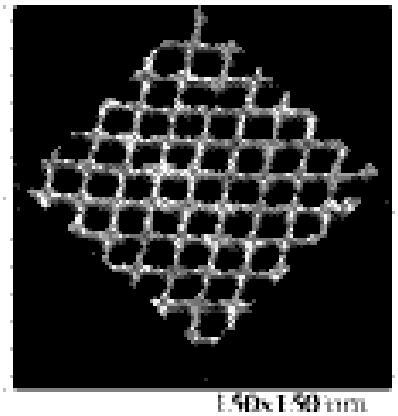
Fundamentals and Limits of DNA Nanotechnology

Jonas Berggren



MASTERS THESIS

CARRIED OUT AT

THE DEPARTMENT OF PHYSICS
DIVISION OF SOLID STATE PHYSICS

Front page: Self-assembled 2D DNA lattice. From (Yan 2003).

Acknowledgements

I thank my supervisor Jonas Tegenfeldt for sharing his great knowledge in the field of DNA research and the list of references that he keeps in his head.

Table of contents

1. Introduction

The discovery of the DNA molecule by Watson and Crick in 1953 (Watson 1953) is regarded as a milestone not only in biological history, but in scientific history overall. Their research was followed by the exploration of the protein synthesis in the 60's. In the 70's, spin-offs could be observed that did not directly belong to the fundamental research area. This decade saw the birth of the biotechnology industry.
In this report, two different areas will be studied: *Microbiological examination of DNA* and *DNA nanostructures as building blocks*.
It should be emphasized that these two research areas partly overlap. Nanoparticle probes have been produced using the self-assembly technique developed by Mirkin *et al. (Mirkin 1996; Cao 2002c).* Distance dependence of FRET is examined using DNA as a "ruler" (Deniz 1999). This technique may in turn be used to develop FRET-based probes for single-molecule spectroscopy of DNA itself or other biomolecules.
It might be possible to utilize the positioning techniques developed in nanofluidics to grow DNA lattices, although, of course, in this case there is no question of self-assembly. Lattices including 500 000 molecules have been fabricated using self-assembly, and to make these by individually positioning each molecule one by one would take several decades even if the process were entirely automated. However, the prospects are better for using techniques to build smaller lattices, nucleus lattices, on which self-assembly could continue.

Microbiological examination of DNA.

Biotechnology research involving DNA has been focused on sequencing and gene mapping. The chapter on "DNA sequencing and analysis" treats this. Examining the genetic network is an important task in order to reveal complex mechanisms leading to cancer and other genetic diseases. Such a network consists of regulatory interactions between genes and proteins; the gene may code for a regulatory protein that prevents (or allows) further reading of the gene itself or some other gene. Some examples of those "Genetic switches" will be given in that chapter. The mapping is made in co-operation between theorists who develop regulatory system models and experimental researchers. One task for the experimentalists is to follow the dynamics of a single gene or regulatory protein. This could more easily be done if the DNA molecule, which is normally twisted and looped, could be stretched out. A technique for doing this is presented in the chapter "Nanofluidics and DNA."

DNA nanostructures as building blocks.

As opposed to DNA molecular biology, the approach of using DNA nanostructures as building blocks focuses on using DNA instead of examining it. The idea is to let small DNA molecules self-assemble to build nanostructures. The DNA bases have selective pairing properties, where cytosine (C) exclusively binds to guanine (G), and adenine (A) to thymine (T). Theoretically, this can be used to make the DNA molecules attach to each other in an arbitrarily complex pattern or structure. By clever design, algorithmically assembled patterns and DNA computers could be achieved through self-assembly. These latter could be used for fast parallel computing. However, the practical usefulness is questionable, especially in the light of new emerging fields of quantum computing, which may prove more efficient. DNA self-assembly shows promising results when it comes to scaffolding other molecules and metal particles on a sub-20 nm-level.

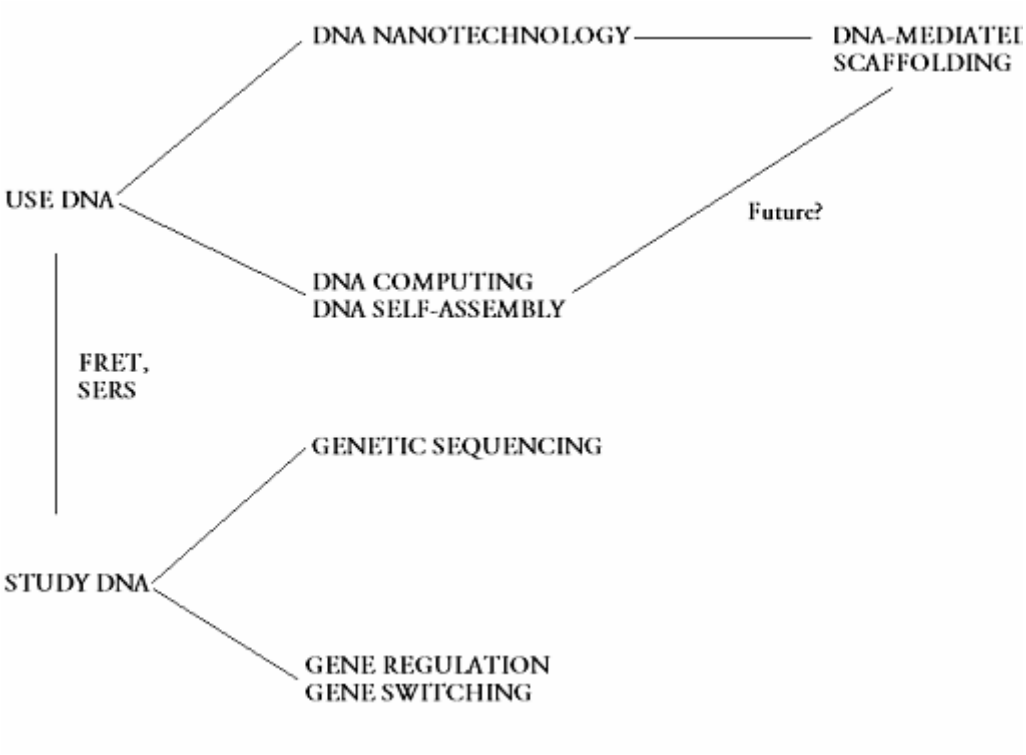


Fig 1.1. *A schematic overview of this report.*

2. The molecular biology of DNA[1]

2.1. DNA, basic concepts

DNA (deoxyribonucleic acid) is the most fundamental molecule of life. It is contained in almost every living cell (there are exceptions, such as red blood cells) and its structure has probably been unchanged for as long as life has existed on earth, about 3.5 billion years. Whereas in the more advanced eukaryotic cells (which is the type of cell in everything from yeast to humans) DNA is contained in a nucleus, this is not the case in prokaryotic, *i.e.* bacterial, cells. The function of DNA, though, is the same everywhere. It consists of a sugar-phosphate backbone on which four kinds of bases; adenine (A), guanine (G), thymine (T), or cytosine (C) are attached at a spacing of 0.34 nm in the physiologically most prevailing B-form. The order in which the bases appear in principle encodes the entire organism. Backbone and bases make a DNA strand. Further, the bases couple in pairs: Adenine to thymine and guanine to cytosine. The natural structure of DNA is the double helix, in which two DNA strands with complementary base chains wind around each other. Each strand has a 3'-end and a 5'-end and the two strands must couple such that the 3'-end of the first meets the 5'-end of the second and *vice versa*. The "wavelength" of the helical twist is 3.4 nm, which corresponds to 10 residues. The diameter of the molecule is 2 nm.

