optimized, wastage due to overlaps will indeed be reducible. (iv) In addition to DNA template length, gel density, and PCR cycles, there are many variables in polony production protocols that may be manipulated and which are actively being explored, e.g. primer densities (see below) and structures. The dependency of i–iii above on these additional variables also needs to be explored.

To gain insight into these factors and trade-offs, we have developed mathematical models of polony growth and interaction. Here we describe two initial models, a single polony growth model (SPGM) and a two polony interaction model (TPIM), and our first results characterizing polony growth profiles and structure, polony yield, and polony invasion and exclusion. In the
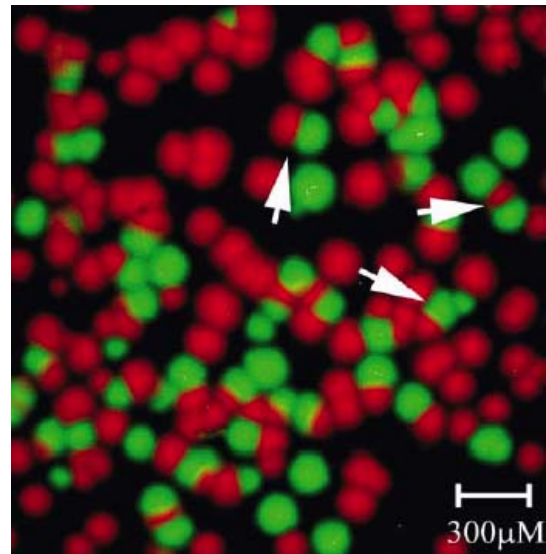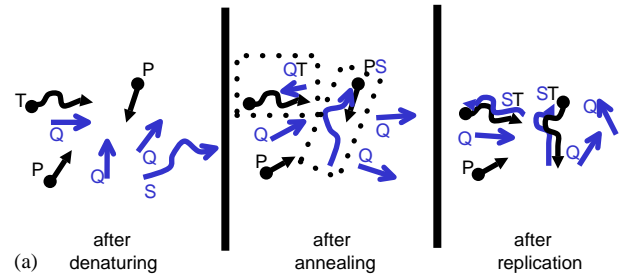
Fig. 1. Polonies and polony generation. (a) Figurative illustration of polony generation showing interactions of molecular species in a polony gel during a single PCR cycle. After a denaturing cycle (first frame) only single-stranded DNA species are present in a gel: P = tethered primer (short black arrows), Q = untethered (free) primer (short blue arrows), S = untethered (free) strand (long blue arrows), T = tethered strand (long black arrows) (see Table 1 for notation and nomenclature). Arrows indicate 5′-to-3′ orientation. Black circles indicate chemical anchoring of tethered species to gel matrix. Only untethered species (blue) are able to diffuse. During the annealing cycle, these untethered species continue to diffuse until they are captured by complementary tethered strands. Therefore, after annealing (second frame) a diffusing S strand has been captured by a tethered P primer, and a diffusing Q primer has been captured by a tethered T strand (indicated in boxes). After the subsequent replication phase (third frame), the captured P primer has extended to a T strand, and the captured Q primer has extended to an S strand, resulting in two ST molecules. These ST molecules will dissociate and the S strands will diffuse away in the following denaturing phase. (b) Image of actual polonies generated by protocol described in the text and illustrated in (a). Each colored spot is the product of the diffusion-constrained PCR amplification protocol generated by a single DNA template molecule. A mixture of two distinct templates was used in this reaction. Polonies were visualized by hybridizing primers specific to the two templates and extending with a single fluorescently labeled base which was different for the two templates, thereby labeling the two different kinds of polonies generated with different colors. Arrows indicate nearby polonies that have apparently deformed against each other rather than interpenetrate, illustrating the polony exclusion phenomenon (see text). Fig. 1b copied from Mitra et al. (2003b) by permission of publisher (see Acknowledgements).

Discussion section we will describe planned refinements to the models and how we plan to integrate them with experimental work. We also discuss how, in addition to their promise as a basis for NA assay technology, polonies may also be of theoretical interest as very simple self-organizing systems, and note that our models may also be applicable to some biological propagation phenomena. Our polony modeling software may be obtained at our web site http://arep.med.harvard.edu/polony_models/ along with supplementary information and full color versions of figures.

2. Models and methods

2.1. Polony generation and usage

Polony generation procedures have been described elsewhere (Mitra and Church, 1999). Briefly, for the contexts considered here, a pair of constant sequences that will be used as PCR primers is first identified. The NAs in the sample and the gel in which polonies will be formed are then separately prepared. The sample NAs are prepared simply by arranging for them to have common 5′ and 3′ end sequences that can be used to amplify them as a diverse library (Mitra and Church, 1999; Singer et al., 1997). Meanwhile, the polony gel is prepared in a way that will tether one of the PCR primers to the gel matrix. This is done by using a version of the primer that is modified so that its 5′ end contains a chemical group that incorporates into the gel matrix as it polymerizes (Rehman et al., 1999). A quantity of these modified primers is added to an acrylamide monomer solution and a thin gel allowed to polymerize on a microscope slide. To make the polonies, all other

ingredients needed for the PCR reaction are supplied to the gel, including the prepared sample NAs, the second (unmodified) PCR primer, and the free nucleotides and required PCR enzymes, and the PCR is performed by normal thermal cycling. The steps of a polony PCR reaction cycle are illustrated in Fig. 1a. To use generated polonies in assays that employ parallel in situ sequencing (Mitra et al., 2003b) or probing (Zhu et al., 2003), the polonies are denatured to dissociate the free and tethered strands and the free strands are diffused out of the gel, leaving only tethered single-stranded DNA behind.

2.2. Model elements

The SPGM and TPIM describe the progress of the PCR reaction under the diffusion-limiting conditions of the polony gel. Diffusion is limited due to both the low diffusion coefficients associated with the gel and the presence of tethered molecules which cannot diffuse at all. The species of molecules considered by the models and the notation used to denote them are given in Table 1. The SPGM considers seven species of molecules, and the TPIM considers an additional five species. A depiction of a PCR cycle is given in Fig. 1a. A list of the chemical reactions covered by each model and the corresponding mathematical equations is given in Fig. 2. Each PCR cycle consists of a succession of three phases, which have distinct mathematical representations. During the denaturing phase, the gel is heated so that double-stranded molecules denature and any resulting single stranded molecules which are not tethered to the gel, and therefore capable of diffusion (Table 1), diffuse freely through the gel. In these initial models, the denaturing phase is simplified by assuming

Table 1
Molecular species considered by polony growth and interaction models and notation for them

Symbol	Model		Diffusible	Description
	SPGM	TPIM		
P	Y	Y		Tethered PCR primer
Q	Y	Y	Y	Untethered (free) PCR primer
S	Y	Y	Y	Untethered strand (sample NA PCR template)
T	Y	Y		Tethered strand (reverse complement of S)
PS	Y	Y		Hybrid of P and S
QT	Y	Y		Hybrid of Q and T
ST	Y	Y		Hybrid of S and T
U		Y	Y	Untethered strand (second sample NA PCR template)
V		Y		Tethered strand (reverse complement of V)
PU		Y		Hybrid of P and U
QV		Y		Hybrid of Q and V
UV		Y		Hybrid of U and V

Symbols may denote either the molecular species themselves or the concentrations of them, depending on context. Model: Y indicates that species is recognized by the single polony growth model (SPGM) or two polony interaction model (TPIM). Diffusible: Y = molecule is not tethered to the gel and so may diffuse through the polony gel. Tethered molecules include the tethered primer (P), any species generated from P by extension (T, V), and any hybrids that include one of these tethered molecules (PS, PU, QT, QV, ST, UV).

