

Java Programming AP Edition

U2C6 Methods

DEMO PROGRAM: DNA ENCODING (GENE.JAVA)

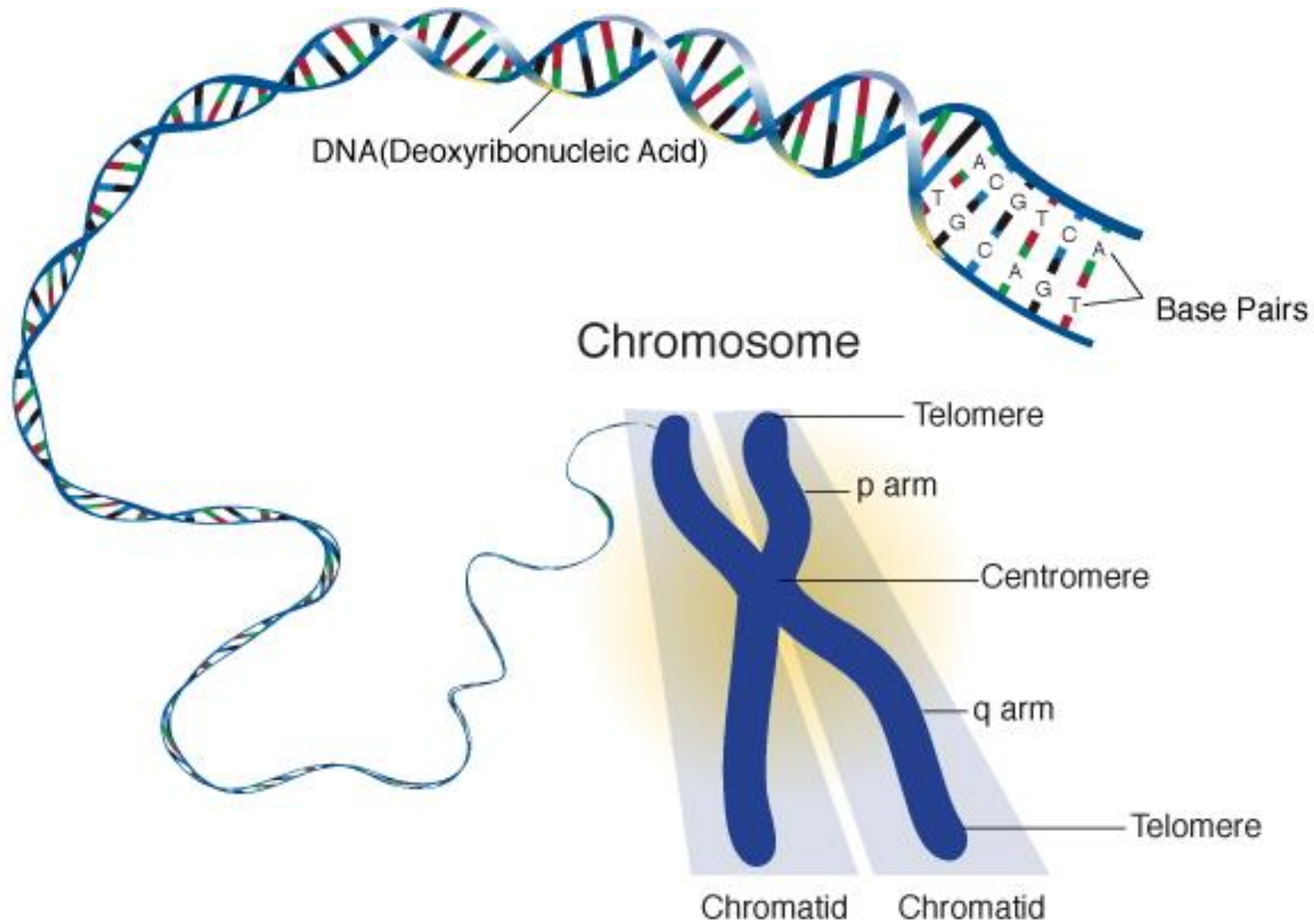
ERIC Y. CHOU, PH.D.

