A Generalized Mathematical Model To Estimate T- and B-Cell Receptor Diversities Using AmpliCot

Irina Baltcheva,[†*△] Ellen Veel,[‡△] Thomas Volman,[‡] Dan Koning,[‡] Anja Brouwer,[‡] Jean-Yves Le Boudec,[†] Kiki Tesselaar,[‡△] Rob J. de Boer,[§△] and José A. M. Borghans[‡△]

[†]Laboratory for Computer Communications and Applications, École Polytechnique Fédérale de Lausanne, Switzerland; and [‡]Department of Immunology, University Medical Center Utrecht, and [§]Department of Theoretical Biology, Utrecht University, Utrecht, The Netherlands

ABSTRACT The efficiency of the adaptive immune system is dependent on the diversity of T- and B-cell receptors, which is created by random rearrangement of receptor gene segments. AmpliCot is an experimental technique that allows the measurement of the diversity of the T- and B-cell repertoire. This procedure has the advantage over other cloning and sequencing techniques of being time- and expense-effective. In previous studies, receptor diversity, measured with AmpliCot, has been inferred assuming a second-order kinetics model. The latter implies that the relation between diversity and concentration × time (Cot) values is linear. We show that a more detailed model, involving heteroduplex and transient-duplex formation, leads to significantly better fits of experimental data and to nonlinear diversity-Cot relations. We propose an alternative fitting procedure, which is straightforward to apply and which gives an improved description of the relationship between Cot values and diversity.

INTRODUCTION

The diversity of T- and B-cell receptors (TCRs and BCRs) is a hallmark of the adaptive immune system, and is responsible for the specific recognition and the defense against a wide variety of pathogens. The structural diversity of BCRs and TCRs is achieved by somatic gene-segment rearrangements and random nucleotide additions and deletions (1). The estimation of the effective size of the human TCR repertoire, both in health and disease, is a fundamental question in immunology. Using single-molecule DNA sequencing, it was estimated that the number of unique TCRβ CDR3 sequences in a healthy adult is 3–4,000,000 (2).

Several experimental techniques have been used to measure the diversity of the TCR or BCR repertoire. Immunoscope (or spectratype) analysis provides qualitative insights into the repertoire's diversity in terms of clone sizes (3,4); high-throughput DNA sequencing exhaustively enumerates the different clonotypes that are present in a sample, thus providing a more detailed picture of the repertoire (5–9). Such deep sequencing techniques are expensive, can be very difficult to interpret because of sequencing and amplification errors, and can therefore not always be applied on large scale. AmpliCot has been introduced as an alternative approach, allowing for the measurement of the diversity of DNA samples through quantification of the rehybridization speed of denatured PCR products (10,11). It has the advantage over cloning and (deep) sequencing methods to be time- and expense-effective.

The AmpliCot experiment is based on the so-called "Cot analysis" (12), according to which the time required for a DNA sample to reanneal (expressed in terms of the product concentration × time, "Cot"), after it has melted, is related to the diversity of the sample. To estimate the diversity of a DNA sample from its annealing curve, Baum and McCune (10) proposed to analyze the Cot values at which, e.g., 50% of the sample is annealed ($Cot_{0.5}$ values). The authors suggested that the relation between $Cot_{0.5}$ values and diversity is linear, which is indeed true if the annealing process obeys second-order kinetics. Accordingly, it is assumed that only perfectly complementary pairs of DNA can associate, i.e., the possibility of heteroduplex formation is neglected. A recent study reported a systematic fluorescence loss at diversities exceeding 4×10^3 (13). Annealing curves of samples with diversity 10^6 and higher did not even reach the 50% annealing point, which made the determination of a $Cot_{0.5}$ value impossible. One explanation that was proposed is that the low concentration of highly similar sequences results in the formation of heteroduplexes (13).

Driven by these observations, we investigated how to deal with heteroduplex formation and its consequences for the interpretation of AmpliCot data. We formally define the previously used model, i.e., second-order kinetics, and we propose a more detailed model that considers the DNA annealing in two steps and takes into account the formation of transient duplexes and heteroduplexes. We then compare the ability of both models to fit AmpliCot annealing time-series. In doing so, we take advantage of the information contained in the entire annealing curves, rather than just the $Cot_{0.5}$ value. We use our model to derive what to our knowledge is a new formula describing the relation between Cot values and diversity. This formula is a generalization of the linear relation based on second-order kinetics. We show that the new generalized Cot expression accurately reproduces Cot

Submitted March 27, 2012, and accepted for publication July 16, 2012.

△Irina Baltcheva and Ellen Veel contributed equally to this work. Kiki Tesselaar, Rob J. de Boer, and Jose A. M. Borghans contributed equally to this work.

*Correspondence: irina.baltcheva@gmail.com

Editor: Leah Edelstein-Keshet.

0006-3495/12/09/0999/12 $2.00

http://dx.doi.org/10.1016/j.bpj.2012.07.017

values of highly diverse samples and leads to better interpretation of experimental data. Finally, we propose a diversity estimation algorithm that is simple to use and that can account for heteroduplex formation.

MATERIALS AND METHODS

AmpliCot assay

Samples containing PCR-amplified DNA or artificially synthesized oligonucleotides were mixed with SYBR green fluorescent dye, which binds to double-stranded DNA. To determine the specific melting point for each analysis, an aliquot of the double-stranded DNA (dsDNA) product (either a PCR product or double-stranded oligomer product) was melted by gradually increasing the temperature and determining the temperature at which the change in SYBR green fluorescence intensity peaked. The annealing temperature for each sample was subsequently set to be 3°C lower than its melting temperature. For AmpliCot analyses, three aliquots of the mixture were placed in a 96-wells plate as the annealing samples and a reference sample. The premelting step consisted of measuring the baseline fluorescence of the samples and reference at annealing temperature (Fig. 1 *A*). Subsequently, the temperature was increased to 95°C for 2 min to aim for total dissociation of the dsDNA strands, whereas the reference stayed at annealing temperature (melting step). The fluorescence intensity of the samples strongly decreased during the melting step as the double-stranded DNA dehybridized. During the annealing step, the temperature of the samples was set back to annealing temperature and the time-varying fluorescence intensity was measured every 5–20 s (Fig. 1 *A*). For any given total concentration, the resulting reannealing speed is expected to be dependent on the diversity of the sample, because in samples of high diversity, each sequence is present at a low concentration.

Experimental data

We tested our new mathematical model using four experimental data sets (see Table 1): the original oligonucleotide data set of Baum and McCune (10), two new data sets with diversities ranging from 1 to 48 and from 10 to 40, and the recently published data set of Baum et al. (14) that includes highly diverse samples.

Oligonucleotides that were used to create data sets 2 and 3 (Table 1) were synthesized according to the following format:

5′-GCTGGCGCAGAAATATACAGGTCGGACCTCAGCTG-**(NNNN)$_4$** CCTCAGCACCTCC-3′,

in which NNNN represents one of the eight nucleotide combinations AATC, ATCA, TCTA, CAAA, TTAC, TACT, ACAT, or CTTT (Eurofins MWG Operon, Huntsville, AL). Samples of the desired diversities were created by mixing the required amount of oligonucleotides at equimolar ratios. To slow down the annealing kinetics of the low diversity samples, some samples were diluted (see Table 1). There are two equivalent alternatives for handling concentration differences between samples. The first one is to adjust the annealing data by using Cot scaling (multiplication of time with the sample's concentration (Cot values)). The second consists of adjusting the concentration differences in the model equations by scaling the DNA association rates (see Eqs. 1 and 2).

