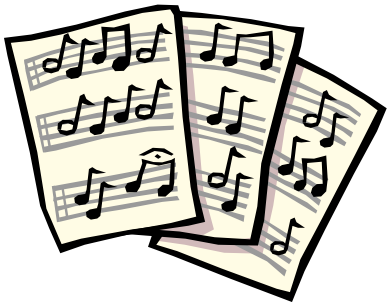
Information Storage and Retrieval using Macromolecules as Storage Media

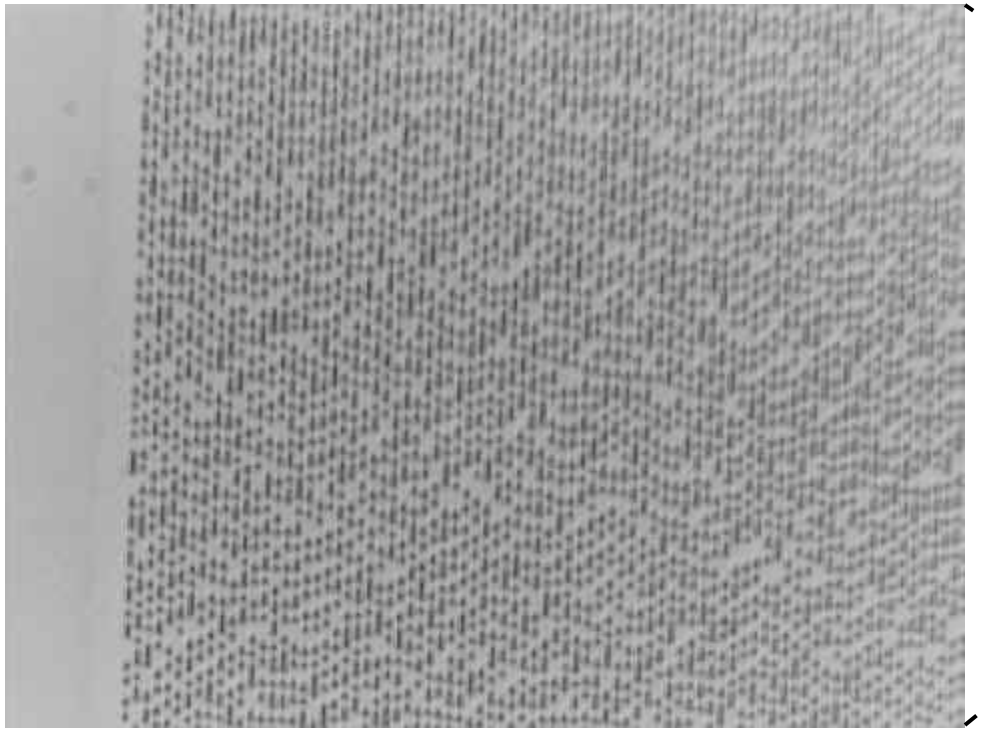
Masud Mansuripur

Optical Sciences Center, The University of Arizona, Tucson, Arizona 85721

March 20, 2003

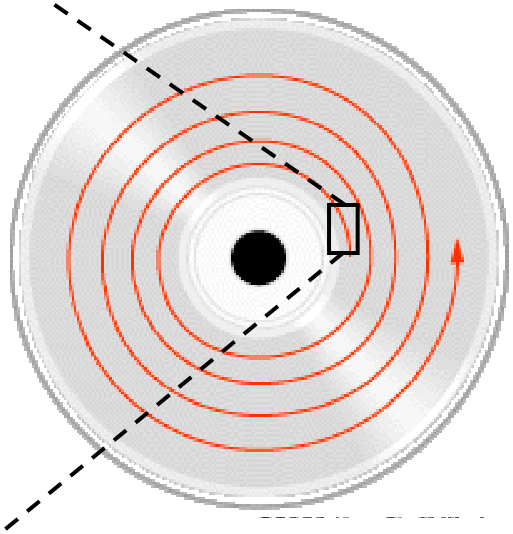
CD Under a Microscope





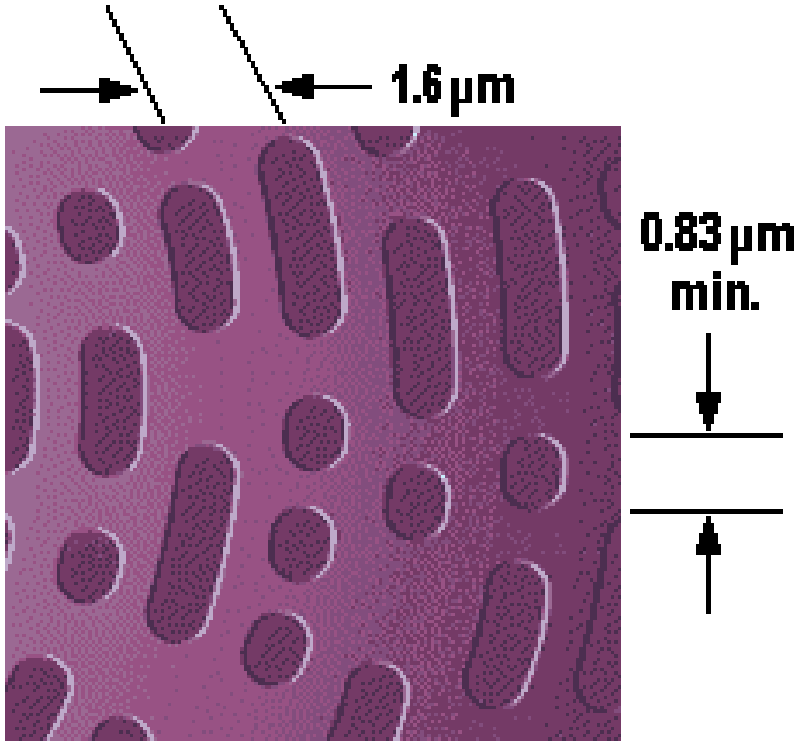
Track direction

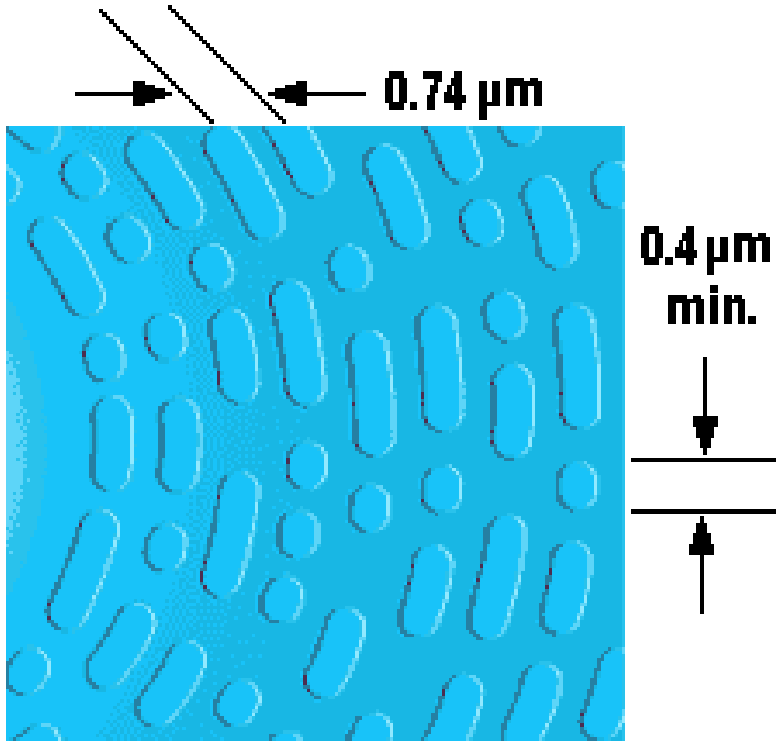
Spiral track



Low-magnification (×32) image of a CD showing an edge of the data zone.

Information Pits on CD and DVD

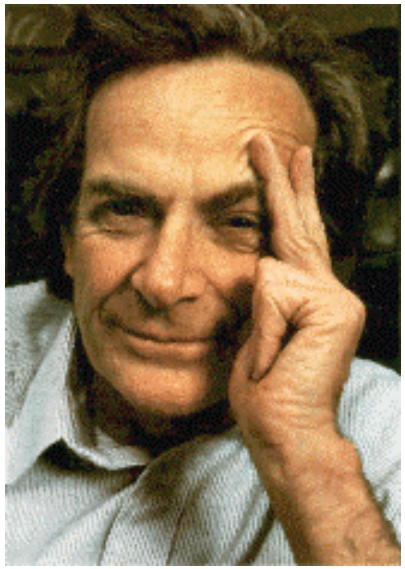




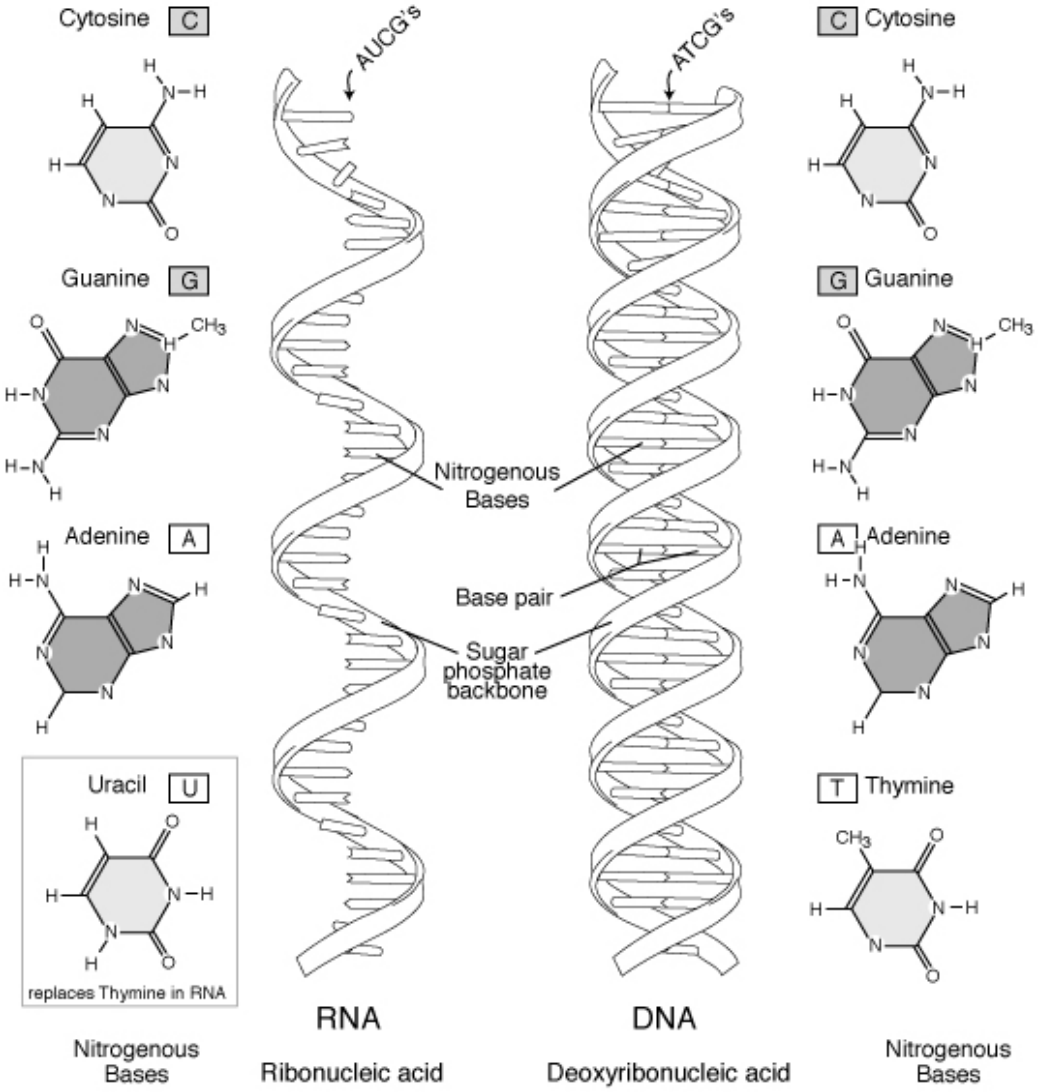
Plenty of Room at the Bottom

Suppose, to be conservative, that a bit of information is going to require a little cube of atoms $5 \times 5 \times 5$, that is 125 atoms. Perhaps we need a hundred and some odd atoms to make sure that the information is not lost through diffusion, or through some other process. I have estimated how many letters there are in the Encyclopaedia, and I have assumed that each of my 24 million books is as big as an Encyclopaedia volume, and have calculated, then, how many bits of information there are (10^{15}). For each bit I allow 100 atoms. And it turns out that all of the information that man has carefully accumulated in all the books in the world can be written in this form in a cube of material one two-hundredth of an inch wide -- which is the barest piece of dust that can be made out by the human eye. So there is *plenty* of room at the bottom! Don't tell me about microfilm!

Richard P. Feynman, December 1959



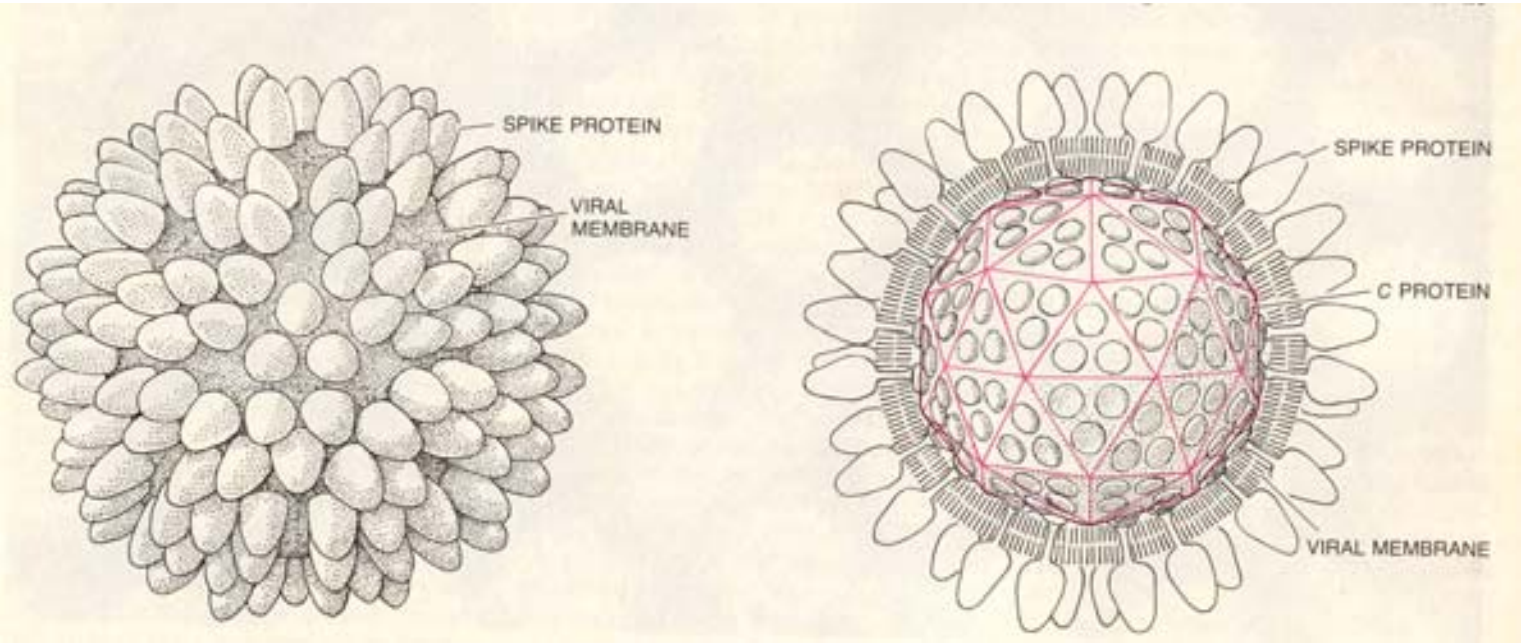
Structure of RNA and DNA



Ribonucleic acid (RNA) is a chemical similar to a single strand of DNA. In RNA the letter U (for uracil) is substituted for T in the genetic code. Messenger RNA delivers DNA's genetic message to the cytoplasm of a cell where proteins are made. Three-letter sequences of messenger RNA, known as codons, code for specific amino acids.