SINGLE POLONY GROWTH MODEL
(SPGM)

DENATURING	ANNEALING	REPLICATION
	Reactions	

$ST \rightarrow S + T$

$$P + S \xrightarrow{k_{PS}} PS$$

$$Q + T \xrightarrow{k_{QT}} QT$$

$$S + T \xrightarrow{k_{ST}} ST$$

$PS \rightarrow ST$
$QT \rightarrow ST$

Equations

First moment:

At all points in space

$S \leftarrow S + ST$
$T \leftarrow T + ST$
$ST \leftarrow 0$

Rest of phase:

$$\frac{\partial Q}{\partial t} = D_Q \nabla^2 Q$$

$$\frac{\partial S}{\partial t} = D_S \nabla^2 S$$

(a)

$$\frac{\partial P}{\partial t} = -k_{PS} P \cdot S \quad \text{(A1)}$$

$$\frac{\partial Q}{\partial t} = -k_{QT} Q \cdot T + D_Q \nabla^2 Q \quad \text{(A2)}$$

$$\frac{\partial S}{\partial t} = -k_{PS} S \cdot S - k_{ST} S \cdot T + D_S \nabla^2 S$$

$$\frac{\partial T}{\partial t} = -k_{QT} Q \cdot T - k_{ST} S \cdot T$$

$$\frac{\partial PS}{\partial t} = k_{PS} P \cdot S$$

$$\frac{\partial QT}{\partial t} = k_{QT} Q \cdot T$$

$$\frac{\partial ST}{\partial t} = k_{ST} S \cdot T$$

At all points in space

$ST \leftarrow PS + QT + ST$
$QT \leftarrow 0$
$PS \leftarrow 0$

TWO POLONY INTERACTION MODEL
(TPIM)

DENATURING	ANNEALING	REPLICATION
	Reactions	

Add reaction

$UV \rightarrow U + V$

Add reactions

$$P + U \xrightarrow{k_{PU}} PU$$

$$Q + V \xrightarrow{k_{QV}} QV$$

$$U + V \xrightarrow{k_{UV}} UV$$

Add reactions

$PU \rightarrow UV$
$QV \rightarrow UV$

Equations

Add equations

First moment:

At all points in space

$U \leftarrow U + UV$
$V \leftarrow V + UV$
$UV \leftarrow 0$

Rest of phase:

$$\frac{\partial U}{\partial t} = D_U \nabla^2 U$$

(b)

Replace equations (A1) and (A2) with

$$\frac{\partial P}{\partial t} = -k_{PS} P \cdot S - k_{PU} P \cdot U$$

$$\frac{\partial Q}{\partial t} = -k_{QT} Q \cdot T - k_{QV} Q \cdot V + D_Q \nabla^2 Q$$

Add equations

$$\frac{\partial U}{\partial t} = -k_{PU} P \cdot U - k_{UV} U \cdot V + D_U \nabla^2 U$$

$$\frac{\partial V}{\partial t} = -k_{QV} Q \cdot V - k_{UV} U \cdot V$$

$$\frac{\partial PU}{\partial t} = k_{PU} P \cdot U$$

$$\frac{\partial QV}{\partial t} = k_{QV} Q \cdot V$$

$$\frac{\partial UV}{\partial t} = k_{UV} U \cdot V$$

Add equations

At all points in space

$UV \leftarrow PU + QV + UV$
$QV \leftarrow 0$
$PU \leftarrow 0$

Fig. 2. Single polony growth and two polony interaction models. Each table describes the chemical reactions and the mathematical equations used to describe a PCR cycle. Each cycle comprises a series of three phases: denaturing, annealing, and replication. (a) Single polony growth model. The complete set of reactions and equations is described. (b) Two polony interaction model, described as a set of additions and changes to the single polony growth model. Note that diffusion equation boundary conditions are always as follows: S and U are 0 at the boundary of solution space. Q is always eQ (see Table 2) at the boundary of solution space (see text).

that all double-stranded molecules completely dissociate in the first instant (see Discussion) and do not renature during the phase. In the annealing phase, the temperature of the gel is lowered so that complementary single stranded molecules can now associate. This process is characterizable as diffusion-to-capture (Berg, 1993, Chapter 3) and is modeled by reaction–diffusion equations that combine rate equations for renaturation with diffusion. The annealing phase will generate a population of double-stranded molecules containing a tethered or free primer hybridized to a template NA strand or its reverse complement. In the replication phase, the primers in these hybrids will be extended against these strands. As with the denaturing phase, this process is simplified in these initial models, here by assuming that extension is complete and immediate (see Discussion).

2.3. Model variants, initial conditions, parameters

Although polony generation takes place in the three-dimensional (3D) reaction environment provided by the polony gel, we developed variants of both the SPGM and TPIM that simulate polony generation in one (1D) and two (2D) spatial dimensions as well as 3D. These lower dimensional models are not only useful analytically (see Results), but can sometimes be good approximations to actual polony generation conditions. For instance, while a typical polony gel is $\approx 10 \,\mu m$ thick, polonies may be generated with a diameter of $> 100 \,\mu m$ (Fig. 1b); in such a case a 2D model may simulate conditions better than a 3D model. The 1D model has potential application to microfluidic environments in which PCR takes place in a linear channel (Kopp et al., 1998).

All polony models assume that polony growth begins with single template molecules. For SPGM, a single molecule of ST is placed at the solution space origin, while for TPIM, a single molecule of ST and a single molecule of UV are placed at equal distances in opposite directions from the origin. These model initial conditions represent the end result of several events that take place early in actual polony reactions: an individual molecule of S placed in the gel will hybridize with a P primer to form a PS molecule, and the P in this PS will subsequently extend to a T to create an ST molecule (similarly for U in TPIM). These events involve the consumption of individual P molecules that are converted to T and U. Therefore, polony model initialization also assumes that single P molecules have been removed from the locations of any initial ST and UV molecules. In actual polony gels, it is possible that two different template molecules S and U do not become converted to ST and UV at the same time, and may not even convert during the same PCR cycle. However, our initial TPIM assumes the synchronous initiation and development of both polonies.

The parameters required by the models are listed in Table 2. These include the diffusion coefficients of free primers and strands, initial concentrations of tethered and free primers, hybridization rate constants, the number of PCR cycles, and the lengths of each PCR denaturing and annealing phase. For the initial modeling results reported here, we developed settings for these parameters ('standard' parameters) based on values used in or calculable from actual polony experiments, or, where information was lacking, from experimental literature on DNA physical chemistry. These values also given in Table 2, and the explanations for them are given in Appendix A. For TPIMs, it is possible to specify different diffusion coefficients and hybridization rate constants for the S and U template strands; however, all simulations reported here assume that S and U share parameter values. Additional parameters beyond the standard parameters given in Table 2 include

Table 2
Standard model parameters

Parameter	Value	Description
3D simulations		
D_Q	20 μm^2/s	Diffusion coefficient of Q
D_S	0.035 μm^2/s	Diffusion coefficient of S
D_U	0.035 μm^2/s	Diffusion coefficient of U
k_{PS}	0.008 μm^3/molecule-s	Rate constant of PS hybridization reaction
k_{QT}	0.008 μm^3/molecule-s	Rate constant of QT hybridization reaction
k_{ST}	0.026 μm^3/molecule-s	Rate constant of ST hybridization reaction
k_{PU}	0.008 μm^3/molecule-s	Rate constant of PU hybridization reaction
k_{QV}	0.008 μm^3/molecule-s	Rate constant of QV hybridization reaction
k_{UV}	0.026 μm^3/molecule-s	Rate constant of UV hybridization reaction
T_d	30 s	Denaturing time
T_a	45 s	Annealing time
cyc	40	Number of PCR cycles
iP	300 molecules/μm^3	Initial concentration of P (0.5 μM)
eQ	300 molecules/μm^3	Initial concentration of Q (0.5 μM)
2D simulation: 1 μm thick gel		
k_{PS}	0.041 μm^2/molecule-s	Rate constant of PS hybridization reaction
k_{QT}	0.041 μm^2/molecule-s	Rate constant of QT hybridization reaction
k_{ST}	0.087 μm^2/molecule-s	Rate constant of ST hybridization reaction
k_{PU}	0.041 μm^2/molecule-s	Rate constant of PU hybridization reaction
k_{QV}	0.041 μm^2/molecule-s	Rate constant of QV hybridization reaction
k_{UV}	0.087 μm^2/molecule-s	Rate constant of UV hybridization reaction
iP	44.8 molecules/μm^2	Initial concentration of P (0.5 μM)
eQ	44.8 molecules/μm^2	Initial concentration of Q (0.5 μM)
2D simulation: 10 μm thick gel		
k_{PS}	0.0041 μm^2/molecule-s	Rate constant of PS hybridization reaction
k_{QT}	0.0041 μm^2/molecule-s	Rate constant of QT hybridization reaction
k_{ST}	0.0087 μm^2/molecule-s	Rate constant of ST hybridization reaction
k_{PU}	0.0041 μm^2/molecule-s	Rate constant of PU hybridization reaction
k_{QV}	0.0041 μm^2/molecule-s	Rate constant of QV hybridization reaction
k_{UV}	0.0087 μm^2/molecule-s	Rate constant of UV hybridization reaction
iP	448 molecules/μm^2	Initial concentration of P (0.5 μM)
eQ	448 molecules/μm^2	Initial concentration of Q (0.5 μM)
1D simulation: linear gel with 1 μm^2 cross-section		
k_{PS}	0.2 μm/molecule-s	Rate constant of PS hybridization reaction
k_{QT}	0.2 μm/molecule-s	Rate constant of QT hybridization reaction
k_{ST}	0.3 μm/molecule-s	Rate constant of ST hybridization reaction
k_{PU}	0.2 μm/molecule-s	Rate constant of PU hybridization reaction
k_{QV}	0.2 μm/molecule-s	Rate constant of QV hybridization reaction
k_{UV}	0.3 μm^2/molecule-s	Rate constant of UV hybridization reaction
iP	6.7 molecules/μm^2	Initial concentration of P (0.5 μM)
eQ	6.7 molecules/μm^2	Initial concentration of Q (0.5 μM)

Parameters used by single polony growth model or two polony interaction models, notation for them, and values used for most simulations ("standard" parameter values). Complete standard parameter values are given for 3D polony simulations, while for lower dimensional simulations, only those values different from the 3D values are listed. Derivations of the standard 3D values are given in Appendix A. Values for lower dimensional simulations are 3D values normalized for the lower dimension (see Methods).

the solution space dimensions, mesh sizes, time increments, and output window sizes used by the algorithms implemented by the models. These are described in our supplemental material at http://arep.med.harvard.edu/polony_models/.

