The Fidelity of the Tag-Antitag System III

Robustness in the excess limit: the stringent temperature

John A. Rose[1]

Department of Computer Science, The University of Tokyo, and
Japan Science and Technology Corporation, CREST
johnrose@is.s.u-tokyo.ac.jp

Abstract. The importance of DNA microarrays and Tag-Antitag (TAT) systems has prompted the recent development of various approaches for high-fidelity design, including analytical methods based on an ensemble average error probability per conformation, or *computational incoherence* (ϵ). Although predictions for dilute inputs indicate the easy attainment of excellent fidelity, recently a sharp phase transition from the low-error ϵ-behavior predicted for dilute inputs, to a high-error ϵ-behavior was predicted to accompany an asymmetric (*i.e.*, single-tag) excess input. This error-response, which is likely to be the critical test of TAT system robustness for DNA-based computing applications that employ non-linear amplification, is examined more closely, via derivation of an approximate expression, $\epsilon_{e(i)}$ for the single-tag, excess limit. The temperature-dependence of this expression is then characterized, and applied to derive an expression for a novel TAT system error-parameter, $T_i^\dagger$ which defines the temperature of minimal $\epsilon_{e(i)}$. $T_i^\dagger$ is taken to provide a precise definition of the stringent reaction temperature previously discussed conceptually in the literature. A similar analysis, undertaken for a uniform excess multi-tag input, indicates the absence of a phase transition in ϵ. The validity of each expression is discussed via simulation, with comparison to the general model. Applicability of $\{T_i^\dagger\}$ to both TAT system design and selection of an optimal reaction temperature is discussed.

1 Introduction

DNA microarrays are indexed arrays of single-stranded (ss) DNA probes which are immobilized on a solid substrate. When exposed to a set of unbound target ssDNA strands, the chip essentially performs an exhaustive, parallel search for complementary sequences between the immobilized probes and target species. DNA chips have been successfully applied to gene expression profiling (GEP) and genotyping on a genomic scale [1], and have also been suggested for applications in DNA computing [2], and DNA computing-based biotechnology [3]. Notably, design for computational application simplifies word selection, since the ssDNA species need not be correlated to a genome of interest, but may be selected arbitrarily. The resulting set forms a Tag-Antitag system [4], constrained only by

the requirement that each anchored probe, or 'antitag' species be the Watson-Crick complement of exactly one corresponding target, or 'tag' species. The anchored component is often referred to as a 'universal' DNA chip.

Although a number of design goals exist for TAT systems (*e.g.*, uniform energetics, minimal folding, maximal specific-affinity, etc. [5]), the design for maximal specific affinity is the most challenging, due to the highly-coupled nature of the hybridizing system. Various heuristic methods for TAT system design have been proposed [4, 6, 7, 2, 8] for this purpose. In addition, a statistical thermodynamic approach for TAT system error analysis and design has also been reported [9, 5], which is attractive due to physical motivation, the availability of energetic parameters, and the generation of a quantitative, well-defined measure of system performance. Fundamental to this approach is the modelling of system error in terms of an ensemble average probability of error hybridization per conformation (the *computational incoherence*), so that the inverse problem of system design is equated to the process of measure minimization. Thus far, results include approximate expressions for dilute single-tag and multi-tag inputs [9], and general single-tag inputs[5]. Note that an equilibrium approach has also been applied to investigate the fidelity of DNA-protein interactions [10], nucleic acid-based antisense agents [11], and via the computational incoherence, the annealing [12] and annealing-ligation biosteps of DNA-based computing [13].