IEEE SENIOR MEMBER

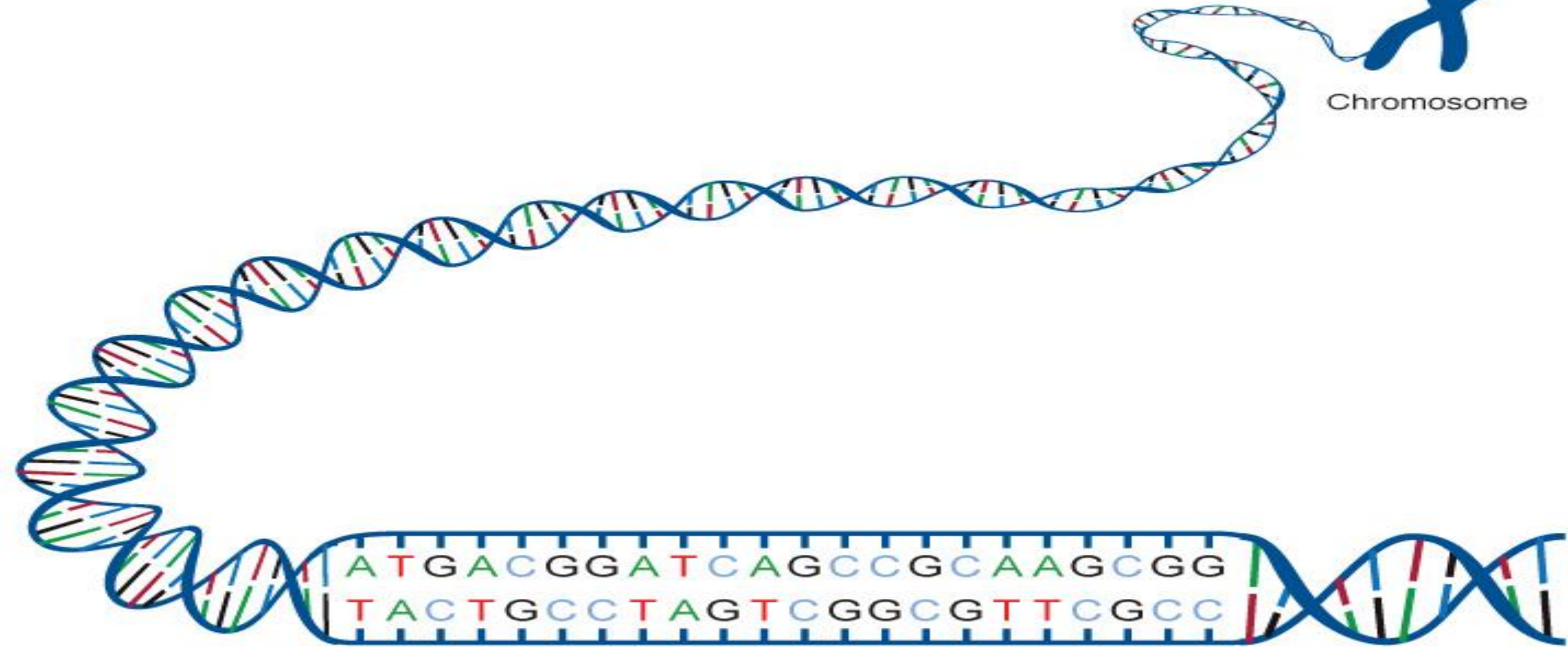


Purpose of this Demo Program:

- (1) Get exposure to a good application field for **Computer Science**.
- (2) Demonstrate **string** manipulation methods (beyond Java API) for some application field.
- (3) know more about methods and calling methods.
- (4) dual I/O using a **log** method.



