

# DNA 计算中核酸序列设计方法比较研究

张 凯<sup>1)</sup> 耿修堂<sup>2)</sup> 肖建华<sup>3)</sup> 赵东明<sup>1)</sup>

<sup>1)</sup>(北京大学信息科学技术学院 北京 100871)

<sup>2)</sup>(中国兵器工业集团公司 陕西 咸阳 712099)

<sup>3)</sup>(南开大学现代物流研究中心 天津 300071)

**摘 要** DNA 计算是将现实问题进行编码,映射到 DNA 分子上,然后通过分子生物实验产生出代表问题解的 DNA 分子,最后通过检测技术提取出该 DNA 分子.高质量的 DNA 编码可以尽可能避免或减少计算过程中出现的错误,并使检测阶段易于提取出代表问题解的 DNA 分子.文中对基于汉明距离和基于自由能的 DNA 核酸编码方法进行研究,分析了两类方法的约束条件对 DNA 编码质量的影响,比较了两类方法排除非特异性杂交的完备性和计算量,进一步分析了两类方法编码 DNA 序列的效率.通过分析和比较得到,两类 DNA 计算编码方法都能有效地限制 DNA 分子间的非特异性杂交,其中基于汉明距离的 DNA 编码方法的计算量比较小,但是它仅能近似地估计 DNA 分子间杂交的热力学稳定性,不能完全替代最小自由能的编码方法.在满足 DNA 计算试验精度要求的条件下,采用基于汉明距离的 DNA 编码设计方法不仅能有效地挑选出特异性杂交和非特异性杂交的 DNA 序列,还能有效地减少计算量,从而提高 DNA 序列设计的效率.

**关键词** DNA 计算;自由能;汉明距离;DNA 序列设计

**中图法分类号** TP301

## Research on Nucleic Acid Sequence Design Methods for DNA Computing

ZHANG Kai<sup>1)</sup> GENG Xiu-Tang<sup>2)</sup> XIAO Jian-Hua<sup>3)</sup> ZHAO Dong-Ming<sup>1)</sup>

<sup>1)</sup>(School of Electronics Engineering and Computer Science, Peking University, Beijing 100871)

<sup>2)</sup>(China North Industries Group Corporation, Xianyang, Shaanxi 712099)

<sup>3)</sup>(Research Center of Logistics, Nankai University, Tianjin 300071)

**Abstract** DNA computing maps the instances of a reality problem onto specific nucleic acid molecules and protocols so that the result contains the answers to the problem's instances to enable successful extraction. Good DNA sequences prevent unwanted hybridization errors during the computation and enable easy retrieval the answers in the extraction phase. At first, the paper introduces two typical nucleic acid sequence methods, free energy based method and Hamming distance based method, and analyzes the influence of constraints on the quality of DNA sequences. In addition, the paper has compared the integrity and computational cost of two methods for excluding the non-specific hybridization sequence. At last, the paper analyzes the efficiency of two nucleic acid sequence design methods. Through the comparison we can see that the Hamming distance based DNA sequence design method can only approximately estimate the thermodynamic stability of DNA hybridization, and can not be a perfect substitute for the minimum free energy method. Further analysis shows that when only consider the free energy of Watson-crick base pairs, the calculation of free energy will be equal to the Hamming distance method. So, if the precision satisfied the requirement of experiment, Hamming distance based DNA sequence design

收稿日期:2008-04-30;最终修改稿收到日期:2008-10-28. 本课题得到国家自然科学基金(60533010,30670540,60874036,60503002)、国家“八六三”高技术研究发展计划项目基金(2006AA01Z104)、中国教育部博士点基金(20070001020)及中国博士后科学基金(20060400344)资助. 张 凯,男,1979 年生,博士,主要研究方向为 DNA 计算、优化计算与运筹学. E-mail: zhangkaicn@sina.com. 耿修堂,男,1975 年生,博士,主要研究方向为智能计算、航空电子综合化系统、高射炮总体设计. 肖建华,男,1979 年生,博士,讲师,主要研究方向为仿生计算和物流优化. 赵东明,男,1977 年生,博士,主要研究方向为 DNA 计算机和仿生计算.

method, not only can effectively distinguish between specific and non-specific hybridization, and effectively reduce the amount of calculation, improve the efficiency of the DNA sequence designing.