1.1 Recent Work and Motivation

In [5], an approximate solution for the computational incoherence of the TAT system in response to a single-tag input, ϵ_i was derived for the error-response over a wide range of input tag concentrations. For all error conditions, the simulated dependence of ϵ_i on total input tag concentration (C_i^o) indicated a sharp phase-transition between high-error and low-error operation, in the vicinity of an equimolar input ($C_i^o = C_a^o$, the total concentration of each antitag species), for temperatures, T beneath the melting transition of the planned TAT species. In particular, TAT system fidelity was predicted to abruptly transition between: (1) a monotonically increasing function of T (dilute inputs; $C_i^o \ll C_a^o$), characterized by low-error operation; and (2) a convex function of T (excess inputs; $C_i^o \gg C_a^o$), characterized by an error minimum at temperature, $T_i^\dagger$, with exponentially-increasing ϵ_i away from this temperature. Intuitively, this transition signals saturation of the target antitag, i^* which naturally accompanies an excess single-tag input, beneath the melting transition of the target TAT duplex.

For simple, 1-step TAT system applications in biotechnology, dilute conditions may generally be safely assumed. Given the ease of attaining high-fidelity performance at low temperatures, predicted in the dilute regime [9, 5], the biasing of DNA-computers to ensure dilute regime operation of an associated TAT system component is clearly desirable. However, given the tendency for DNA computing architectures to implement repeated linear strand growth, via *merge* operations, as well as species-specific, non-linear strand growth via PCR amplification, over the course of multiple steps/rounds [8], there appears to be a strong

potential for computational processes to generate an *asymmetric input*, consisting of a dilute component combined with an excess component of one (or more) input species. In this case, consistent high-fidelity operation at low temperatures is predicted to become substantially more problematic, for even the best encodings (see Sec. 4). For these architectures, consideration of the associated TAT system's $|i|$ single-input, excess-error response curves yields valuable information for selection of a reaction temperature, T_{rx} appropriately robust to a range of asymmetric excess-inputs. To support this analysis, the current work undertakes a closer examination of the single-tag error behavior in the excess regime, with the aim of identifying a design principle which renders the implemented TAT systems maximally robust to asymmetric, excess inputs.

Following an overview of the general model for predicting TAT system single-input, error-response (Sec. 2), an approximate expression is derived in Sec. 2.1 for ϵ_i in the limit of excess input ($\epsilon_{e(i)}$). Sec. 2.2 then discusses the temperature-dependence of this expression, followed in Sec. 2.3 by identification of a new TAT system parameter, $T_i^\dagger$, which estimates the temperature of optimal fidelity, given an excess input of tag species, i. $T^\dagger$ is taken to provide a novel, precise definition of the *stringent* T_{rx} previously discussed conceptually in the literature for the TAT system [4]. For completeness, Sec. 3 describes a parallel analysis for the uniformly excess input. Sec. 4 reports a set of statistical thermodynamic simulations undertaken to explore the validity and implications of derived expressions for $\epsilon_{e(i)}$ and $T_i^\dagger$. In closure, Sec. 5 discusses applicability to TAT system design.

2 The Single-tag Input

The error probability per hybridized tag for a TAT system, in response to an input of a single tag species, i is estimated by the expression [5],

$$\epsilon_i = \frac{\sum_{j^* \neq i^*} C_{ij^*}}{\sum_{j^*} C_{ij^*}} = (1 + SNR_i)^{-1}. \tag{1}$$

Here, SNR_i is the signal-to-noise ratio,

$$SNR_i = \frac{C_{i^*} K_{ii^*}}{\sum_{j^* \neq i^*} C_{j^*} K_{ij^*}}, \tag{2}$$

where K_{ij^*} denotes the net equilibrium constant of duplex formation between tag species i and antitag species j^*, while K_{ii^*} distinguishes that of matching TAT pair, $\{i, i^*\}$. For approximation purposes, it is typical to assume a small overall error-rate, so that (*i.e.*, $\sum_{j^*} C_{ij^*} \approx C_{ii^*}$)[9, 5]. At equilibrium, this condition, here referred to as *weak orthogonality* takes the convenient form, $\sum_{j^*} C_{j^*} K_{ij^*} \approx C_{i^*} K_{ii^*}$. Although this approximation will begin to fail for an excess (but not dilute) input, as T_{rx} is reduced to the vicinity of the melting temperature of the most stable error TAT pair, it nevertheless facilitates an investigation of the approximate functional form of ϵ_i. Furthermore, upon failure, this approximation

will overestimate ϵ_i, and thus provide a bounding value, which, as simulations indicate, is not too far off the mark [5].

