Interactive DNA sequence and structure design for DNA nanoapplications

Mingjun Zhang
Chaman Sabharwal
Missouri University of Science and Technology, chaman@mst.edu
Weimin Tao
Tzyh-Jong Tarn
Ning Xi

See next page for additional authors

Follow this and additional works at: http://scholarsmine.mst.edu/faculty_work

Part of the Computer Sciences Commons

Recommended Citation
Zhang, Mingjun; Sabharwal, Chaman; Tao, Weimin; Tarn, Tzyh-Jong; Xi, Ning; and Li, Guangyong, "Interactive DNA sequence and structure design for DNA nanoapplications" (2004). Faculty Research & Creative Works. Paper 53.
http://scholarsmine.mst.edu/faculty_work/53

This Article is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. For more information, please contact weaverjr@mst.edu.
Interactive DNA Sequence and Structure Design for DNA Nanoapplications

Mingjun Zhang*, Member, IEEE, Chaman L. Sabharwal, Weimin Tao, Member, IEEE, Tzyh-Jong Tarn, Fellow, IEEE, Ning Xi, Member, IEEE, and Guangyong Li, Student Member, IEEE

Abstract—DNA sequence and structure design is very important for DNA nanoapplications. A computer-aided design tool is needed for exploring DNA sequence and structure of interests before experimental synthesis, which is a time- and labor-consuming process. In this paper, an interactive DNA sequence and structure design software tool called DNA shop is proposed and implemented. The visualization tool can generate DNA structures by specifying, selecting, and moving DNA sequences around and display corresponding structures. Using the tool, DNA sequence and structure can be visually inspected in three-dimensional space before experimental studies.

Index Terms—DNA computation, DNA machine, DNA nanotechnology, interactive design, visualization.

I. INTRODUCTION

Computers are expected to play important roles in enabling technologies for both theoretical and practical applications in nanotechnology. Some interesting work has been done for nanorobots assembly simulation in nanomedicine [2]. In this paper, we will show how computers help to interactively generate DNA sequence and display structure for DNA nanoapplications.

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 associations of DNA are highly specific and can be used to direct specific intermolecular associations of DNA complexes into more intricate arrangements. Seeman [12] reported that the combination of branched DNA molecules and sticky ends may create a powerful molecular assembly kit for structural DNA nanotechnology. Mao et al. [9] constructed a two-dimensional (2-D) DNA crystal using Holliday junction analogues that contain two helical domains twisted relative to each other. Polyhedra, complex topological objects, nanomechanical devices, and 2-D arrays with programmable surface features have already been produced in this way.

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 [3] 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 detectable output molecules that encode the computational results. In the future, there may be a need for fully organic computer devices implanted in a living body that can integrate signals from several sources and compute a response to an organic molecular-delivery device for drugs or signals. DNA molecule computing may pave the way in this direction [8].

For the above DNA applications, DNA sequence and structure design is a critical step. Currently, one must obtain proper conditions, refine design, and determine experimental windows for DNA structure through tedious and often expensive processes of trials and errors [14]. No software tool is available for interactive DNA structure design before experimental studies. 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. The idea is to structure of branched DNA motifs can provide basis for dynamic assembling switchable molecular machines. Yurke et al. [16] reported that the construction of a DNA machine in which the DNA is used not only as a structural material, but also as fuel. The proposed machine made from three strands of DNA has a pair of tweezers. The DNA machine may be closed and opened by addition of auxiliary strands of fuel DNA; each cycle produces a duplex DNA waste products. The Watson–Crick complementary association of sticky ends may be used to direct specific intermolecular associations of the DNA complexes into more intricate arrangements. Seeman [12] pointed out that the combination of branched DNA molecules and sticky ends may create a powerful molecular assembly kit for structural DNA nanotechnology. Mao et al. [9] constructed a two-dimensional (2-D) DNA crystal using Holliday junction analogues that contain two helical domains twisted relative to each other. Polyhedra, complex topological objects, nanomechanical devices, and 2-D arrays with programmable surface features have already been produced in this way.

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 [3] 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 detectable output molecules that encode the computational results. In the future, there may be a need for fully organic computer devices implanted in a living body that can integrate signals from several sources and compute a response to an organic molecular-delivery device for drugs or signals. DNA molecule computing may pave the way in this direction [8].

For the above DNA applications, DNA sequence and structure design is a critical step. Currently, one must obtain proper conditions, refine design, and determine experimental windows for DNA structure through tedious and often expensive processes of trials and errors [14]. No software tool is available for interactive DNA structure design before experimental studies. 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. The idea is to
let researchers interactively assemble DNA bases and generate DNA machines as if using symbols A, T, G, and C and visually inspect the three-dimensional (3-D) structures before experimental studies. This may help researchers get better ideas before experimental studies, and do better design of experiments.

