Engineering multistate DNA molecules: a tunable thermal band-pass filter

John A. Rose1✉, Ken Komiya2, Satoshi Kobayashi3

1Culture, Society and Media Cluster, Ritsumeikan Asia Pacific University, Beppu, Oita 874-8556, Japan
2School of Computing, Tokyo Institute of Technology, Yokohama, Kanagawa 226-8503, Japan
3Department of Computer Science, The University of Electro-Communications, Chofu, Tokyo 182-8585, Japan
✉ E-mail: jarose@apu.ac.jp

Published in Micro & Nano Letters; Received on 25th May 2016; Accepted on 19th July 2016

Engineering of biopolymer ‘partner folds’ that exist in competitive equilibrium with the native state to produce exotic behaviours remains a relatively unexplored area of molecular engineering. Previously, a temperature-sensitive DNA nanodevice that operates by harnessing such a partner fold to implement a thermal band-pass filter was proposed, modelled, and experimentally validated. Due to its peculiar hill-shaped efficiency profile, which differs markedly from the sigmoidal melting curves of simple DNA hairpins, this device could be used to implement temperature-specific control of other molecular machines, and thus represents a promising biotechnological advance. However, no effort was made to examine the detailed dependencies of the peak temperature $T^\dagger$, width $\Delta T^\dagger$, and maximum efficiency $\varepsilon_{\text{max}}$ on the stabilities of device components. In this work, closed-form expressions for $T^\dagger$ and $\varepsilon_{\text{max}}$ are derived and validated. The functional behaviours of these expressions are then examined and harnessed to construct an efficient algorithm for producing designs with target $T^\dagger$ and $\Delta T^\dagger$ values and optimised $\varepsilon_{\text{max}}$, thereby establishing the feasibility of algorithmic device design. Method effectiveness is validated via production of a target filter, with detailed simulations of device behaviour. Finally, a discussion is presented regarding model effectiveness, extension, and scope.

1. Introduction: The DNA molecule has recently attracted substantial attention as a material for constructing functional devices, allowing the programming of structural changes and self-assembly at the nanometre scale via sequence design according to Watson–Crick (WC) base-pairing rules [1, 2]. Numerous molecular algorithms and devices based on hairpin folding have been proposed for various applications, including the Whiplash polymerase chain reaction [3], molecular memory [4], and as simple sensors [5]. Furthermore, DNA hairpin formation, along with extension by polymerase, is one of the most important operations in the theory of DNA computing models [6]. However, attention to subsystem ‘partner folds’ has largely been restricted to error estimation and prevention [7].

Meanwhile, although multistate RNAs are regarded to play a role in RNA evolution [8], relatively few studies have addressed the design of such partner folds, which may exist in equilibrium with the native state, to achieve custom behaviours. In this regard, in [9], a temperature-sensitive bistable DNA hairpin-based nanodevice was proposed, modelled in terms of partner-fold occupancy, and experimentally validated as a potential platform for exploiting partner-fold behaviour to implement a thermal band-pass filter. In related work [10], an algorithm was proposed for designing a bistable RNA to implement a $T$-dependent structural switch, although the development focused on the energy barrier between the alternative structures, rather than tuning the occupancy of the partner fold to support custom application, as in [9].

As shown in Fig. 1, this nanodevice consists of a single DNA strand capable of transforming between two competing hairpins (dominant fold, $S_1$; less stable partner fold, $S_2$, which is the device target) in response to changes in $T$. As shown in Fig. 2 for the device encoding in [9] under an improved parametrisation (cf. Section 2.4), $S_1$ exhibits a hill-shaped thermal stability characterised by a peak value $\varepsilon_{\text{max}}$ at temperature $T^\dagger$ and width $\Delta T^\dagger$. This behaviour, which differs remarkably from the S-shaped melting curves of simple hairpins, was experimentally observed in [9] via fluorescence measurements (0.95 correlation with predicted behaviour), confirming both the hill-shaped device efficiency and the ability of an equilibrium model to provide quantitative predictions.

As noted in [9], the unusual thermal response of this DNA device is promising, as it might be used to implement a tunable thermal band-pass filter. If practically desirable to operate at target $T^\dagger$ and $\Delta T^\dagger$ values, this device could be applied to control DNA self-assembly, or as a multi-purpose, $T$-dependent modulator of chemical circuits (e.g. control a polymerisation reaction via gated activation), similar to the use of RNA hairpins in nature as in-vivo thermometers to regulate transcription [11]. Accordingly, in the current work, the statistical thermodynamic model of this device is extended to derive closed-form expressions for $T^\dagger$ and $\varepsilon_{\text{max}}$. Based on the predicted scaling behaviours, an algorithm is then constructed for producing designs with targeted $T^\dagger$ and $\Delta T^\dagger$ values and optimised $\varepsilon_{\text{max}}$, establishing the feasibility of algorithmic design. Both model validity and method effectiveness are verified in the context of production of a target filter design, with simulations of device behaviour at each step. A preliminary version of partial results was presented at IEEE-NEMS 2016 [12].