Heteroduplex formation

We tested whether heteroduplexes tend to fluoresce less than homoduplexes, which may explain why highly diverse samples in which heteroduplex formation occurs tend to attain lower levels of fluorescence than homogeneous samples. The oligonucleotides used for these tests were synthesized according to the following format (Eurofins MWG Operon).

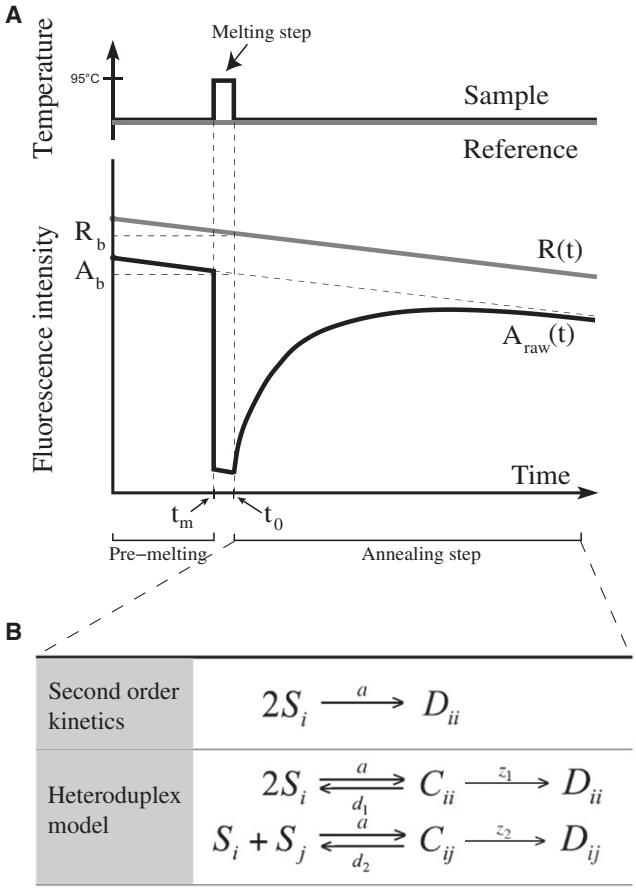

FIGURE 1 AmpliCot assay and model. (*A*) Samples containing PCR-amplified TCR genes or oligonucleotides are placed on both extremities of a 96-well plate as the samples and the reference. The baseline fluorescence intensity of samples and reference is measured at annealing temperature (premelting step). Samples are then melted at 95°C and their fluorescence drops (melting step). After 2 min of melting, the temperature of the samples is quickly set back to the annealing temperature to allow for reannealing of DNA strands (annealing step). $A_{raw}(t)$ and $R(t)$ are, respectively, the fluorescence intensities of the samples and reference at time t (minutes), or at the start of the annealing step (A_b and R_b, respectively). (*B*) Two possible models of the biochemical reactions occurring during the annealing step of AmpliCot. Second-order kinetics (*top line*) is the minimal model in which only homoduplexes are formed. The heteroduplex model (*bottom line*) considers the reaction in more detail. The association occurs in two steps (a first encounter followed by a zipping reaction), and includes the possibility of heteroduplex formation.

Main strand:
5′-GCTGGCGCAGAAATATACAGGTCGGACCTCAGCTGTTACTT ACACAT-CAAACCTCAGCACCTCCGCC-3′
Complementary strand:
5′-GGCGGAGGTGCTGAGGTTTGATGTGTAAGTAACAGCTGAG-GTCCGACCTGTATATTTCTGCGCCAGC-3′
Three mismatches:
5′-GGCGGAGGTGCTGAGGTTTGATGTGTAAGTAACA**TAC**GAG GT-CCGACCTGTATATTTCTGCGCCAGC-3′
Five mismatches:
5′-GGCGGAGGTGCTGAGGTTTGATGTGTAAGTAAC**TTACC**AG GTCC-GACCTGTATATTTCTGCGCCAGC-3′

TABLE 1 Four data sets of known diversity templates used in the analysis

Data set	Diversities	Number of replicates	Dilution factor
1 (10)	$n = 1, 2, 5, 10, 30, 48, 96$	1	Same for all n
2	$n = 1, 4, 8, 16, 32, 48$	2 ($n = 1, 4, 8, 16$) 1 ($n = 32, 48$)	1:4 ($n = 1, 4$) 1:2 ($n = 8, 16, 32, 48$)
3	$n = 10, 20, 30, 40$	2	Same for all n
4 (14)	$n = 1, 4, 16, 64, 128, 512, 896, 1568, 2744, 4900, 8750, 15{,}625, 25{,}000$	3	Same for all n

These oligonucleotides were directly mixed at high concentrations with SYBR green dye and subjected to the AmpliCot procedure. For these experiments, samples were melted at 95°C and subsequently annealed at 40°C. We chose this low annealing temperature because under these nonstringent conditions both homoduplexes and heteroduplexes will be formed (15,16).

MODEL

We considered two models describing the biochemical reaction of the annealing step of AmpliCot: second-order kinetics and the heteroduplex model (Fig. 1 *B*). We assumed that samples contain a large amount of DNA and that the material is well mixed, so both models could be described by ordinary differential equations. The main difference between the models is the level of detail incorporated in the description of the underlying biochemical reaction.

Second-order kinetics

Second-order kinetics is the simplest model describing AmpliCot (Fig. 1 *B*). It describes the association (at rate a) of two perfectly complementary single DNA strands under the assumption that the encounter of two complementary strands is the rate-limiting step, and that the subsequent hybridization is fast compared to the former process. Under these assumptions, the hybridization of DNA is a second-order reaction (10). Consider a DNA sample of diversity n. Let S_i be the concentration of single-stranded DNA (ssDNA) molecules of type i, where, for simplicity, a certain ssDNA and its complementary strand are both denoted by i. Consequently, D_{ii} is the concentration of homoduplexes of type i, where $i = 1,\ldots,n$. The following differential equations describe the second-order kinetics model:

$$\begin{aligned}\frac{dS_i}{t} &= -2aS_i^2,\\ \frac{dD_{ii}}{t} &= aS_i^2.\end{aligned} \tag{1}$$

Let $t_0 = 0$ be the beginning of the annealing phase of AmpliCot and let T be the total concentration of DNA strands in a sample (i.e., twice the concentration of dsDNA premelting). Let f_i be the proportion of ssDNA of type i at the beginning of the annealing phase. Ideally, there would be f_iT single-stranded molecules of type i at the beginning of the annealing phase. Because a small fraction of the DNA molecules may remain in the double-stranded form, we let α be the proportion of melted molecules at t_0 ($\alpha \in [0,1]$). Thus, the initial conditions for the above system are $S_i(0) = \alpha f_iT$ and $2D_{ii}(0) = (1 - \alpha)f_iT$, where $i = 1,\ldots, n$.