**Keywords** DNA computing; free energy; Hamming distance; DNA sequence design

1 Introduction

In 1994, Adleman<sup>[1]</sup> presented the experiment of using molecular biology to solve a 7-vertex instance of Hamiltonian path problem. Then Lipton<sup>[2]</sup> presented a RNA-based approach for the solution of another famous SAT problems. Biomolecular computing shows a great potential to solve the NP-complete problems. In 1997, Garzon<sup>[3]</sup> identified that biomolecular computing maps the instances of a reality problem onto specific DNA molecules and protocols so that the result contains the answers to the problem’s instances to enable successful extraction. Obviously, the most important criteria to influent the efficiency of biomolecular computing is the designing of DNA sequences.

Nowadays, most DNA computing models are based on the specific hybridization between DNA molecules, but the probability of a good encoding in a randomly chosen sample goes to zero fairly quickly with the number of errors for arbitrary encoding lengths. In order to prevent the interference between different DNA molecules, and improve the reliability and effectiveness of the experiment, DNA computation requires reliable libraries of DNA sequences to be designed so that specific duplexes are formed during annealing. In recent years, many scholars proposed various DNA encoding algorithms and constraints to design DNA sequences. These DNA sequences design approach can be divided into two categories, a category is based on Hamming distances<sup>[4-6]</sup>, the other category is based on the free energy<sup>[7-8]</sup> that single-strands release during hybridization in passing to the lower energy states of double-strands.

We will introduce two typical nucleic acid sequence methods, free energy based method and Hamming distance based method, and analyze the influence of constraints on the quality of DNA sequences. Then, we compared the integrity and computational cost of two methods for excluding the non-specific hybridization sequence. In addition, we will analyze the efficiency of two nucleic acid sequence design methods.

2 DNA Encoding Constraints Analysis

In order to define the DNA sequence design

problem precisely, some definitions are necessary. In DNA computing experiment, single-strand DNA molecules dismissed randomly in vitro. DNA molecular, corresponding complementary sequence, reverse sequence and reverse complementary sequence exist simultaneity, they may be involved in the calculation of DNA hybridization reaction. In the following discussion, let  $X=5'-x_1x_2...x_n-3'$  be a DNA sequence, and  $X^C=3'-\overline{x_1}\overline{x_2}...\overline{x_n}-5'$ ,  $X^R=3'-x_n...x_2x_1-5'$ ,  $X^{RC}=5'-\overline{x_n}...\overline{x_2}\overline{x_1}-3'$  be the corresponding complement sequence, reverse sequence and reverse complement sequence, respectively.

The current models of DNA computing are based on the specific hybridization between a given sequence and its unique Watson-Crick complement. Non-specific hybridization can introduce errors, such as false positives and negatives, and degrade efficiency. Obviously, the key of DNA sequences design is to avoid the non-specific hybridization. The specific hybridization and non-specific hybridization are described in Fig. 1.

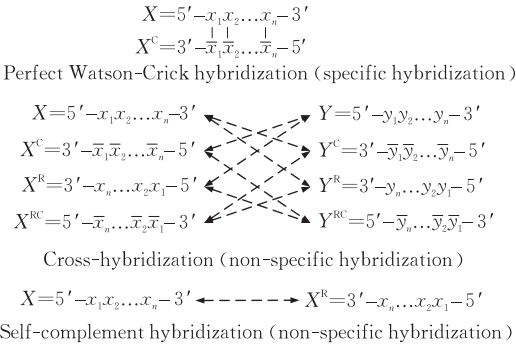


Fig. 1 DNA molecular specific hybridization and non-specific hybridization

2.1 Hamming Based Constraints Analysis

The DNA sequence design methods based on Hamming distance choose low similarity DNA sequences to exclude non-specific hybridization sequence, so that DNA molecular can be bound with imperfect matching of complementary base pairs. There are 5 constraints: (1) Hamming distance constraint; (2) similarity constraint; (3) H-measure constraint; (4) reverse complement Hamming distance; (5) self-complement Hamming distance, the specific is described as follows.