Figure 2.1: *The schematic structure of the DNA molecule.*

2.2. Packaging DNA

DNA in eukaryotic cells is contained in chromosomes, of which there is a diploid set of 2*23 in each human cell. Every chromosome contains a single linear molecule of DNA. The length of the human genome, *i.e.* the content of all 23 chromosomes taken together, is about 1 m, made up of $3.9*10^9$ base pairs. For comparison, in yeast there are about $1.2*10^7$ base pairs and 2*16 chromosomes.
DNA in chromosomes is wound up by histones, which are small cylindrical proteins specifically designed for packaging. This leads to a very high genetic information density, as DNA of 200 base pairs wounds around the cylinder-shaped histone of diameter 11 nm and height 6 nm. In the language of a computer scientist, 8 binary bits corresponds to 1 byte. As the DNA bases are 4 (C, G, A or T instead of 0 or 1), DNA requires half the amount of base

[1] This chapter relies primary on a few basic text books in biochemistry and genetics:

Lewin, B. (2000). Genes VII. New York, Oxford University Press.

Stryer, L. (1995). Biochemistry. New York, Freeman, Cop.

Voet, D. (1999). Fundamentals of Biochemistry. New York, John Wiley & Sons, Inc.

pairs, *i.e.* 4, instead of 8 binary bits, to make one "genetic byte". This makes 50 "genetic bytes" on every histone. The volume of a histone is $V=(11/2)^2*\pi*5$ nm^3~100 nm^3, which means that there are on the order of 10 Pb/mm^3 (Petabyte, 10^{15} bytes!) . For comparison, an ordinary DVD disk (4.7 GB) contains about 1 Mb/mm^3.

2.3. Protein synthesis, genes

The fundamental task of DNA is to code for proteins, the building blocks of life. The protein synthesis as invented by nature is undoubtedly a far more sophisticated nanomechanical system than any so far created by humans. The crucial role is played by RNA, of which there are three classes: messenger RNA (mRNA), transfer RNA (tRNA), and ribosomal RNA (rRNA).

A *gene* is a sequence of DNA coding for one or more proteins. It consists of several exons and introns. The introns are believed to have catalytic and other functions. The exons contain genetic information that code for the proteins.

The formation of mRNA from DNA is known as transcription. It is a reaction catalyzed by RNA polymerase, which is an enzyme. The polymerase recognizes a starting point for the transcription, an initiation site, and starts synthesizing RNA while it decodes the DNA from initiation to termination sites. This gives an RNA transcript, the pre-mRNA, which is identical to one of the strands in the DNA sequence and complementary to the other, with one exception: in RNA thymine (T) is replaced by uracil (U). In the next step, the intron sections of pre-mRNA are removed and the exon parts are ligated. This part of the procedure is called splicing, and allows a certain shuffling among the exons. Thus it is possible for several different proteins to be the outcome of transcription of the same original DNA sequence. The ligated exon parts of the pre-mRNA form the actual mRNA. The genome contained in mRNA is called the structural gene.

The next step is translation of the genetic information in the mRNA into the building of peptide chains of amino acids that eventually will fold into proteins. Every triplet of bases, every codon, of the mRNA corresponds to one amino acid, and if this correspondence could be reversed, there would thus be $4^3=64$ amino acids. However, due to degeneracy and other functions, there are only 20. Several codons thus correspond to the same amino acid. This actually works as a "security system". When the codons are read, errors may sometimes occur, particularly for the last base. That is why, in many cases, the last base of the codon is not of any significance.

The ribosome provides the environment for reading the codons as it moves along the mRNA. Responsible for building the peptide chains as well as for the reading itself is the tRNA.

The ribosome has two sites for binding tRNA; the A site (or acceptor site) and the P site (the donor site). The two sites are reserved for two different types of tRNA, peptidyl-tRNA at the P site and aminoacyl-tRNA at the A site, respectively. The former holds the successively formed chain of amino acids, the peptide chain, while the latter "reads" the mRNA by base pairing and carries the right amino acid, see fig 2.2.

Not every region on the mRNA will be translated for amino acids. Because of this there has to be initiation and termination sites for the procedure as well. The codons AUG, and less often, GUG or UUG, have shown to be present at the ribosomal initiation site (although these code for amino acids as well). Further, less than 10 bases upstream of this site, the sequence AGGAGG (The Shrine-Dalgarno sequence) is found. These two features are significant for an initiation site.

The termination site is always recognized by one of the codons UAA, UAG, and UGA (TAA, TAG, and TGA for DNA). These are included in the standard genetic code, fig 2.2, and do not code for amino acids.

Even before the ribosome reaches its termination site, the amino acids in the peptide chain start to fold into proteins. This is an area of research in its own right.

	U	**C**	**A**	**G**	
U	UUU **Phenylalanine** (Phe)	UCU **Serine** (Ser)	UAU **Tyrosine** (Tyr)	UGU **Cysteine** (Cys)	**U**
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	**C**
	UUA **Leucine** (Leu)	UCA Ser	UAA **STOP**	UGA **STOP**	**A**
	UUG Leu	UCG Ser	UAG **STOP**	UGG **Tryptophan** (Trp)	**G**
C	CUU **Leucine** (Leu)	CCU **Proline** (Pro)	CAU **Histidine** (His)	CGU **Arginine** (Arg)	**U**
	CUC Leu	CCC Pro	CAC His	CGC Arg	**C**
	CUA Leu	CCA Pro	CAA **Glutamine** (Gln)	CGA Arg	**A**
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	**G**
A	AUU **Isoleucine** (Ile)	ACU **Threonine** (Thr)	AAU **Asparagine** (Asn)	AGU **Serine** (Ser)	**U**
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	**C**
	AUA Ile	ACA Thr	AAA **Lysine** (Lys)	AGA **Arginine** (Arg)	**A**
	AUG **Methionine** (Met) or **START**	ACG Thr	AAG Lys	AGG Arg	**G**
G	GUU **Valine** Val	GCU **Alanine** (Ala)	GAU **Aspartic acid** (Asp)	GGU **Glycine** (Gly)	**U**
	GUC (Val)	GCC Ala	GAC Asp	GGC Gly	**C**
	GUA Val	GCA Ala	GAA **Glutamic acid** (Glu)	GGA Gly	**A**
	GUG Val	GCG Ala	GAG Glu	GGG Gly	**G**

Fig 2.2. *The standard genetic code; the codons and their amino acids. Some codons code for start and stop. In many cases the third base is of no significance. (From web page: http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/C/Codons.html)*

2.4. Genes can be turned on and off

It is obvious that in a complex biological system inside or outside a living cell, not every protein is needed at every time. It would be a waste of resources or even harmful if proteins were made of which the organism had no use. That is why on-off mechanisms operate in the cell in a complex regulatory manner. *Regulator genes* are a class of genes that code for proteins called *transcription factors* whose only task is to regulate other genes in the transcription phase. It is done by the transcription factor attaching itself to the initiation site thus either promoting or preventing transcription of the gene. The transcription factor can be regarded as a kind of signal, which turns the gene on and off. This signal is in turn triggered by another, coming from outside the cell, which may be a reaction to macroscopic changes in temperature, chemical environment, concentrations of certain compounds such as signalling molecules, energy and water supply. There will be examples in the chapter "*Genetic switches"* that two transcription factors may co-operate; giving rise to what is similar to a logical gate.
Since the regulator genes themselves could be turned on and off by other transcription factors, many genes are interconnected to form an intricate genetic network.

3. DNA sequencing and analysis

Several severe diseases have their origin in errors that may occur in the genetic network, caused by *e.g.* mutations or viruses. Cancer is a classical example where genes, which should normally be in an off-mode too often or always are turned on (or *vice versa)*. This is an important reason for studying the behaviour of the transcription factors on genes. It could be done by attaching fluorescent tags to the factors, which can be studied by microscope at the initiation sites of the genes.
Fluorescent tags may also be used in order to find *methylation sites* of the genes (Anderson 2004). Methylation is a process controlled by enzymes that sometimes attach CH_3 to the C (cytosine) base, but only if this is followed by G (guanine). These sites are called CpG sites where "p" represents the normal phosphoric bond in the strands[2]. Methylation of a transcription site prevents transcription factors from attaching.
It has been proved that the occurrence of CpG sites on the human genome is far less common than it would be had it occurred randomly. This is known as CG suppression, and is due to the fact that these sequences mutate very easily because of methylation; since they cannot be transcribed, they are useless genetic material. Anyhow, in mammal genome, there has been evidence of large CG sequences in very small distances along the genes, so-called CpG islands. See fig 3.1. These play a crucial role in gene silencing, which means that the gene is permanently turned to off mode. Although normally *not* methylated, they may under certain conditions be.

