

# 用于 DNA 数值计算的一种信息传递模式

刘 伟 孙守霞

(鲁东大学数学与信息学院 山东 烟台 264025)

**摘 要** 数值计算是 DNA 计算的一个重要的研究方向,它直接导致了世界上第一台 DNA 计算机的诞生. 而设计一个可以在较大范围内使用的计算机的一个前提条件是它执行数值计算的能力. 这里引入一种通用的信息传递模式,利用这种模式的生化反应对 DNA 单链和不完全双链执行剪接操作,设计了一种  $N$  进制各位同时运算的并行计算的加法和减法的通用模型,可以实现数值计算的 DNA 自装配,使用 DNA 计算机进行数值计算比使用传统电子计算机进行数值计算的优势在于算法的巨大并行性.

**关键词** 信息传递;剪接操作;自装配;DNA

**中图法分类号** TP301

## An Information Transmission Mode for DNA Numerical Calculation

LIU Wei SUN Shou-Xia

(College of Mathematics and Information, Ludong University, Yantai, Shandong 264025)

**Abstract** Numerical calculation is an important research direction in DNA computing, which lead to the naissance of the first DNA computer in the world. A prerequisite for designing a computer useful in a wide range of applications is the ability to perform numerical calculations. In this paper, the authors propose a kind of general mode of information transmission in DNA computing, with which we can use splicing operations to implement the  $N$ -scale self-assembly addition and subtraction model. The advantage of numerical calculation with DNA molecules computer liken to the traditional electronic computer lie on the great parallelism DNA computing.

**Keywords** information transmission; splicing operation; self-assembly; DNA

### 1 Introduction

During the study of DNA computing, we commonly consider DNA computer does not precede electronic computer in the numerical calculation. But set up DNA numerical calculation model is one of the essence problems of DNA computer. Since Guarnieri<sup>[1]</sup> first proposed the addition model of DNA computing in 1996, many groups have been joined the study of numerical calculation DNA model. For example, Oliver<sup>[2]</sup> proposed the DNA model of matrix multiplication, Leete<sup>[3]</sup> proposed

the algorithmic method of symbolic determinant deploy with DNA model, Guarnieri and Bancroft solved the horizontal chain reaction for DNA-based addition<sup>[4]</sup>, Qiu and Lu proposed the numerical calculation mode with surface-based DNA computing model<sup>[5]</sup>, LaBean<sup>[6]</sup> et al. implemented logical computation using algorithmic self-assembly of DNA triple-crossover molecules, and Barua, Misra<sup>[7]</sup> brought binary addition in solution into effect. Although many works have been done, there still absence an ideal model to implement the numerical calculation perfectly. So the further

study is needed.

Generally speaking, the main function of computer is the expression, processing and transmission of information. The mode of information transmission not only is in connection with the purpose of the computation, but also in connection with the method of operation which we adopt to deal with DNA molecules. As we can see, in the surface-based DNA computing which proposed by Guarnieri, LaBean and Qiu, respectively, due to they adopted different molecular structures and operations, the computation mode they adopted to implement the addition carry information transmission from low-order to high-order is dissimilarity. We set up a greatly suitable self-assembly computation model of information transmission with splicing operation. It will be provided with DNA highly-parallel computation capacity, furthermore the DNA molecules which we use are simple, and the self-assembly operation shows all-right computational automation.

The paper is organized as follows: Section 2 describes the general mode of information transmission with splicing operation, section 3 proposes a symbol expression rule which is compliant with the addition and subtraction DNA computational model in section 4 and section 5, respectively. Finally, a conclusion will be drawn in section 6.

## 2 Information Transmission of Splicing Mode

A formal definition of the splicing operation was proposed in 1996 by Salomaa. In terms of this definition, we specify the basic splicing rule from DNA computing:

$$R_i \Rightarrow_i R_i R_{i+1} : R_i \# X \$ Y \# R_{i+1} \mapsto_r R_i \# R_{i+1} \$ Y \# X \quad (1)$$

The special symbol “#” denotes a connection of two information codes in one and the same DNA molecule, and the symbol “\$” denotes a spacer of two different molecules, while symbol “r” denotes a kind of restricted enzyme. The production denotes information  $R_i$  and  $R_{i+1}$  are cutting from the molecules which they belong to, respectively, and then be bonded each other so that they come into being one molecule, we call this operation is splicing operation.