### 2.1.1 Hamming distance constraint

The Hamming distance between two binary strings is the number of corresponding places where two characters differ. For every pair of distinct words  $X=5'-x_1x_2...x_n-3'$ ,  $Y=5'-y_1y_2...y_n-3'$  in the set,  $H(X,Y) \geq d$ , here,  $H(X,Y)$  represents the Hamming distance between words  $X$  and  $Y$ , namely, the number of positions  $i$  at which the  $i$ th letter in  $X$  differs from the  $i$ th letter in  $Y$ .  $H(X,Y)$  is given by

$$H(X,Y) = \sum_{i=1}^n h(x_i, y_i), \quad h(x_i, y_i) = \begin{cases} 0, & x_i = y_i \\ 1, & x_i \neq y_i \end{cases} \quad (1)$$

In DNA coding, Hamming distance is used to describe the non-similar degree between two DNA sequences, with the greater the Hamming distance, the less similar the degree of two base pairs and the less likely for mismatch hybridization.

### 2.1.2 Similarity constraint

The similarity constraint is used to describe a similar degree between two DNA sequences. The similarity constraint computes the similarity in the same direction to keep each sequence as unique as possible including the position shift. Similarity between two binary strings is the number of the corresponding places where two characters are the same. Similarity between two DNA words  $X$ ,  $Y$  is given by

$$\text{Similarity}(X,Y) = \min_{-n < k < n} H(X, \sigma^k(Y)) \quad (2)$$

Where  $k > 0$ ,  $\sigma^k$  is the left-shift by  $k$  positions, if  $k < 0$ ,  $\sigma^k$  is the right-shift by  $k$  positions.  $H(*,*)$  is the ordinary Hamming distance between binary strands.

### 2.1.3 H-measure constraint

Garzon<sup>[9]</sup> proposed the H-measure constraint which considers two sequences as complementary ones. H-measure computes how many nucleotides are complementary between the given sequences to prevent cross-hybridization of two sequences. H-measure takes the minimum of all the Hamming distances obtained by successively shifting and lining up the Watson-Crick complement of  $Y$  against  $X$ . H-measure between two DNA words  $X$ ,  $Y$  is given by

$$\begin{aligned} H\text{-measure}(X,Y) &= \min_{-n < k < n} H(X, \sigma^k(Y^c)) \\ H(X,Y) &= \sum_{i=1}^n bp(x_i, y_i), \end{aligned} \quad (3)$$

$$bp(x_i, y_i) = \begin{cases} 1, & x_i \neq \overline{y_i} \\ 0, & x_i = \overline{y_i} \end{cases}, \quad x_i, y_i \in \{A, C, G, T\}$$

Where  $k > 0$ ,  $\sigma^k$  is the left-shift by  $k$  positions, if  $k < 0$ ,  $\sigma^k$  is the right-shift by  $k$  positions.  $H(*,*)$

is the ordinary Hamming distance between binary strands. This  $H$ -measure successively lines up the reverse of  $Y$  against  $X$  for hybridization.

### 2.1.4 Reverse complement Hamming distance constraint

In DNA computing, single-stranded DNA molecule randomly diffused in tube,  $X$  and  $Y^R$  may form non-specific hybridization. Let  $X$ ,  $Y$  denote  $5'-x_1x_2...x_n-3'$ ,  $5'-y_1y_2...y_n-3'$  respectively,  $H(X, Y^{RC})$  denote the reverse complement Hamming distance, and its calculation can be divided into two steps.

First, let  $W$  be  $Y^{RC}$ ,  $W=5'-w_1w_2...w_n-3'=5'-\overline{y_n}...\overline{y_2}\overline{y_1}-3'$ . Second, we can calculate the Hamming distance  $H(X,W)$  between  $X$  and  $W$ .

### 2.1.5 Self-complement Hamming distance constraint

DNA molecules with a certain concentration present in vitro, because of spatial nearness, DNA sequence hybridizes itself frequently. Tanaka<sup>[10]</sup> presented the self-complement constraint to avoid the hybridization between DNA molecular and its reverse DNA strand. Self complement Hamming distance is given by

$$H\text{-measure}(X,X) = \min_{-n < k < n} H(X, \sigma^k(X^R)) \quad (4)$$

The constraint is similar to the H-measure constraint, Where  $k > 0$ ,  $\sigma^k$  is the left-shift by  $k$  positions, if  $k < 0$ ,  $\sigma^k$  is the right-shift by  $k$  positions.  $X^R$  is the reverse sequence of  $X$ .

