Taphonomical Security: (DNA) Information with a Foreseeable Lifespan

Fatima-Ezzahra El Orche, Marcel Hollenstein, Sarah Houdaigoui, David Naccache, Daria Pchelina, Peter B Rønne, Peter y A Ryan, Julien Weibel, Robert Weil

Taphonomical Security: (DNA) Information with a Foreseeable Lifespan

Fatima-Ezzahra El Orche[1,2], Marcel Hollenstein[3] Sarah Houdaigoui[1], David Naccache[1], Daria Pchelina[1], Peter B. Rønne[2], Peter Y.A. Ryan[2], Julien Weibel[1], and Robert Weil[4]

[1] ENS, CNRS, PSL Research University, Paris, France
Département d'informatique, École Normale Supérieure Paris,
first_name.family_name@ens.fr
[2] SnT, FSTC, University of Luxembourg
{fatima.elorche,peter.roenne,peter.ryan}@uni.lu
[3] Institut Pasteur, Paris
marcel.hollenstein@pasteur.fr
[4] Sorbonne Universiy, Inserm, UMR1135, CNRS, ERL8255, CIMI, Paris, France
robert.weil@pasteur.fr

Abstract This paper introduces the concept of information with a foreseeable lifespan and explains who to achieve this primitive via a new method for encoding and storing information in DNA-RNA sequences.
The storage process can be divided into three time-frames. Within the first (life), we can easily read out the stored data with high probability. The second time-frame (agony) is a parameter-dependent state of uncertainty; the data is not easily accessible, but still cannot be guaranteed to be inaccessible. During the third (death), the data can with high probability not be recovered without a large computational effort which can be controlled via a security parameter. The quality of such a system, in terms of a foreseeable lifespan, depends on the brevity of the agony time-frame, and we show how to optimise this.
In the present paper, we analyse the use of synthetic DNA and RNA as a storage medium since it is a suitable information carrier and we can manipulate the RNA nucleotide degradation rate to help control the lifespan of the message embedded in the synthesized DNA/RNA molecules. Other media such as Bisphenol A thermal fax paper or unstable nonvolatile memory technologies can be used to implement the same principle but the decay models of each of those phenomena should be re-analysed and the formulae given in this paper adapted correspondingly.

Keywords: Cryptography, Information with foreseeable lifespan, Data Storage, Information theory, DNA, RNA.

1 Introduction

Over time, the physical media on which we store information degrades. Traditionally, much effort has been put into *protecting* media against degradation to achieve more robust and durable storage mechanisms.

In this paper, instead of resisting the time's unavoidable effects, we try to *exploit* them: rather than allowing information to slowly and progressively get destroyed, we aim at a swift and complete erasure. Just as a thermal fax machine paper that fades with time, we propose to synthesize DNA and RNA molecules whose lifetime can be approximately tuned. Such a "time fuse" can guarantee, for instance, that a cryptographic secret (typically a plaintext encrypted under a hash of the DNA information) cannot be used or recovered beyond some expiry date.

Since DNA is a reasonably stable molecule, we assume in this paper that DNA does not degrade at all. By contrast, RNA nucleotides quickly decay over time. We hence propose to synthetically incorporate RNA nucleotides in DNA molecules. The DNA nucleotides will store the cryptographic secret whereas RNA will serve as a natural countdown mechanism. This technique guarantees that, with high probability, the whole secret will be recoverable before some target time $t_{\text{target }1}$, but will not be reconstructible after $t_{\text{target }2}$. A mathematical analysis allows tuning the t_i as a function of the molecules' molecular decay probability distribution and the storage environment parameters such as temperature and exposure to radiation.

Structure of the paper. We start (Section 2) by a general overview of the biochemical notions necessary to understand the concept. The method itself is presented in Section 3. In Section 4, we introduce a probabilistic model and mathematically determine bounds on the t_is. We analyse our method's efficiency and explain how to tune the t_is in Section 5. Appendix A provides proofs and Appendix B, lists numerical values.

2 Biochemical Preliminaries

For the article to be understandable, it is necessary to present a few biochemical notions about DNA and RNA [16]. The following sections will deal with DNA and RNA composition and degradation. The acronym "NA" (Nucleic Acid) will denote both DNA and/or RNA.

2.1 NA Composition

DNA and RNA belong to the category of NAs, which are bio-macro-molecules; both are chains of nucleotides. A nucleotide is composed of a nucleobase, a pentose sugar and one phosphate group. In nature, there exist five different nucleotides: adenine (A), thymine (T), uracil (U), cytosine (C) and guanine (G). DNA contains A,T,C,G, whereas RNA contains A,U,C,G. Figure 1 shows the structures of NAs, while Figure 2 details the four DNA nucleotides.

A fundamental difference between DNA and RNA is their pentose composition. DNA pentose has a deoxyribose sugar which has no substituent at position C2', whereas the RNA sugar is a ribose which contains a 2'-hydroxyl (OH) moiety as shown in Figure 3. Another chemical difference between DNA and RNA appears when considering hydrolysis of RNA in buffer: cleavage can occur by

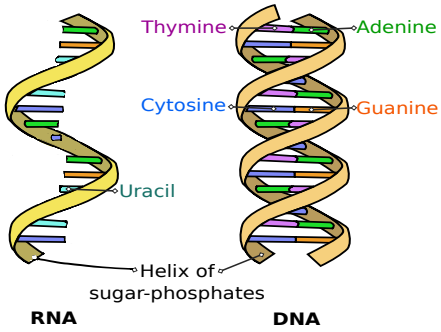

Figure 1: RNA and DNA structures

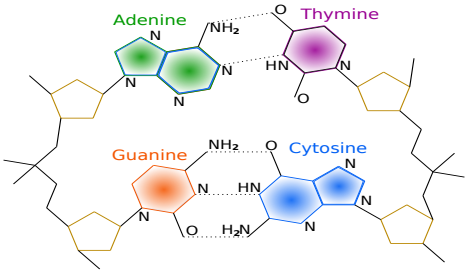

Figure 2: Closeup structure of DNA with the four nucleotides represented: Adenine, Thymine, Cytosine, and Guanine.

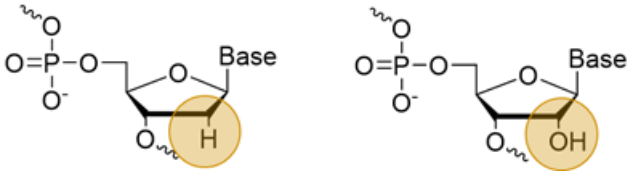

Figure 3: RNA and DNA pentose.

an intramolecular attack of the 2'-OH unit which is on the sugar moiety on the phosphorous center of the phosphodiester unit. Since DNA lacks a 2'-OH and RNA has a 2'-OH such a reaction is favored in RNA as displayed by the differences in the rates of uncatalyzed hydrolysis [15] (see Section 2.2).