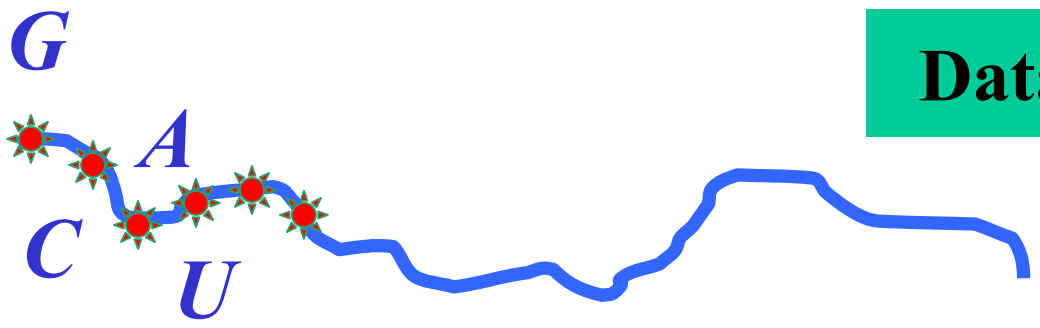
Reproduced with permission from http://www.nhgri.nih.gov

Semliki Forest Virus (SFV)



The SFV particle is 65 nm in diameter. The RNA inside the capsid is a chain of 12,700 nucleotides, the order of which provides the information needed for making the viral proteins. The capsid consists of 180 molecules of a protein, forming a regular polygon with 60 triangular faces. The outer viral membrane is composed of two layers of lipid molecules, with their hydrophobic heads facing outward and their hydrophilic tails facing inward. Inserted into this membrane are 180 spikes, each made up of three linked proteins.

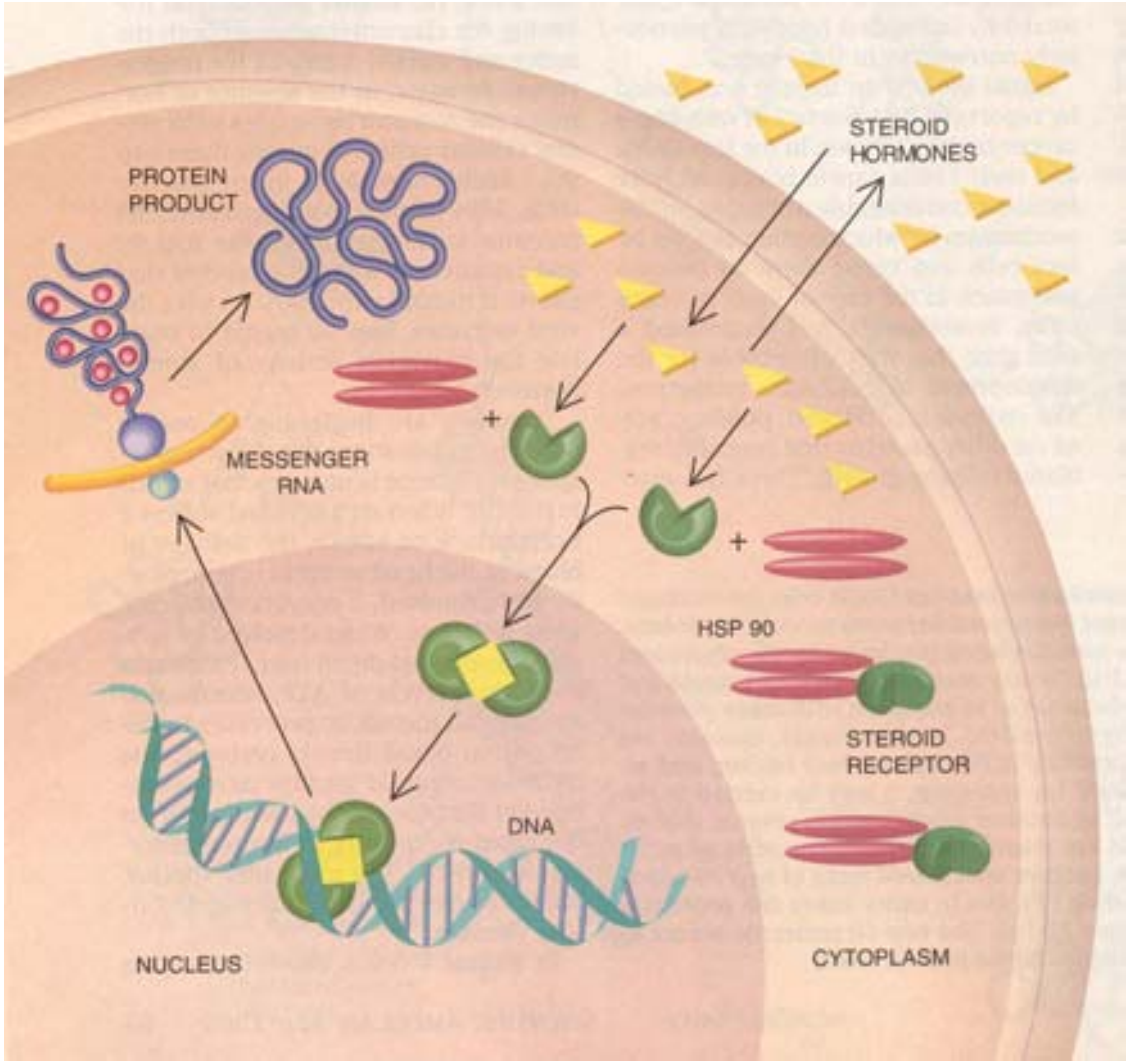
Copyright © 1982, estate of Bunji Tagawa.



Data Storage in RNA

- Number of bases in the RNA strand inside SFV: 12,700
- Number of bits per base (A = 00, G = 01, C = 10, U = 11): 2
- Smallest pit on DVD surface: $400 \times 400 \times 100$ nm*3*
- SFV particle diameter: 65nm
- Number of SFV particles that fit inside a DVD pit: ~ 40
- Number of RNA bits that can be placed inside a DVD pit: 106

Inside the Cell



Responses to steroid hormones are controlled in part by hsp 90. This stress protein helps to maintain steroid receptors in their inactive form. When hormones are present, they bind to the receptor, and the hsp 90 is released. The activated receptor complex can then interact with DNA and initiate the expression of genes for certain proteins.

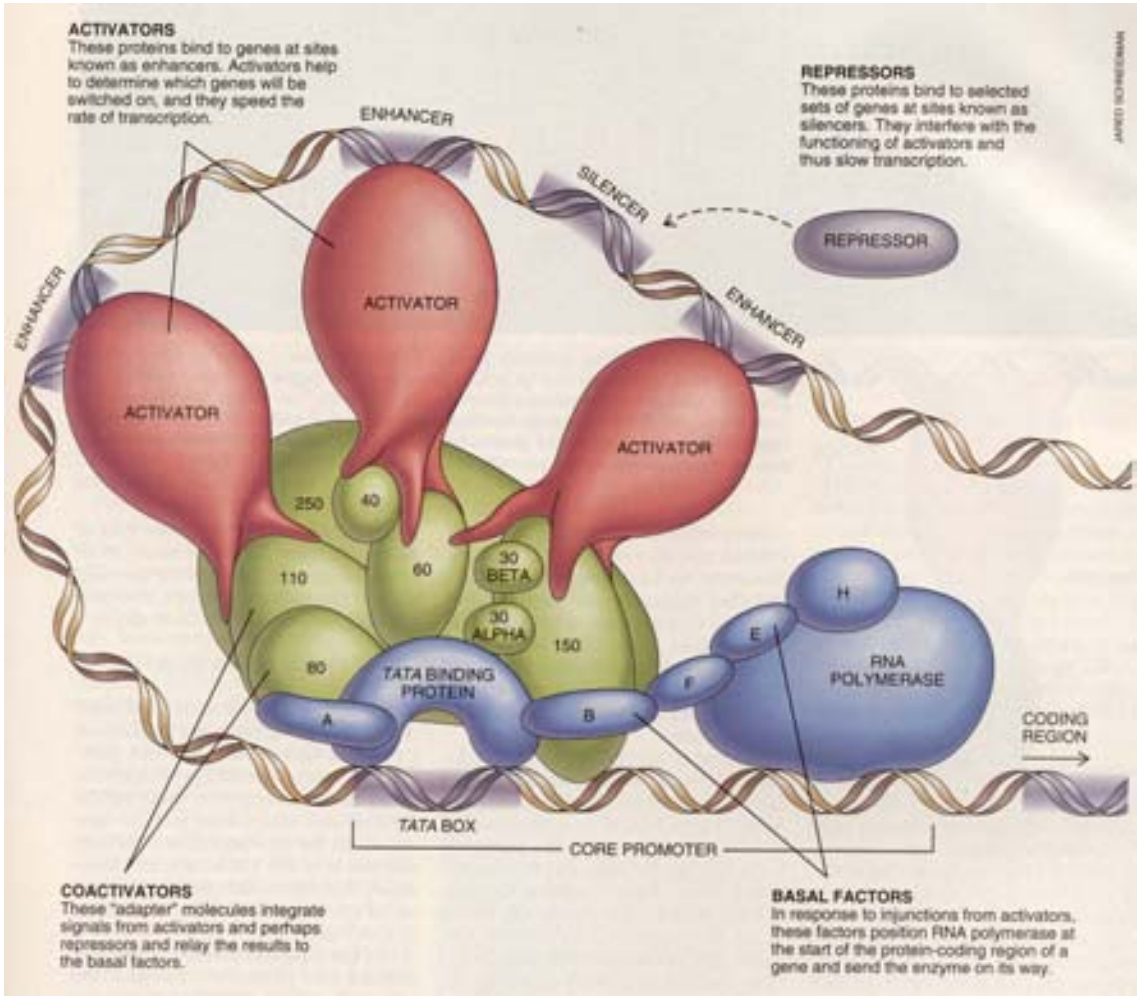
W. J. Welch, "How Cells Respond to Stress," *Scientific American* **268**, 56-64, May 1993.

Biology and Physics

"According to legend, Leo Szilard's baths were ruined by his conversion to biology. He had enjoyed soaking for hours while thinking about physics. But as a convert he found his pleasure punctuated by the frequent need to leap out and search for a fact. In physics – particularly theoretical physics – we can get by with a few basic principles without knowing many facts; that is why the subject attracts those of us cursed with poor memory."

Michael Berry in "Reference Frame: Why are special functions special?" Physics Today, April 2001.

