

利用脱氧核酶调控基因表达的 DNA 计算模型研究

张凤月¹⁾ 宋海刚²⁾ 温俊海¹⁾ 张 强³⁾ 刘 波⁴⁾

¹⁾(北京理工大学生命科学与技术学院生物医学与工程系 北京 100081)

²⁾(中国科学技术部基础研究管理中心 北京 100862)

³⁾(大连大学辽宁省智能信息处理重点实验室 辽宁 大连 116622)

⁴⁾(中国检验检疫科学研究院 北京 100029)

摘 要 用于逻辑调控基因表达分子自动机的研究是 DNA 计算的重要研究领域. 文中将脱氧核酶技术应用于 DNA 计算研究当中, 利用脱氧核酶的特性, 特别是可以作为反义药物的特点, 作为构建分子自动机的主要材料, 设计了调控基因 H-ras 表达的 DNA 计算模型, 而且模型也可适用于其它过表达基因的调控. 结合 DNA 计算具备的高度并行性和智能性的优点, 该模型为 DNA 计算在基因表达调控方面的应用做了进一步探索.

关键词 DNA 计算; 脱氧核酶; 基因表达调控

中图法分类号 TP301

DNA Computing Model with DNzyme to Control Gene Expression

ZHANG Feng-Yue¹⁾ SONG Hai-Gang²⁾ WEN Jun-Hai¹⁾ ZHANG Qiang³⁾ LIU Bo⁴⁾

¹⁾(Department of Biomedical Engineering, School of Life Science and Technology, Beijing Institute of Technology, Beijing 100081)

²⁾(Fundamental Research Management Center of Ministry of Science and Technology of China, Beijing 100862)

³⁾(Liaoning Key Laboratory of Intelligent Information Processing, Dalian University, Dalian, Liaoning 116622)

⁴⁾(Chinese Academy of Inspection and Quarantine, Beijing 100029)

Abstract A molecular computer for logical control of gene expression is an essential research field of DNA computing. In this paper, with its characteristic especially being antisense medication, DNzyme technique is used in construction of DNA computing model as basis material. Furthermore, a DNA computing model is designed to control H-ras gene expression, and this model also has potential to control other over expressed genes. Benefited from the massive parallelism and intelligence feature of DNA computing, this model explores further application of DNA computing to control of gene expression.

Keywords DNA computing; DNzyme; control of gene expression

1 Introduction

Scientific researcher attempted to use the organism's biological macromolecule (DNA, RNA, proteins, etc.) to improve the electronic computer or construct a new type of computer vision was al-

ready long-standing, but it was not until 1994 when Adleman successfully solved a 7 vertexes of the Hamilton Road Problem with DNA molecules, this aspect's research only then obtained a major breakthrough^[1]. Afterward, the DNA computing has made the outstanding contributions on solving

收稿日期:2008-04-20;最终修改稿收到日期:2008-11-02. 本课题得到国家自然科学基金(60403001,30740036,30870573)、辽宁省智能信息处理重点实验开放课题(2005-1)以及新世纪优秀人才支持计划资助. 张凤月,男,1973年生,博士,讲师,主要研究方向为生物计算和生物信息学. 宋海刚,男,1975年生,博士,助理研究员,主要研究方向为系统工程、计算机应用. 温俊海,男,1969年生,博士,副教授,主要研究方向为图像处理与重建. 张 强,男,1971年生,博士,教授,主要研究领域为 DNA 计算、神经网络、计算机动画. 刘 波,女,1973年生,博士,副研究员,主要研究方向为 DNA 计算和植物保护. E-mail: liubobj@126.com.

hard computational problems such as NP problem, which is difficult to solve by electronic computer currently. The representative of the study include: Lipton proposed a solution of Satisfiability problem (SAT) by using a simple contact network in a tube in 1995^[2]; Ouyang imitated Adleman's method to put forward a solution for the Maximal Clique Problem in 1997^[3]; Eric tried to solve SAT by a cellular automata model as well as by the plasmid suggested in reference [4]; Qihuang Liu, Wisconsin-Madison University, did the surface computing and solved a 4 variables SAT problem in 2000, this experiment indicated the possibility of constructing DNA computer on the surface^[5]; Braich solved a 20 variables of the 3-SAT problem based AcryditeTM separation technology in 2002^[6]. In addition, for the general model of DNA computing, researchers take the Sticker Systems^[7] and the Splicing System model^[8] as representative, and many related research results has been reported.