For each polony model, the variables specifically represented are concentrations of molecular species at each point in the space assumed by the model. Because polony modeling starts with single molecule seeds, concentrations are more intuitively expressed in terms of molecules/volume than molarities, and yields are correspondingly expressed in terms of numbers of molecules. Rate constants are expressed accordingly in units of volume/molecule-s. The unit of volume may be μm, μm^2, or μm^3, depending on the model's dimensionality. As standard parameter values derived from actual experiments are intrinsically 3D, lower dimensional simulations are based on standard 3D parameters that are normalized to the lower dimension. For instance, the standard initial concentration of P is 300 molecules/μm^3 ($\approx 0.5\,\mu M$, similar to concentrations actually used in experiments); in 2D models, the corresponding value is 44.8 molecules/$\mu m^2 = 300^{2/3}$ molecules/$(\mu m^3)^{2/3}$. This setting will allow a gel that is 1 μm thick to be simulated by a 2D model. To model a gel with a typical 10 μm thickness in a 2D fashion requires an initial P concentration of 448 molecules/μm^2. Again, rate constants must be similarly adjusted (Table 2).

2.4. Algorithms and software

Models were implemented as C programs. For all models, the diffusion partial differential equations (PDEs) employed in the denaturing and annealing phases are solved by implicit methods (Ames, 1992); however, the finite difference equations that represent these PDEs depend closely on the coordinate system used by the model, and models differ substantially in their coordinate systems. The PDEs in denaturing phases are solved as pure diffusion equations. The reaction–diffusion equations that represent the PCR annealing phase are solved by alternating between steps of pure diffusion and pure annealing (operator splitting, Press et al., 1996, Chapter 19). Diffusion equations involving multiple spatial coordinates (2D SPGM, 2D TPIM, 3D TPIM) are solved either by operator splitting or by the alternate directions method (Ames, 1992, Eq. (5-77)). In all diffusion equations, free strands S and U are assumed to have zero concentrations at the solution space boundary, while free primer Q is assumed to be maintained at its initial value concentration there; i.e. Q is assumed to have source at the boundary that represents the reservoir of Q in the gel beyond solution space that may diffuse in to replenish the Q used up in the polony reaction. To improve performance of multiple spatial coordinate models, while supporting

adequate spatial resolution of polony features whose scale may vary ~ 100-fold over the course of a 40 PCR cycle simulation, dynamic rescaling logic was developed that automatically increases spatial variable mesh and solution space sizes as the polony grows. Polar coordinates are used in 3D polony models. The 3D SPGM assumes spherical symmetry and employs a single radial coordinate, while the 3D TPIM employs cylindrical coordinates comprising a 2D radial coordinate and a Z-axis Cartesian coordinate. Difference equations for polar coordinates are implicit versions of (Crank, 1956, Eqs. (10.36–7) and (10.41–2)). Additional details on the algorithms, dynamic rescaling, program usage, and algorithm performance and accuracy, are available on our web site http://arep.med.harvard.edu/polony_models/, along with the software itself.

2.5. Model output and analysis

Each program implementing a model may report, subject to user specification, the concentrations of each molecular species at every point in the model's solution space after any phase of any PCR cycle. The total yield of each molecule, computed by integrating the concentrations over solution space, may also be reported after any phase of any PCR cycle. For analysis of yields and the overall spatial distribution of free and tethered strands, it is convenient to look at values after denaturing cycles when these species exist exclusively as single strands. As polony assays use only the tethered strands (see above), the T and V yields are of particular interest. For analysis of growth, it is convenient to look at hybrids PS and QT (and, for TPIMs, PU and QV) after annealing cycles, as the distribution of these species determines where new free and tethered strands will be generated during the subsequent replication step. Output concentration data sets may be large, especially for models that require multiple spatial coordinates. By reporting average values over coordinate intervals, the output windowing option can make these data sets more manageable by reducing the number of reported variable values. However, by dint of this averaging, assessments of where and when variables may reach threshold values of interest may yield slightly different results in simulations run with identical model parameters, if the output window sizes are different.

3. Results

We analysed SPGM simulations to characterize the main features of polony growth, and TPIM simulations to characterize polony exclusion and invasion.

3.1. Main features of single polony growth

3.1.1. Single polony morphology and development

Key features of single polony morphology and development can be seen in Fig. 3, which presents selected results from a 3D SPGM simulation using standard parameters (Table 2). We distinguish two phases of polony growth and describe the transition between them as polony "maturation." The first phase of growth is characterized by rapid growth of concentrations from initial single molecule levels throughout a steadily enlarging spherical region around the origin. Polony growth in all phases is fueled by conversion of tethered primer P to tethered strand T, where P is initially present in a uniform concentration and is never replenished. The second phase of growth begins when P becomes completely consumed in the interior of the polony. Following this, polony growth can only continue by the depletion of P and corresponding accretion of T in spherical shells around the interior of ever larger radius. This transition takes place in cycle 22 in Fig. 3, at which point the concentration P becomes 0 and T becomes 300 molecules/μm^3 at the origin, a plateau level that can never increase. After cycle 22, the radius of the spherical region that reaches this plateau expands with each cycle. The graphs for P and T in Fig. 3 are arithmetic inverses that sum to 300 molecules/μm^3 at every radius and every cycle. As concentrations for P, Q, S, and T are presented after PCR

denaturing phases, the corresponding concentrations of all hybrid molecules are all zero.

Unlike P, Q does not become completely consumed at any time or location during polony development and the maximum depletion from its initial concentration is ~2.3% (concentration at the origin at cycle 40 = 293.221 vs. initial concentration of 300 molecules/μm^3). This is because Q is replenished inside the polony by diffusion from the region outside of the polony, which is in turn replenished by the Q source at the solution space boundary. In the course of polony PCR, Q becomes converted to S by virtue of hybridizing with T and subsequent extension. Because, unlike P, Q does not become limiting, the concentrations of its extension product S do not plateau in the interior of the polony as do the concentrations of P's extension product T. S concentrations thus continue to increase in the polony interior as the polony grows in radius.

These developments in polony morphology may also be seen by looking at the concentrations of the hybridized species PS and QT after PCR annealing phases (Fig. 3). PS and QT are the only species contribute to the generation of new S and T strands. In the first phase of growth the concentrations of these species are maximal at the origin, reflecting the generation of both S and T throughout the spherical region occupied by the as yet immature polony. After polony maturation, PS and QT concentrations are maximal at non-zero radii from the origin, and the
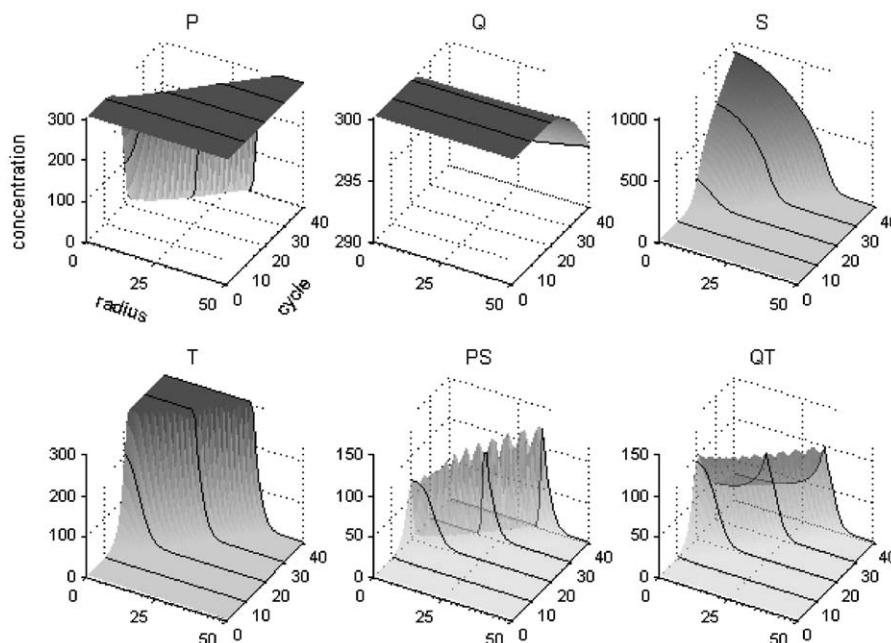

Fig. 3. Results of a 3D SPGM simulation using standard parameters (Table 2) that illustrate key features of polony morphology and development. Each graph displays the concentration of a molecular species as a function of distance from the origin (radius), which is the location of the original polony template strand seed molecule, at the end of a particular phase of each of the 40 PCR cycles. Shown are P, Q, S, and T at the end of PCR denaturing phases, and PS and QT after PCR annealing phases. Axis labels are indicated on the P graph. The solid lines in each graph are the concentration curves for cycles 10, 20, 30, and 40. Concentrations are in units of molecules/μm^3. See text for details (Results).

radius of maximal concentration increases with each cycle. The maximal values of PS and QT are comparable. For PS, the region of significant concentration is narrowly constrained around the maximal concentration radius, defining the narrow spherical shell in which new T strands are generated. For QT, concentration drops off sharply to insignificant levels outside of the maximal concentration radius, but remains significant inside this radius, reflecting the saturating concentration of T inside the polony and its absence outside. The concentration of QT inside the polony diminishes spatially with decreasing radius and temporally with increasing cycle due to the ever increasing concentration of S, which competes against Q for T during annealing.