Heteroduplex model

The heteroduplex model (Fig. 1 *B*) takes into account the fact that hybridization involves two distinct processes: the association of short, homologous sites on two single strands, followed by a reversible hybridization (17). Two perfectly complementary single strands S_i form a partially hybridized homoduplex C_{ii}. Two partially complementary strands S_i and S_j can form a partially hybridized heteroduplex C_{ij} (where $j \neq i$). Partially hybridized homoduplexes (respectively, heteroduplexes) can dissociate at rate d_1 (respectively, d_2), or hybridize completely at rate z_1 (respectively, z_2) to form the final product D_{ii} (respectively, D_{ij}, $j \neq i$). Note that $C_{ij} = C_{ji}$ and $D_{ij} = D_{ji}$. The differential equations describing the change in time of the above-mentioned concentrations are:

$$\begin{aligned}\frac{dS_i}{dt} &= -2aS_i^2 - aS_i\sum_{j\neq i}S_j + 2d_1C_{ii} + d_2\sum_{j\neq i}C_{ij},\\ \frac{dC_{ii}}{dt} &= aS_i^2 - (d_1 + z_1)C_{ii},\\ \frac{dC_{ij}}{dt} &= aS_iS_j - (d_2 + z_2)C_{ij},\\ \frac{dD_{ii}}{dt} &= z_1C_{ii},\\ \frac{dD_{ij}}{dt} &= z_2C_{ij}.\end{aligned} \tag{2}$$

We assume that the melting process is fast compared to the reannealing, and that the melting temperature is so high that no rehybridization is occurring during the melting phase. Under these assumptions, the sample contains only ssDNA or unmelted dsDNA homoduplexes at the beginning of the annealing phase. The initial conditions for the above system are thus $S_i(0) = \alpha f_iT$, $2D_{ii}(0) = (1 - \alpha)\, f_iT$, and $C_{ii}(0) = C_{ij}(0) = D_{ij}(0) = 0$, where $i = 1,\ldots,n$, $j \neq i$, and $\alpha \in [0,1]$. Note that the heteroduplex model is a generalization of second-order kinetics; when setting $d_1 = d_2 = z_2 = 0$ and $z_1 \to \infty$ in the heteroduplex model (Eq. 2), one obtains the second-order kinetics model (Eq. 1).

Annealing kinetics

From the above model definitions, we define the kinetics of fluorescent DNA strands, $F(t)$. We assume that the latter are proportional to the concentration of double-stranded molecules at time t. In the case of second-order kinetics (SOK),

$$F^{\mathrm{SOK}}(t) = 2\sum_{i=1}^{n} D_{ii}(t) = T - \sum_{i=1}^{n} S_i(t). \tag{3}$$

In the case of the heteroduplex model, we allow heteroduplexes to have a decreased fluorescence intensity compared to homoduplexes (see Results below). This is modeled by weighting their level of fluorescence by a factor $\varphi \in [0,1]$. The concentration of fluorescent molecules under the heteroduplex model is hence defined as

$$F(t) = 2\left(\sum_{i=1}^{n} D_{ii}(t) + \varphi \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} D_{ij}(t)\right). \tag{4}$$

From the above expression (Eq. 4), we define the theoretical annealing curve, $A(t)$, as the proportion of fluorescent material in a sample, i.e., $A(t) = F(t)/T$, where T is the total concentration of DNA strands in a sample. We present here three expressions of the annealing kinetics: with $A(t)$ we denote the solution of the heteroduplex model; $A^{\mathrm{SOK}}(t)$ denotes the solution of the second-order kinetics model (i.e., a special case of $A(t)$); and $A^{\mathrm{data}}(t)$, the annealing kinetics of the experimental data.

To obtain a closed form solution of $A(t)$ and $A^{\mathrm{SOK}}(t)$, we solved the ordinary differential equation (ODE) systems analytically (Eqs. 1 and 2) for the case where all DNA species have the same concentration in the sample, i.e., under the equal molarity assumption (see Appendix 1 for the definition of the resulting mean-field systems). The equimolarity assumption makes the level of diversity (n) a parameter of the system. Moreover, to solve the heteroduplex model analytically, we applied a quasi-steady-state assumption for the transient complexes (see Appendix 2). The above transformations and the definition of $F(t)$ in Eq. 4 yield the expression

$$A(t;n) = \frac{F(t)}{T} = \left(\alpha - \frac{\alpha}{1 + 2\frac{a}{n}\left(\xi_1 + \xi_2\left(\frac{n-1}{2}\right)\right)\alpha Tt}\right) \times \left(\frac{\xi_1 + \varphi\xi_2\left(\frac{n-1}{2}\right)}{\xi_1 + \xi_2\left(\frac{n-1}{2}\right)}\right) + (1-\alpha), \tag{5}$$

where $\xi_1 = z_1/(z_1+d_1)$, $\xi_2 = z_2/(z_2+d_2)$ and n has been highlighted as an argument of the function $A(\bullet)$. Note that $A(t;n) \in [1-\alpha,1]$. The expression $A^{\mathrm{SOK}}(t)$ is a particular case of Eq. 5 and is obtained by setting $\xi_1 = 1$ and $\xi_2 = 0$ in Eq. 5:

$$A^{\mathrm{SOK}}(t;n) = \frac{F_{\mathrm{SOK}}(t)}{T} = 1 - \frac{\alpha}{1 + 2\frac{a}{n}\alpha Tt}. \tag{6}$$

To obtain the annealing kinetics from the raw experimental data, the experimental data were first normalized by correcting for the baseline fluorescence discrepancies of the reference and the sample and by correcting for the time-dependent fluorescence decline (see Fig. 1),

$$A^{\mathrm{data}}(t;n) = \frac{R_b}{A_b}\left(\frac{A_{\mathrm{raw}}(t)}{R(t)}\right), \tag{7}$$

where A_b and R_b are the fluorescence intensities of the sample and the reference at the start of the annealing step, which were estimated as the mean of the last 10 measurements of the premelting phase, assuming that the melting phase was short enough to ensure little loss of fluorescence during melting.

Cot values and annealing kinetics

The acronym "Cot" stands for "concentration × time" (12). In terms of our model, Cot = Tt. Cot values were used in the original AmpliCot article (10) to compare the annealing speed of samples with different DNA concentrations.

Let $s = Tt$ be a Cot value. The annealing kinetics can be expressed as a function of the Cot value s, by replacing the product Tt with the new variable s in Eqs. 5 and 6:

$$A(s;n) = \left(\alpha - \frac{\alpha}{1 + 2\frac{a}{n}\left(\xi_1 + \xi_2\left(\frac{n-1}{2}\right)\right)\alpha s}\right) \times \left(\frac{\xi_1 + \varphi\xi_2\left(\frac{n-1}{2}\right)}{\xi_1 + \xi_2\left(\frac{n-1}{2}\right)}\right) + (1-\alpha), \tag{8}$$

$$A^{\mathrm{SOK}}(s;n) = 1 - \frac{\alpha}{1 + 2\frac{a}{n}\alpha s}. \tag{9}$$

Model fitting

The models (Eqs. 5 and 6) were fitted to experimental data (Eq. 7) using a least-squares procedure (implemented in MATLAB version 7.10.0; The MathWorks, Natick, MA), applied to the log-transformed annealing curves. The 95% confidence intervals on parameter values were computed using 999 bootstrap replicates of each original data set. The bootstrap was done by sampling points $(t_i, A_{\mathrm{raw}}(t_i))$ from the raw annealing curves with replacement. The

bootstrap replicates were fitted in the same way as the original data set. The confidence intervals were computed using order statistics of the bootstrap distribution (18).

RESULTS

Heteroduplexes emit a lower fluorescence signal than homoduplexes

It was previously observed that samples of very high diversity may not reach the 50% annealing point (13). One hypothesis that would explain these observations states that the low concentration of perfectly complementary strands inside a huge excess of highly similar sequences results in the rapid formation of heteroduplexes, which will give a lower SYBR green fluorescence signal than homoduplexes (19,20). This would result in overestimation of the diversity of a given sample. Indeed, when we mixed oligonucleotides that were either perfectly complementary or contained three or five mismatches (i.e., a mismatch of 5% or 7.5% of the oligonucleotide length, respectively) at a temperature (40°C) well below their melting points, we observed the formation of heteroduplexes with a lower fluorescence intensity than homoduplexes (Fig. 2). The fluorescence level of the sample decreased as the number of mismatches in the complementary strand increased. These results show that heteroduplex formation may significantly influence the results of an AmpliCot experiment.

