01 Jan 2004

Interactive DNA Sequence and Structure Design for DNA Nanotechnology and DNA Computation

Mingjun Zhang

Chaman Sabharwal
Missouri University of Science and Technology, chaman@mst.edu

Weimin Tao

Tzyh-Jong Tarn

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/comsci_facwork/265

Recommended Citation

M. Zhang et al., "Interactive DNA Sequence and Structure Design for DNA Nanotechnology and DNA Computation," *Proceedings of the 4th IEEE Conference on Nanotechnology, 2004*, Institute of Electrical and Electronics Engineers (IEEE), Jan 2004.
The definitive version is available at https://doi.org/10.1109/NANO.2004.1392331

2004 4th IEEE Conference on Nanotechnology

Interactive DNA Sequence and Structure Design for DNA Nanotechnology and DNA Computation

Mingjun Zhang*, Chaman L. Sabharwal†, Weimin Tao‡, Tzyh-Jong Tarn§,Ning Xi¶ and Guangyong Li¶
*Life Sciences and Chemical Analysis Division, Agilent Technologies, CA. Email: mingjunzhang@ieee.org
†Department of Computer Science, University of Missouri-Rolla, MO. Email: chaman@umr.edu
‡Brooks Automation, CA. Email: david_tao@brooks-pri.com
§Department of Electrical and Systems Engineering, Washington University in St. Louis, MO. Email: tarn@wuauto.wustl.edu
¶Department of Electrical Engineering, Michigan State University, MI. Email: {xin,liguangy}@egr.msu.edu

Abstract— **DNA sequence and structure design is very important for DNA nanotechnology and DNA computation. A computer aided design tool is needed for exploring DNA sequence and structure of interests before experimental synthesis, which is a very time and labor consuming process. In this paper, an interactive DNA sequence and structure design software tool called DNA shop is proposed and implemented.**

I. Introduction

As the carrier of genetic information for all living species, DNA has been well-known for its unique properties of information encoding, structure self-recognition and self-assembly. DNA computation and DNA nanotechnology are two emerging fields aiming to use these properties.

DNA nanotechnology takes advantage of the fact that the intermolecular interactions of DNA are highly specific and readily programmed through Watson-Crick complementary property. Possible applications include scaffolds for molecular electronic devices and nanometer-scale robots construction. Seeman [7] proposed DNA as scaffolds to organize structures of other molecules. Mao et al. [6] showed that self-assembly structure of branched DNA motifs can provide basis for dynamic assembling switchable molecular machines. Yurke et al. [9] reported that the construction of a DNA machine in which the DNA is used not only as a structural material, but also as fuel.

In addition, the combination of DNA information-encoding and recognition properties, and the enzymatic machinery capability for DNA manipulation facilitate the emergence field of DNA computation. The feasibility of DNA computation was first demonstrated by Adleman in 1994 [1]. Benenson [2] discussed a programmable finite automation comprising DNA and DNA-manipulating enzymes that solves computational problems autonomously. The automation's hardware consists of a restriction nuclease and ligase, the software and inputs are encoded by double-stranded DNA, and programming amounts to choose appropriate software molecules. Upon mixing solutions containing these components, the input molecules are processed via a cascade of restriction, hybridization and ligation cycles, producing a detectable output molecules that encode the computational results.

For above DNA nano-applications, DNA sequence and structure design is a critical step. Currently, one must obtain proper conditions, refine designs and determine experimental windows for DNA structure design through tedious and often expensive processes of trials and errors [8]. A visualization tool is needed for DNA sequence and structure design. The purpose of this paper is to propose a software tool for DNA sequence and structure design.

II. DNA Properties

DNA's unique biological properties include the specificity of the base pairing that holds two strands of double helix together: Adenine (A) pairs with Thymine (T) and Guanine (G) pairs with Cytosine (C). Usually, DNA double helix generated from the complementary interactions is a linear molecule. Its axis is not branched in the biological sense. However, branched DNA molecules do occur as key intermediate state in DNA metabolism, particularly in the processes of replication, recombination.

Properties of DNA double helix are unlike those of any other natural or synthetic polymers. The molecule's characteristic base stacking and braided architecture lend it unusual stiffness: it takes about 50 times more energy to bend a double-stranded DNA molecule into a circle than to perform the same operation on single-stranded DNA. Furthermore, the double-stranded DNA molecule is very stable. These features make the double-stranded DNA molecule a great candidate of scaffolds for other molecules. DNA electrical properties make DNA one of the most interesting bio-molecules for molecular electronics [10].

DNA molecules are in the nanometer scale and encode binary information. A DNA double helix is about 2 nm in diameter with helical repeat of about 10 base pairs, which produces a pitch of 3.4-3.6 nm. A small volume of DNA contains a vast number of molecules. DNA in weak solution of one liter of water can encode 10^7 to 10^8 tera-bytes information, which makes DNA molecule a potential material for information storing. DNA sequence may be used for encoding information that can be read externally by proteins and nucleic acids. Most current data storing media has a life around 100 years. DNA molecules may have much longer, stable and larger capacity for information storage.

