# Private and Public Key DNA steganography

Christoph Richter[†], André Leier[†], Wolfgang Banzhaf[†], Hilmar Rauhe[†‡]

February 11, 2000
Revised April 20, 2000

{leier,richter,banzhaf,rauhe}@ls11.cs.uni-dortmund.de

In this paper steganographic approaches to DNA cryptography are presented. The first approach shows how digital DNA strands can be used for steganography to provide rapid encryption and decryption. The second approach is based on a method of graphical subtraction of gel-images. It can be used to constitute a molecular checksum and can be combined with the first approach to support encryption. The second part of this paper explains a public key steganographic system. It is based on the usage of a certain double stranded DNA ring molecule which can be constructed by means of grammar rule molecules.

**Keywords:** Digital DNA, DNA Cryptography, DNA Steganography, Grammar, Private Key, Public Key

## Private key encryption

In DNA steganography [1, 3] secret messages encoded by DNA strands are encrypted by hiding them among a multitude of additional DNA strands. For decryption, unique PCR priming sites were attached to the message strand such that the PCR primers worked as secret key by amplifying the message strand selectively. Clelland et al. [1] have demonstrated the feasibility of this concept, using a substitution cipher for plaintext encoding and random DNA to disguise the secret message DNA. A similar cryptosystem based on DNA encoded binary strings (digital DNA strands), was shown and analysed in [5, 7]. The binary encoding provides rapid decryption utilizing a method of digital DNA typing [4, 6]. Using this method the information content can be decrypted and read directly by PCR and subsequent gel-electrophoresis, requiring no additional subcloning or sequencing (figure 1).

[†]University of Dortmund, Dept. of Computer Science, Chair of Systems Analysis, 44221 Dortmund, Germany
[‡]To whom correspondence should be addressed.

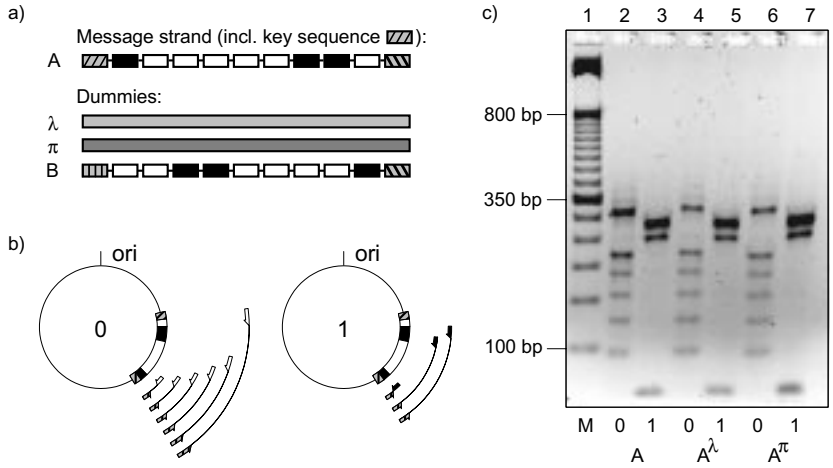


Figure 1: **Steganography with digital DNA.** a) The information content of a digital DNA strand is shown by black and white boxes representing the 1- and 0-bit. Such a message DNA strand ($A$) containing a unique key sequence (start) can be hidden among additional strands, either using random DNA such as bacteriophage $\lambda$ ($\lambda$) or herring sperm DNA ($\pi$), or using digital DNA strands with different key sequences ($B$). b) Sketch of readout/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. The sequences of the DNA bits are considered to be public, whereas the terminator sequences are considered to be secret. 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 $\lambda$ DNA, lanes 6 and 7 show decryption of $A$ encrypted with herring sperm DNA ($\pi$). Lane 1 shows a molecular weight marker (50bp GIBCO BRL). In both cases the DNA used for encryption is equimolar to the amount of message strands of type $A$. The encrypting DNA does not interfere with the primers during readout.