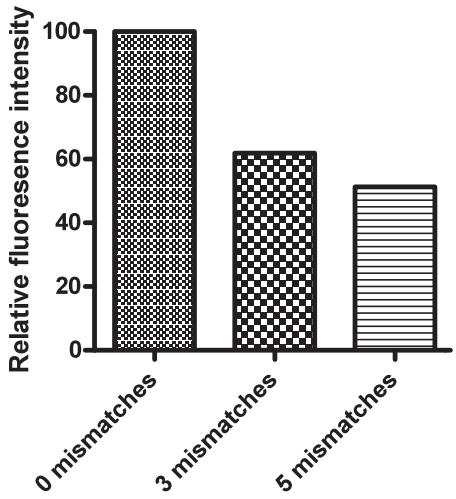

FIGURE 2 Heteroduplexes have a lower fluorescence intensity than homoduplexes. Formation of dsDNA products was analyzed at an annealing temperature of 40°C for ssDNA samples with 0, 3, or 5 nucleotide mismatches. The fluorescence signal of two complementary strands (homoduplex) was set to 100% and the fluorescence intensity of heteroduplexes was expressed as a percentage of the fluorescence intensity of homoduplexes.

Generalized expression for Cot values as function of diversity ($Cot_p(n)$)

The relation between Cot (concentration × time) values and diversity is important for the correct interpretation of the AmpliCot assay. Cot values of templates of known diversity are used to calibrate the assay, and are the benchmark for the inter- or extrapolation to unknown diversities. The procedure proposed in the original AmpliCot article (10) presumes the validity of second-order kinetics, i.e., it assumes that no heteroduplexes or temporary complexes are formed. We present here a mathematical expression that describes how Cot values depend on the diversity of the sample (n). The expression is based on the relaxed assumption that the annealing kinetics behave according to the heteroduplex model (Eq. 2), which is a generalization of second-order kinetics.

Let s^* be the Cot value for which a fraction p of a sample has annealed. We computed the formula presented hereafter by setting $A(s^*; n) = p$ in Eq. 8 and by solving for s^*. We call the generalized Cot expression $\mathrm{Cot}_p(n) = A^{-1}(s^*; n)$ (Eq. 10). Here, the fraction annealed, p, is considered as a parameter and the diversity, $n \geq 1$, is the independent variable, for

$$\text{Generalized Cot} = \mathrm{Cot}_p(n) = \frac{1}{2a\alpha}\left(\frac{\alpha-(1-p)}{\xi_1(1-p)+\xi_2[(1-p)+\alpha(\varphi-1)]\left(\frac{n-1}{2}\right)}\right)n, \tag{10}$$

where $\xi_1 = z_1/(z_1 + d_1)$ (respectively, $\xi_2 = z_2/(z_2 + d_2)$) is the proportion of homoduplexes (respectively, heteroduplexes) that hybridize completely. A list of all parameters is provided in Table 2. The expression of $\mathrm{Cot}_p(n)$ in the case of second-order kinetics can be derived either from Eq. 10 by setting $\xi_1 = 1$ and $\xi_2 = 1$, or by finding the value s^* for which $A^{\mathrm{SOK}}(s^*) = p$ in Eq. 6:

$$\mathrm{Cot}_p^{\mathrm{SOK}}(n) = \frac{1}{2a\alpha}\left(\frac{\alpha}{1-p}-1\right)n. \tag{11}$$

Importantly, the latter expression is linear in n, whereas the generalized Cot expression (Eq. 10) is a rational (nonlinear) function of n.

To illustrate the difference between the dynamics of second-order kinetics and the heteroduplex model, we plotted the annealing kinetics of both models for one set of parameter values and three diversities (Fig. 3 *A*). Although

TABLE 2 Parameters, their meaning, and typical ranges

Parameter	Meaning	Range	Typical value
a	Association rate of two single DNA strands	>0	—
d_1	Dissociation rate of a partially hybridized homoduplex	>0	—
d_2	Dissociation rate of a partially hybridized heteroduplex	>0	—
z_1	Hybridization rate of a homoduplex	>0	—
z_2	Hybridization rate of a heteroduplex	>0	—
ξ_1	Composite parameter $= z_1/(z_1 + d_1)$	[0,1]	Close to 1
ξ_{12}	Composite parameter $= z_2/(z_2 + d_2)$	[0,1]	Close to 0
α	Proportion of melted molecules at start of annealing	[0,1]	>0.5
ϕ	Weight factor for the fluorescence of heteroduplexes	[0,1]	>0.5
n	Diversity	>0	—
p	Annealing proportion	[0,1]	—
T	Total concentration of single DNA strands in a sample	>0	—
A_b	Baseline fluorescence of a sample	>0	—
R_b	Baseline fluorescence of a reference	>0	—

Some parameters can differ in each experiment; in that case, typical values are not provided.

the diversity increases 10-fold between the curves ($n = 10$, $n = 100$, $n = 1000$), the annealing speed (reflected in the Cot 50% value) in the case of the heteroduplex model does not decrease 10-fold, as it does under second-order kinetics, because the function $\mathrm{Cot}_{0.5}(n)$ of Eq. 10 exhibits a concave (saturating) shape (Fig. 3 *B*). Note that the discrepancies between both Cot curves are small for low diversities, but the deviation from linearity becomes more apparent as diversity increases. Indeed, the higher the diversity, the more heteroduplexes are expected to be formed. Note that for $n = 1$, the heteroduplex model ($\mathrm{Cot}_p(n)$, Fig. 3 *B*) reveals a slightly higher Cot value even though heteroduplexes cannot be formed. This is due to the formation of temporary complexes (C_{ii}) in this model that delay the annealing process.

Annealing time-series data: heteroduplex model fits significantly better than second-order kinetics

To compare the validity of both models, we fitted Eqs. 5 and 6 to the time-series of data sets 1–3 (Table 1). The fits of both models to the annealing curves are depicted in Fig. 4 where, due to space limitation, only three diversities per data set are presented. The fits to the full data sets can be found in Section S1 in the Supporting Material; the corresponding best-fitting parameters and their confidence intervals are given in Section S2 in the Supporting Material. Note that the horizontal axes of the annealing curves are given in time units. We corrected for concentration differences in the data by adjusting the DNA association rate a in the model (a was multiplied by T, the estimated total ssDNA concentration in a sample, which, under the quasi-steady-state assumption, is equivalent to using a Cot scale in the data).

Visual inspection of these fits revealed a small difference between the performance of both models on the data set 1 (Fig. 4 *A*). The models clearly differ in fitting the time course of data set 2 (Fig. 4 *B*), where second-order kinetics was unable to reproduce the correct curvature and the apparent asymptotic value of the data, especially for high diversities. Similarly, second-order kinetics failed to give the correct asymptote value in the fit of data set 3 (Fig. 4 *C*). A statistical analysis, accounting for the different number of parameters in each model (likelihood ratio test for nested models, based on the χ-square distribution (18)), indicated that the improvement brought by the heteroduplex model was significant for all three data sets (p-value $< 10^{-3}$).

