ELSEVIER

BioSystems 57 (2000) 13–22

BioSystems

www.elsevier.com/locate/biosystems

Cryptography with DNA binary strands

André Leier, Christoph Richter, Wolfgang Banzhaf, Hilmar Rauhe *

University of Dortmund, Department of Computer Science, Chair of Systems Analysis, 44221 Dortmund, Germany

Received 21 January 2000; received in revised form 10 April 2000; accepted 14 April 2000

Abstract

Biotechnological methods can be used for cryptography. Here two different cryptographic approaches based on DNA binary strands are shown. The first approach shows how DNA binary strands can be used for steganography, a technique of encryption by information hiding, to provide rapid encryption and decryption. It is shown that DNA steganography based on DNA binary strands is secure under the assumption that an interceptor has the same technological capabilities as sender and receiver of encrypted messages. The second approach shown here is based on steganography and a method of graphical subtraction of binary gel-images. It can be used to constitute a molecular checksum and can be combined with the first approach to support encryption. DNA cryptography might become of practical relevance in the context of labelling organic and inorganic materials with DNA 'barcodes'.

Keywords: DNA computing; Cryptography; Steganography; Graphical decryption; DNA binary strands; Molecular checksum; DNA barcodes

1. Introduction

As a medium with high information density, DNA was proposed for computational purposes (Adleman, 1994). Since then several approaches have been investigated like implementations of combinatorial (Adleman, 1994; Lipton, 1995; Ouyang et al., 1997) and functional (Guarnieri et al., 1996) algorithms and approaches based on self-assembly (Winfree et al., 1996, 1998; Rauhe et al., 1999). Theoretical considerations dealt with Turing machines, associative memory and cryptanalysis.

Cryptography has been shown recently as a new application of DNA Computing: Clelland et al. (1999) have demonstrated an approach to steganography by hiding secret messages encoded as DNA strands among a multitude of random DNA. Steganography means hiding of secret messages among other information to conceal their existence (Kahn, 1967; Schneier, 1996) and is known as a simple cryptographic method. Clel-

* Corresponding author.

E-mail addresses: leier@ls11.cs.uni-dortmund.de (A. Leier), richter@ls11.cs.uni-dortmund.de (C. Richter), banzhaf@ls11.cs.uni-dortmund.de (W. Banzhaf), rauhe@ls11.cs.uni-dortmund.de (H. Rauhe).

0303-2647/00/$ - see front matter
PII: S0303-2647(00)00083-6

land et al. (1999) have used a substitution cipher for plaintext encoding where a unique base triplet is assigned to each letter of the alphabet, each numeral and some special characters.

Instead, as digital messages usually correspond to 0–1-series, a binary DNA representation has been used here (see Fig. 1). The binary encoding is in particular suitable for the construction of datastructures and for simple and rapid decryption. Decryption can be done by an adapted method of digital DNA typing originally developed for minisatellite analysis (Jeffreys et al., 1991) (see Fig. 2). Using this method the information content can be decrypted and read directly by PCR and subsequent gel-electrophoresis, requiring no additional work such as subcloning or sequencing.

DNA binary strands were assembled by concatenation of short double stranded DNA molecules representing 0 (0-DNA bit), 1 (1-DNA bit), start or end as described earlier (Rauhe et al., 1999). The DNA molecules contain overlapping sequences ('sticky ends') and were polymerized to DNA binary strands by annealing and ligation (see Fig. 1). In order to isolate single molecules from the pool of all generated DNA strands, the molecules were ligated into plasmids (see Fig. 2b) and cloned in bacteria. Then the informational content of every cloned strand could be read individually by PCR and subsequent gel-electrophoresis. For the readout PCR a strand's start terminator and its bits were used as priming sites (see Fig. 2).

2. DNA steganography — method I

As the readout procedure is based on the knowledge of the primer sequences the primers are essential if there is no other way of reading the binary strand. Thus mixing a certain binary strand with other DNA becomes a steganographic approach to encryption as it prevents reading the binary strand by sequencing.