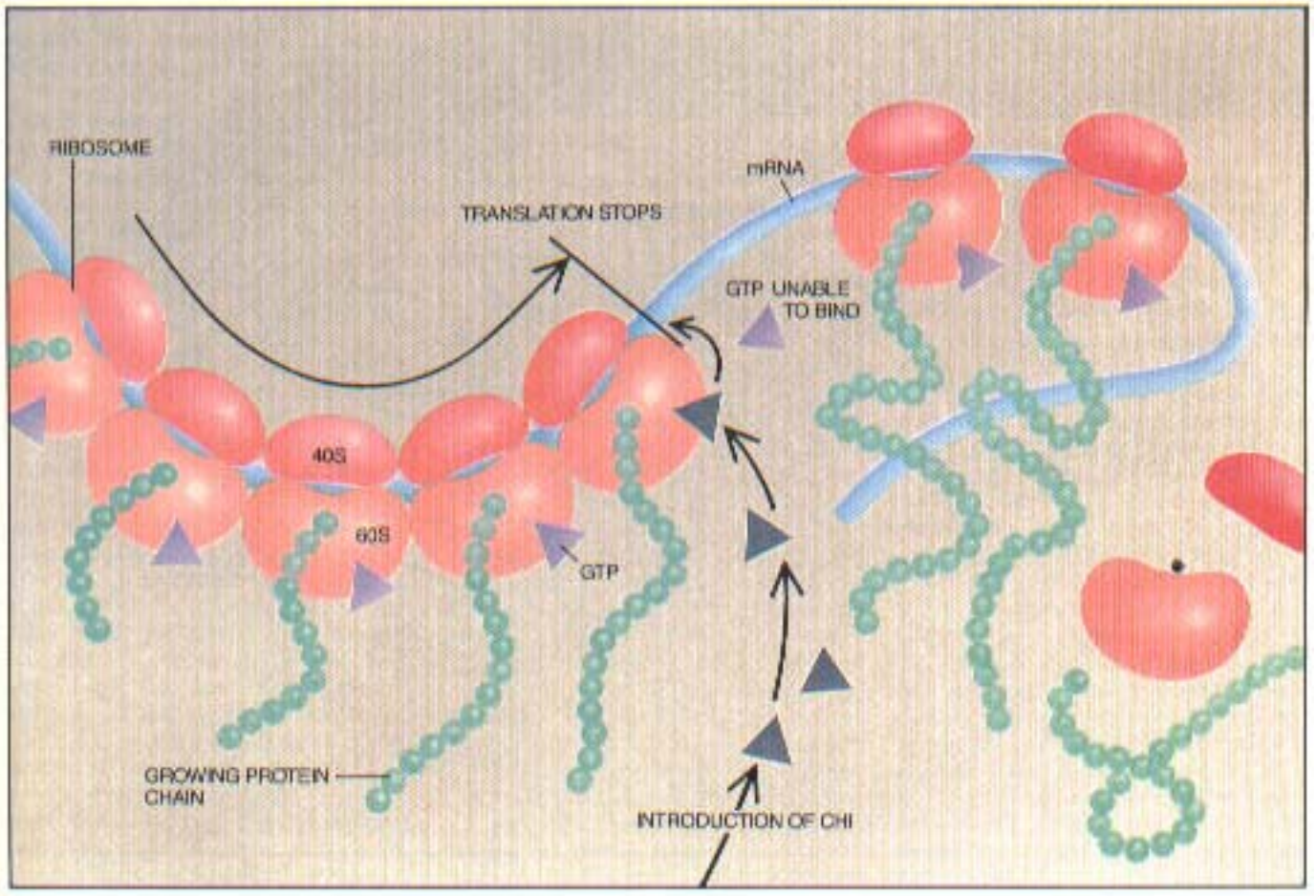
Anatomy of the Transcription Apparatus



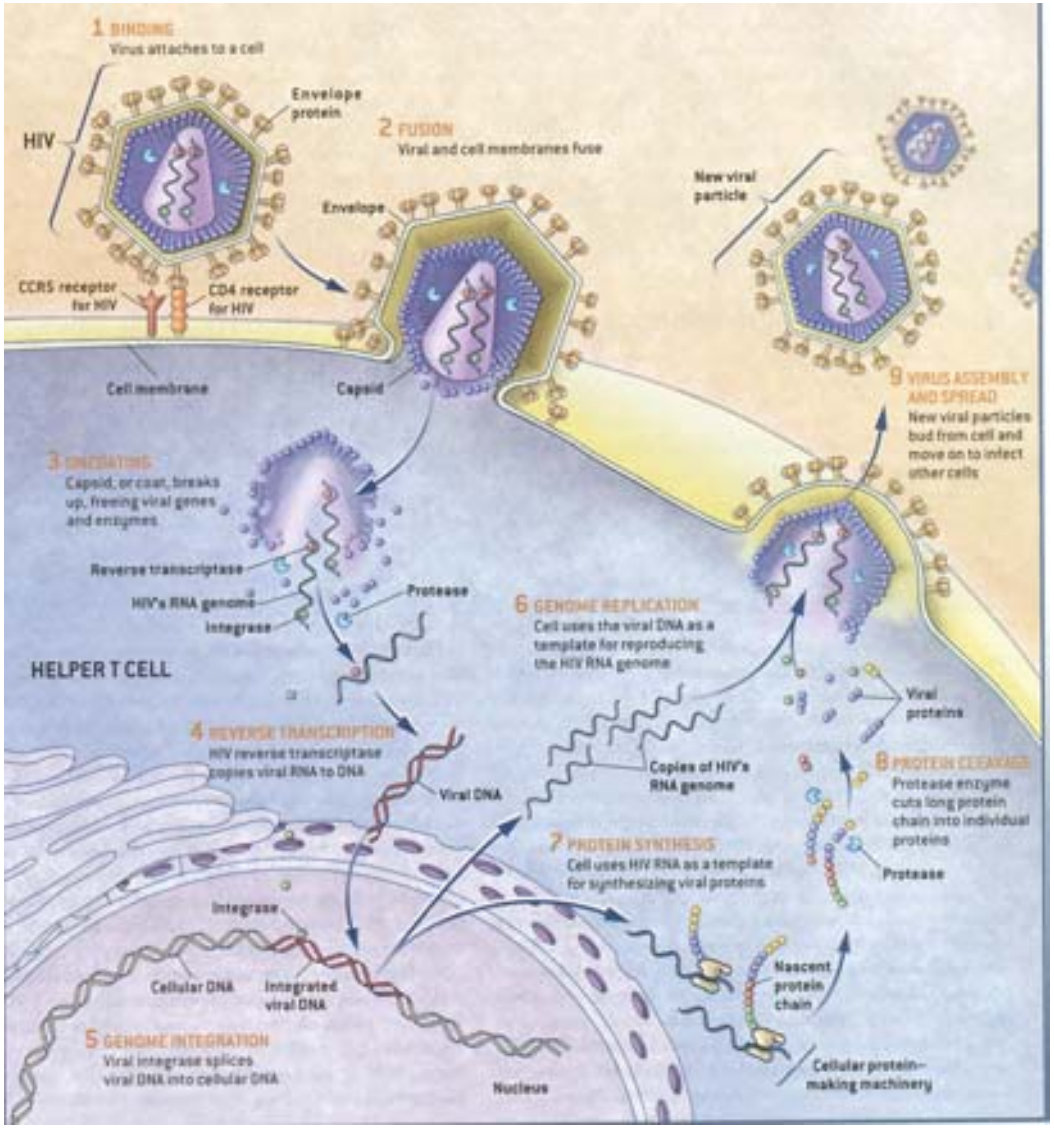
The molecular apparatus controlling transcription in human cells consists of four kinds of components. Basal factors (*blue shapes at bottom*), generally named by single letters, are essential for transcription but cannot by themselves increase or decrease its rate. That task falls to regulatory molecules known as activators (*red*) and repressors (*gray*); these can vary from gene to gene. Activators, and possibly repressors, communicate with the basal factors through coactivators (*green*) – proteins that are linked in a tight complex to the TATA binding protein (TBP), the first of the basal factors to land on a regulatory region of genes known as the core promoter. Coactivators are named according to their molecular weights (in kilodaltons).

From: Robert Tjian, "Molecular Machines That Control Genes," *Scientific American* **272**, 54-61, February 1995.

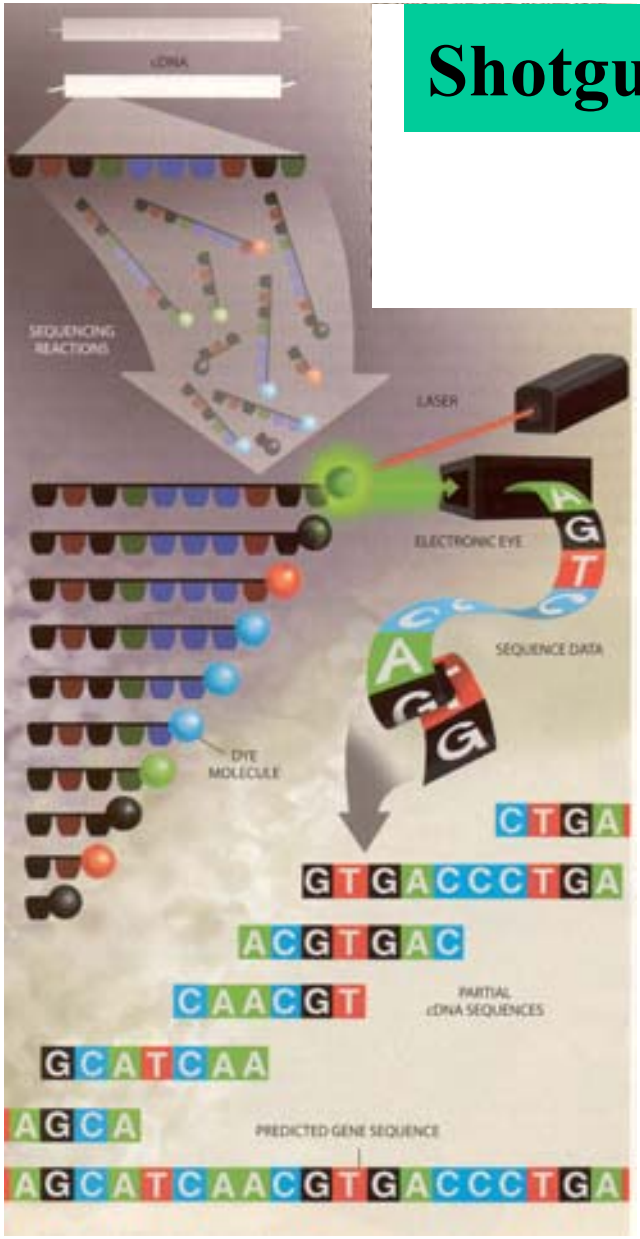
RNA Reading in Nature



HIV Infection



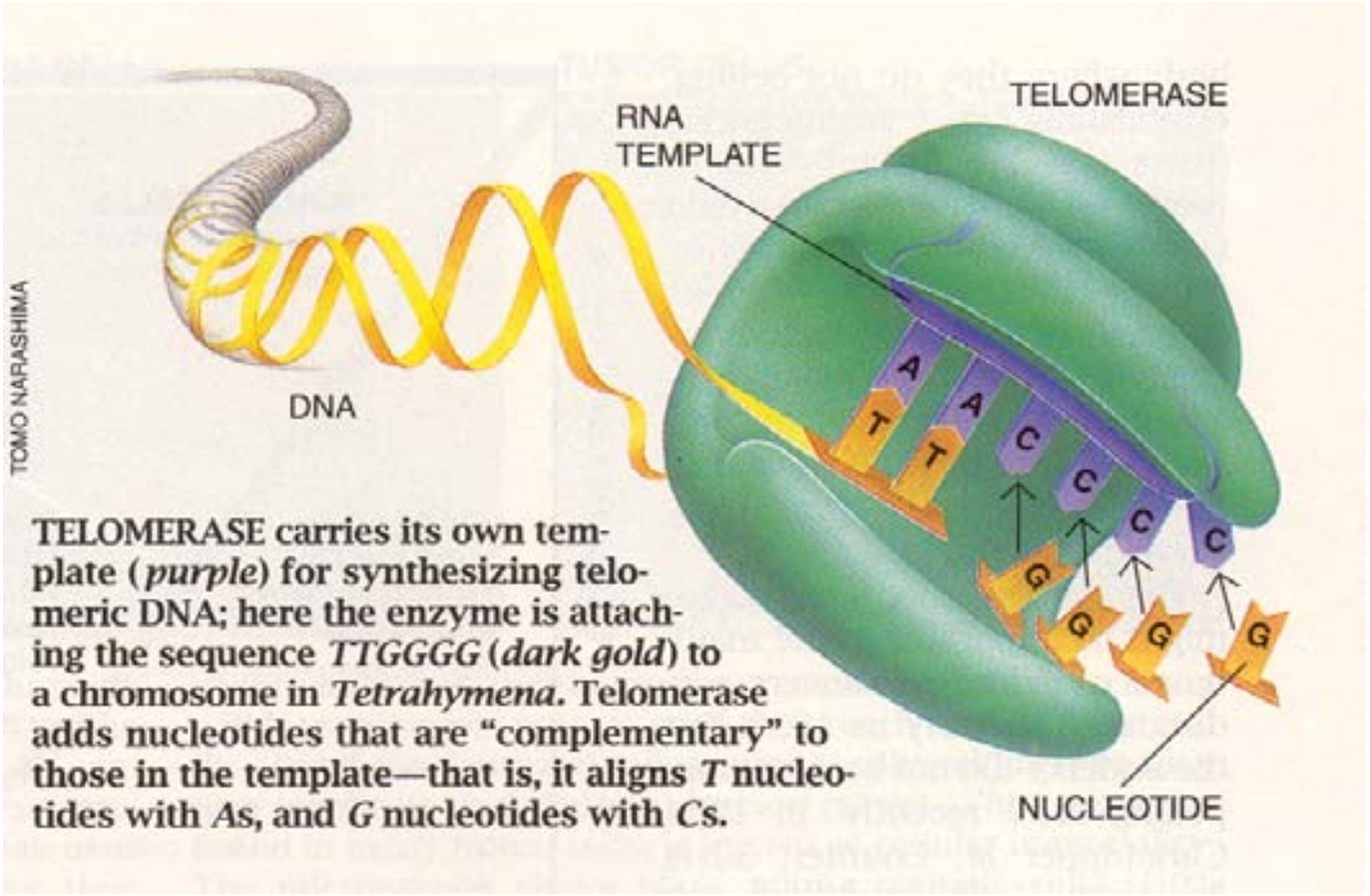
Shotgun Approach to DNA Sequencing



Researchers find partial cDNA sequences by chemically breaking down copies of a cDNA molecule to create an array of fragments that differ in length by one base. In this process, the base at one end of each fragment is attached to one of four fluorescent dyes, the color of the dye depending on the identity of the base in that position. Machines then sort the labeled fragments according to size. Finally, a laser excites the dye labels one by one. The result is a sequence of colors that can be read electronically and that corresponds to the order of the bases at one end of the cDNA being analyzed. Partial sequences hundreds of bases in length can be pieced together in a computer to produce complete gene sequences.

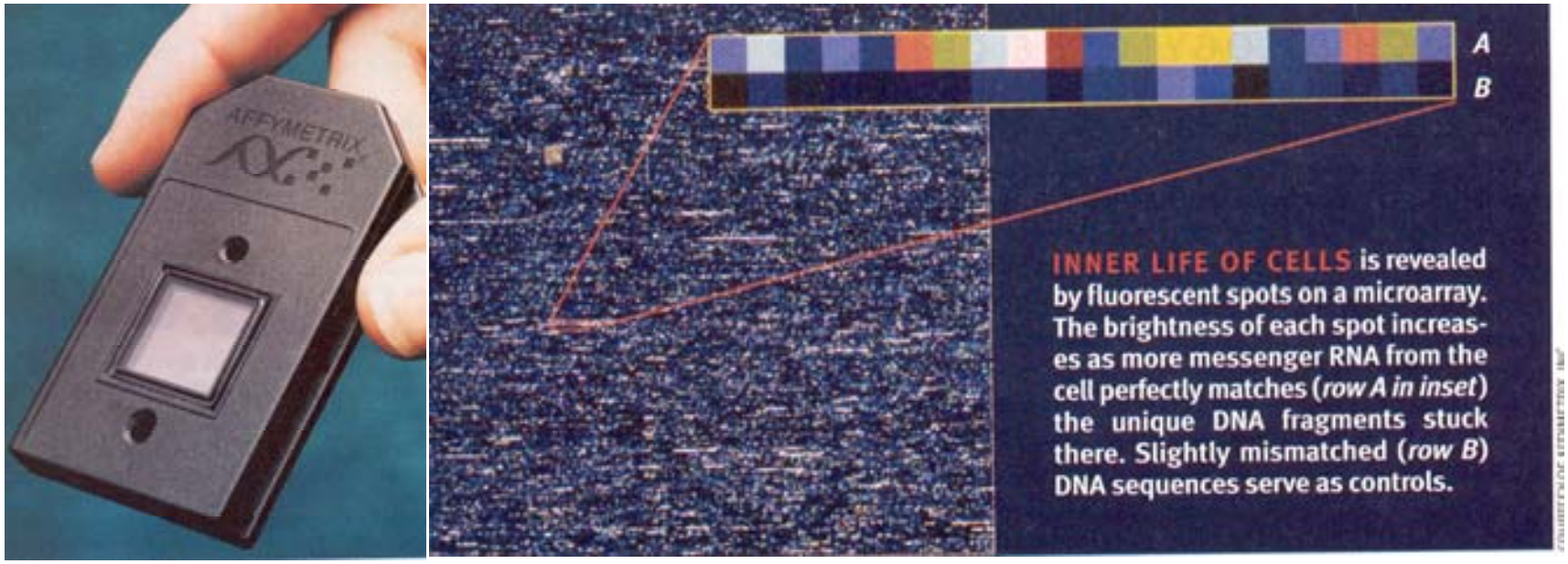
From W. A. Haseltine, "Discovering Genes for New Medicines," *Scientific American* **276**, March 1997.

DNA Writing in Nature



From C. W. Greider and E. H. Blackburn, "Telomeres, Telomerase, and Cancer," *Scientific American* **274**, February 1996.

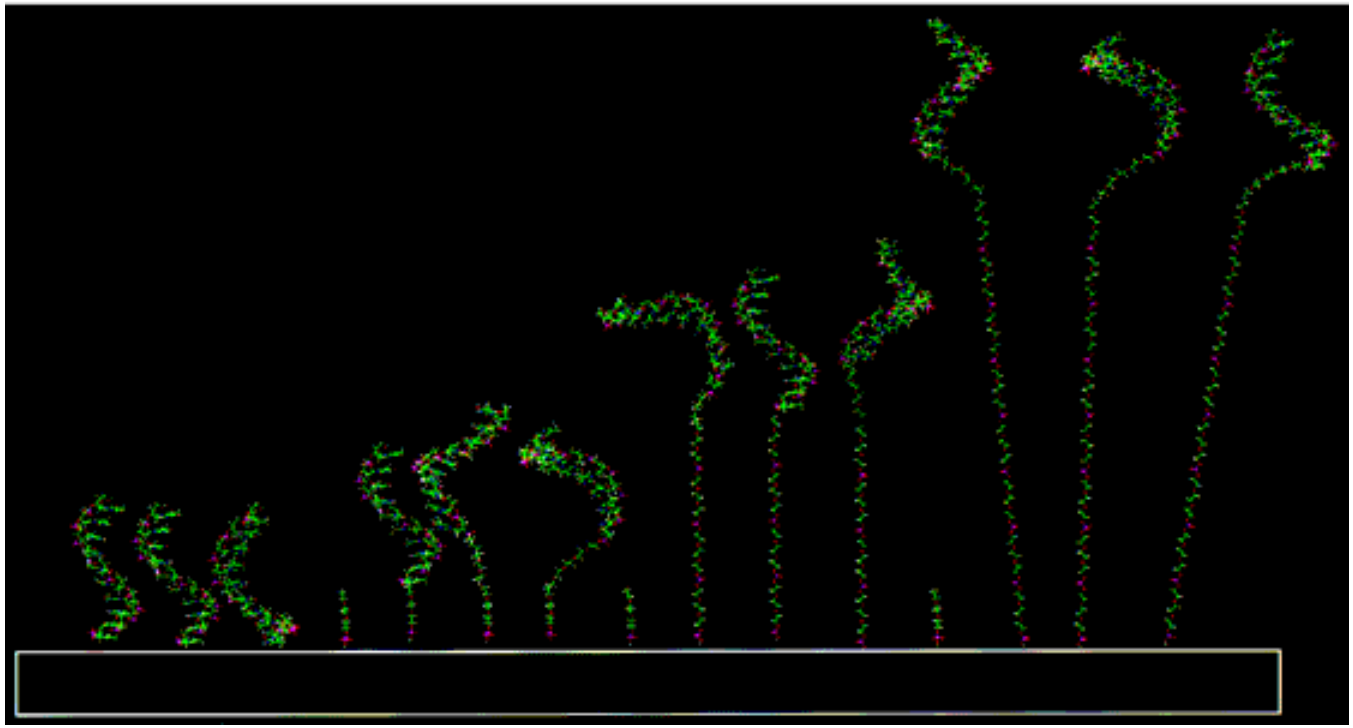
Gene-chip Technology



DNA chip can sense the state of up to 400,000 genes in a tissue sample. The number of probes on a single glass wafer may soon exceed 60 million, in which case the entire human genome will fit on 200-300 wafers.