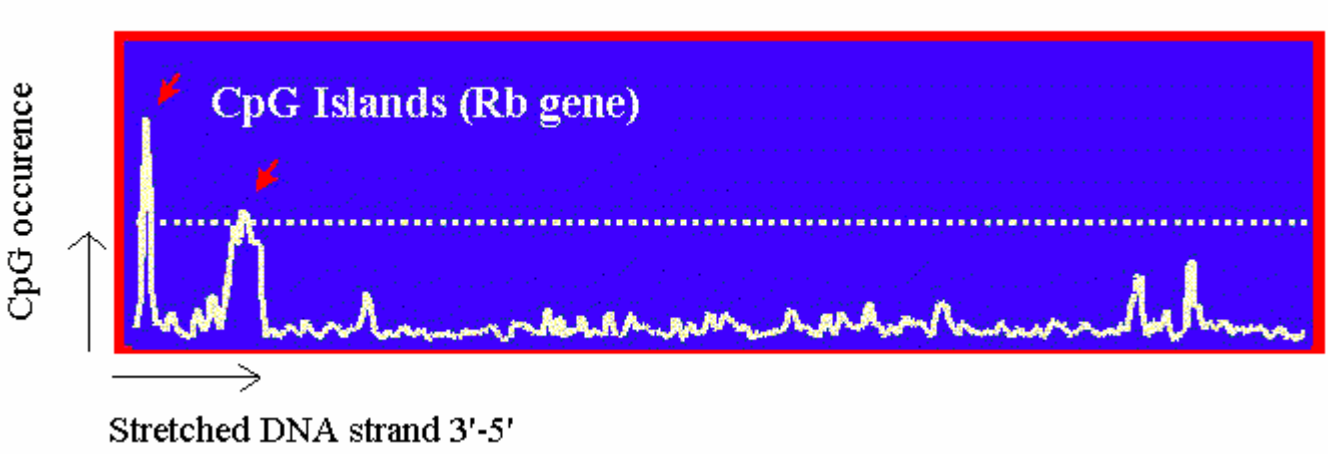


Figure 3.1 *From (Anderson 2004).*

On the contrary, so-called imprinted genes, where either parental gene is expressed and the other repressed, are always methylated. Classical examples of imprinted genes from the human genome are *e.g.* genes for blood group and eye color. If error occurs and a normally repressed gene starts to reproduce in an uncontrolled way, it may cause cancer. One concrete example of this concerning the human genome is Wilms' tumor, where both copies of the gene *IFM2* are expressed, but only the father's should be.
The importance of studying CpG islands is evident. From a biophysical point of view, this may be done by footprinting methods, which involves massive replication of genes. Fom our point of view, however, direct examination of single genes through attachment of fluorescent tags is more important.

[2] Not to be confused with C–G as a complimentary bond between two strands.

3.1. The physics behind fluorescent tags

The phenomenon of fluorescence is due to the fact that a material absorbs light of a certain wavelength, exciting electrons to higher energy states, which in their turn deexcite after a certain period of time, possibly to other states, which may render the wavelength of the emitted light different from the absorbed. Everyday examples of this are ordinary light switches at night. The deexcitation makes the material gain energy, which is often transformed into vibration. Fluorescence has its name from the mineral flourspar, CaF_2, which exhibits these properties. As far as fluorescent tags are concerned, however, Ethidium Bromide or Green Fluorescent Protein (GFP) (Tsien 1998) is often used. When this material is illuminated by UV-light, it will start to emit photons from the visible range, in this case orange or green light.

3.2. Single nucleotide polymorphisms (SNPs)

In recent years, much attention has been drawn towards research on Single Nucleotide Polymorphisms (SNPs, pronounced and sometimes spelled "snips"). (Chakravarti 2001). As can be concluded from the name, SNPs are single base-pair changes from one base to another that occur particularly easy somewhere along a DNA sequence, due to mutations. The fact that these SNPs are very commonly occurring makes them particularly interesting; in the human genome, there are about $1.4*10^6$ SNPs. There is thus in the order of 1 SNP to every 1000 base pairs. SNPs can occasionally alter protein production in a manner that causes severe diseases. Sickle cell anemia is a concrete example of this. The oxygen-transporting protein haemoglobin, which should be smooth, under these conditions takes the shape of a sticky rod, partly destroying the blood cells carrying it. Searching for this type of SNPs is obviously of high priority, but fortunately, most SNPs cause no damage. These harmless SNPs may however still become useful as genetic "beacons". Since they are so common, there is high probability that there are SNPs in the proximity of a gene of interest. It has been shown from SNP mappings that groups of people with a particular genetic disease or mutation have more SNPs in common near the modified gene than those without. Although not the actual cause of the disease, the SNPs are inherited together with the modified gene. This means that SNPs are fingerprints to look for when localizing genes on a chromosome. Apart from its obvious applications in genetics and medicine, SNP mapping also applies to the field of *molecular anthropology*. It may be used for tracking colonialization of Polynesia and the Americas by Asian tribes thousands of years ago, or to find out whether the modern human race is related to the Neanderthals (Stoneking 2001).

3.3.Haplotypes

Haplotypes are closely related to SNPs. A haplotype is simply a combination of SNPs in a gene. Consider, for example a particular SNP base site with C or T and another base site with A or T. If there are no other SNPs on this gene, it may have a maximum of 2^2 haplotypes. A gene containing 10 SNPs would thus have at most 2^{10}=1024 haplotypes. It has been found, though, that the actual number of haplotypes is much lower. A gene containing about 10 SNPs actually has no more than a dozen.
It is easily realized that a haplotype, a combination of SNPs, can alter the function of a gene in a more fundamental manner than is the case with a single SNP.

3.4. Detecting SNPs, Haplotypes and other interesting gene sequences

Several methods have been developed for detection. Some use so-called FISH (fluorescence *in situ* hybridization) methods; *i.e.* fluorescent tags are attached to small sequences of DNA complimentary to target sequences. These could be haplotypes, the vicinity of a SNP, or entire genes.

3.4.1. Padlock probes

The padlock probe uses such a hybridization technique (Antson 2001; Landegren 2004). It consists of single-stranded DNA at a length of about 90 bases. The bases near the 3'- and 5'-ends are complimentary to the target sequence, but the bases in the middle are not. To these, fluorescent molecules are attached. When reaching the target sequence, the probe forms a loop, and the fluorescent molecules reveal the site of the sequence. The 3'- and 5'-ends of the probe DNA will meet and be ligated. The probe DNA will form a loop. See fig. 3.2. The mechanism is very sensitive; even if there were one single base not complimentary, the two ends would not legate. It also sits strongly in place. Padlock probes are thus suitable for detecting SNPs.

