Published in final edited form as:

Biophys Rev Lett. 2007 April ; 2(2): 123–137. doi:10.1142/S1793048007000490.

SELF-RECOGNITION OF DNA FROM LIFE PROCESSES TO DNA COMPUTATION

WEI DA NG and

National Institute of Education, Nanyang Technological University, 1 Nanyang Walk, Singapore 637616

CHEE KEONG BENJAMIN WONG

World Scientific Publishing Co Pte Ltd, 5 Toh Tuck Link Singapore 596224

WEI DA NG: nwd04@pacific.net.sg; CHEE KEONG BENJAMIN WONG: ckwong@wspc.com.sg

Abstract

Ever since the first appearance of deoxyribose nucleic acid (DNA) in 1953, it has fascinated multitudes with its simplicity. With a modest syllabus of four nucleotides (adenine, thymine, cytosine and guanine), it codes for the complexity of life around us. In this paper, we investigate how the structure of DNA codes for life processes and how we can take advantage of its minuscule size, mechanism of self-recognition and self-assembly for “bottom-up” nanotechnology. High hopes are also placed on miniaturizing present computing technology using DNA computing based on two fundamental features; massive parallelism of DNA strands and Watson–Crick complementarity. Advances in DNA-based computation and algorithmic assembly are then used to complement researches in DNA nanotechnology.

Keywords

DNA nanotechnology; DNA computation; self-assembly

1. Introduction

Like most scientific breakthroughs, discovering the structure of DNA and how it functions required the sustained efforts of many scientists over a period of time. In the early days, little was known of its structure apart from the knowledge of its molecular composition; that it consists of subunits called nucleotides which in turn are comprised of three parts: a phosphate group, a sugar called deoxyribose and one of four possible nitrogen-containing bases — adenine (A), thymine (T), guanine (G) or cytosine (C). In the 1940s, Erwin Chargaff analyzed the relative amounts of the nucleotides in DNA from various species and found that there were always equal amounts of cytosine and guanine, as well as equal amounts of adenine and thymine. Although he offered no explanation for his observation, this discovery proved to be a milestone in the final prediction of the DNA’s structure.

2. Structure of DNA and Its Biological Functions

In 1953, Watson and Crick created history when they became the first to decipher the structure of DNA. They suggested that within each DNA strand, the phosphate group of one nucleotide joins to the sugar group of the adjacent nucleotide, forming a backbone that is linked by strong covalent phosphodiester bonds with the nucleotide bases protruding from this sugar–phosphate backbone. Their most significant proposal was that two nucleotide strands actually wind about each other to form a double-helical structure. They are held

together by hydrogen bonds which form between the protruding bases of the two strands, and these bonds confer a ladder-like structure to the DNA molecule with the sugar–phosphate backbone on the outside and the nitrogenous bases on the inside. In addition, the two DNA strands in the double helix are oriented in opposite directions, one running in the 5′-3′ direction and the other in the 3′-5′ direction.

At each rung of the double helix ladder, adenine forms hydrogen bonds (two of them) only with thymine while guanine forms hydrogen bonds (three) only with cytosine. These A–T and C–G pairs are called complementary base pairs, or Watson–Crick base pairs, and their presence fully justify Chargaff’s results. Besides providing the chemical foundation for genetics, this specific bonding of DNA base pairs also presents a powerful molecular recognition system that can be tapped in various fields, examples of which will be discussed in later sections.

The standard structure of DNA is that of B-DNA which is predominant in humid conditions and standard aqueous solutions. However, at low humidity, or in the presence of alcohols and salts, DNA can also adopt alternative structures such as the A-DNA and Z-DNA (Fig. 3). The former is a shorter, broader right-handed variation of B-DNA which results from the high inclination and displacement of the base pairs relative to the helical axis, while the latter is a narrower and longer left-handed variant of the DNA duplex with a zigzag backbone.