These general features of polony morphology and development also appear in 1D and 2D simulations and are robust to changes in some polony parameters. For instance, lowering the diffusion coefficients of Q and S, even by two orders of magnitude, does not alter these general features, although it has substantial effects on the shape and size of the S concentration distribution interior to the polony and the rate of growth of the polony radius. (See supplemental information http://arep.med.harvard.edu/polony_models/)

3.1.2. Single polony yield

We also investigated the yield of S and T as a function of PCR cycle. The two phases of growth described in Section 3.1.1 can be plainly seen in Fig. 4a, a 3D SPGM simulation with standard parameters (Table 2) except for being carried out to 100 cycles: Both S and T exhibit exponential growth during the first phase of polony development, seen in the linear increase in their $\log_{10}$ yields, while the second phase exhibits sublinear increases in $\log_{10}$ yield. Early simulations of polonies based on very simple models had already suggested that polony yield would first increase exponentially and later shift to polynomial growth; additionally, focusing on T, the species used by polony assays (see Methods) and therefore of greatest interest, the expectation that T should accrete in mature polonies by spherical shells of constant thickness suggested that polynomial growth should be of degree 3 (cubic) (Mitra and Church, 1999). With our more sophisticated SPGM in hand, we re-examined these issues by performing two kinds of regressions on our computed $\log_{10}$ T yields.

We first dissected the $\log_{10}$ T graph into an exponential and a power law part by means of a least-square regression of the form

$$\log_{10}(T\ yield) = (A + B \cdot \log(C \cdot cycle))(\eta(cycle - D))$$
$$+ (E + F \cdot cycle)(1 - \eta(cycle - D)), \quad (1)$$

where $\eta(x) = 1$ for $x > 0$ and 0 for $x < 0$ (approximated in our regression by a sharp logistic function), and found $(A, B, C, D, E, F) = (-0.4, 4.3, 1.9, 23.3, -0.2, 0.3)$. At 23.3, the value of D, the regression estimate of
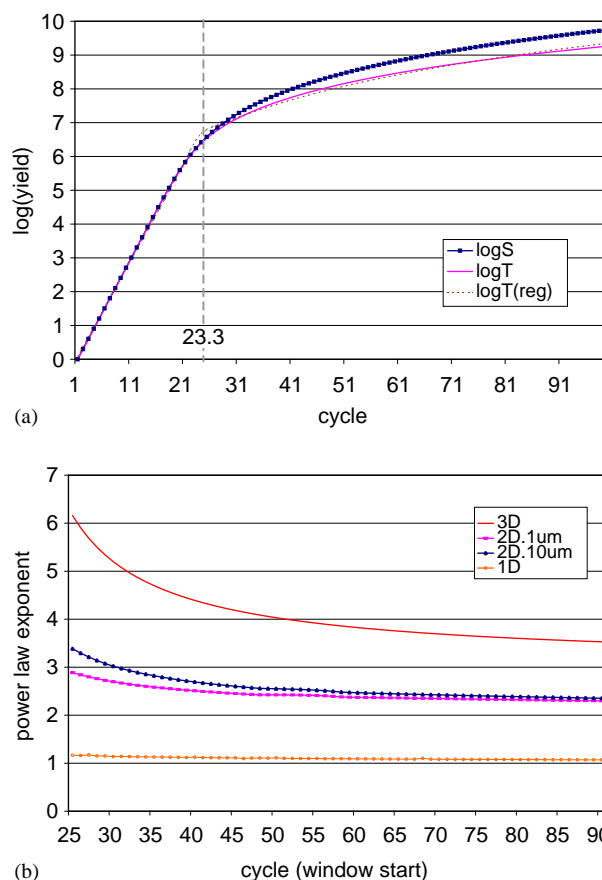

(a)

(b)

Fig. 4. Analysis of estimated polony yield by cycle. (a) Graph of $\log_{10}$ yield of S ($\log$ S) and T ($\log$ T) after every denaturing cycle from a 3D SPGM simulation with standard parameters (Table 2) except for being carried out to 100 cycles. Also shown ($\log$ T(reg)) is the graph of the regression of $\log$ T described by Eq. (1) in the text. In this regression, the cycle at which yield switched from exponential to power law growth was 23.3 (indicated by the vertical dashed line). The power law exponent was 4.3. (b) Evolution of power law exponent for $\log$ T as determined by power law regressions over windows of 10 consecutive $\log_{10}$ T yields, showing gradual decline of power law exponents moving towards apparent asymptotes that depend on dimension. The 3D curve is the evolution of the power law exponent for the 3D simulation shown in (a). The 2D curves are power law exponent curves for 2D simulations of with standard parameters (Table 2) for gels that are 1 μm thick (2D.1 μm) and 10 μm thick (2D.10 μm). The 1D simulation assumes a 1D linear gel with a 1 μm² cross section (Table 2).

the cycle at which the polony matures, compares well with our identification above of cycle 22 as the break-point based on the cycle at which P is first exhausted in the polony interior. The estimated power law exponent $B = 4.3$ is well above the cubic growth we expected, but regression (1) assumes a fixed power law exponent after the break from exponential growth and inspection of the $\log$ T curve in Fig. 4a suggested that the exponent might be gradually decreasing. We therefore performed straight power law regressions

$$\log_{10}(T\ yield) = A + B \log(C \cdot cycle) \quad (2)$$

for windows of every 10 cycles starting with cycle 25. When we plotted the computed power law exponents (*B*) against the window start cycle (Fig. 4b), we observed that these exponents do in fact exhibit a steady decline. Surprisingly, however, even at 91–100 cycles, the exponent is still considerably above the expected value of 3 (3.52). A reasonable hypothesis is that the exponent asymptotically approaches 3 from above. Consistent with this, similar power law regressions performed on the results of two 2D SPGM and one 1D SPGM also show an apparent asymptotic approach to exponents 2 and 1, respectively (Fig. 4b), values that would be expected based on applying the argument used in the 3D case to these lower dimensions. That these apparent asymptotes depend on the dimension of the model and not on the higher concentrations of growth-fueling primers P and Q available in higher dimensional models (see Table 2) may be seen from the fact that the two 2D SPGM simulations have P and Q concentrations that differ by 10× but yet tend to the same power law exponent. While we can as yet offer no analytic proof that these are the actual exponent asymptotes, we

provide an argument for this result along with additional analysis of these results on our web site http://arep.med.harvard.edu/polony_models/.