Following application of weak orthogonality, approximate solution of Eq. 1 involves re-expression of C_{j^*} in terms of equilibrium constants and initial concentrations, via combination of the $|j^*|+1$ equations of stand conservation, with an equation of mass action for each component equilibrium. In particular, strand conservation yields an equation of the form,

$$C_a = C_{j^*}(1 + K_{j^*}^{hp} + C_i K_{ij^*}), \tag{3}$$

for each antitag species, j^*, and equation,

$$C_i^o = C_i(1 + K_i^{hp} + \sum_{j^*} C_j^* K_{ij^*}), \tag{4}$$

for the single input tag species, i, where the impact of tag-tag interaction has been neglected. Strict estimation of C_{i^*} then proceeds via numerical solution of the $|j^*| + 1$ coupled equations formed by Eqs. 3 and 4. In [5], an approximate approach was used to derive a general solution for ϵ_i, applicable over a wide range of input concentrations. Readers are referred to the original paper, for a detailed development and discussion. In the current work, attention is restricted to a more detailed analysis of TAT system behavior in the excess-input limit.

2.1 Behavior in the Excess Limit

A simple, approximate expression for the single-tag error-response, ϵ_i in the limit of excess input (*i.e.*, $C_i^o \geq 10C_a$) may be derived straightforwardly, by noting that the impact of hybridization on the equilibrium concentration of the input tag species, C_i may be neglected. This allows Eq. 4 to be approximated as $C_i \approx C_i^o(1 + K_i^{hp})^{-1}$. Substitution of this expression into Eq. 3 yields,

$$C_{j^*} \approx \frac{C_a(1 + K_i^{hp})}{(1 + K_i^{hp})(1 + K_{j^*}^{hp}) + C_i^o K_{ij^*}}, \forall\, j^*. \tag{5}$$

Invoking weak orthogonality, followed by insertion of these expressions reduces Eq. 1 to the desired approximate form,

$$\epsilon_{e(i)} \approx \sum_{j^* \neq i^*} \frac{K_{ij^*}[(1 + K_i^{hp})(1 + K_{i^*}^{hp}) + C_i^o K_{ii^*}]}{K_{ii^*}[(1 + K_i^{hp})(1 + K_{j^*}^{hp}) + C_i^o K_{ij^*}]}, \tag{6}$$

which applies to the case of excess input. In the absence of significant hairpin formation, this reduces to a simple ratio,

$$\epsilon_{e(i)} \approx \sum_{j^* \neq i^*} \frac{C_i^o + K_{ii^*}^{-1}}{C_i^o + K_{ij^*}^{-1}}. \tag{7}$$

For comparison purposes, the approximate expression for the converse limit of dilute input, without hairpinning was given in [5] as the simple ratio,

$$\epsilon_{d(i)} \approx \sum_{j^* \neq i^*} \frac{K_{ij^*}}{K_{ii^*}}. \tag{8}$$

2.2 Temperature Dependence

The temperature-dependence of $\epsilon_{e(i)}$ may be investigated by straightforward differentiation. Neglecting hairpin formation, this process yields

$$\frac{d\epsilon_{e(i)}}{dT} \approx \frac{\epsilon_{e(i)}}{RT^2} \left(\left\langle \frac{\Delta H^\circ_{ij^*}}{1 + C^o_i K_{ij^*}} \right\rangle_e - \frac{\Delta H^\circ_{ii^*}}{1 + C^o_i K_{ii^*}} \right), \tag{9}$$

where $\langle x_{ij^*} \rangle_e$ denotes an ensemble average taken *only* over the set of error conformations, defined formally by the expression,

$$\langle x_{ij^*} \rangle_e = \frac{\sum_{j^* \neq i^*} C_{ij^*} x_{ij^*}}{\sum_{j^* \neq i^*} C_{ij^*}} = \frac{\sum_{j^* \neq i^*} C_{j^*} x_{ij^*} K_{ij^*}}{\sum_{j^* \neq i^*} C_{j^*} K_{ij^*}} \tag{10}$$