2. Statistical thermodynamic analysis

2.1. Device efficiency: An expression for the device efficiency, $\varepsilon$ at equilibrium has previously been derived [9], using the tools of equilibrium chemistry in [13]. For clarity, an abbreviated treatment is also presented here. The derivation begins by invoking strand conservation for the coupled equilibrium shown in Fig. 1, yielding $C_{\text{tot}} = C_{\text{ss}} + C_1 + C_2$, where $C_{\text{tot}}$ denotes the total strand concentration, and $C_{\text{ss}}, C_1,$ and $C_2$ are the equilibrium concentrations of melted coils and hairpins $S_1$ and $S_2$, respectively. The mass action expression for each simple folding equilibrium has the form $C_{\text{tot}} = C_{\text{ss}} K_0$ where $K_0$ is the equilibrium constant for formation of fold $S_0$. Equating device efficiency to the fractional occupancy of the target fold, $\varepsilon = C_1/C_{\text{tot}}$, followed by insertion of the above expressions with rearrangement yields

$$\varepsilon = \left[ 1 + \frac{(1 + K_1) T^\dagger}{K_2} \right]^{-1}. \quad (1)$$

The equilibrium constant for formation of each folded hairpin,
Fig. 1  Bistable DNA nanodevice formed by two competing hairpins, $S_1$ and $S_2$ with stem lengths $L_1 = 13$ and $L_2 = 12$ bps, and codewords for hybridisation separated by poly-T spacers, $sp_3$ and $sp_5$, with lengths $L_{sp3} = 28$ and $L_{sp5} = 29$ bases. The less stable partner fold, $S_2$ is the targeted fold of the device

$$i = \{1, 2\}$$ may be estimated via the Gibbs factor

$$K_i = \exp\left(-\Delta G_i^\circ / RT\right). \quad (2)$$

where $R$ is the gas constant, $T$ is the absolute temperature, and $\Delta G_i^\circ$ is the Gibbs free energy, which is estimated via $\Delta G_i^\circ = \Delta H_i^\circ - T \Delta S_i^\circ$ from the enthalpy, $\Delta H_i^\circ$ and entropy, $\Delta S_i^\circ$ of formation of structure $i$. Estimation of $\epsilon$ then reduces to an estimation of $\Delta H_i^\circ$ and $\Delta S_i^\circ$ for each hairpin $i$, using the methods in Section 2.4.

2.2. Device peak temperature, $T^\dagger$: A closed-form expression for $T^\dagger$, which corresponds to the peak of the device efficiency curve, may be derived from (1) using standard analytical methods for locating an extremum, as follows. The derivation begins by defining the quantity, $\xi = (1 + K_1)/K_2$, so that

$$\epsilon = (1 + \xi)^{-1},$$

Calculation of the derivative of $\epsilon$ then proceeds, as follows

$$\frac{d\epsilon}{dT} = -(1 + \xi)^{-2} \frac{d\xi}{dT}. \quad (3)$$

and

$$\frac{d\xi}{dT} = K_2^{-1} \frac{d}{dT} K_1 + (1 + K_1) \frac{d}{dT} K_2^{-1}.$$

This expression may be reduced via the following basic relations

$$\frac{d}{dT} K = K \frac{\Delta H^\circ}{RT^2} \quad \text{and} \quad \frac{d}{dT} K^{-1} = -K^{-1} \frac{\Delta H^\circ}{RT^2}.$$

Insertion of these expressions for $K_1$ and $K_2$, with rearrangement then yields

$$\frac{d\xi}{dT} = \frac{1}{RT^2} K_1 (\Delta H_2^\circ - (1 + K_1^{-1}) \Delta H_1^\circ).$$

Insertion into (3), followed by evaluation at $T^\dagger$ and equating the result with zero to correspond to an extremum yields

$$\frac{d\xi}{dT} |_{T^\dagger} = -\frac{K_1}{RT^2} \left[(1 + K_1^{-1}) \Delta H_2^\circ - \Delta H_1^\circ \right] = 0$$

where ‘$\dagger$’ denotes evaluation at $T^\dagger$. As the first factor is strictly positive, the last factor must be equal to zero

$$(1 + K_1^{-1}) \Delta H_2^\circ - \Delta H_1^\circ = 0.$$ 

Expansion of $K_1^{-1}$ via (2), and taking the logarithm of both sides with rearrangement then yields

$$\frac{\Delta H_2^\circ}{RT^2} = \frac{\Delta S_2^\circ}{R} + \ln \left(\frac{\Delta H_2^\circ}{\Delta H_1^\circ}\right).$$

where $\Delta \Delta X^\circ = \Delta S_2^\circ - \Delta S_1^\circ$ for $X \in \{H, S\}$. Finally, rearrangement to isolate $T^\dagger$ yields the desired result

$$T^\dagger = \frac{\Delta H_1^\circ}{\Delta S_1^\circ + R \ln (\alpha - 1)}, \quad (4a)$$

where $\alpha = \Delta H_1^\circ / \Delta H_2^\circ$ is a useful parameter for discussing system design. This equation may also be expressed in terms of $T_m(1) = \Delta H_1^\circ / \Delta S_1^\circ$, the melting temperature of $S_1$ in isolation, as

$$T^\dagger = T_m(1) \left[1 + R \Delta S_1^\circ \ln(\alpha - 1) \right]^{-1}. \quad (4b)$$

As discussed in Section 3.2, the tendency of $T^\dagger$ to approach $T_m(1)$ with increasing $\alpha$ provides insight towards algorithmic filter design. In this regard, the value, $\alpha = 2$, at which point $T^\dagger = T_m(1)$, represents an important special case, and is termed as the design zero-point (ZP) for a given design process with fixed $T_m(1)$.