In 2001, the programmable semiautomatic tube DNA computer based on the biological macromolecule, which proposed by Benenson from Israeli Weitzman Academy of Science caused the DNA computation research to obtain unprecedented progress^[9]. In 2002, a DNA computer was developed for the analysis of gene expression by Suyama. This computer was mainly composed of the molecular computation module and the diagnostic part. The former selected the correct result responded through the biochemistry reactions, the latter carried on analysis to the obtained result^[10]. After proposed the programmable semiautomatic tube DNA computer and unified automata theory and life science, Benenson put forward a molecular automaton for the logical regulation gene expression to open up a new research direction in DNA computation in 2004^[11], which related test simulation has been carried on *in vitro*^[12]. This automaton is mainly composed of three programmable modules, respectively were the computation module, the input module and the output module. The elementary function of this automaton is to diagnose the specific mRNA density automatically in tube according to the hypothesis logical rule, and to produce the suppressor for gene expression corresponding the diagnosis result (or release drugs). Thereby it can be used as gene therapy, i. e. anti-sense therapy^[13]. And the essential of this computation module is using endonuclease Fok I as the hardware molecule, and after operation and transition, this automaton have two conditions of "yes"

and "no", corresponds in the biological "positive diagnosis" and "negative diagnosis" separately. Afterward, the related Turning machine, the biological Turning machine and the molecular automaton model also obtained further elaborated^[14]. In 2006, Georg designed a logic operation model without enzyme, which using short oligonucleotides as input and output and implement Boolean logic operation function ("AND" "OR" and "NO")^[15]. Carrying on the multistage logic operations on the chip surface, a DNA logic computing model was proposed with more complex structure^[16], in addition, according to the Boolean logic operation, the use of oligonucleotide-based distribution so as to produce output fluorescent molecule, also caused using DNA to carry out the logic operation examination easier^[17]. Harvard University system biology's scholar further designed a DNA logic operation model with RNAi technology in 2007, this model was confirmed that it could be applied in artificial cultivation of renal cell^[18]. With the use of molecular release's configurationally entropy, to design specific nucleotide sequences to catalyze the release of the output molecular can amplify the logic operation signal^[19]. In 2008, Peng Yin designed a variety of complex structures like linear, branched, circuit and autonomous locomotion by the use of DNA strand self-assembling and disassembling, which has established more solid foundation for DNA computing^[20].

In this paper, a DNA computing model is proposed by using DNAzyme technology of molecular biology, after integrated molecular automata for logic control of gene expression with cancer research related to polygenes. This study, DNAzyme-based molecular automata for logic regulation of gene expression, is expected to improve its application in tumor treatment, gene regulation and so on. Furthermore, this research provides a potential means for research of regulation of gene expression and establishes a basis for its application in living organisms.

2 DNAzyme-Based DNA Computing Model for Regulation Gene Expression

2.1 DNAzyme

DNAzyme (Deoxyribozyme, DNA), is a catalytic DNA molecules, also known as DNA enzyme or enzyme-DNA, and is a new member of biological catalysts together with protein enzymes and ribozyme. The main enzyme activity are ligase, metal chelating and phosphatase activity.

It has been proved that DNAzyme 10-23 can snip mRNA in the AU point, thereby, in theory only through adjusting the nucleotide components of its substrate's recognition site, any of the mRNA start code can be cleaved, which implies a master key for the control protein expression^[21]. So DNAzyme has potential applications in many areas such as gene therapy, disease diagnosis and analysis tools of molecular biology^[22-23]. As a special kind of DNA molecules, the special features of DNAzyme make it suitable for DNA computing, especially as a active molecular for the regulation of gene expression; and it is possible of using DNAzyme to design logical operation due to the independence of the its domains. The DNAzyme 10-23 molecular model is shown in Fig. 1, and its DNAzyme sequence is indicated in Fig. 2.