The security of this cryptosystem is far enhanced when digital DNA strands are used as distractor strands also. For a maximum security these "dummy strands" should be equal in length to the secret message strand [5, 7].

Digital DNA strands can also be used for an alternative approach of private key steganography which was called "graphical decryption" [5, 7]. It is based on steganography and a method of graphical subtraction of binary gel-images using techniques of digital image processing (figure 2). It can be used to constitute a kind of a molecular checksum and can be combined with other cryptographic techniques to support encryption.

For graphical decryption, a 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 (figure 3).

Instead, the whole pool of dummy strands used for encryption was used as decryption

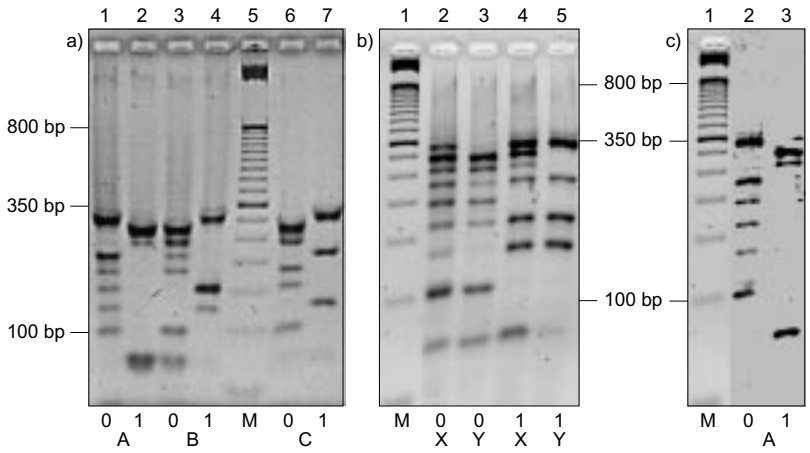


Figure 2: **Graphical Decryption.** a) Gel-electrophoresis of three 9-bit numbers ($A$, $B$ and $C$). Read bottom up, $A$ (lanes 1 and 2) equals $100000110_2 = 262_{10}$, $B$ (lanes 3 and 4) equals $001100001_2 = 97_{10}$, $C$ (lanes 6 and 7) equals $101001001_2 = 329_{10}$. $M$ (lane 5) is a 50bp molecular 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 lane) is the result of graphical subtraction of lane 3 from lane 2 in b), lane 3 (1-bit lane) 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).

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 *Materials and methods*) these gel-images were then subtracted graphically and yielded the original message strand's binary sequence (figure 2).

When mixing graphical decryption with other methods it 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.

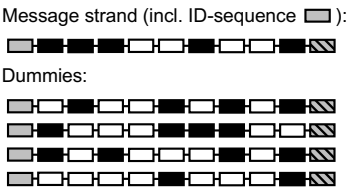


Figure 3: **DNA steganography based on graphical decryption.** The message strand is encrypted by mixing it with dummy strands containing the same key sequence. Not the key sequence but the dummy pool was used as decryption key.

A drawback of graphical decryption is that using the same dummy pool more than one time means a loss in security, as an interceptor can get more and more information about the key by performing digital image processing on all available gel-images.

### A public key system

The DNA cryptosystems described above are so-called private key systems. The construction of a *public key* DNA steganography system can be done following the principle of a combination lock. The key is represented by a double stranded circular molecule, that contains a sticky end as the public key and open H-bonds within the ring as the private key (figure 4). Inner and outer ring of the molecule can be "rotated" against each other. Examinations of potential key molecules have shown that in contrast to other DNA molecules (like linear or branched DNA strands) this key molecule keeps privacy under use of current DNA manipulation techniques [7]. It is to remark that the introduced system is based on theoretical considerations and is not put into practice up to now.