3.2. Two polony interaction: exclusion and invasion

As already noted, experiments with polonies had yielded the observation that nearby polonies sometimes deform against each other rather than interpenetrate (Fig. 1b), and there is interest in exploiting this phenomenon to increase the density of distinct polonies. We found we could simulate 'polony exclusion' using the TPIM and provide an example in Fig. 5. Fig. 5 exhibits results from the end of the denaturing phase of cycle 40 of a 3D TPIM simulation using standard parameters (Table 2). The simulation began with single ST and UV seed molecules at +12 and −12 μm, respectively, from the origin along the *Z*-axis. Three-dimensional TPIM simulations employ 3D cylindrical coordinates that use a Cartesian *z* coordinate (*Z*-axis) and a 2D polar radial coordinate (*R*-axis). Figs. 5a and b show the concentrations of T (5a) and V (5b) as a
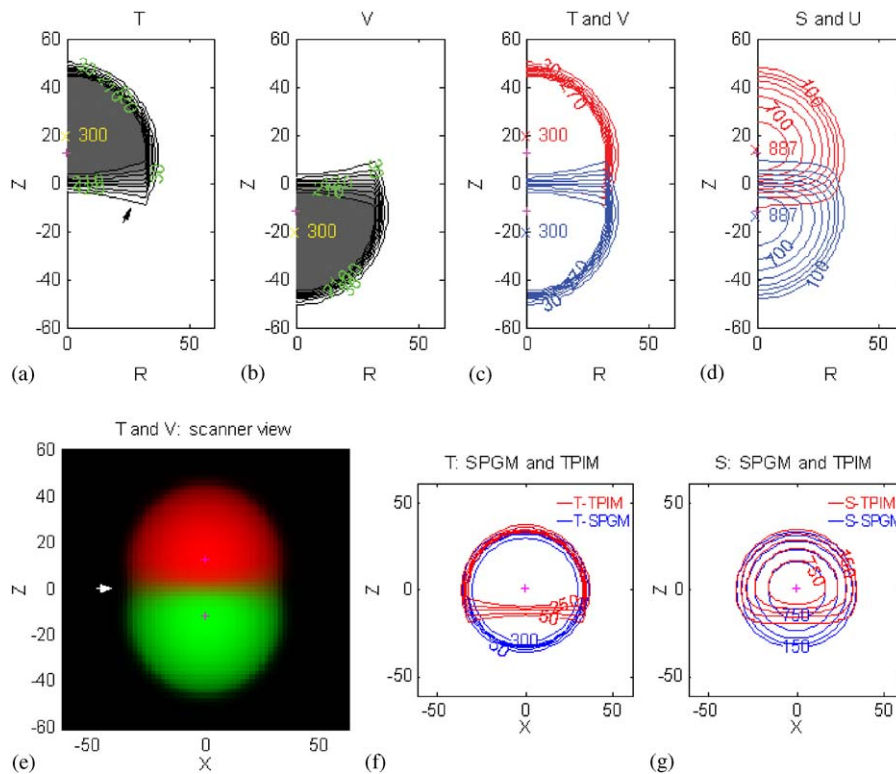

Fig. 5. Polony interaction simulation illustrating the polony 'exclusion' phenomenon. Graphs show results from the end of the denaturing phase of cycle 40 from a 3D TPIM simulation using standard parameters (Table 2) that began with single ST and UV seed molecules at +12 and −12 μm from the origin along the *Z*-axis. See text for details. (a) and (b) Filled contour plots showing concentrations of tethered strands T (a) and V (b). (c) and (d) Contour plots showing overlays of T and V (c) and overlays of corresponding free strands S and U (d). (e) Scanner view (see text) of interacting T and V polonies, generated by computationally rotating, projecting, and integrating the concentration data of (c), that may be compared with scanned images of actual polonies obtained by fluorescence microscopy (Fig. 1b). T density is indicated by red, V by green. (f) Comparison of T from this TPIM simulation with T from a corresponding SPGM simulation, where there is no interaction with another polony. (g) Corresponding comparison of S from this TPIM and the corresponding SPGM simulation. Arrowheads: see text. ×: maximum concentration (a–d). + (magenta):location of initial polony seed molecule.

function of R and Z. T and V are the tethered strands associated with the two polonies, and the concentrations observed in these figures define what we call the T and V polonies. We refer to the half-space $Z > 0$, where the initial T seed was located and most of the final T product is found, as the region 'owned' by the T polony, and likewise the region $Z < 0$ is owned by the V polony; we also call the boundary plane $Z = 0$ the midplane. Here we see that only a small amount of T has infiltrated the space owned by the V polony (arrowhead, Fig. 5a), and this T density remains close to the Z-axis. Due to equality of rate constants and diffusion coefficients involving S and T, and U and V in this simulation (Table 2), the T and V polonies are reflections of each other about the Z-axis and so an equally small amount of V has infiltrated the T region. The small degree of overlap is shown in Fig. 5c, in which plots of the T and V polonies are superimposed, and also in Fig. 5d, which shows the concentration densities of S and U, the free strands corresponding to T and V.

Figs. 5a–d are all $R−Z$ plots of concentrations, which represent the concentrations that would be seen on any half-plane extending from the Z-axis. In actual polonies, only a confocal microscope focused on such a plane could see these concentrations, while the non-confocal fluoroscopic scanned images of polonies (e.g. Fig. 1b) used for polony assays instead collect light emitted from a substantial volume of the polony. In order to generate the perspective of an image like Fig. 1b from the $R−Z$ plot of Fig. 5c, the density in the $R−Z$ plot must be rotated in space around the Z-axis, projected to the $X−Z$ plane, and integrated. The result of this transformation is shown in the generated image Fig. 5e (scanner view), which compares well with the images of 'excluded' polonies indicated Fig. 1b. The narrow area of T and V polony overlap seen in Figs. 5a–c corresponds to a thin but relatively well-defined band of merging that separates the two polonies (Fig. 5e, white arrowhead). The degree of deformation associated with the interaction of the polonies can be assessed in Figs. 5f and g. In Fig. 5f, compares the T polony in the TPIM simulation with a T polony from an SPGM 3D simulation by overlaying 'confocal' (i.e. *not* projected and integrated) views of T concentrations in the $X−Z$ plane. As can be seen, the TPIM T polony looks like an ordinary SPGM T polony whose edge has been pushed back by the interaction with the V polony. Fig. 5g shows a corresponding comparison of the S concentrations in the TPIM T polony with those of an SPGM T polony.

These results suggest a simple explanation for the polony exclusion phenomenon: As the T and V polonies develop, they individually approach the state of maturity (see Section 3.1) in which tethered primer P is exhausted in the polony interior by virtue of being completely converted to tethered strand. In consequence, by the time the T and V polonies interact, they

have each created individual regions that are invasion-proof: There is no longer any P available in the interior of the T polony by which V could form by pairing and extension of P against U strands diffusing in from the V polony, and likewise, no P in the V polony to support the generation of T. Indeed, in this case, the T and V polonies are so far apart that they are hardly in contact at all when they mature at cycle 22 (see above). Three dimensional SPGM results indicate that if the T polony were grown in isolation, only 1.1% of the T yield and 1.4% of the S yield would have crossed a plane 12 μm away from the seed location, which, in the 3D TPIM context, represents the amount of T that could have crossed the midplane into the V polony's space. Consistent with this low amount of tethered strand near the midplane, the concentration of P available at the origin at cycle 22 in the TPIM simulation is $\sim 72\%$ of its initial value, indicating that there is still plenty of tethered primer left there to be converted to T and V in subsequent cycles. This pool of P in the midplane region supports the development there of a mixed population of tethered strands, but two factors delimit the mixture away from the midplane: (i) More P will have been already used up by the nearest polony leaving a smaller pool available for mixture, e.g. already in cycle 22, the fraction of available P drops to 45% at 4 μm away from the origin in the direction of either polony, and to 1% at 8 μm, according to the TPIM. (ii) On the T polony side, higher concentrations of S vs. U will preferentially convert this P to T than V, and vice versa. The consequence will be that, as cycles continue, the mixture region will be bounded by sharp gradients of tethered strands, seen in Figs. 5a and b as closely spaced contours at the frontiers between the T and V polonies.

If this explanation is true, we should expect to see more mixture and less well-defined exclusion if interacting polonies are seeded more closely together. This is shown in Fig. 6, a version of Fig. 5 based on a TPIM simulation in which polony seeds were at ± 4 μm along the Z-axis rather than ± 12 μm. Compared to the ± 12 μm case, the ± 4 μm simulation results in more merging of the T and V polonies. Contours of V now protrude more deeply into the T polony region (see arrows in Figs. 5a and 6a), and the scanner view shows more merging and less delineation between the polonies (see arrows in Figs. 5e and 6e). Rather than exclusion, the results in Fig. 6 may be better described as polony invasion. This invasion exhibits a complex geometry: The projecting contour at the arrow in Fig. 6a suggests that the T polony invades the V polony at its flanks, but this is not an accurate interpretation of the $R−Z$ plot. Rather, the indicated contour represents a surface in 3D space obtained by rotating the contour around the Z-axis that is shaped like the interior of a wine glass, and the nested projecting contours indicate an onion-like nesting of volumes bounded by such surfaces. Polonies

Fig. 6. Polony interaction simulation illustrating the polony invasion phenomenon. Graphs show results from the end of the denaturing phase of cycle 40 from a 3D TPIM simulation using standard parameters (Table 2) that began with single ST and UV seed molecules at $+4$ and $-4\,\mu m$ from the origin along the Z-axis. See text for details and Fig. 5 for figure layout and notation.

invade each other by enveloping each others' centers. Whereas in the $\pm12\,\mu m$ case, the T and V polonies were far enough apart to mature without interference from the neighboring polony, in this case the polonies never get a chance to mature completely in the sense that neither T nor V anywhere achieves the maximum possible concentration level of $300\,molecules/\mu m^3$, the initial concentration of P. Rather T and V are mixed everywhere. Compared to the $\pm12\,\mu m$ case, the $\pm4\,\mu m$ polonies show more deformation relative to SPGM polonies (compare Figs. 6f and g with 5f and 5g). Instead of the tightly bunched contours T of Fig. 5f, the contours fan out in Fig. 6f.