2.3. Device peak efficiency, $\epsilon_{\max}$: Equation (4) provides a convenient method of obtaining an exact value for $\epsilon_{\max}$ via insertion into (1) as $\epsilon(T^\dagger)$ for any encoded instance, and also enables visualisation of the accompanying scaling behaviours. However, characterisation of the functional forms of these scaling tendencies requires a closed-form expression for $\epsilon_{\max}$. A useful approximate form may be produced via successive approximation, as follows. First, evaluating (1) at $T^\dagger$ and noting that $K_1 \gg K_2$, by design, yields the form

$$\epsilon_{\max} \simeq K_2^2 / (1 + K_1^{-1}). \quad (5)$$

For the purpose of deriving a closed-form expression for $\epsilon_{\max}$ which
is useful in clarifying the scaling behaviour of designs in the primary design range, viz., beneath ZP (i.e. $1 < \alpha < 2$), this expression can be further simplified via application of a second, less general approximation, $K_1^0 \gg 1$, which yields

$$\epsilon_{\text{max}} \approx K_1^0 / K_1^1. \quad (6)$$

This approximation can be expected to perform reasonably well for designs operating well below ZP (i.e. with large $K_1^0$ values). As a concrete example, for the reference encoding, $K_1^0 \approx 1.2$ and $K_1^1 \approx 10$, which corresponds to a relative error of $\approx 0.22$ between the two approximate forms. However, at ZP, $K_1 = 1$ and $\eta \approx 1$, so that agreement has degraded to order-of-scale.

Combining the exponentials in (6) and taking the logarithm of both sides, followed by isolation of $T^4$ on the left-hand side and equating the result with (4a) for $T^4$ then yields

$$T^4 = \frac{\Delta H_I^T}{\Delta S^T + R \ln (\alpha - 1)} \approx \frac{\Delta S^T}{\Delta S^T + R + \ln \epsilon_{\text{max}}}. \quad (7)$$

Elimination of $T^4$, with rearrangement to isolate $\ln \epsilon_{\text{max}}$ yields

$$\ln \epsilon_{\text{max}} \approx \frac{\Delta S^T}{\Delta H_I^T} \beta - \frac{\Delta S^T}{R},$$

where $\beta$ is defined for convenience as $\beta = \ln (\alpha - 1) + \Delta S^T / R$. However, as the magnitude of the logarithmic term in $\beta$ is very small relative to that of the trailing term for all relevant designs, $\beta \approx \Delta S^T / R$. Insertion of this approximation then yields

$$\ln \epsilon_{\text{max}} \approx \frac{\Delta S^T}{\Delta H_I^T} \beta - \frac{\Delta S^T}{R}. \quad (8)$$

Finally, this expression can be simplified as

$$\ln \epsilon_{\text{max}} \approx \frac{1}{R} \left[ \Delta S^D = \frac{\Delta H_I^T}{T_m(1)} \right]. \quad (9)$$

2.4. Model implementation: A statistical thermodynamic approach was used to model the $K_{\text{eq}}$ for each device hairpin [13], employing a nearest-neighbour (NN) model of duplex energetics for each hairpin stem, and a Jacobson–Stockmeyer (J–S) power law to model the loop entropy. The NN model was applied using the parameters in [14] for estimating the stacking enthalpy and entropy of WC base-pair (bp) doublets. Based on conditions in [9], a net ionic strength of $[\text{Na}^+] = 0.165 \text{ M}$ was adopted for all simulations, necessitating an entropic correction for variation from 1.0 M Na$^+$. The energetic penalty for helix unwinding at the two duplex ends was estimated using the helix initiation parameters in [14], which estimate the cooperativity parameter, $\sigma$, when applied as a statistical weight. As an improvement over the original model parameterisation in [9], the statistical weight of hairpin loop closure was modelled using a J–S inverse 2.4 power law, as suggested by Goddard et al. [15], and applied in [14] in a piecewise manner for the combined modelling of all types of DNA loops. However, for theoretical clarity, a pure J–S power law employing this exponent was selected, as $\Delta S^D_{\text{loop}} = -2.44R \ln (\alpha + 1)$.

Parameterisation for mean-case NN energetics (hereinafter, ‘mean-case’) simulations was performed similarly, using: (i) mean WC doublet stacking parameters, $\Delta H_I^T$ and $\Delta S^T$ obtained via the respective averages over the NN parameters in [14]; and (ii) mean initiation parameters, $\Delta H_I^T$ and $\Delta S^T$ obtained via respective averages over the accompanying parameters for initiation of helices with GC and AT terminal bps. Note that in terms of the net $\Delta H_I^T$ and $\Delta S^T$ of hairpin formation, this yields the linear form

$$H^I = \Delta H_I^{\text{nuc}} + (L - 1)\Delta H_I^{\text{nn}}; \quad (9a)$$

$$\Delta S^T = \Delta S^T_{\text{nn}} + (L - 1)\Delta S^T_{\text{nn}} + \Delta S^T_{\text{loop}}. \quad (9b)$$

For clarity, each hairpin $K_{\text{eq}}$ was estimated as a product of the subweights for loop closure, initiation, and stacking, from the respective $\Delta H_I^T$ and $\Delta S^T$ values, as $K_{\text{eq}} = \sigma \exp (-\Delta G_{\text{nn}}^T / RT) (n + 1)^{-2.44}$, where $\Delta G_{\text{nn}}^T$ is the net free energy of stacking. These equilibrium constants were then combined with the derived model equations to simulate behaviours for both the original device encoding and mean-case scaled implementations, under their respective parametrisations. Simulations were performed via Mathematica [Wolfram Research].