For encryption, the message strand that corresponds to the binary encoded plaintext was mixed with other DNA, so-called dummy strands, in equimolar amounts (see Fig. 2). To achieve better security, the dummy strands should have the same binary format as the message strand. For decryption, a unique identification sequence (key sequence) attached to the message strand is required. This can be any of the terminator sequences, normally the start sequence (see Fig. 2). Thus decryption was done by readout of the message strand using the appropriate key sequence as one primer of a PCR reaction (see Fig. 2). The other primer is either the corresponding 0-DNA bit or the corresponding 1-DNA bit. Performing both PCR reactions separately and visualizing the results by gel-electrophoresis yielded complementary patterns of bands that were read from the gel (see Fig. 2).

3. Security

For the formal analysis of the security of the encryption it is assumed that the plaintext, given

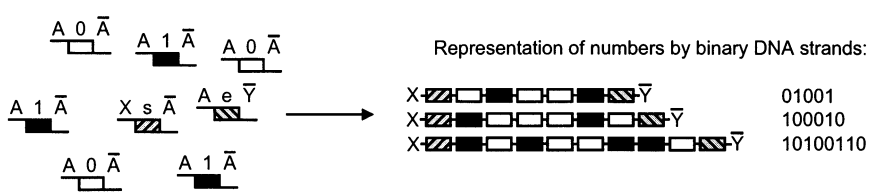

Fig. 1. Assembly of DNA binary strands. All binary strands are of the form $s\{0|1\}e$ (EBNF after ISO14977), yielded by concatenation of two terminators s (start) and e (end) and an arbitrary number of DNA bits in between. Concatenation was performed by annealing and ligation. Terminators and DNA bits are made of annealed complementary oligonucleotides having sticky ends (A, $\bar{A}$, X, $\bar{Y}$) for concatenation on both sides. The sticky end A ($\bar{A}$) works as a variable for correct concatenation of bits and terminators, whereas X and $\bar{Y}$ are sticky ends required for subsequent cloning. Bits and terminators were represented by unique double stranded DNA sequences which overlap in 26 bp and contain sticky ends of 4 nucleotides length. Bit strands containing up to 32 bits were yielded from the subsequent ligation reaction (Data not shown, see Section 9).

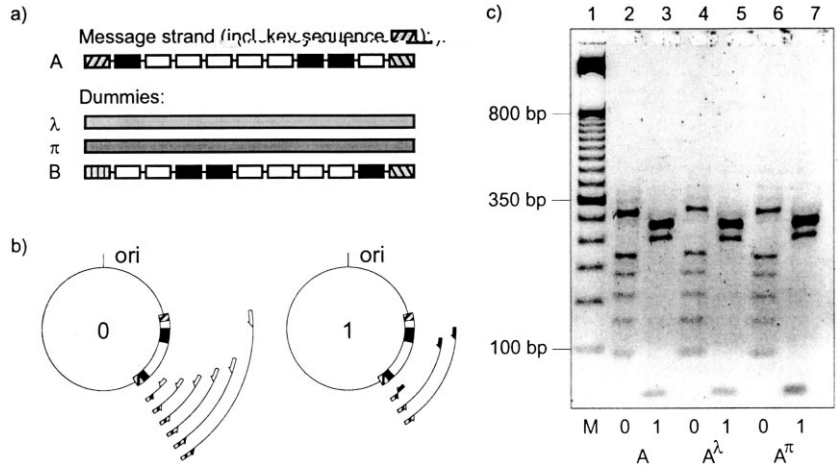

Fig. 2. Steganography with DNA binary strands, Method I. (a) A message strand (A) consisting of 9 bits and containing a unique key sequence (start) can be hidden among dummy strands, either using random DNA such as bacteriophage λ (λ) or herring sperm DNA (π), or using DNA binary strands with different key sequences (B). (b) Sketch of decryption of A. The message strand can be decrypted only if the key sequence is known because the PCR readout is based on knowledge of both primers. (c) Decryption of message strand A. Lanes 2 and 3 show readout of unencrypted A, lanes 4 and 5 show decryption of A encrypted with DNA, lanes 6 and 7 show decryption of A encrypted with herring sperm DNA (π). Lane 1 shows a molecular weight marker (50 bp GIBCO BRL). All primers used for readout were 30 nts long, the forward primer priming in the secret start-sequence and the reverse primer priming in the 0-sequence or in the 1-sequence, respectively. The readout results in a complementary ladder pattern showing bands in discrete steps of 30 bp beginning at 60 bp (see Section 9). The DNA used for encryption is in both cases equimolar to the amount of message strands of type A.

as the message strand, has the described binary structure and is being encrypted and decrypted as described above. The underlying communications protocol is described as usual by means of the A(lice)-B(ob)-scenario:

1. A and B exchange the generated key over a secure communication channel. It is expected that only A and B have the knowledge about the key.
2. A generates the message following the binary pattern (see above) and adds the key in a ligation step.
3. A generates a certain number of dummies and puts the message strand among them.
4. A sends the resulting solution to B, using an open communication channel.
5. B decrypts the message.