## 2.2 Free Energy Based Constraint Analysis

Free energy change  $\Delta G$  refers to the energy change when two single-stranded DNA molecules form hybrid double-stranded. DNA hybridization usually emit heat, so free energy changes are usually negative, that is  $\Delta G < 0$ .  $\Delta G$  is a measure of DNA double-stranded stability, the higher the absolute value, the DNA double-stranded more stable. In order to prevent the non-specific hybridization, every DNA sequences must satisfy the minimum free energy constraint. When the  $\Delta G$  is greater than the given threshold  $\Delta G_{\min}$  for any two DNA molecules, then can not form a stable double-stranded structure to prevent the occurrence of non-specific hybridization.

The methods based on free energy used the nearest-neighbor thermodynamic model, the formula is given by

$$\Delta G = \sum_i n_i \Delta G(i) + \Delta G(iniGC) + \Delta G(iniAT) + \Delta G(sym) \quad (5)$$

Where  $\Delta G(i)$  are the standard free energy changes for the 10 possible Watson-crick NNs (e.g.,  $\Delta G(1) = \Delta G(AA/TT)$ ,  $\Delta G(2) = \Delta G(TA/AT)$ , ...

etc.),  $n_i$  is the number of occurrences of each nearest neighbor,  $i$ , and  $\Delta G(\text{sym})$  equals  $+0.43 \text{ kcal/mol}$  ( $1 \text{ cal} = 4.184 \text{ J}$ ) if the duplex is self-complementary and zero if it is non-self-complementary<sup>[11]</sup>.

### 3 Analysis and Comparison for Free Energy and Hamming Distance Method

#### 3.1 Analysis of Restrictions on Non-Specific Hybridization

Among these Hamming based DNA encoding constraints, Hamming distance, similarity and H-measure constraints can prevent the non-specific hybridization between  $X$  and  $Y^C$ . Self-complement Hamming distance can prevent the non-specific hybridization between  $X$  and  $X^R$ . Reverse complement Hamming distance constraint can prevent the non-specific hybridization between  $X$  and  $Y^R$ . So the Hamming based DNA encoding constraints includes all the relationship to restrict the non-specific hybridization, as shown in Fig. 2.

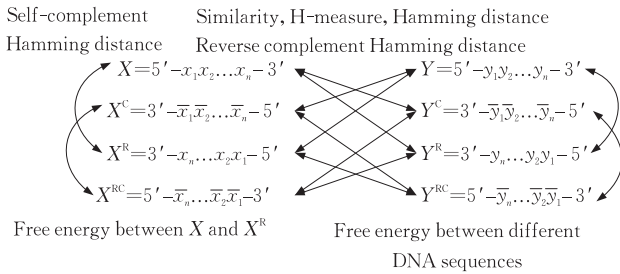


Fig. 2 Non-specific hybridization relationship

The free energy threshold  $\Delta G_{\min}$  between different DNA sequence can prevent the non-specific hybridization between  $X \leftrightarrow Y^C$  and  $X \leftrightarrow Y^R$ . The free energy threshold between  $X$  and its reverse sequence  $X^R$  can prevent the non-specific hybridization between  $X \leftrightarrow X^R$ . Obviously, the free energy constraints also include all the relationship to restrict the non-specific hybridization, as shown in Fig. 2.

#### 3.2 Analysis of Calculation on Non-Specific Hybridization

##### 3.2.1 Calculation analysis of Hamming distance method

For any two DNA sequences  $X$  and  $Y$ , let  $WC(X, Y)$  be overall complement base pairs, the formula as following

$$WC(X, Y) = \sum_{i=1}^n wc(x_i, y_i),$$

$$wc(x_i, y_i) = \begin{cases} 1, & x_i = \overline{y_i} \\ 0, & x_i \neq \overline{y_i} \end{cases}, \quad x_i, y_i \in \{A, C, G, T\} \quad (6)$$

Let the number of complement basepairs between  $X$  and  $Y^C$  is  $k$ , i. e.  $WC(X, Y^C) = k$ . Obviously, the number of complement basepairs between

corresponding reverse DNA sequence  $X^C$  and  $Y^C$  is also  $k$ ,  $WC(X^C, (Y^C)^C) = k$ .  $WC(X^C, (Y^C)^C) = WC(X^C, Y) = k$ .

The number of complement basepairs between corresponding complement DNA sequence  $X$  and  $Y^C$  is also  $k$ ,  $WC(X^R, (Y^C)^R) = k$ .  $WC(X^R, (Y^C)^R) = WC(X^R, Y^{RC}) = k$ .