**Definition 1.** Let  $\Rightarrow_n^*$  be a series of splicing operations as follows:

$$\begin{aligned} & \Rightarrow_1, \Rightarrow_2, \dots, \Rightarrow_n \\ & \text{If } R_1 \Rightarrow_1 R_1 R_2 \\ & \text{Then } R_1 \Rightarrow_n^*, R_1 R_2 \dots R_n \end{aligned} \quad (2)$$

In fact, each splicing operation may generate double strands or imperfection double strands. That is to say, some duplex in the information transmission self-assembly model own unilateral sticky end, others own bilateral sticky end. These DNA strands are come forth as the in-process product. There are maybe some imperfection double strands in both sides of symbol “\$”. The production of splicing will be more complexity if exist more than one cutting sites. For example, the following splicing operations will be used in this paper:

$$R_1 \Rightarrow_1, R_1 R_2 : R_1 \# Y_1 \$ X_2 \# R_2 Y_2 \mapsto_r R_1 \# R_2 \$ Y_1 \$ Y_2 X_2 \quad (3)$$

$$\begin{aligned} & R_i \Rightarrow_i, R_{i-1} R_i R_{i+1} : \\ & X_{i+1} \# R_{i-1} \# Y_{i-1} \$ x_i \# R_i \# Y_i \$ X_{i+1} \# R_{i+1} \# Y_{i+1} \end{aligned} \quad (4)$$

$$i = 2, 3, \dots, n-1$$

$$\begin{aligned} & R_n \Rightarrow_n, R_{n-1} R_n : \\ & X_{n-1} \# R_{n-1} \# Y_{n-1} \$ x_n \# R_n \mapsto_r R_{n-1} \# R_n \$ X_n \$ Y_{n-1} \# X_{n-1} \end{aligned} \quad (5)$$

The structure figure of  $R_i$  and the process of splicing acts is shown as Fig. 1:

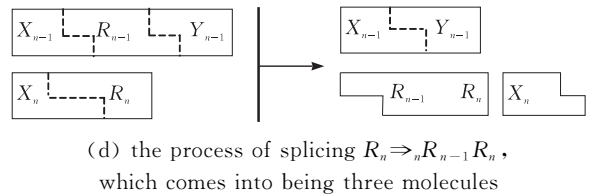
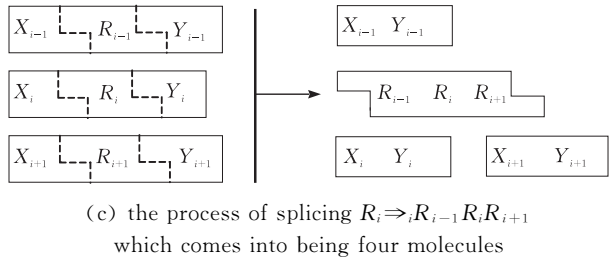
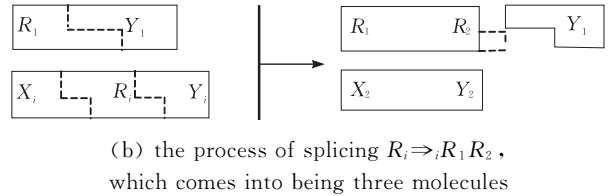
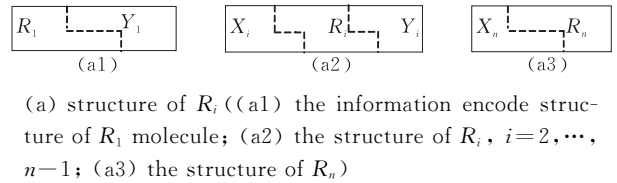


Fig. 1 The structure figure of  $R_i$  and the process of splicing acts

### 3 Symbol Expression Manner of DNA Molecule

To express DNA molecules and the alteration of biochemical reaction front-to-rear simply, this paper takes the same rule of symbol expression to denote DNA single-strands, imperfection double-strands, interrelated biochemical reactions and operations.

(1) The symbol “( )” and “[ ]”

Symbol “( )” denotes the unalloyed DNA single-strands, and symbol “[ ]” denotes the unalloyed DNA double-strands. We can use not only base {A, T, C, G} to denote encode, but also use symbols and literals to express what's the information and meaning of this DNA molecule fragment with. When we need to express different contents with some DNA fragments, the symbol “|” may be used. These DNA molecules denotes 5'→3' base sequence, while “( )<sup>T</sup>” and “[ ]<sup>T</sup>” denotes 3'→5' base sequence. In addition, sequence “(ATGC)” denotes the Watson-Crick complementary of sequence “(ATGC)”.