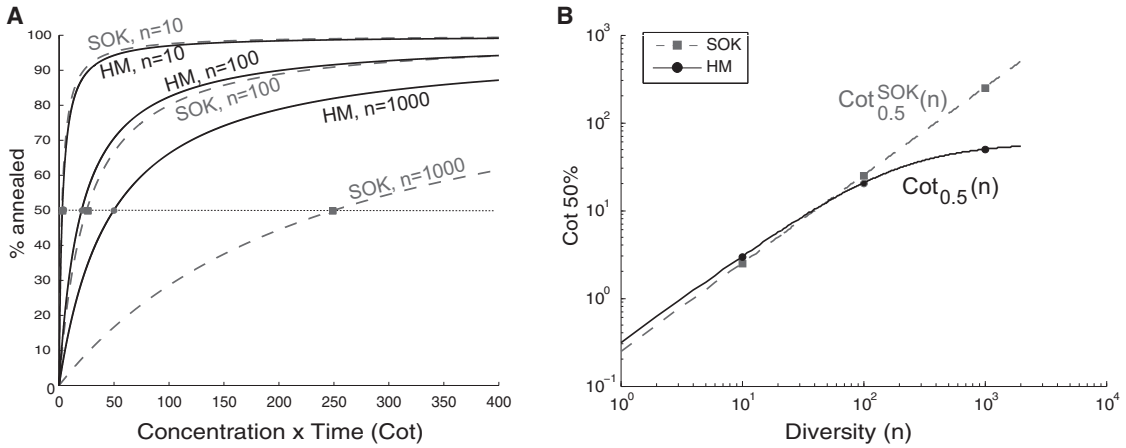

FIGURE 3 Second-order kinetics (SOK) and the heteroduplex model (HM) exhibit different annealing kinetics and diversity-Cot relations. (*A*) Annealing kinetics as a function of Cot values for three diversities ($n = 10$, $n = 100$, $n = 1000$) and both models (SOK, *dashed*; HM, *solid*). The chosen parameter values are similar to the best-fitting parameters for data set 1 (see Fig. 4): $a = 2$, $\xi_1 = 0.8$, $\xi_2 = 0.009$, $\varphi = 0.97$, $a = 1$, and $T = 1$. (*B*) Cot 50% values were computed using Eq. 10 (*solid*) and Eq. 11 (*dashed*) and were plotted as a function of diversity for the same parameter values as in panel *A*. The relation between diversity and Cot values is linear under second-order kinetics, whereas the heteroduplex model reveals a saturating $\mathrm{Cot}_{0.5}(n)$ relation. Note that for $n = 1$, both models reveal slightly different Cot values even though heteroduplexes cannot be formed. This is due to transient duplex formation.

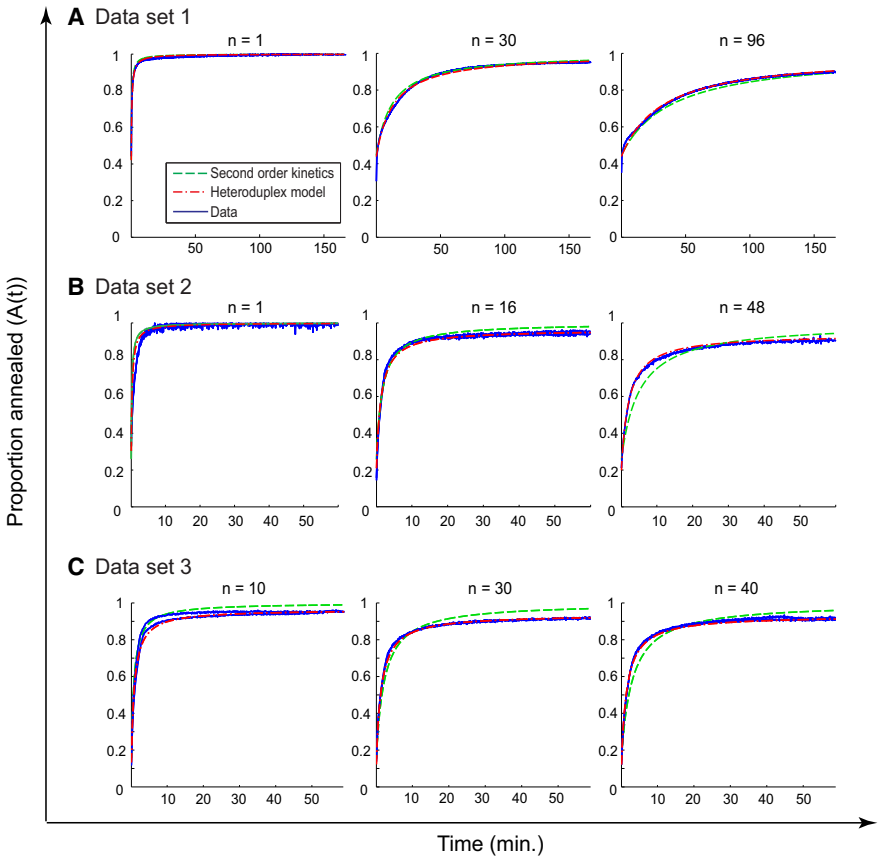

FIGURE 4 Best-fits of data sets 1–3 (*A–C*) for known diversity templates (with only the lowest, an intermediate, and the highest diversities of each data set shown; all other diversities are given in Section S1 in the Supporting Material. (*Solid blue*) Data sample (one or two replicates). (*Dashed green*) Best-fit of the second-order kinetics model (Eq. 6). (*Dash-dotted red*) Best-fit of the heteroduplex model (Eq. 5). For the best-fitting parameters and their confidence intervals, see Table S1 in the Supporting Material. The heteroduplex model results in a significantly better fit to the data than the second-order kinetics model (p-value $<10^{-3}$ for all three data sets).

Cot values as function of diversity: heteroduplex model captures nonlinear relationship, second-order kinetics does not

Although the heteroduplex model gave a significantly better fit to all three AmpliCot time-series data, in some cases the visual difference between the fit of the second-order kinetics and the heteroduplex model was not very large. Small differences in the fit to the full annealing curve may, however, lead to large differences in the estimated Cot value, especially for higher Cot-values that fall in the saturating part of the annealing curve. We therefore investigated the relationship between Cot values and the diversity n under second-order kinetics and the heteroduplex model. We used data sets 1–3 to estimate Cot 50% and Cot 80% values ($\mathrm{Cot}_p{}^{\mathrm{data}}(n)$, $p = \{0.5, 0.8\}$), which we plotted against the diversity n (Fig. 5). For comparison, we also computed $\mathrm{Cot}_p(n)$ and $\mathrm{Cot}_p{}^{\mathrm{SOK}}(n)$, $p = \{0.5, 0.8\}$, using Eqs. 10 and 11, given the best-fitting parameters to each full data set (see Table S1 in the Supporting Material), described in the previous section.

Interestingly, Cot values that were directly estimated from the experimental data ($\mathrm{Cot}_p{}^{\mathrm{data}}$) presented a clear deviation from linearity (for all data sets) and exhibited a concave shape, similar to the one predicted by the heteroduplex model. As a result, the $\mathrm{Cot}_p{}^{\mathrm{data}}(n)$ curves were in general better captured by the heteroduplex model ($\mathrm{Cot}_p(n)$) than by second-order kinetics ($\mathrm{Cot}_p{}^{\mathrm{SOK}}(n)$). The only exception is the description of the Cot 50% values of data set 1, which is poor for both models (Fig. 5 *A*), because the Cot 50% value could hardly be read-out for this data set. The results of Fig. 5 suggest that, in general, Cot analyses based on the generalized $\mathrm{Cot}_p(n)$ expression (Eq. 10) yield more-accurate diversity estimates than those based on second-order kinetics ($\mathrm{Cot}_p{}^{\mathrm{SOK}}(n)$, Eq. 11).