The number of complement basepairs between corresponding complement DNA sequence  $X^R$  and  $Y^{RC}$  is also  $k$ ,  $WC((X^R)^C, (Y^{RC})^C) = k$ .  $WC((X^R)^C, (Y^{RC})^C) = WC(X^{RC}, Y^R) = k$ .

$WC(X, Y^C) = WC(X^C, Y) = WC(X^R, Y^{RC}) = WC(X^{RC}, Y^R) = k$ . The four group of basepairs complement relationship is equivalent, as shown in Fig. 3.

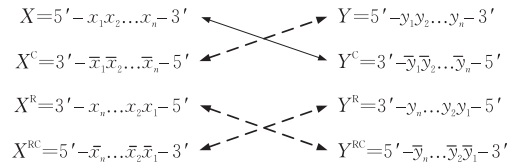


Fig. 3  $X$  and  $Y^C$  complement basepairs equivalent relationship

And for the same reason,  $X$  and  $X^R$  also contain four group equivalent relationships,  $WC(X, Y^R) = WC(X^R, Y) = WC(X^{RC}, Y^C) = WC(X^C, Y^{RC})$ , as shown in Fig. 4.

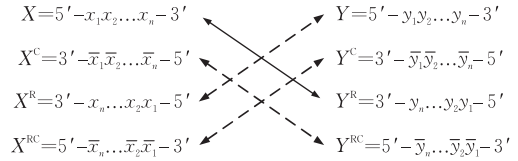


Fig. 4  $X$  and  $Y^R$  complement basepairs equivalent relationship

Through the above analysis, for any given Hamming distance constraints threshold value, the calculation of  $WC(X, Y^R)$  and  $WC(X, Y^C)$  can contain 8 specific types of hybridization relationship.

##### 3.2.2 Calculation analysis of free energy method

There is a significant different between free energy calculation and Hamming distance calculation, which is the nearest-neighbor model has the  $5'-3'$  direction and the thermodynamic parameters are not symmetry. Therefore Hamming distance equivalent DNA sequences, it will not necessarily equal free energy.

For example, DNA sequences  $X = 5' - \text{CACT-CATGAA} - 3'$ ,  $Y = 5' - \text{CACACCTGAA} - 3'$ . According to the Hamming distance constraints character, there are equivalence Hamming distance relations  $WC(X, Y^C) = WC(X^R, Y^{RC}) = WC(X^C, Y) = 8$ . However, its corresponding free energy is different, as shown in Fig. 5.

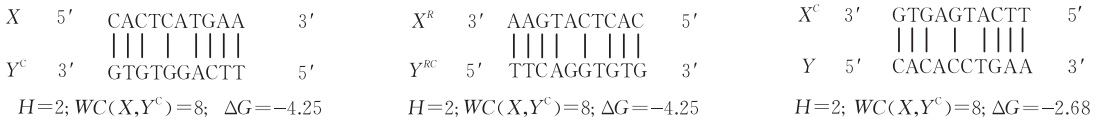


Fig. 5 Hamming distance is equal, but free energy is different

① the free energy of  $X$  and  $Y^C$

$$\begin{aligned}
 X^C \quad 5' - \text{CACTCATGAA} - 3' &= \text{CA} + \text{AC} + \text{CT} + \text{GT} + \text{CA} + \text{AG} + \text{TG} + \text{GA} + \text{AA} + \Delta G(\text{ini GC}) + \Delta G(\text{ini AT}) \\
 Y^C \quad 3' - \text{GTGTGGACTT} - 5' &= \text{GT} \quad \text{TG} \quad \text{GT} \quad \text{CT} \quad \text{GG} \quad \text{TA} \quad \text{AC} \quad \text{CT} \quad \text{TT} \\
 &= -1.45 - 1.44 - 0.12 + 0.45 + 0.03 + 0.02 - 1.45 - 1.30 - 1.00 + 0.98 + 1.03 \\
 &= -4.25
 \end{aligned} \tag{7}$$