3. Results

3.1. Functional behaviour of $T^4$: An examination of the behaviour of $T^4$ against the stabilities of hairpins $S_1$ and $S_2$ is essential for gaining traction in development of an algorithmic method for device design. First, (4) displays a surprising lack of dependence on $\Delta S^T$. In addition, the form of (4b) indicates that $T^4 \propto T_m(1)$, suggesting a strong sensitivity to variations in the thermal stability of the dominant hairpin, $S_1$. In contrast, the placement of $\Delta H_I^T$ inside the logarithm indicates a much lower sensitivity to variations in the stability of $S_2$. Finally, (4b) indicates that $T^4$ reduces to $T_m(1)$ at ZP, where $\Delta H_I^T = \Delta H_I^T / 2$. This represents a useful special case that divides the $T^4$ values of designs obtained by successive decreases in $L_2$ into two very roughly linear regimes in the vicinity of $T_m(1)$: (i) for $\Delta H_I^T < \Delta H_I^T / 2$ (more-stable $S_2$), $T^4 \propto T_m(1)$ from below; and (ii) for $\Delta H_I^T > \Delta H_I^T / 2$ (less-stable $S_2$), $T^4 > T_m(1)$, and diverges from $T_m(1)$ above.

To illustrate these behaviours, Fig. 3 presents a mean-case simulation of variations in $T^4$ against changes in the system’s component hairpin stem lengths, which confirms the predicted relative insensitivity to changes in $L_2$. In addition, Fig. 3b shows the scaling behaviour of $T^4$ for a specific set of fixed $L_1$ values, with $T^4$ values plotted relative to the fixed $T_m(1)$ value for each curve, indicating that $T^4$ slowly approaches $T_m(1)$ from below as $\Delta L_2$ is increased, reaching $T_m(1)$ near $L_2 = L_2^*/2$, as predicted.

3.2. Functional behaviour of $\epsilon_{\text{max}}$: The form of (8) indicates a log-linear relationship between $\epsilon_{\text{max}}$ and each of the bulk thermodynamic parameters of $S_2$. However, since $\Delta H^T$ and $\Delta S^T$ directly oppose, the net dependence on $L_2$ is not immediately evident. Given adoption of a mean-case NN model (9), this dependence can be clarified via differentiation, yielding

$$\frac{d(\ln \epsilon_{\text{max}})}{dL_2} \approx \frac{\Delta S^T_{\text{nn}}}{T_m(1)} \left[ 1 - \frac{T_m(\infty)}{T_m(1)} \right] \approx \frac{\Delta S^T_{\text{nn}}}{R} \gamma, \quad (10)$$

where $\gamma = 1 - \frac{T_m(\infty)}{T_m(1)}$, and $T_m(\infty) = \frac{\Delta H_I^T}{\Delta S^T_{\text{nn}}}$ corresponds to the $T_m$ value of a hairpin with mean NN energetics in the infinite stem-length limit, and thus bounds $T_m(1)$, above. Accordingly, $\gamma$ is strictly negative, so that $\ln \epsilon_{\text{max}} \propto L_2$ with a constant positive slope. Expressed in terms of finite changes, this equation indicates that an isolated unit decrease in $L_2$ will be accompanied by a constant log-linear efficiency decrease of

$$\Delta (\ln \epsilon_{\text{max}}) \approx -\frac{\Delta S^T_{\text{nn}}}{R} \gamma. \quad (11)$$

The behaviour of $\ln \epsilon_{\text{max}}$ against $L_1$ can be addressed similarly,
Accordingly, given that $T^\dagger$ is a complex function of the independent variable, so that a discussion in terms of $\ln (T)$ is primarily isolated, small variations in $L_2$ will act to oppose like variations in $L_1$, and is thus deferred to future study. However, it indicates that isolated, small variations in $L_1$ will act to oppose like variations in $L_2$ in terms of effect on $\ln (\varepsilon_{\text{max}})$ and will be dampened by a factor of $1/\alpha$, which approaches $1/2$ at $ZP$. Accordingly, given that $T^\dagger$ is primarily fixed by $L_1$, algorithmic attempts to offset efficiency losses due to repeated $L_2$ decreases during design via a compensating increase in $L_1$ appear unlikely to be well motivated.

To illustrate these behaviours, Fig. 4 presents a mean-case simulation of variations in $\ln (\varepsilon_{\text{max}})$ against changes in $L_1$ and $L_2$, as predicted by (8), along with exact values obtained by $\epsilon(T^\dagger)$ via (4b). As shown in log$_{10}$ scale for clarity, both curve sets predict $\varepsilon_{\text{max}}$ to scale log-linearly against $L_2$, but to have a more complex dependence on $L_1$. Based on these curves along with values listed in Table 1, (8) provides an effective means of examining the scaling behaviour of $\varepsilon$, and a reasonable approximation of $\ln (\varepsilon_{\text{max}})$ for specific designs (within $\simeq 10\%$). However, when converted to $\varepsilon_{\text{max}}$ values, agreement deteriorates to order of scale at ZP, as predicted.