0-7803-8536-5/04/$20.00 ©2004 IEEE

A. DNA sticky ends

DNA can be cut at precise locations by using restriction enzymes. Another enzyme–DNA ligase–can then be used to reassemble the pieces into any desired order. Together, these two enzymes allow researchers to assemble customized DNA structures. For example, restriction enzymes typically recognize a symmetrical sequence of DNA, such as the site of EcoRI:

$$---GAATTC---\quad (1)$$

$$---CTTAAG---\quad (2)$$

For the DNA sequences, the top strand is the reverse of the bottom strand. Using restriction enzymes to react at this cite and break hydrogen bonds holding the overlapping single stranded complementary strands, which are the strands between G and A, overhanging chains as follows can be obtained:

$$---G\qquad AATTC\quad ---\quad (3)$$

$$---CTTAA\qquad G\quad ---\quad (4)$$

The segments $AATT$ in (3) and $TTAA$ in (4) are called sticky ends, which are complementary each other. DNA sticky ends provide a predictable, diverse, reliable and programmable set of intermolecular interactions. DNA molecules can be manipulated by commercially available enzymes – they can be joined by DNA ligases, cleaved at specific sites by restriction enzymes, phosphorylated by kinases and have their topology altered by DNA topoisomerases.

B. DNA self-assembly

The capability of fabricating individual molecules and atoms is the key for nanotechnology. Self-assembly, which is a method for constructing structures by spontaneously self-ordering of substructures, is an attractive approach for nano-structure fabrication. The technique works by simulating the way biological systems build molecules, viruses, and cells. DNA self-assembly presents a bottom-up approach to fabricate nano-scale objects. DNA self-assembly uses artificially synthesized single strand DNA to self-assemble into different DNA crossover molecules (tiles), which have sticky ends that preferentially match the sticky ends of certain other DNA tiles, facilitating the further assembly into tiling lattices – DNA machines.

III. A SOFTWARE TOOL FOR DNA SEQUENCE AND STRUCTURE DESIGN: DNA SHOP

Two steps usually are involved in the DNA sequence and structure design. The first step is called sequence selection and the second step can be regarded as interactive DNA structure design. Sequence selection is the critical part of DNA structure design. Sequence needs to be proper defined so that the Watson-Crick complementarity and sticky ends can generate desired structure. Branched target molecules correspond to an excited state must be taken to ensure that the excited product obtained is the one that is sought. During the structure design, users need to have a flexibility to move DNA segments around interactively. Users may specify the length, start and ending bases of the sequence. Complementary sequences will be generated automatically following the Watson-Crick complementarity. Sticky ends are specified by users interactively.

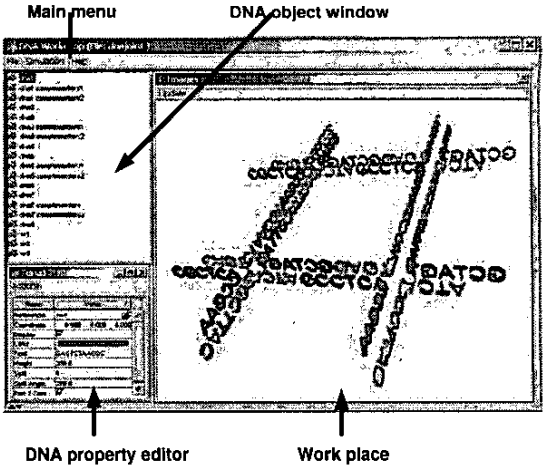

Fig. 1. User interface of the DNA shop

To facilitate DNA sequence and structure design, a software tool called DNA shop is developed. DNA shop is a java visualization tool for interactive DNA sequence and structure design. The software tool contains object management module, interactive user input module, graphics display module, coordinate transformation module, and DNA object module. The system is based on Java 3D class and Java J2SE SDK. Fig. 1 shows the user interface of the tool. It contains a "main menu", a "work place", a "DNA object window" and a "DNA property editor". Components and functions of each windows are as follows:

- The main menu contains "File", "Simulation" and "Help" three submenus. The "File" submenu is used to open and save a file, or exit the program. The "simulation" submenu is used to activate the work space.
- The work space shows the designed DNA structure. Users may select bases and change them. Users may move any part of a sequence to different locations by changing the coordinates.
- The DNA object window shows all DNA objects in the work space. For each DNA object, users can make copy, generate complementary or change the name.
- The DNA property editor shows each DNA's object-oriented properties including "Reference", "Coordinate", "Display", "Color", "Text", "Height", "Left split", "Split angle", and "Left complementary".
  - **Reference** represents origin of the DNA sequence. It could be any pre-defined world coordinate or another DNA object. In Fig. 1, W1 is the world coordinate.
  - **Coordinate** is the world coordinate of the DNA with respect to the "Reference" point.
  - **Display** property controls whether displays the ob-

ject or not.