This quantity is distinguished from $\langle x_{ij^*} \rangle$ by the absence of a contribution from C_{ii^*} in both numerator and denominator, due to the restriction that measurements are over the error ensemble. In contrast with the monotonically increasing form reported for $d\epsilon_{d(i)}/dT$ [9], the form of Eq. 9 suggests that $\epsilon_e(i)$ behaves as a convex function of T, with a minimum between the melting temperatures of the planned and error TAT pairs. This is discussed via simulation, in Sec. 4.

2.3 Robustness in the Excess Limit: the Stringent Temperature

Eq. 9 may also be used to derive an approximate expression for the temperature, $T^\dagger$, at which $\epsilon_{e(i)}$ assumes a minimum value. This is accomplished by noting that at $T^\dagger_i$, $d\epsilon_{e(i)}/dT = 0$, so that

$$\left\langle \frac{\Delta H^\circ_{ij^*}}{1 + C^o_i K_{ij^*}} \right\rangle^\dagger_e = \frac{\Delta H^\circ_{ii^*}}{1 + C^o_i K_{ii^*}}, \tag{11}$$

where the superscript, '†' denotes strict evaluation at $T = T^\dagger_i$, followed by the application of three well-motivated approximations. First of all, as simulations [5] predict that $T^\dagger_i$ is consistently located substantially above the melting transition of all error TAT species, the statement $C^o_i K_{ij^*} \ll 1$ is expected to hold, so that $1 + C^o_i K_{ij^*} \approx 1$. Secondly, the ensemble average enthalpy of formation for error species $\langle \Delta H^o_{ij^*} \rangle^\dagger_e$ is assumed to be approximated to first order by the enthalpy of the single most-dominant error species, $\Delta H^o_{ij^*}(err) \equiv Inf\{\Delta H^o_{ij^*}; j^* \neq i^*\}$, given the usual dominance of this term in the weighted average. Finally, the statement $C^o_i K_{ii^*} \gg 1$ should hold, since $T^\dagger$ is also expected to be located beneath the melting temperature of the planned TAT species, ii^* [5], at least for

the case of excess input. In this case, $C_i^o K_{ii^*} + 1 \approx C_i^o K_{ii^*}$. Application of each of these expressions to Eq. 2.3, followed by rearrangement yields,

$$T_i^\dagger \approx \frac{\Delta H_{ii^*}^o}{\Delta S_{ii^*}^o + R \ln C_i^o + R \ln \left[\frac{\Delta H_{ij^*}^o(err)}{\Delta H_{ii^*}^o} \right]} \tag{12}$$

which defines the T_{rx} for optimum-fidelity operation, given an excess input of species, i. This new TAT system parameter is taken to provide a novel, precise definition of the intuitive concept of *stringent* T_{rx} previously discussed conceptually in the literature [4]. Applicability of the parameter set, $\{T_i^\dagger\}$ to both TAT system design and selection of optimal T_{rx} is discussed in Sec. 5.

3 The Excess Multi-tag Input

An approximate expression for the error-response due to a multi-tag input in the excess limit may be derived similarly, beginning with the standard expression for the *computational incoherence* [12, 13], as applied to the TAT system[9, 5]:

$$\epsilon = \frac{\sum_i \sum_{j^* \neq i^*} C_{ij^*}}{\sum_i \sum_{j^*} C_{ij^*}}, \tag{13}$$

and proceeds via approximation of the equilibrium concentrations, in a process similar to the single-tag development presented in Sec. 2. First, the impact of hybridization on each excess C_i is again neglected, so that the equation of strand conservation for each input tag, i again takes the approximate form, $C_i \approx C_i^o(1 + K_i^{hp})^{-1}$. Using this expression, the equation of strand conservation for each antitag species, j^* may then be written as,

$$C_{j^*} \approx C_a \left(1 + K_{j^*}^{hp} + \sum_i \frac{C_i^o K_{ij^*}}{1 + K_i^{hp}} \right)^{-1}. \tag{14}$$