Both polony exclusion and invasion are seen in TPIM simulations in lower dimensions (Fig. 7). In 2D simulations (Fig. 7a), invasion exhibits a geometry very similar to 3D invasion. Scanner views of 2D invasion (Fig. 7b) and 3D invasion (Fig. 6e) are very similar despite the fact that the TPIM employs $X-Y$ Cartesian coordinates and contours do not need to be rotated around a Z-axis as in the 3D case (see above). In the 1D case (Figs. 7c and d), a striking phenomenon is observed at close seed distances that could be construed as paradoxical: Concentrations of the invading strand V are depressed underneath the large density peaks of S and T in the middle of the T polony, but recover on the far side of the T polony, a phenomenon we call

'tunneling' (arrows, Fig. 7d). While we do not have a complete explanation for tunneling, evidence suggests that it reflects transient conditions in effect at the advancing T polony edge during the course of polony development that are preserved in the invading V spatial concentration profile (see supplemental material at http://arep.med.harvard.edu/polony_models/). 1D tunneling may illuminate a question concerning 3D invasion profiles: Do protruding invasion contours such as shown by the arrow in Fig. 6a arise by diffusion and amplification of density that moves outward in the manner indicated by the red arrow in Fig. 6b? While this is a plausible hypothesis, the 1D results suggest that the contours could also arise by tunneling from the polony core (black dashed arrow in Fig. 6b). At this time we have no account of how much either of these mechanisms contributes to the development of invading density.

4. Discussion

The mathematical models we have developed for the diffusion-constrained PCRs used to generate polonies (Mitra and Church, 1999) have clarified several aspects of polonies important in their practical use and development. In several cases model-based analysis accords with reported experimental observations or
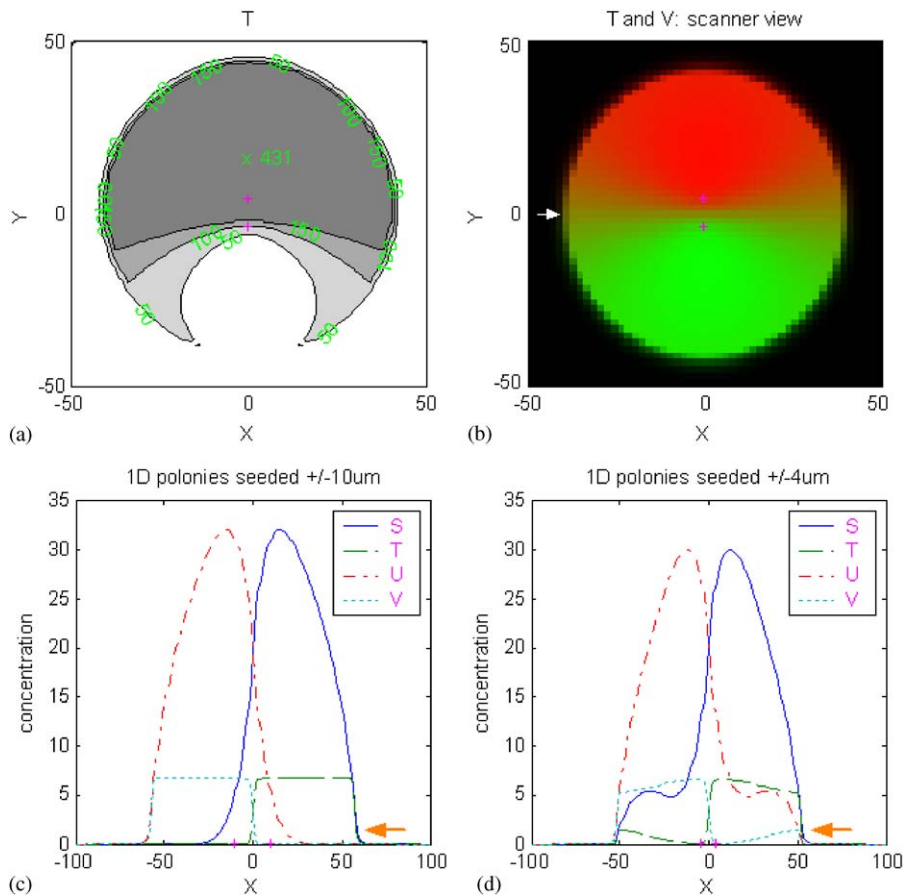
Fig. 7. Polony exclusion and invasion in two dimensions (a and b) and one dimension (c and d). (a) T concentration at end of denaturing phase of cycle 40 for 2D TPIM simulation of a 10 μm thick gel (Table 2) with polony seed molecules for ST and UV at ±4 μm on the Y-axis (indicated by magenta + signs). T concentration profile shows complex invasion geometry similar to 3D invasion geometry of Fig. 6a. (b) Scanner view of T and V polonies from simulation in (a) shows merging along the midline similar to Fig. 6e (arrow). (c) S, T, U, and V concentrations at end of denaturing phase of cycle 40 for 1D TPIM simulation (Table 2) with polony seed molecules ST and UV at ±10 μm, showing well-defined exclusion of polonies. T density is indicated by red, V by green. (d) Same variables as in (c) for 1D TPIM simulation with polony seeds at ±4 μm, showing polony invasion with 'tunneling' described in text (compare arrows in (d) and (c)). + (magenta): polony seed locations.

extends previous models and explanations. For instance, the growth of polonies in two phases, an initial exponential phase followed by a cubic growth phase, reported in Mitra and Church (1999) is analysed in much greater detail: we can now associate the break-point between phases with the development of an advancing growth face in mature polonies, and see the mature phase as exhibiting a more complex power law growth that apparently only approaches cubic growth asymptotically. The models also provide a detailed account of events that may underlie the observed phenomenon of polony exclusion (Mitra et al., 2003b) based on the inability of forming tethered strands for one polony in a region of another polony that has already exhausted tethered primer P. Also, model-generated scanned images of excluding polonies compare well with actual images (Figs. 6e and 1b).

Meanwhile, the models suggest new directions for experimental research on polonies and development of polony technology. On one level, model-based predictions of polony density distributions and morphology may spur more careful microscopic examination of polonies, possibly including confocal microscopy, and may thus improve both empirical and theoretical understanding of their fine structure. Of particular interest would be the observation of the predicted complex geometry of polony invasion: so far unreported, this has not been a research target because overlapping polonies are usually simply passed over as 'rejects' in polony assays and are not considered interesting; however, experimental observation of these geometries could provide useful confirmations of modeling assumptions or point the way to important corrections.

On another level, model predictions may suggest directions for polony protocol development. For instance, results that consistently show that yields of free strand S are much higher than those of tethered strand T

may encourage exploration of techniques for post-generation capture of S in polony gels, allowing S, the polony product of greatest yield, to be directly used in polony assays rather than simply diffused away, as current protocols direct.

Our most immediate interest is to understand the trade-offs between polony yield, density, and size, and the factors that govern polony exclusion, so that polony assays can be made as high-throughput, accurate, and inexpensive as possible (see Introduction). The initial models described here are not yet developed to the point where definitive answers to these questions are possible, but the path to improvements is clear: (i) currently the models focus on diffusion coefficients that are only indirectly controlled experimentally rather than parameters that are directly manipulated such as DNA strand lengths, monomer concentrations, temperature, etc. One set of improvements will be to incorporate relationships that will allow the model to deal with these directly manipulated values, so that predicted and experimental results can be more easily compared. (ii) Several idealizations incorporated into the models can be refined or replaced with more accurate formulations. For instance, currently the models do not consider that diffusion and annealing may continue to take place in PCR replication cycles, that dissociation of double-stranded DNA molecules during denaturing does not take place instantaneously but is subject to rate constants, or that some amount of replication may actually take place during annealing. Also, in some circumstances the TPIM should take into account the possibility of hybridization between S and V, and between U and T, that is currently ignored. For the most part, these changes will require determination of new rate constants and their incorporation into the model equations of Fig. 2. (iii) Improvements will also

come from more accurate determination of the strand hybridization rates already recognized in the initial model so that they directly take into account the physical and chemical characteristics of polony gels (see Appendix A for the basis of current estimates).