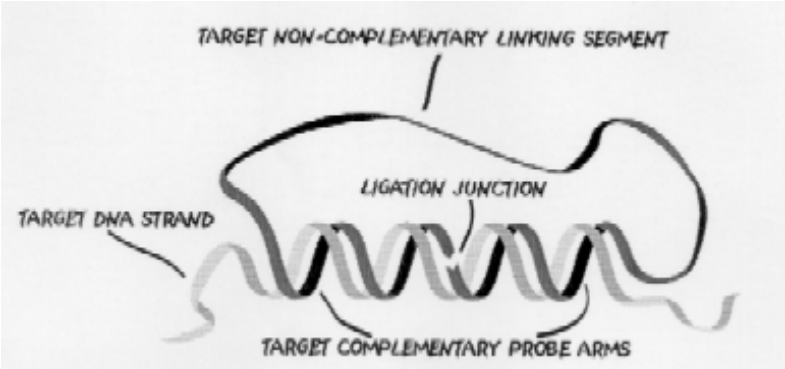


Figure 3.2. *A schematic view of the padlock probe. Adopted from (Antson 2001).*

3.4.2. Molecular beacons

Molecular beacons are similar to padlock probes but more advanced since they only are fluorescent when attached to their target sequences (Kramer 2004). This is possible due to an ingenious combination where on one hand thermodynamic properties and on the other co-operation between a fluorophore and a quencher are exploited. Molecular beacons consist of a "DNA loop" where the center of the beacon attaches to the target and the ends are used for fluorescence, contrary to a padlock probe. To the two ends, a fluorophore and a quencher are attached. When these two are near each other they transiently share electrons by the FRET mechanism, which will be described in later chapters. The quencher will inhibit fluorescence of a certain color, which has been pre-defined for detection, *e.g.* green. When the two tags move apart, the fluorescent tag will start emitting light. The molecular beacon is constructed in such a way that different molecular forms are thermodynamically preferred depending on whether it is attached to the target or not. See fig 3.3.

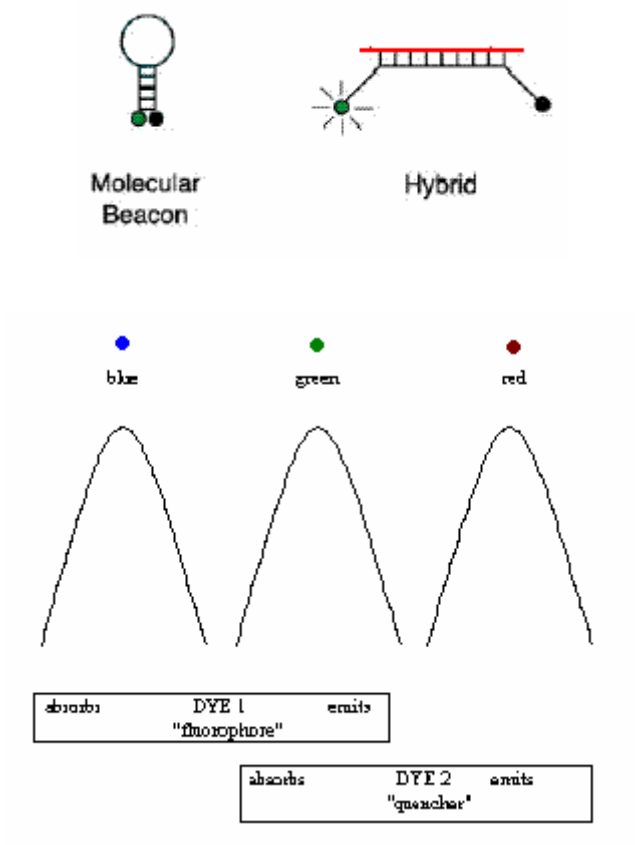


Figure 3.3. *(Above) The two thermodynamical states of a molecular beacon. When no target is present, base pairs at the end of a DNA sequence will pair, forming a loop of the DNA complimentary to the target. The quencher will be very close to the fluorophore, preventing it from emitting light. The molecular beacon is thermodynamically designed in such a way that in the presence of a target sequence (red) the binding of the beacon loop to the target will be preferred to the former binding, thus breaking the base pair bonds at the end. This in turn removes the fluorophore from the quencher, allowing it to emit light. Adopted from (Kramer 2004).*
(Below) The quenching mechanism. When no target is recognized, the two dyes will be close enough to interact. When lit upon by blue light (the colors are naturally arbitrarily chosen), dye 1 is constructed to emit green light if alone. However, in the presence of a quencher, which is constructed to absorb green light and emit red, the green light (which in this case is the significant color for detection) will be quenched.

The switching between the two states allows very accurate target identification. If only a single nucleotide differs, the molecular beacon discriminates the target sequence and remains in the looped form, hence does not emit light. Thus, they may be used when detecting SNPs.

Because molecular beacons are able to turn fluorescence on and off, they have been referred to as “molecular switches”. The fluorescence-quenching properties could prove to be of use in molecular electronics as well.

3.4.4. Fiber-FISH

The fiber-FISH concept is developed from the fact that DNA can be stretched out for linear analysis on surfaces or in nanochannels by nanofluidic methods described in another chapter. It simply means using FISH methods on these linear strands of DNA. The word “fiber” relates to a linear chromosome.

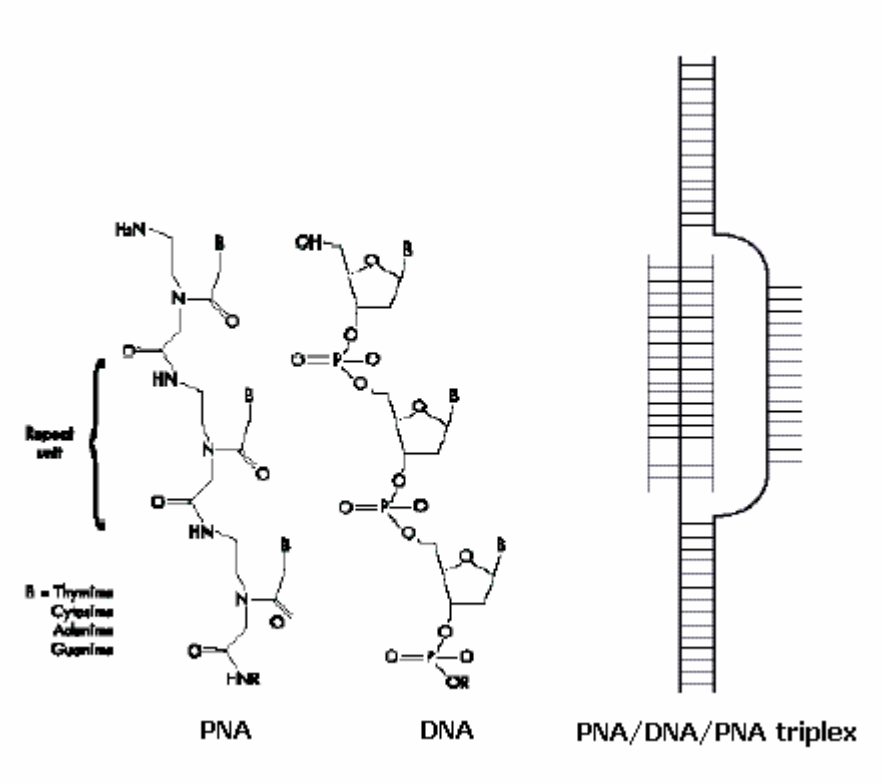


Figure 3.4. *The structure of PNA. From (Nielsen 1991).*

3.5. Peptide Nucleic Acids (PNA)

Every detection and analysis method described above can be utilized with PNA (Peptide Nucleic Acid) instead of DNA as the probe molecule. PNA is an artificially produced molecule that resembles DNA in many ways but has certain advantages (Nielsen 1991; Eurogentec 2004):

- As opposed to DNA, which is charged, PNA is uncharged. The consequences are two: 1) After hybridization, binding between DNA/PNA or PNA/PNA strands become stronger than between DNA/DNA strands, du to lack of electrostatic repulsion. 2) The specificity in binding to complimentary DNA is stronger, making PNA a very suitable detection probe for SNPs.

- PNA and DNA can together form triplexes, see fig 3.4. Forming such a probe molecule, more tags can be attached. Triplexes could possibly also be used when building lattices or 3D-objects.

- Unlike ordinary DNA, PNA is unaffected by enzymes. This allows for *in vivo* investigations.

3.6. Surface-enhanced Raman scattering and its applications to DNA analysis

In this chapter the phenomenon of Raman scattering will be explored, followed by an investigation how the scattering signal might be improved near certain surfaces (Surface-enhanced Raman scattering, SERS). This proves helpful in single-molecule spectroscopy of DNA and other biomolecules.