This paper is organized as follows. Section II summarizes DNA properties that can be used for automatic structure construction. In Section III, an example of DNA sequence and structure design is discussed. An interactive DNA sequence and structure design tool called DNA shop is presented in Section IV. Conclusions are given in Section V.

II. DNA Properties

DNA’s unique biological property includes the specificity of the base pairing that holds two strands of double helix together. 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, and repair [12], [13].

The physical 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 for scaffolds of other molecules. Moreover, the phosphates in DNA’s backbone make it one of the most highly charged polymers [4].

DNA’s electrical properties make DNA one of the most interesting biomolecules for molecular electronics [17].

A DNA double helix is about 2 nm in diameter with helical repeat of about ten 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 $10^7$–$10^8$ terabytes 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 will have much longer, stable, and larger capacity for information storage.

A. DNA Sticky Ends

DNA can be cut at precise locations by restriction enzymes. Another enzyme—DNA ligase—can then be used to reassemble the pieces into a 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 binding site of EcoRI containing $\text{--GAATTC--}$ and $\text{--CTTAAG--}$. Using restriction enzymes to react at this site and break hydrogen bonds holding the overlapping single-stranded complementary strands, which are the strands between $G$ and $A$, the following overhanging chains can be obtained:

$\text{---G AATTC ---}$ \hspace{1cm} (1)

$\text{---CTTAAG G ---}$ \hspace{1cm} (2)

The segments AATT and TTAA in the above chains are called sticky ends, because the base pairs glue the two pieces together at proper conditions. The sticky end is an essential part of genetic engineering. It allows researchers to cut little pieces of DNA and place them at specific locations, where the sticky ends match.

DNA sticky ends provide a predictable, diverse, reliable, and programmable set of intermolecular interactions. Various DNA structure molecules can be created in this way for possible nanaplications. The use of stable-branched DNA molecules permits one to make stick figures. DNA double-crossover molecules are rigid DNA motifs for nanostructural construction, which contain two double helices linked at two different points. The sticky ends that hold the array together can vary to include diverse periodic arrangements of molecules in the crystal. DNA branched junctions have been constructed that contains three, four, five, and six arms. The combination of branched DNA and sticky-ended ligation results in the ability to form stick figures whose edges consist of double helical DNA, and whose vertices are the branch points of the junctions [12].

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 nanostructure 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 nanoscale 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.

Various DNA structures have been built using the DNA self-assembly property, including DNA knots [14], cubes, truncated octahedron, and Borromean rings. Two-dimensional crystalline forms of DNA that self-assemble from synthetic DNA double-crossover molecules have been observed by atomic force microscopy by Winfree et al. [15]. Intermolecular interactions between the structure units are programmed by the design of sticky ends according to Watson–Crick complementarity. The results demonstrate potential of using DNA to create self-assembling periodic nanostructures.

DNA self-assembly generally includes annealing single-strand DNA into tiles and assembling tiles into superstructures. Direct lattices assembly from single-strand DNA is possible and has been demonstrated for noncomputational DNA lattices [11].

III. Example of DNA Sequence and Structure Design

The above DNA properties are fundamental principles that need to be followed for designing DNA sequence and structure. Two steps are usually involved in the DNA sequence and structure design. The first step is called sequence selection. 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 properly 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 [12].

To illustrate the idea of DNA sequence and structure design, a simple four-arm branched DNA molecule is shown in Fig. 1, where “○” represents hydrogen bonding between the DNA bases. To design the four-arm branched DNA, only two single-stranded DNA need to be specified as shown in Fig. 2. The rest strands and sequences can be generated through Watson–Crick complementarity automatically. A 2-D lattice can then be constructed from the four-arm junction with sticky ends as shown in Fig. 3. Using the four-arm branched DNA molecule, 2-D lattice can be generated as shown in Fig. 4, where \( X \) and \( Y \) stand for sticky ends and \( X' \) and \( Y' \) represent their complements.

More complicated structures can be built in a similar way. A software tool can automate some of the above design work. For example, users may only need to specify the length, start, and ending bases of a predefined sequence. Complementary sequences can be generated automatically following the Watson–Crick complementarity principles. By introducing interactive visualization concepts, users can have the flexibility
to move DNA segments around during the design stage, and the sticky ends can be specified interactively. To facilitate these design ideas, a software tool called DNA shop has been proposed and implemented.

