Information Encryption Scheme Based on Chaotic Map and Hybridization Chain Reaction

Zhen Tang\textsuperscript{1}, Zhixiang Yin\textsuperscript{1,2*}, Jing Yang\textsuperscript{1}, Rujie Xu\textsuperscript{1}, Xiyuan Wang\textsuperscript{1}, Xue Pang\textsuperscript{1} and Congcong Liu\textsuperscript{1}

\textsuperscript{1}School of Mathematics and Big Data, Anhui University of Science & Technology, Huainan 232001, Anhui, China
\textsuperscript{2}School of Mathematics, Physics and Statistics, Shanghai University of Engineering Science, Shanghai 201620, China
Email: zxyin66@163.com

Abstract. Chaotic maps are characterized by randomness, unpredictability and sensitivity to initial values, and are suitable for information encryption. The natural characteristics of DNA molecules have advantages over conventional methods in the process of mass information storage and processing. In this paper, we propose an information encryption scheme based on Logistic map and hybridization chain reaction. Firstly, the plaintext is XOR-operated by chaotic sequences generated by Logistic map, and it is scrambled by hybridization chain reaction, and finally the plaintext is transformed into encrypted text. The analysis of an example shows that the information encryption scheme has good encryption effect.

Keywords. Information encryption; chaotic map; hybridization chain reaction.

1. Introduction
With the development of network communication technology, the storage and transmission of information has changed greatly, and people are facing the severe challenge of information security. One of the cores of information security is information encryption technology. Chaos is defined as the occurrence of random or unpredictable behavior in a deterministic system. In 1963, Lorenz discovered chaos in deterministic nonlinear systems by studying the general circulation model, and then came up with the argument that chaos is sensitive to initial conditions (the famous butterfly effect). Since then, chaos theory has developed rapidly [1]. The randomness, unpredictability and sensitivity to initial value of chaos are related to the field of information encryption [2-4]. Information encryption schemes based on chaotic maps usually use scrambling and diffusion methods to encrypt information, but if only chaotic maps are used to implement scrambling and diffusion in encryption schemes, the encryption schemes are very simple, and it could be vulnerable. Researchers actively explore the use of cross-disciplinary knowledge to combine chaotic maps to design the new information encryption schemes. In the process of mass information storage and processing, DNA molecules have the characteristics of molecular specificity, high parallelism and minuteness, and can be stored and operated in parallel. In 1994, Adleman made his first foray into DNA computing, using DNA coding to solve the Hamilton path problem of digraphs [5]. With the rapid development of DNA computing and biotechnology, the storage, hiding and encryption of information based on DNA computing have been greatly developed, it requires complex biological experiments, which undoubtedly increases the cost of encrypting information [6-8]. Therefore, the combination of DNA computing and chaotic maps can not only enrich encryption methods, but also save the cost of encryption [9-12].
In this paper, we propose an information encryption scheme based on Logistic map and hybridization chain reaction (HCR). The plaintext information is translated into one-dimensional decimal matrix by ASCII, and then XOR operation is carried out with the chaotic sequences generated by Logistic map, and the second encryption is carried out through HCR, and the plaintext is finally transformed into encrypted text. The case analysis shows that the encryption scheme has better encryption feasibility, and improves the key space of single chaotic map encryption scheme.

2. Logistic Map and HCR

2.1. Logistic Map
The equation of one dimensional Logistic map is as follows:

\[ x_{n+1} = rx_n(1-x_n), \quad x_n \in (0,1), \quad r \in [0,4] \]  

(1)

\( r \) is the parameter of the Logistic map. Figure 1 is a bifurcation diagram of the Logistic map. The one-dimensional Logistic map is chaotic only when \( 3.569945972 \leq r \leq 4 \).

