Supporting Information

Aviran et al. 10.1073/pnas.1106541108

SI Text

Convexity and Simplification of the Optimization Problems. Before we start, we note that we replace n' with n throughout this section for notational convenience. To simplify the presentation, we also treat the *third* optimization subproblem first.

In the third optimization problem, we fix Γ and c at their most recent estimates, $\Gamma^\star$ and $c^\star$, respectively, and seek a probability distribution Θ that maximizes $\log \mathscr{L}(\Theta,\Gamma^\star,c^\star)$. We call this problem **P3** and establish its strong convexity in the following lemma.

Lemma 1. ***P3** is a feasible convex optimization problem.*

Proof: It is easy to see that maximizing $\log \mathscr{L}(\Theta,\Gamma^\star,c^\star)$ over Θ is equivalent to maximizing

$$\mathscr{F}(\Theta)=\sum_{k=1}^{n} X_k\left[c^\star\left(\sum_{l=k+1}^{n}\theta_l-1\right)+\log(e^{c^\star\theta_k}-(1-\gamma_k^\star))\right]. \quad \textbf{[S1]}$$

We now show that $\mathscr{F}(\Theta)$ is strictly concave in Θ on its feasible convex domain $\mathscr{D}_{\mathscr{F}}=\{(\theta_1,\ldots,\theta_n)|\theta_k>\frac{1}{c^\star}\log(1-\gamma_k^\star)\ \forall\ 1\le k\le n\}$. This can be seen from Eq. **S1** by observing that $\mathscr{F}(\Theta)$ is a sum of a linear function of Θ and of log functions, where each logarithm depends only on one θ_k. For convenience, we denote the latter functions by $f_k(\Theta)=f_k(\theta_k)=\log(e^{c^\star\theta_k}-(1-\gamma_k^\star))$ and infer their strict concavity from the sign of $f_k''(\theta_k)=\frac{-(1-\gamma_k^\star)(c^\star)^2e^{c^\star\theta_k}}{f_k(\theta_k)^2}<0$. We also wish to stress that $\mathscr{F}(\Theta)$'s domain includes negative θ_k's because $\log(1-\gamma_k^\star)\le\log(1)$, with a strict inequality for at least one k. Finally, the convexity and feasibility of **P3** follow directly from the linearity of the imposed equality and inequality constraints (1).

Next, we provide the complete details of the proof of Theorem 3 in the main text.

Proof of Theorem 3: For a feasible convex optimization problem, the Karush–Kuhn–Tucker (KKT) constraints provide necessary and sufficient conditions on its unique solution (1). From Lemma 1's proof, it follows that the KKT conditions for **P3** take the form

$$\theta_k^\star\ge 0,\qquad \lambda_k^\star\ge 0,\qquad \lambda_k^\star\theta_k^\star=0,\qquad k=1,\ldots,n,$$
$$\sum_{k=1}^{n}\theta_k^\star=1, \quad \textbf{[S2]}$$

$$\frac{\partial\mathscr{F}(\Theta^\star)}{\partial\theta_k}+\lambda_k^\star-\nu^\star=0,\qquad k=1,\ldots,n \quad \textbf{[S3]}$$

where $\Theta^\star$ is the optimal solution, the $\lambda_k^\star$'s and $\nu^\star$ are the KKT multipliers that solve Eqs. **S2** and **S3**, and $\mathscr{F}(\Theta)$ is defined in Lemma 1's proof. Eq. **S3** can then be explicitly written as

$$c^\star\left[\sum_{i=1}^{k-1}X_i+\frac{X_k}{1-(1-\gamma_k^\star)e^{-c^\star\theta_k^\star}}\right]+\lambda_k^\star=\nu^\star,\qquad 1\le k\le n. \quad \textbf{[S4]}$$

Clearly, if the solution of the original, less-constrained problem is nonnegative, then it is also the sought solution and satisfies the KKT conditions with $\lambda_k^\star=0$ for all k.

Now, assume without loss of generality (*w.l.o.g.*) that $X_k>0$. Consider the case where $\theta_k^\star>0$, implying that $\lambda_k^\star=0$ and that the following equation must hold:

$$c^\star\sum_{i=1}^{k-1}X_i+X_kf_k'(\theta_k^\star)=\nu^\star, \quad \textbf{[S5]}$$

where we use the previously introduced notation $f_k(\theta)=\log(e^{c^\star\theta}-(1-\gamma_k^\star))$ and where $f_k'(\theta)$ stands for $f_k(\theta)$'s derivative. The left-hand side of Eq. **S5** consists of a nonnegative constant and of a strictly monotonously decreasing function of $\theta_k^\star$, as follows from Lemma 1's proof and as illustrated in Fig. S1*A*. For $\theta_k^\star>0$, $f_k'(\theta_k^\star)$ is bounded by $[f_k'(\infty)f_k'(0)]=[c^\star,\frac{c^\star}{\gamma_k^\star}]$, and, therefore, given $\nu^\star\in(c^\star\sum_{i=1}^{k}X_i,c^\star(\sum_{i=1}^{k-1}X_i+\frac{X_k}{\gamma_k^\star}))$, one can solve Eq. **S5** for a positive $\theta_k^\star$. The solution then satisfies

$$e^{c^\star\theta_k^\star}=(1-\gamma_k^\star)\frac{\nu^\star-c^\star\sum_{i=1}^{k-1}X_i}{\nu^\star-c^\star\sum_{i=1}^{k}X_i}>1. \quad \textbf{[S6]}$$

However, if $\nu^\star$ exceeds the value of the right boundary point of the interval $[c^\star,\frac{c^\star}{\gamma_k^\star})$, then $\theta_k^\star>0$ is impossible, and thus we set $\theta_k^\star=0$ and $\lambda_k^\star=\nu^\star-c^\star(\sum_{i=1}^{k-1}X_i+\frac{X_k}{\gamma_k^\star})\ge 0$. Note that if $\gamma_k^\star=0$, $f_k'(\theta_k^\star)$ is unbounded over $\mathbb{R}^+$ and so any $\nu^\star>c^\star\sum_{i=1}^{k}X_i$ yields a positive solution. Also note that the explicit expression for $\hat{\Theta}$ that is computed by Algorithm 1 is obtained when plugging $\nu^\star=\hat{c}\sum_{i=1}^{n+1}X_i$ into Eq. **S6**, with $c^\star=\hat{c}$ and $\gamma_k^\star=\hat{\gamma}_k$.

For notational convenience, we introduce the constants $\alpha_k=c^\star(\sum_{i=1}^{k-1}X_i+\frac{X_k}{\gamma_k^\star})$, $1\le k\le n$, and obtain the following relations between $e^{c^\star\theta_k^\star}$ and $\nu^\star$:

$$e^{c^\star\theta_k^\star}=\begin{cases}1 & \text{if } \nu^\star\ge\alpha_k\\ (1-\gamma_k^\star)\frac{\nu^\star-c^\star\sum_{i=1}^{k-1}X_i}{\nu^\star-c^\star\sum_{i=1}^{k}X_i} & \text{if } c^\star\sum_{i=1}^{k}X_i<\nu^\star<\alpha_k\\ \text{undefined} & \text{if } \nu^\star\le c^\star\sum_{i=1}^{k}X_i\end{cases}. \quad \textbf{[S7]}$$

Note that when $\nu^\star\le c^\star\sum_{i=1}^{k}X_i$, there is no feasible solution to the KKT conditions because the $\theta_k^\star$ that solves Eq. **S5** lies outside of $\mathscr{F}(\Theta)$'s domain. As will become clear later, this regime is irrelevant to the problem's solution. Hence, we restrict attention to $\nu^\star>c^\star\sum_{i=1}^{k}X_i$ and simplify Eq. **S7** to

$$e^{c^\star\theta_k^\star}=\max\left\{1,(1-\gamma_k^\star)\frac{\nu^\star-c^\star\sum_{i=1}^{k-1}X_i}{\nu^\star-c^\star\sum_{i=1}^{k}X_i}\right\}. \quad \textbf{[S8]}$$