IV. DNA SHOP

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, DNA object module, and design engine. The system is primarily based on Java 3-D class and Java J2SE SDK. The design components include four DNA bases: A, T, G, and C. Fig. 5 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 three submenus: “File,” “Simulation,” and “Help.” The “File” submenu can be used to open and save a file or to exit the program. The “Simulation” submenu can be used to activate the workspace.
- The workspace shows the designed DNA structure. Users may select bases and change them or delete or insert bases. 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 workspace. In Fig. 5, 11 DNA objects are shown. Users can easily add new DNA objects to the list. 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 the origin of the DNA sequence. It could be any predefined world coordinate or another DNA object. In Fig. 5, \( W_1 \) is the world coordinate.
  - **Coordinate** is the world coordinate of the DNA with respect to the “Reference” point.

![Workflow of DNA sequence and structure design](image)

Fig. 5. DNA shop user interface.

---

**Fig. 6.** Workflow of DNA sequence and structure design.

---

**Display** controls whether the object displays or not.

**Color** represents the color used to draw the DNA object.

**Text** is used to specify DNA sequences by the four DNA letter bases.

**Height** defines height of DNA characters shown in the screen.

**Left split** specifies the bending position of the DNA strand by counting the number of DNA bases from the left.

**Split angle** defines the bending angle of the DNA strand.

**Left complementary** is used to control the sequence for complimentary generation. If it is not checked, the complimentary of the right split sequence will be generated. Otherwise, the complimentary for the left split sequence will be used.
Right-clicking on an object in the object window can select the object. Users may then select “delete” or “copy” the object. An object can be moved by changing coordinate in the property editor. Right-clicking an object in the workspace will show the name of the object. Users can modify the object through the object window.

A workflow of DNA sequence and structure design using the DNA shop is shown in Fig. 6. Users can start to design work from scratch or continue from previous work. First, DNA objects need to be created. Properties of each DNA object can then be specified. Users can view, rotate, and move 3-D structure of the DNA molecule interactively in the workspace. If users prefer expand certain base components or connect different components together, users may specify them and let the tool conduct the work automatically. Fig. 7 shows the results of DNA sequence and structure design for the example given in Section III. The top left figure shows a linear DNA sequence after specifying the DNA sequence. The top right figure shows the structure after bending an angle. The middle left figure shows complimentary of the two bending strands. The middle right shows complete complimentary of all strands. The bottom left figure shows the DNA structure after specifying sticky ends. Self-assembling of the strands shows an aggregated DNA structure in bottom right, which is a 2-D lattice.

The DNA sequence is first shown in a linear form. Users may define as many sequences as they want and change the form by specifying turning angles. Only a single sequence needs to be specified. After right-clicking a DNA object in the object window, an action menu will pop up. Clicking on the “Complimentary” menu item will generate a complimentary sequence of the object.

Additional features of the DNA shop are the following.
- Modification of any single base in a strand will automatically adjust corresponding binding pair.
- Sticky ends can be specified by users interactively. Users may specify multiple sticky ends.
- Moving the mouse while pressing the left button cause the DNA structure move around.
- Holding the “Shift” key while pressing the left mouse button will rotate the DNA structure and show in different views.
- Holding the “Ctrl” key while pressing the left mouse button will zoom in and out of the DNA structure.
- The 3' end of each strand is marked with > in the workspace.

Users may change sequence bases any time during the design process. The DNA shop will automatically synchronize corresponding changes. Complicated structures using base components of 2-D lattice or 3-D structure as shown in Fig. 8 can be generated using the DNA shop.

A. Design Engine and Automation

The DNA shop is aimed to help researchers to design DNA sequences and structures interactively and efficiently. The automatic design engine plays a key role for this goal. In the DNA shop, the design engine mainly consists of a knowledge-based system. The following types of rules have been implemented:
- Watson–Crick complementary binding properties;
- the sticky end association property.

In addition, component-based design automation mechanisms have been implemented to improve design efficiency. Two types of design automation have been implemented in the DNA shop.

- Automated sequence expansion. Users can define a base DNA component to be expanded and set the number of the base component expanded. A new DNA sequence and structure can then be generated automatically based on the
V. CONCLUSION

Physical observations for DNA structures are expensive. They should be applied only if necessary. For some cases, computer visualizations may provide appropriate solutions. This paper has proposed a software tool called DNA shop, which is aimed for interactive DNA sequence and structure design for drug delivery. Recently, the authors have been using the tool to design DNA sequence for research on DNA computation [18]. All of the above research requires significant amount of DNA sequence and structure design work before experimental studies. The DNA sequences and structures involved here are too complicated to be determined by brain imaging or pen-paper drawing. The proposed software tool can be used in these applications.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for their constructive comments and suggestions.