(2) The expression of imperfection double-strand

The 5'→3' sticky end sequence of imperfection double-strand is expressed by symbol “>” or “<”, and 3'→5' sticky end sequence is expressed by symbol “><sup>T</sup>” or “<<sup>T</sup>”. If there is single-strand fragments come forth inside the imperfection double-strands, we can mark it with symbol “( )”. But if symbol “( )” and symbol “[ ]” come forth in pairs, they can't mutually nest with each other.

(3) The operational sign of biochemical reaction

The DNA molecule's main biochemical reaction which used in this paper is paste and splicing.

#### Definition 2.

(1) Symbol “+” denotes the mix of DNA molecules, but the reaction is not occurs.

(2) Symbol “⊕” denotes mutual paste of DNA molecules, and “⊕<sup>K</sup>” denotes mutual paste and bonded with ligase.

(3) Symbol “|<sub>r</sub>” denotes restricted enzyme r from splicing site with 5'→3' direction interspace  $x$  bases cut a sticky with  $y$  bases.

(4) Symbol “≈” denotes the prolong reaction which can generate DNA double-strand structure under the primer.

### 4 Self-Assembly Model of Addition

Here we use above information mode to set up DNA computing self-assembly addition model. We take the algorithm of parallel addition computation

which each bit of  $N$  scale can calculate at one time, then design those DNA molecules that will be used to execute self-assembly calculation and accomplish parallel addition calculation with fitter operation program.

#### 4.1 Parallel Calculation of Addition

We get adapted to the addition mode of bit by bit additive from low bit to high order, i. e. a mode of “acceptance carry - additive by bit - carry”. In stead of the traditional calculation, we introduce some of a parallel calculation mode which we can compute additive by bit and carry digit additive on each bit at one time. Suppose there are two  $N$  scale expression, one is  $A = (a_n a_{n-1} \cdots a_0)_N$ , the other is  $B = (b_n b_{n-1} \cdots b_0)_N$ , and  $a_i, b_i$  denotes the value of  $N^i$  bit.

#### Algorithm.

First, let

$$a_i + b_i = S'_i \cdot N + R'_i \quad (6)$$

Where  $S'_i = 0$  or 1 denotes the pre-carry value from No.  $N^i$  bit to No.  $N^{i+1}$  bit. It is decided only by the value of  $a_i, b_i$ . Obviously, when  $a_i + b_i \geq N$ , then  $S'_i = 1$ , otherwise,  $S'_i = 0$ .  $R'_i$  is the pre-outcome of No.  $N^i$  bit in this step.

Second, let  $R_i$  denotes the outcome of No.  $N^i$  bit, and  $S_i$  denotes carry value from No.  $N^i$  bit to No.  $N^{i+1}$  bit. If  $R'_i + 1 \neq N$ , then  $R_i = R'_i + S'_{i-1}$

$$R_i = \begin{cases} R'_i, & S'_{i-1} = 0 \\ R'_i + 1, & S'_{i-1} = 1 \end{cases} \quad (7)$$

And the value of  $S_i$  is invariability. If  $R'_i + 1 = N$ , then  $R_i = 0$ , and  $S_i = S'_i + 1$ .

#### 4.2 Library of DNA Molecule

**Definition 3.** Test tube is a certain molecule set in which there is no biochemical reaction occurs.

Through such definition, test tube not only denotes those actual tubes which we use in experiment, but also denotes the DNA molecule set in specially independence area of surface method. In fact, both the model of addition and subtraction in this paper can be executed with surface technique. For a DNA computing system which at most can compute  $M+1$  bits of  $N$  scale addition, we design  $M+1$  test tubes for the input DNA molecule library:  $T_i^{(1)} (i = 0, 1, \cdots, M)$ . The DNA molecules in these test tubes have the following structure:

$$T_0^{(1)} = \{ (r | R_0 | S_1 | a_0 | b_0) \} \quad (8)$$

$$T_i^{(1)} = \{ (q | S_i | R_i | \bar{r}^T | r | R_i | S_{i+1} | a_i | b_i), \\ i = 1, 2, \cdots, M-1 \} \quad (9)$$

$$T_M^{(1)} = \{ (q | S_M | R_M | r | R_{M+1} | a_M | b_M) \} \quad (10)$$

$$a_i, b_i = \{ 0, 1, \cdots, 9 \}, i = 0, 1, \cdots, M,$$

$$S_i = \{0, 1\}, i = 1, 2, \dots, M, R_i = 0, 1, \dots, 9.$$

Where  $r$  is the identify site (GACGC,  $r^{(5,5)}$  cutting pattern) of restricted enzyme HgaI.