The security of this technique is based on the concept that there are no mark or characteristics that help a potential interceptor to distinguish between dummies and the message strand and thus to restrict the search space (the entire DNA in the solution). Decryption is allowed to be successful only with the knowledge about the key. Any single attack must not yield more information than a random extraction of one strand and its duplicates. In order to achieve this aim, two essential aspects have to be considered. First, it is necessary that each dummy has the same structure as the message strand. Second, analyzing the DNA-coded messages, which follow a linguistic structure, by methods based on linguistic statistics must be prevented. On the assumptions that the dummy strands are random base pair-sequences and that the plaintext comes from a natural source such as English, the original plaintext part of the message strand is distinguishable from the equivalent part of the dummy strands. An interceptor can take advantage of this by iteratively reducing the ratio of the number of message strands to the number of dummies through separation by use of the distinguishing mark (Gehani et al., 1999). So the cryptosystem described here is only using dummies consisting of a DNA binary strand concatenated with a sequence of length L and being equivalent to the key sequence of the

message strand with the same length. The length of the DNA binary strand corresponds to the length of the message strand, too. Further, let D be the set of different dummies and $d = |D|$ the number of its elements. To counteract an analysis based on linguistic statistics it has to be ensured that the message, normally coded as a 0–1-sequence, has about the same probability of occurrence as any other 0–1-sequence of same length. In practice compressing the message on a digital computer is used for that purpose. For the following analysis, it is assumed that the message can be encoded in a way that no attack based on linguistic statistics is successful.

For the cryptanalysis the following assumptions are made:

1. The interceptor has the same technical abilities and resources as the sender and the receiver of the message.
2. The interceptor knows that a message is sent and he has access to the open communication channel.
3. The interceptor knows the cryptographic system (in particular the encryption method).
4. The encrypted solution of molecules can not be copied. Thus, if the interceptor wants to hide the attack he or she has to feed the solution back to the communication channel.
5. The interceptor does not know the key of the message strand.
6. Each possible key has the same probability of occurrence, which does not mean that each sequence inevitably occurs.
7. All DNA strands are randomly and evenly distributed in the solution.
8. The individual frequency (number of duplicates) of each strand admits no conclusions about it being the message strand or not. Either all occurring strands have the same frequency or they are completely random.
9. The interceptor is able to filter out all duplicates of a strand if its sequence is known. Therefore, for the cryptanalysis, it is no longer necessary to take the individual frequencies into consideration.

From these assumptions it follows that the interceptor has to separate the message strand from the intercepted solution to be successful. After separation he or she can read out the message as described above.

For the interceptor there is a rare chance to distinguish between the dummies and the message strand. The only way to get the message strand is to take it by chance or to guess the key sequence. The method has the security σ ($0 \leq \sigma \leq 1$) if the probability for randomly selecting the message strand is $1 - \sigma$. Due to the upper conditions the probability is smaller the more possibilities exist to select a strand. A systematic procedure, that is separating every strand equally depends on the number of strands. The definition corresponds with the intuitive understanding of security: the more effort is needed to crack the system, the more secure it is.

The security of the cryptosystem described here is equal to the probability of grabbing the message strand out of the dummy set. In general, there are two different approaches: In case A the interceptor's only chance to take a strand out of the solution is to generate a potential key sequence. In case B the interceptor is able to specifically take any strand out of the solution. These two different approaches lead to different values of security (σ_1 for case A and σ_2 for case B).