2.1. Biological functions

The role of DNA as a molecule of heredity is augmented by its double helical structure which confers stability to the molecule. Within a gene, each sequence of bases defines a protein which will be expressed or manufactured in an organism’s lifetime via transcription and translation. Since the four types of bases can be permuted in any linear order within each DNA strand, a vast amount of information can thus be encoded. The high specificity exhibited by the base pairings also ensures that the integrity of the information is maintained during replication. This is critical because as a molecule that carries hereditary information, the same set of genetic code must be passed on from the parent to the daughter cells.

3. Principles of DNA Self-Recognition and Assembly

Apart from its biological role as genetic material, DNA can also be used to construct new materials on the nanometer ($1\text{ nm} = 10^{-9}\text{ m}$) scale. This is because B-DNA, which is the standard right-handed form of DNA, has a double helix with a width of 2 nm, a helical repeat of ten base pairs per turn, and a rise of 0.34 nm per base pair.

There are two basic approaches to nanotechnological constructions, namely the “top-down” approach, whereby microscopic manipulations of small numbers of atoms or molecules fashion elegant patterns, and the “bottom-up” approach, whereby several molecules self-assemble in parallel steps, as a function of their molecular recognition properties. Self-assembly is defined as the spontaneous self-ordering of substructures into superstructures. Since artificially synthesized single-stranded DNA has the capability of self-assembling into DNA crossover molecules, DNA has an important role to play in “bottom-up” nanotechnological constructions.

The key to the use of DNA in material applications is the specificity of the Watson–Crick base pairing. The double helical arrangement of the two antiparallel polynucleotide chains in natural DNA leads to a linear, unbranched helix axis. However, this unbranched helix does not allow synthesis in multiple dimensions, which is required in order to produce interesting nanomaterials from DNA.

Branched DNA occurs naturally in living systems. During meiosis, a process whereby haploid nuclei with unpaired chromosomes are produced from diploid parent nuclei with paired chromosomes, homologous chromosomes line up side by side and exchange segments of DNA. Next, as recombination occurs, the aligned strands of DNA break and cross over one another, forming Holliday junctions. An example of such a junction is shown in Fig. 4.

The Holliday junction consists of four DNA strands 1, 2, 3 and 4 bound together to form four double helical arms I, II, III and IV flanking a branch point. In a natural Holliday structure, the branch point can relocate throughout the molecule by virtue of the homologous sequences. However, in a synthetic Holliday structure, the branch point can be fixed in position and designed to contain varied numbers of arms. The use of branched intermediates such as Holliday structures enables the synthesis of N -connected objects from DNA, as well as periodic and aperiodic arrays. The design of branched nucleic acid motifs is based on the idea of maximizing base pairing, regardless of the presence of branch points.

In addition to having strands that are completely paired with one another, it is also possible to have one strand a little longer than its complement, leading to an overhang, which is known as a sticky end. Sticky ends are typically four to eight bases long, and two molecules with complementary sticky ends will cohere to form a molecular complex. The use of sticky ends enables predictable control of intermolecular associations and predictable geometry at the point of cohesion. The formation of a two-dimensional lattice from a four-arm junction with sticky ends is shown in Fig. 5.

A is a sticky end and A' is its complement. The same relation exists between B and B'. Four of the monomeric junctions on the left side of Fig. 5 are complexed in parallel orientation to produce the structure on the right. In addition, the complex has maintained open valences so that it can be extended by the addition of more monomers. One important advantage of sticky-ended cohesion over the similar affinity properties of antibodies and antigens is that sticky ends form B-DNA when they bind to each other, resulting in a known local geometry of the cohesive system.

3.1. DNA as a scaffold

One possible application of spatially periodic DNA arrays is the use of DNA crystals as scaffolding to organize other molecules. For example, if stick-figure crystalline DNA cages could be constructed on the nanometer scale, they could be used to orientate other biological macromolecules as guests inside the cages, thus enabling the three-dimensional structure of the biological macromolecules to be studied using X-ray crystallography. This concept is illustrated in Fig. 6.