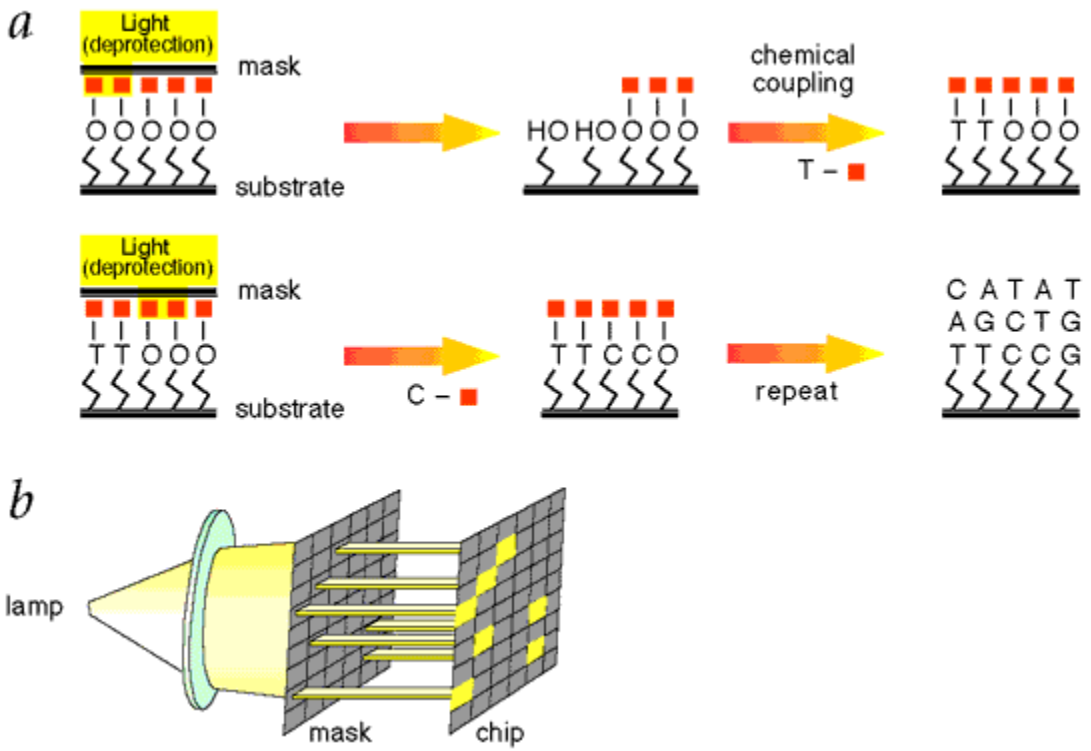
Copyright © 2001, Affymetrix, Inc.

Inside the Gene-chip



The density of oligonucleotides on the surface is approximately 10 pmol per mm^2 on aminated polypropylene, approximately 0.1 pmol per mm^2 on glass after ammonia deprotection——equivalent to one molecule per 39 square angstroms. The oligonucleotides are just about within reach of each other on glass, but rather closely packed on polypropylene supports. Spacers help to overcome steric interference, which can take a number of forms: the ends of the probes closest to the surface are less accessible than the ends furthest away; tethered molecules may crowd each other. Oligonucleotides on long spacers are better able to extend away from their neighbours and from the surface to allow interaction with the target. In this and other figures, the molecules are shown in a stretched conformation. It is likely that the molecules are in a dynamic state which may include this as one extreme, but in which the average state is somewhat more condensed. The linkers illustrated are oligoethylene glycols 26, 60 and 105 atoms in length.

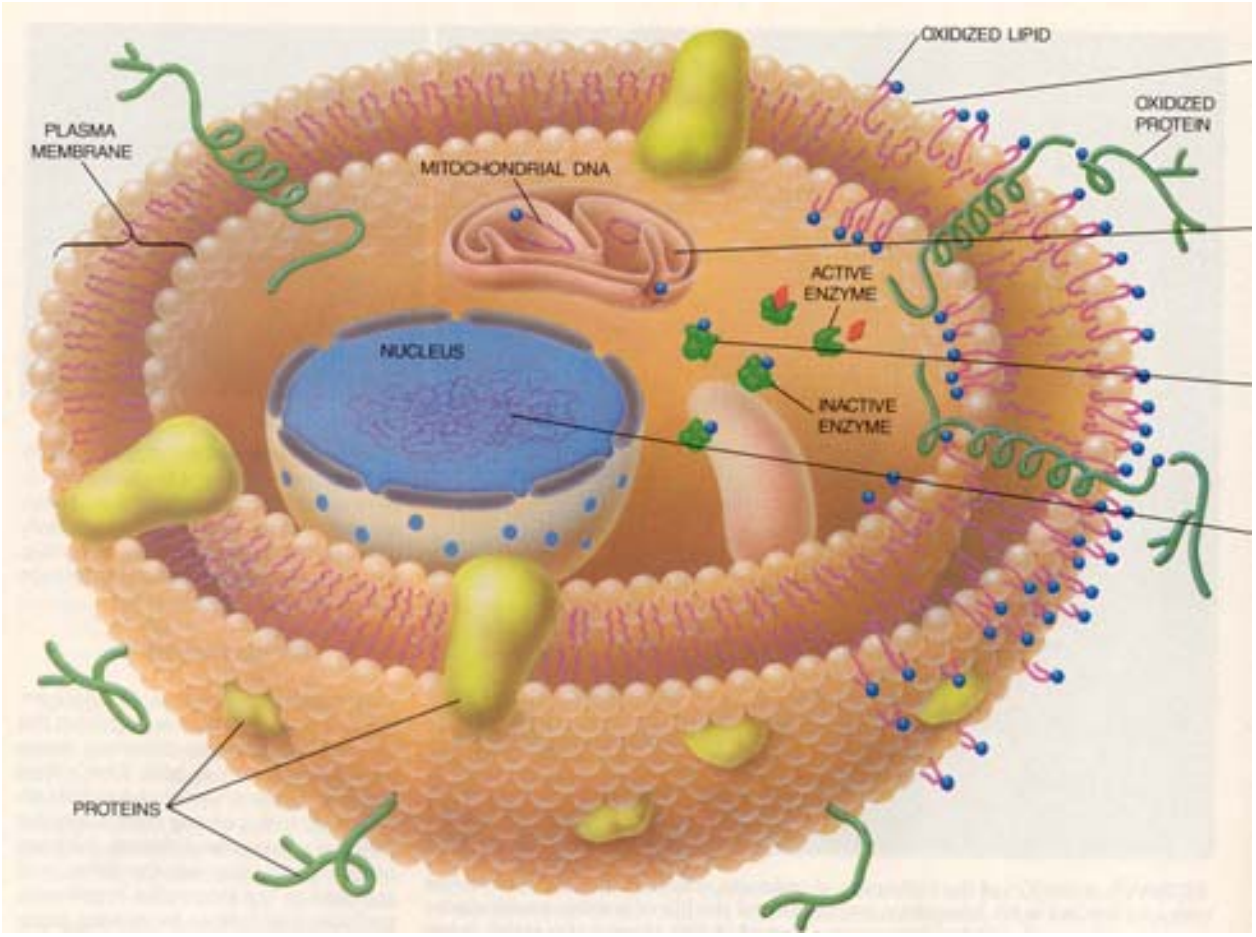
Synthetic Oligonucleotide Arrays



(a) Light directed oligonucleotide synthesis. A solid support is derivatized with a covalent linker molecule terminated with a photolabile protecting group. Light is directed through a mask to deprotect and activate selected sites, and protected nucleotides couple to the activated sites. The process is repeated, activating different sets of sites and coupling different bases allowing arbitrary DNA probes to be constructed at each site. (b) Schematic representation of the lamp, mask and array.

Robert J. Lipshutz *et al., Nature Genetics* 21, 20 – 24 (1999).

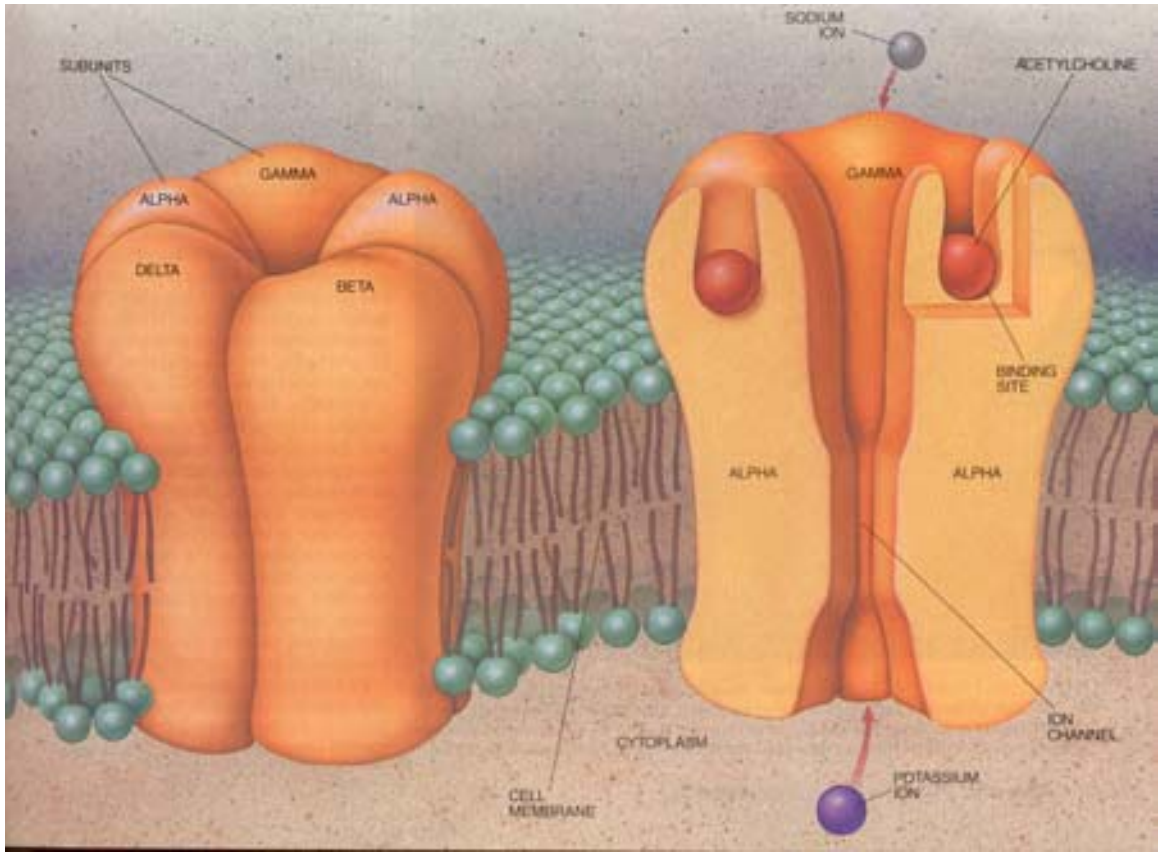
Proteins within the Cell Membrane



Oxidative attack on proteins (shades of green), lipids (pink) and DNA (lavender) -indicated schematically by attachment of a small, bright blue sphere- can impair the functioning of many cellular constituents (only some are depicted). One leading theory of aging, originally set out in the mid 1950s, holds that the human body deteriorates over time because it continuously generates potentially destructive oxidizing agents known as free radicals.

From: R. L. Rusting, "Trends in Biology: Why Do We Age," *Scientific American* **267**, 86-95, December 1992.

Gated Ion Channel



ACETYLCHOLINE RECEPTOR, which consists of five subunits (left), was the first neurotransmitter receptor to be isolated. Later work showed it to include not only neurotransmitter binding sites but also an ion-transporting channel (right). (The beta and delta subunits and part of one alpha subunit have been cut away for clarity.) The channel is closed when the receptor is at rest, but it opens rapidly when the two alpha subunits both combine with acetylcholine.

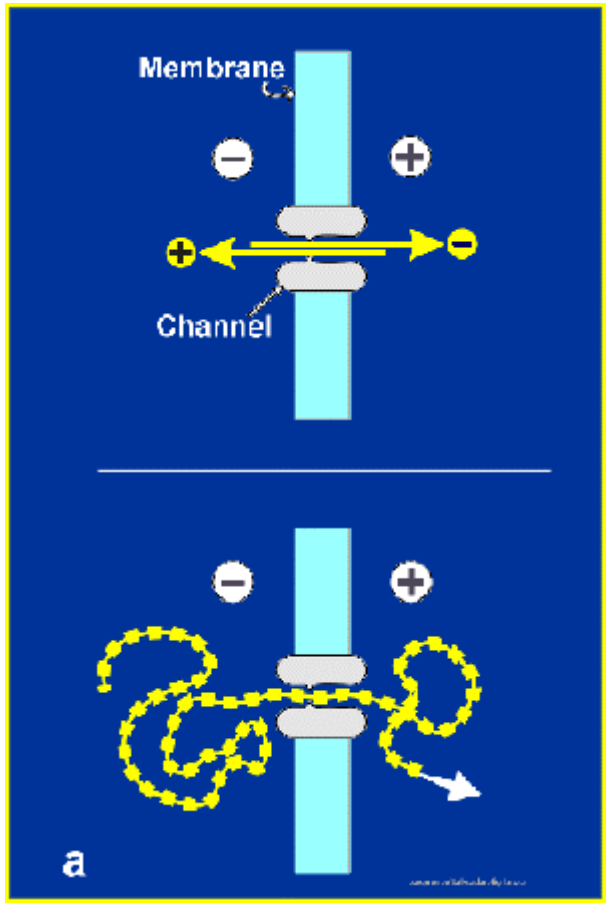
From: J-P. Changeux, "Chemical Signaling in the Brain," *Scientific American* 269, 58-62, November 1993.