The scaling behaviour of $\ln (\varepsilon_{\text{max}})$ against $n_2$ may also be addressed similarly. Adoption of a mean-case NN model and approximating $\ln (n)$ as $\ln (n)$ in the J–S loop entropy, due to this architecture’s large loops, followed by differentiation yields

$$\frac{d(\ln \varepsilon_{\text{max}})}{dL_2} \simeq -\frac{\Delta S_{\text{ent}}}{\alpha R} \gamma \simeq -\frac{1}{\alpha} \frac{d(\ln \varepsilon_{\text{max}})}{dL_2}. \quad (12)$$

Unlike (10), this expression is a complex function of the independent variable, so that a discussion in terms of finite changes is more involved, and is thus deferred to future study. However, it indicates that isolated, small variations in $L_1$ will act to oppose like variations in $L_2$ in terms of effect on $\ln (\varepsilon_{\text{max}})$ and will be dampened by a factor of $1/\alpha$, which approaches $1/2$ at $ZP$. Accordingly, given that $T^\dagger$ is primarily fixed by $L_1$, algorithmic attempts to offset efficiency losses due to repeated $L_2$ decreases during design via a compensating increase in $L_1$ appear unlikely to be well motivated.

To illustrate these behaviours, Fig. 4 presents a mean-case simulation of variations in $\ln (\varepsilon_{\text{max}})$ against changes in $L_1$ and $L_2$, as predicted by (8), along with exact values obtained by $\epsilon(T^\dagger)$ via (4b). As shown in log$_{10}$ scale for clarity, both curve sets predict $\varepsilon_{\text{max}}$ to scale log-linearly against $L_2$, but to have a more complex dependence on $L_1$. Based on these curves along with values listed in Table 1, (8) provides an effective means of examining the scaling behaviour of $\varepsilon$, and a reasonable approximation of $\ln (\varepsilon_{\text{max}})$ for specific designs (within $\simeq 10\%$). However, when converted to $\varepsilon_{\text{max}}$ values, agreement deteriorates to order of scale at ZP, as predicted.

The scaling behaviour of $\ln (\varepsilon_{\text{max}})$ against $n_2$ may also be addressed similarly. Adoption of a mean-case NN model and approximating $\ln (n)$ as $\ln (n)$ in the J–S loop entropy, due to this architecture’s large loops, followed by differentiation yields

$$\frac{d(\ln \varepsilon_{\text{max}})}{dn_2} \simeq -\frac{4.44}{n_2}. \quad (13)$$

which indicates a log–log relationship between $\varepsilon_{\text{max}}$ and $n_2$. This equation may be re-expressed in terms of finite changes via separation of variables and definite integration of the respective sides over $d(\varepsilon_{\text{max}})$ and $dn_2$, yielding

$$\Delta \varepsilon_{\text{max}} \simeq \varepsilon_{\text{max}}(0) \left[ \left( 1 + \frac{n_2}{n_2(0)} \right)^{-2.44} - 1 \right]. \quad (14)$$

where ‘(0)’ denotes initial design values. This equation predicts an isolated fractional decrease in $n_2$ to be accompanied by a fractional increase in $\varepsilon_{\text{max}}$, as discussed more fully in the next section, in the concrete context of the proposed design method, viz., Algorithm 1 (see Fig. 5).

3.3. Algorithm for filter design: Although no closed-form expression for predicting the scaling behaviour of $\Delta T_{50}$ is available, prospective simulations in [9] indicate that $\Delta T_{50}$ consistently narrows with increasing $\Delta L$. Combined with this empirically determined behaviour, the behaviours predicted by (4b) for $T^\dagger$ suggest an efficient method for targeted filter design. In particular, if the device is designed to initialise $T_{\text{init}}(1)$ to roughly the desired $T_{\text{target}}$ value, then $\Delta L$ can be iteratively

---

**Fig. 3** Peak temperature scaling behaviour

- $T^\dagger$ against changes in the stem lengths of hairpins $S_1$ and $S_2$, as expressed by $L_1$ and $L_2$.
- Curves from (a) for a range of $S_1$ stem lengths, with $T^\dagger$ values plotted relative to the fixed $T_{\text{init}}(1)$ value for each curve. The curve for $L_1 = 13$ bps also shows the trajectory taken by the illustrative implementation of Algorithm 1 (see Fig. 5), where points represent designs produced on the path to convergence.

**Fig. 4** Peak efficiency scaling behaviour

- $a \log_{10} \varepsilon_{\text{max}}$ against variations in the stem lengths of hairpins $S_1$ and $S_2$, as predicted via (4b), as $\epsilon(T^\dagger)$ (lower curve; exact), and (8) (upper curve; approx.).
- Exact (solid) against approx. (dashed/dotted) curves from (a) for implementations with a range of $S_1$ stem lengths; approx/exact curve pairs from top-to-bottom depict curves for $L_1 = \{16, 13, 10\}$ bps. The exact curve for $L_1 = 13$ bps also shows designs produced on the path to convergence (red points) by the illustrative implementation of Algorithm 1 (see Fig. 5).