2.2 NA Bonds Degradation Over Time

NAs degrade over time. The main degradation reaction of RNA nucleotides is called *transesterification*, while the main degradation phenomena for DNA are *phosphodiester hydrolysis*, *oxidative cleavage*, and *cleavage* as a result of *depurination*. Whilst we will not dive deeper into the particularities of those biochemical processes, the reader may wonder why the same degradation reactions do not apply to both DNA and RNA. This results from the fact that the difference in the pentose has a considerable influence on the reactions leading to degradation. Briefly, in DNA a hydroxide ion (OH -) will attack the phosphorous center which eventually will lead to hydrolysis of the phosphodiester bond. Such a mechanism can also occur in RNA but since RNA displays a 2'-OH moiety, this hydroxyl group can be activated (by deprotonation under basic conditions or by metal coordination) and attack the phosphorous center in an intramolecular rather than in an intermolecular manner [12,18]. In particular, there is a big difference between the two degradation speeds: under representative physiological conditions, RNA hydrolysis is 10^5 times faster than DNA hydrolysis. DNA degradation speed is hence almost negligible compared to the RNA degradation. In what follows, we

are chiefly interested in considering the RNA degradation in our mathematical analysis. Figure 4 illustrates the way in which RNA degrades.

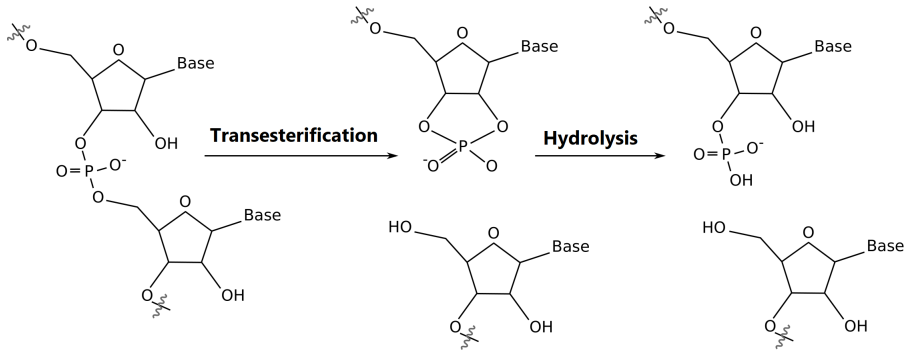

Figure 4: RNA degradation through transesterification

2.3 On the Synthesis of RNA-DNA Chimeric Oligonucleotides

Oligonucleotides in general can be synthesized by two main approaches: chemical synthesis based on solid-phase methods and enzymatic synthesis. In this section, we will briefly describe both methods and then highlight how DNA-RNA chimeric oligonucleotides required for our purposes could be made.

Chemical synthesis The main synthetic access to RNA and DNA oligonucleotides is granted by automated DNA and RNA solid-phase synthesis. In this approach, activated nucleoside units called phosphoramidites are sequentially added on a first nucleoside bound to a solid support. Each cycle encompasses a coupling step (where the incoming phosphoramidite is reacted with a free 5'-OH unit of a solid support bound nucleoside) followed by a capping step (to avoid reaction of unreacted hydroxyl moieties in subsequent steps). The coupling and capping steps are followed by oxidation of the newly created linkages (P(III) to P(V)) and removal of the next protecting group to enable continuation of the synthesis (the interested reader is directed to more comprehensive review articles dedicated to this topic [8, 11]). Such syntheses are usually carried out on synthesizers (Figure 5) on scales ranging from µmoles to moles. This method is routinely used to synthesize DNA and RNA oligonucleotides either based on standard chemistry or encompassing chemical modifications required for in vivo applications. However, this method is restricted to rather short (i.e. around 100-150 nucleotides for DNA and around 100 nucleotides for RNA [2, 8, 11] oligonucleotides due to low yields for fragments exceeding 100 nucleotides because of folding on solid support during synthesis and due to the inherent nature of the coupling yields (even with a 99% coupling efficiency, the maximum theoretical yield that can be

obtained for 100 nucleotide long sequence would be 0.99100 37%). Hence, this approach is ideal for synthesizing short DNA-RNA chimeric oligonucleotides but is unlikely to be applicable to longer sequences.

Figure 5: DNA synthesizer used for solid-phase synthesis of DNA and RNA oligonucleotides.

Enzymatic Methods The most popular enzymatic methods for the synthesis of oligonucleotides include polymerase-assisted synthesis using nucleoside triphosphates and ligation of shorter fragments into long oligonucleotides. In the first strategy, nucleoside triphosphates are recognized by enzymes called polymerases which add these nucleotides onto a growing chain of DNA or RNA. For DNA synthesis, the presence of a primer and a template are strictly required since the polymerase will add nucleotides at the 3'-end of the primer while the sequence composition of the template will dictate the polymerase which nucleotide needs to be incorporated. On the other hand, RNA polymerases are primer independent and only require the presence of a DNA template to mediate transcription of DNA into RNA. Such a method can be coupled with chemical modifications to generate mRNA vaccines [4, 13] and other functional nucleic acids [1, 7]. This method is not restricted to any size limitation and is compatible with numerous chemical modifications. On the other hand, the sequence specific incorporation of distinct RNA nucleotides in long DNA oligonucleotides will be difficult to achieve by this method.

In the second method, DNA or RNA ligases mediate the formation of phosphodiester linkages between the terminal 3'-OH residue of an oligonucleotide with the 5'-end (usually phosphorylated) of a second oligonucleotide [17] . Often a "splint" oligonucleotide is required as a template since this guide oligonucleotide is partially complementary to the termini of both oligonucleotides that need

to be ligated together. This method is compatible with the synthesis of longer oligonucleotides [14] as well as with different chemistries in oligonucleotides [5,10], and hence is deemed as the method of choice for this project.

Synthesis of RNA-DNA Chimeric Oligonucleotides To synthesize long DNA oligonucleotides containing RNA nucleotides at distinct and specific positions in the future we propose to synthesize short DNA-RNA sequences using solid phase synthesis and combine these fragments by (repeated) ligation reactions as highlighted in Figure 6. This protocol will circumvent the drawbacks associated with all the different methods and should yield the desired oligonucleotides.

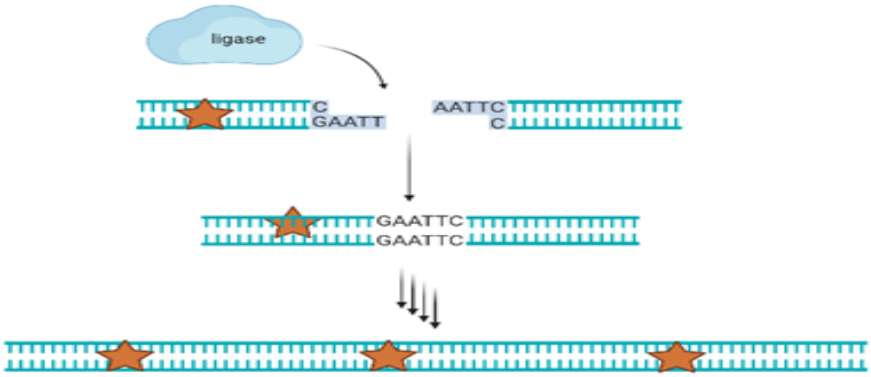

Figure 6: Schematic representation of the synthesis of long DNA oligonucleotides containing RNA nucleotides (star symbols) using a combination of solid-phase synthesis (short blue fragments) and DNA ligation reactions.