The detailed structure of the key molecule is the following (see also figure 4):

- A double stranded DNA-molecule forms a ring. This DNA ring contains a 3-arm junction, whereby one arm forms a joining piece for connecting the secret message strand and the DNA ring. This connection segment is the public part of the key. Everything else belongs to the private part of the key.
- The DNA ring is sectioned into $K$ sectors. Each sector is sectioned into two segments of different type, a variable segment (**V-segment**) and a fixed segment (**F-segment**). All F-segments (V-segments) have the same fixed length $L_F$ ($L_V$).
- Incident segments on the ring are of different types.
- $K-1$ F-segments are identical. $K$ single stranded parts of the F-segments belonging to the inner ring are equal. Because of the connection segment only $K-1$ single stranded parts of the F-segments belonging to the outer ring are equal.
- V-segments are never identical and do not contain any restriction site.
- The connection segment does not contain any restriction site.

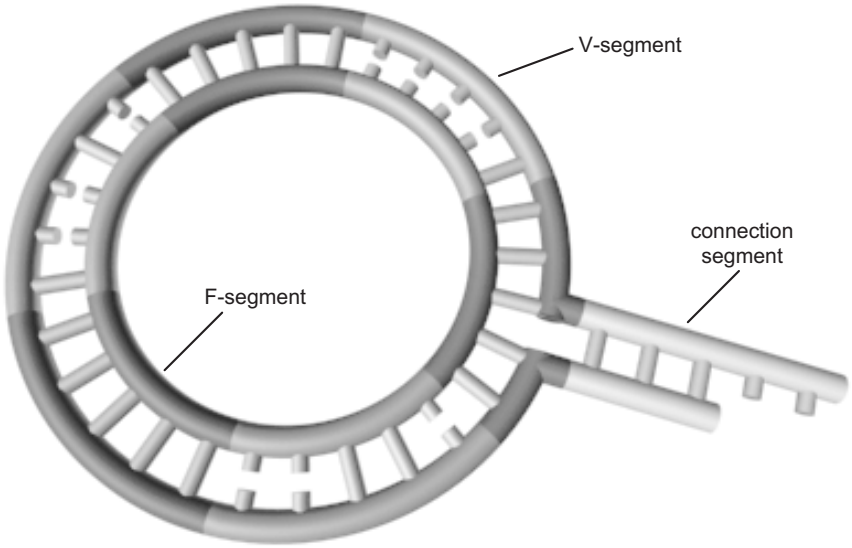


Figure 4: **Key or complementary key, respectively.** Both key and complementary key have the same construction, which is shown here schematically.

The F-segments are important for the stability of the DNA molecule. It is their purpose to enable the key construction. Except the F-segment with 3-arm junction, a single F-segment is a completely complementary double strand. The V-segments are the carrier of the real secret key information. Knowing all V-segments, i.e. their base sequences, and their order in the key, the secret message strand concatenated with this key molecule can be separated from other molecules. Each single V-segment consists of two DNA strands which must not be completely complementary. The pattern of complementary and not complementary bases, i.e. the pattern of primings and misprimings, in all V-segments is the secret key information. With it, it is possible to identify the key molecule and therefore to decrypt the secret message (see figure 4 for the spatial structure of the key molecule).

The principle of encryption is again based on hiding a secret message strand among a multitude of additional DNA. These additional dummy molecules match to the structure of the key with the concatenated message strand, i.e. each dummy consists of a key-like molecule, the "dummy key", annealed to a dummy message strand. The dummy keys are "variants" of the real key. In dummy keys the inner single stranded DNA ring is rotated as against the outer ring (figure 5). Without knowledge about the real key the dummy keys can not be distinguished from it.