Next, we seek $\nu^\star$ such that $\sum_{k=1}^{n}\theta_k^\star=1$, or alternatively, that $e^{c^\star\sum_{k=1}^{n}\theta_k^\star}=e^{c^\star}$. Using Eq. **S8**, we can write

$$\prod_{k=1}^{n}\max\left\{1,(1-\gamma_k^\star)\frac{\nu^\star-c^\star\sum_{i=1}^{k-1}X_i}{\nu^\star-c^\star\sum_{i=1}^{k}X_i}\right\}=e^{c^\star\sum_{k=1}^{n}\theta_k^\star}=e^{c^\star}. \quad \textbf{[S9]}$$

The product in Eq. **S9** forms a piecewise continuous and monotonously decreasing function of $\nu^\star$, with breakpoints at the finite α_k's. To observe these properties, we assume *w.l.o.g.* that all α_k's

are finite and inspect the function's behavior when starting at $\nu^\star_{ub} = \max_{1\le k\le n}\{\alpha_k\}$ and then gradually *decreasing* $\nu^\star$. Clearly, the product yields 1 at $\nu^\star_{ub}$, but once $\nu^\star < \nu^\star_{ub}$, it increases due to at least one site k_0 for which $\nu^\star < \alpha_{k_0}$ and which contributes $(1-\gamma^\star_k)\frac{\nu^\star - c^\star\sum_{i=1}^{k_0-1} X_i}{\nu^\star - c^\star\sum_{i=1}^{k_0} X_i} > 1$ to the total product. As we continue decreasing $\nu^\star$, additional terms increase their contribution to more than 1, with transitions taking place at the α_k's. Moreover, each of these "flexible" terms forms a strictly monotonously decreasing function of $\nu^\star$ (once it exceeds 1) and so it increases its contribution as $\nu^\star$ decreases. Importantly, each increasing term reflects a continuous increase in its corresponding θ. This way, the growth of the total product facilitates the expansion of certain $\theta^\star_k$'s while keeping others fixed at 0. Once $\sum_{k=1}^n \theta^\star_k = 1$, Eq. **S9** holds and a unique assignment for $\Theta^\star$ is determined as follows: Suppose the intersection occurs at $\alpha_l \le \nu^\star_{opt} < \alpha_m$, then $\theta^\star_k = 0$ for all k such that $\alpha_k \le \alpha_l$, and $\theta^\star_k = \frac{1}{c^\star}[\log\frac{\nu^\star_{opt} - c^\star\sum_{i=1}^{k-1} X_i}{\nu^\star_{opt} - c^\star\sum_{i=1}^{k} X_i} + \log(1-\gamma^\star_k)]$ otherwise. By replacing $\sum_{i=1}^k X_i$ with $\sum_{i=1}^{n+1} X_i - \sum_{i=k+1}^{n+1} X_i$ we obtain the expression for $\theta^\star_k$ in Theorem 3 in the main text. At this point, we wish to note that this solution technique is known as *water filling* (1, 2), and thus we call the product in Eq. **S9** the *water-filling function*. The technique is illustrated in Fig. S1*B* and is visualized as the flooding of a region with varying surface levels up to a constant amount of water.

Finally, we argued earlier that the domain $\nu^\star \le c^\star\sum_{i=1}^k X_i$ is irrelevant to the solution. We also mentioned that the formula for the reactivities that is computed by Algorithm 1 forms a special case of Eq. **S6** when $\nu^\star = c^\star\sum_{i=1}^{n+1} X_i$. Now, if all estimated reactivities are nonnegative when $\nu^\star = c^\star\sum_{i=1}^{n+1} X_i$, then $\nu^\star$ is the desired threshold. However, if some reactivities are negative, then the sum of the positive ones must exceed 1, and hence $\nu^\star$ should be increased to more than $c^\star\sum_{i=1}^{n+1} X_i$ until the sum reaches 1. This implies that $\nu^\star \ge c^\star\sum_{i=1}^{n+1} X_i > c^\star\sum_{i=1}^{k} X_i\ \forall k$ and also justifies setting $\theta^\star_k = 0$ whenever $X_k = 0$, because Eq. **S3** now takes the form $c^\star\sum_{i=1}^{k-1} X_i + \lambda^\star_k = \nu^\star$ and must hold with $\lambda^\star_k > 0$.

The *second* subproblem, hereafter called **P2**, entails the maximization of $\log\mathscr{L}(\Theta^\star,\Gamma,c^\star)$ when both $\Theta^\star$ and $c^\star$ are kept fixed and when Γ lies in the unit hypercube $[0,1]^n$. **P2**'s strong convexity is straightforward, as shown below.

Lemma 2. ***P2*** *is a convex optimization problem.*

Proof: From the expression for the likelihood function in the main text we have

$$\begin{aligned}\log\mathscr{L}(\Theta^\star,\Gamma,c^\star) = &\sum_{k=1}^{n} Y_k\left[\sum_{i=1}^{k-1}\log(1-\gamma_i) + \log\gamma_k\right]\\ &+\sum_{k=1}^{n} X_k\left[c^\star\left(\sum_{l=k+1}^{n}\theta^\star_l - 1\right) + \sum_{i=1}^{k-1}\log(1-\gamma_i)\right.\\ &\left.+\log(e^{c^\star\theta^\star_k} - 1 + \gamma_k)\right] + (X_{n+1}+Y_{n+1})\\ &\times\sum_{i=1}^{n}\log(1-\gamma_i) - c^\star X_{n+1}, \qquad \textbf{[S10]}\end{aligned}$$

which consists of constants and of a positively weighted sum of logarithms of affine functions of Γ. It is well known that convexity is preserved under affine transformation and that logarithms are strictly concave (1), thus establishing the log-likelihood's concavity. This fact, together with the imposed box constraints, completes the proof.

We are now in a position to prove Theorem 2 in the main text.

Proof of Theorem 2: A key observation here is that **P2** can be transformed into n independent *unconstrained* optimization problems, each in one variable. This is because $\log\mathscr{L}(\Theta^\star,\Gamma,c^\star)$ is separable in $\gamma_1,\ldots,\gamma_n$ as follows:

$$\begin{aligned}\log\mathscr{L}(\Theta^\star,\Gamma,c^\star) = \sum_{k=1}^{n}\left[\left(\sum_{i=k+1}^{n+1}(X_i+Y_i)\right)\log(1-\gamma_k) + Y_k\log\gamma_k\right.\\ \left.+X_k\log(e^{c^\star\theta^\star_k} - 1 + \gamma_k)\right] + C, \qquad \textbf{[S11]}\end{aligned}$$

where C is a constant. For simplicity, we introduce the constants $S_k = \sum_{i=k+1}^{n+1}(X_i+Y_i)$ and functions

$$l_k(\gamma_k) = S_k\log(1-\gamma_k) + Y_k\log\gamma_k + X_k\log(e^{c^\star\theta^\star_k} - 1 + \gamma_k) \qquad \textbf{[S12]}$$

for $1\le k\le n$, such that $\log\mathscr{L}(\Theta^\star,\Gamma,c^\star) = C + \sum_{k=1}^n l_k(\gamma_k)$. Now, not only each function can be optimized separately, but the box constraints can be removed as well, as long as $Y_k > 0$ or $X_k > 0$ jointly with $\theta^\star_k = 0$. To see this, note that $S_k > Y_{n+1} > 0$ and hence $\log(1-\gamma_k)$ serves as a natural barrier that upper bounds $l_k(\gamma_k)$'s domain at 1. Similarly, $Y_k > 0$ (or $X_k > 0$, $\theta^\star_k = 0$) imposes zero as a lower bound via $\log\gamma_k$. Optimization in the absence of the $\log\gamma_k$ term is also straightforward, as explained below.