The sum over i may now be simplified by invoking the dual of 'weak orthogonality', $\sum_i C_i^o K_{ij^*} \approx C_j^o K_{jj^*}$ which holds for all but the worst TAT encodings, but *only* under conditions of excess input for *all* tag species. Insertion of these expressions into Eq. 13 via mass action and invoking weak orthogonality yields the desired approximation,

$$\epsilon_e \approx \frac{\sum_i \frac{C_i^o}{1+K_i^{hp}} \sum_{j^* \neq i^*} \frac{K_{ij^*}(1+K_j^{hp})}{(1+K_j^{hp})(1+K_{j^*}^{hp})+C_j^o K_{jj^*}}}{\sum_i \frac{C_i^o K_{ii^*}}{(1+K_i^{hp})(1+K_{i^*}^{hp})+C_i^o K_{ii^*}}}, \tag{15}$$

where subscript, e denotes excess input, for *all* i. The form of this expression is similar to that reported for a dilute, multi-tag input [9].

The temperature-dependence of ϵ_e may be investigated by a rather tedious process of differentiation. Here, only the result is presented:

$$\frac{d\epsilon_e}{dT} \approx \frac{\epsilon_e}{RT^2} \langle \Delta\Delta H_{ij^*}^\circ \rangle_e, \tag{16}$$

where hairpin formation has been neglected, $\Delta\Delta H_{ij^*} \equiv \Delta H^{\circ}_{ij^*} - \Delta H^{\circ}_{jj^*}$, and $\langle x \rangle_e$ denotes the ensemble average of quantity x, computed over all error TAT pairs, $\{i, j^* \neq i^*\}$, respectively. Again, this expression is functionally similar to the temperature-dependence reported in [9] and [5] for the dilute multi-tag and single-tag inputs, respectively, although in [9], $\langle x \rangle_e$ was mistakenly identified in the text as the sum, rather than ensemble average over the enthalpies of formation for error conformations. From the form of Eq. 16, the error-response due to an excess input is a monotonically increasing function of T, with no error minimum between the melting temperatures of the planned and dominant error duplexes (*i.e.*, no stringent temperature). This typical behavior is in marked contrast with the logarithmically convex TAT system error behavior predicted in the vicinity of the error minimum, in response to an *asymmetric* input, composed of an excess of a single (or several, but not all) tag species.