Decryption is done by hybridizing the open H-bonds of the key molecule to the open H-bonds of the "complementary key". This is a DNA ring molecule, having the same structure

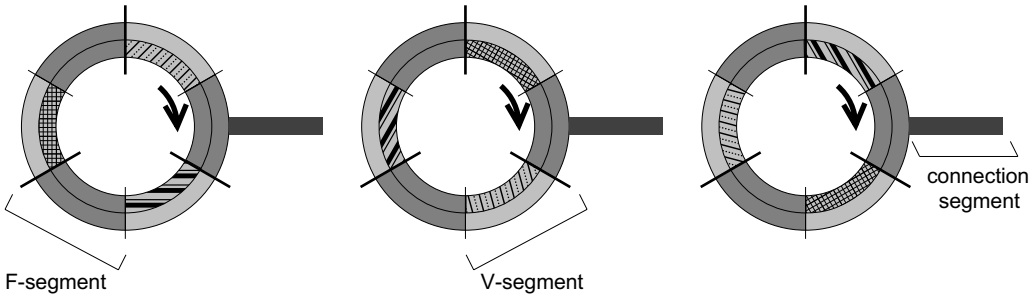


Figure 5: **Sketch of the key variants.** By rotation of the inner DNA ring different variants of the key molecule are created. The number of possible key variants (three in this figure) equals the number of sectors of the ring. The V-segments of the sectors are shown in different patterns.

and exactly the same pattern of primings and misprimings as the key molecule, a connection segment inclusive. In particular the bases of key and complementary key are complementary to each other at the mispriming positions of both molecules allowing annealing of key and complementary key (figure 6).

By means of the connection segment the complementary key can be fixed on a surface. After annealing, key and complementary key can be separated from other molecules. The separated molecule contains the secret message strand. The readout of this message depends upon the used DNA message representation, for instance is the DNA binary coding (figure 1).

### Key construction

To create a key, at first the number of its sectors (K) has to be determined. In the next step the inner and outer sequences of the V-segments have to be defined. This segments will enable the key identification. Therefore, it is important that different combinations resulting from rotations of the inner ring are leading to different patterns of complementarity. Let the outer sequences of the V-segments be $v_i$ and the inner sequences be $w_i$ $(i = 1, \ldots, K)$.

Self assembly of molecules as described by a certain grammar [8] can be used to construct a key. The rules of the grammar are represented by "rule molecules" referred to as algomers, which will assemble to key molecules [6]. A suitable grammar is $G = (\Sigma, V, R, S)$ with $\Sigma := \{a_1, a_2, \ldots, a_K, E\}, V := \{S, X_1, \ldots, X_K\}$ and $R := \{S \to a_1 X_1, X_1 \to a_2 X_2, \ldots, X_{K-1} \to a_K X_K, X_K \to E\}$.

Converting the rules to algomers the variables $X_1, \ldots, X_K$ are represented by sticky ends. The terminal symbols $a_1, \ldots, a_K$ in particular contain the V-segments. As sticky ends are used as variables, it is important that the terminal symbols are extended in a way that the demanded F- and V-segments are produced. The V-segments were defined already (see above). To keep the structure of the key molecule it is essential that the V-segments do not contain the sequences of the variables $X_1, \ldots, X_K$. Thus, these sequences have to be included into the F-segments. With it a certain structure of the F-segments results which is shown in figure 7.