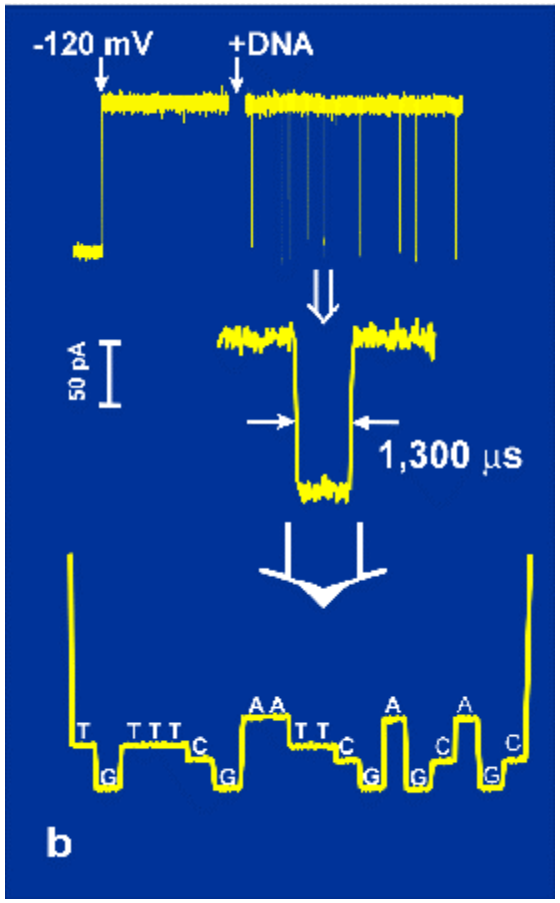
Probing Polynucleotides with a Nanopore: High Speed, Single Molecule DNA Sequencing

A revolutionary technology for probing, and eventually sequencing, individual DNA molecules using single-channel recording techniques.

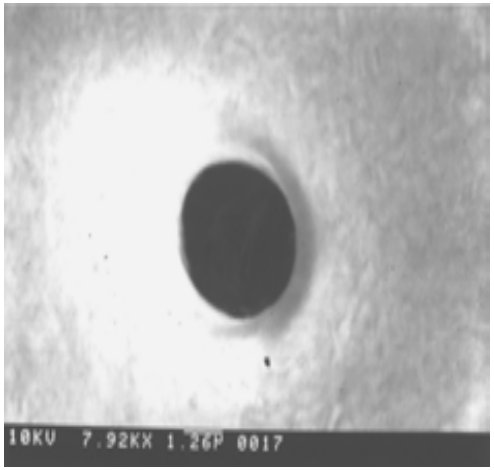
- Translates the characteristics of a polynucleotide into electronic signals;
- Has the potential to sequence DNA at unprecedented speeds;
- Can probe very long stretches of DNA or RNA;
- Is a high throughput, single-molecule technique compatible with high levels of nano-fabrication.

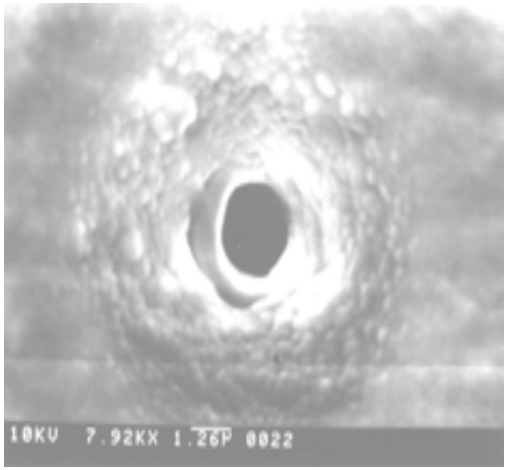
These design criteria are met by an instrument which draws single molecules of DNA through a small channel or pore.

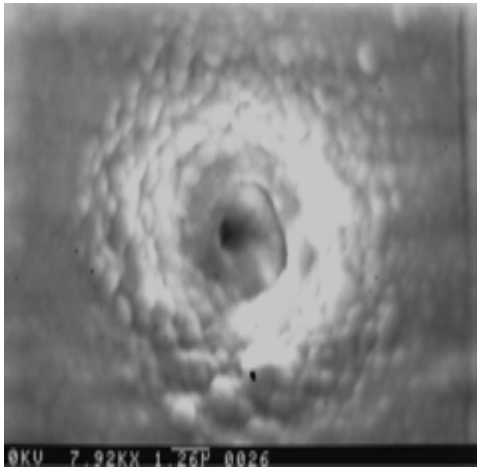




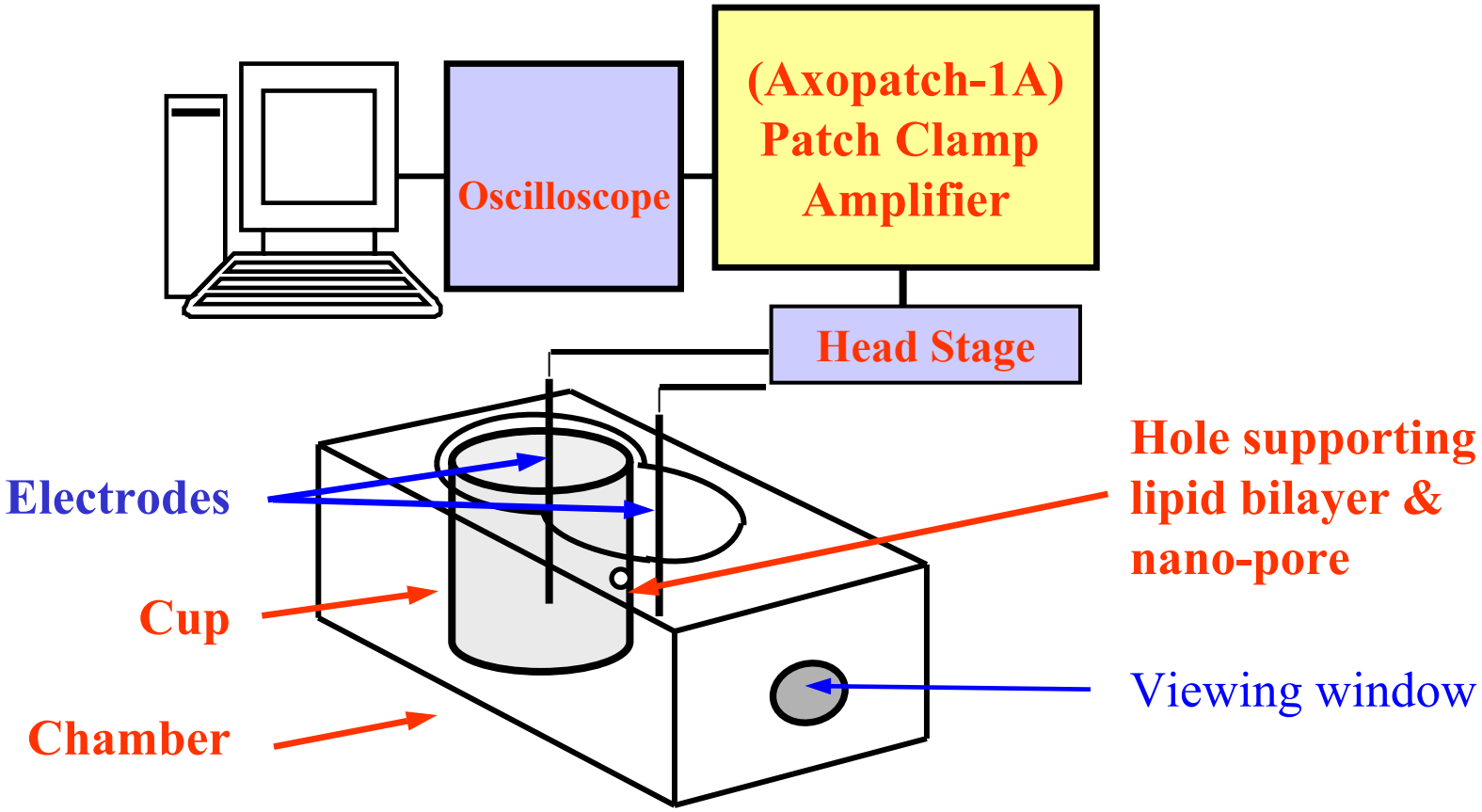
Solid-state nano-pores



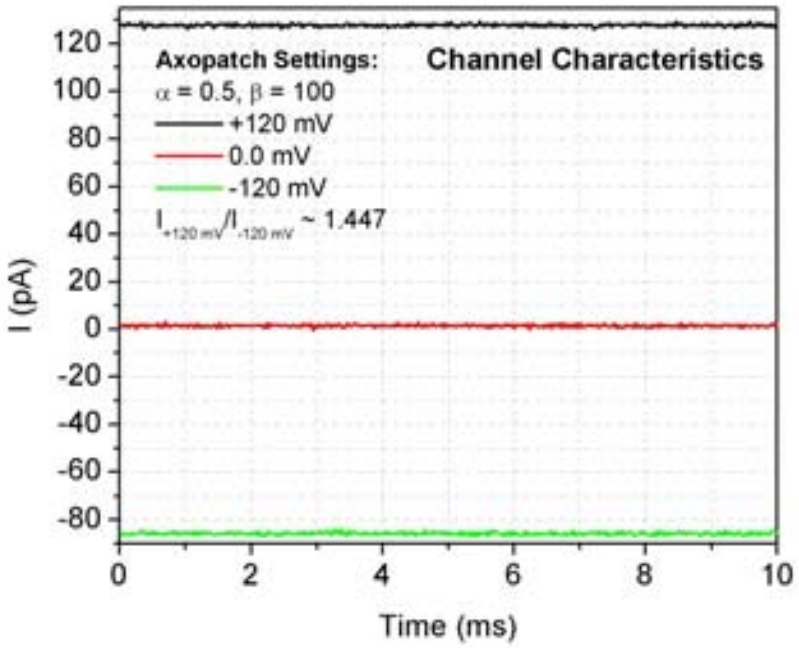




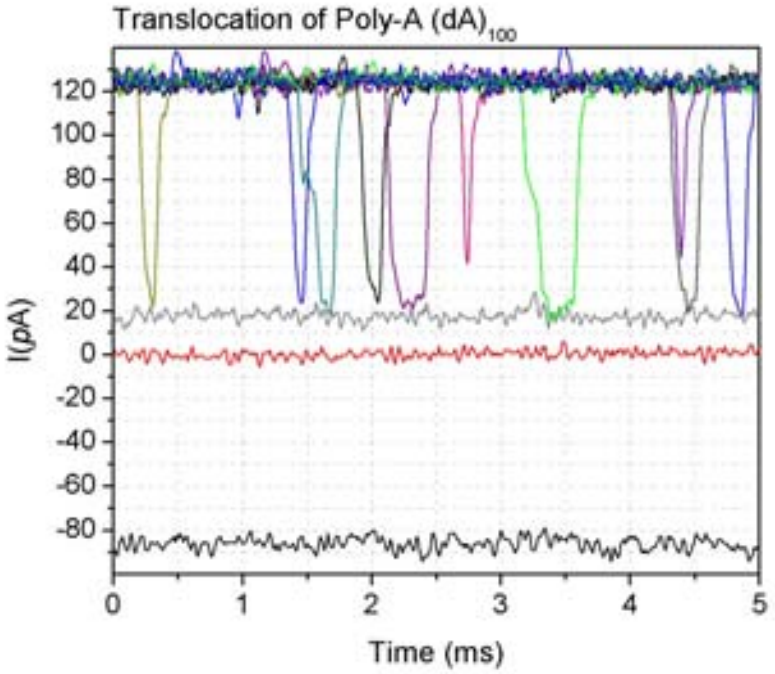
Experimental Setup



DNA Translocation Experiments

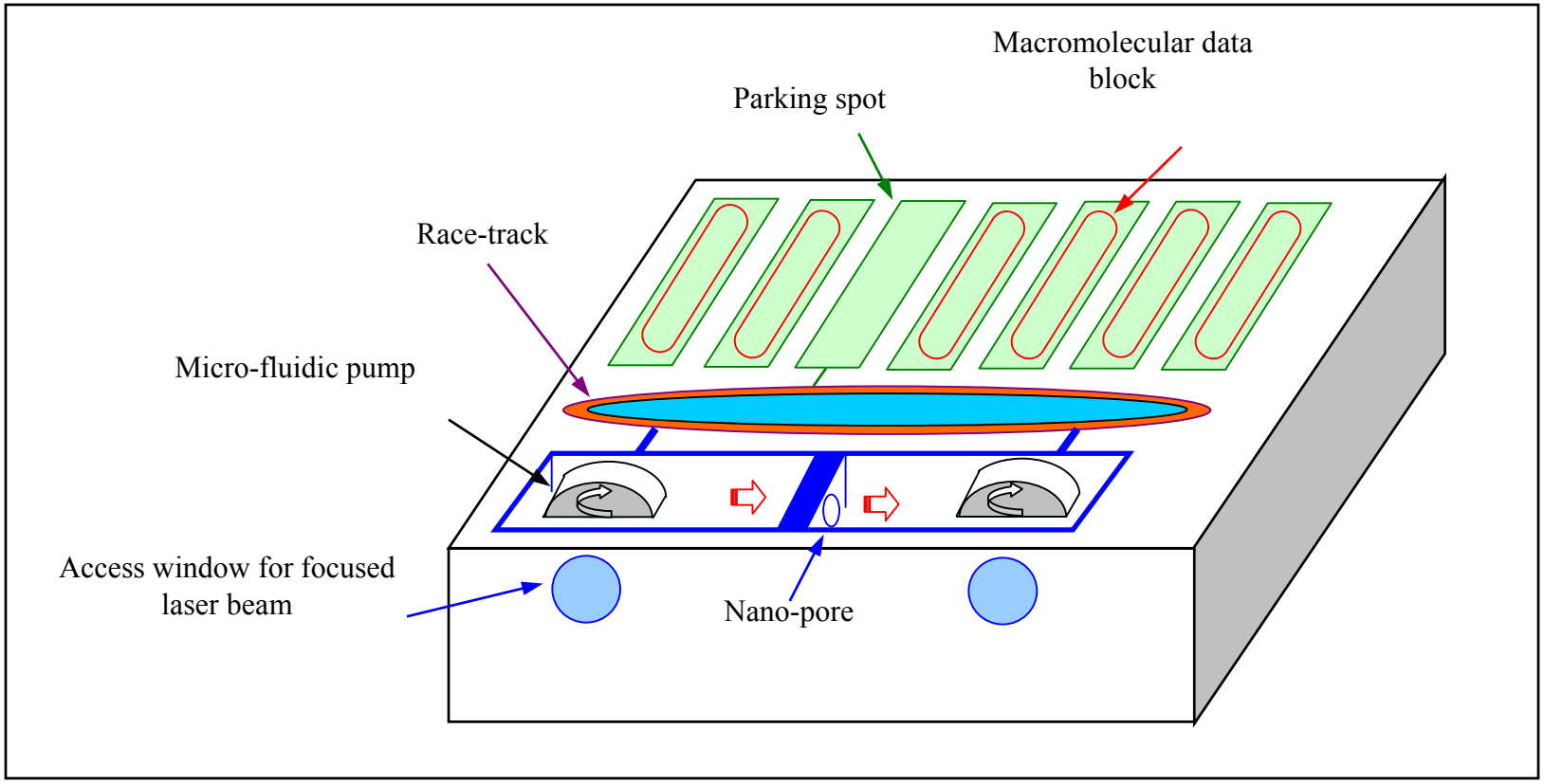


Characteristics of a nano-pore ion-channel in the absence of DNA molecules.



Translocation of ploy-A(5'A_{100}3') with positive and negative voltages across the channel.