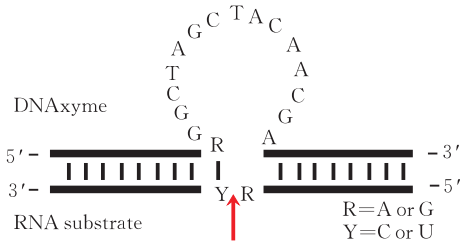


Fig. 1 10-23 DNAzyme molecular model

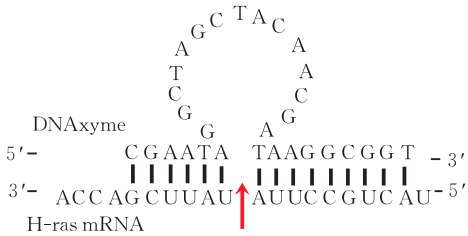


Fig. 2 DNAzyme sequence of the proposed DNA computing model and its active site in H-ras mRNA molecules

2.2 Selection of the Target Gene

With the in-depth research in the life sciences, in particular the completion of human genome project (HGP) and the startup of post HGP, lots of disease genes have been revealed and some diseases will be cured through gene therapy. It is now unambiguous that the tumor is a polygenic disease, due to the inactivation of tumor suppressor genes and the activation of cancer genes, as well as the disorder of cell cycle regulation gene on the level of gene expression. For example, when p16 (tumor suppressor genes) and CyclinD1 (cancer gene) gene bind CDK4 (cyclin-dependent kinase), the former inhibits cell growth and the latter promotes cell growth. That missing p16 or arising CyclinD1 expression will cause uncontrolled cell growth and in-

duce cell cancer finally^[24]. In human tumors, a class of important genes, ras gene family(including three functional genes that H-ras, K-ras and N-ras), has to be taken seriously, the current researches have shown that malignant tumors are associated with disorder of p16, CyclinD1 and H-ras gene expression in gastric cancer, colorectal cancer, and other common gastrointestinal cancer.

Therefore, the rules to determine these cancer cells can be summed up as: p16 ↓ CyclinD1 ↑ H-ras ↑ (↓ lower expression; ↑ over expression).

The human H-ras gene was chosen as controlled target to construct a DNA computing model for regulation of gene expression. Human H-ras is positioned in chromosome 11, its family encode highly similar proteins of which the molecular weight is 2100 Dalton and called p21 protein. The p21, located in the inner membrane, plays an important role in adenosine receptor cyclase signaling pathway with the GTP activity. Searched through bioinformatics methods, the entire length of human H-ras gene sequence is 41026 bp. An EST sequence, which is a gene transcription fragment of the mRNA, is selected as the target for design of the DNA computing model, and the corresponding cDNA sequences is following:

1 ccggcctnngg ncccggcctt ggncccgggg gcagtcgcgc
ctgtgaacgg tggggcagga gaccctgtag aggaccccg 81
ggccgcaccg tggaggagcg atgacggaat ataagctggt ggtg-
gtgggc gccggcggtg tgggcaagag tgcgctgacc 161 atc-
cagctga tccagaacca ttttgtggac gaatacgacc ccactataga
gggttntctac cggaagcagg tgg.

2.3 Molecular Automaton Model

The molecular automata model shown in Fig. 3 includes three modules. The hardware module is a restrictive endonuclease which can recognize and snip a specific sequence of double-stranded DNA and then form a sticky or blunt ends; The software module contains the restrictions enzyme site, the interval sequence and a double-stranded DNA with sticky ends.

The input and output module are DNA molecules, and encoded DNA sequences are taken as state set and symbol set. The state set include state of S0 and S1, as well as the symbol set is “a” and “b”; the transit rules are described. If the current state is S0 and the symbol read is “b”, then the unit, which is composed of the hardware module and the software module, will change its state to S1 and the symbol to “a”, and move one position to the right as well. The output state is S0 and the symbol is “t” (terminator) after