2.4 RNA Degradation in Further Detail

RNA nucleotide degradation probability follows an exponential distribution of parameter λ. λ depends on numerous factors such as temperature, pH and the concentration of some ions[5]. Degradation also depends on the sequence context and the 3D structure of the RNA oligonucleotide. We will use Equation (1) of [9] to model λ:

$$\lambda = \lambda_0 \cdot 10^e \cdot c_K \cdot [K^+]^{d_K} \cdot c_{\mathrm{Mg}} \cdot [\ \mathrm{Mg}^{2+}]^{d_{\mathrm{Mg}}} \tag{1}$$

where

$$e = a_{\mathrm{pH}}(\mathrm{pH} - b_{\mathrm{pH}}) + a_K([K^+] - b_K) + a_T(T - b_T)$$

$\lambda_0 = 1.3 \cdot 10^{-9}$ min^{-1} and the constants are indicated in Table 4 in Appendix B with their corresponding units.

Note that Equation (1) is an approximation and λ_0 needs to be updated depending on the considered range of physical parameters. See [9] for more details. We give in Table 7 in Appendix B several λ values under different conditions

[5] $[K^+]$, $[\mathrm{Mg}^{2+}]$

and the corresponding λ_0 values. The expected time for one RNA nucleotide to degrade (which is $1/\lambda$) is given in the table to get an idea of the order of magnitude of time. This will prove helpful in the rest of the paper for determining which λ to work with when tuning the information's lifetime. The chemical parameters mentioned in this table are taken from Table 1 and using Equation (e) from [9].

3 The Proposed Method

This section presents our new method for encoding and storing information using DNA and RNA nucleotide. We propose a way on how to synthesize a new DNA/RNA molecule, and we show that we can reach a good security level with our method.

3.1 Description of the Method

The idea is to incorporate RNA fragments into DNA oligonucleotides using standard solid-phase synthesis and produce DNA-RNA chimeric sequences to form a new DNA/RNA chimeric oligonucleotide. A DNA fragment can be composed of one nucleotide base (A, C, T, G) or by a juxtaposition of several nucleotides bases linked to each other $(AA, CC, TT, GG, AC, AT, GT, ACT, \cdots$ etc). The chain's length depends on the size of the key that we want to encode and store. This DNA/RNA chimeric oligonucleotide will contain k RNA nucleotides and $k+1$ DNA fragments. We synthesize n copies of this molecule and keep it in a fluid.

To understand the insertion/encoding mechanism, we refer the reader to Figure 7 which illustrates the insertion of the key `SECRET`. In this example, we have 5 RNA nucleotides and 6 DNA fragments and an alphabetic substitution for each letter in which we arbitrarily assigned different fragments to different letters of the English alphabet.

Note that below we will actually use distinct DNA molecule fragments and encode the stored information into the permutation of these.

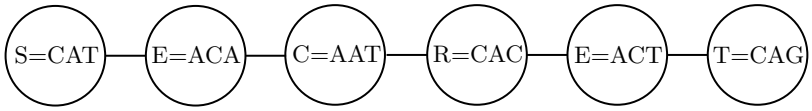

Figure 7: An explanatory illustration of one copy of the DNA/RNA oligonucleotide encoding the key `SECRET`. Beads represent DNA fragments and inter-bead links are RNA nucleotides.

3.2 Encrypting Information

Encrypt the information to be time-protected using some symmetric cipher (e.g. AES [3]) and encode the key as an DNA/RNA oligonucleotide using a permutation of pairwise distinct DNA fragments.

Erase the plaintext and the electronic version of the key and assume that the ciphertext is accessible by the opponent. Hence, as long as we can reconstruct the key from the DNA/RNA oligonucleotide the plaintext is recoverable.

We will now focus our attention on the recovery of the key from the DNA/RNA oligonucleotide, as long as this molecule is physically reconstructible.

3.3 Key Reconstruction

We assume that the DNA fragments are pairwise distinct by construction. Call the DNA/RNA oligonucleotide w and remember that it contains k RNA nucleotides. Remember as well that there are n copies of w floating in a liquid. Suppose that all the copies of w were cut randomly in pieces. We are given the set of these pieces, and we seek to restore the initial oligonucleotide w if such a reconstruction is still possible. Figure 8 represents the evolution of 3 copies of a key made of 9 DNA fragments.

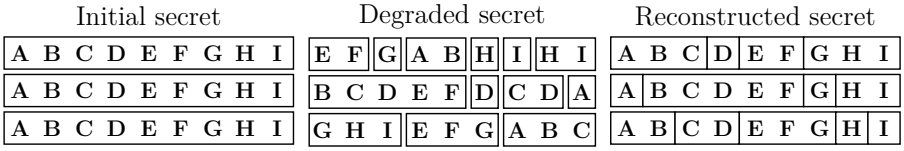

Figure 8: An evolution of a secret with 3 copies and 9 fragments in each copy

The following algorithm outlines how we can recover the information after it has begun degrading if such a reconstruction is possible at all.

- ① First, we can easily obtain all fragments of w by analysing the pieces we are given. For each fragment x, we will try to find the "next" fragment next$[x]$ (the one which follows x in the molecule w). If for all fragments except one (which is w's last fragment) the next fragment is found, we can restore w.
- ② For any two fragments x and y, if there exists a third piece where y follows x, then in the initial molecule w, y also follows x, and thus next$[x] = y$. Therefore, we have just to treat each piece as follows: for every fragment x except the last, define next$[x]$ as the fragment following x inside that piece. Since all fragments of w were distinct, there is at most one possible value of next$[x]$ for all fragments x. Figure 9 represents this relation in a graph.
- ③ After this procedure, we have to find a fragment which follows nothing, and if it's unique, we set it as the first and then add next fragments one by one until there is nothing to add. This allows us to reconstruct w. If there are several such fragments, w cannot be recovered without brute-force guessing.

If we got several fragments following nothing at the end of the algorithm, it is impossible to recover the initial molecule having no information except the input. We can only obtain separated pieces of w applying the last step of the algorithm

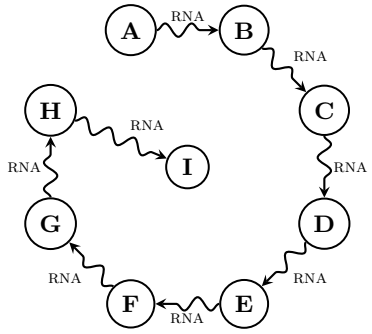

Figure 9: The graph representing the next$[x]$ relation in the algorithm

to each of the "first" fragments. This situation happens if and only if at least one *cut* occurred during degradation, i.e. in all molecules, the RNA nucleotides at the same specific location in all the molecule broke. The probability of this happening will be investigated in the next section.

3.4 Security

We now turn to measure our method's security, but before doing so, let us define what the term *cut* means.

Definition 1. *A* cut *happens at the i^{th} position if for all $1 \leq i \leq k$, the i^{th} bond of each of the n copies is broken.*

Assume that the secret contains cuts. For each cut, the next fragment after this cut follows nothing according to the reconstruction algorithm. Simultaneously, for all other fragments except the first one of the initial molecule, there is at least one piece where this fragment follows another one. This means that the only pieces which can be recovered are any piece delimited by two cuts or the pieces delimited by one cut and an endpoint of the initial molecule.

Since no link can be guessed between two described pieces, the only strategy to recover the whole initial molecule is to test all reconstructed pieces' permutations. If there are k cuts, then $(k+1)!$ possibilities need to be browsed. This fact defines security: a given number of cuts guarantees a security of $\approx$ 80, 100, 128, 256 bits. Table 3 (a trivial $\log_2(k+1)!$ lookup table given in Appendix B for the ease of quick reference) gives the correspondence between the number of security bits versus the number of cuts and DNA fragments. The number of security bits, which is called the security parameter and that we denote a, is simply: $a = \log_2(n_{\mathrm{D}}!)$, where n_{D} is the number of the DNA fragments.