Translocating macromolecular data-blocks through read station



DNA Storage Chip

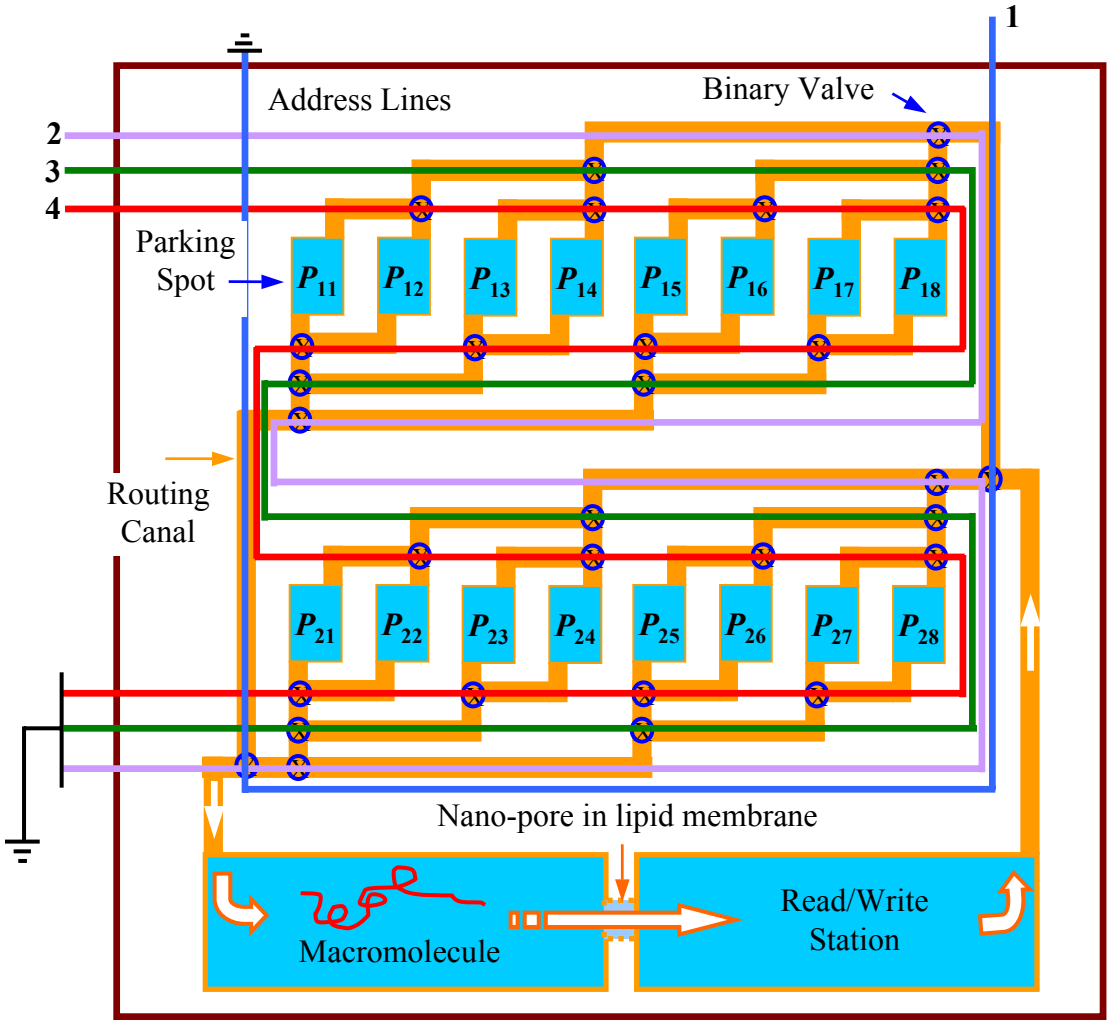
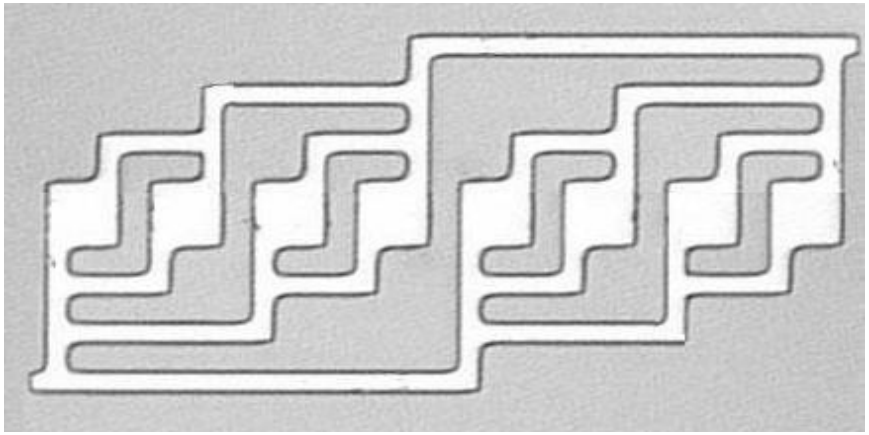


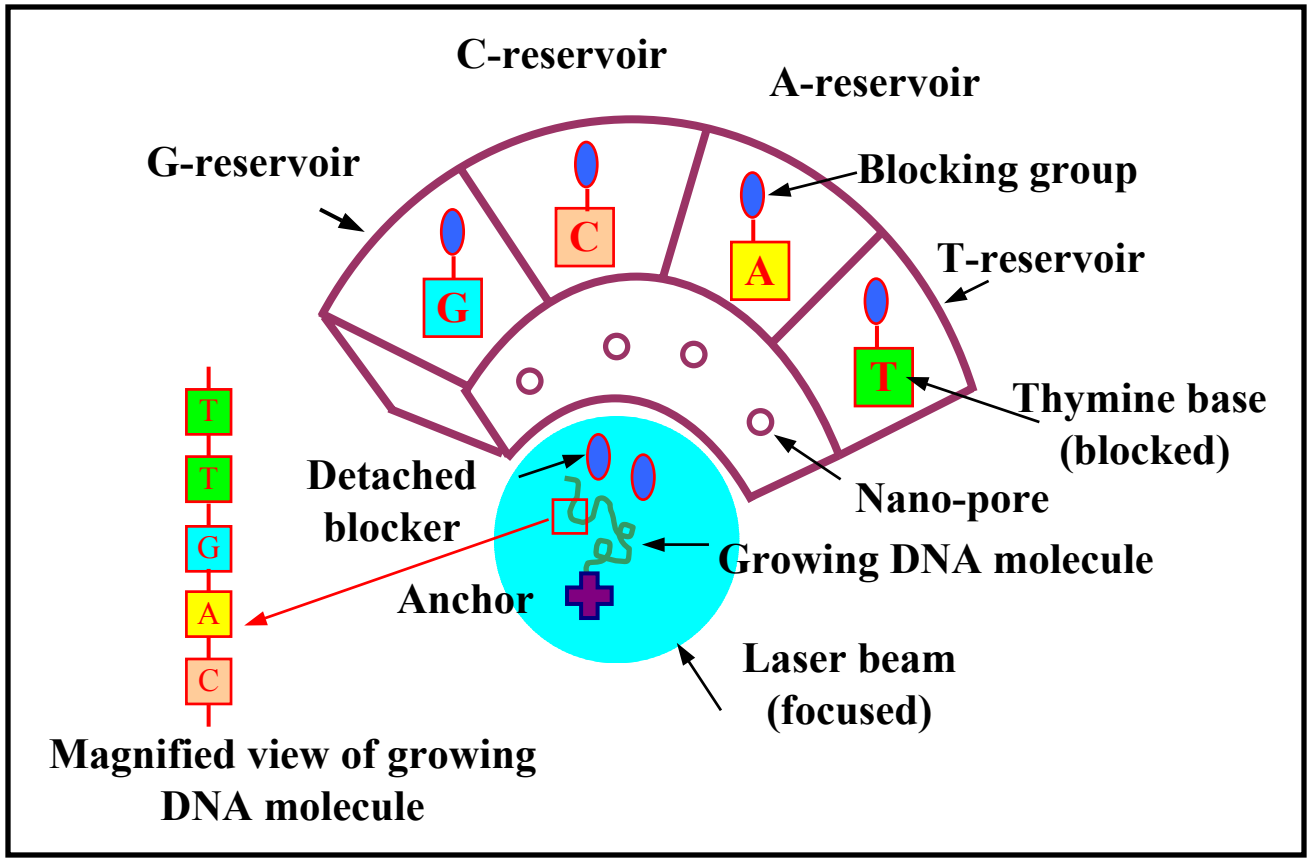
Diagram showing the patterned surface of the proposed data storage chip. Once a parking spot is selected, its macro-molecular content will be transferred to the Read/Write Station under the influence of an applied electric voltage (i.e., electro-phoretic transfer). Following the completion of the read/write operation, the macromolecule should be returned to its designated parking spot.

Blue-laser Writer



Scanning Photolithography **system built around a conventional optical microscope using a blue laser diode (λ = 405 nm). The SEM micrograph above shows 8 parking spots, each $50 \times 50 \mu m^2$ (5.5μm depth) produced by blue laser writing, followed by freon ion-milling.**

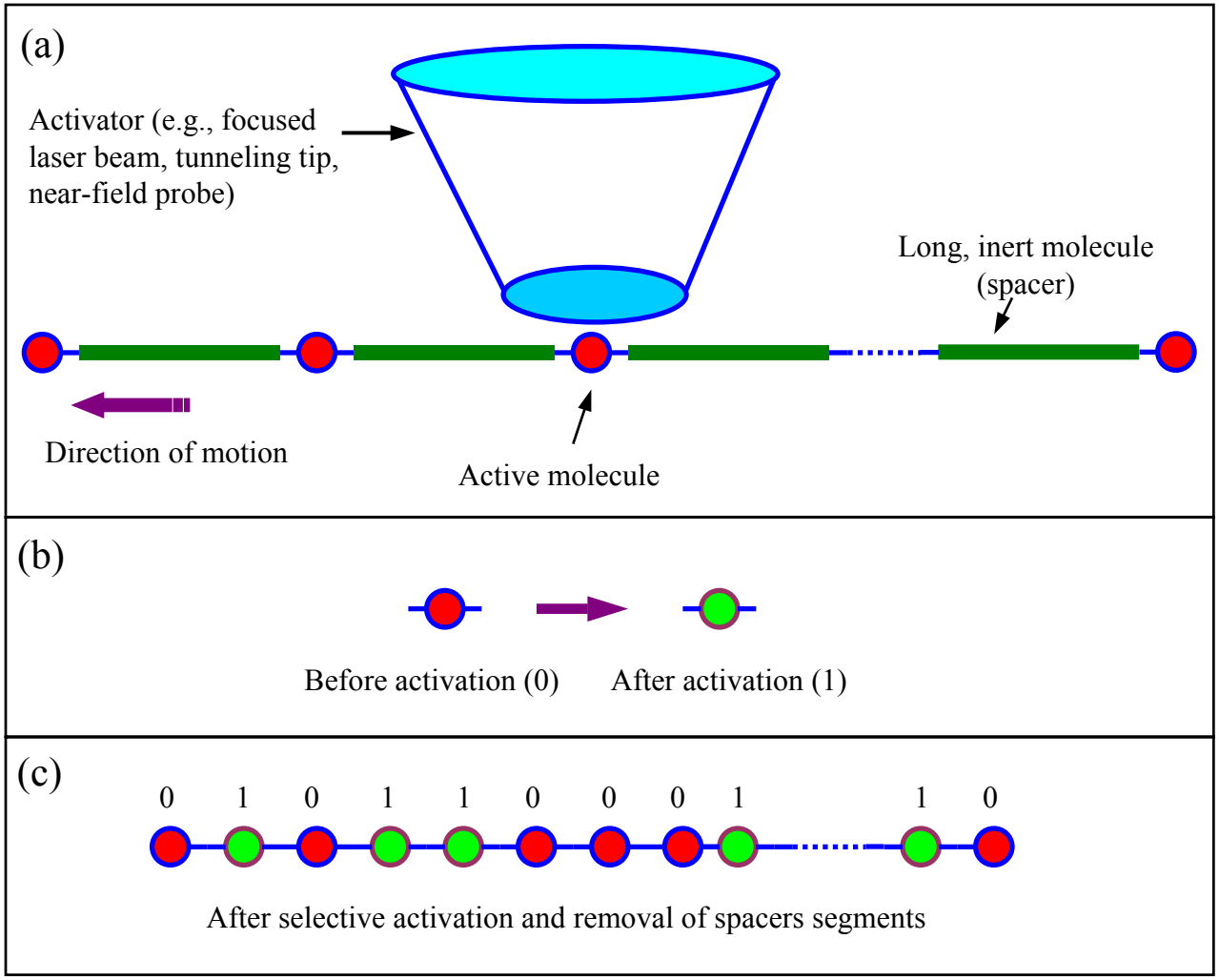
DNA Writing Scheme



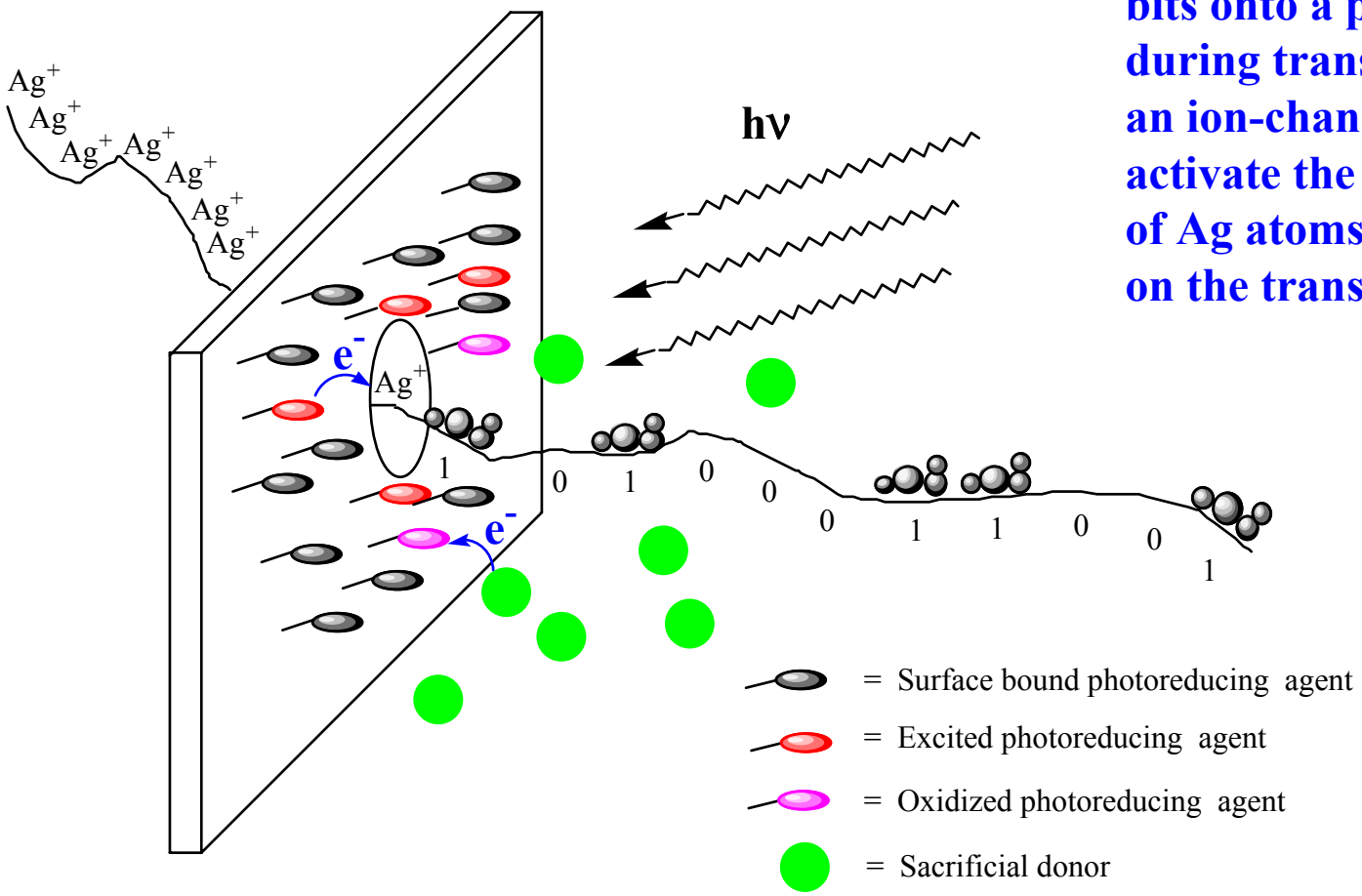
A write station for creating DNA molecules of arbitrary base sequence. Four isolated chambers act as reservoirs for the four nucleic acid bases; all base molecules stored in these reservoirs are blocked by a chemical group to ensure the attachment of a single base to the growing DNA strand at each stage of the process. The reservoirs are connected through single nano-pores to the main chamber, where a growing DNA molecule is anchored at one end but is freely floating otherwise. When the growing DNA strand enters individual reservoirs through the corresponding nano-pore, it adds the desired base to its growing end, then returns to the main chamber. A photo-activation process in the main chamber (initiated by the focused laser beam) removes the cap from the growing end of the strand. Upon completion of recording, the DNA molecule is detached from the anchor and parked in a designated spot.