But while these improvements are being made, the process of optimizing polony protocols can begin with the current initial models. For instance, as a step towards our goal of optimizing polony exclusion, we have developed two simple measures of polony invasion: (a) half-space $V/(V + T)$ (designated VVT) is the ratio of the total amount of V over the total amount of tethered strand product $(V + T)$ in the T polony's space. (b) Fractional loss of T yield (FLTY) is $1 - r$, where r is the ratio of total T yield over all space in a TPIM simulation over the total T yield in an SPGM simulation with the same model parameters. While VVT directly measures V infiltration, FLTY measures the total impact on the T polony of a neighboring polony. Both should be directly measurable in actual polonies with suitable labeling and scanning. Graphs of VVT and FLTY as functions of PCR cycle and seed distance are shown in Fig. 8. These figures give at a glance the effect of one directly manipulable parameter on polony exclusion: the number of PCR cycles. Consistent with our observations above (Section 3.2), FLTY shows a sharp rise at a threshold cycle that represents the cycle at which the polonies have grown large enough to overcome the seed distance originally separating them and begin interacting, while VVT shows a sharp but smaller rise for larger seed distances. To optimize polony exclusion, we can track the effect on these measures of other manipulable parameters such as diffusion coefficients, controllable through DNA template length and gel density (Mitra and Church, 1999), or tethered and free primer concentrations. A more sophisticated measure that can
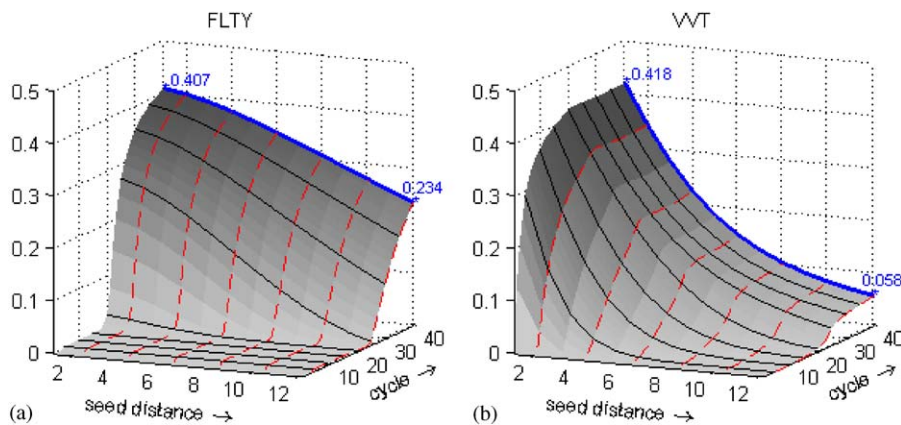

Fig. 8. Measures of the extent of polony invasion in 3D TPIM simulations. Two measures are shown as functions of polony seed distance from the origin (in μm) and PCR cycle, and are based on concentration profiles and yields of tethered strand T after denaturing phases. Blue lines show each measure as a function of seed distance from origin at PCR cycle 40, the last cycle for standard simulations (Table 2). (a) FLTY measures the fractional loss in T yield due to the presence of a neighboring polony compared to an SPGM simulation in which the T polony develops without interference. (b) VVT is the total amount of V, the tethered strand of a neighboring invading polony, divided by the total of the amounts of V and T, in the half-space owned by the T polony.

be computed with the current initial models, but which has not yet been attempted, will involve computing the amount of yield of a polony that is at or above a certain threshold purity. This measure will improve on FLTY and VVT because it directly relates to factors that make polonies usable, i.e. they must have enough material in them to be detected by a scanner, and they must be pure enough that parallel in situ sequencing (see Introduction and Mitra et al., 2003b) will not confound base calls from multiple template molecules. The dependency of this measure on diffusion coefficients, primer concentrations, and PCR cycles and cycle times, will provide an even clearer basis for the optimization of exclusion.

A second priority is to generate variants of the models that keep pace with new experimental protocols for polony development. For instance, assembly of sequences obtained by in situ sequencing (Mitra et al., 2003b) will benefit from the ability to sequence both ends of long DNA molecules, a standard technique in sequencing (IHGS, 2001; Venter et al., 2001). One way to do this would involve tethering both strands of a double-stranded template molecule, and this could be accomplished by adding untethered variants of P and tethered variants of Q to the polony gel, along with the usual tethered P and untethered Q. The required modifications to the model are straightforward and can be used to gauge the effect of these protocol changes on polony growth and development.

Aside from their application to polonies and polony technology, the models described here may be of general interest because polonies represent extremely simple abiotic systems that exhibit lifelike attributes of self-organized growth, development (maturity), functional morphology (growth faces), self-preservation (exclusion), and invasion behavior that is not only surprisingly complex (invasion geometry) but in some cases nearly paradoxical ('tunneling'). Of course, the degree of self-organization and complexity achieved by polonies is primitive compared to bacteria such as *E. coli* and *S. typhimurium*, or the yeast *S. cerevisiae*, which can form very complex and structured colonies (Budrene and Berg, 1995; Reynolds and Fink, 2001; Woodward et al., 1995). Our analysis adds polonies to the list of simple reaction–diffusion systems which have been found to exhibit complex, dynamic, and even self-replicating patterns (Lee et al., 1993, 1994; Reynolds et al., 1994; Turing, 1951; Zaikin and Zhabotinsky, 1970). Compared with most of these reaction–diffusion systems, many of which evolve to oscillating or steady-state behavior, it is striking that polony generation seems endowed with a kind of directionality that appears to bring it a step closer to biological systems, manifested in our vocabulary of development and maturation. What accounts for this difference? Partly it is simply our focus. Most of these other systems are studied with respect to the diverse patterns that may arise from different and

often randomized initial conditions. With polonies, we study a particular pattern starting from fixed initial conditions that reflect our technological goals. However, there are also elements in polony systems not present in most of these reaction–diffusion systems that may contribute to their apparent directionality. In addition to molecular species with different diffusion coefficients, a common feature of lifelike reaction–diffusion systems, some species in polonies are actually tethered and do not diffuse at all, and the accretion of tethered product molecules may add a novel element of direction. Also, compared to these other systems, polony protocols are programmed by the PCR cycle so that three different reactions take place successively, possibly adding a novel element of synchronization (although the programming in polonies, unlike the genetic programming of living organisms, is completely devoid of environmental or internal regulation). The fact that some polony molecules have the capacity to self-replicate is not a major difference from these other systems, as auto-catalysed production is generally assumed at some level.

But these differences from other reaction–diffusion systems, though subtle, may have broad biological significance. Certainly, sequences of reactions in the natural world may be programmed by natural cycles such as diurnal cycles. But we focus here on tethering, which is clearly important to a great many biological processes. Through tethering, sets of molecules, cells, or organisms are kept in close enough proximity for them to functionally interact, and, indeed, in this capacity tethering has been accorded a critical role in the origin of life itself (Ferris and Ertem; 1992; Wachtershauser; 1994). In polonies tethering does not function in this way, but may instead play a more primitive role in establishing a developmental direction simply by promoting accretion. In cells or organisms, accretion may set the stage for more complex environmental and developmental regulation, and the onset of tethering itself may part of a genetic program. But there may also be instances in biology where a more primitive polony-like mode of unregulated accretional development still obtains. Biofilms present an interesting case. Though it is clear that biofilms are heavily influenced by genetically programmed interactions between the organisms that compose them, and between the organisms and their environment, the question of whether they represent products of an integrated developmental process vs. globally unregulated aggregations of microbes is the subject of current lively debate (Ghigo, 2003; Kjelleberg and Molin, 2002; O'Toole et al., 2000; Stoodley et al., 2002; Watnick and Kolter, 2000).

Finally, we note that the ideas in our polony models may also apply to biological propagation which feature tethering and diffusion, for instance to organisms such which have a sessile phase and release spores or larvae which drift by diffusion (e.g. mushrooms). In such cases,

P might represent habitats needed by the sessile phases, and Q might represent nutrients needed for reproduction, or sperm. We are searching for candidate systems to which this framework may be applied.

In summary, we have developed several initial models for diffusion-constrained PCRs that are a basis for new high-throughput NA assay technologies. These models have clarified several aspects of the growth and interaction of the polymerase colonies (polonies) by these reactions, and will be used to help optimize the technologies and their applications. Polonies are also of interest because they are simple abiotic systems that exhibit primitive forms of self-organization and directed development characteristic of living organisms. The polony models may also have application to biological propagation systems which feature tethering and diffusion.

Appendix A. Derivations of 'standard' parameters

The following standard parameter values apply to 3D models. Values for 1D and 2D models are based on the dimensional normalization described in Models and Methods. As diffusion coefficients and rate constants for the U and V strands in TPIM have been assumed equal to those for S and T in SPGM, only the S and T parameters are described here.