REFERENCES

Mingjun Zhang (S’98–M’01) received the Ph.D. degree in industrial automation from Zhejiang University, Hangzhou, China, in 1996 and the D.Sc. degree in systems science and mathematics from Washington University in St. Louis, St. Louis, MO, in 2000. Since 2001, he has been working on DNA genechip and protein microarray technology at the Life Sciences and Chemical Analysis Division of Agilent Technologies, Palo Alto, CA. His research interests include DNA nanotechnology, automation for the life sciences, system dynamics and control, robotics and mechatronics.

Dr. Zhang was awarded the Early Industrial Career Award by the IEEE Robotics and Automation Society in 2003, and was the 2004 Boeing Distinguished Speaker for Bioinformatics at the University of Missouri-Rolla. Currently, he serves as an Associate Editor for the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING.

Chaman L. Sabharwal received the Ph.D. degree from the University of Illinois, Urbana-Champaign, in 1967. He was a Professor at St. Louis University and a Senior Specialist at McDonnell Douglas Corporation. He has been a consultant for McDonnell Douglas Research Laboratories and VanHouten Technologies, Inc. He is currently a Professor of Computer Science at the University of Missouri-Rolla, Rolla. He has taught various courses on mathematics and computer science, including algorithmic, artificial intelligence, computer vision, computer graphics, Java graphical user interfaces, Java Beans, Java JINI, machine learning, neural networks, parallel processing, and robotics. He has published research papers on software engineering, geometric modeling, computer graphics, computer vision, machine learning, parallel processing, and pattern recognition. His research interests include artificial intelligence, computer graphics, computer vision, geometric modeling, image databases, parallel processing, pattern recognition, robotics, machine learning, and neural networks.

Tzyh-Jong Tarn (M’71–M’83–F’85) received the D.Sc. degree in control system engineering from Washington University in St. Louis, St. Louis, MO. Currently, he is a Professor in the Department of Electrical and Systems Engineering and the Director of the Center for Robotics and Automation at Washington University.

Prof. Tarn received the NASA Certificate of Recognition for the creative development of a technical innovation on “Robot Arm Dynamic Control by Computer” in 1987. The Japan Foundation for the Promotion of Advanced Automation Technology presented him with the Best Research Article Award in March 1994. He also received the Best Paper Award at the 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. He is the first recipient of both the Nakamura Prize and the Ford Motor Company best paper award at the Japan/USA Symposium on Flexible Automation in 1998. In addition, he is the recipient of the prestigious Joseph F. Engelberger Award of the Robotic Industries Association in 1999, the Auto Soft Lifetime Achievement Award, the 2004 Boeing Distinguished Speaker Award in Biomedical Engineering, the Best Research Article Award in March 1994. He also received the Best Paper Award in the 1998 Japan-USA Symposium on Flexible Automation. He was featured in the Special Report on Engineering of the 1998 Best Graduate School issue of U.S. News and World Report and his recent research accomplishments were reported in the Washington Times, the Financial Times, Le Monde, and the Chicago Sun-Times, among others. He served as the President of the IEEE Robotics and Automation Society from 1992 to 1993, the Director of the IEEE Division X from 1995 to 1996, and was a Member of the IEEE Board of Directors from 1995 to 1996. At present, he serves as the Vice President for Conferences of the IEEE Robotics and Automation Society.

Ning Xi (M’94) received the B.S. degree in electrical engineering from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1982, the M.S. degree in computer science from Northeastern University, Boston, MA, in 1988, and the D.Sc. degree in systems science and mathematics from Washington University in St. Louis, St. Louis, MO, in 1993. Currently, he is a Professor in the Department of Electrical and Computer Engineering at Michigan State University, East Lansing. His research interests include mobile robot control, computer vision, manufacturing automation, micro/nano systems, and intelligent control and systems.

Dr. Xi received the Best Paper Award at the IEEE/RSJ International Conference on Intelligent Robots and Systems in 1995. He also received the Best Paper Award in the 1998 Japan–USA Symposium on Flexible Automation. He was awarded the first Early Academic Career Award by the IEEE Robotics and Automation Society in 1999. In addition, he is a recipient of National Science Foundation CAREER Award.

Guangyong Li (S’01) received the B.S. degree in mechanical engineering from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 1992 and the M.S. degree in control and automation from Beijing Institute of Control Engineering, China Academy of Space Technology, in 1999. He is currently working toward the Ph.D. degree in the Department of Electrical and Computer Engineering, Michigan State University, East Lansing.

From 2000 to 2001, he was a Research Scholar in the Department of Electrical and Computer Engineering, National University of Singapore. Singapore, working on mobile robot control and neural network control. His research interests include mobile robot control, neural networks control, micro/nano robotics and systems, nanoelectronics, and nanobiotechnology.