3.6.1. Raman scattering

Raman scattering stems from interaction between an external electromagnetic field and the vibrational energy levels of a molecule (Kaiser 2004). Although an overwhelming number of the incoming photons are elastically scattered by the molecule, a few (about 1 in 10^7) interact with the vibrational modes of the molecule. This will either increase on decrease their initial energy, giving rise to a Stokes shift (in the case where $E_{Final}<E_{Initial}$) or Anti-Stokes shift (where $E_{Final}>E_{Initial}$), see fig1. The process works as follows:

An external electromagnetic field **E** affects a molecule with polarizability α in that it induces a dipole moment **P** according to:

$$\mathbf{P}=\alpha\mathbf{E}$$

The formula obviously means that the electron cloud of a molecule with high α will be highly sensitive to external fields in that it creates a large dipole, which is the source of radiation of photons with the same energy as the incoming. This is Rayleigh (elastic) scattering.

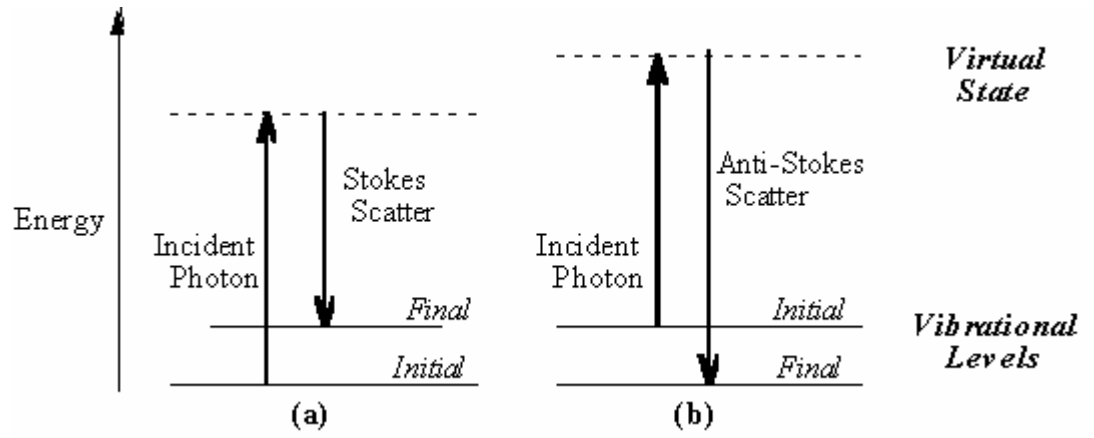


Figure 3.5. Raman scattering. From (Kaiser 2004)

However, molecules also vibrate. The vibration will itself cause a distortion of the electron cloud, which means a change in polarizability and dipole moment. This, in turn, means a slight shift in dipole radiation and an energy difference between the incoming and outgoing photons. Raman scattering is thus inelastic. The energy is either given to the molecule as heat (Stokes shift) or taken from the vibrational energy (Anti-Stokes shift), but since the amount is very small on the molecular level, it hardly affects the target at all.

Because the Raman scattered photons interact thoroughly with the molecule, they contain very valuable information for single molecule spectroscopy. A decisive obstacle, though, is the fact that they are extremely few; even if a molecule were inserted in a laser-created field with the large intensity of $1W/\mu m^2$ there would only be one Raman scattered photon with an energy shift large enough to detect every few hours. So, for this phenomenon to be useful for single molecule spectroscopy, it has to be considerably improved.
Surface-enhanced Raman scattering is a means of strongly enhancing the electromagnetic field recognized by the molecule and thus its vibrational modes.

3.6.2. Surface-enhanced Raman scattering (SERS)

If the target molecule is placed near the conducting surface of certain metals, Raman scattering will increase considerably (Campion 1998). The Surface-enhanced Raman scattering is due to electromagnetic interaction between the molecule and the metal surface:
The electron gas of a conductor may get excited in an electromagnetic field, which makes the electrons oscillate collectively. This is known as plasma oscillation and its corresponding quantified entity is the plasmon. It induces a new electromagnetic field related to that of the excitation. Due to properties of the metal, there will be a resonance in the induced field at a certain frequency, making it very large. The incoming excitation field is enhanced by 10^3-10^6. (Kaiser 2004).
Now consider a molecule in the proximity of the surface. The point is that not only will the incoming field be enhanced by the plasmons, but the Raman scattered field from the molecule as well. Vibrational modes normal to the surface are considerably amplified, and so is consequently the polarizability and Stokes' shift. The SERS effect has been compared to that of a radio antenna, which by the same resonance principle amplifies radio signals.

The initial Raman scattering process, which is quite inefficient for single molecule (biomolecular) spectroscopy may thus become useful when improved using SERS. The fabrication of SERS-probes for single molecular spectroscopy is a vivid area of research and experiments including DNA mapping have been conducted.

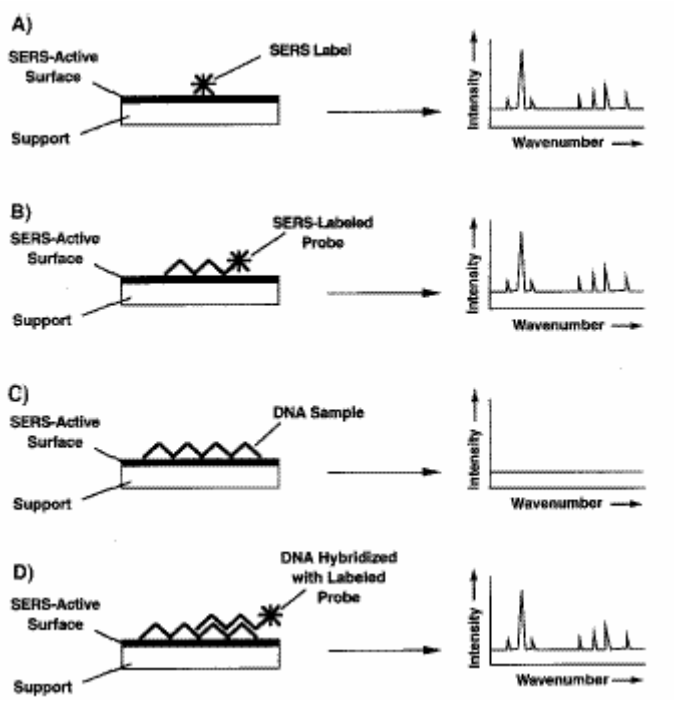


Figure 3.7. *The principle of DNA sequencing with SERS probes. From (Vo-Dinh 1999)*

3.6.3. DNA and Gene analysis using SERS

SERS probes for DNA mapping and gene sequencing have been investigated (Vo-Dinh 1999). The idea is described in fig. 3.7. A SERS-active molecule (*i.e.* a good Raman scatterer) is attached to a probe string of DNA. The DNA sample to be analyzed is stretched out on a SERS-active surface. In case it contains a sequence complimentary to that of the DNA probe this latter will attach, bringing the SERS-molecule into proximity of the SERS-active surface and SERS light will be emitted when lit upon by a laser. Up to, possibly, on the order of a hundred SERS labels at different wavelengths may therefore be used simultaneously since there is a significantly smaller risk of spectral overlapping compared to when fluorescent labels are used. For comparison, the half-width of a fluorescent peak measures 50-60 nm whereas that of a SERS peak measures less than 0.5 nm (Vo-Dinh 1999). The SERS probe method was involved when part of the RNA sequence of the HIV virus was localized (Isola 1998).
SERS-probe based DNA analysis has also been undertaken by the Mirkin group. See pp. 43-45.