Current biological limitations. It is currently technically feasible to have an NA chain of about 100 nucleobase pairs [2, 8, 11]. Each DNA chunk is linked by an RNA pair. In the security analysis in this paper we assume that each DNA

chunk is unique, since copies reduce the security (and would makes the following analysis harder).

Since the security stems directly from the number of RNA fragments, we should construct chains containing the maximal number of RNA bonds while observing distinctness of the DNA fragments.

Hence, we start by generating all 1-digit integers in base 4 (there are $u_1 = 4$ of them), then all two digit integers in base 4 (there are $u_2 = 4^2 = 16$ of them) and finally we will fill in with 3-digit integers in base 4.

Linking the $u_1 + u_2 = 20$ 1- and 2-digit pairs requires $u_1 + u_2 - 1 = 19$ RNA nucleobase pairs. Hence, all in all we are already at a molecule comprising:

$$u_1 + 2u_2 + u_1 + u_2 - 1 = 4 + 2 \times 16 + 4 + 16 - 1 = 55 \text{ pairs}$$

To proceed, the 45 remaining pairs, permitted by the current technological synthesis capacity, must be constructed using 3-digit integers linked with 1 RNA bond.

We can hence solve $45 \geq 3u_3 + u_3$ to get $u_3 = 11$ (we need one RNA bond for each 3-digit fragment and one RNA to bind to the rest of the string).

The security level of the resulting scheme is $\log_2((u_1 + u_2 + u_3)!) = \log_2 31! \simeq 113$ bits.

Note that it is possible to artificially construct new types of DNA molecules, see [6] where two new types have been constructed. Assuming that we have 6 different types of DNA molecules at hand the analysis above can be repeated. First we have 6 1-digit integers in base 6. We now have 6 1-digit integers in base 6 and 36 2-digit integers in base 6. We first choose $u'_1 = 6$ and solve $100 - (u'_1 + u'_1 - 1) = 89$ molecules left. We then solve $89 \geq 2u'_2 + u'_2$ to get $u'_2 = 29$. In this case don't need 3-integers. In total we now have 34 RNA bonds, $\log_2((u'_1 + u'_2)!) = \log_2 35! \simeq 133$ bits.

4 Controlling the Information Lifetime

In order to understand and control the lifetime of the information embedded in the DNA/RNA molecules, we introduce a probabilistic model and mathematically determine the bounds on the information lifetime.

4.1 Probabilistic Model

Recall that k denotes the number of RNA nucleotides in each of the n identical copies of the initial molecule w. Denote by $L_{i,j}$ the random variable giving the degradation time of the j^{th} RNA nucleotide of the i^{th} copy. Our main assumption is that the $L_{i,j}$, for all i, j, are independent and identically distributed random variables following the exponential distribution of parameter λ. Denote by T_j the random variable representing the time for the cut at the j^{th} position to appear and by t_x the random variable giving the x^{th} cut time to appear. t_x is the x^{th} order statistic of $(T_j)_{1 \leq j \leq k}$, i.e. the x^{th} smallest element of $\{T_1, \cdots T_k\}$. By definition, $T_j = \max\limits_{1 \leq i \leq n} L_{i,j}$ and in compactified notation, $t_x = T_{(x)}$.

4.2 The Information Lifetime Bounds

We consider that the information stored in the NA molecule goes through three different periods that we call: *life*, *agony* and *death*. In this section, we describe each period separately, and give the mathematical model allowing us to determine the bounds of each one of them. Figure 10 shows the different periods represented on a time axis:

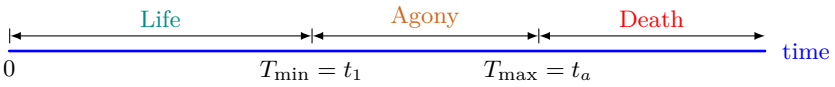

Figure 10: The information lifespan

Life: The information embedded in the NA molecule is fully accessible during the first phase. This happens when no cut has occurred, i.e. for $t \in [0, t_1[$. Let $T_{\min} = \min_{j \leq k} T_j$ be the random variable giving the time at which the first cut to occur. We have $t_1 = T_{min}$ and :

$$\mathbb{P}(T_{\min} > t) = (1 - (1 - \exp(-\lambda t))^n)^k \tag{2}$$

Agony: Agony starts after the first cut has appeared. We can only recover the information by at least brute-force guessing. For each guess, the probability p that a guess gives the correct secret is equal to $\frac{1}{(x+1)!}$, where x is the number of cuts at the time t. Agony ends when the a^{th} cut appears, i.e. we have $t \in [t_1, t_a]$.

Death: After the a^{th} cut, we consider the information to be *dead* – it is no longer feasible to brute-force a recovery of the information. We have $t_a = T_{\max}$ and:

$$\mathbb{P}(T_{\max} \leq t) = 1 - \sum_{i=0}^{a-1} \binom{k}{i} p(t)^i (1 - p(t))^{k-i} \tag{3}$$

with $p(t) = (1 - \exp(-\lambda t))^n$. In this case $t \in]t_a, \infty[$ and it is computationally infeasible within the chosen security parameter to recover the secret information.

Proof. See Appendix A for the derivations of Equation (2) and Equation (3).

We note that $\mathbb{P}(T_{\min} > t)$ is increasing in n and decreasing in k, t and λ, and $\mathbb{P}(T_{\max} \leq t)$ is increasing in k, in $p(t)$ and thus in t and λ, but decreasing in n.

Lemma 1 (Evolution of the number of cuts over time). *Let $C(t)$ denote the number of cuts at the time t. C is a random variable and we have:*

$$\mathbb{E}[C(t)] = k \times (1 - \exp(-\lambda t))^n$$

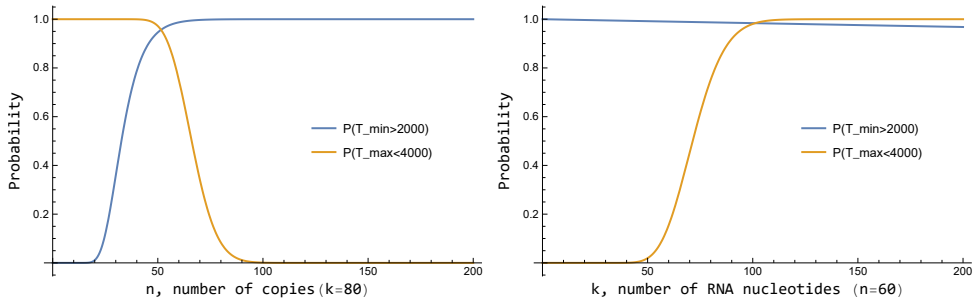

Figure 11: Probabilities as functions of the number of copies n, with a fixed number of RNA bonds $k = 80$ (left), and as functions of the number of RNA bonds k with a fixed number of copies $n = 60$ (right). Here $a = 24$ and $\lambda = 0.001$ min^{-1}.

Proof. See Appendix A

The following Lemma allows us to calculate the expected life spans and their variance.