![Figure 1. Bifurcation diagram of logistic map.](image)

2.2. DNA Coding Rules and HCR

2.2.1. DNA Coding Rules. DNA sequences usually contain four nucleic acid bases, A (adenine), T (thymine), C (cytosine), G (guanine), which strictly follow the principle of base complementary pairing, that is, A and T are complementary, G and C are complementary. These four bases can be used to encode binary numbers 00, 01, 10, 11. Due to the need to satisfy the base complementary pairing principle, there are eight DNA coding rules, as shown in table 1.

| Rule | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------|---|---|---|---|---|---|---|---|
| A    | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| T    | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| C    | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| G    | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |

2.2.2. HCR. The principle of HCR is to use initial strand to induce DNA hybridization of two different types of hairpin DNA strands, alternately open the hairpin structure to form double-stranded DNA polymer with gap [13]. The basic reaction principle for HCR is shown in figure 2. The initial strand specifically binds to the hairpin DNA strand k1 through the toehold-mediated, the hairpin of K1 is
opened, the exposed region binds specifically to the hairpin DNA strand K2, the hairpin of K2 is opened.

Figure 2. The basic reaction principle for HCR.

3. Information Encryption Scheme Based on Logistic Map and HCR

The flow chart of the information encryption scheme is shown in figure 3. The encryption steps are as follows:

Step 1: using ASCII to convert the plaintext information into a $1 \times L$ decimal matrix A.

Step 2: the initial value $x_0$ and parameter $r$ of Logistic chaotic map are used as the key. After iterating for 1000+L times, the first 1000 chaotic sequences are removed, and the values in the chaotic sequences are normalized to the interval $[0, 255]$ to obtain the $1 \times L$ decimal matrix B.

Step 3: XOR the decimal matrices A and B to get the $1 \times L$ decimal matrix C.

Step 4: convert decimal matrix C to 8-bit binary matrix, and then convert to DNA sequence matrix D by DNA coding rule 2. Then the DNA sequence matrix D is divided into blocks according to columns, and every 16 columns is divided into blocks to get the DNA sequence matrix $D_i, D_2, D_3, \ldots, D_L$. If $\frac{L}{4}$ is not an integer, add a Space after the plaintext information to make the final plaintext information length an integer multiple of 4. Each matrix $D\left(i = 1 - \frac{L}{4}\right)$ contains a DNA sequence (16nt), and these DNA sequences are set from left to right at 5’ to 3’ ends.

Step 5: the 16nt DNA sequence in matrix $D\left(i = 1 - \frac{L}{4}\right)$ is scrambled by using HCR, as follows:

We suppose the DNA sequence of the first 16 columns of matrix $D_1$ is 5’-TCCACCTCCACG-3’. Next, the sequence is displaced with HCR, and the length of toehold of the K1, K2, and K3 is set to 4nt. When HCR is used for encryption, the hairpin DNA strands K1 and K2, and the DNA complex K3 could be used as three keys, but not all of the sequences on the DNA strands could be chosen at random. As shown in figure 4, the initial DNA sequence determines the toehold region (green region, followed by the AGAG from left to right) and the stem region (blue region, bottom blue region TGGTAAGTGC, top blue region ACCATTCCCACG) on the hairpin DNA strand K1. K1’s hairpin region (red region) is the key region, and its four bases are randomly chosen, so there are 256 possibilities. Here, the sequence assigned to K1’s hairpin region is TACG. Similarly, the DNA strand
K1 determines the toehold region and the stem region of the DNA strand K2. K2’s hairpin region (purple region) is the key region, its four bases can be randomly selected, K2’s hairpin region is assigned to the sequence TAAT. The DNA strand K2 determines the toehold region and the complementary region of the DNA complex K3. The black region of K3 is the key region, and it is assigned to the sequence CGCG. The initial DNA sequence, 5’-TCTCACCATTCCCACG-3’, is successfully encrypted into 5’-ACCATTACCAGAGCT-3’ by HCR.

Step 6: after the encryption method in step 5, the new encrypted DNA sequence matrices $E_i \left( i = 1 \sim \frac{L}{4} \right)$ are obtained, and the matrices $E_i \left( i = 1 \sim \frac{L}{4} \right)$ are merged into the $1 \times 4L$ DNA sequence matrix $E$.

Step 7: using the DNA coding rule 2 to decrypt the matrix $E$, convert it into decimal matrix $F$, and finally translate the matrix $F$ into encrypted text by ASCII.