② the free energy of  $X^R$  and  $Y^{RC}$

$$\Delta G(X^R, Y^{RC}) = \Delta G(X, Y^C) = -4.25.$$

③ the free energy of  $X^C$  and  $Y$

$$\begin{aligned}
 Y \quad 5' - \text{CACACCTGAA} - 3' &= \text{CA} + \text{AC} + \text{CA} + \text{GA} + \text{CC} + \text{AT} + \text{TG} + \text{GA} + \text{AA} + \Delta G(\text{ini GC}) + \Delta G(\text{ini AT}) \\
 X^C \quad 3' - \text{GTGAGTACTT} - 5' &= \text{GT} \quad \text{TG} \quad \text{GA} \quad \text{CA} \quad \text{GT} \quad \text{TC} \quad \text{AC} \quad \text{CT} \quad \text{TT} \\
 &= -1.45 - 1.44 + 0.43 + 0.17 + 0.62 + 0.73 - 1.45 - 1.30 - 1.00 + 0.98 + 1.03 \\
 &= -2.68
 \end{aligned} \tag{8}$$

In addition, for any two DNA sequences  $X$  and  $Y$ , let  $\Delta G(X, Y^C)$  be the energy of  $X$  and  $Y$ , the corresponding reverse sequences free energy is the equal to  $\Delta G(X, Y^C)$ , i. e.  $\Delta G(X, Y^C) =$

$\Delta G(X^R, Y^{RC})$ . And for the same reason,  $\Delta G(X^C, Y) = \Delta G(X^{RC}, Y^R)$ ;  $\Delta G(X, Y^R) = \Delta G(X^R, Y)$ ;  $\Delta G(X^{RC}, Y^C) = \Delta G(X^C, Y^{RC})$ , as shown in Fig. 6.

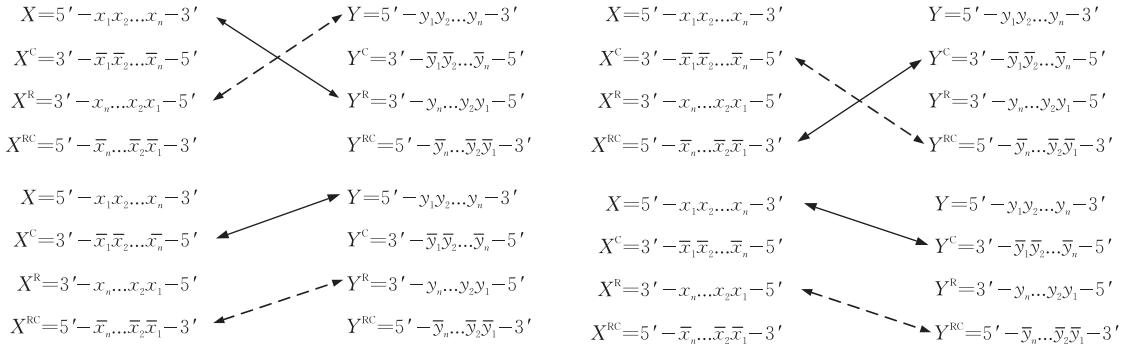


Fig. 6 Equivalent free energy calculation relationship

### 3.3 Comparison of Nucleic Acid Encoding

Through the above analysis we can see that the two catalogs DNA sequences design method can both restrict the 8 type non-specific hybridizations. The Hamming distance based method need only calculate 2 non-specific hybridization combinations, while free energy based method must calculate 4 non-specific hybridization combinations. It is clear that the computationally of free energy based DNA sequence method greater than Hamming distance based DNA sequence design method.

However, through the comparison of formula 7 and formula 8, we can see that, if only consider the free energy of Watson-Crick base pairs without the free energy of internal single mismatch and

other DNA secondary structure,  $\Delta G(X, Y^C)$  will be equal to  $\Delta G(X^C, Y)$ . At this point, the calculation of free energy equivalent to the Hamming distance,  $\Delta G(X, Y^C) = \Delta G(X^C, Y) = \Delta G(X^R, Y^{RC}) = \Delta G(X^{RC}, Y^R)$ , and  $\Delta G(X, Y^R) = \Delta G(X^R, Y) = \Delta G(X^{RC}, Y^C) = \Delta G(X^C, Y^{RC})$ . When only consider the free energy of Watson-crick base pairs, the calculation of free energy will be equal to the Hamming distance method, as shown in Fig. 3 and Fig. 4.