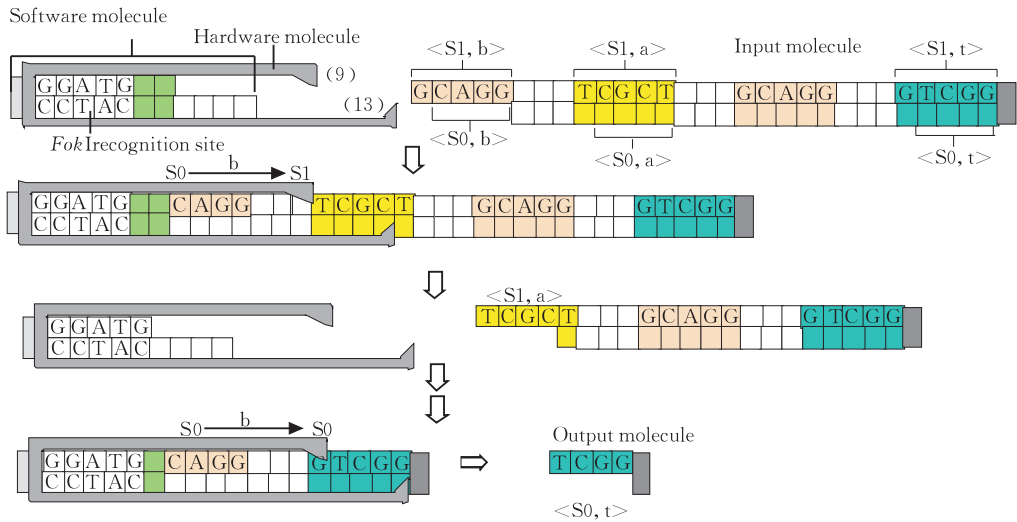


Fig. 3 The theoretical model of molecular automata

operations.

2.4 Model Design

Based on the research about the theoretical model of molecular automata, A DNA computing model is designed to regulate H-ras gene expression after the target gene is selected and the related DNAzyme is constructed.

2.4.1 Basic principles of model

The model encoding has been checked by related bioinformatics software in order to ensure the specificity between the designed DNA computing models to the regulated target. The specified region, from 181 to 210 nucleotide of the target gene, serves as a template for model design, and

the snip site of DNAzyme is the AU sequence of between 115 to 120 nucleotide.

2.4.2 Regulate rules of model

The DNA computing model recognizes the level condition of gene H-ras expression, and releases inhibiting molecules if the gene is over expression, or maintains the original state if it is normal.

2.4.3 Composition and function of DNA computing model

The designed DNA computing model includes three parts of which are computing module, diagnostic module and output module. The respective structure and their relation are described in Fig. 4, and encode of these sequences are listed in Table 1.

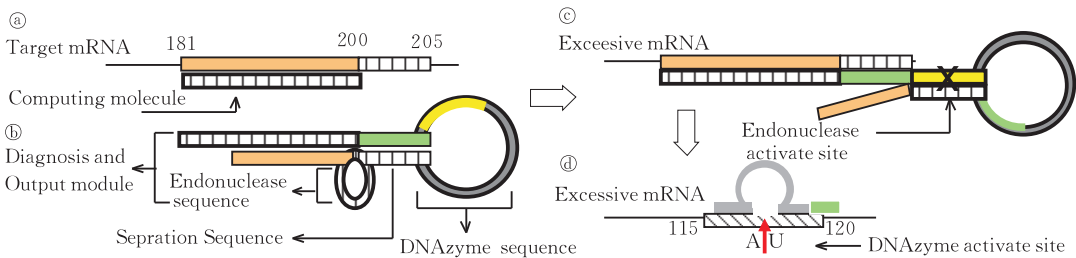


Fig. 4 The structure of DNA computing model based DNAzyme 10-23 for regulation H-ras gene expression and the computing processes

Table 1 DNA Computing Model Components and Encode

Components	Strands	Sequences(5'-3')
computing module	sequence	AAAACACCTG CTTATGCTGG
	activate site	UUUUGUGGAC GAAUACGACC
diagnostic module	diagnostic strand	AAAACACCTG CTTATGCTGG
	complement strand	AC GAATACGACC
	separation strand	CCACT
	DNAzyme sequence	TGATCA
output module	DNAzyme sequence	ACTAGT
	regulation region	CGAATA TAAGGCGGT
	activate site	ACUGCCUUA UAUUCG
	coral site	GGCTAGCTACAACGA

The purpose of computing module is for the detection of target mRNA. It is a ssDNA strand and the encode is complementary to the region from 181 to 200 nucleotide of H-ras cDNA.