4. System specific security, case A

Let S_E be the special key sequence and S_T an arbitrary (test-)sequence. Let $P(S_T \bowtie S_E)$ be the probability that S_T and S_E bind together. With the normalized length L of a sequence and $B = \{A, T, G, C\}(|B| = 4)$ the probability is

$$P(S_T \bowtie S_E) = \frac{1}{|B|^L} \tag{1}$$

This is valid only if S_T has to be complementary with S_E in every single base, i.e. their Hamming distance equals 0 ($h(S_T, S_E) = 0$). However it has to be considered that annealings occur even between not absolutely complementary sequences. The number of sequences S_T with Hamming distance i is

$$\binom{L}{i}(|B|-1)^i \tag{2}$$

Taken a dummy set with size d let $\lceil H(d, L)\rceil$ be the maximum number of misprimings such that an unambiguous binding is still possible. Hence, the probability is rising to

$$P(S_T \times S_E) = \frac{1}{|B|^L} \sum_{i=0}^{\lceil H(d,L)\rceil} \binom{L}{i}(|B|-1)^i \tag{3}$$

The sum yields the number of sequences S_T with $h(S_T, S_E) \leq \lceil H(d, L)\rceil$. Usually $H(d, L)$ is not integer and so $\lceil H(d, L)\rceil$ is used to consider the worst case. It does not matter that S_E is not known, as this value is the same for all sequences S_E. Altogether, the security σ_1 becomes

$$\sigma_1(d,L) = 1 - \frac{1}{|B|^L} \sum_{i=0}^{\lceil H(d,L)\rceil} \binom{L}{i}(|B|-1)^i \tag{4}$$

In order to get a specific value at this point it is necessary to determine the values of L and $H(d, L)$.

Corresponding to the PCR conditions the key sequence S_E should be of sufficient size. Considering typical conditions a size of 16 bases is realistic (Sambrook et al., 1989). To receive a higher security it is not only sufficient to increase the length of the key sequence L. Moreover it is necessary to take care for a small value of $H(d, L)$. To achieve this the dummy set D has to be enlarged in an appropriate way. Thus the security is limited, with its limit depending on the maximum length of the synthetically producible molecules only. The security is not affected by the message length but by L.

$\lceil H(d, L)\rceil$ is intended to specify the maximum number of misprimings allowing an unambiguous annealing of an arbitrary sequence S_T to the key sequence S_E. This means that the Hamming distance to the key sequence S_E may not be smaller than this maximum number for any of the d dummy sequences.

For an analysis the intuitive idea is used that S_T is binding to S_E whenever the Hamming distance between S_T and S_E is smaller than the Hamming distances between S_T and all dummy sequences S_i ($i = 1, \ldots, d$). This idea is expanded to the case of the existence of dummy sequences whose Hamming distance to S_T equals that of S_T to S_E. This leads to a worst-case security, because S_T will also anneal to S_E if there are dummy sequences that can not be distinguished with respect to their Hamming distance to the key sequence S_E. The specific Hamming distances are not calculable, because the sequences in the dummy set and the key sequence can be chosen arbitrarily. Therefore, it is necessary to consider $H(d, L)$ to be the average minimal Hamming distance between all S_i and S_E of all possible dummy sets, i.e. subsets of all $|B|^L$ possible sequences of size d. The average minimal Hamming distance $\min_{S\in D}\{h(S_E, S)\}$ is independent from the reference sequence S_E. Hence it is $h(S) := h(S_E, S)$. With the number $\binom{|B|^L-1}{d}$ of d-sized subsets of a set with the size $|B|^L - 1$—the reference sequence must not be in any set D—it is

$$H(d,L) = \frac{1}{\binom{|B|^L-1}{d}} \sum_{D,|D|=d} \min_{S\in D}\{h(S)\} \tag{5}$$

In particular the minima determination of all subsets can be put in specific terms. The following combinatorial statements will be needed:

The number of d-sized subsets of a set with size $|B|^L$ equals

$$\binom{|B|^L}{d} \tag{6}$$

The number of d-sized subsets, which contain a certain sequence equals

$$\binom{|B|^L-1}{d-1} \tag{7}$$

The number of d-sized subsets which contain at least one certain sequence from a n-sized set equals

$$\sum_{i=1}^{n}\binom{|B|^L - i}{d-1} \tag{8}$$

The number of d-sized subsets which contain at least one of n given sequences but none of m given sequences, since both sets are disjunctive, equals

$$\sum_{i=1}^{n}\binom{|B|^L - i - m}{d-1} \tag{9}$$

Using this in Eq. (5) leads to

$$H(d,L) = \frac{1}{(|B|^L_d - 1)} \sum_{i=1}^{L} i \sum_{j=1}^{\binom{L}{i}(|B|-1)^i} \binom{|B|^L - a(i,L) - j}{d-1} \tag{10}$$

with $a(i,L) = \Sigma_{j=0}^{i-1}\binom{L}{j}(|B|-1)^j$, i.e. $a(i, L)$ is the number of sequences with Hamming distance less than i. The first sum considers all subsets with a minimal Hamming distance i, i.e. the subsets containing at least one sequence with Hamming distance i. The single addends are weighted with the distance i to allow averaging. Every addend itself is a sum equal to the number of d-sized subsets, which contain at least one of $\binom{L}{i}(|B|-1)^i$ given sequences (see Eq. (2)), but none of $a(i,L)$ given sequences (see Eq. (9)).