Assume *w.l.o.g.* that $X_k, Y_k, \theta^\star_k > 0$ and consider setting

$$l'_k(\gamma_k) = -\frac{S_k}{1-\gamma_k} + \frac{Y_k}{\gamma_k} + \frac{X_k}{e^{c^\star\theta^\star_k} - 1 + \gamma_k} = 0, \qquad \textbf{[S13]}$$

which reduces to a quadratic equation when $\gamma_k \notin \{0,1,1-e^{c^\star\theta^\star_k}\}$ (recall that $e^{c^\star\theta^\star_k} > 1$). $l_k(\gamma_k)$'s concavity asserts that the equation's solution (if it exists) corresponds to its global maximum. Now, $l'_k(\gamma_k)$ is well-defined and continuous over (0, 1) and approaches ∞ near 0 and $-\infty$ near 1. Therefore, it must cross zero inside the unit interval at

$$\begin{aligned}\gamma^\star_k = \frac{1}{2(X_k+Y_k+S_k)}\left[X_k + Y_k - T_k(S_k+Y_k)\right.\\ \left.+\sqrt{[X_k+Y_k-T_k(S_k+Y_k)]^2 + 4T_kY_k(X_k+Y_k+S_k)}\right], \qquad \textbf{[S14]}\end{aligned}$$

where $T_k = e^{c^\star\theta^\star_k} - 1 > 0$ and where

$$X_k + Y_k + S_k = \sum_{i=k}^{n+1}(X_i+Y_i) = S_{k-1} \qquad \textbf{[S15]}$$

stands for the total count of fragments that contain nucleotide k. When $Y_k = 0$ but $\theta^\star_k > 0$, the solution simplifies to $\gamma^\star_k = \frac{X_k - T_kS_k}{X_k+S_k}$, but $l_k(\gamma_k)$'s domain is now extended into $\mathbb{R}^-$ with $l'_k(\gamma_k)\to\infty$ near $-T_k$. For this reason, we set $\gamma^\star_k = \max\{0, \frac{X_k - T_kS_k}{X_k+S_k}\}$ in this case. The last case of interest is when $\theta^\star_k = 0$ (i.e., $T_k = 0$), and then $\gamma^\star_k = \frac{X_k+Y_k}{X_k+Y_k+S_k} > 0$, which generalizes the initial estimate $\hat\gamma_k = \frac{Y_k}{\sum_{i=k}^{n+1} Y_i}$ to data that were aggregated from both channels.

The *first* optimization problem entails setting $(\Theta,\Gamma) = (\Theta^\star,\Gamma^\star)$ and seeking a positive c that maximizes $\log\mathscr{L}(\Theta^\star,\Gamma^\star,c)$. We refer to this problem as **P1** and establish its strong convexity next.

Lemma 3. ***P1*** *is a convex optimization problem.*

Proof: The concavity of $\log\mathscr{L}(\Theta^\star,\Gamma^\star,c)$ with respect to c can be seen by following the same derivation as in Lemma 1. Specifically, it suffices to maximize

$$\mathscr{H}(c) = \mathscr{F}(\Theta^\star,\Gamma^\star,c) - cX_{n+1}, \quad \textbf{[S16]}$$

where $\mathscr{F}(\Theta^\star,\Gamma^\star,c)$ stands for $\mathscr{F}(\Theta = \Theta^\star)$, with $c = c^\star$ (see Eq. **S1**). One can readily see that $\mathscr{H}(c)$ consists of a sum of a linear function of c and of the functions $h_k(c) = \log(e^{c\theta_k^\star} - (1-\gamma_k^\star))$ for all k for which $\theta_k^\star \neq 0$. Because each of the latter functions is symmetric with respect to θ_k and to c, their second-order derivatives with respect to c are all negative (see Lemma 1), and so is the derivative of their positively weighted sum. This establishes $\mathscr{H}(c)$'s concavity. At this point, it is worth noting that $\mathscr{H}(c)$'s domain takes the form $\mathscr{D}_\mathscr{H} = \{c|c > c_{lb}\}$, where $c_{lb} = \max_{k:\ \theta_k^\star \neq 0}\{\frac{1}{\theta_k^\star}\log(1-\gamma_k^\star)\}$. We remind the reader that $\log\mathscr{L}(\Theta,\Gamma,c)$ with $Y_{n+1} > 0$ constrains the γ_k's to $0 < 1-\gamma_k \leq 1$, and therefore $c_{lb} < 0$ if and only if $\gamma_k^\star > 0$ for each k for which $\theta_k^\star \neq 0$. In addition, because the logarithmic terms are not well-defined whenever $\theta_k^\star = \gamma_k^\star = 0$, the function pertains only to those positions where $\theta_k^\star \neq 0$ or $\gamma_k^\star \neq 0$. Importantly, we do not expect to encounter positions where $\theta_k^\star = \gamma_k^\star = 0$. This is because when $\theta_k^\star = 0$, the kth term in problem **P2** becomes $\log\gamma_k^\star$, thus ensuring $\gamma_k^\star > 0$ whenever $X_k \neq 0$. If $X_k = 0$, then we have $Y_k \neq 0$ (recall that all double-zero positions are excluded from optimization), in which case the term $Y_k\log\gamma_k^\star$ serves the same purpose. Finally, note that the constraint $c > 0$ does not comply with the standard (stricter) definition of a convex optimization problem, which involves weak inequalities, i.e., $c \geq 0$ (1). Yet, this subtlety is accounted for in the next lemma, and we refer to **P1** as a convex optimization problem in the broader sense.

Problem **P1** does not pose a significant computational challenge as it involves one variable. Yet, its solution can be further simplified by relaxing the constraint $c > 0$. This gives rise to an unconstrained optimization problem, where a concave $\mathscr{H}(c)$ is maximized over the semibounded interval $\mathscr{D}_\mathscr{H}$. Because $\mathscr{D}_\mathscr{H}$ may contain negative values, the relaxation might result in a negative maximizing argument. The following lemma shows that a maximum for the unconstrained problem is indeed attained within its semibounded domain and that it can be readily used to infer **P1**'s solution.

Lemma 4. *$\mathscr{H}(c)$ attains a unique maximum over its domain $\mathscr{D}_\mathscr{H} = \{c|c > c_{lb}\}$. If* $\arg\max\mathscr{H}(c) \leq 0$*, then problem **P1** is infeasible. Otherwise,* $\arg\max\mathscr{H}(c)$ *is also the solution of **P1**.*