The function of the diagnostic module is transits state according to the proceeded results of the computing module; the sequences of the diagnostic module are composed of diagnostic strand and its complementary, separation stand and restriction enzyme strand. The diagnostic strand forms double-stranded structure with its complementary. The diagnostic strand and the computing module share the same encoding strategy. The structure of the separate stand is double-stranded and its encode is as same as that of the region of 200 to 210 nucleotide of H-ras cDNA. Located in the middle of the diagnostic and the separate stand, the restriction enzyme sequence is only single-stranded and forms a circle structure.

An endonuclease sequence and a DNAzyme sequence constitute the output module. The endonuclease sequence is complement to the endonuclease sequence in the diagnostic module. As a suppressor, the structure and activate site of DNAzyme is indicated in Fig. 3, of which the strand of CGAATA TAAGGCGGT recognize and complete to the sequence of 115 to 120 nucleotide of target mRNA. The stand of GGCTAGCTACAACGA is the coral site because it can snip AU site specifically.

2.5 Biological Computing Processes of the Model

The first step, a certain amount of computing module (equivalent to the number of mRNA expressed by gene H-ras) was inserted into the target cell's cytoplasm, in there is the transcript including the mRNA of gene H-ras. According to the principle of Watson-Crick, the computing module is complementary to the region of 181 to 200 nucleotide of mRNA. Then

(1) If the amount of computing module \geq the amount of transcription mRNA of gene H-ras, all the region of 181 to 200 nucleotide of these mRNA will be double-stranded structure (20bp), which is shown in Fig. 4(a);

(2) If the amount of computing module $<$ the amount of transcribed mRNA of gene H-ras, the region of 181 to 200 nucleotide of those redundant mRNA will remain the original single-strand structure.

The second step, a certain amount of the diagnostic module enter the cell, and because the diagnostic strand of this module is the same as the computing module's, then:

(1) If the first step produces the results of (1), the diagnostic module will remain the original structure indicated by Fig. 4(b);

(2) If the first step produces the results of (2), the diagnostic module will form complementary structure with these redundant mRNA in the region of 181 to 200 nucleotide (25bp), which is shown in Fig. 4(c);

The third step, if the second step produces the results of (2), the complementary sequence (12 bases) will be single strand due to its molecular force less than double-stranded force. So the exposed single-stranded restriction enzyme sequence form a completed double-stranded site with the complementary sequence of the output module, and the diagnostic module changes its state indicated by Fig. 4(c).

The fourth step, the DNAzyme is released by adding restriction enzyme Bcl I, which snip the output module in site. Furthermore, the DNAzyme cleavers mRNA molecules in AU site after it is attached to the region of 115 to 120 nucleotide, so the over expressed gene, H-ras, is properly regulated, and the operation is indicated in Fig. 4(d).

After these four steps operations, the DNA computing based DNAzyme fulfills the designed purpose and maintains the normal function of cells through diagnose target gene and degradation excessive mRNA when gene is over expressed.

3 Conclusion and Prospect

Gene expression and regulation of organisms are the essential function, any out of control in this function will no doubt have adverse effects, such as has been shown that cancer is the outcome of a polygenes interaction, of which oncogene activation and over expression play an important role. In this paper, a DNA computing model with DNAzyme to regulate gene expression of H-ras gene is constructed in theory. This model has a good versatility in the study of Gene expression and regulation, especially over expressed genes, since the output molecules is DNAzyme. For example, only taken different encode, the model will adapt to regulate over expressed CyclinD1 gene.

Despite that the current DNA computing models for regulation of gene expression are not very complicated as well as the target genes are less, it is enough for scientists to create biosensors and drug transport systems for specific human cells. In addition to the treatment of cancer, these research-

ches have good prospects in functional genomics, especially to identify complex signaling pathway in a comprehensive relations of different genes. From that point of view that the non-linear trait is mainly characteristic of life system itself, DNA computation with its high parallelism, intelligent and affinity to organisms will be a potential research tool for complex phenomenon of life research, further study will have a major significance both in the field of information science and life science.