For simplification (for mathematical background see Graham et al., 1990), Eq. (10) can be transformed to

$$H(d,L) = \frac{1}{\binom{|B|^L - 1}{d}} \sum_{i=1}^{L} \binom{|B|^L - a(i,L)}{d} = \sum_{i=1}^{L} \prod_{j=0}^{a(i,L)-2} \frac{|B|^L - d - j - 1}{|B|^L - j - 1} \tag{11}$$

Table 1 shows some values of $\sigma_1(d, L)$ with the corresponding $\lceil H(d, L) \rceil$ (in brackets) for different keylengths L and sizes d of D.

5. System specific security, case B

If the interceptor is able to isolate a certain strand out of the solution he or she is not forced to synthesize an arbitrary (test-)sequence. Rather the interceptor can sequence the isolated strand and read the strand directly. Thus the systems security now becomes directly dependent on the number of different strands in the solution. The total number of different strands in the solution is $d+1$ (dummies plus message strand). The probability isolating the message strand by chance is calculated as follows:

$$P(\text{isolation of message strand by chance}) = \frac{1}{d+1} \tag{12}$$

Hence the security σ_2 becomes

$$\sigma_2(d) = 1 - \frac{1}{d+1} \tag{13}$$

Table 1
Security σ_1[a]

L/d	1	100	$2^{10}-1$	$\lvert B\rvert^L/2$	$\lvert B\rvert^L-1$
5	0.23730 (4)	0.89648 (2)	0.98438 (2)	0.89648 (2)	0.98438 (1)
10	0.24403 (8)	0.98027 (4)	0.99649 (3)	0.99958 (2)	0.99997 (1)
20	0.22515 (16)	0.98614 (10)	0.99606 (9)	0.99999 (2)	0.99999 (1)

[a] σ_1 is the result of a step function. An increase of d always means an increase of the practical security also. The same is happening if the dummy set D is cleverly chosen without changing the set size d. An altered dummy can reduce the Hamming distance to the key sequence. At any rate a newly added dummy will be an 'interfering element' for unauthorized decryption.

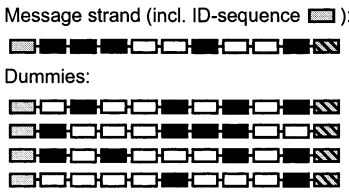

Fig. 3. Second approach to DNA steganography. The message strand is encrypted by mixing it with dummy strands that contain the same key sequence. In contradiction to the first approach (see Fig. 2a) not the key sequence but the dummy pool was used as decryption key.

Corresponding to the two different attempts to break the encryption there are two security values σ_1 and σ_2. Considering the particular case that the dummy set D is of the maximal size $|B|^L - 1$, cases A and B are identical. Then both securities become the same σ for any L and $d = |B|^L - 1$:

$$\sigma = \sigma_1(d,L) = \sigma_2(d) = 1 - \frac{1}{d+1} \quad (14)$$

6. DNA steganography — method II

An alternative steganographic approach was used for a cryptographic technique allowing a graphical message decryption: The message strand was encrypted by mixing it with a multitude of dummy strands containing an identical key sequence. As a result the key sequence could not be used as a distinctive feature for the readout process anymore (see Fig. 3).