4. FRET measurements using DNA as a "ruler"

FRET (Fluorescence Resonance Energy Transfer) is an electromagnetic phenomenon that has been used for spectroscopic investigations of single molecules, even in real-time, as opposed to averaged behaviour. In the FRET case, the fluorescence of two tags interacts constructively if they are close enough and their fluorescence spectra overlap. This results in a non-radiative energy transfer from one tag to the other. FRET tags hitherto explored are all organic molecules, but the use of quantum dots has been proposed (Guha 2003). The FRET transfer efficiency was predicted by Förster to be:

$$E = \frac{1}{1 + \left(\frac{R}{R_0}\right)^6} \qquad (4.1)$$

where R_0 is the so-called Förster distance, giving 50% transfer efficiency in the system:

$$R_0^6 = 8.785 \times 10^{-5} \left[m^2\right] \frac{\kappa^2 \phi J}{n^4} \qquad (4.2)$$

κ is a uniteless orientation factor between the donor and acceptor; it can be intuitively understood that FRET interaction is dependent on how the dye molecules are oriented towards each other. For organic molecules kappa is mostly set to 2/3.

ϕ is the uniteless quantum yield between the dyes, *i.e.* the magnitude of electromagnetic interaction between them.

J is the overlap integral between the fluorescence spectra of the molecules. It is schematically shown in fig. 4.1. Unit [m^4].

n is the refractive index of the medium.

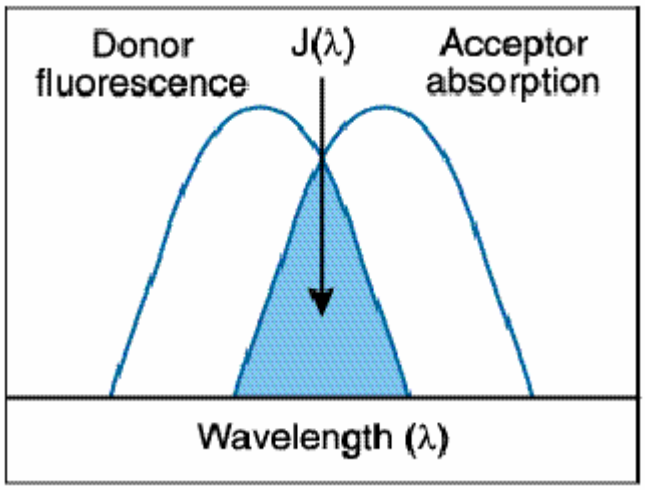


Figure 4.1: *Qualitative description of the parameter J. From (Guha 2003)*

The prediction (4.1) was verified by Deniz *et al*. in 1999 in an ingenious manner using DNA as a "ruler" (Deniz 1999). FRET-sensitive molecules may be attached to the bases of the DNA. As the distance between the base pairs of the DNA is exactly 0.34 nm, very accurate *efficiency vs. distance* measurements can be made within the order of persistence length (about 50 nm) of the DNA molecule. A donor (TMR) was attached to the end of the DNA and the acceptor was placed at different base pair distances downwards the DNA[3]. That double-stranded DNA is helix-shaped must be taken into account when calculating the distance between two sets of base pairs. The distance R in nm depends on the two terms

$$R_a = 0.34N + L \qquad (4.3)$$

and

$$R_b = a^2 + d^2 + 2ad\cos(\theta) \qquad (4.4)$$

where

$$\theta = (36\ N + \varphi) \qquad (4.5)$$

according to

$$R = \sqrt{R_a^2 + R_b^2} \qquad (4.6)$$

N is the number of base pairs, L and φ are the distances and angles between the dyes at the beginning. The distances of the acceptor and donor dyes to the axis of the DNA molecule, are labelled a and d, respectively. See fig. 4.2 (inlet). The transfer efficiency of the FRET interaction was measured for a number of distances according to fig. 4.2. The drawn curve is the Förster prediction.

The measurements are marred by errors due to noise and fluctuations and the FRET efficiency peaks will be broadened. For example, dye-pair measurement at 12 DNA base pair's distance (about 5 nm) will look as in figure 4.3 (left). If more than one acceptor is involved, the relative occurrence of each subpopulation may be calculated by comparing the areas occupied by each distribution, see again fig 4.3 (right). This could be used in relative concentration measurements of molecules or in examinations of the structure of a particular molecule.

[3] This was in practice achieved by preparing one DNA molecule for each measurement.

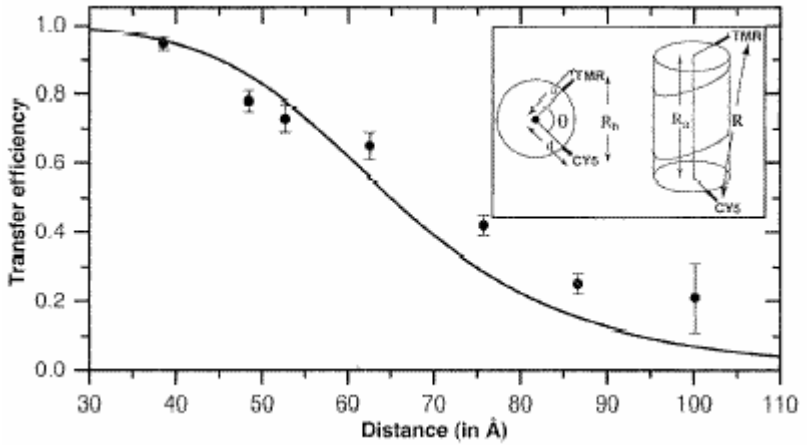


Figure 4.2: *Dots: Measurement of efficiency using DNA as a "ruler". Drawn curve: Efficiency as predicted by Förster theory. Inlet: Calculation of distance with the helical structure of DNA taken into account. From (Deniz 1999).*

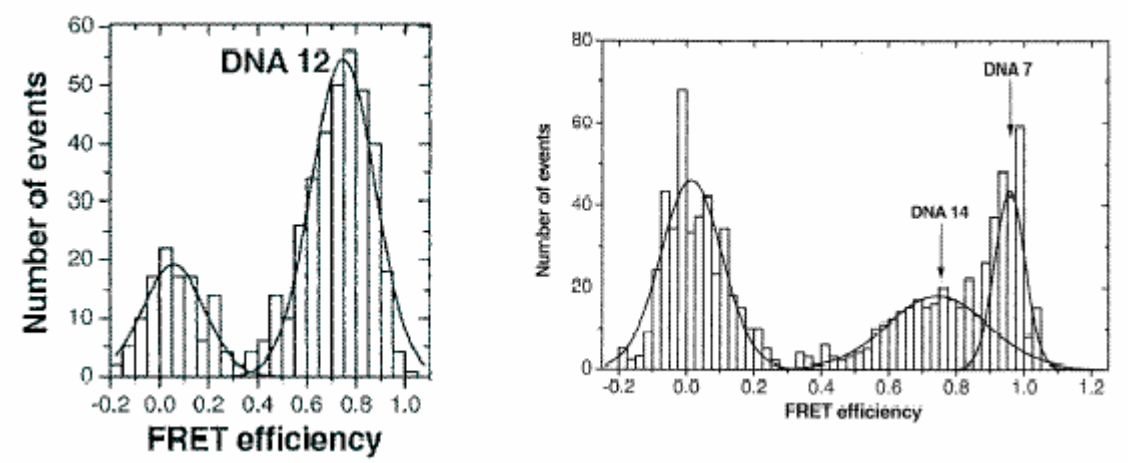


Figure 4.3 *(Left) The peak at 0.0 is a noise peak and the peak at 0.7 corresponds to the second measured value from left in fig 4.2. Ideally, this would have been singular but the broadening is due to noise and fluctuations in the donor-acceptor system. (Right) Two efficiency peaks corresponding to a donor and two acceptors in the system. The areas under each peak obviously correspond to the occurrence of each donor. If two molecules were dyed the relative concentration of each could have been calculated. Figures 4.2 and 4.3 from (Deniz 1999).*

4.1. FRET of DNA-based single molecular photonic wires.

Recently, FRET effects in chains of dye molecules, so-called single molecular photonic wires, have been investigated using DNA as a base (Heilemann 2004). The idea of the experiment was to attach five FRET-sensitive molecules at even spacing on a DNA strand, creating a "wire" of length 13.6 nm. It showed quite efficient fluorescent quantum yield through the wire, up to 90%.