Because the encryption algorithm is reversible, the decryption process is the inverse process of the above steps.

![HCR encryption principle](image-url)
4. Case Analysis and Discussion

4.1. Case Analysis
In order to better understand the above encryption scheme, we take the specific plaintext “Hello, Beautiful Sun” as an example, using the above encryption scheme for information encryption. The corresponding matrices for each step are shown in table 2.

Step 1: using ASCII to convert the plaintext information “Hello, Beautiful Sun” into a $1 \times 20$ decimal matrix $A$.

Step 2: Set the initial value $x_0 = 0.1$ and parameter $r = 4$ of Logistic chaotic map are used as the key. After iterating for $1000 + 20$ times, the first 1000 chaotic sequences are removed, and the values in the chaotic sequences are normalized to the interval $[0, 255]$ to obtain the $1 \times 20$ decimal matrix $B$.

Step 3: XOR the decimal matrices $A$ and $B$ to get the $1 \times 20$ decimal matrix $C$.

Step 4: convert decimal matrix $C$ to 8-bit binary matrix, and then convert to the $1 \times 80$ DNA sequence matrix $D$ by DNA coding rule 2. Then the DNA sequence matrix $D$ is divided into blocks according to columns, and every 16 columns is divided into blocks to get the DNA sequence matrices $D_1, D_2, D_3, D_4, D_5$. Each matrix $D_i (i = 1 \sim 5)$ contains a DNA sequence (16nt), and these DNA sequences are set from left to right at 5’ to 3’ ends.

Step 5: the 16nt DNA sequence in matrix $D_i (i = 1 \sim 5)$ is scrambled by using HCR, and the key sequences are shown in table 2.

Step 6: after the encryption method in step 5, the new encrypted DNA sequence matrices $E_i (i = 1 \sim 5)$ are obtained, and the matrices $E_i (i = 1 \sim 5)$ are merged into the $1 \times 80$ DNA sequence matrix $E$.

Step 7: using the DNA coding rule 2 to decrypt the matrix $E$, convert it into decimal matrix $F$, and finally translate the matrix $F$ into encrypted text by ASCII.

4.2. Discussion
The above encryption scheme consists of two parts: the first encryption using chaotic map and the second encryption using HCR. The key space is also composed of chaotic key and DNA sequence key. When the correct chaotic keys are used, the matrix $C$ can be successfully decrypted into the corresponding matrix $A$. When the initial value $x_0 = 0.1$ of the chaotic keys is slightly changed ( $x_0 = 0.10000000000001$), and all the other keys are not changed, matrix $C$ cannot be successfully decrypted into matrix $A$, but instead gets an incorrect matrix $A' = [79 \ 126 \ 73 \ 121 \ 38 \ 245 \ 153 \ 113 \ 22 \ 206 \ 150 \ 80 \ 125 \ 29 \ 85 \ 30 \ 206 \ 191 \ 249]$. Therefore, it can be concluded that the solution space of the chaotic keys in the encryption scheme can reach $10^{12}$, and the solution space of two chaotic keys is equal to $10^5 \times 10^5 = 10^{10}$. In the HCR encryption part, 15 DNA sequences (the length of sequence is 4nt) are used, and the solution space is equal to $256^{15}$. Therefore, the whole key solution space of the encryption scheme is $10^{12} \times 256^{15}$. This key space is large and sensitive enough to resist brute force attacks and statistical attacks.
Table 2. The details of the encryption process.