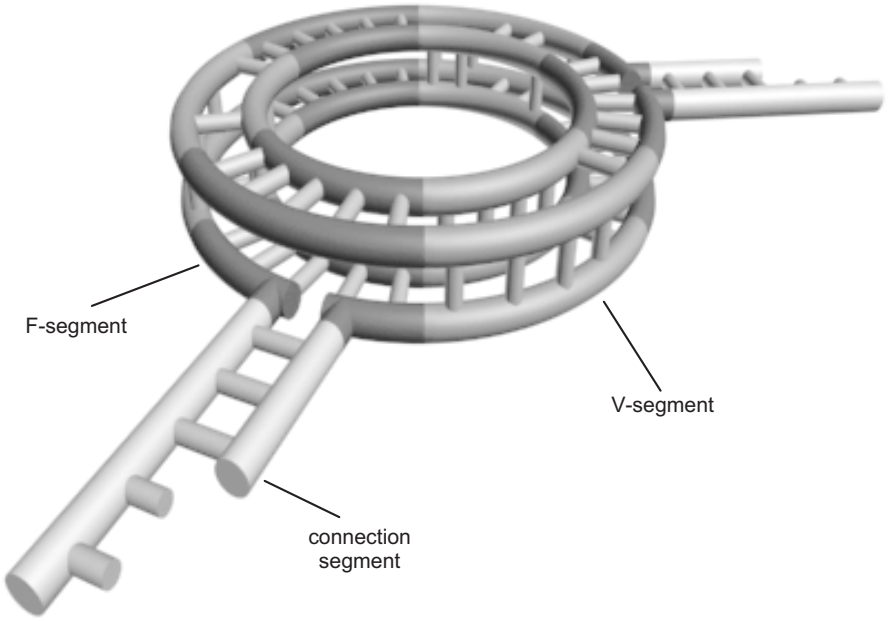


Figure 6: **Key and complementary key.** Owning the secret key, it is possible to anneal it to the message strand's secret key. The "free" bases in the V-segments of both molecules allow exact annealing of key and complementary key.

The area $X_A$ in the F-segment is needed for the implementation of the connection segment. In the F-segments, where a connection segment is necessary, this area contains a 3-arm junction. In all other F-segments this area has the same sequence but the junction is not implemented.

The algomers are designed to guarantee a correct structure of the key molecule. The F- and V-segments have to appear alternately. Because the stability of the V-segments is not warranted, it is not sufficient to provide certain algomers for each segment type. Therefore every V-segment is enclosed by F-segments. The structure of the algomers is shown in figure 8.

$K$ algomers are necessary to transfer the rules of the grammar. The first one of this algo-

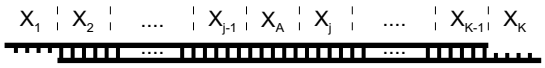


Figure 7: **Sketch of the F-segment structure.** The segment consists of a series of variables $X_1, \ldots, X_K$, which is interrupted by the area $X_A$. This corresponds to the area, which enables the junction for the connection segment in the first F-segment.

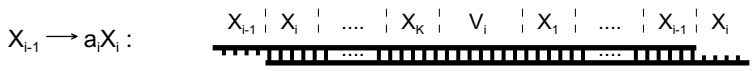


Figure 8: **Sketch of the structure of the $i$-th algomer.** $V_i$ represents the $i$-th V-segment. The area $X_A$ (figure 7) is located somewhere between $X_1$ and $X_K$ and is not shown here.

mers which represents the rule $S \rightarrow a_1 X_1$ contains the junction for the connection segment of the key (figure 9). Because of their sticky ends these $K$ algomers can be concatenated to a key molecule. The given grammar is using the abstract end symbol $E$, where the constructed key is going to form the ring.

The used grammar is a regular grammar. Due to the ringform which has to be constructed it is possible to describe the process of construction with other grammars also. For example the ring may be build in both directions at once. Therefore a rule of the form $S \rightarrow X_K a_1 X_1$ is necessary. With a rule of this form the grammar becomes a context-free one.

For the generation of the required DNA sequences a DNA sequence compiler [2] can be used.

Examinations concerning the security of this public key system showed that it is the best to classify this security using the probability of extracting the message strand out of the cryptogram. Considering all information about the key an interceptor may gain, the resulting probability is $P = 1/K$, where $K$ is the number of sectors of the key. This is very high, so to reduce this probability to $P = 1/K^n$ the key has to be extended. It is possible to bind $n - 1$ other key molecules to the "free" F-segments, i. e. to the F-segments, which are not joining pieces to connection segments.