Proof: We first demonstrate that $\mathscr{H}(c)$ attains a maximum over $\mathscr{D}_\mathscr{H}$ by showing that its derivative crosses zero inside the domain. The derivative takes the form

$$\mathscr{H}'(c) = \sum_{k=1}^{n} X_k\left[\sum_{l=k+1}^{n}\theta_l^\star + \frac{\theta_k^\star}{1-(1-\gamma_k^\star)e^{-c\theta_k^\star}} - 1\right] - X_{n+1} \quad \textbf{[S17]}$$

and is well-defined and continuous over $\mathscr{D}_\mathscr{H}$. When c approaches c_{lb} from the right, we have $\frac{\theta_k^\star}{1-(1-\gamma_k^\star)e^{-c\theta_k^\star}} \to \infty$ for at least one position k with $\theta_k^\star \neq 0$, implying that $\mathscr{H}'(c) > 0$ for sufficiently small c values. In contrast, when $c \to \infty$, we have $\frac{\theta_k^\star}{1-(1-\gamma_k^\star)e^{-c\theta_k^\star}} \to \theta_k^\star$ for each k with $\theta_k^\star > 0$ (and a zero-term otherwise), and therefore $\mathscr{H}'(c) \to \sum_{k=1}^{n} X_k(\sum_{l=k}^{n}\theta_l^\star - 1) - X_{n+1}$. The expression on the right-hand side must be negative, because $\sum_{l=k}^{n}\theta_l^\star \leq 1$ with strict inequality for some k's and because we expect to have $X_{n+1} > 0$ (see *Modeling Chemical Modification* in the main text for reasoning). It now follows from the intermediate value theorem that $\mathscr{H}'(c^\star) = 0$ for some $c^\star \in \mathscr{D}_\mathscr{H}$, which is where the unique maximum is located. Finally, $\mathrm{sign}(c^\star)$ determines whether $c^\star$ solves **P1** as well. In case it is negative, we have $\mathscr{H}(c) < \mathscr{H}(0)$ for all $c > 0$, and because $\mathscr{H}(c)$ monotonously decreases over $\mathbb{R}^+$ **P1** is infeasible. Otherwise, $c^\star$ maximizes $\mathscr{H}(c)$ over $\mathbb{R}^+$ and is **P1**'s sought solution.

Concavity of the log-Likelihood in the Entire Parameter Space. Although the decomposition of the global optimization problem into local optimization steps lends itself to an efficient semianalytic solution, it is also of interest to investigate the function's concavity with respect to the entire coordinate system. Such global concavity has two implications: first, every local maximum is also global, and thus Algorithm 2 should converge to a global maximum even when initialized at a single starting point. Second, it facilitates the solution of the entire optimization problem using numerical interior-point methods. Yet, it is not obvious that the latter approach is advantageous over Algorithm 2, as it might be more computationally intensive and/or yield less accurate solutions. Hence, even if the function is globally concave, the performance of interior-point methods needs to be carefully examined and compared to that of Algorithm 2's.

Before we study $\log\mathscr{L}(\Theta,\Gamma,c)$ as a function of all parameters, we set $c = c^\star$ and investigate its concavity with respect to (Θ,Γ). We can further simplify analysis by following the lines of Theorem 2's proof and observing that $\log\mathscr{L}(\Theta,\Gamma,c^\star)$ is separable in the pairwise variables (θ_k,γ_k) as follows:

$$\log\mathscr{L}(\Theta,\Gamma,c^\star) = \sum_{k=1}^{n} l_k(\theta_k,\gamma_k) + L(\Theta) + C. \quad \textbf{[S18]}$$

Here, $L(\Theta)$ represents a linear function of Θ, whereas C and $l_k(\theta_k,\gamma_k)$ pertain to the constant and functions that were introduced in Theorem 2's proof, respectively. Specifically,

$$l_k(\theta_k,\gamma_k) = S_k\log(1-\gamma_k) + Y_k\log\gamma_k + X_k\log(e^{c^\star\theta_k} - 1 + \gamma_k), \quad \textbf{[S19]}$$

where $S_k = \sum_{i=k+1}^{n+1}(X_i + Y_i)$. Because $\log\mathscr{L}(\Theta,\Gamma,c^\star)$ is concave in (Θ,Γ) if and only if $l_k(\theta_k,\gamma_k)$ is concave in (θ_k,γ_k) for all k, it suffices to investigate the properties of the two-dimensional function $l_k(\theta_k,\gamma_k)$. We start by showing that its right-hand term, that is,

$$f_k(\theta_k,\gamma_k) = \log(e^{c^\star\theta_k} - (1-\gamma_k)), \quad \textbf{[S20]}$$

is *not* concave in (θ_k,γ_k). Before we proceed, we stress that $f_k(\theta_k,\gamma_k)$'s domain, consisting of all (θ_k,γ_k) such that $\gamma_k > 1 - e^{c^\star\theta_k}$, does *not* form a convex set. This can be seen from the fact that the domain consists of the epigraph of the strictly concave function $\gamma_k(\theta_k) = 1 - e^{c^\star\theta_k}$. In light of the importance of a domain's convexity in determining a function's concavity (1), we artificially restrict the function's domain to the following convex subset of the original domain, which contains the entire feasible parameter space:

$$\mathscr{DS} = \{(\theta_k,\gamma_k):\ 0 \leq \theta_k \leq 1, 0 \leq \gamma_k \leq 1\}.$$

Lemma 5. *The function $f_k:\mathbb{R}^2 \to \mathbb{R}$, with domain $\mathscr{DS} = \{(\theta_k,\gamma_k):\ 0 \leq \theta_k \leq 1, 0 \leq \gamma_k \leq 1\}$ and given by $f_k(\theta_k,\gamma_k) = \log(e^{c^\star\theta_k} - (1-\gamma_k))$, is not concave in (θ_k,γ_k).*

Proof: Because $f_k(\theta_k,\gamma_k)$ is continuous and twice differentiable in its domain, its concavity can be inferred from the negative semidefiniteness of its Hessian on the interior of $\mathscr{DS}$. The Hessian is given by

$$H_{f_k} = -[e^{c^\star\theta_k} - (1-\gamma_k)]^{-2}\begin{bmatrix} c^{\star 2}e^{c^\star\theta_k}(1-\gamma_k) & c^\star e^{c^\star\theta_k} \\ c^\star e^{c^\star\theta_k} & 1 \end{bmatrix}, \quad \textbf{[S21]}$$

and when multiplied by $V = (x,y)$ from both sides we obtain

$$VH_{f_k}V^T = -[e^{c^\star\theta_k} - (1-\gamma_k)]^{-2}[c^{\star 2}e^{c^\star\theta_k}(1-\gamma_k)x^2 + y^2 + 2c^\star e^{c^\star\theta_k}xy] \quad \textbf{[S22]}$$

$$= -[e^{c^\star\theta_k} - (1-\gamma_k)]^{-2}[(c^\star e^{c^\star\theta_k}x + y)^2 + c^{\star 2}e^{c^\star\theta_k}(1-\gamma_k - e^{c^\star\theta_k})x^2]. \quad \textbf{[S23]}$$

Consider (x,y) such that $x \neq 0$ and $c^\star e^{c^\star\theta_k}x + y = 0$, in which case

$$VH_{f_k}V^T = -[e^{c^\star\theta_k} - (1-\gamma_k)]^{-2}c^{\star 2}e^{c^\star\theta_k}(1-\gamma_k - e^{c^\star\theta_k})x^2 \quad \textbf{[S24]}$$

$$= [e^{c^\star\theta_k} - (1-\gamma_k)]^{-1}c^{\star 2}e^{c^\star\theta_k}x^2 > 0, \quad \textbf{[S25]}$$

where the positive sign follows from the fact that $f_k(\theta_k,\gamma_k)$ is defined only when $e^{c^\star\theta_k} - (1-\gamma_k) > 0$ and because $c^\star > 0$. Eq. **S25** implies that H_{f_k} is not negative semidefinite and hence $f_k(\theta_k,\gamma_k)$ is not concave. Note that H_{f_k} is not positive semidefinite as well, because a choice of $V = (0,y)$ with $y > 0$ yields a negative product in Eq. **S22**. Therefore, $f_k(\theta_k,\gamma_k)$ is neither convex nor concave on the domain of interest.