Genetic Code



Replication



DNA replication is the process by which a molecule of DNA is duplicated. When a cell divides, it must first duplicate its genome so that each daughter cell winds up with a complete set of chromosomes.



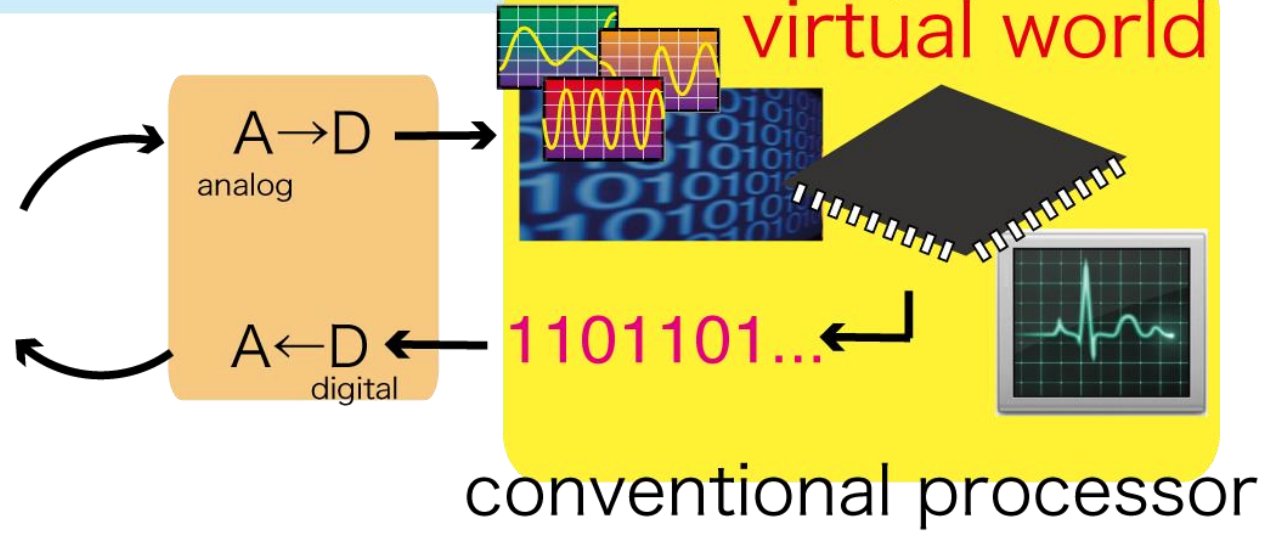
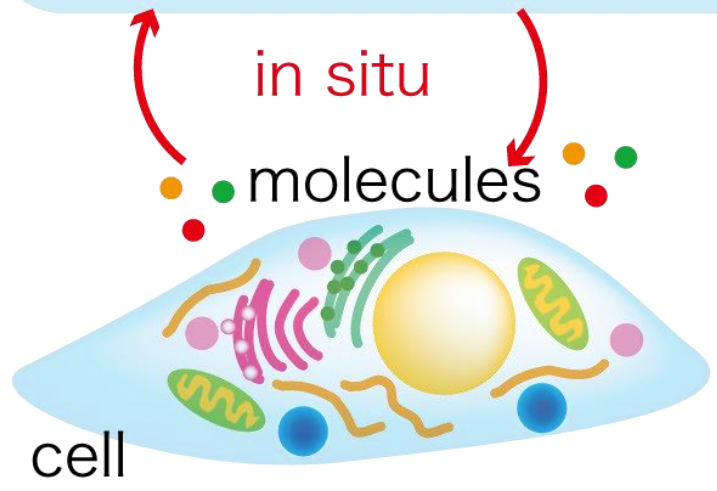
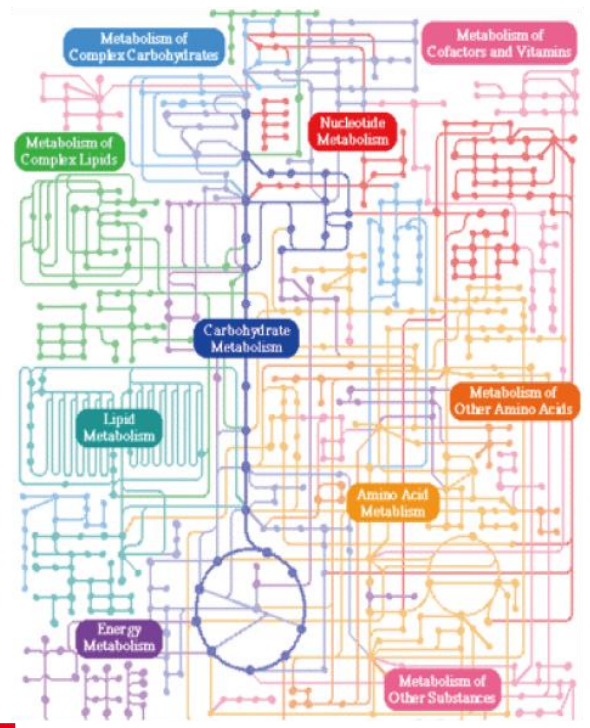
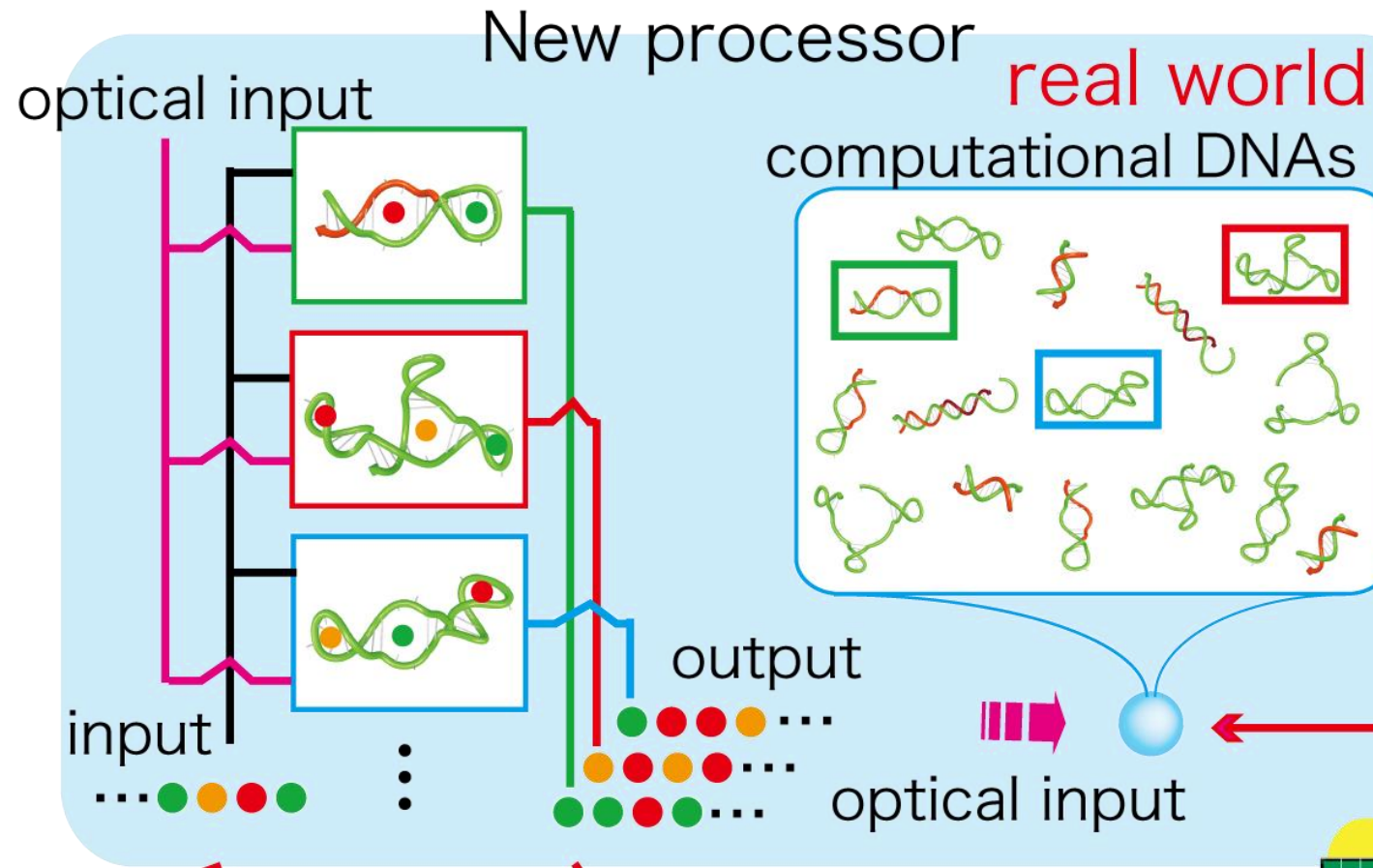
Purpose of Studying DNA