Lemma 2 (The average time and the variance for the x^{th} cut to appear). *The average time when the x^{th} cut appears, $\mathbb{E}(t_x)$, and the corresponding variance, $V(t_x)$, are given by the following formulas:*

$$\mathbb{E}(t_x) = \frac{1}{\lambda}\sum_{s=1}^{kn} C_x(k,n,s), \mathbb{V}(t_x) = \frac{1}{\lambda^2}\left[2\sum_{s=1}^{kn}\frac{C_x(k,n,s)}{s} - \left(\sum_{s=1}^{kn} C_x(k,n,s)\right)^2\right]$$

where:

$$C_x(k,n,s) = \frac{(-1)^{s+1}}{s}\sum_{m=x}^{k}\sum_{p=0}^{s}\sum_{i=0}^{k-m}(-1)^i\binom{k}{m}\binom{mn}{p}\binom{k-m}{i}\binom{ni}{s-p}$$

Proof. See Appendix A

$\mathbb{E}(t_x)$ is increasing in n under fixed k and it is decreasing in k under fixed n.

5 Parameter Choice and Efficiency Analysis

This section presents three different methods on how to tune the lifetime of the information embedded in the NA molecule. The first method consists of determining the best (n, k) pair to choose. We want the data to still be accessible before some given time t and destroyed after some given time t', both with a chosen tolerance level for the probability. The second method seeks the optimal (n, k) pair yielding the shortest agony time phase compared to the total life time, i.e. being able to determine the lifespan as clearly as possible. Finally, in the third method, we describe how to find the best (n, k) pair, such that the expected value $\mathbb{E}(n, k, t_a)$ is very close to some target time t_{target} with minimal variance, giving the best guarantee that the data will be destroyed after this time.

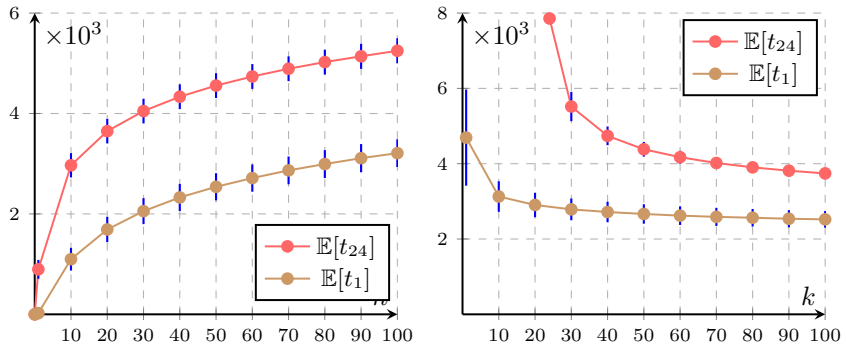

Figure 12: $\mathbb{E}[t_{24}]$ and $\mathbb{E}[t_1]$ as functions of n for $k = 40$ (left) and as functions of k for $n = 60$ (right).

5.1 Finding (n, k) for Target Times t and t'

For specific t and t' values, we want our data to still be accessible up to t and completely destroyed after t'; what is the (n, k) pair to consider? To answer this question, n and k should satisfy the following criteria:

$$\mathbb{P}(T_{\min} > t) \simeq 1 \quad \text{and} \quad \mathbb{P}(T_{\max} \leq t') \simeq 1$$

To this end, we fix a tolerance level, Δ, and require:

$$\mathbb{P}(T_{\min} > t) \geq 1 - \epsilon_\Delta \quad \text{and} \quad \mathbb{P}(T_{\max} \leq t') \geq 1 - \epsilon_\Delta$$

with $\epsilon_\Delta = \Delta \cdot 10^{-2}$.

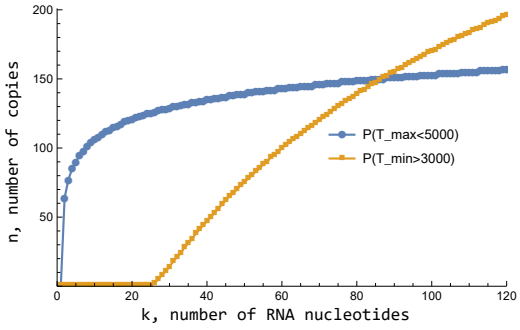

Figure 13: Domains bounds of existing solutions satisfying $\mathbb{P}(T_{\min} > 3000) \geq 96\%$ (blue indicates lower bound) and $\mathbb{P}(T_{\max} \leq 5000) \geq 96\%$ (orange indicates upper bound). Here $\Delta = 4$ and $\lambda = 0.001$ min^{-1}. The time is in min.

Figure 13 shows that a solution exists at the intersection point of the two curves. For the example in Figure 13, the solution for $t = 3000$ min and $t' =$

5000 min is: $(n, k) = (150, 86)$. This means that if we manufacture 150 NA copies containing 86 RNA nucleotides in each, there is a chance of 96% that the secret is still accessible before 3000 minutes and completely destroyed after 5000 minutes. Other solutions for other target t and t' are given in Table 1.

$t \setminus t'$	2000	3000	4000	5000
1000	$(22, 819)$	$(17, 74)$	$(15, 38)$	$(15, 30)$
2000	-	$(71, 1243)$	$(53, 83)$	$(48, 40)$
3000	-	-	$(206, 1492)$	$(150, 86)$
4000	-	-	-	$(572, 1584)$

Table 1: (n, k) solutions for different values of target t and t' (in minutes). Here $\lambda = 0.001$ min^{-1} and $\Delta = 4$.

The results of Figure 13 and Table 1 were obtained after running a search code in Python available from the authors. Note that these (n, k) values represent the solutions having the lowest cost in terms of n and k.

5.2 Finding (n, k) with Lowest Agony Ratio

What if we want that the data stored in the NA molecule to be fully accessible before some time t and then gets quickly destroyed after some time t' ? Depending on a specified risk level, expressed through α, we want that the time from $\mathbb{E}(t_1) - \alpha\delta_{t_1}$, where we are confident to have the information fully available, until time $\mathbb{E}(t_a) + \alpha\delta_{t_a}$, where we are confident that it is destroyed, is as short as possible. Here δ is the standard deviation. We thus define the *agony ratio* as:

$$f(n, k) = \frac{\mathbb{E}(t_a) + \alpha\delta_{t_a}}{\mathbb{E}(t_1) - \alpha\delta_{t_1}} \ ,$$

and aim to find the (n, k) pair giving the smallest ratio, i.e. being as close to one as possible.

Note that the agony ratio f does not depend on λ. This is particularly useful if we can adjust the fluid's chemical properties to determine the actual life span, refer to Table 7 in Section 2.4 to have an idea about the order of magnitude of the time for different values of λ.

We expect, at least for large k and n, that the probabilities $p_1 = \mathbb{P}(\mathbb{E}(t_1) - \alpha\delta_{t_1} < t_1)$ and $p_2 = \mathbb{P}(\mathbb{E}(t_a) + \alpha\delta_{t_a} > t_a)$ are close to the ones derived from a normal distribution. This is confirmed by Table 6 in Appendix B, which gives numerical values of p_1 and p_2 for the cases $\alpha = 1$ and $\alpha = 2$.

Table 5 in Appendix B shows that we effectively have lower agony ratios when n and k are significant, the best (n, k) pair is then the one with the largest values of n and k. This suggests to go further with more significant values of n and k, when the resources allow it, and when actual acceptable timings can be found depending on λ.