Importantly, the fact that $f_k(\theta_k,\gamma_k)$ is not concave does not imply that $l_k(\theta_k,\gamma_k) = S_k\log(1-\gamma_k) + Y_k\log\gamma_k + X_k f_k(\theta_k,\gamma_k)$ is not concave as well, because the first two terms are strictly concave functions. In what follows, we attempt to investigate $l_k(\theta_k,\gamma_k)$'s concavity by examining its Hessian, which takes the form

$$H_{l_k} = -\begin{bmatrix} \frac{X_k c^{\star 2}e^{c^\star\theta_k}(1-\gamma_k)}{[e^{c^\star\theta_k}-(1-\gamma_k)]^2} & \frac{X_k c^\star e^{c^\star\theta_k}}{[e^{c^\star\theta_k}-(1-\gamma_k)]^2} \\ \frac{X_k c^\star e^{c^\star\theta_k}}{[e^{c^\star\theta_k}-(1-\gamma_k)]^2} & M \end{bmatrix}, \quad \textbf{[S26]}$$

where $M = \frac{X_k}{[e^{c^\star\theta_k}-(1-\gamma_k)]^2} + \frac{S_k}{(1-\gamma_k)^2} + \frac{Y_k}{\gamma_k^2}$. We now exploit the fact that H_{l_k} is two-dimensional and hence its determinant's sign is indicative of its definiteness. If the sign is negative, then the two eigenvalues must be of different signs and the matrix is neither positive nor negative semidefinite, or equivalently, the function is neither concave nor convex. The determinant is given by

$$\det H_{l_k} = \frac{X_k c^{\star 2}e^{c^\star\theta_k}}{[e^{c^\star\theta_k} - (1-\gamma_k)]^2}\left[\frac{-X_k}{e^{c^\star\theta_k} - (1-\gamma_k)} + \frac{S_k}{1-\gamma_k} + \frac{Y_k(1-\gamma_k)}{\gamma_k^2}\right], \quad \textbf{[S27]}$$

and if $X_k > 0$ then

$$\operatorname{sign}(\det H_{l_k}) = \operatorname{sign}\left(\frac{-X_k}{e^{c^\star\theta_k} - (1-\gamma_k)} + \frac{S_k}{1-\gamma_k} + \frac{Y_k(1-\gamma_k)}{\gamma_k^2}\right). \quad \textbf{[S28]}$$

Recall that $S_k > 0$ and let us assume for simplicity that X_k and Y_k are also positive. Recall also that we are concerned with all the points in $\mathcal{DS}$'s interior where $0 < \theta_k < 1$ and $0 < \gamma_k < 1$. We thus wish to explore the existence of a feasible point for which the expression in Eq. **S28** assumes a negative value. Although this expression reduces to a cubic equation in γ_k, finding its real root analytically is intractable. Instead, we observe that $\frac{X_k}{e^{c^\star\theta_k}-(1-\gamma_k)}$ depends on both θ_k and γ_k, whereas the other two terms depend only on γ_k. Additionally, this term increases in magnitude as $\gamma_k \to 0$ and as $e^{c^\star\theta_k} \to 1$. However, we also have $\frac{Y_k(1-\gamma_k)}{\gamma_k^2} \to \infty$ when $\gamma_k \to 0$ and $\frac{S_k}{1-\gamma_k} \to \infty$ when $\gamma_k \to 1$. This means that for any given $c^\star\theta_k$ value, $\frac{X_k}{e^{c^\star\theta_k}-(1-\gamma_k)}$ needs to offset an increasing (and unbounded) positive term at both limits of the interval $0 < \gamma_k < 1$. It is therefore not obvious whether there exist θ_k and γ_k for which $\frac{-X_k}{e^{c^\star\theta_k}-(1-\gamma_k)} + \frac{S_k}{1-\gamma_k} + \frac{Y_k(1-\gamma_k)}{\gamma_k^2} < 0$.

Instead, we resort to considering the limiting case where $\theta_k = 0$ (on $\mathcal{DS}$'s boundary), because this is where $\frac{X_k}{e^{c^\star\theta_k}-(1-\gamma_k)}$ is maximized for each γ_k. Here, we can test whether the simplified condition

$$\frac{-X_k}{\gamma_k} + \frac{S_k}{1-\gamma_k} + \frac{Y_k(1-\gamma_k)}{\gamma_k^2} < 0 \quad \textbf{[S29]}$$

holds for some $0 < \gamma_k < 1$. By converting the latter condition into the equivalent inequality

$$(S_k + Y_k + X_k)\gamma_k^2 - (2Y_k + X_k)\gamma_k + Y_k < 0, \quad \textbf{[S30]}$$

it is easy to show that it is satisfied within some subinterval of (0, 1), provided that

$$X_k^2 > 4Y_k S_k. \quad \textbf{[S31]}$$

This is because its left-hand side is positive at $\gamma_k \in \{0,1\}$ and its minimum is attained at $0 < \bar{\gamma}_k = \frac{2Y_k + X_k}{2(S_k + Y_k + X_k)} < 1$. Now, if the expression's discriminant is positive, it must cross zero twice inside (0, 1) and attain negative values between the two intersection points. Finally, constraining the discriminant in this manner yields inequality **S31**. When **S31** is satisfied, we can choose a sufficiently small $c^\star\theta_k$ such that $e^{c^\star\theta_k} - 1 < \epsilon$ and $\frac{X_k}{\gamma_k} \approx \frac{X_k}{e^{c^\star\theta_k}-(1-\gamma_k)}$. This, in turn, guarantees that $\frac{-X_k}{e^{c^\star\theta_k}-(1-\gamma_k)} + \frac{S_k}{1-\gamma_k} + \frac{Y_k(1-\gamma_k)}{\gamma_k^2} < 0$ for the chosen values.

When **S31** does not hold, it follows that $\frac{X_k}{e^{c^\star\theta_k}-(1-\gamma_k)} < \frac{X_k}{\gamma_k} \leq \frac{S_k}{1-\gamma_k} + \frac{Y_k(1-\gamma_k)}{\gamma_k^2}$, or that $\operatorname{sign}(\det H_{l_k}) > 0$. Notably, a positive determinant implies that $l_k(\theta_k,\gamma_k)$ is either convex or concave, but because $f_k(\theta_k,\gamma_k)$ is neither concave nor convex (see the proof of Lemma 5) and $\log(1-\gamma_k)$ and $\log\gamma_k$ are strictly concave, $l_k(\theta_k,\gamma_k)$ must also be concave in this case. Finally, when **S31** does *not* hold for every $1 \leq k \leq n$, we can conclude that $\log\mathscr{L}(\Theta,\Gamma,c^\star)$ is concave in (Θ,Γ).

Our analysis thus suggests that the global concavity of $\log\mathscr{L}(\Theta,\Gamma,c^\star)$ is data-dependent, as captured by the condition in **S31**. In light of this condition, we consider the case where $Y_k = 0$ but $X_k > 0$, which, in our experience, is commonly observed, albeit for very few positions. For such positions, **S31** clearly holds, thereby eliminating the log-likelihood's global concavity. Although one may arbitrarily set $Y_k = 1$ for such positions, **S31** may or may not be satisfied following such correction, depending on the corresponding values of X_k and S_k.

As final remarks, our analysis considered the domain $\mathcal{DS} = \{(\theta_k,\gamma_k)\colon 0 \leq \theta_k \leq 1, 0 \leq \gamma_k \leq 1\}$, but, in fact, optimization takes place over a convex subset of this domain, in which $\sum_{k=1}^n \theta_k = 1$. It is possible that under a fixed-sum restriction, $\log\mathscr{L}(\Theta,\Gamma,c^\star)$ exhibits global concavity. However, analysis of the type we conducted above does not apply in this case due to two reasons: (*i*) only directions $V = (v_1,\ldots,v_n)$ such that $\sum_{k=1}^n v_k = 0$ are feasible under the additional restriction, and (*ii*) one cannot leverage on the separability of $\log\mathscr{L}(\Theta,\Gamma,c^\star)$ into two-dimensional functions, as these are now linked by the fixed-sum constraint. Therefore, such analysis exceeds the scope of this work and is a topic for

future research. In addition, we did not include analysis of concavity in (Θ,Γ,c) (i.e., when c is also allowed to vary), because lack of concavity in a lower dimension implies the lack of it in the entire coordinate system.