However, the potential of this application awaits the successful self-assembly of three-dimensional DNA lattices. DNA scaffolds can also be used to position and orientate components of molecular electronic devices with nanometer-scale precision. An example of this type of application is shown in Fig. 7.

In this technique, nanowires are attached to two branched DNA junctions. When the sticky ends of the two branched DNA junctions cohere, the nanowires are organized as well, with a positively charged ion between them. The positively charged ion will stabilize the resulting coordination complex.

3.2. Advantages of using DNA as a chemically based assembly system

The existence of several commercially available modifying enzymes has contributed significantly to the growth of DNA nanotechnology. For example, ligases are used to

establish covalent bonds between the cohesive ends of branched DNA molecules. Restriction endonucleases play a role in the manufacture of sticky ends, cutting DNA at specific 4–8 base long sequences. Exonucleases are used to digest failure products that do not contain cyclic strands. Topoisomerases are enzymes which are capable of breaking the phosphodiester bonds between nucleotides of one or both strands of the DNA double helix, allowing the helical strain to be relieved. Therefore, topoisomerases are used to change and analyze the topological linking of DNA strands.

After hydrogen bonds have formed between complementary base pairs, the DNA sequence can still be read from the outside. This enables points on a DNA construct to be identified uniquely. Furthermore, DNA is a stiff molecule, with a persistence length of about 50 nm under standard conditions. Over this distance, the central axis of the DNA molecule bends very little. Certain DNA motifs, such as DNA double crossover molecules, are even less flexible. In addition, DNA has a high functional group density. This means that large DNA crossover molecules can be derivatized down to separations of 0.34 nm, which is the rise of a single base pair.

4. DNA Nanotechnologies and 3D Molecular Circuits

4.1. DNA nanotechnological constructions

The first artificial DNA structure was a three-connected DNA molecule whose helix axes were connected like the edges of a stick cube. This is shown in Fig 8.

The stick cube consists of six cyclic interlocked single strands of DNA, one for each face of the polyhedron. Each edge of the cube consists of two turns of the DNA double helix. The DNA strand which corresponds to the front face of the cube is linked twice to the strand on the right, the strand on the top, the strand on the left, and the strand on the bottom. The DNA strand corresponding to the front face is linked only indirectly to the DNA strand at the rear.

Since individual branched junctions are floppy, with flexible angles between the four rigid arms, the overall conformation of the branched junctions can be analyzed or designed, but the geometry of the branched junctions is fixed. In this case, the topology of the branched junctions was determined by electrophoretic analysis of cubes cut by specific restriction endonucleases.

The production of two-dimensional periodic arrays with defined features has been a major success of structural DNA nanotechnology.

Figure 9 shows a meiotic double crossover recombination intermediate (MDX), which consists of a pair of homologous chromosomes. Each homologous chromosome consists of two DNA strands which align and cross over in order to swap equivalent portions of genetic information. Each meiotic double crossover recombination intermediate has two Holliday junctions (HJ).

Figure 10 shows the structure of an analogue double crossover unit (ADX), which is used as a motif in the construction of two-dimensional DNA arrays. Each ADX unit consists of two DNA strands, two DNA crossover strands, and a central crossover strand.

Figure 11 shows the strand structure and base pairing of an analogue ADX molecule labeled A on the left, and a variant labeled B* on the right. The B* ADX molecule contains an extra DNA domain which extends from the central crossover strand. The extra DNA domain protrudes roughly perpendicular to the plane of the rest of the DX molecule, and the domain facilitates the visualization of the two-dimensional crystalline array of DX molecules under an atomic force microscope (AFM).

Figure 12 is a schematic representation of the A and B* double crossover molecules, where the perpendicular domain of B* is represented as a circle. The sticky-ended complementarity is represented geometrically. Since the dimensions of both A and B* are about 4×16 nm in this projection, the B* protrusions, which are visible as white stripes in the AFM image in Fig. 13, lie about 32 nm apart.