(1) DNA modeling and computer simulation for DNA repair, DNA replication (birth control), DNA therapy, DNA analysis, DNA matching and DNA engineering.

(2) DNA reading (understand all its purpose)

(3) DNA computing. (Photonic Computing, Neural computing, and ...)

(Legal issues?)





Advantage of DNA Computing

Massively parallel computing.

Low Cost. (Cost of protein)

Fast computing speed. (due to randomness)

(1 sec DNA computing can be equivalent to 1 year simulation time)

Disadvantage:

Hard to guarantee results so far.

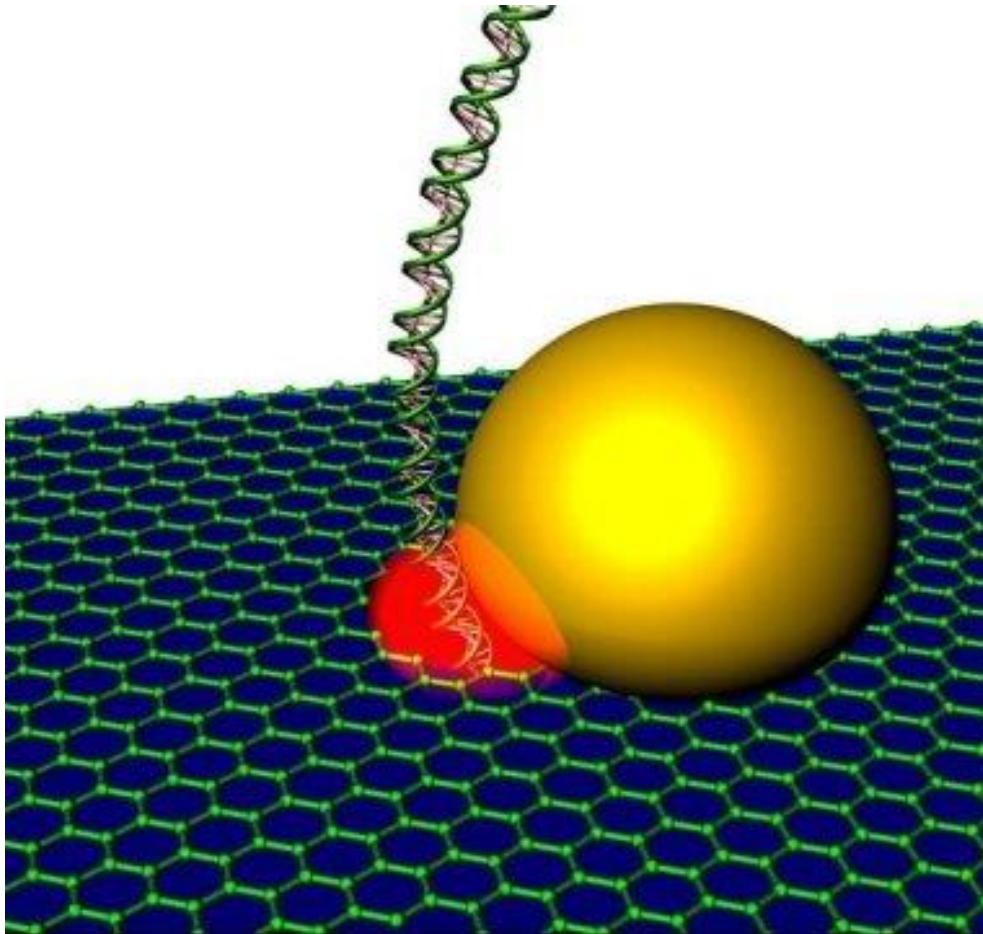
Not practical yet.