Implementation Details. The following technical detail pertains to the analysis procedure:

1. Each nucleotide included in the molecular construct probed by the SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) experiment should be associated with reactivity and natural drop-off parameters. This is because nucleotides outside of the studied domain, such as those found in the structure cassette and bar codes, may react with the agent. This effectively increases n to a few tens more than the length of the investigated RNA strand.

Below we provide additional details regarding efficient implementation of our proposed approach to numerical maximum-likelihood (ML) estimation. Our goal in pointing these out is twofold: reducing running time and facilitating the usage of accessible, general-purpose, optimization software.

1. The implementation of the water-filling (WF) routine in problem **P3** can be simplified by incorporating the following observations:

- Before calling the WF routine at each iteration, it is worthwhile testing whether the explicit solution (i.e., Algorithm 1's result) is valid. This is because it is possible that a feasible solution becomes achievable following the updates made to $c^\star$ and $\Gamma^\star$. However, from a run-time perspective, the benefits are negligible because the entire WF step is computationally undemanding. Note that alternatively, one can sort all α_k's and verify whether any of them is smaller than $c^\star \sum_{i=1}^{n+1} X_i$. Each $\alpha_k < c^\star \sum_{i=1}^{n+1} X_i$ corresponds to a negative entry.
- It suffices to evaluate the WF function only at the breakpoints, and in fact, one needs only evaluate it for a subset of them. This is because the points can be sorted in increasing order such that the function monotonously decreases as we progress in the sorted list. One can start evaluating at the first α_k that is greater than $\nu^\star = c^\star \sum_{i=1}^{n+1} X_i$, at which point, all negative entries (whose α_k's are smaller than $c^\star \sum_{i=1}^{n+1} X_i$) are set to zero. As we progress along the evaluation points, additional θ_k's are being zeroed. Once $e^{c^\star}$ is crossed, evaluation is stopped.
- Such evaluation procedure also identifies the specific interval where intersection takes place. This allows one to narrow down the numerical search to this interval.
- Since the WF function is continuous, standard root-finding algorithms that do not require any derivative information (e.g., Matlab's *fzero* routine) can be used to efficiently perform the search.
- The analytical expression for the WF function

$$\prod_{k=1}^{n} \max\left\{1,(1-\gamma_k^\star)\frac{\nu^\star - c^\star \sum_{i=1}^{k-1} X_i}{\nu^\star - c^\star \sum_{i=1}^{k} X_i}\right\} = e^{c^\star} \qquad \text{[S32]}$$

can be vastly simplified by observing that for any two consecutive nonzero terms, $\theta_k^\star$ and $\theta_{k+1}^\star$, the kth denominator and the $(k+1)$th numerator cancel out and thereby decrease the degree of the polynomials in both the numerator and denominator. Hence, every run of consecutive nonzero $\theta_k^\star$'s yields an expression of the type $(1-\gamma_l^\star)(1-\gamma_m^\star)\frac{(\nu^\star - c^\star \sum_{i=1}^{l-1} X_i)}{(\nu^\star - c^\star \sum_{i=1}^{m} X_i)}$, where l is the smallest index in the run and m is the largest. The runs should be updated at every breakpoint, but allow one to maintain relatively low complexity of function evaluation.
- Positions for which $\gamma_k^\star = 0$ should not pose a numerical problem because their α_k's will end up at the bottom of the sorted breakpoint list, and water filling should be completed before reaching them.
- The complexity of the WF routine is quadratic in n as a worst-case scenario. This is because the number of breakpoints where the function is in fact evaluated is less than n, and if the data are of reasonable quality, this number is expected to be significantly less than n. At each evaluation point, we need to compute at most n factors as well as their product.

2. Problem **P1** also does not require any specialized convex or constrained-optimization software. Based on Lemmas 3 and 4, we used a general-purpose unconstrained optimization routine to find the unique optimum of the relaxed problem. Lemma 4 also informs us that problem **P1** may not always be feasible and that its feasibility can be easily checked by testing for the sign of $\mathscr{H}'(0) = \sum_{k=1}^{n} X_k[\sum_{l=k+1}^{n} \theta_l^\star + \frac{\theta_k^\star}{\gamma_k^\star} - 1] - X_{n+1}$. In fact, we can avoid solving the relaxed problem once we observe that $\mathscr{H}'(0) < 0$. Such observation suggests that the current combination of $\Theta^\star$ and $\Gamma^\star$ estimates does not make sense and could allude to problems in the dataset. However, it is also possible that the iterative routine is converging to a local optimum that is different from the global one. In an attempt to address such possibility and find another local minimum, we set $c = 1$ and continue to the next iteration.

Robustness Analysis Details. ***Robustness of estimates with respect to sequencing depth.*** Estimation accuracy at different sequencing depths was assessed by sampling our dataset (without replacement) to create smaller datasets, which were then analyzed using Algorithms 1 and 2. The sampling was done by randomly permuting the list of observed fragments and then recording the first m fragment lengths, where m is the subset size. This process was repeated 200 times per each sequencing depth (i.e., subset size), in order to assess the average performance and the variation around it. Accuracy was measured as the fraction of sites that were assigned reactivities within 15% of the full-dataset estimate, as depicted in Fig. S2. In addition, we verified that the estimated reactivity profile for smaller datasets still retains the main features of the full-dataset profile (see Fig. S3). We performed this analysis for both the RNase P and the pT181 molecules and observed similar size-dependence characteristics.

Sensitivity to approximation of repeated modification per site. To study the effects of the assumed approximation on estimation accuracy, we simulated many model scenarios in conjunction with the modified sites being sampled without replacement. This is in contrast to sampling with replacement, which amounts to the approximation we made when deriving the model-induced fragment length distribution. Because it is difficult to derive an explicit form for the induced fragment length distribution in this case (note that we accommodated any possible form of Θ), we used Monte Carlo simulations to estimate it to high accuracy. In these simulations, we set a maximum of seven modifications, because we used Poisson rates ranging from $c = 0.25$ to $c = 2$, and the Poisson-induced probabilities of having eight or more exposures are negligible for $c = 2$. Per each number of modifications i, we simulated successive drawings of i modified sites, while revising the sampling distribution after each draw accordingly. This sequential drawing was repeated 10^6 times to obtain an estimate of the probability of encountering the first adduct at site k given that there were i exposures. These estimates were then summarized over all considered numbers of modifications. For simplicity, we disregarded natural drop-off effects and then recovered the estimated reactivities from the estimated empirical length distribution, using Algorithm 1's formula. We then assessed the scheme's accuracy by computing the fraction of sites that were assigned reactivities within 10%, 5%, and 1% of the true reactivity. Intervals of 15% around the true Θ were also con-

sidered, but they consistently yielded 100% hits, which is indicative of overall low sensitivity to the approximation and is also the reason why we decreased the resolution of our accuracy measure.

We tested a variety of model scenarios. First, we assumed that Θ is uniform and changed the Poisson rate within the range $c = 0.25$–2. We also considered two uniform distributions: one with $n = 10$ reactive sites and one with $n = 50$ reactive sites. Table S1 shows that accuracy improves with decreasing modification rate and with increasing molecule size (more precisely, with increasing number of reactive sites). This is expected, because, in general, the statistics of sampling with and without replacement become similar when the number of sites to draw from is significantly larger than the number of draws (i.e., modifications). Second, for a uniform Θ, we varied n while keeping the rate fixed at $c = 1$ and at $c = 2$. One can see the same trend, where increased lengths lead to lower sensitivity, and that high quality is achieved for molecules with 40 or more reactive sites, a scenario which, in our experience, is realistic. Third, we considered four other Θ distributions, as follows: A decreasing exponential and an increasing exponential were chosen as extreme examples of unbalanced distributions, although we do not think they are representative of realistic reactivity profiles as these typically display some form of symmetry due to base pairing. A decreasing exponential of length $n = 50$ takes the form $a \times 0.2 \times (0.9)^i$ (a is a normalization contant), starting at approximately 0.1 and reaching very close to zero at $i = 50$, where the increasing exponential is its flipped version. We also generated two profiles that reflect our understanding of the crude form of a single hairpin ($n = 20$) and a concatenation of two hairpins ($n = 50$). It can be seen from Table S1 that the model had low sensitivity to the approximation under the first three distributions, whereas estimation was less accurate in the single hairpin case.