Heteroduplex model also captures nonlinear trend of highly diverse samples

Driven by our finding that Cot values are better described with the heteroduplex model, we assessed our new formula for $\mathrm{Cot}_p(n)$ by fitting it directly to the diversity-Cot relationships of data sets 1, 2, and 4, without first fitting the annealing time-series data. We omitted data set 3 because it contains too few different diversities to fit the five parameters of the generalized Cot expression. On the contrary, the recently published data set 4 (14) contains diversities that differ by several orders of magnitude and is thus particularly well suited for testing our new formula.

In Fig. 6 are depicted the fits of Eqs. 10 and 11 to the Cot 50% and Cot 80% values of the different data sets. The annealing duration in data set 4 was too short to

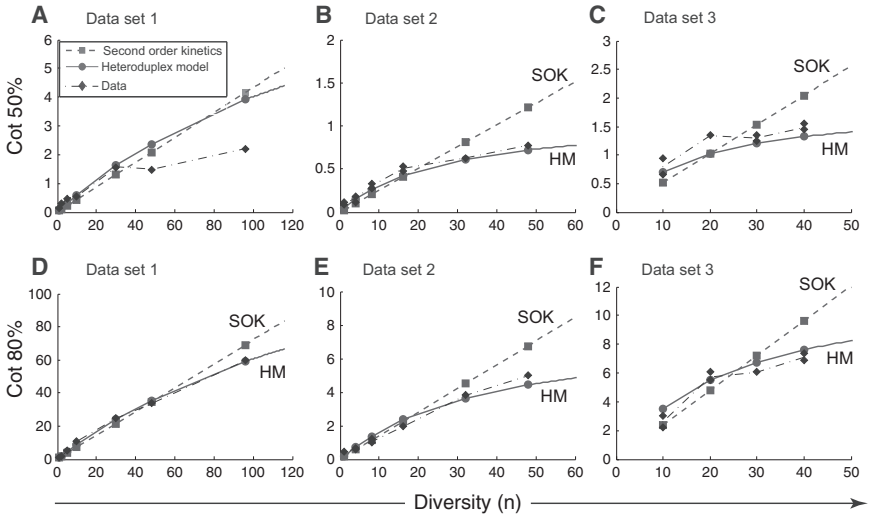

FIGURE 5 Generalized (nonlinear) $Cot_p(n)$ expression reproduces Cot values of the experimental data better than second-order kinetics. The behavior of Cot 50% (*A–C*) and Cot 80% (*D–F*) as a function of diversity was computed under both models: second-order kinetics (■, Eq. 11) and the heteroduplex model (●, Eq. 10). The best-fitting parameters of the time-series fits (see Table S1) were used. Cot 50% and 80% of the experimental data are also plotted (◆). The difference between both models is amplified as diversity increases. Connecting lines are shown to help the visualization of the trend.

compute Cot 80% values, so we used Cot 70% values instead. Note that the experimental data in panels *A*, *B*, *D*, and *E* of Fig. 6 are the same as the data in the corresponding panels of Fig. 5. Similarly to data sets 1–3, Cot values of data set 4 revealed a clear deviation from linearity at high diversities. Such deviation is clearly observed in all data sets and is well captured by the Cot expression based on the heteroduplex model (Eq. 10), in contrast to the Cot expression based on second-order kinetics (Eq. 11). Note that a convex shape was observed for Cot 50% values of data set 4, whereas Cot 70% and Cot 80% values exhibited a concave curvature. Both were well captured by the generalized Cot expression (Fig. 6, *C* and *F*). Indeed, Eq. 10 is a rational function of diversity and hence allows the reproduction of both convex and concave shapes. These correspond, respectively, to both asymptote-bounded arms of the function.

To test whether the heteroduplex model (Eq. 10) fits the observed Cot values significantly better than second-order kinetics (Eq. 11), we applied a likelihood ratio test for nested models (18). Statistical significance was reached for all fits (see *p*-values in *upper-left corner* of each panel of Fig. 6). Hence, in addition to better describing time-series annealing data, the generalized Cot expression based on the heteroduplex model is also better at fitting Cot values directly, especially for highly diverse samples, such as those of data set 4.

Generalized Cot analysis: diversity estimation procedure

We formally define here an alternative to the original Cot analysis for the interpretation of AmpliCot experimental data. Our method allows the estimation of an unknown

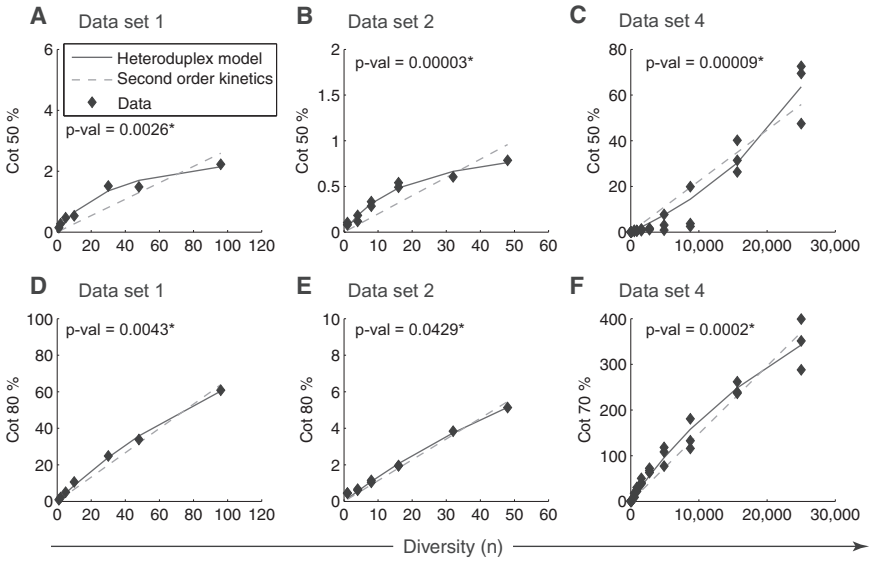

FIGURE 6 Generalized (nonlinear) $Cot_p(n)$ expression (Eq. 10) reproduces the diversity-Cot 50% and 80% relationships of data sets 1, 2, and 4 (*A–F*) better than the $Cot^{SOK}{}_p(n)$ expression (Eq. 11). The Cot expression based on the heteroduplex model (*solid line*) and the Cot expression based on second-order kinetics (*dashed line*) were fitted to Cot 50% or Cot 80% values (*diamonds*), without calibrating the model on time-series annealing data. The highest Cot value assessable in data set 4 was Cot 70%. The best-fitting parameters to the data can be found in Table S2 and Section S3 in the Supporting Material. The *p*-values of a likelihood ratio test for nested models are indicated in each panel. The fit of the generalized Cot expression was considered significantly better than the fit of $Cot^{SOK}{}_p$ at level 95% when the *p*-value was <0.05 (indicated by *).

diversity from a library of known diversities and provides a more general alternative to the original method (10). The method consists of four steps. We first suggest to use not only one, but several annealing proportions for better calibration. Second, the raw annealing data of the templates with known diversity are normalized to estimate the Cot_p values necessary for the calibration of the generalized Cot expression. Third, the parameter values of Eq. 10 are determined by fitting this equation to Cot values of the data (for all predetermined annealing proportions). Finally, the unknown diversity is estimated using the inverse of the calibrated $Cot_p(n)$ relation and the measured Cot value of the sample to assess. The algorithm of our diversity estimation procedure is given below.