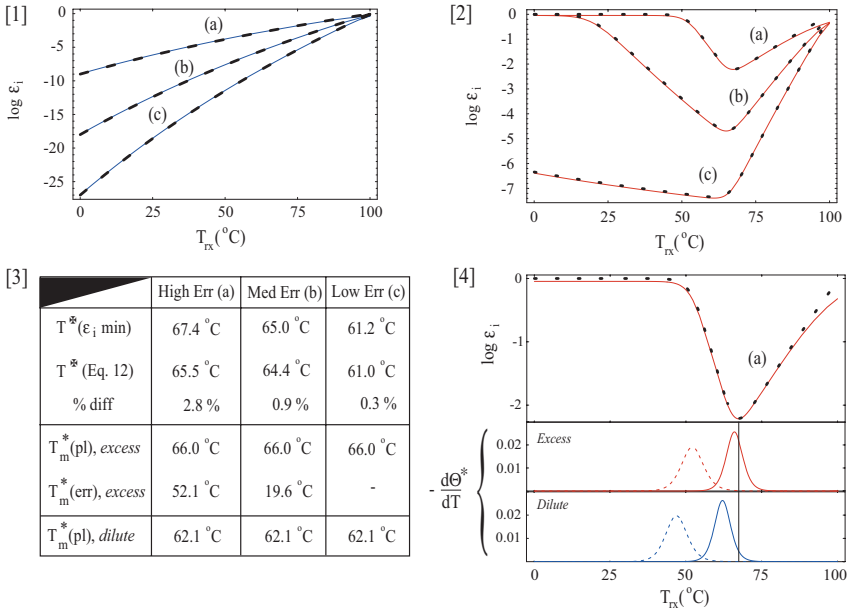
4 Simulations

In order to investigate the applicability of $\epsilon_{e(i)}$ (Eq. 7) for approximating the error-response due to an asymmetric input of a single, excess tag species, a set of simulations was implemented via (*Mathematica*$^{\text{TM}}$). Fig. 1 illustrates simulation results, which predict ϵ_i for the minimal complexity (*i.e.*, 2-probe) DNA chip, composed of ssDNAs of length 20 bases, in which the input target species may participate in a full-length planned hybrid, or a single error duplex of length (a) 15 base-pairs (bps), (b) 10 bps, or (c) 5 bps. Predictions presented as a function of T_{rx}, and in response to specific dilute ($C^o_i = 10^{-10}M$; panel 1, solid blue lines), and excess ($C^o_i = 10^{-8}M$; panel 2, solid red lines) input tag concentrations. Each antitag present at total concentration, $C_a = 10^{-9}M$; pH = 7.0, and [Na^+] = 1.0 M. Each K_{ij^*} was estimated via the Gibbs factor, using a Watson-Crick, two-state model, assuming mean doublet stacking energetics (ΔH° = -8.36 kcal/mol; ΔS° = -22.4 cal/(mol K) [14]). The impact of dangling ends, hairpin formation and tag-tag interaction were also neglected. Dashed lines in panels 1 and 2 present corresponding predictions provided by the approximate expressions presented in Sec. 2.1 for the limiting cases of dilute ($\epsilon_{d(i)}$) and excess ($\epsilon_{e(i)}$) single-tag input, respectively.

For each error condition, Panel 3 compares the predicted temperature for optimal fidelity excess operation, $T^{\dagger}$ obtained via (1) visual inspection of plotted ϵ_i curves, and (2) approximate expression, Eq. 12. For comparison purposes, melting temperatures, $T^*_m(pl)$ for each planned duplex (under both excess and dilute input conditions), and each error duplex, $T^*_m(err)$ (excess conditions only) are also illustrated, as predicted *in isolation* (denoted by, '*'). Each listed T^*_m value corresponds to the temperature which maximizes the corresponding differential melting curve, generated via a statistical, two-state model of DNA melting [15, 5]. Panel 4, top inset shows a blow-up of high-error curve, panel 2(a); Middle and bottom insets illustrate isolated, differential melting curves predicted in isolation for the planned (solid curves) and dominant error (dashed curves), for excess ('10x') and dilute ('0.1x') input, respectively.

5 Discussion and Conclusion

As shown in Fig. 1 (panel 2), simulations for all error conditions indicate that $\epsilon_{e(i)}$ (Eq. 7) is in good agreement with the predictions of the general model, ϵ_i reported in [5] for excess input, with only minor deviations at high and low temperatures. In each case, excess-input error-response is predicted to assume the expected logarithmically convex function of T_{rx}, with a minimum at distinguished temperature, $T^{\dagger}$. This behavior is in stark contrast to the low-error, monotonically T_{rx}-dependent error-response predicted for both single-tag, dilute inputs (panel 1) and multi-tag inputs which are either uniformly dilute [9, 5], or uniformly at excess (Eq. 15).