Alternative Writing Scheme

(a) A precursor strand consisting of long, inert segments separating active molecules in their native (or ground) state, is exposed to an activator. The strand is pulled past the activator by a micro-actuator. (b) When the activator is energized, the active molecules are transformed from their native (ground) state to an excited state. When the activator is turned off, however, the active molecules remain in their ground state. (c) Once the entire precursor strand is written (i.e., its active molecules selectively placed in the excited state), the inert spacer molecules are removed, and the active molecules spliced together in the same order in which they were switched by the activator.

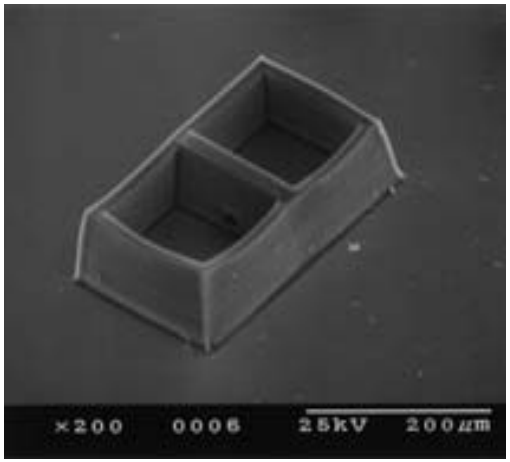


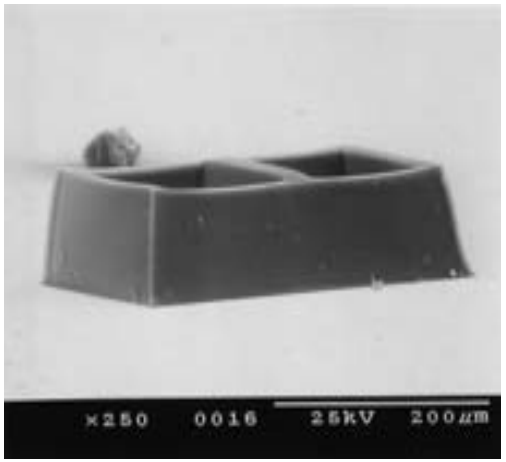
Silver nano-particles on DNA backbone

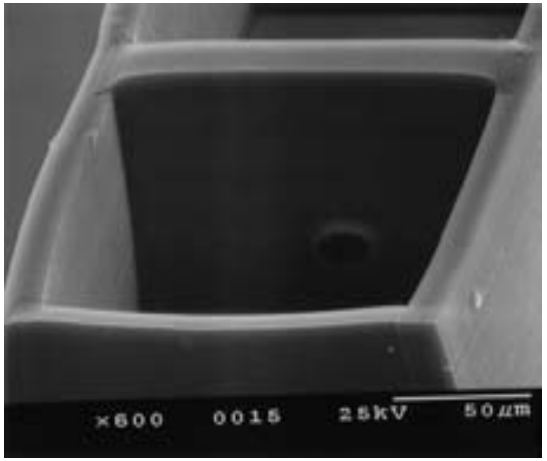


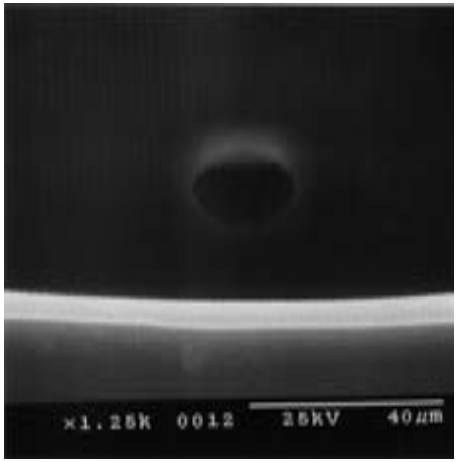
Writing of metal nanocluster bits onto a polynucleotide chain during translocation through an ion-channel. Laser pulses activate the photo-generation of Ag atoms that form clusters on the translocating chain.

Micro-chambers





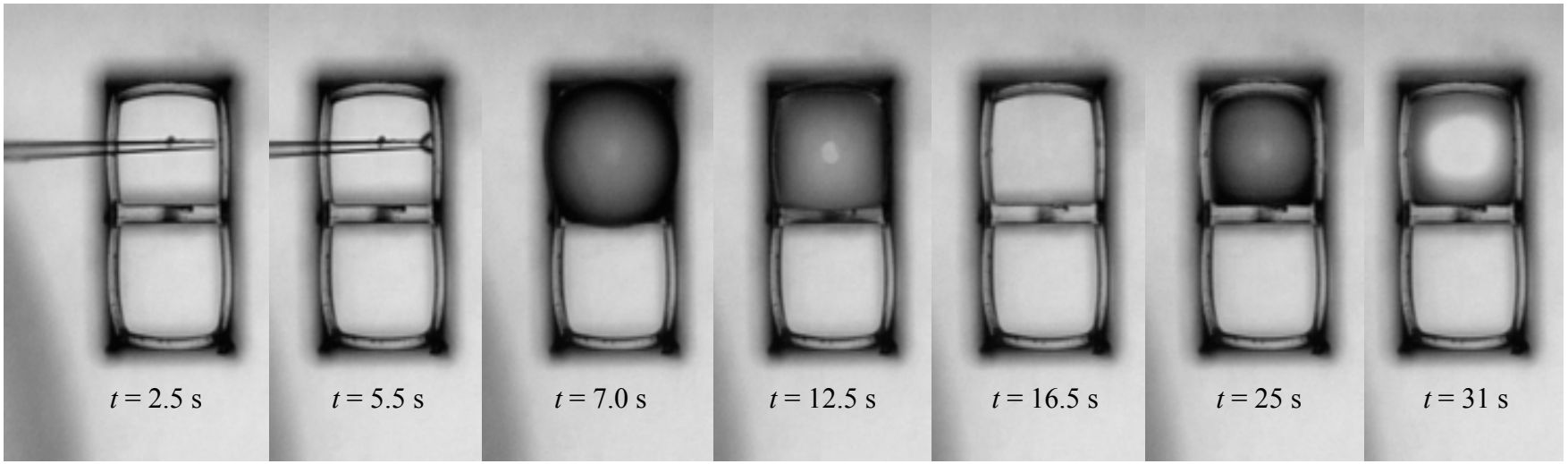




Two views of the micro-chamber fabricated by two-photon lithography. Each chamber is approximately 150μm on each side, and 150μm deep.

Two views of the μ-hole in the wall separating the two chambers. The μ-hole is ~20 μm in diameter.

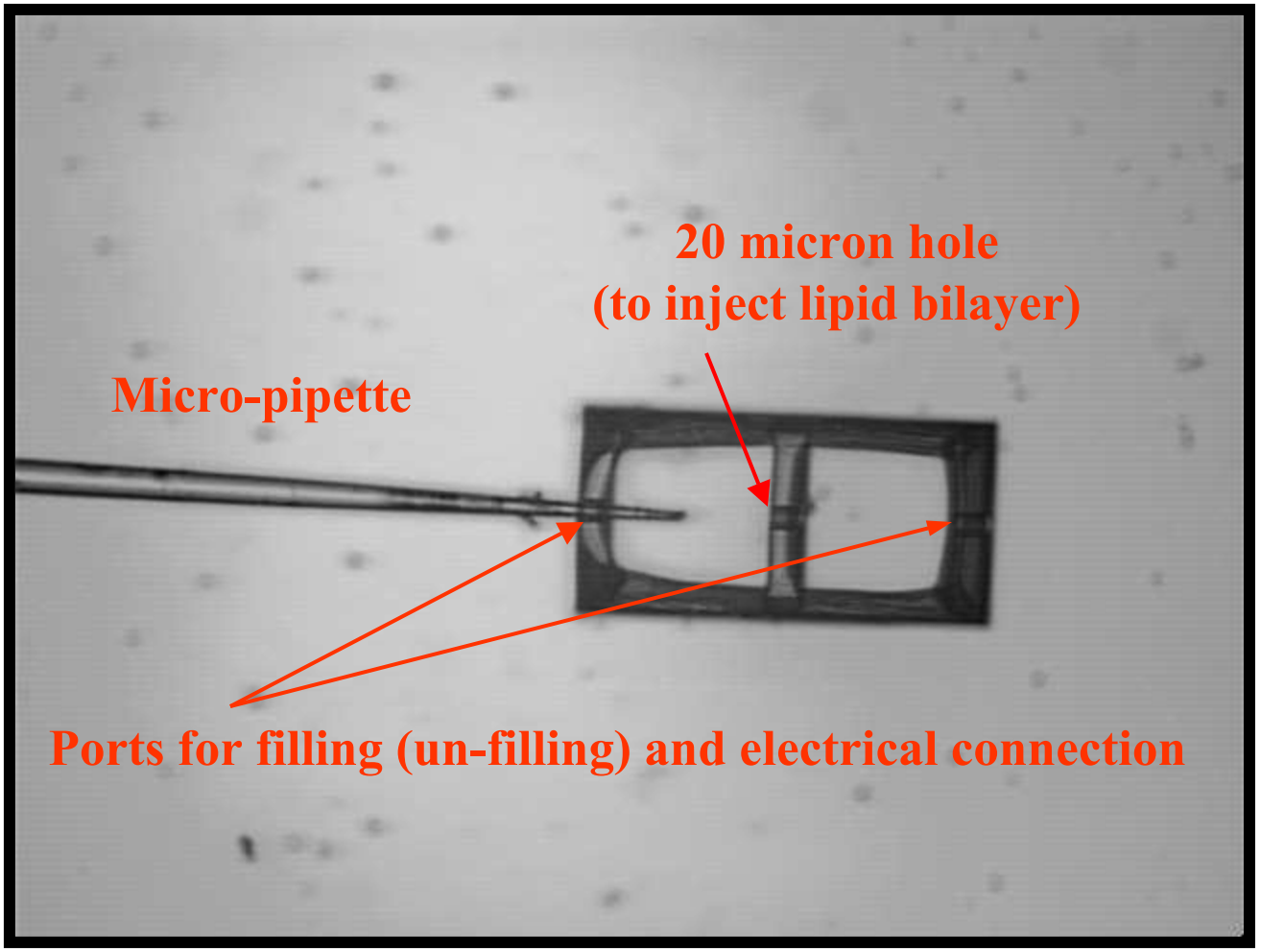
Water-filling of Micro-chambers



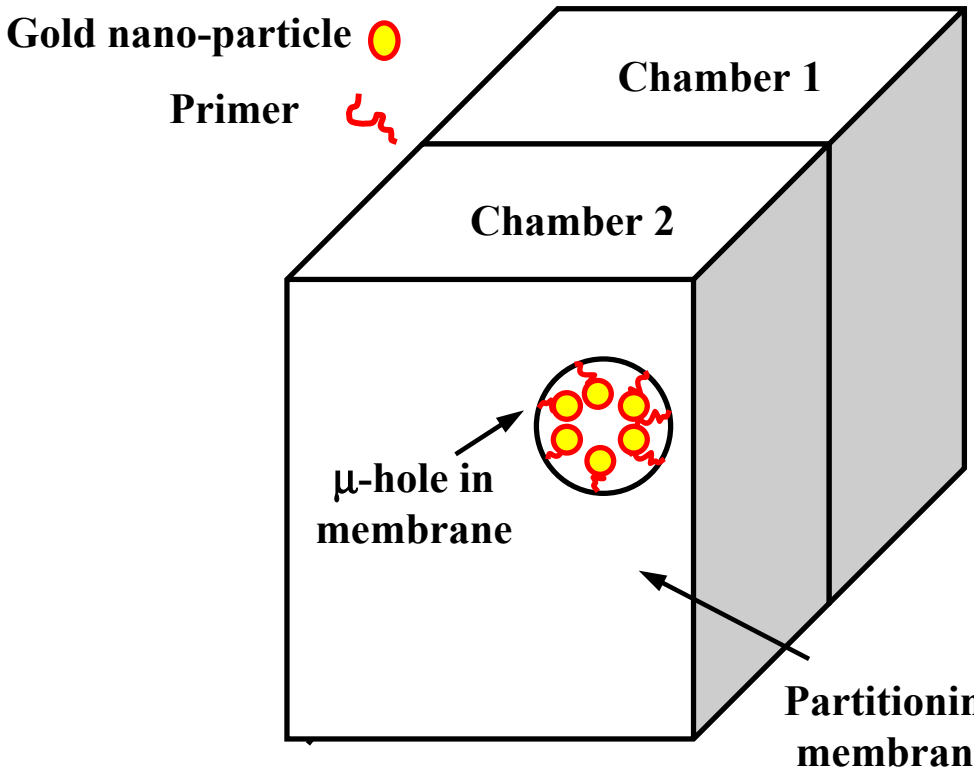
Filling the chambers of the structure shown in Fig. 5 with water using a 10μm diameter pipette. The top surface of the water is convex at *t* = 7.0s, flat at *t* = 16.5s, and concave at *t* = 25s.