---

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/)

Micro & Nano Letters, 2016, Vol. 11, Iss. 10, pp. 595–601
doi: 10.1049/mnl.2016.0345
increased from unity by decreasing \( L_2 \), while the stability of \( S_1 \) is maintained via fixed \( L_1 \) and \( L_{gpt} \). Successive increases in \( \Delta L \) will then result in a gradual decrease in \( \Delta T_{50} \), with a simultaneous approach to \( T_f^\ast \). Convergence occurs when a suitably narrow \( \Delta T_{50} \) value is obtained, given that \( T_f \) remains suitably close to \( T_f^\ast \). Herein, each \( L_2 \) reduction is implemented via a 3′-word-5′-trim, which deletes the 5′ base of the 3′ DNA codeword for \( S_2 \) formation. While this also increases \( n_2 \) by unity, causing a small change in \( \Delta S_2 \) as a side-effect, \( T_f \) is predicted to be invariant to this change.

The scaling behaviour predicted by (14), combined with the invariance of \( T_f \) to changes in \( \Delta S_2 \) suggests the utility of a post-convergence maximisation of \( \epsilon_{\max} \) via an isolated reduction in \( n_2 \). However, in the absence of a closed-form expression for \( \Delta T_{50} \), the accompanying effect on curve width is unclear. Accordingly, this optimisation step is implemented via an isolated, single-base trimming of \( sp3 \), while monitoring \( \Delta T_{50,design} \), and halting if \( W \)-convergence is lost, which ensures that the performance achieved during convergence is maintained.

A substantial simplification may also be achieved by adopting a mean-case approach, which exploits the dependency of device behaviour on the encoded bulk thermodynamic properties, rather than specific sequence. By effectively classifying large sets of encodings as thermodynamically equivalent, this approach obviates the need for repetitive, computationally intensive sequence-specific search. This approach is notably less fruitful in the related problem.

### Table 1: Values of \( \epsilon_{\max} \), \( T_f^\ast \), \( T_\sigma \), and \( \Delta T_{50} \) for simulated device efficiency curves

| \( \Delta L \) (bps) | \( L_{gpt} \) (nt) | \( \epsilon_{\max} \) | \( \log_{10} \epsilon_{\max} \) | \( T_f^\ast \) (°C) | \( T_\sigma \) (°C) | \( T_f \) (°C) | \( \Delta T_{50} \) (°C) |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Fig. 2            | 1               | 28              | 0.10            | -1.00           | -0.821          | 62.3            | 62.3            | 43.3            | 69.0            | 25.7            |
|                   | 2               | 28              | 0.031           | -1.50           | -1.36           | 55.7            | 55.7            | 36.3            | 62.5            | 26.2            |
|                   | 3               | 28              | 9.73 × 10^{-1}  | -2.01           | -1.81           | 57.5            | 57.5            | 46.3            | 63.4            | 17.1            |
|                   | 4               | 28              | 3.10 × 10^{-1}  | -2.51           | -2.26           | 58.6            | 58.6            | 50.2            | 64.3            | 14.1            |
|                   | 5               | 28              | 1.00 × 10^{-1}  | -2.98           | -2.70           | 59.5            | 59.5            | 52.5            | 65.5            | 13.0            |
|                   | 5               | 0               | 1.09 × 10^{-1}  | -3.49           | -3.15           | 60.4            | 60.4            | 54.1            | 66.2            | 12.1            |
|                   | 5               | 0               | 1.09 × 10^{-1}  | -2.96           | -2.66           | 60.4            | 60.4            | 54.1            | 66.2            | 12.1            |

\( L_1 = 13 \) bps for all curves; ‘num’: numerical.

DNA Strand Design [16], which seeks to design strand sets with minimal mis-hybridisation, which is highly sequence dependent.

This design strategy is expressed formally as Algorithm 1 (see Fig. 5). The initial seed of the design process is taken as the reference structure in Fig. 1, implemented using mean-case NN energetics and appropriate variations of \( L_1 \) and \( L_2 \). The process goal is production of a system design which satisfies the following two properties: (i) T-convergence: \( \Delta T_f \equiv |T_{target} - T_{design}| \leq \sigma_T \), and (ii) W-convergence: \( \Delta T_{50,design} \leq \Delta T_{50,thr} \), where \( \sigma_T \) is a tolerance for \( T_f \) from its target value and \( \Delta T_{50,thr} \) is a threshold maximum width. It is also necessary to account for the possibility that input condition \( \Delta T_{50,thr} \) is too narrow to permit W-convergence. In this case, \( T_f \) will rise above \( T_m \) for \( T_f \) to convergence is lost, a detectable condition referred to as ‘T-divergence’, which serves as a test for non-convergent halting. Note that Algorithm 1 (see Fig. 5) also employs several helper functions: \( \text{InitLG} (T_{target}) \), which uses the model in Section 2.4 to determine the \( L_1 \) value producing the hairpin \( T_m \) nearest to \( T_{target} \); \( \text{TDAg}(L_1, L_2) \), which computes \( T_f \) using (4a); and, \( \text{Width}(L_1, L_2, L_{gpt}) \), which computes \( \Delta T_{50,design} \) via numerical estimation of \( T_f \) and \( \sigma_T \) (cf. Fig. 2) using (1) and (4a).