| STEP | Description |
|------|-------------|
| 1    | Plaintext: Hello, beautiful sun<br>A=[72 101 108 108 111 44 32 98 101 97 117 116 105 102 117 108 32 115 117 110] |
| 2    | B=[13 51 163 235 74 210 149 248 28 99 242 50 160 238 64 191 192 190 193 187]<br>Logistic: x₀ = 0.1 and r = 4 |
| 3    | C=[69 86 207 135 37 254 181 154 121 2 135 70 201 136 53 211 224 205 180 213] |
| 4    | 8 bit binary stream: 010001010101011011001111100001111011101101101101110111000111001100110011011001101101111000110110110110011 |
|      | D=[CACCCCGTATTGACTAGGCTTGGTGGCGGCGCTGCAAAGGACTCACGTAGCGGAATCCTCATTGAAATATCGTCATCC] |
|      | D₁ : 5’- CACCCCGTATTGACTAGGCTTGGTGGCGGCGCTGCAAAGGACTCACGTAGCGGAATCCTCATTGAAATATCGTCATCCC -3’ |
|      | D₂ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAATATCGTCATCCC -3’ |
|      | D₃ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAATATCGTCATCCC -3’ |
|      | D₄ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAATATCGTCATCCC -3’ |
|      | D₅ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAATATCGTCATCCC -3’ |
| 5    | Key sequences: K₁₁: ATCG; K₁₂: TATA; K₁₃: ACTG<br>K₂₁: CTCG; K₂₂: CAT; K₂₃: TCTA<br>K₃₁: ACTG; K₃₂: TTGA; K₃₃: CCTG<br>K₄₁: ATGG; K₄₂: TGGG; K₄₃: ACCG<br>K₅₁: GACG; K₅₂: GGTC; K₅₃: GAAT |
| 6    | Encrypted DNA sequence matrices:<br>\( E₁ : 5’- CACCGTATTGACTACTGATTGAGGACTACGTGGTGGGCTGCTA -3’ \)<br>\( E₂ : 5’- CACCGTATTGACTACTGATTGAGGACTACGTGGTGGGCTGCTA -3’ \)<br>\( E₃ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAAT -3’ \)<br>\( E₄ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAAT -3’ \)<br>\( E₅ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAAT -3’ \)<br>\( E₆ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAAT -3’ \)<br>\( E₇ : 5’- CTCGGAAGACTCACGTAGCGGAATCCTCATTGAAAT -3’ \)<br>E₇ = [CACCGTATTGACTACTGATTGAGGACTACGTGGTGGGCTGCTA] |
| 7    | F=[86 207 135 30 254 181 154 220 2 135 70 94 136 53 211 22 205 180 213 131]| |

5. Conclusion
In this paper, we combine HCR with Logistic map to encrypt the text information, and the random arrangement of DNA sequences can greatly increase the key space. Compared with previous encryption algorithms, this encryption scheme can improve the key space based on chaotic map encryption, and can resist exhaustive attack and statistical attack more effectively. How to incorporate more DNA-based reaction networks into chaotic encryption schemes is the next step of our work.

Acknowledgments
This work was supported by National Natural Science Foundation of China (62072296), Natural Science Foundation of Anhui Province (1808085MF193), Sub-Project of CST Forward Innovation Project (18163ZT00500901).

References
[1] Lorenz E N 1963 J. Atmos. Sci. 20 130-141.
[2] Zhou N, Pan S, Cheng S and Zhou H 2016 *Opt. Laser Technol.* **82** 121-133.
[3] Zhang Y and Tang Y 2018 *Multimed. Tools Appl.* **77** 6647-6669.
[4] Zhu S, Zhu C and Wang W 2018 *Entropy-Switz* **20** 716.
[5] Adleman L M 1994 *Science* **266** 1021-1024.
[6] Farzadfar F and Lu T K 2014 *Science* **2346** 1256272.
[7] Zhang Y, Wang F, Chao J, Xie M, Liu H, Pan M, Kopperger E, Liu X, Li Q, Shi J, Wang L, Hu J, Wang L, Simmel C and Fan C 2019 *Nat. Commun.* **10**.
[8] Zhang Y, Min F, Mao X, Jing X and Liu X 2019 *J. Am. Chem. Soc.* **141** 17861-17866.
[9] Wu J, Liao X and Bo Y 2018 *Signal Processing* **153** 11-23.
[10] Qiang Z and Wei X 2013 *Optik* **124** 6276-6281.
[11] Wang X, Wang Y, Zhu X and Luo C 2020 *Opt. Laser. Eng.* **125** 105851-105851.
[12] Zou C, Wei X, Zhang Q, Zhou C and Zhou S 2021 *IEEE T. Nanobiosci.* **99**.
[13] Dirks M and Pierce N 2004 *P. Natl. Acad. Sci. USA.* **101** 15275-15278.