4.2. DNA-based nanomechanical devices

Structural DNA nanotechnology has also led to the development of nanomechanical devices. The first DNA nanomechanical device which is capable of controlled movement consisted of two DNA double crossover (DX) molecules connected by a DNA double helix. The connecting DNA double helix contains a special sequence of DNA that can be converted from the standard right-handed B-DNA to the narrower and longer left-handed Z-DNA.

In standard B-DNA promoting conditions, the two DX molecules are on the same side of the connecting helix, as shown in Fig. 14. When the complex ion hexaamminecobalt (III), $[\text{Co}(\text{NH}_3)_6]^{3+}$ is added to the solution, giving rise to Z-DNA promoting conditions, an apparent rotary motion of about half a revolution occurs, causing the two DX molecules to lie on opposite sides of the connecting helix. This motion was demonstrated by fluorescent resonance energy transfer, and atomic displacements ranging from 2 nm to 6 nm were observed, depending on the location of the atom relative to the axis of the connecting helix. The problem with this device is that it is activated by a complex ion. This results in a lack of programmability, whereby an ordered collection of DX molecules all produce the same response when subjected to the same stimulus. In order to produce effective nanorobotics, devices need to be addressable individually.

A solution to this problem was formulated by the development of a sequence-control device with tweezers-like motion. The machine cycle of this device is shown in Fig. 15.

This rotary DNA mechanical device is based on a motif known as the paranemic crossover (PX) system, which leads to sequence-specific nanomechanical motion. In the PX state, two DNA strands each have a top double helix and a bottom double helix. The top double helix is connected to the bottom double helix by a single DNA strand which crosses over at a number of places. This results in two rigid arms with a flexible hinge in between them. There exists another related motif, which is known as the JX_2 state. The ends of this JX_2 molecule, which is shown on the right in Fig. 15, and the ends of the PX molecule, which is shown on the left in Fig. 15, are the same at the top (A and B), but rotated 180° at the bottom (C and D). The two discrete states PX and JX_2 are set by the presence of set strands, which bind to the frame in different ways to produce different conformations. Set strands are DNA strands with small extensions that enable their removal when complementary strands are added.

In Step I of the machine cycle described above, complete complementary DNA strands of the set strands are added to the solution containing the PX molecule. These complementary strands have biotin groups (represented as black dots) attached to their ends. The biotin groups can be bound to magnetic streptavidin beads so that the complementary strand and the attached set strand can be removed from the solution. After the set strands have been removed from the PX molecule, an intermediate is formed. In Step II, set strands are added to the intermediate so as to form the JX_2 molecule, rotating the bottom of the molecule. Subsequently, Steps III and IV restore the original PX conformation. The operation of this device has been demonstrated by electrophoresis and atomic force microscopy. A variety of sequence-specific devices can be developed by changing the sequences of the strands, and the sections of the device to which they bind.

4.3. Application of DNA assemblies to the layout of molecular-scale circuit components

Molecular-scale circuits have the potential to replace traditional micro-electronics with densities up to millions of times current circuit densities. Recent efforts to design molecular circuit components have led to the use of organic molecules as conducting wires, switches, dynamic random access memory (DRAM) cells and rectifying diodes with negative differential resistance. These molecular electronic components change their conformation when they perform electrical switching. One major challenge in molecular electronics is the development of molecular electronic components that can restore a signal to binary values. A possible solution to this problem would be the use of multi-component assemblies that exhibit cooperative thresholding.

The molecular circuit assembly problem refers to the problem of assembling molecular electronic components into a molecular scale circuit. At the molecular scale, top-down techniques of laying out circuits, which involve the extraction of devices from a single large-scale building block, may not be practical. Therefore, a bottom-up approach has to be adopted, whereby the whole system is composed of small basic building blocks with recognition, structuring, and self-assembly properties. DNA tiling assemblies have the capability to contribute to such bottom-up approaches. By using annealing, the selective attachment of molecular electronic components to particular tiles of the DNA tiling array can be achieved.