Two-photon Lithography

Micro-well structure with input/output ports

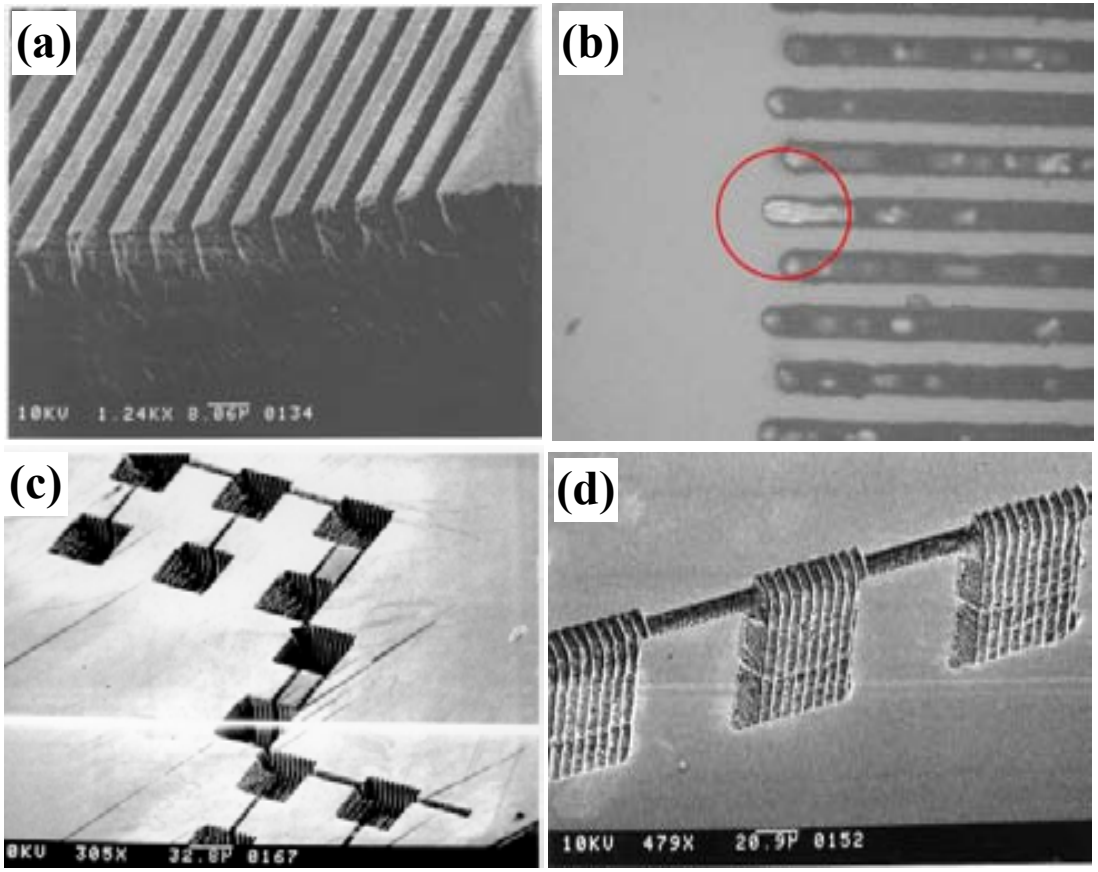


Nano-pore formation in Micro-chambers



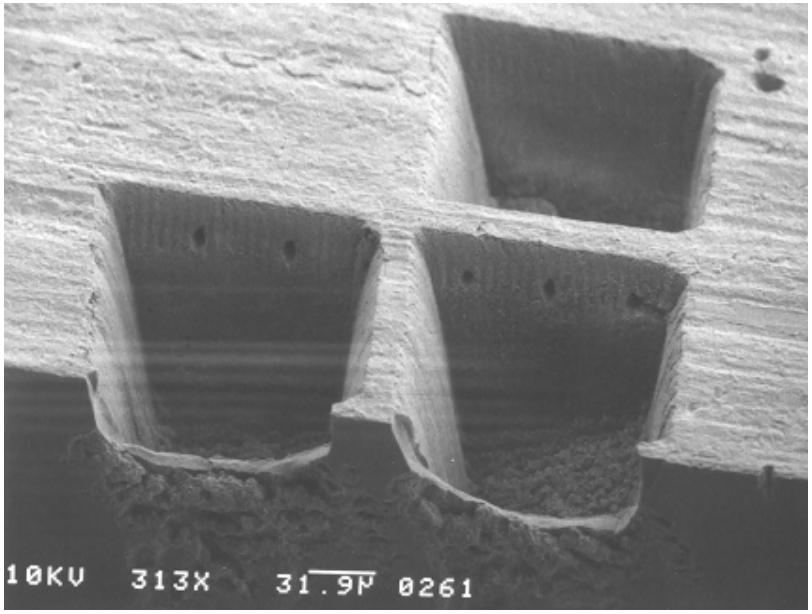
Method of forming a solid-state nano-pore in the micron-sized perforation within a partitioning wall.

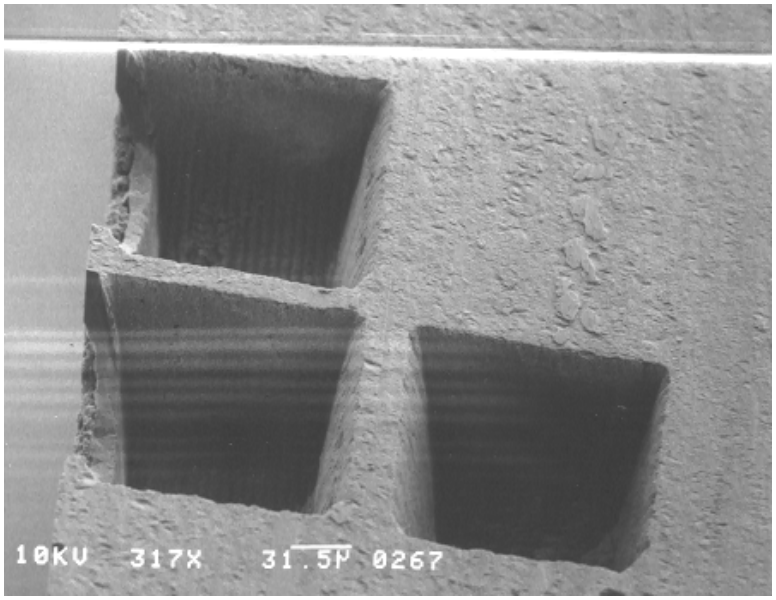
Femto-second Pulsed Laser Micro-machining



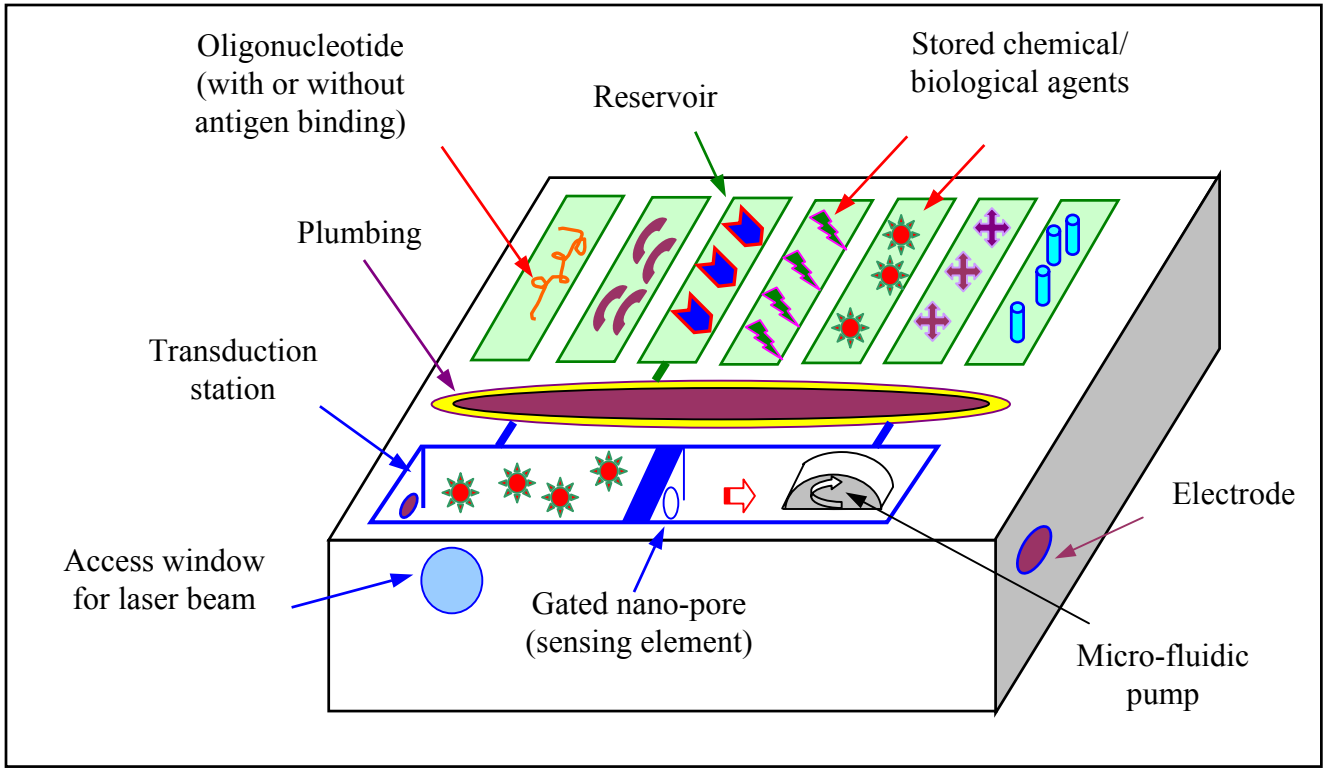
SEM images of μ-channels and μ-chambers carved on glass slides using femto-second pulsed laser micro-machining. (a) Grooves, ~20μm wide and 10μm deep. (b) Same grooves observed under optical microscope. (c) Square μ-chambers, each 50μm on the side and ~10μm deep, connected by μ-channels of the same depth. (d) Close-up of some of the micro-chambers.

Chambers cut with femto-second pulsed laser

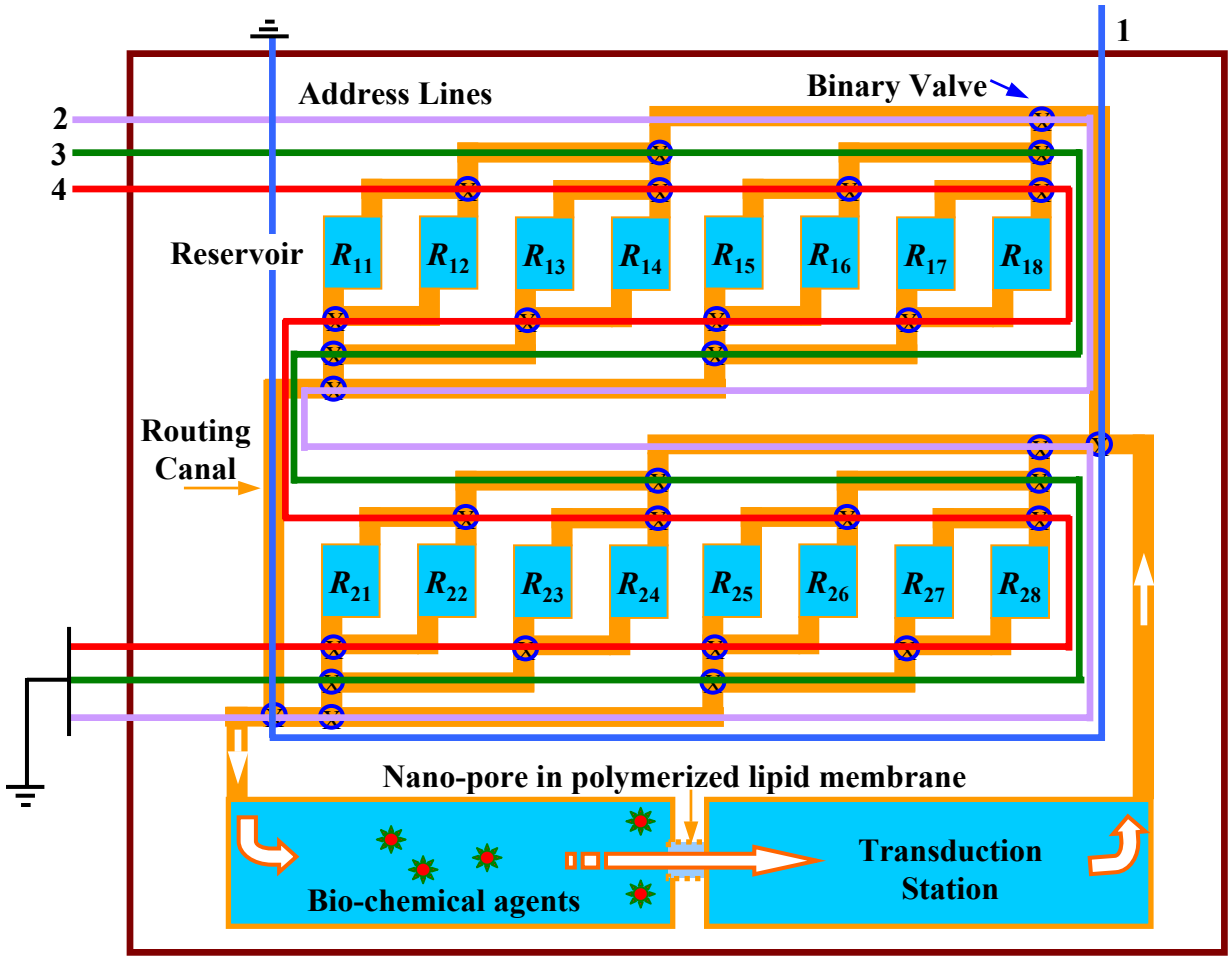


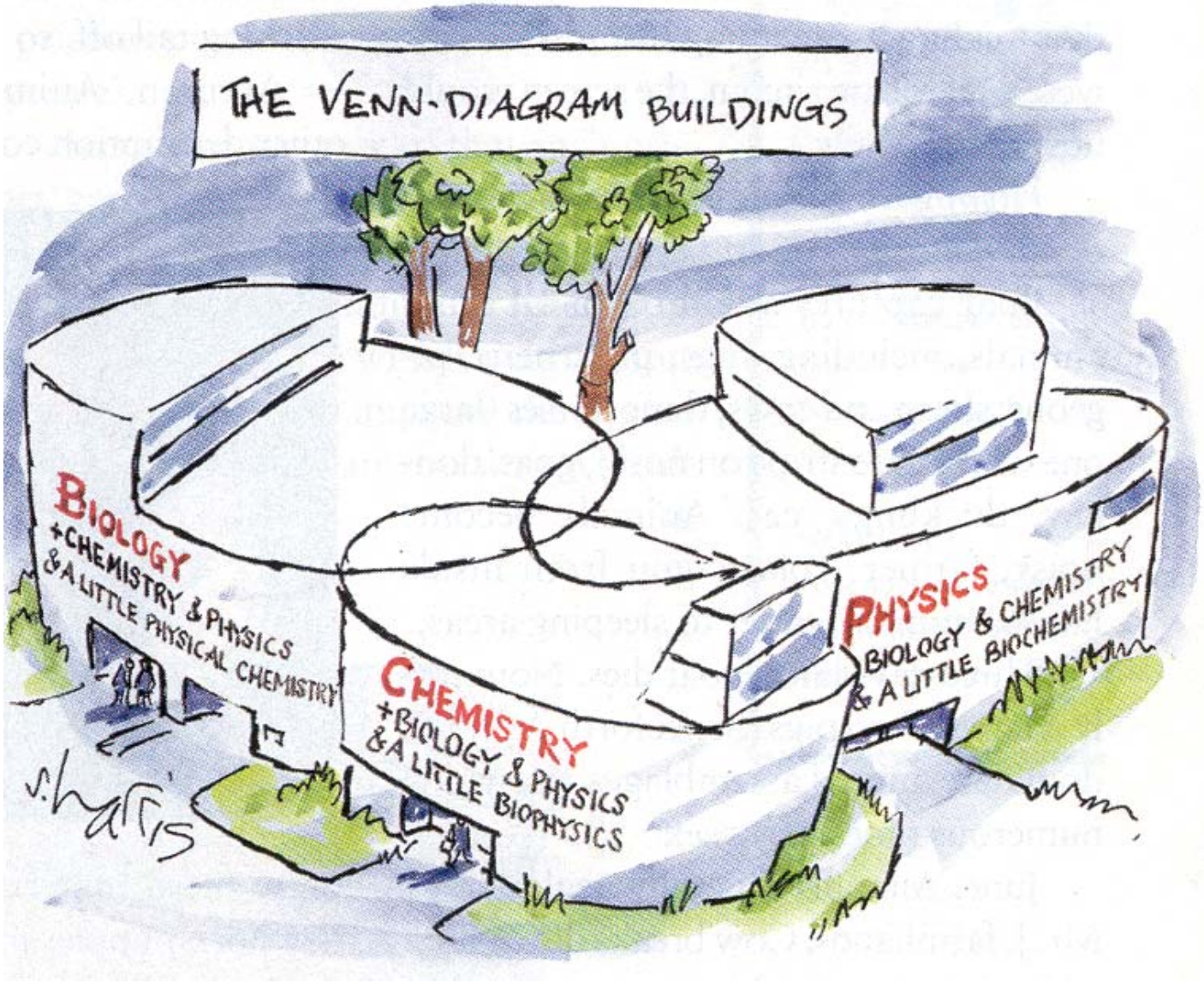


Chemical/Biological Sensor



Integrated Biochip for Chemical/Biological Sensing and Drug Delivery





Venn Diagram Building, *Physics today*