Robustness of estimates under different model distributions. We tested estimation quality under different model parameters by simulating a total of 5 million draws from the induced fragment length distributions in the (+) and (−) channels, and then processing the counts with Algorithms 1 and 2. We repeated this process 500 times. Here, we considered the average fraction of hits at resolutions of 15%, 10%, and 5% of the true Θ as well as the standard deviation around the average (reported in Table S2). We also computed the standard deviation of the estimated reactivity per site. The standard deviation (per site and per interval) was negligible is all cases considered. We first assumed that $\gamma_k = 0.01$ for all k and simulated the model for the five Θ distributions mentioned in the previous subsection as well as for a uniform, a decreasing exponential, and an increasing exponential, all of size $n = 10$. Subsequently, we relaxed the fixed-Γ assumption and added spikes of magnitude 0.15 on top of the $\gamma_k = 0.01$ background. The spike magnitude was chosen based on spikes we observed in the control experiments of the RNase P and the pT181 molecules. We tested the effects of two and four such spikes (distributed evenly across the molecule) when Θ is uniform and when Θ represents two concatenated hairpins. The results are summarized in Table S2.

Read Alignment Details. Reads for reverse transcriptase (RT) fragments were first split into 1M7-treated and untreated pools by examining the 4 nucleotide handle sequence on the 5′ end of the read generated from the 3′ end of each RNA probed in the experiment. Reads with an RRRY handle identified (+) fragments and those with YYYR identified (−) fragments. This handle was then trimmed from each read to allow alignment of the reads to probed RNAs. Reads were then trimmed for A_adapter_b and A_adapter_t using the FASTX toolkit, because RT products shorter than the length of a sequencing read will produce reads with adapter at their 3′ ends.

Paired reads were optimally aligned to the probed RNAs using Bowtie 0.12.8 (3) with the parameters (–best X 2000 y), and allowing no mismatches in the alignments with parameters (-v 0). Bowtie suppressed alignments for reads that mapped ambiguously to the probed RNAs (-m 1), though because the bar code in each read identified each read with its target RNA and alignments were required to be perfect, this step was not strictly necessary. The 3′ end of each fragment alignment (toward the 5′ end of the probed RNA) corresponds to the point at which RT stopped. Two counters (one for the 1M7 condition and one for the control condition) at each position of each probed RNA position were used to track RT stopping points. The (+) counter for a probed RNA at position i was incremented when a fragment from the (+) pool aligned to the RNA starting at position $(i + 1)$ and ending at the RNAs 3′ end. The (−) counter was incremented for (−) pool fragment alignments. These RT-stop counts were then used to calculate maximum likelihood reactivities.

Source Code. Source code to implement read mapping and ML estimation can be found at http://bio.math.berkeley.edu/SHAPE-Seq/ as either a .tar.gz or .zip file. After unpacking with the command `tar -xzvf Aviran_Trapnell_SHAPE-Seq_analysis_code_PNAS_2011.tar.gz` or `unzip Aviran_Trapnell_SHAPE-Seq_analysis_code_PNAS_2011.zip` a directory called `spats-0.0.1` will be created. Please follow instructions in the README file to compile and execute the read mapping code (spats). Matlab source code to implement the ML estimation can be found in `spats-0.0.1/src/matlab`.

1. Boyd S, Vandenberghe L (2004) *Convex Optimization*. (Cambridge Univ Press, Cambridge, UK).
2. Cover TM, Thomas JA (2006) *Elements of Information Theory*. (John Wiley, Hoboken, NJ), 2nd Ed.
3. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.

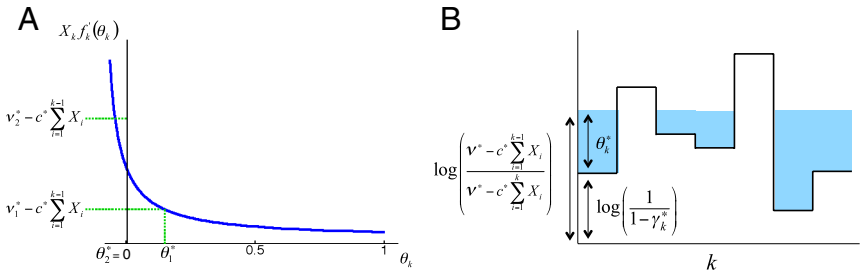

Fig. S1. Illustration of the use of water filling to optimize Θ. (*A*) Illustration of the function $f'_k(\theta_k)$ and of the solution obtained for a single KKT condition under two instances of ν^*. In the first instance ($\nu_1{}^*$), a strictly positive solution is obtained, whereas in the second instance ($\nu_2{}^*$), a zero reactivity is assigned. (*B*) Illustration of the water-filling technique following ref. 1. The height of each bar is computed from the corresponding parameter $\gamma_k{}^*$. The gradual decreasing of ν^* can be viewed as flooding the bars up to a level of $\log(\nu^* - c^* \sum_{i=1}^{k-1} X_i) - \log(\nu^* - c^* \sum_{i=1}^{k} X_i)$, until the total quantity of water equals c^*. The level of water above each bar (in light blue) corresponds to the optimal $\theta_k{}^*$.

1. Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge Univ Press, Cambridge, UK).

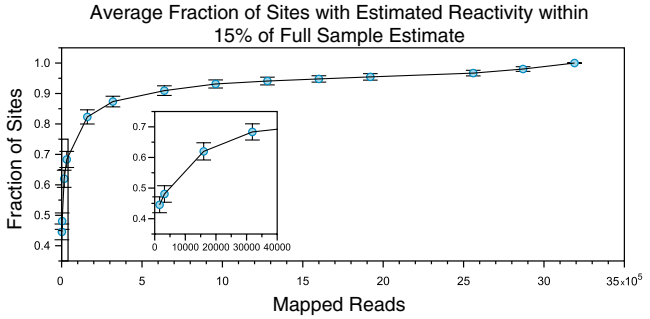

Fig. S2. Dependence of the ML estimates on the number of mapped reads for the pT181 molecule. Varying numbers of reads were drawn at random from the observed dataset and used to generate the ML reactivity estimate for each nucleotide. The fraction of these reactivities that were within 15% of the ML estimate of the full dataset was recorded. This was repeated 200 times for each specified number of reads. Points represent the average over these 200 draws and error bars represent standard deviations. Very accurate estimates are obtained with only 300,000 reads (10% of the full number of reads), and over half of the nucleotide positions show accurate reactivities with as little as 16,000 reads (0.5% of the full number of reads).