Obviously, we encode  $S_i$  and  $R_i$  with five bases is enough. Because there are only two values of  $S_i$ , so the hanming distance of five bases' encoding may be greater than 2. Furthermore, the mispairing ratio of hybridization when  $R_i$  does not engage in hybridization reaction seldom comes forth. The DNA segment  $q$  denotes a certain molecule which can make sure the splicing operation unique.

In addition, to identify each bit of  $N^i$  in the DNA molecule, we need to design some single aided molecules as follows:

$$K_1 = \{S_1 | P(1) | S_1\} \quad (11)$$

$$K'_i = \{(S_{i+1} | P(i+1) | S_{i+1})\} \quad (12)$$

$$K''_i = \{(P(i+1) | \bar{S}_i | \bar{S}_i)\} \quad (13)$$

$$K_n = \{(P(n) | \bar{S}_n | \bar{S}_n)\} \quad (14)$$

Where  $P(i)$  ( $i = 1, 2, \dots, n+1$ ) are the identify of  $N^i$  bit.

### 4.3 Operation Process of Addition Calculate

Let each pair value  $(a_i, b_i)$  of two addend strings  $A = (a_n a_{n-1} \dots a_0)_N$  and  $B = (b_n b_{n-1} \dots b_0)_N$  ( $n \leq M$ ) correspond with a single DNA molecule encode  $(a_i | b_i)$  ( $i = 0, 1, \dots, n$ ), then the addition  $A+B$  can be executed as follows:

Step1. Put the given  $(a_i | b_i)$  into molecular library test tube  $T_i^{(1)}$ , then reduce the temperature to occur hybridize reaction and prolong reaction. At last, divert those double stranded DNA molecule into  $T_i^{(2)}$ . The concretely reactions take placed in each test tube are denoted with the following forms:

$$(r | R_0 | S_1 | a_0 | b_0) \oplus (a_0 | b_0) \approx [r | R_0 | S_1 | a_0 | b_0] \quad (15)$$

$$T_0^{(1)} = \{(r | R_0 | S_1 | a_0 | b_0) | a_0, b_0 = 0, 1, \dots, 9\} \quad (16)$$

$$T_0^{(2)} = \{[r | R_0 | S_1 | a_0 | b_0], \text{ both } a_0 \text{ and } b_0 \text{ are given}\} \quad (17)$$

$$T_0^{(1)} \Rightarrow T_0^{(2)} \quad (18)$$

For  $i = 1, 2, \dots, n-1$

$$(q | S_i | R_i | \bar{r}^T | r | R_i | S_{i+1} | a_i | b_i) \oplus (a_i | b_i) \approx [q | S_i | R_i | \bar{r}^T | r | R_i | S_{i+1} | a_i | b_i] \quad (19)$$

$$T_i^{(1)} = \{(q | S_i | R_i | \bar{r}^T | r | R_i | S_{i+1} | a_i | b_i) | a_i, b_i = 0, 1, \dots, 9; S_i = 0, 1\} \quad (20)$$

$$T_i^{(2)} = \{[q | S_i | R_i | \bar{r}^T | r | R_i | S_{i+1} | a_i | b_i], \text{ both } a_i \text{ and } b_i \text{ are given; } S_i = 0, 1\} \quad (21)$$

$$T_i^{(1)} \Rightarrow T_i^{(2)} \quad (22)$$

$$(q | S_n | R_n | \bar{r}^T | R_{n+1} | a_n | b_n) \oplus (a_n | b_n) \approx [q | S_n | R_n | r | R_{n+1} | a_n | b_n] \quad (23)$$

$$T_n^{(1)} = \{(q | S_n | R_n | \bar{r}^T | R_{n+1} | a_n | b_n) | a_n, b_n = 0, 1, \dots, 9; S_n = 0, 1\} \quad (24)$$

$$T_n^{(2)} = \{[q | S_n | R_n | \bar{r}^T | R_{n+1} | a_n | b_n],$$

$$\text{both } a_i \text{ and } b_i \text{ are given; } S_n = 0, 1\} \quad (25)$$