Instead the pool of dummy strands used for encryption was used as decryption key: Readout of both the dummy pool and the encrypted pool, dummy pool plus message strand, resulted in two different gel-images. Using techniques of digital image processing (see Section 9) these gel-images were then subtracted graphically and yielded the original message strand's binary sequence (see Fig. 4). In general three variations of the method shown here seem possible:

1. The dummy pool is the key. The sender uses a certain amount of the dummy pool to hide the

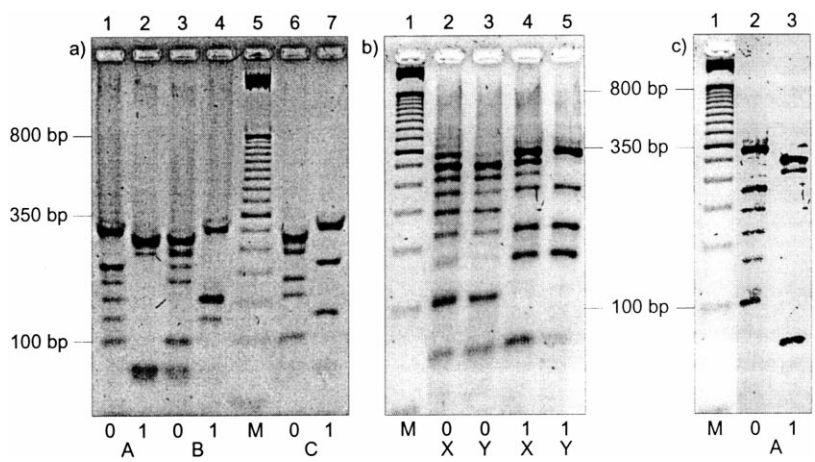

Fig. 4. Graphical Decryption. (a) Gel-electrophoresis of three 9-bit numbers (A, B and C). Read bottom up, A (lanes 1 and 2) equals $100\,000\,110_2 = 262_{10}$, B (lanes 3 and 4) equals $001\,100\,001_2 = 97_{10}$, C (lanes 6 and 7) equals $101\,001\,001_2 = 329_{10}$. M (lane 5) is a 50 bp molecular weight marker. (b) Gel-image of readout of encrypted message X (lanes 2 and 4) and dummy pool Y (lanes 3 and 5). X contains A that was mixed with B and C for encryption. Y contains only B and C as dummy pool. Both X and Y were read by PCR, 0-bits and 1-bits separately. Lane 1 is the marker lane. (c) Result of graphical decryption. The gel-image (b) was processed such that the 0-bit-lanes and the 1-bit-lanes were subtracted ($X - Y$). As the result the binary pattern of A becomes visible: Lane 2 (0-bit lanes) is the result of graphical subtraction of lane 3 from lane 2 in (b), lane 3 (1-bit lanes) is the result of graphical subtraction of lane 5 from lane 4 in (b). M (lane 1) is the same as lane 1 in (b). For confirmation of the result of decryption refer to A's binary pattern as shown in (a).

message. The receiver gets another amount of the dummy pool as secret key. Decryption is done by the receiver by performing a PCR on the encrypted solution and the key solution independently and decrypting the message strand by graphical subtraction as shown in Fig. 4.

2. A gel-image of the dummy pool is the key. The procedure is similar to variation 1 except that the gel-image is used directly.
3. The sequence information of the dummy pool is the key. The procedure is similar to variation 1 except that sender and receiver are not sharing the dummy pool but all information that is relevant to create it: the dummy sequences, their frequencies and the physical parameters.

7. Graphical decryption-usage

Graphical decryption can be considered as an independent cryptographic system. Using a dummy pool as key only once, the system is as secure as the DNA steganography system described above assuming no bit position on the 0- and 1-lane of the gel-image is unmarked. On the other hand, with multiple use of the same dummy pool, the interceptor gets more and more information about the key. This can be done by performing digital image processing on all available gel-images.

Mixing graphical decryption with other methods, e.g. DNA steganography as described above, is possible and can be regarded as a simple molecular checksum. A message interception always means a physical interception of the solution containing the message. After the interception the interceptor is forced to forward the solution to avoid the attack being noticed. If the dummies are not used during the decryption process of a message, like in the DNA steganographic system above, an attack manipulating the solution will not be noticed as long as the message is still in the solution. But if the dummies are used during decryption it should be possible to detect manipulations of the solution as a modified solution is leading to an altered gel-image. If the receiver of the message detects irregularities in the difference picture he or she has to assume that an attack has been tried.

With DNA bits in a fixed length it is to expect that a higher resolution achieved on the gel will enable encoding of a higher number of bits.