So far, still mostly used for DNA Study.



DNA Reader

Source: DOE/Lawrence Berkeley National Laboratory



High-speed reading of the genetic code should get a boost with the creation of the world's first graphene nanopores -- pores measuring approximately 2 nanometers in diameter -- that feature a "built-in" optical antenna. Researchers have invented a simple, one-step process for producing these nanopores in a graphene membrane using the photothermal properties of gold nanorods.



DNA representation

The complementarity between base pairs (**A = T** and **G = C**) implies that if you know one sequence you can deduce the complementary sequence.

It is common to represent DNA sequences by 4-letter strings:

TGCTAATGCCGCTACTCTATCTGC

By convention, we write sequences **from 5' to 3'** end.

5' – TGCTAATGCCGCTACTCTATCTGC – 3'

DNA representation



Don't forget the second strand!

When we analyze a DNA sequence represented by

ATGCGCGGATG

we should keep in mind that the corresponding molecule is a double strand helix with the following base pairs:

5' - ATGCGCGGATG - 3' (upper strand)

|||||

3' - TACGCGCCTAC - 5' (lower strand)

Note that the concept of *upper strand* and *lower strand* are purely artificial. A DNA molecule is a 3D structure and there is no reason to consider preferentially one or the other strand.

This does not mean however that the two strands are functionally equivalent: in coding regions for example, only one strand will serve as a template for the synthesis of RNA.

The **5' and 3'** mean "five prime" and "three prime", which indicate the carbon numbers in the **DNA's** sugar backbone. The **5'** carbon has a phosphate group attached to it and the **3'** carbon a hydroxyl group. This asymmetry gives a **DNA** strand a "direction".

5' and 3' means directions



Reverse complementarity

Reverse complementary sequences represent the two strands of the same DNA molecule.

The reverse complement is obtained by transposing each nucleotide into its complementary nucleotide (**A** → **T**, **T** → **A**, **C** → **G**, **G** → **C**), and then reversing the string.

For example the sequences **ATGCGCGGATG** and **CATCCGCGCAT** are mutually reverse complementary. These strings describe the two strands of the same DNA molecule. Consequently, the two following double strand schemes represent the same molecule:

```
5' - ATGCGCGGATG - 3'
      |||||
3' - TACGCGCCTAC - 5'
```

```
5' - CATCCGCGCAT - 3'
      |||||
3' - GTAGGCGCGTA - 5'
```



Symmetries in DNA sequences

Tandem repeat

GATAAGATAAGATAAGATAA = 2 x GATAAGATAA = 4 x GATAA

GATAAGATAAatgtagGATAAGATAA = 2 x GATAAGATAA separated by a non repeated sequence.