Step 2. Put restricted enzyme HgaI into  $n+1$  test tubes  $T_i^{(2)}$ ,  $i = 0, 1, 2, \dots, n$  respectively, then splicing reaction come forth. At last, take DNA sequence  $r$  and  $\bar{r}^T$  as primers to extract DNA molecules into test tubes  $T_i^{(3)}$ ,  $i = 1, 2, \dots, n$ . The DNA molecules in  $T_i^{(3)}$  show as follows:

$$[r | R_0 | S_1 | a_0 | b_0] \vdash_{\text{HgaI}} [r | R_n] \bar{S}_1 > + < S_1 [a_0 | b_0] \quad (26)$$

$$T_0^{(3)} = \{[r | R_0] \bar{S}_1 >\} \quad (27)$$

$$[q | S_i | R_i | \bar{r}^T | r | R_i | S_{i+1} | a_i | b_i] \vdash_{\text{HgaI}} [q] \bar{S}_i > + < S_i [R_i | \bar{r}^T | r | R_i] \bar{S}_{i+1} > + < S_{i+1} [q] \quad (28)$$

$$T_i^{(3)} = \{< S_i [R_i | \bar{r}^T | r | R_i] \bar{S}_{i+1} > | S_i = 0, 1\} \quad (29)$$

$$[q | S_n | R_n | \bar{r}^T | R_{n+1} | a_n | b_n] \vdash_{\text{HgaI}} [q] \bar{S}_n > + < S_n [R_n | \bar{r}^T | R_{n+1} | a_n | b_n] \quad (30)$$

$$T_n^{(3)} = \{< S_n [R_n | \bar{r}^T | R_{n+1} | a_n | b_n] | S_n = 0, 1\} \quad (31)$$

$T_i^{(3)}$ ,  $i = 1, 2, \dots, n$  has two kinds of DNA molecule which correspond with  $S_i = 0, 1$  respectively, While  $T_0^{(3)}$  only has one kind of DNA molecule. The value of  $\bar{S}_i$  is decided only by  $a_0, b_0$ . The molecular structure sketch map of splicing reactions is shown as Fig. 2.

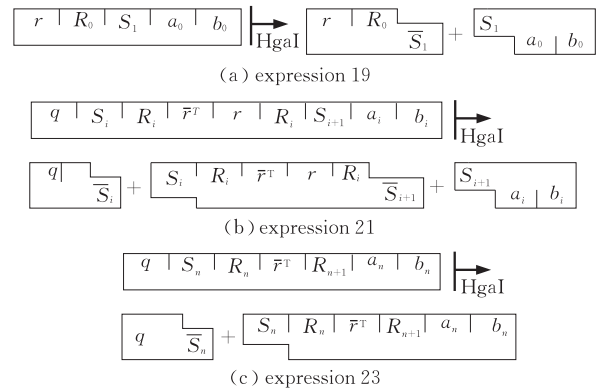


Fig. 2 The molecular structure sketch map of splicing reactions

Step 3. Mix  $n+1$  test tubes  $T_i^{(3)}$ ,  $i = 0, 1, 2, \dots, n$ , reduce the temperature and add ligase at the same time, then all the sticky ends which satisfy Watson-Crick complementary principle will take place connexion reaction. At last, it will come into being all kinds of imperfection double-strands and perfect double-strands through the agency of K ligase. Here we mark this process as:

$$T^{(4)} = T_0^{(3)} \oplus^K T_1^{(3)} \oplus^K \dots \oplus^K T_n^{(3)} \quad (32)$$

In order to simplify the description, we introduce a symbol  $f_i(R_i)$ :

$$f_0(R_0) = [r | R_0] \quad (33)$$

$$f_i(R_i) = [S_i | R_i | \bar{r}^T | r | R_i], i = 1, 2, \dots, n-1 \quad (34)$$

$$f_n(R_n) = [S_n | R_n | \bar{r}^T | R_{n+1} | a_n | b_n] \quad (35a)$$

There is only one kind of DNA strand is the longest in  $T_i^{(4)}$ :

$$[f_0(R_0) | f_1(R_1) | \cdots | f_n(R_n)] \quad (35b)$$

And  $R_0, R_1, \dots, R_n, R_{n+1}$  of (35b) just are the values which make  $A+B=R((R_{n+1}R_nR_{n-1}\cdots R_0)_N)$ .