As an example, we take $(n, k) = (280, 280)$ and we give in Table 2 numerical values for $(t, t') = (\mathbb{E}(t_1) - 2\delta_{t_1}, \mathbb{E}(t_a) + 2\delta_{t_a})$ for different values of λ and $\alpha = 2$. In this case, $f(280, 280) \simeq 1.41$ and $(p_1, p_2) \simeq (0.96, 0.97)$. Remember that getting other values for (t, t') requires choosing other values for λ and hence adjusting the chemical properties of the fluid accordingly.

λ(in mins^{-1})	10^{-3}	$4.5 \cdot 10^{-4}$	$4.06 \cdot 10^{-5}$	$3.3 \cdot 10^{-6}$	$1.22 \cdot 10^{-7}$
(t, t')	(57.5, 81.5) hours	(5.3, 7.5) days	(2, 2.8) months	(2, 2.8) years	(53.3, 75.7) years
(p_1, p_2)	(0.96, 0.97)	(0.96, 0.97)	(0.96, 0.97)	(0.96, 0.97)	(0.96, 0.97)
agony ratio	1.41	1.41	1.41	1.41	1.41

Table 2: The best (t, t') solutions in terms of the lowest agony ratio for different values of λ. Here $(n, k) = (280, 280)$, $\alpha = 2$ and $a = 24$.

5.3 Finding (n, k) for Target Time t_{target} with the Least Variance

For a target time t_{target} we want that the secret data is inaccessible after t_{target}, what is the best (n, k) to consider? To answer this question n and k should satisfy the following approximation:

$$\mathbb{E}(n, k, t_a) - t_{target} \simeq 0$$

We are therefore looking for (n, k) pairs minimizing the distance between $\mathbb{E}(t_a)$ and t_{target}. However, we also want to be as confident as possible that this is the time that the information is destroyed. Hence, we would prefer the (n, k) pair for which $\mathbb{E}(t_a)$ has the least variance. Figure 14 represents the optimal (n, k) solutions verifying $\mathbb{E}(n, k, t_a) \simeq 2000$ min. We see that we have the least variance when n and k are large. Table 8 gives the corresponding k for each n.

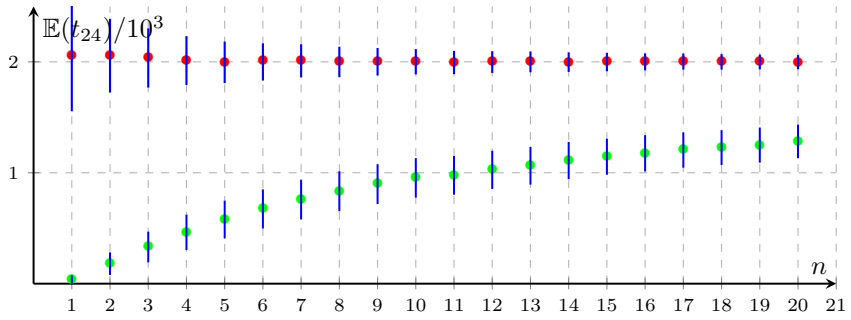

Figure 14: Optimal Solutions for $t_{\text{target}} = 2000$ *mins*. The blue lines represent the standard deviation, the red points represent the expected values $\mathbb{E}(t_{24})$, and the green ones represent $\mathbb{E}(t_1)$. Here $\lambda = 0.001$ min^{-1}.

As seen in the last section, when n and k are significant, we have a lower agony ratio as well.

Search Algorithm: Our search for finding the optimal solutions consists of finding the optimal k for each n, which we call k_n. The pair (n, k_n) ensures: $t_{target} \simeq \mathbb{E}(n, k_n, t_a)$. We can use the monotonicity properties of $\mathbb{E}(n, k_n, t_a)$ to proceed efficiently as follows:

- Initialisation: Start with $n = 1$ and $k = a$ and compute $\mathbb{E}(n, k, t_a)$. For fixed n, $\mathbb{E}(t_a)$ is decreasing with k, and for fixed k, $\mathbb{E}(t_a)$ is increasing with n. Hence if $\mathbb{E}(t_a) < t_{target}$, increase n until $\mathbb{E}(t_a) > t_{target}$.
- Increasing k: We can now increase k until we find $\mathbb{E}(n, k, t_a) \simeq t_{target}$ and we can take $k_n = k$. Since the expectation value is monotonically decreasing with k this is easily determined and can be accelerated via the bisection method. For an integer Δ, consider the interval $I_\Delta = [k, k + \Delta]$ and compute $c = g(n, k) \cdot g(n, k + \Delta)$, where $g(n, k) = \mathbb{E}(n, k, t_a) - t_{target}$. If $c < 0$, $k_{n+1} \in I_\Delta$. In this case bisect I_Δ and repeat the same operation. If not, $k_n > k + \Delta$ and we can choose a new interval from $k + \Delta$.
- Increasing n: Bearing in mind that $t_{target} \simeq \mathbb{E}(n, k_n, t_a) \leq \mathbb{E}(n + 1, k_n, t_a)$, k_{n+1} is either k_n or bigger. Hence, in the next iteration for n, we initialise k to k_n and proceed using last step.
- Finally, the optimal n, k_n value is chosen to minimise the variance of $\mathbb{E}(n, k, t_a)$.

Remark 1. We can get a wide range of values of λ if we can adjust the chemical parameters such as temperature, PH and the concentration of particular ions. Since the $\mathbb{E}(n, k, t_a)$ and its standard deviation are inversely proportional to λ this can help us to find even better solutions.

Note that we got all of our numerical values using a Python simulation since it is faster than working with the theoretical formula of $\mathbb{E}(t_a)$ directly. Table 8 in Appendix B illustrates our findings of optimal k of n for $t_{target} = 2000$ min.

6 Conclusion and Future Work

This paper presented a new method for encoding and storing information using synthetic DNA and RNA. We showed that our method allows having information with a foreseeable lifespan. Moreover, we analyzed its security, discussed parameter choice and efficiency. We proposed three different algorithms on how to tune the information lifetime.

Other media supports such as bisphenol A thermal fax paper or unstable nonvolatile memory technologies can be used to implement the same principle but the decay models of each of those phenomena should be re-computed and the formulae given in this paper adapted. For instance, in the case of thermal paper, for instance, the number of copies can be replaced by pixel size.

Future Work: Being aware of the fact that having very long oligonucleotides is a synthetic challenge and that it can yield to different rates of hydrolysis compared to small-length oligonucleotides due to the formation of intra- and inter-molecular interactions, our theoretical analysis works as a proof of principle and the answer to these research questions is left for a future work.