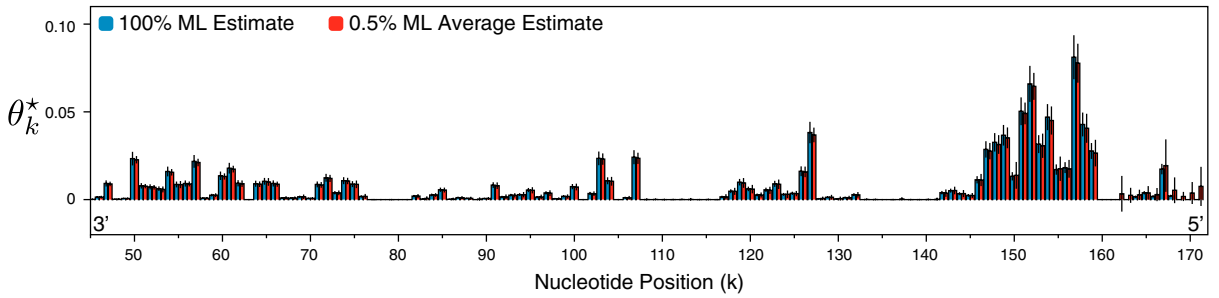

Fig. S3. Sample ML estimates based on a fraction of the observed data for the pT181 molecule. The full ML reactivity estimates for each nucleotide of the pT181 molecule (see Fig. 2 in the main text) were calculated from the 3.2 million mapped reads and are shown in blue with ±15% of the estimate plotted as error bars. The red bars represent an average of the ML reactivity estimate at each nucleotide, calculated from 200 independent random draws of 16,000 observed reads (0.5% of the total) with error bars representing the standard deviation of the estimate at each nucleotide.

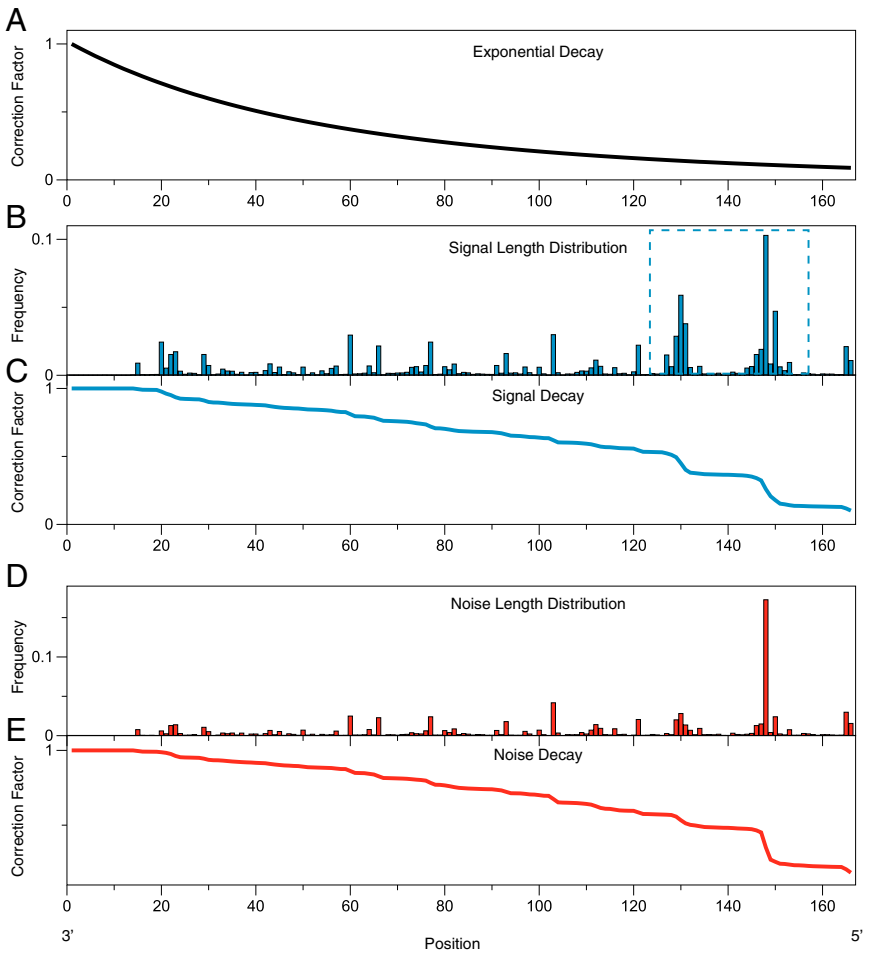

Fig. S4. Signal decay deviates from an exponential model for the specificity domain of *Bacillus subtilis* RNase P RNA. (*A*) The correction currently applied to SHAPE-capillary-electrophoresis (SHAPE-CE) output is based on assuming exponential-like signal decay. The panel depicts such decay, as obtained from our model when the true reactivities (Θ) are uniformly distributed over n' nucleotides and when the drop-off rate (γ_k) is constant and equals 0.005. Here, $n' = 166$ is the number of nucleotides for which k-fragments were observed, out of 198 nucleotides in the probed molecule. Out of the 198 nucleotides, the RNase P RNA spans 154, with the rest comprising one structure cassette and identifier bar codes. The support set includes 151 nucleotides from the RNase P domain, spanning indices 14 to 164. (*B*) The empirical (+) channel fragment length distribution (i.e., $\hat{w}_1,\ldots,\hat{w}_{n'}$) of RNase P does not display a clear trend of signal decay (from left to right). (*C*) The (+) channel signal decay is computed as the ratio between the input signal frequency and the corrected signal portion in Eq. **30** in the main text. The decay rate is clearly nonuniform, with the decay accelerating over a stretch of about 20 nucleotides toward the 5′ end (highlighted in a box). The accelerated decline reflects a highly reactive region that amounts to a significant portion of the distribution. (*D*) The empirical (–) channel fragment length distribution (i.e., $\hat{p}_1,\ldots,\hat{p}_{n'}$) of RNase P displays an outstanding spike near the 5′ end. (*E*) The noise decay is computed as the ratio between the input noise frequency and the corrected noise portion in Eq. **30** in the main text. The expedited decay near the 5′ end implies that the input noise frequencies should be significantly amplified in this region before they are subtracted from the corrected signal frequencies.

Table S1. Model sensitivity with respect to the approximation of repeated modification per site

n	Θ	c	Average fraction of reativities within ± 10% of known value	Average fraction of reativities within ± 5% of known value	Average fraction of reativities within ± 1% of known value
10	uniform	0.25	1	1	0.7
		0.50	1	1	0.4
		1.00	1	0.9	0.2
		1.50	1	0.6	0.1
		2.00	0.9	0.4	0.1
50		0.25	1	1	1
		0.50	1	1	1
		1.00	1	1	0.86
		1.50	1	1	0.72
		2.00	1	1	0.44
20		1	1	1	0.35
40			1	1	0.83
60			1	1	0.98
80			1	1	0.99
20		2	1	0.9	0.2
40			1	1	0.35
60			1	1	0.48
80			1	1	0.61
20	one hairpin	1	1	0.85	0.35
		2	0.95	0.55	0.15
50	uniform	1	1	1	0.86
	decreasing exponential		1	1	0.92
	increasing exponential		1	1	0.06
	two hairpins		1	1	0.54

Table S2. Robustness of estimates under different model distributions

			Fraction of reativities within ± 15% of known value		Fraction of reativities within ± 10% of known value		Fraction of reativities within ± 5% of known value	
n	Θ	Γ	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
50	uniform	fixed (0.01)	1	0	1	0	1	0
	decreasing exponential		0.91	0.03	0.86	0.04	0.76	0.04
	increasing exponential		0.97	0.02	0.93	0.03	0.83	0.04
	two hairpins		0.99	0.01	0.97	0.02	0.86	0.04
10	uniform	fixed (0.01)	1	0	1	0	1	0
	decreasing exponential		1	0	1	0	1	0
	increasing exponential		1	0	1	0	1	0
	one hairpin		1	0	1	0	0.99	0.02
50	uniform	two spikes (0.01,0.15)	1	0	1	0	0.99	0.01
		four spikes (0.01,0.15)	1	0	1	0	0.99	0.01
	two hairpins	two spikes (0.01,0.15)	0.98	0.02	0.95	0.03	0.83	0.04
		four spikes (0.01,0.15)	0.96	0.02	0.92	0.03	0.81	0.04