- **Color** represents the color used to draw a DNA object.
- **Text** specifies DNA sequences using four DNA letter bases.
- **Height** defines height of DNA characters.
- **Left split** specifies the bending position of the DNA strand by counting number of DNA bases from left.
- **Split angle** defines the bending angle of the DNA strand.
- **Left complementary** controls the sequence for complimentary generation. If it is not checked, complimentary of the right split sequence will be generated. Otherwise, the complimentary for the left split sequence will be generated.

Right click on an object in the object window will select the object. Users can then select delete or copy the object. An object can be moved by changing coordinate in the property editor. Right click an object in work space will show the name of the object.

The DNA sequence is first shown in linear form. Users may define as many sequences as they want and change the form. Only single sequence needs to be specified. After right clicking a DNA object in object window, an action menu will pop up. Clicking on "Complimentary" menu item will generate a complimentary sequence of the object. Additional features of the DNA shop are:

- Modification of any single bases in one strand will automatically adjust corresponding binding pairs.
- Stick ends can be specified by users interactively.
- Moving the mouse while pressing left button down cause the DNA structure move around.
- Hold "Shift" key while pressing mouse left button rotate the DNA structure and show in different views.
- Hold "Ctrl" key while pressing mouse left button will zoom in and out of the DNA structure.

Users may change sequence bases at any time during the design process. The DNA shop will automatically synchronize corresponding changes.

The DNA shop is aimed to facilitate researchers to design DNA sequences and structures. In addition to provide manual interactive design, additional design automation mechanisms are also attractive to improve design efficiency. The purpose of design automation is to relieve the designers from tedious repeatable design work and allow them focus on high-level strategic planning. Two types of design automation have been implemented in the DNA shop. The first one is called automated sequence expansion, which means users can define a base DNA component to be expanded, and set the number of base component expanded. A new DNA sequence and structure can then be generated automatically based on the base component. The other one is called automated sequence connection. Through automated sequence expansion, a cluster of DNA blocks can be generated. Automated sequence connection will allow designers to connect two or more DNA blocks

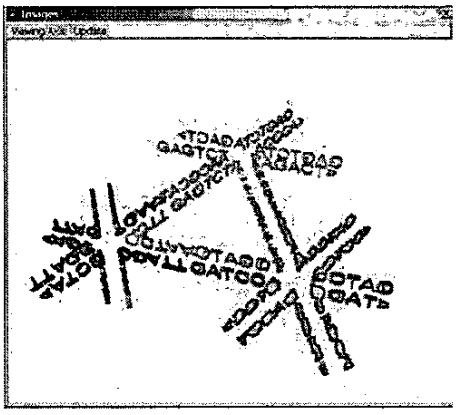

Fig. 2. DNA structure generated by automated expansion

automatically and generate more complicated structures. The user only needs to specify DNA blocks that need to be connected and the position of sticky end of a connecting block. The DNA blocks will then be connected automatically. Fig. 2 shows a DNA structure obtained by automatically connect three DNA blocks created by automated sequence expansion.

IV. Conclusions

This paper has proposed a software tool called DNA shop for DNA nano-applications. It has demonstrated the concept of interactive DNA nano-structure design using computers. The purpose is to improve DNA structure design efficiency and have better idea about the designed DNA sequences and structures before tedious and time-consuming laboratory experiments. The DNA shop may offer potential applications for interactive molecular machine design.

References

[1] L. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, vol. 266, no. 5187, pp. 1021-1025, November 1994.

[2] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keian, Z. Livneh and E. Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, vol. 414, no. 22, pp. 430-434, November 2001.

[3] K. E. Drexler. Nanosystems: molecular machinery, manufacturing and computation. *Wiley Interscience*, 1992.

[4] Q. Liu, L. Wang, A. G. Frutos, A. E. Condon, R. M. Corn and L. M. Smith. DNA computing on surfaces. *Nature*, vol. 403, no.13, pp. 175-179, January 2003.

[5] F. Liu, R. Sha and N. C. Seeman. Modifying the surface features of two-dimensional DNA crystals. *J. Am. Chem. Soc.*, vol. 121, pp. 917-922, August 1998.

[6] C. Mao, W. Sun, Z. Shen and N. C. Seeman. A nanomechanical device based on the B-Z transition of DNA. *Nature*, vol. 397, no. 14, pp. 144-146, January 1999.

[7] N. C. Seeman. DNA nanotechnology: novel DNA constructions. *Annu. Rev. Biophys. Biomol. Struc.*, vol. 27, pp. 225-248, 1998.

[8] N. C. Seeman, et al. New motifs in DNA nanotechnology. *Nanotechnology*, vol. 9, pp. 257-273, 1998.

[9] B. Yurke, A. J. Turberfield, A. P. Mills Jr, F. C. Simmel, and J. L. Neumann. A DNA-fuelled molecular machine made of DNA. *Nature*, vol. 406, no. 10, pp. 605-608, Agust 2000.

[10] Mingjun Zhang and T. J. Tarn. DNA Electrical Properties and Potential Nano-applications. *Proceedings of the 2003 IEEE Conference on Nanotechnology*, pp. 512-515, San Francisco, CA, Agust 2003.