References

1. Cheung, Y.W., Röthlisberger, P., Mechaly, A.E., Weber, P., Levi-Acobas, F., Lo, Y., Wong, A.W., Kinghorn, A.B., Haouz, A., Savage, G.P., et al.: Evolution of abiotic cubane chemistries in a nucleic acid aptamer allows selective recognition of a malaria biomarker. Proceedings of the National Academy of Sciences 117(29), 16790–16798 (2020)
2. Flamme, M., McKenzie, L.K., Sarac, I., Hollenstein, M.: Chemical methods for the modification of rna. Methods 161, 64–82 (2019)
3. Joan Daemen, V.R.: The Design of Rijndael: AES — The Advanced Encryption Standard (Information Security and Cryptography). Springer (2002)
4. Karikó, K., Muramatsu, H., Welsh, F.A., Ludwig, J., Kato, H., Akira, S., Weissman, D.: Incorporation of pseudouridine into mrna yields superior nonimmunogenic vector with increased translational capacity and biological stability. Molecular therapy 16(11), 1833–1840 (2008)
5. Kestemont, D., Renders, M., Leonczak, P., Abramov, M., Schepers, G., Pinheiro, V.B., Rozenski, J., Herdewijn, P.: Xna ligation using t4 dna ligase in crowding conditions. Chemical Communications 54(49), 6408–6411 (2018)
6. Kimoto, M., Hirao, I.: Genetic alphabet expansion technology by creating unnatural base pairs. Chemical Society Reviews 49(21), 7602–7626 (2020)
7. Kodr, D., Yenice, C.P., Simonova, A., Saftić, D.P., Pohl, R., Sýkorová, V., Ortiz, M., Havran, L., Fojta, M., Lesnikowski, Z.J., et al.: Carborane-or metallacarborane-linked nucleotides for redox labeling. orthogonal multipotential coding of all four dna bases for electrochemical analysis and sequencing. Journal of the American Chemical Society (2021)
8. Kumar, P., Caruthers, M.H.: Dna analogues modified at the nonlinking positions of phosphorus. Accounts of Chemical Research 53(10), 2152–2166 (2020)
9. Li, Y., Breaker, R.R.: Kinetics of rna degradation by specific base catalysis of transesterification involving the 2¢-hydroxyl group (1999)
10. McCloskey, C.M., Liao, J.Y., Bala, S., Chaput, J.C.: Ligase-mediated threose nucleic acid synthesis on dna templates. ACS synthetic biology 8(2), 282–286 (2019)
11. McKenzie, L.K., El-Khoury, R., Thorpe, J.D., Damha, M.J., Hollenstein, M.: Recent progress in non-native nucleic acid modifications. Chemical Society Reviews (2021)
12. Mikkola, S., Lönnberg, T., Lönnberg, H.: Phosphodiester models for cleavage of nucleic acids. Beilstein journal of organic chemistry 14(1), 803–837 (2018)
13. Polack, F.P., Thomas, S.J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., Perez, J.L., Pérez Marc, G., Moreira, E.D., Zerbini, C., et al.: Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. New England Journal of Medicine 383(27), 2603–2615 (2020)
14. Renders, M., Miller, E., Hollenstein, M., Perrin, D.: A method for selecting modified dnazymes without the use of modified dna as a template in pcr. Chemical Communications 51(7), 1360–1362 (2015)
15. Richard, W., Mark J., S.: The depth of chemical time and the power of enzymes as catalysts. American Chemical Society (2001)
16. Sadava, Hillis, H., Berenbaum: Life, the science of biology (2009)
17. Verma, S., et al.: Modified oligonucleotides: Synthesis and strategy for users. biochem. 1998. 67: 99-134. 1998 by Annual Reviews
18. Zhou, D.M., Taira, K.: The hydrolysis of rna: from theoretical calculations to the hammerhead ribozyme-mediated cleavage of rna. Chemical reviews 98(3), 991–1026 (1998)

A Proofs

Proof of Equation (2) For time t, the probability that the information is still accessible at this moment is given by $\mathbb{P}(T_{\min} > t)$ and we have: $\mathbb{P}(T_{\min} > t) = \mathbb{P}(\forall j \leq k,\ T_j > t) = \prod_{j=1}^{k} \mathbb{P}(T_j > t) = \mathbb{P}(T_1 > t)^k = (1 - (1 - \exp(-\lambda t))^n)^k$, and this is true by definitions of T_i and $L_{i,j}$, and by independence and uniform distribution of $\{T_i\}_{i \leq k}$ and $\{L_{i,j}\}_{i \leq n, j \leq k}$.

Proof of Equation (3) For time t, the probability that the information is completely destroyed after this time is given by: $\mathbb{P}(T_{\max} < t)$ where: $T_{\max} = T_{(a)}$ and a is the security parameter. We introduce the random variable $Z = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t)$ and we define $p(t) = \mathbb{P}(T_i \leq t) = (1 - \exp(-\lambda t))^n$. We have then: $\mathbb{P}(T_{\max} \leq t) = \mathbb{P}(\#\{i \mid T_i \leq t\} \geq a) = \mathbb{P}(Z \geq a)$ and thus $1 - \mathbb{P}(T_{\max} \leq t) = \sum_{i=0}^{a-1} \binom{k}{i} p(t)^i (1 - p(t))^{k-i}$

Proof of Lemma 1 The number of cuts as a function of the time is a random process, which we will denote by $C(t)$. We can get the expected number of cuts at time t as follows: $\mathbb{E}[C(t)] = \sum_{i=1}^{k} \mathbb{E}[\mathbb{1}_{T_i \leq t}] = \sum_{i=1}^{k} \prod_{j=1}^{n} \mathbb{P}(L_{i,j} \leq t) = k \times \mathbb{P}(L_{1,1} \leq t)^n = k \times (1 - \exp(-\lambda t))^n$ which is true using independence and uniform distribution of random variables $\{L_{i,j}\}_{j \leq n, i \leq k}$.

Proof of Lemma 2 This proof is done by calculating three different elements: the cumulative distribution function, the density function and the expected value of t_x:

– Cumulative distribution function of t_x:

$$
\begin{aligned}
F_{t_x}(t) &= \mathbb{P}(\exists i_1, \ldots, i_x \in [1,k] : T_{i_1}, \ldots, T_{i_x} \leq t) \\
&= \sum_{m=x}^{k} \binom{k}{m} \mathbb{P}(T_1 \leq t, \ldots, T_m \leq t, T_{m+1} > t, \ldots T_k > t) \\
&= \sum_{m=x}^{k} \binom{k}{m} \mathbb{P}(T_1 \leq t)^m \cdot (1 - \mathbb{P}(T_1 \leq t))^{k-m} \\
&= \sum_{m=x}^{k} \binom{k}{m} (1 - e^{-\lambda t})^{m \cdot n} \cdot (1 - (1 - e^{-\lambda t})^n)^{k-m} \\
&= \sum_{m=x}^{k} \binom{k}{m} \cdot \left(\sum_{p=0}^{m \cdot n} \binom{m \cdot n}{p} (-1)^p e^{-\lambda p t} \right) \cdot
\end{aligned}
$$

$$\left(\sum_{a=0}^{k-m}\binom{k-m}{a}(-1)^a\cdot\left(\sum_{b=0}^{n\cdot a}\binom{n\cdot a}{b}(-1)^b e^{-\lambda bt}\right)\right)$$
$$=\sum_{m=x}^{k}\sum_{p=0}^{m\cdot n}\sum_{a=0}^{k-m}\sum_{b=0}^{n\cdot a}\binom{k}{m}\binom{m\cdot n}{p}\binom{k-m}{a}\binom{n\cdot a}{b}(-1)^{p+a+b}e^{-\lambda(p+b)t}$$
$$=\sum_{m=x}^{k}\sum_{p=0}^{m\cdot n}\sum_{a=0}^{k-m}\sum_{s=p}^{n\cdot a+p}\binom{k}{m}\binom{m\cdot n}{p}\binom{k-m}{a}\binom{n\cdot a}{s-p}(-1)^{s+a}e^{-\lambda st}$$
$$=\sum_{s=0}^{kn}\tilde{C}_x(k,n,s)e^{-\lambda st}$$

This result follows from the independence of T_i, for $i\in[1,k]$, and using the Newton binomial formula three times. Here:

$$\tilde{C}_x(k,n,s)=\sum_{m=x}^{k}\sum_{p=0}^{m\cdot n}\sum_{a=0}^{k-m}\sum_{p''=p}^{n\cdot a+p}\binom{k}{m}\binom{m\cdot n}{p}\binom{k-m}{a}\binom{n\cdot a}{b-p}(-1)^{b+a}\delta_{b,s}$$
$$=\frac{(-1)^{s+1}}{s}\sum_{m=x}^{k}\sum_{p=0}^{s}\sum_{a=0}^{k-m}(-1)^a\binom{k}{m}\binom{mn}{p}\binom{k-m}{a}\binom{na}{s-p}$$

where $\delta_{s,b}$ is the Kroenecker delta function.