As a concrete example, Fig. 6 shows the results of application of Algorithm 1 (see Fig. 5) to produce a filter design with target values: \( T_{target} = 60.0 \) °C, which is intentionally near that of the mean-case implementation of the reference system in Fig. 1, \( \sigma_T =2.0\) °C, and \( \Delta T_{50,thr} =12.5\) °C. As shown in Fig. 6 for efficiency plots for \( \Delta L = 1\)–5 bps, application of Algorithm 1 (see Fig. 5) results in the narrowing of \( \Delta T_{50} \) for successive structures, while the accompanying \( T_f \) values simultaneously approach \( T_m \). For clarity, the trajectory taken by the algorithm during this process is also shown in Figs. 3 and 4. Algorithm 1 (see Fig. 5) produced an initial \( S_1 \) stem length of \( L_1 = 13 \) bps (\( T_m (1) = 61.0 \) °C) and converged after five iterations, yielding a first design with \( |L_1 - L_2| = 8 \) bps, \( \Delta T_{50} = 5\) °C, and operating characteristics: \( T_f \approx 60.4 \) °C, \( \sigma_T =12.0\) °C, and \( \epsilon_{\max} =0.0036 \). The detailed thermal profile of this untrimmed design is shown in the panel inset of Fig. 6. As shown in Table 1, T- and W-convergence were achieved after three and five iterations, respectively. For completeness, \( T_f \) values provided by (4a), along with numerical (exact) values for \( T_f \), \( \epsilon_{\max} \), \( T_\sigma \), and \( \Delta T_{50} \) are shown in Table 1 for the sequence-specific simulation in Fig. 2 as well as the mean-case simulations in Fig. 6. As expected, the agreement between \( T_f \) values produced by (4a) and numerical values is exact. As anticipated, numerical values obtained for \( \Delta T_{50} \) with fixed \( L_1 \) decrease consistently with decreasing \( L_2 \).

Table 1 also shows both exact and approximate values for \( \log_{10} \epsilon_{\max} \) provided by (8) and (8), respectively. Although the iterative 3′-word-5′-trim of Algorithm 1 (see Fig. 5) converged successfully, it also resulted in a log-linear reduction in \( \epsilon_{\max} \) as expected from the scaling behaviour of (8). The magnitude of this efficiency loss was predicted at \( \Delta \log_{10} | \epsilon_{\max} | \simeq -0.50 \) per step by \( \epsilon(T_f) \), which was reasonably well approximated by (8) as \( \epsilon(T_f) \). The individual contributions due to the base \( L_2 \) decrease and \( n_2 \) increase inherent in each 3′-word-5′-trim were resolved into approx. \(-0.43\) and \(-0.014\) per step, respectively, via appropriate

**Fig. 5** Algorithm for thermal band-pass filter design

---

This is an open access article published by the IET under the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/).
application of (11) and (14), showing that the latter change makes only a minor contribution of about 3% to the net efficiency decrease, as expected.

The effect of the final sp3-trim is shown in Fig. 7a. W-convergence was maintained throughout, resulting in complete loss of sp3, yielding a net increase in $\epsilon_{\max} > 200\%$, and a finalised design with $L_1 = 13$ bps and $\Delta L = 4$ bps, and $T^\ast = 60.4 \, ^\circ\text{C}$, $\Delta T_{50} = 12.1 \, ^\circ\text{C}$, and $\epsilon_{\max} = 1.09 \times 10^{-3}$. As shown in Fig. 7b, although (8) was less accurate, (14) with an initial seed value from $\epsilon(T')$ provided a nearly exact estimate $[213\% \text{ (approx.) against } 207\% \text{ (exact)]}. The invariance of $\Delta T_{50}$ to this trim also suggests that it may be invariant to $\Delta S_2$.

4. Discussion: By exploiting the functional behaviour of $T'$ with respect to its tendency to approach $T_m(1)$ with an iterated decrease in $L_2$, and its invariance to changes in $\Delta S_2$, which suggests the effectiveness of an $\epsilon_{\max}$-optimising final sp3 trim, Algorithm 1 (see Fig. 5) establishes the feasibility of algorithmic device design. Based on the success of the design process in Figs. 6 and 7, this method is expected to be effective in producing filters with tailored $T'$ and $\Delta T_{50}$ values, and optimised $\epsilon_{\max}$ values, given reasonably narrow target widths. Nevertheless, the scaling behaviour and tunability limits of $\Delta T_{50}$ require further investigation. While a monitored stepwise reduction in $n_2$, as implemented via the final sp3 trim of Algorithm 1 (see Fig. 5) is well motivated given lack of information regarding the functional form of $\Delta T_{50}$, its insensitivity to the trim suggests that like $T'$, it may be invariant to $\Delta S_2$. If so, this operation could be replaced with simple elimination of $n_2$ regions with no other function, which would aid the design of device variants with different word placement. Accordingly, derivation of a closed-form expression for $\Delta T_{50}$ is required to clarify these points. In addition, the co-tuning of $n_1$ with $L_1$ during initialisation of ZP remains an unexplored avenue to achieve more flexible device tuning, which bears investigation.