Parameter	Explanation
iP, eQ	300 molecules/$\mu m^3 \approx 0.5\,\mu M$ (pers. comm. R. Mitra; cf. 1 μM in Mitra and Church, 1999).
T_d	30 s (pers. comm. R. Mitra; cf. 30 s in Mitra and Church, 1999).
T_a	45 s (pers. comm. R. Mitra; cf. 45 s in Mitra and Church, 1999).
cyc	40 (pers. comm. R. Mitra; cf. 40 cycles in Mitra and Church, 1999).
D_S	Derived from Fig. 3 of Mitra and Church (1999), considering a polony of $\sim 100\,\mu m$ radius that develops from a 200 nt DNA strand (a radius that corresponds to a gel containing between 10% and 15% acrylamide). Assuming that a polony radius grows each cycle by roughly an average diffusion length estimated by sqrt($6D_S T_d$) (Berg, 1993), D_S may be solved as $\approx 0.035\,\mu m^2/s$ from cyc·sqrt($6D_S T_d$) = 100 μm. Note that simulations based on this value typically yield polonies with radii of 40–50 μm.
D_Q	Also derived from Fig. 3 of Mitra and Church (1999). The 10% acrylamide line in this figure has the approximate equation $\Delta\log_{10}(R) \approx -1.4\Delta\log_{10}(L)$. Assuming the length of primer Q is 10% of the assumed 200 nt length of S, the radius of a polony derived from this 20 nt primer would be $\approx 2500\,\mu m$. Consistent with this, apparent primer–dimer polonies have, in fact, been observed that have radii on the order of mm. As with D_S, D_Q may be estimated as $\approx 20\,\mu m^2/s$ from cyc·sqrt($6D_Q T_d$) = 2500 μm.
k_{ST}	A renaturation rate of 2.6×10^7/M-s for the two strands of a double-stranded 564 nt DNA molecule is cited in Sikorav and Church (1991). For a mixture of DNA strands of complexity N, the renaturation rate constant k of two complementary strands, the shorter of which has length L_s, may be estimated by $k = k'_N L_s^{0.5}/N$, where k'_N is a nucleation rate constant (Wetmur, 1975). Therefore, the renaturation rate constant for the 200 bp ST molecule can be estimated by $2.6 \times 10^7 \cdot (200/564)^{0.5}$/M-s = 0.026 μm^3/molecule-s. Note that the original 2.6×10^7/M-s renaturation rate constant pertained to hybridization conditions different from a polony gel. Therefore this k_{ST} estimate must be considered only a rough approximation.
k_{PS}, k_{QT}	Same derivation as k_{ST} except that $L_s = 20$ for hybridization of a 20 nt primer against a 200 nt strand, yielding rate constants of $2.6 \times 10^7 \cdot (20/564)^{0.5}$/M-s = 0.008 μm^3/molecule-s. As with k_{ST}, these k_{PS} and k_{QT} estimates must be considered rough approximations.

References

Ames, W.F., 1992. Numerical Methods for Partial Differential Equations. Academic Press, San Diego, CA.

Badarinarayana, V., Estep, P.W., Shendure III, J., et al., 2001. Selection analyses of insertional mutants using subgenic-resolution arrays. Nat. Biotechnol. 19 (11), 1060–1065.

Berg, H.C., 1993. Random Walks in Biology. Princeton University Press, Princeton, NJ.

Braslavsky, I., Hebert, B., Kartalov, E., et al., 2003. Sequence information can be obtained from single DNA molecules. Proc. Natl Acad. Sci. USA 100 (7), 3960–3964.

Brenner, S., Johnson, M., Bridgham, J., et al., 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat. Biotechnol. 18 (6), 630–634.

Budrene, E.O., Berg, H.C., 1995. Dynamics of formation of symmetrical patterns by chemotactic bacteria. Nature 376 (6535), 49–53.

Bulyk, M.L., Huang, X., Choo, Y., et al., 2001. Exploring the DNA-binding specificities of zinc fingers with DNA microarrays. Proc. Natl Acad. Sci. USA 98 (13), 7158–7163.

Crank, J., 1956. Mathematics of Diffusion. Oxford University Press, London.

DeRisi, J.L., Iyer, V.R., Brown, P.O., 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278 (5338), 680–686.

Ferris, J.P., Ertem, G., 1992. Oligomerization reactions of ribonucleotides: the reaction of the $5'$-phosphorimidazolide of nucleosides on montmorillonite and other minerals. Origins Life Evol. Biosphere 22 (6), 369–381.

Ghigo, J.M., 2003. Are there biofilm-specific physiological pathways beyond a reasonable doubt? Res. Microbiol. 154 (1), 1–8.

IHGS (International Human Genome Sequencing Consortium), 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Ito, T., Chiba, T., Ozawa, R., et al., 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl Acad. Sci. USA 98 (8), 4569–4574.

Kjelleberg, S., Molin, S., 2002. Is there a role for quorum sensing signals in bacterial biofilms? Curr. Opin. Microbiol. 5 (3), 254–258.

Kopp, M.U., Mello, A.J., Manz, A., 1998. Chemical amplification: continuous-flow PCR on a chip. Science 280 (5366), 1046–1048.

Lee, K.J., McCormick, W.D., Quyang, Q., et al., 1993. Pattern formation by interacting chemical fronts. Science 261, 192–194.

Lee, K.J., McCormick, W.D., Pearson, J.E., et al., 1994. Experimental observation of self-replicating spots in a reaction–diffusion system. Nature 369, 215–218.

Lee, T.I., Rinaldi, N.J., Robert, F., et al., 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298 (5594), 799–804.

Lockhart, D.J., Dong, H., Byrne, M.C., et al., 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. 14 (13), 1675–1680.

Merritt, J., DiTonno, J.R., Mitra, R.D., et al., 2003. Parallel competition analysis of *Saccharomyces cerevisiae* strains differing by a single base using polymerase colonies. Nucleic Acids Res. 31 (15), e84.

Mitra, R.D., Church, G.M., 1999. In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res. 27 (24), e34.

Mitra, R.D., Butty, V.L., Shendure, J., et al., 2003a. Digital genotyping and haplotyping with polymerase colonies. Proc. Natl Acad. Sci. USA 100 (10), 5926–5931.

Mitra, R.D., Shendure, J., Olejnik, J., et al., 2003b. Fluorescent in situ sequencing on polymerase colonies. Anal. Biochem. 320 (1), 55–65.

O'Toole, G., Kaplan, H.B., Kolter, R., 2000. Biofilm formation as microbial development. Annu. Rev. Microbiol. 54, 49–79.

Press, W.H., Flannery, B.P., Teukolsky, S.A., et al., 1996. Numerical Recipes in C. Cambridge University Press, Cambridge.

Rehman, F.N., Audeh, M., Abrams, E.S., et al., 1999. Immobilization of acrylamide-modified oligonucleotides by co-polymerization. Nucleic Acids Res. 27 (2), 649–655.

Reynolds, T.B., Fink, G.R., 2001. Bakers' yeast, a model for fungal biofilm formation. Science 291 (5505), 878–881.

Reynolds, W.N., Pearson, J.E., Ponce-Dawson, S., 1994. Dynamics of self-replicating patterns in reaction diffusion systems. Phys. Rev. Lett. 72 (17), 2797–2800.

Sikorav, J.L., Church, G.M., 1991. Complementary recognition in condensed DNA: accelerated DNA renaturation. J. Mol. Biol. 222 (4), 1085–1108.

Singer, B.S., Shtatland, T., Brown, D., et al., 1997. Libraries for genomic SELEX. Nucleic Acids Res. 25 (4), 781–786.

Stoodley, P., Sauer, K., Davies, D.G., et al., 2002. Biofilms as complex differentiated communities. Annu. Rev. Microbiol. 56, 187–209.

Turing, A.M., 1951. The chemical basis of morphogenesis. Philos. Trans. R. Soc. London Ser. A 237, 37–72.

Uetz, P., Giot, L., Cagney, G., et al., 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 403 (6770), 623–627.

Velculescu, V.E., Zhang, L., Vogelstein, B., et al., 1995. Serial analysis of gene expression. Science 270 (5235), 484–487.

Venter, J.C., Adams, M.D., Myers, E.W., et al., 2001. The sequence of the human genome. Science 291 (5507), 1304–1351.

Wachtershauser, G., 1994. Life in a ligand sphere. Proc. Natl Acad. Sci. USA 91 (10), 4283–4287.

Watnick, P., Kolter, R., 2000. Biofilm, city of microbes. J. Bacteriol. 182 (10), 2675–2679.

Wetmur, J.G., 1975. Acceleration of DNA renaturation rates. Biopolymers 14, 2517–2524.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., et al., 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. Science 285 (5429), 901–906.

Woodward, D.E., Tyson, R., Myerscough, M.R., et al., 1995. Spatio-temporal patterns generated by *Salmonella typhimurium*. Biophys. J. 68, 2181–2189.

Zaikin, A.N., Zhabotinsky, A.M., 1970. Concentration wave propagation in two-dimensional liquid-phase self-oscillating system. Nature 225, 525–537.

Zhu, J., Shendure, J., Mitra, R.D., et al., 2003. Single molecule profiling of alternative pre-mRNA splicing. Science 301 (5634), 836–838.