8. Conclusions and outlook

It has been shown how molecular encryption can be performed on the basis of DNA binary strands using two independent approaches to steganography. It has been shown that the first method is secure under certain assumptions about the parameters, in particular equal technical capabilities of sender, receiver and interceptor. The second method can be used as a kind of molecular checksum and help to strengthen security.

In comparison to the approach of Clelland et al. (1999) the use of DNA binary strands has some advantages. Decryption can be done easier and more rapidly requiring only PCR and subsequent gel-electrophoresis, while subcloning and sequencing is not necessary. Compared to triplet coding (Clelland et al., 1999) there is no need for an additional coding table.

Although the approach of generating bitstrands shown here has advantages such as rapid readout, it has also practical limitations. One of the limitations is the resolution of the used agarose-gels. The used gels could detect bitchains of at least up to 32 bits. Longer bitchains require gels with higher resolution such as PAGE and more sensitive detection methods of the DNA bands. For that purpose the bits can be read like nucleotides with an automatic sequenzer. However, in order to increase the amount of information substantially, a more sophisticated encoding of information is required. This can be addressed utilizing the programmability of this self-assembly approach to implement larger and more complex data structures. Further research is currently done in that direction.

Using DNA binary strands supports feasibility and applicability of DNA-based cryptography. Although molecular encryption systems may not

yet be of direct interest to computer technology they may become highly significant in another context: it was shown that DNA strands can be used for labelling of various substances and materials (Rauhe et al., 1999). As a kind of artificial 'genetic' fingerprints those DNA 'barcodes' have a broad range of potential applications in authentication, quality checking and contamination detection, for instance labelling of paint, oil and paper-based materials. Even labelling of genetically engineered products such as food seems possible. In this context molecular cryptography has a lot of aspects ranging from identification and authentication to the protection of molecular data.

9. Materials and methods

9.1. Preparation of DNA binary strands

DNA binary strands were assembled as follows (described in more detail previously in Rauhe et al., 1999):

1. Construction and synthesizing of DNA oligos: For representation of the bits and terminators unique double stranded DNA sequences with sticky ends were used. Every DNA molecule consisted of a 26 bp long double stranded core sequence and two sticky ends of 4 nucleotides length. All oligos were ordered PAGE-purified from a commercial supplier (ARK Scientific, Darmstadt, Germany).
2. Assembly of DNA bits and terminators: For assembly of the bits and terminators, corresponding oligos (upper and lower strand) were mixed and annealed in a thermocycler (PTC-100 MJ Research). Annealing was done for at least 45 min starting at 95°C and decreasing to 50°C in steps of 1°C/min.
3. Polymerization of DNA bits and terminators to DNA binary strands: Bits were phosphorylated with Polynucleotide Kinase (PNK, NEB) to be ligatable. After deactivation of PNK, bits and terminators were mixed and incubated with T4 DNA Ligase (NEB) at 16°C for 12 h. Bit strands containing up to 32 bits were yielded from polymerization.
4. Isolation by cloning: Single molecules out of the pool of binary strands were isolated by cloning in pBluescriptIIKS+ plasmid (Stratagene).

9.2. Readout

DNA binary strands were read out by PCR: two PCR reactions were set up each containing the 5′ start-primer and either the 3′-0-5′ primer or the 3′-1-5′ primer. The PCRs resulted in complementary ladder patterns of DNA fragments when visualized by gel-electrophoresis. Each PCR was prepared in 200 reaction volume by mixing 144 μl H_2O, 20 μl PCR buffer 10 × (100 mM Tris–HCl, 500 mM KCl, pH 8.3 at 20°C), 20 μl $MgCl_2$ (25 mM), 4 μl dNTPs (10 mM, Pharmacia Biotech), 2 μl Taq-Polymerase (5 u/μl, Gibco-BRL), 4 μl 5′ primer (10 μM), 4 μl 3′ primer (10 μM), 2 μl template (message strand or message strand with dummies, 10^6 molecules/μl). PCR was performed in a thermocycler (PTC-100, MJ Research) using the following protocol: 5′ 95°C, 30 cycles of 30″ 95°C, 30″ 69.5°C, 30″ 72°C; stop at 4°C.

9.3. Gel-electrophoresis

Gel-electrophoresis was done in 4% agarose. The gels were stained in 0.0005% ethidium bromide.