References

- [1] Adleman L M. Molecular computation of solution to combinatorial problems. *Science*, 1994, 266: 1021-1024
- [2] Lipton Richard J. DNA solution of hard computational problems. *Science*, 1995, 268: 542-545
- [3] Ouyang Qi. DNA solution of the maximal clique problem. *Science*, 1997, 278: 446-449
- [4] Head T, Lommerce P H M, Spaink H P. Computing with DNA by operating on plasmids. *BioSystems*, 2000, 57: 87-93
- [5] Liu Qing-Hua, Wang Li-Man, Anthony G F, Anne E C et al. DNA computing on surfaces. *Nature*, 2000, 403: 175-178
- [6] Braich R S, Chelyapov N, Johnson C et al. Solution of a 20-Variable 3-SAT problem on a DNA computer. *Science*, 2002, 296: 499-502
- [7] Kari L, Paun G, Rozenberg et al. DNA computing, sticker systems, and universality. *Acta Informatica*, 1998, 35: 401-420
- [8] Paun G, Rozenberg G, Salomaa A. *DNA Computing: New Computing Paradigms*. Berlin: Springer-Verlag, 1998
- [9] Benenson Y, Paz-Elizur T, Adar R et al. Programmable and autonomous computing machine made of biomolecules. *Nature*, 2001, 414: 430-434
- [10] Sakakibara Y. Intelligent DNA chips: Logical operation of gene expression profiles on DNA computers. *Genome Informatics*, 2002, 11: 33-42
- [11] Benenson Y, Shapiro E. *Molecular computing machines*// Schwarz James A, Contescu Cristian I, Putyera Karol eds.

Dekker Encyclopedia of Nanoscience and Nanotechnology. New York: Marcel Dekker, Inc., 2004: 2043-2056

- [12] Benenson Y, Gil B, Ben-Dor U, Adar R et al. An autonomous molecular computer for logical control of gene expression. *Nature*, 2004, 429: 423-429
- [13] Anne C. Automata make antisense. *Nature*, 2004, 427: 351-352
- [14] Shapiro E, Benenson Y. Bringing DNA computers to life: Tapping the computing power of biological molecules gives rise to tiny machines that can speak directly to living cells. *Scientific American*, 2006, 5: 45-51
- [15] Georg S, David S, David Y Zh et al. Enzyme-free nucleic acid logic circuits. *Science*, 2006, 314: 1585-1588
- [16] Brian M F, Scott L C, Ghadiri M. Modular multi-level circuits from immobilized DNA-based logic gates. *Journal of the American Chemical Society*, 2007, 129: 14875-14879
- [17] Wataru Y, Yohei Yokobayashi. Photonic boolean logic gates based on DNA aptamers. *Chemical Communications*, 2007 (2): 195-197
- [18] Keller R, Leonidas B, Rohan M et al. A universal RNAi-based logic evaluator that operates in mammalian cells. *Nature Biotechnology*, 2007, 25: 795-801
- [19] David YZh, Andrew J T, Bernard Y et al. Engineering entropy-driven reactions and networks catalyzed by DNA. *Science*, 2007, 318: 1121-1125
- [20] Peng Yin, Harry M T Choi, Colby R Calvert. Programming biomolecular self-assembly pathways. *Nature*, 2008, 451, 318-322
- [21] Santoro S W, Joyce G F. Mechanism and utility of an RNA-cleaving DNA enzyme. *Biochemistry*, 1998, 37: 13330-13342
- [22] Achenbach J C, Chiuman W, Cruz R P, Li Y. DNazymes: From creation in vitro to application *in vivo*. *Current Pharmaceutical Biotechnology*, 2004, 5: 321-336
- [23] Cairns M J, Hopkins T M, Witherington C, Sun L Q. The influence of arm length asymmetry and base substitution on the activity of the 10-23 DNA enzyme. *Antisense and Nucleic Acid Drug Development*, 2000, 10: 323-332
- [24] Azechi H, Nishida N, Fukuda Y et al. Disruption of the p16/Cyclin D1/Retinoblastoma protein pathway in the majority of human hepatocellular carcinomas. *Oncology*, 2001, 60: 346-354



ZHANG Feng-Yue, born in 1973, Ph. D., lecturer. His research interests include biological computing and bioinformatics.

SONG Hai-Gang, born in 1975, Ph. D., assistant researcher. His research interests include systems engineering

and computer applications.

WEN Jun-Hai, born in 1969, Ph. D., associate professor. His research interests include image processing and image reconstruction.

ZHANG Qiang, born in 1971, Ph. D., professor. His research interests include DNA computing, neural networks and computer animation.

LIU Bo, born in 1973, Ph. D., associate researcher. Her research interests include DNA computing and plant protection.