- Density function of t_x: $f_{t_x}(t)=F'_{t_x}(t)=\sum_{s=0}^{kn}-\lambda s\tilde{C}_x(k,n,s)e^{-\lambda st}=\sum_{s=1}^{kn}-\lambda s\tilde{C}_x(k,n,s)e^{-\lambda st}$ where we start from $s=1$ since the constant term vanishes after differentiation.
- Expected value of t_x: $E(t_x)=\int_0^{+\infty}tf_{t_x}(t)dt=\sum_{s=1}^{kn}-\lambda s\tilde{C}_x(k,n,s)\int_0^{+\infty}te^{-\lambda st}dt=\frac{1}{\lambda}\sum_{s=1}^{kn}C_x(k,n,s)$ where: $\int_0^{+\infty}te^{-\lambda st}dt=\dfrac{1}{\lambda^2s^2}$ and $C_x(k,n,s)=\dfrac{-\tilde{C}_x(k,n,s)}{s}$. This result follows from the fact that we have finite sums.
- Variance of t_x: $V(t_x)=E(t_x^2)-E(t_x)^2=\int_0^{+\infty}t^2f(t_x)dt-\left(\int_0^{+\infty}tf(t_x)dt\right)^2=\sum_{s=1}^{kn}\dfrac{-2}{\lambda^2s^2}\tilde{C}_x(k,n,s)-\left(\dfrac{1}{\lambda}\sum_{s=1}^{nk}C_x(k,n,s)\right)^2=\dfrac{1}{\lambda^2}\left[2\sum_{s=1}^{kn}\dfrac{C_x(k,n,s)}{s}-\left(\sum_{s=1}^{kn}C_x(k,n,s)\right)^2\right]$

B Numerical Values

Security bits $a=\log_2(k+1)!$	Number of cuts
$a=84$	24
$a=103$	28
$a=133$	34
$a=260$	57

Table 3: Number of security bits versus number of cuts and DNA fragments. Note that the number of DNA fragments needed is always the number of cuts plus one.

constant	a_{pH}	b_{pH}	a_{K}	b_{K}	c_{K}
value	0.983	6	0.24	3.16	3.57
unit	none	none	$\mathrm{L.mol}^{-1}$	$\mathrm{mol.L}^{-1}$	$(\mathrm{mol.L}^{-1})^{-d_k}$

constant	d_K	c_{Mg}	d_{Mg}	a_{T}	b_{T}
value	-0.419	69.3	0.80	0.07	23
unit	none	$(\mathrm{mol.L}^{-1})^{-d_{\mathrm{Mg}}}$	none	$^{\circ}\mathrm{C}^{-1}$	$^{\circ}\mathrm{C}$

Table 4: Values of the constants in Equation (1).

$n \setminus k$	120	160	200	240	280
120	1.50	1.46	1.44	1.42	1.41
160	1.46	1.42	1.40	1.38	1.37
200	1.43	1.40	1.37	1.36	1.34
240	1.41	1.38	1.35	1.34	1.33
280	1.40	1.37	1.34	1.33	1.31

$n \setminus k$	120	160	200	240	280
120	1.67	1.62	1.58	1.56	1.54
160	1.60	1.56	1.53	1.50	1.49
200	1.56	1.52	1.49	1.47	1.45
240	1.53	1.49	1.46	1.44	1.43
280	1.52	1.48	1.44	1.43	1.41

Table 5: $f(n,k)$ values for different values of n and k. Left $\alpha = 1$ and right $\alpha = 2$

$n \setminus k$	120	160	200	240	280
120	$(0.84, 0.84)$	$(0.84, 0.84)$	$(0.84, 0.82)$	$(0.83, 0.82)$	$(0.84, 0.84)$
160	$(0.83, 0.83)$	$(0.82, 0.83)$	$(0.82, 0.84)$	$(0.84, 0.84)$	$(0.85, 0.86)$
200	$(0.84, 0.84)$	$(0.85, 0.86)$	$(0.85, 0.84)$	$(0.82, 0.84)$	$(0.83, 0.82)$
240	$(0.85, 0.86)$	$(0.84, 0.83)$	$(0.85, 0.82)$	$(0.83, 0.84)$	$(0.84, 0.83)$
280	$(0.84, 0.83)$	$(0.83, 0.84)$	$(0.84, 0.84)$	$(0.84, 0.83)$	$(0.84, 0.84)$

$n \setminus k$	120	160	200	240	280
120	$(0.97, 0.97)$	$(0.97, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$
160	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.95, 0.97)$
200	$(0.96, 0.97)$	$(0.97, 0.98)$	$(0.96, 0.97)$	$(0.95, 0.97)$	$(0.96, 0.97)$
240	$(0.97, 0.98)$	$(0.96, 0.97)$	$(0.97, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$
280	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$	$(0.96, 0.97)$

Table 6: (p_1, p_2) probabilities. Top $\alpha = 1$ and bottom $\alpha = 2$. Here $\lambda = 0.001$.

$T(^{\circ}\mathrm{C})$	pH	$[K^+]$	$[\mathrm{Mg}^{2+}]$	$\lambda_0(\ \mathrm{mins}^{-1})$	$\lambda(\ \mathrm{mins}^{-1})$	$1/\lambda$
23	13	0.1	0	$1.3 \cdot 10^{-9}$	10^{-3}	1000 minutes
23	12.5	0.03	0	$1.3 \cdot 10^{-9}$	$4,5 \cdot 10^{-4}$	37 hours
37	7.4	0.25	0.005	$1.4 \cdot 10^{-7}$	$4.06 \cdot 10^{-5}$	17.1 months
4	10.7	0.25	0.005	$1.3 \cdot 10^{-9}$	$3.3 \cdot 10^{-6}$	210.43 days
23	7	0.25	0.005	10^{-8}	$1.22 \cdot 10^{-7}$	15.5 years

Table 7: Order of magnitude of the time for different values of λ

n	1	2	3	4	5	6	7	8	9	10
Optimal k	27	31	36	42	49	57	65	76	87	101

n	11	12	13	14	15	16	17	18	19	20
Optimal k	118	136	156	182	210	243	280	324	374	435

Table 8: Optimal k for n verifying $E(n, k, t_a) - t_{target} \simeq 0$. $\lambda = 0.001$ and $a = 24$.