	High Err (a)	Med Err (b)	Low Err (c)
$T^{*}(\epsilon_i \text{ min})$	67.4 °C	65.0 °C	61.2 °C
T^{*} (Eq. 12)	65.5 °C	64.4 °C	61.0 °C
% diff	2.8 %	0.9 %	0.3 %
T_m^{*}(pl), *excess*	66.0 °C	66.0 °C	66.0 °C
T_m^{*}(err), *excess*	52.1 °C	19.6 °C	-
T_m^{*}(pl), *dilute*	62.1 °C	62.1 °C	62.1 °C

Fig. 1. Behavior and validity of approximate models: Panels 1 and 2: estimates of ϵ_i provided by $\epsilon_{d(i)}$ (Eq. 8) and $\epsilon_{e(i)}$ (Eq. 7) for limiting dilute and excess inputs (dashed curves), respectively, vs. full-model predictions (solid curves) for specific dilute ('0.1x': $C_i^o = 0.1C_a^{\circ}$; blue) and excess ('10x': $C_i^o = 10C_a^{\circ}$; red) inputs. For all cases, $C_a^{\circ} = 10^{-9}M; [Na^+] = 1.0M$. Curve sets (a), (b), and (c) depict error-responses due to a single dominant error-duplex of length 15/20, 10/20, and 5/20 bps, respectively. Panel 3: Optimal-fidelity temperatures for excess-input operation, $T^{\dagger}$ estimated by visual inspection of (a-c) (row 1), and Eq. 12 (row 2); Melting temperatures for the planned duplex (excess and dilute inputs) and error duplexes (excess only), predicted *in isolation* (denoted '*') also listed for comparison (rows 4-6). Panel 4: (Top sub-panel) Blow-up of high-error curve; Middle, bottom sub-panels: isolated differential melting curves for the planned (solid curves) and dominant error species (dashed curves), for excess ('10x') and dilute ('0.1x') inputs, respectively.

For TAT systems which form a component of a DNA computer, the potential for excessive error due to an *asymmetric input*, consisting of a dilute component combined with an excess component of one (or more) input species may be evaluated by examining that system's set of single-tag, excess input values, $\{\epsilon_e(i)\}$, at the operating temperature of interest, T_{rx}. For such systems, the mean value of $\epsilon_e(i)$ over i is proposed as a well-defined measure for high-fidelity design. As indicated by Fig. 1 (panel 3), Eq. 12 provides a good approximation for $T_i^\dagger$. Note that minimization of $\epsilon_{e(i)}$ has the additional desirable effect of decreasing the sensitivity of the excess error response to variations away from $T^\dagger$, since this process broadens the width of $\epsilon_{e(i)}$ around $T_i^\dagger$.

The evident dominance of target duplex formation on the inflection point, as evinced by the general proximity of $T_i^\dagger$ to the melting temperature of the isolated planned TAT pair, $T_m^*(pl)$ under excess conditions (see Fig. 1(panel 3)) deserves further discussion. As illustrated in Fig. 1(panel 4; top inset), in the context of the high-error case (panel 2, (a)), the sigmoidal portion of $\epsilon_{e(i)}$ beneath $T_i^\dagger$ (vertical line) is seen to just span the interval between $T_m^*(pl)$ and $T_m^*(err)$, as indicated by the differential melting curves of the isolated planned and error TAT pairs, predicted under excess-input conditions (Panel 4, middle inset; solid and dashed red curves, respectively)). Overall duplex formation in this regime, predicted to accompany successive decreases in T_{rx} beneath $T_i^\dagger$, is thus characterized by increasing concentrations of error TAT species, compensated for by increasingly smaller fractional increases in the concentration of planned TAT species, due to the onset of planned antitag saturation (thus, the sigmiodal shape).

From a fidelity perspective, for systems in which the potential for asymmetric, excess input is unavoidable, the most robust operating condition is approximated by the mean value of the set $\{T_i^\dagger\}$. For this reason, design for uniform $T_i^\dagger$ values, enabling uniformly error-resistant operation at the mean, is proposed as a second well-defined criterion for guiding high-fidelity statistical-thermodynamic TAT design. On the other hand, several points of care are required in interpreting the mean $T_i^\dagger$ value as an optimal operation T_{rx}. First of all, if an architecture can be verifiably biased to ensure operation of any associated TAT system strictly in the dilute regime, then a lower temperature of much greater fidelity may be employed, according to the temperature-dependence of ϵ_d, as shown in Fig. 1, panel 1 [9]. If non-dilute operating conditions cannot be strictly avoided, then simulations strongly suggest the utility of selecting a higher operating temperature, for which $\{T^\dagger\}$ should provide a guide. However, additional care is still required.