A practical consideration is the need to produce specific encodings for the mean-case designs generated by Algorithm 1 (see Fig. 5). Fortunately, selecting DNA sequences to encode oligonucleotide helices with desired $\Delta F$, $\Delta S$, and $T_m$ values is known to be easy [14]. Accordingly, the selection of device encodings which closely satisfy a set of thermodynamic characteristics determined via mean-case simulations is not likely to be challenging. Another issue involves the practicability of filter applications, given a relatively low $\epsilon_{\max}$. As shown in Table 1, the narrowing of $\Delta T_{50}$ implemented by Algorithm 1 (see Fig. 5) carries the cost of a reduced occupancy of $S_2$ in solution. However, the very high copy number typical of DNA experiments indicates that this efficiency level will be adequate, in practice. For instance, the wet-lab implementation in [3] employed $\geq 1.2 \times 10^{13}$ DNA strands immobilised onto streptavidin beads in a 400 μL reaction volume. An equivalent implementation of the designed filter operating at $T^\ast = 60.4 \, ^\circ\text{C}$ at $\epsilon_{\max} = 1.09 \times 10^{-3}$ would contain a plentiful $1.31 \times 10^6$ properly folded target structures at equilibrium, which would suffice for applications that do not require a large excess of product.

As noted in Section 1, the engineering of partner folds to implement exotic behaviours, such as a thermal band-pass filter, is largely unexplored. In this regard, the bulk equilibrium model developed herein ($\epsilon, T'$, $\epsilon_{\max}$) is general in scope for multistate biopolymer engineering. Accordingly, this model has potential applications beyond the current device, for engineering the behaviours of partner folds in other contexts and biopolymers (e.g. RNA, proteins).

5. Acknowledgment: The authors are grateful to Prof M. Yamamura (Tokyo Institute of Technology) and Prof A. Suyama (University of Tokyo) for their kind arrangements on the present study. This work was financially supported, in part by a JSPS Kakenhi Grant-in-Aid for Challenging Exploratory Research 26540151, MEXT Kakenhi for Scientific Research on Innovative Areas 24104003, and a Grant-In-Aid from Ritsumeikan APU.

6. References

[1] Bath J., Turberfield A.J.: ‘DNA nanomachines’, Nat. Nanotechnol., 2007, 2, pp. 275–284
[2] Liedl T., Sobey T.L., Simmel F.C.: ‘DNA-based nanodevices’, Nano Today, 2007, 2, pp. 36–41
[3] Komiya K., Sakamoto K., Gouzu H., *et al.*: ‘Successive state transitions with I/O interface by Molecules’, *Lect. Notes Comput. Sci.*, 2001, **2054**, pp. 17–26

[4] Takinoue M., Suyama A.: ‘Hairpin-DNA memory using molecular addressing’, *Small*, 2006, **2**, pp. 1244–1247

[5] Jonstrup A., Fredsøe J., Andersen A.: ‘DNA Hairpins as temperature switches, thermometers and ionic detectors’, *Sensors*, 2013, **13**, pp. 5937–5944

[6] Paun G., Rozenberg G., Yokomori T.: ‘Hairpin languages’, *Int. J. Found. Comput. Sci.*, 2001, **12**, pp. 837–847

[7] Rose J.A., Deaton R.J., Suyama A.: ‘Statistical thermodynamic analysis and design of oligonucleotide based computers’, *Nat. Comput.*, 2004, **3**, pp. 443–459

[8] Schultes E.A., Bartel D.P.: ‘One sequence, two ribozymes: implications for the emergence of new ribozyme folds’, *Science*, 2000, **287**, pp. 448–452

[9] Komiya K., Yamamura M., Rose J.A.: ‘Quantitative design and experimental validation for a single-molecule DNA nanodevice transformable among three structural states’, *Nucl. Acids Res.*, 2010, **38**, pp. 4539–4546

[10] Flamm C., Hofacker I.L., Maurer-Stroh S., *et al.*: ‘Design of multistable RNA molecules’, *RNA*, 2001, **7**, pp. 254–265

[11] Narberhaus F., Waldmninghaus T., Chowdhury S.: ‘RNA thermometers’, *FEBS Microbiol. Rev.*, 2006, **30**, pp. 3–16

[12] Rose J.A., Komiya K.: ‘Analysis and design of a single-molecule DNA nanodevice for thermal band-pass filters’. Proc. 11th IEEE Annual Int. Conf. on Nano/Micro Engineered and Molecular Systems (NEMS), 2016

[13] Cantor C., Shimmel P.: ‘Biophysical chemistry, part III’ (W. H. Freeman, New York, 1983)

[14] Santalucia J., Hicks D.: ‘The thermodynamics of DNA structural motifs’, *Annu. Rev. Biophys. Biomol. Struct.*, 2004, **33**, pp. 415–440

[15] Goddard N.L., Bonnet G., Krichevsky O., *et al.*: ‘Sequence dependent rigidity of single stranded DNA’, *Phys. Rev. Lett.*, 2000, **85**, pp. 3400–3403

[16] Rose J.A., Suyama A.: ‘Physical modeling of biomolecular computers: Models, limitations, and experimental validation’, *Nat. Comput.*, 2004, **3**, pp. 411–42