DNA lattices can be used as a foundation upon which nanoscale wires are grown, as shown in Fig. 16. In this process, gold is deposited from colloid onto nano-spheres which are immobilized on DNA tiling lattices. Molecular probe devices are then used to test the electrical properties of the resulting molecular circuit attached to the DNA tiling array.

5. DNA Computing

DNA computing together with quantum computing is the next step to miniaturizing current computer technology. There are high hopes in DNA computing based on the two fundamental features that DNA offers, *massive parallelism* and *Watson–Crick complementarity*.

Most intractable problems in computing are not solvable by efficient algorithms, except for an exhaustive search that amounts to exponential computing time. With current technology, this problem can only be alleviated slightly using parallel computing. However, using the high density of information stored in DNA, and the ability to make large amounts of exactly similar copies through Polymerase Chain Reaction (PCR), DNA computing offers a large parallelism that can be used to run through an exhaustive search simultaneously.

Another feature of DNA that makes it viable for computing is the ability of the bases to participate in complementary pairing, A with T, and C with G. There are many operations that could be carried out using DNA molecules. With the help of DNA exonucleases and endonucleases, we can cut DNA at specific sites and create a DNA strand of choice. Pasting of DNA can be achieved with DNA ligase. Using appropriate modifying enzymes, we can also modify the nucleotides according to choice. Finally, the desired nucleotide sequence can be multiplied using PCR.

5.1. NP-complete problems

There exists a series of problems that are termed NP-complete in complexity theory. A typical NP-complete problem is unlikely to possess an efficient algorithm (i.e. working in polynomial time). In theory, if one can be solved by an algorithm, the algorithm can be

extended to all other NP-complete problems. However, NP problems are very difficult to solve and will require exhaustive searches that amount to exponential resources used.

One NP-complete problem is the directed Hamiltonian Path Problem (HPP). The problem is to find a directed path from the input vertex to the output vertex by traversing through each vertex once.

It has been found that there is no efficient way to determine whether there is a solution to this problem except to run through all possible paths. In 1994, Leonard Aldeman solved the HPP by using the massive parallelism of DNA. He used strands of 20 nucleotides to represent each vertex, s_0 to s_6 , in the above example.

Using the function $h()$ to map the Watson–Crick complement, we can code for the directed edge of the graph using the following:

$$\begin{aligned}\text{Vertex: } s_i &= s_i' s_i'' \\ S_j &= s_j' s_j'' \\ \text{Edge: } e_{i \rightarrow j} &= h(s_i'' s_j')\end{aligned}$$

To determine whether the above graph is Hamiltonian, large amounts of oligonucleotides that code for all the vertices and paths are used in the presence of enzymes to yield all possible combinations. Aldeman used the following algorithm to solve the HPP.

Input: A directed graph G with n vertices, among which are designated vertices v_{in} and v_{out} .

- Step 1: Generate paths in G randomly in large quantity.
- Step 2: Reject all paths that do not begin with v_{in} and end in v_{out} .
- Step 3: Reject all paths that do not involve exactly n vertices.
- Step 4: For each of the n vertices v , reject all paths that do not involve v .

Output: “Yes” if any path remains, “No” otherwise.

Using DNA computing, the satisfiability problem and the Data Encryption Standard (DES) code can also be worked out. The satisfiability problem is to check if a propositional formula is satisfiable. An example of such a formula is given by

$$\alpha = (x_1 \vee \sim x_2 \vee x_3) \wedge (x_2 \vee x_3) \wedge (\sim x_1 \vee x_3) \wedge \sim x_3. \quad (1)$$

This problem can be solved by a directed graph with two edges from each vertex representing 0 or 1. For the above proposition α , it can be shown that it cannot be satisfied.

5.2. Sticker model

In this model, a single stranded nucleotide can be used as a memory strand that is distinctly divided into k non-overlapping regions of m bases each. Using complementary strands, termed stickers, to such k regions, we can form a memory complex to represent Boolean variables. Hence, the memory complex will be partially double stranded with the double stranded region representing bit 1 and single stranded region representing bit 0.