9.4. Steganography — method I

DNA binary strands were encrypted by mixing a certain cloned strand (the message strand) with DNA dummy strands in equimolar amounts. As dummy strands either Bacteriophage λ DNA (GIBCO BRL, Cat. no. 25250) or herring sperm DNA (Sigma D6898, Deoxyribonucleic Acid Type XIV) was used. Decryption was done by adaptation of a method of digital DNA typing originally developed for DNA minisatellite analysis (Jeffreys et al., 1991). For that purpose two independent PCR reactions were performed. Both reactions contained the single stranded 5′ secret key sequence as first and either the 3′ 0-bit sequence or the 3′ 1-bit sequence as second primer. The PCRs resulted in a binary complementary

ladder of bands in steps of 30 bp starting at 60 bp that was visualized by gel-electrophoresis as described above (see Fig. 2, Fig. 4).

9.5. Steganography — method II

Digital messages were encrypted by mixing the message strand with other binary strands in equimolar amounts. As a first step of decryption, the 0-bits and 1-bits were read out as described above. This was done for the solution containing the encrypted message and for the dummy pool that was used as decryption key. Graphical subtraction (see Fig. 4) then was done using Photoshop (Version 5.0 for Apple Macintosh, Adobe Systems Inc.; other image processing programs, e.g. Gimp can be used as well). In particular the decrypted gel-image (Fig. 4c) was created as follows:

1. The corresponding lanes 2 and 3 (0-bits) and the corresponding lanes 4 and 5 (1-bits) of the original gel image (Fig. 4b) were copied in separate layers by the 'Rectangular Marquee-Tool' and the 'Layer via Copy' command.
2. In case of the 0-bits as well as in case of the 1-bits the unencrpyted lane was put congruently on top of the corresponding encrypted lane such that bands, visible in both lanes, covered each other.
3. The corresponding encrypted and unencrypted lanes were subtracted graphically by applying the 'calculation' command to the two layers using the following settings: blending = subtraction, invert sources = true. Changing the offset leaded to a better resulting image.
4. Contrast and brightness of this image were modified such that the single bands became clearly visible (Adjust-Brightness/Contrast).
5. Using the 'variations' command shadows, midtones and highlights were modified and applied to the image as a whole to enhance perceptibility.

Acknowledgements

The first authorship is equally shared by C. Richter and A. Leier. The authors wish to thank Jonathan C. Howard for his friendly support and Robert E. Keller for help with the manuscript.

References

Adleman, L.M., 1994. Molecular computation of solutions to combinatorial problems. Science 266, 1021–1024.

Clelland, C.T., Risca, V., Bancroft, C., 1999. Hiding messages in DNA microdots. Nature 399, 533–534.

Gehani, A., LaBean, T.H., Reif, J.H., 1999. DNA-based Cryptography. In: 5th DIMACS Workshop on DNA Based Computers, MIT, June, 1999.

Graham, R.L., Knuth, D.E., Patashnik, O., 1990. Concrete Mathematics. 6. Addison–Wesley, Reading, MA print, with corr.

Guarnieri, F., Fliss, M., Bancroft, C., 1996. Making DNA add. Science 273, 220–223.

Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L., Monckton, D.G., 1991. Minisatellite repeat coding as a digital approach to DNA typing. Nature 354, 204–209.

Kahn, D., 1967. The Codebreakers. Macmillan Publishing Company, New York.

Lipton, R.J., 1995. DNA solution of hard computational problems. Science 268, 542–545.

Ouyang, Q., Kaplan, P.D., Liu, S., Libchaber, A., 1997. DNA Solution of the maximal clique problem. Science 278, 446–449.

Rauhe, H., Vopper, G., Banzhaf, W., Howard, J.C., 1999. Programmable polymers http://ls11-www.cs.uni-dortmund.de/molcomp/publications/publications.html.

Sambrook, J., Fritsch, E.F., Maniatis, T., 1989. Molecular Cloning, second ed. Cold Spring Harbor Laboratory Press, New York.

Schneier, B., 1996. Applied Cryptography, second ed. John Wiley, New York.

Winfree, E., Yang, X., Seeman, N.C., 1996. Universal Computation via Self-assembly of DNA. Some Theory and Experiments. In: Proceedings of the Second DIMACS Meeting on DNA Based Computers, Princeton University, June 1996.

Winfree, E., Liu, F., Wenzler, L.A., Seeman, N.C., 1998. Design and self-assembly of 2-dimensional DNA crystals. Nature 394, 539–544.