Diversity estimation algorithm

1. Choose an appropriate set of values of p (annealing proportions).
2. Normalize the raw data using Eq. 7 and estimate the Cot_p values of the templates with known diversity.
3. Fit the parameters (a, α, ξ_1, ξ_2, φ) of the generalized Cot expression (Eq. 10) to the Cot data of the templates with known diversity.
4. Using the Cot value of the sample with unknown diversity, estimate its diversity from the generalized Cot curve fitted above.

DISCUSSION

A framework for better understanding and analysis of AmpliCot data

By means of mathematical modeling, we developed a general framework for the understanding and interpretation of AmpliCot data. We showed that the initially assumed underlying model, second-order kinetics (10), might not always be the best way to describe DNA annealing kinetics. This was revealed by the model-fit of annealing time-series data and by the deviation from linearity of the Cot-diversity relation. We developed an alternative, the heteroduplex model, which describes the underlying biochemical reaction in further detail and reproduces the nonlinear nature of Cot values as a function of diversity.

In the original AmpliCot article, the authors assumed a linear relation between Cot values and calibrating diversities (10). We showed that this linear relation is indeed correct under second-order kinetics, i.e., in the absence of heteroduplexes and temporary duplexes (Eq. 11). However, under the heteroduplex model, the generalized $Cot_p(n)$ expression is not linear. Indeed, Eq. 10 is a rational function of n. Intuitively, the possibility of formation of partially fluorescent heteroduplexes results in a faster annealing for a given diversity and concentration. Instead of only binding to perfectly matching strands, some DNA molecules may associate to partially complementary molecules. The resulting heteroduplexes still contribute to the observed fluorescence but to a lesser extent, as we showed experimentally (Fig. 2).

The presence of heteroduplexes with lower fluorescence levels can also explain the observation of Schütze et al. (13), who noted that reannealed samples did not reach their preanneal fluorescence intensity, even after correction for the fluorescence decline due to dye degradation. We also considered two alternative explanations for this phenomenon. The first hypothesis is that no heteroduplexes are formed, but homoduplexes may constantly associate and dissociate because the annealing temperature is very close to the melting temperature. We fitted such a model to the data and although it accounted for the above-mentioned loss of fluorescence, it did not explain the early time-course of the annealing curves (results not shown) and it yielded a significantly lower quality of fit compared to the heteroduplex model. The second alternative explanation that we tested is that the intensity of the SYBR green dye is diminished after melting. However, this explanation did not account for the observed dependence on diversity of the fluorescence loss. The heteroduplex formation leading to a lower SYBR green signal was therefore the most likely explanation.

Generic and easy-to-use diversity estimation procedure

We propose what to our knowledge is a novel procedure allowing for the estimation of a sample's diversity from a library of known calibration diversities. Our procedure is based on the result that the heteroduplex model is the one that best describes AmpliCot data. The advantage of our new method over the second-order kinetics-based approach is that it encompasses both underlying models. Indeed, the Cot expression of Eq. 10 is a generalization of the expression based on second-order kinetics. Therefore, it can be applied both to samples that exhibit few or no heteroduplexes (10,14), as well as to samples in which heteroduplex formation is suspected (13). The data themselves will determine the degree of deviation from linearity (if any) of the diversity-Cot relation. Our new method is simple to use, as it requires the manipulation of one single formula (Eq. 10). It is also computationally efficient (complexity similar to the one of the second-order kinetics-based method), because it is directly calibrated on Cot values.

Limitations

When using our diversity estimation procedure, one should be careful in the parameter calibration step based on Cot values. Being a rational function of diversity, the generalized Cot expression (Eq. 10) has one vertical and one horizontal asymptote. When extrapolating unknown diversities

that are expected to be very different from the calibration set, one should be aware that the horizontal asymptote may render the estimation impossible. For example, this could happen if the calibrated parameters result in an asymptote below the Cot value of the sample with unknown diversity. To circumvent such problems, one could alternatively use the time-series data to calibrate parameters of the heteroduplex model in step 3 of the estimation procedure. The larger amount of information contained in time-series data is expected to result in more robust parameter estimates, and may reduce the number of calibrating diversities that are needed to make a sound estimation.

Applications

The correct calibration of AmpliCot is crucial for the estimation of an unknown diversity. If one uses a linear approximation by assuming second-order kinetics, the diversity estimation may be biased, as revealed by our nonlinear fits of the heteroduplex model to experimental data. In their recent articles, Baum et al. (11,14) proposed a novel method for estimating the absolute number of unique TCRβ chain rearrangements in a blood sample. AmpliCot is part of this integrated method and the assay was used to estimate the absolute diversity of several independent VβJβ pairs of CD4+ naive T cells. The overall procedure resulted in highly reproducible estimates, but the authors consistently reported lower diversities than expected. Instead of the anticipated 100,000 or 200,000 cells with unique TCR sequences, the authors measured approximately twofold lower diversities (see Fig. 5 of Baum et al. (11)). The authors suggested several reasons for this discrepancy: the potential existence of expanded clones, the phenotype reversion of atypical memory cells, and the higher probability of occurrence of some TCR rearrangements (11), which all seem entirely plausible. Our analysis of the calibration set published by Baum et al. (11) (data set 4) revealed a clear deviation from linearity (Fig. 6, *C* and *F*) that, in the case of Cot 50%, could be another reason for the underestimation of the true diversity.

CONCLUSION

In summary, we show that deviations from linearity are well represented by the heteroduplex model. The use of a linear model could lead to under- or overestimation of unknown diversities, which could be improved by the use of the heteroduplex model.

APPENDIX 1: MEAN-FIELD MODELS

The mean-field models take advantage of the assumption that all DNA strands in the sample are present in equal concentration in the sample. This equimolarity assumption allows us to reduce the dimension of the ordinary differential equations and to render them independent of diversity.

Mean-field second-order kinetics

Let

$$S(t) = \sum_{i=1}^{n} S_i(t) \text{ and } D(t) = \sum_{i=1}^{n} D_{ii}(t) .$$

If all DNA strands are present in equimolar concentrations in the mixture, we have $S_i(t) = S_j(t)$, $\forall t$. Therefore, $S(t) = nS_1(t)$, $D(t) = nD_{11}(t)$ and the ODE system of Eq. 1 becomes

$$\begin{aligned} \frac{dS}{dt} &= -2a\frac{S^2}{n}, \\ \frac{dD}{dt} &= a\frac{S^2}{n}, \end{aligned} \tag{12}$$

where we have used the fact that $nS_1{}^2 = S^2/n$. The initial conditions are

$$S(0) = \alpha T \text{ and } 2D(0) = (1-\alpha)T,$$

and the fluorescent molecules are

$$F(t) = 2D(t).$$

Mean-field heteroduplex model

Assuming equimolar concentrations of each species, we define the quantities

$$\begin{aligned} S(t) &= \sum_{i=1}^{n} S_i(t) = nS_1(t), \\ C(t) &= \sum_{i=1}^{n} C_{ii}(t) = nC_{11}(t), \\ H(t) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} C_{ij}(t) = \frac{n(n-1)}{2} C_{12}(t), \\ J(t) &= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} D_{ij}(t) = \frac{n(n-1)}{2} D_{12}(t), \\ D(t) &= \sum_{i=1}^{n} D_{ii}(t) = nD_{11}(t), \end{aligned} \tag{13}$$

where indices 1 and 2 have been chosen arbitrarity to design one species. $S(t)$ denotes ssDNA, $C(t)$ partially hybridized homoduplexes, $H(t)$ partially hybridized heteroduplexes, $J(t)$ final product heteroduplexes, and $D(t)$ the final homoduplexes. The differential equations of Eq. 2 can be written in terms of the above variables as