4.2. FRET using quantum dots

The use of quantum dots instead of organic molecules as fluorescent material has been proposed (Guha 2003). Guha lists some possible advantages that these might have:

- High emission quantum yield. The electromagnetic interaction providing the FRET is larger between quantum dots. According to equations (4.1) and (4.2), this makes the Förster distance larger and the FRET more far-reaching.

- Sharp emission spectra.

- It is easier to take advantage of the different colors of the quantum dots, because of the sharp emission spectra.

The two latter features reduce the risk of two different FRETs overlapping each other. It opens the possibility of monitoring multiple FRET interactions simultaneously, using different colors. This argument is similar to the one presented in the SERS section, where up to 100 SERS probes could possibly be used in the investigation of a biomolecule.

4.3. Possible experiments using DNA as a base for FRET analysis

-A further step in the research of photonic wires would be the insertion of a photoswitch in a photonic wire as described above. Such switches have been accomplished (Irie 2002; Terazono 2004).

-No photonic wire using the FRET effect has been made of quantum dots. Experiments so far (Heilemann 2004) were conducted using organic flourophores.

-The behaviour of FRET when quantum dot photonic wires are attached to branched DNA or even more advanced structures has not been investigated. See fig 4.4.

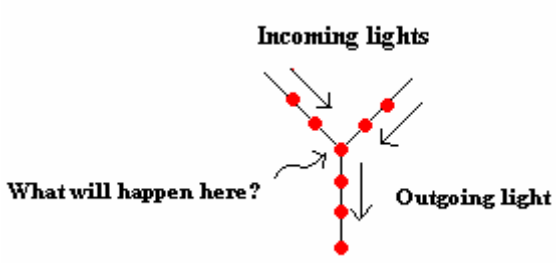


Figure 4.4: *Y-shaped (branched) DNA form the basis on which quantum dots can be attached. What will happen in the photonic wire at the point of interference and what will be the spectrum of the outgoing light as compared to the incoming? This might prove interesting when it comes to developing switches and gates.*

-The empirical value of the parameter κ in eq. 4.2 has to be investigated. If the average value is 2/3 for organic molecules, this is not the case for quantum dots. Investigations of κ could be part of an experiment that consists of repeating the investigations by Deniz *et al*., except using

quantum dots instead of organic molecules. Establishing the values of κ is crucial to all applications mentioned above involving quantum dots.
One long-term goal of FRET single molecular spectroscopy is obviously to fabricate probes for DNA analysis similar to those suggested in the SERS chapter. This is an example of the double usage that DNA might have in nano- and biophysics: Developing tools for analyzing DNA involves methods that use DNA in the other fashion; as building blocks.

5. Mechanical and electrochemical manipulation of DNA

5.1. Elongating coiled DNA through micro- and nanofluidics

As described in other chapters of this report, there are several reasons for analyzing very long strands of DNA. One is to localize the transcription sites of the genes, which is a prerequisite for developing methods to intervene in the protein synthesis process. Another is to recognize methylation sites. In order to do this, the DNA has to be elongated from its ordinary, high-entropy state where it is complexly coiled and looped (Cao 2002b). This has been achieved by letting DNA molecules flow through nanochannels fabricated by nanoimprinting lithography (Tegenfeldt 2004a; Tegenfeldt 2004b). Obviously, the size of the channels must be smaller than the persistence length of DNA, about 50 nm. Dimensions of 10×50 nm have been fabricated (Cao 2002a). A considerable problem was how to overcome the steep entropy barrier at the interface between the micro- and nanofluidic areas, see fig 5.1a. It means that the DNA molecules would not readily move into the channels but get stuck at the beginning of the channel region. However, if the molecule were forced to gradually stretch out already in the microfluidic region, this would reduce the slope (although, of course, not the total height) of the entropy barrier. See fig 5.1b.

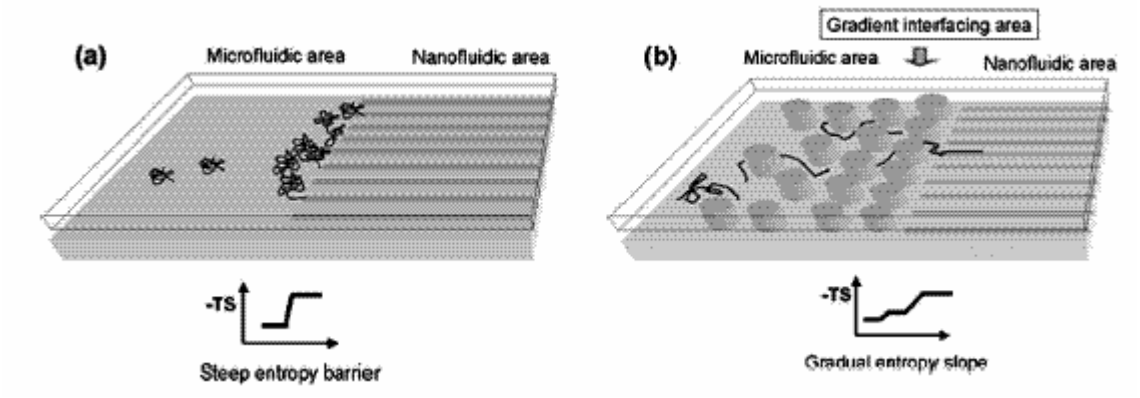


Figure 5.1: *a): Steep entropy barrier; the molecules become clogged at the border between the micro- and nanofluidic areas. b): Gradual entropy slope; the molecules are successively prolonged before they enter the nanochannel area. From (Cao 2002b).*

5.2. Dielectrophoresis

In general, dielectrophoresis (DEP) means achieving movement of polarized particles, which includes DNA, in inhomogeneous electric fields. The force affecting the molecule is:

$$F_d = -grad(U) = \frac{\alpha}{2} grad(E^2) \qquad (5.1)$$

Here, α is the polatizability of the molecule, and E is obviously the field strength. This is why the field is required to be inhomogeneous; in case $grad(E)=0$, $F_d=0$.
Since the particle is prepared to move in the direction of the field gradient, varying the field vector can control the movement (Chou 2002). DEP is also suitable for separation (fractionation) or concentration, as particles with different polarizabilities respond differently

to the field. In a potential slope, the most polarizable molecule will go into the trap, thereby separating it from the rest. Concentrating particles (*e.g.* for easier detection) can be achieved by using areas with different field strengths. The particles will prefer one of these areas.

5.3. A "DNA prism"

A "DNA prism" for separating DNA molecules of different size has been made (Huang 2002). The idea is to lead the molecules into an array and expose the structure to an electric field whose direction alternates between two directions 120° apart.