An additional concern is that operating conditions be selected which not only ensure high fidelity, but also allow substantial process completion, for all potential input conditions of interest (*i.e.*, both excess and dilute). For this reason, a comparison of $T^\dagger$ with the melting temperatures of each planned TAT species (denoted, $T_m^*(i)$, for a TAT system with $|i|$ distinct, single-tag inputs) as expected under dilute conditions is also indicated. Based on the predictions provided by Fig. 1 (panel 3, row 6), adoption of $T_i^\dagger$ as the optimal T_{rx} for general

system operation, although attractive due to its robustness to error-prone excess inputs, will always come at the cost of reduced completion of the planned TAT pair, and according to Fig. 1, is strictly satisfactory only for well-encoded TAT systems (as $T^{\dagger}$ is located beneath $T_m^*(pl)$, for both dilute and excess inputs). This is illustrated more clearly in Fig. 1 (panel 4, bottom inset), which compares the depressed melting transition of the planned duplex under dilute input (('0.1x'), solid blue curve; compare with the same transition, under excess input (solid red curve, middle inset)) with the elevated $T_i^{\dagger}$ characteristic of a high-error system (vertical line), indicating a substantial lack of completion of planned duplex, at $T_i^{\dagger}$ under dilute conditions. If the potential for operation in the non-dilute regime cannot be avoided (so that a suitable, high-fidelity, lower T_{rx} cannot be selected), the best compromise is probably to select $T_{rx} = Inf\{\{T_i^*\}\bigcup\{T_m^*(i)\}\}$, where the melting temperatures of planned TAT pairs are assessed under the most dilute practical conditions of interest. Furthermore, to minimize this problem, it is evident that a third well-motivated design criterion is to encode for uniform T_m^*'s of planned interaction, as suggested previously [8].

Acknowledgements

Financial support generously provided by Grants-in-Aid for Scientific Research B (15300100 and 15310084), from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, from Nihon Tokutei, and by JST-CREST.

References

1. D. Lockhart and E. Winzeler, *Nature* **405**, 827 (2000).
2. Q. Liu, et al., *Nature* **403**, 175 (2000).
3. A. Suyama, et al., in *Currents in Computational Molecular Biology*, S. Miyano, et al., Eds. (Univ. Acad. Press, 2000), 12.
4. A. BenDor, et al., *J. Comput. Biol.* **7**, 503 (2000).
5. J. Rose, M. Hagiya, and A. Suyama, in *Proc. 2003 Cong. Evol. Comp., Vol. IV*, R. Sarker, et al., Eds., (IEEE Press, 2003), 2740.
6. R. Deaton, et al., *Phys. Rev. Lett.* **80**, 417 (1998).
7. Q. Liu, et al., *Biosystems* **52**, 25 (1999).
8. H. Yoshida and A. Suyama, in *DNA Based Computers V*, E. Winfree and D. Gifford, Eds., (Am. Math. Soc., 2000), 9.
9. J. Rose, et al., in *DNA Computing*, N. Jonaska and N. Seeman, Eds., (Springer, Berlin, 2001), 138.
10. P. von Hippel and O. Berg, *Proc. Natl. Acad. Sci. USA* **83**, 1608 (1986).
11. B. Eaton, et al., *Chemistry and Biology* **2**, 635 (1995).
12. J. Rose, et al., in W. Banzhaf, et al., eds., *Proc. GECCO '99*, (Morgan Kauffman, San Francisco, 1999), 1829.
13. J. Rose and R. Deaton, in *DNA Based Computers*, A. Condon and G. Rozenberg, Eds., (Springer, Berlin, 2001), 231.
14. J. SantaLucia, Jr., *Proc. Natl. Acad. Sci.* **95**, 1460 (1998).
15. R. Wartell and A. Benight, *PHYSICS REPORTS* **126**, 67 (1985).