$$\begin{aligned} \frac{dS}{dt} &= -a(n+1)\frac{S^2}{n} + 2d_1C + 2d_2H, \\ \frac{dC}{dt} &= a\frac{S^2}{n} - (d_1 + z_1)C, \\ \frac{dH}{dt} &= a\left(\frac{n-1}{2}\right)\frac{S^2}{n} - (d_2 + z_2)H, \\ \frac{dJ}{dt} &= z_2H, \\ \frac{dD}{dt} &= z_1C, \end{aligned} \tag{14}$$

with initial conditions

$$S(0) = \alpha T,$$

$$2D(0) = (1-\alpha)T,$$

$$C(0) = H(0) = J(0),$$

and fluorescent molecules

$$F(t) = 2(D(t) + \varphi J(t).$$

APPENDIX 2: MODEL SOLUTION

Second-order kinetics

The solution of the ODE system of Eq. 12 with initial conditions

$$S(0) = \alpha T \text{ and } 2D(0) = (1-\alpha)T$$

is

$$S(t) = \frac{\alpha T}{1 + 2\frac{a}{n}\alpha T t},$$

$$D(t) = \frac{1}{2}\left(1 - \frac{\alpha}{1 + 2\frac{a}{n}\alpha T t}\right)T.$$

Heteroduplex model

We present here an analytical solution of the heteroduplex model of Eq. 14 under a quasi-steady-state condition. If the association/dissociation rates a,d_1,d_2 of transient complexes $C(t)$ and $H(t)$ are large compared to the final duplex formation rates z_1 and z_2, we can assume that the transient complexes quickly reach a steady state. By setting their corresponding time-derivatives to 0, we get

$$C = \frac{a}{n(d_1 + z_1)}S^2, \tag{15}$$

$$H = \frac{a}{n(d_2 + z_2)}\left(\frac{n-1}{2}\right)S^2. \tag{16}$$

By inserting Eqs. 15 and 16 into the initial system (Eq. 14), we get

$$\begin{aligned}\frac{dS}{dt} &= -2K(n)S^2,\\ \frac{dJ}{dt} &= \frac{a}{n}\left(\frac{z_2}{d_2 + z_2}\right)\left(\frac{n-1}{2}\right)S^2,\\ \frac{dD}{dt} &= \frac{a}{n}\left(\frac{z_1}{d_1 + z_1}\right)S^2,\end{aligned} \tag{17}$$

where

$$K(n) = \frac{a}{n}\left(\frac{z_2}{d_1 + z_1} + \frac{z_2}{d_2 + z_2}\left(\frac{n-1}{2}\right)\right). \tag{18}$$

By using the initial conditions, we obtain the following solution of Eq. 17:

$$\begin{aligned}S(t) &= \frac{\alpha T}{1 + 2K(n)\alpha T t},\\ J(t) &= \frac{a}{n}\left(\frac{z_2}{d_2 + z_2}\right)\left(\frac{n-1}{2}\right)\frac{1}{2K(n)}(\alpha T - S(t)),\\ D(t) &= \frac{a}{n}\left(\frac{z_1}{d_1 + z_1}\right)\frac{1}{2K(n)}(\alpha T - S(t)) + \left(\frac{1-\alpha}{2}\right)T.\end{aligned} \tag{19}$$

SUPPORTING MATERIAL

Three figures and two tables and the MATLAB code allowing to fit all types of AmpliCot data are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(12)00792-8.

The authors thank Prof. Paul Baum and colleagues for allowing the use of their data.

This work was financially supported by the Netherlands Organization for Scientific Research (grants No. 836-07-002 and 917-96-350).

REFERENCES

1. Goldsby, R., T. Kindt, …, J. Kuby. 2003. Immunology. W. H. Freeman and Company, New York.
2. Robins, H. S., P. V. Campregher, …, C. S. Carlson. 2009. Comprehensive assessment of T-cell receptor β-chain diversity in $\alpha\beta$ T cells. *Blood.* 114:4099–4107.
3. Pannetier, C., M. Cochet, …, P. Kourilsky. 1993. The sizes of the CDR3 hypervariable regions of the murine T-cell receptor β-chains vary as a function of the recombined germ-line segments. *Proc. Natl. Acad. Sci. USA.* 90:4319–4323.
4. Currier, J. R., and M. A. Robinson. 2001. Spectratype/immunoscope analysis of the expressed TCR repertoire. *In* Current Protocols in Immunology. Wiley, New York Chapter 10, Unit 10.28.
5. Shendure, J., and H. Ji. 2008. Next-generation DNA sequencing. *Nat. Biotechnol.* 26:1135–1145.
6. Mardis, E. R. 2008. Next-generation DNA sequencing methods. *Annu. Rev. Genomics Hum. Genet.* 9:387–402.
7. Boyd, S., E. Marshal, …, A. Z. Fire. 2009. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel V-D-J pyrosequencing. *Sci. Transl. Med.* 1:12ra23. http://dx.doi.org/10.1126/scitranslmed.3000540.
8. Freeman, J. D., R. L. Warren, …, R. A. Holt. 2009. Profiling the T-cell receptor β-chain repertoire by massively parallel sequencing. *Genome Res.* 19:1817–1824.
9. Wang, C., C. M. Sanders, …, J. Han. 2010. High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc. Natl. Acad. Sci. USA.* 107:1518–1523.
10. Baum, P. D., and J. M. McCune. 2006. Direct measurement of T-cell receptor repertoire diversity with AmpliCot. *Nat. Methods.* 3:895–901.
11. Baum, P. D., J. J. Young, and J. M. McCune. 2011. Measurement of absolute T cell receptor rearrangement diversity. *J. Immunol. Methods.* 368:45–53.

12. Britten, R. J., and D. E. Kohne. 1968. Repeated sequences in DNA. Hundreds of thousands of copies of DNA sequences have been incorporated into the genomes of higher organisms. *Science.* 161:529–540.

13. Schütze, T., P. F. Arndt, …, J. Glökler. 2010. A calibrated diversity assay for nucleic acid libraries using DiStRO—a diversity standard of random oligonucleotides. *Nucleic Acids Res.* 38:e23.

14. Baum, P. D., J. J. Young, …, J. M. McCune. 2011. Design, construction, and validation of a modular library of sequence diversity standards for polymerase chain reaction. *Anal. Biochem.* 411:106–115.

15. Ishii, K., and M. Fukui. 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. *Appl. Environ. Microbiol.* 67:3753–3755.

16. Sipos, R., A. J. Székely, …, M. Nikolausz. 2007. Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targeting bacterial community analysis. *FEMS Microbiol. Ecol.* 60:341–350.

17. Wetmur, J. G., and N. Davidson. 1968. Kinetics of renaturation of DNA. *J. Mol. Biol.* 31:349–370.

18. Le Boudec, J.-Y. 2010. Performance Evaluation of Computer and Communication Systems. EPFL Press, Lausanne, Switzerland.

19. Schneeberger, C., P. Speiser, …, R. Zeillinger. 1995. Quantitative detection of reverse transcriptase-PCR products by means of a novel and sensitive DNA stain. *PCR Methods Appl.* 4:234–238.

20. Colborn, J. M., B. D. Byrd, …, D. J. Krogstad. 2008. Estimation of copy number using SYBR Green: confounding by AT-rich DNA and by variation in amplicon length. *Am. J. Trop. Med. Hyg.* 79:887–892.