Tandem repeats are presumed to occur frequently in genomic sequences, comprising perhaps **10%** or more of the human genome (*Benson, NAR 27:573,1999*).

Tandem repeat are sometimes associated to a **repeated structure of a protein** (Ex: some ABC transporters have been shown to contain a tandem repeat of six transmembrane helices, *Tusnady et al, FEBS Lett 402:1-3,1997*).

In recent years, the discovery of **short tandem repeat polymorphisms** are involved in various diseases (e.g. Cancer, Huntington, Parkinson,..., *Zhang & Yu, Eur J Surg Oncol,33:529-34,2007*).



ATP-binding Cassette (ABC) transporters



Symmetries in DNA sequences

Textual palindromes

ATGGCCGGTA = ATGGC | CGGTA

Note that the corresponding DNA molecule does not contain any axis of symmetry since in 3D space a nucleotide cannot be superimposed on its own image. Therefore, searching for palindromes is not relevant for detecting biological features.

5' - ATGGC - 3' ↔ 5' - CGGTA - 3'

Symmetries in DNA sequences



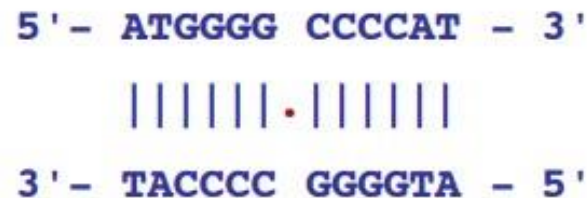
Symmetries in DNA sequences

Reverse complementary palindromes

A reverse complementary palindrome is a sequence identical to its reverse complement.

Example: `ATGGGGCCCCAT`

Reverse complementary palindromes correspond to 3D symmetries in DNA molecules. In the following 2-strand representation, a 180° rotation around the center would swap the two strands, and each letter would take place of an identical letter on the complementary strand.



Note that this sequence is not a textual palindrome.

Note that reverse complementary palindromes can be separated by a stretch of non-symmetrical nucleotides.

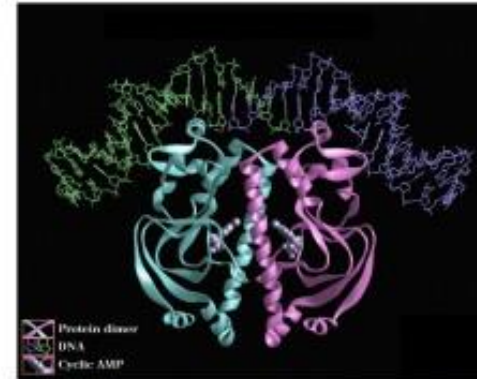
Symmetries in DNA sequences



Symmetries in DNA sequences

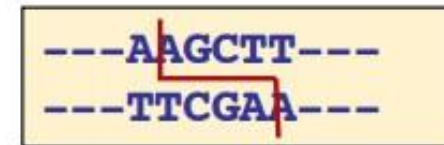
Reverse complementary motifs play important roles in biological mechanisms.

Example 1: some classes of transcription factors (e.g. *helix-turn-helix*) typically form homodimers whose tridimensional structure is symmetrical. These protein complexes specifically recognize reverse complementary motifs in gene promoters.



cAMP Receptor Protein (CRP)
TGTGA-N₆-TCACA

Example 2: In bacteria, hexamers with reverse complementary palindromic structure also play an essential role as recognition sites for *restriction enzymes*.



The restriction enzyme **HindIII**
specifically cuts DNA at instance
of **AAGCTT**

Symmetries in DNA sequences



Symmetries in DNA sequences

Reverse complementary motifs play important roles in biological mechanisms.

Example 3: Reverse complementary motifs separated by a stretch are frequent in RNA, where they mediate the pairing between distant segments of the molecules.

5' - UCGGGcucauaaCCCGA - 3'

folding →



stem loop structure