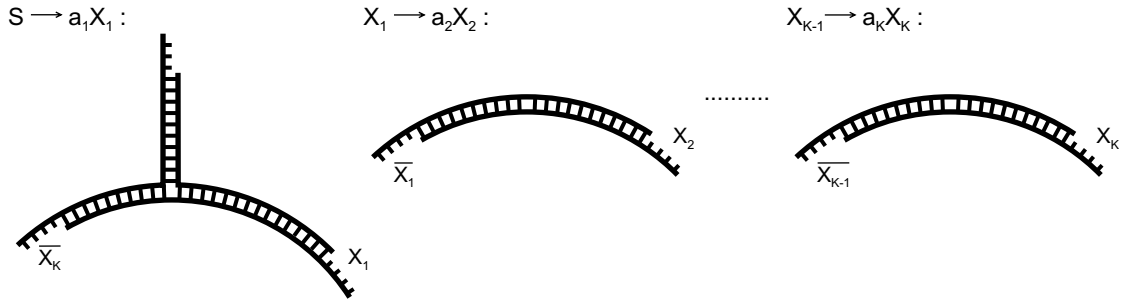


Figure 9: **Sketch of algomers for the key construction.** The sticky ends of the molecules are representing the variables $X_1, \ldots, X_K$. The rest of each molecule contains a V-segment and parts of both adjacent F-segments (figure 8).

## Conclusions and outlook

Approaches to private and public key DNA cryptography were presented. The first private key system based on digital DNA molecules in particular allows rapid decryption. Graphical decryption can be used as a kind of molecular checksum and help to strengthen security of other cryptosystems. In case of the public key system it is theoretically possible to achieve any security because of the extendible key. A drawback is, that every extension will result in more complicated molecules which are harder to handle.

Molecular cyptography will surely not become a competitor to computer cryptography, but in context with labeling of materials and items [6] it is of higher applicability. All investigations of the public key system were of theoretical nature. In the future it would be a reasonable step to test this system in a laboratory to realize practical problems.

## Materials and methods

**Preparation of digital DNA:** [6]
**Readout:** [6]
**Gel-electrophoresis:** [6]
**Steganography with digital DNA:** Digital DNA were encrypted by mixing a certain cloned strand (the message strand) with DNA dummy strands in equimolar amounts. As dummy strands either Bacteriophage $\lambda$ DNA (dam+ dcm+ from E.coli LE597, Cat.No. 25250, GIBCO BRL) 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 [4]. 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 30bp starting at 60bp that was visualized by gel-electrophoresis as described above (figure 1, figure 2).
**Graphical Decryption:** 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 (figure 2) 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 (figure 2c) 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 (figure 2b) 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.

## References


[1] Clelland, C.T., Risca, V., Bancroft, C., 1999: *Hiding messages in DNA microdots.* Nature 399, 533-534.

[2] Feldkamp, U., 1999: *Ein DNA-Sequenz-Compiler*, Diploma-thesis at the department of computer science, University of Dortmund.

[3] Gehani, A., LaBean, T.H., Reif, J.H., 1999: *DNA-based Cryptography.* Proceedings of the 5th DIMACS Workshop on DNA Based Computers, MIT.

[4] 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.

[5] Leier, A., Richter, C., Banzhaf, W., Rauhe, H., 2000: *Cryptography with DNA binary strands.* BioSystems (to appear).

[6] Rauhe, H., Vopper, G., Feldkamp, U., Banzhaf, W., Howard, J.C., 2000: *Digital DNA Molecules* (DNA6).

[7] Richter, C., Leier, A., 1999: *Molekulare Kryptographiesysteme.* Diploma-thesis at the department of computer science, University of Dortmund.

[8] Winfree, E., Yang, X., Seeman, N.C., 1996: *Universal Computation via Self-assembly of DNA: Some Theory and Experiments.* Proceedings of the 2nd DIMACS Meeting on DNA Based Computers, Princeton University.