Step 4. Separate the longest strand from  $T_i^{(4)}$  with gel electrophoresis technique, and read out  $R_i, i=0, 1, \dots, n+1$  at last.

## 5 Self-Assembly Model of Subtraction

Suppose  $A$  and  $B$  are minuend and subtrahend respectively,  $R=A-B$ ,  $a_i, b_i$  denotes the value of  $N^i$  bit as well. Then the self-assembly model of subtraction complete by two steps too:

$$\text{Step 1. } a_i, b_i = (R'_i, S'_i) \quad (36)$$

Where  $R'_i$  is the pre-outcome of No.  $N^i$  bit in this step.  $S'_i=0$  or 1 denotes the pre-borrow digit value of No.  $N^i$  bit from No.  $N^{i+1}$  bit. If  $a_i \geq b_i$ , then  $S'_i=0$ , otherwise,  $S'_i=1$ .

Step 2. Let  $R_i$  denote the outcome of No.  $N^i$  bit, and  $S_i$  denote the pre-borrow digit value of No.  $N^i$  bit from No.  $N^{i+1}$  bit. We get  $R_i$  of each digit by using  $R'_i$  subtract it's pre-borrow digit value, If  $R'_i \neq 0$ , then

$$R_i = \begin{cases} R'_i, & S'_{i-1}=0 \\ R'_i-1, & S'_{i-1}=1 \end{cases} \quad (37)$$

If  $R'_i=0$  and  $S'_{i-1}=1$ , then  $S_i=S'_{i-1}+1$ .

From the above algorithm we can see that both the design of molecule library and domain fragments are ipentity with those design of addition, and the difference only lies in the computation  $R_i$  and  $S_i$  with  $a_i, b_i$ . In particular, both the self-assembly process of addition and subtraction can confirm the carry value or borrow digit value  $S_1$  by the computational result of No.  $N^0$  bit. While the self-assembly binding between the molecules of  $N^i$  digit and  $N^{i+1}$  digit are completeness controlled by the sticky end encode  $S_i$  and  $\bar{S}_i$ . They have the same self-assembly mechanism, and there is no

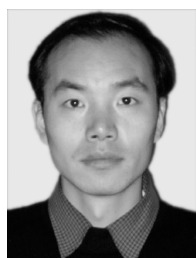
difference in structure between the self-assembly model of addition and subtraction. So we do not to describe it in detail any more.

## 6 Conclusion

The information transmission mode we brought forward by the splicing operation in this paper, not only can be used in the binding control in the carry digit and borrow digit of numerical addition and subtraction respectively, but also to be used in constitution the computation model of multiplication, division and other computational problems, So that we can use laconic way and simple DNA molecules to execute self-assembly model. The further study on the algorithm model is significant to all of us.

## References

- [1] Guarnieri F, Fliss M, Bancroft C. Making DNA add. Science, 1996, 273: 220-223
- [2] Oliver Johns. Computation with DNA: Matrix multiplication//Proceedings of the 2nd DIMACS Workshop on DNA Based Computers. Princeton, USA, 1996: 236-248
- [3] Leete T H, Schwartz M D. Massively parallel DNA computation: Expansion of symbolic determinants//Proceedings of the 2nd DIMACS Workshop on DNA Based Computers. Princeton, USA, 1996: 49-66
- [4] Guarnieri F, Bancroft C. Use of horizontal chain reaction for DNA-based addition//Proceedings of the 2nd DIMACS Workshop on DNA Based Computers. Princeton, USA, 1996: 249-259
- [5] Qiu Z Frank, Lu Mi. A surface-based DNA algorithm for the expansion of symbolic determinants//Proceedings of the Parallel and Distributed Computing and Networks. Brisbane, Australia, 1998: 481-486
- [6] LaBean Thomas H. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. Nature, 2000, 407: 493-496
- [7] Barua Rana, Misra Janardan. Binary arithmetic for DNA computers//Proceedings of the 8th DIMACS Workshop on DNA Based Computers. Sapporo, Japan, 2002: 124-132



**LIU Wei**, born in 1977, M. S., lecturer. His research interests include DNA computing and artificial neural network.

**SUN Shou-Xia**, born in 1977, M. S., lecturer. Her research interests include optimization method and DNA computing.