The potential application of the sticker model is to attack the DES. It is calculated that only approximately 1 gram of DNA is needed. Even though the attack is likely to produce errors, there is a high probability of success.

DNA computing is still in its infant stage. Many improvements are still needed to make it into a viable avenue of computation. This includes the advancement in biophysical techniques of DNA manipulation and techniques in the automation of processes. Although DNA computing shows great promise, it is more likely to be used to complement current silicon based computing rather than to replace it.

6. Conclusion

The fledgling fields of DNA nanotechnology and DNA computing have attracted immense interest from researchers in recent times, largely due to the vast potential benefits that can be tapped from these areas. For the past 50 years, the study of DNA has been limited to the realm of biologists and chemists. However, in the next 50 years, there is a high likelihood that this field of research will shift to the exploitation of DNA properties in a non-biological context by material scientists and computer engineers.

Acknowledgments

The authors would like to thank Lim Yee Woon for his contribution to the section on DNA computing. Lim Yee Woon is now at the Division of Physics and Applied Physics, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore. He can be contacted at the yeewoon@ntu.edu.sg

References

1. Watson JD, Crick FHC. Nature 1953;4356:737. [PubMed: 13054692]
2. Seeman NC. Nature 2003;421:427. [PubMed: 12540916]
3. Garrett, RH.; Grisham, CM. Biochemistry. 2. Saunders College Publications; 1999.
4. Audesirk, T.; Audesirk, G.; Byers, BE. Biology: Life on Earth. 6. Prentice Hall; 2001.
5. van Holde, KE.; Johnson, WC.; Shing Ho, P. Principles of Physical Biochemistry. Prentice Hall; 1998.
6. Seeman NC. Materials Today Jan 2003;2003:24.
7. Reif JH, LaBean Thomas H, Nadrian C. Seeman, Challenges and Applications for Self-Assembled DNA Nanostructures. Lecture Notes in Computer Science Jan;2001 2054:173.
8. Porath D, Cuniberti Gianaurelio, Di Felice Rosa. Top Curr Chem 2004;237:183.
9. Claude, CS.; Paun, G. Computing with Cells and Atoms. Taylor and Francis; 2001.
10. Paun, G.; Rozenberg, G.; Salomaa, A. DNA Computing. Springer; 1998.
11. Reif JH. IEEE Computing in Science & Engineering 2002;4(1):32.

Fig. 1.
Composition of a nucleotide.

Fig. 2.
Watson-Crick base pairs.

Fig. 3.
Structures of B-DNA and the two other variants A-DNA and Z-DNA.

Fig. 4.
A four-arm branched DNA junction formed from four DNA strands.

Fig. 5.
Formation of a two-dimensional lattice from a four-arm junction with sticky ends.

Fig. 6.
Macromolecular guests arranged in a parallel manner in a DNA host lattice.

Fig. 7.
Organization of nanoelectronic components by DNA.

Fig. 8.
A DNA molecule with the connectivity of a cube. The six backbone strands are represented by colored balls, while the bases are drawn in white.

Fig. 9. Schematic drawing of a meiotic double crossover recombination intermediate (MDX).

Fig. 10.
Schematic drawing of an analogue double crossover unit (ADX).

Fig. 11. Strand structure and base pairing of the analogue ADX molecule, labeled A, and a variant, labeled B*.

Fig. 12. Schematic drawings of A and B* where the perpendicular domain of B* is represented as a blue circle.

Fig. 13.
Atomic force microscope image of the B* protrusions, which are visible as stripes in two-dimensional DNA arrays.

Fig. 14.
A nanomechanical device based on the B-Z transition of DNA.

Fig. 15.
Machine cycle of a rotary DNA mechanical device.

Fig. 16. Binding gold nano-spheres to a DNA lattice and fusing these into a molecular wire by metal deposition.

Fig. 17.
A Hamiltonian Path Problem with seven vertices.