In addition, Hamming distance based method can only approximate estimates of the stability of the nucleic acid hybridization. For example,  $X = 5' - \text{GGCTAACT} - 3'$ ,  $Y = 3' - \text{CCGAGAAT} - 5'$ ,  $Z = 3' - \text{CAGTTAGC} - 5'$ .  $WC(X, Y) = WC(X, Z)$ , while  $\Delta G(X, Y) \neq \Delta G(X, Z)$ , as shown in Fig. 7.



Fig. 7 Hamming distance can only approximately estimate the stability of hybridization

Two group of DNA sequence with same Hamming distance have widely different free energy  $\Delta G$ . As a result, the Hamming distance based DNA sequence design method can only approximately estimate the thermodynamic stability of DNA hybridization, and can not be a perfect substitute for the minimum free energy method.

## 4 Conclusion

In order to improve the reliability and effectiveness of the experiment, DNA computation need to design reliable libraries of DNA sequences. The paper introduced two typical nucleic acid sequence methods, free energy based method and Hamming distance based method, and analyzed the influence of constraints on the quality of DNA sequences. Then, the paper compared the integrity and computational cost of two methods for excluding the non-specific hybridization sequence. In addition, the paper analyzed the efficiency of two nucleic acid sequence design methods. Through the comparison we can see that the two catalogs DNA sequences design method can both restrict the all type non-specific hybridizations, and the computationally of free energy based DNA sequence method is greater than Hamming distance based DNA sequence design method. Further analysis shows that when only consider the free energy of Watson-crick base pairs, the calculation of free energy will be equal to the Hamming distance method. So, if the precision satisfy the requirement of experiment, Hamming distance based DNA sequence design method, not only can effectively distinguish between specific and non-specific hybridization, and effectively reduce the amount of calculation, improve the efficiency of the DNA sequence designing.



**ZHANG Kai**, born in 1979, Ph. D. . His current research interests include DNA computing and operational research.

## References

- [1] Adleman L M. Molecular computation of solutions to combinatorial problems. *Science*, 1994, 266: 1021-1024
- [2] Faulhammer D, Cukras A R, Lipton R J et al. Molecular computation: RNA solutions to chess problems. *Proceedings of the National Academy of Sciences*, 2000, 97(4): 1385-1389
- [3] Garzon M, Deaton R, Neathery P et al. On the encoding problem for DNA computing//*Proceedings of the 3rd DIMACS Workshop on DNA-Based Computing*. University of Pennsylvania, Preliminary Proceedings, 1997: 230-237
- [4] Frutos A G, Liu Qinghua, Thiel A T et al. Demonstration of a word design strategy for DNA computing on surface. *Nucleic Acids Research*, 1997, 25: 4748-4757
- [5] Shin S Y, Lee I H, Kim D et al. Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing. *IEEE Transactions on Evolutionary Computation*, 2005, 9(2): 143-158
- [6] Arita M, Nishikawa A, Hagiya M et al. Improving sequence design for DNA computing//Whitley D et al eds. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2000)*. Morgan Kaufmann, 2000: 875-882
- [7] Deaton R, Chen J, Bi H et al. A software tool for generating non-crosshybridization libraries of DNA oligonucleotides//*Proceedings of the 8th International Workshop DNA Based Computing*, 2002: 252-261
- [8] Tanaka F, Kameda A, Yamamoto M et al. Design of nucleic acid sequences for DNA computing based on a thermodynamic approach. *Nucleic Acids Research*, 2005, 33(3): 903-911
- [9] Garzon M, Deaton R, Nino L F et al. Genome encoding for DNA computing//*Proceedings of the 3rd DIMACS Workshop on DNA-Based Computing*. University of Pennsylvania, 1999: 230-237
- [10] Tanaka F, Nakatsugawa M, Yamamoto M et al. Towards a general purpose sequence design system in DNA computing//*Proceedings of the 2002 Congress on Evolutionary Computing*, 2002: 73-78
- [11] SantaLucia J Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences*, 1998, 95: 1460-1465

**GENG Xiu-Tang**, born in 1975, Ph. D. . His fields of interest are soft computing, aviation electron, system modeling and optimization, DNA computing and so on.

**XIAO Jian-Hua**, born in 1979, Ph. D. , lecturer. His research interests include Bio-inspired computation and logistics optimization.

**ZHAO Dong-Ming**, born in 1977, Ph. D. . His research interests include DNA computing and Bio-inspired computation.