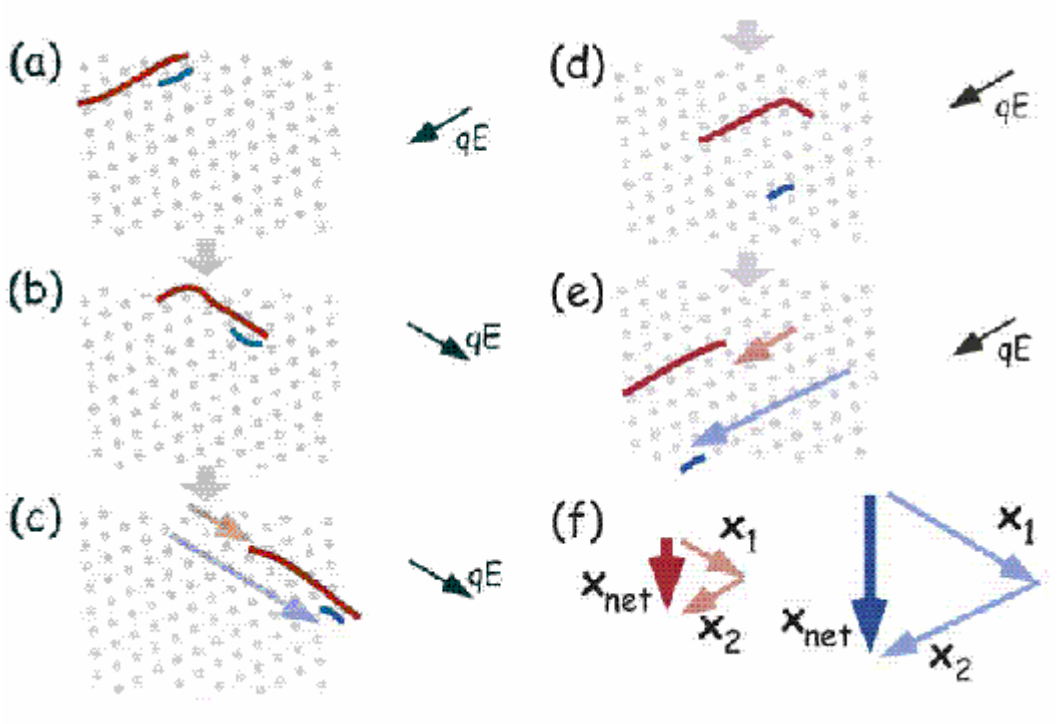


Figure 5.2 *The DNA prism. From (Tegenfeldt 2004b).*

Experiments to be conducted within this area of research

One research project consists of combining the devices described here with the DNA sequencing techniques. It means that optical excitation and detection devices involved in SERS and other methods have to be incorporated, on a chip at the nanoscale level, together with the nanofluidics methods. Such a device should also involve the extraction of DNA from cells. The following features have been proposed for this type of device (Tegenfeldt 2004a):

1) *Cell sorting* based on mechanical or electrical properties of entire cells or on artificial properties such as detection of cells prepared in advance with detection tags.
2) *Cell lysis* by osmotic shock in water.
3) *DNA extraction and purification* through nanochannel trapping described above.
4) *DNA fractionation* through a DNA prism.
5) *Uniform stretching in nanochannels of DNA*
6) *Linear analysis of DNA* with sequencing methods described in chapter 3.

The technique described here regards DNA, but the ultimate goal is to be able to analyze the biochemistry of entire cells and not only short DNA sequences on such chips.

6. DNA self-assembly and scaffolding

The fact that DNA can be used as a self-assembler relies solely on the base-pairing properties. On a short range, less than the persistence length of 50 nm, double-stranded DNA is a very stiff molecule and the idea is to use small pieces that self-assemble into geometrical objects and lattices. Although bigger (the smallest building blocks are about 5*15 nm), DNA is more advanced than other types of self-assemblers because it can pair selectively and form advanced patterns and even objects. During the last 20 years, a number of geometrical objects have been fabricated or theoretically proposed. In what follows, certain structures will be overviewed, going from simpler to more advanced. At the end of the chapter, there will be a thermodynamic treatment of DNA self-assembly.

6.1. The DNA cross

Four different strands of DNA with certain conditions on the base pairings form the cross, see fig. 6.1.

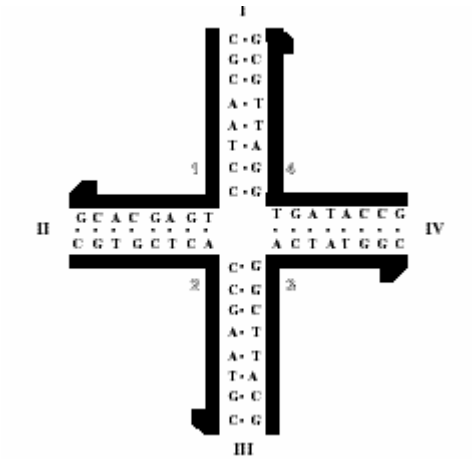


Figure 6.1: *A DNA cross made of four single strands*

Figures adopted from (Seeman 1982).

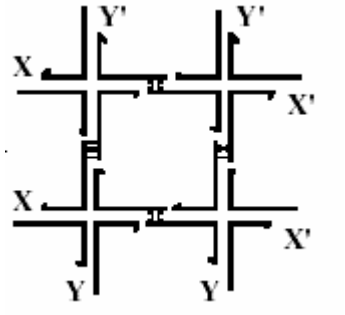


Figure 6.2*: A DNA lattice is in theory formed by crosses which are joined together by sticky ends.*

6.2. A 2D-lattice

A 2D-lattice is formed, at least in theory, by applying so-called "sticky ends" to the DNA crosses and then ligate (Seeman 1991). This means that there will be some bases at the end of each of the four strands that are complementary to one of the other three sticky ends. See fig. 6.2. It is obvious that the design of the sticky ends is crucial to DNA self-assembly.
It has been experimentally confirmed, however, that these structures lack in rigidity. Instead, a DX crossover molecule, which is a modified DNA cross, has shown to be much stiffer building blocks in a 2D-lattice (Seeman 2003), see fig 6.3. DX molecules fabricated so far measure about 5*13 nm. A 2D-lattice is found in fig. 6.4. Yet another structure which proves rigidity and has been manufactured lately is a larger DNA cross. This structure will be discussed at the end of this chapter. Fig. 6.17.

6.3. Wang tiles

In the 90's, Erik Winfree, then PhD Student, now professor at Caltech, proposed that that the properties of the sticky ends of the DX-molecule correspond to the mathematical concept of *Wang tiles* (Winfree 2003). These tiles were proposed by Wang in 1963 and are of interest particularly to computer scientists because they offer a prospect for massive parallel computing. In the theoretical model, Wang tiles consist of geometrical 2D objects, for example squares. Each side of the squares is colored and in a pattern only squares with the same colors on the adjacent sides are allowed next to each other. The color property can be complemented with a shape property or others. See fig 6.5.
A DX crossover molecule has four sticky ends. This corresponds to the four sides of the Wang tile described above. Furthermore, obviously, one sticky end fits into another only if a) their base code is complementary and b) the strands make a 3' vs. 5' pair. These two properties correspond to *color* and *form* of the Wang tiles. It is thus clear that in a theoretical discussion of how DNA self-assemble, the concept of Wang tiles can be adopted.

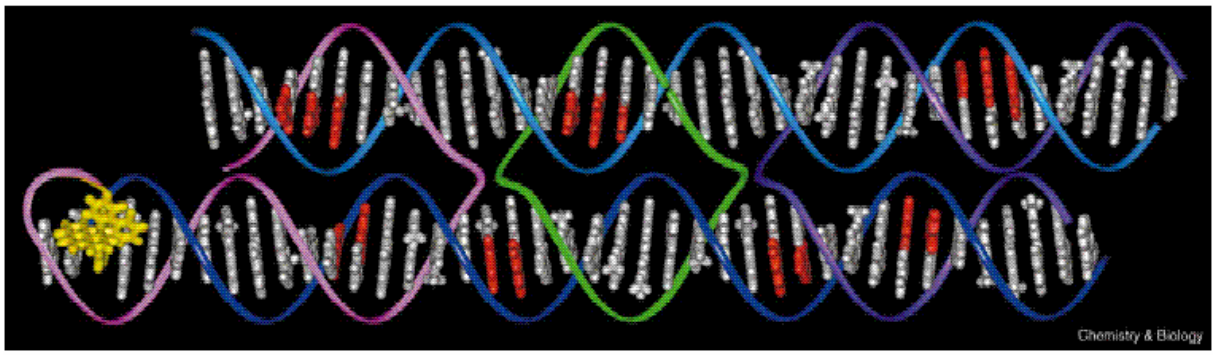


Figure 6.3a. *The double-helical form of a double-crossover (DX) molecule. Adopted from (Odom 2000).*

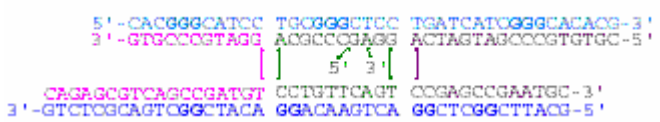


Figure 6.3b. *The structural form of a DX molecule. The four ends can be made "sticky". Adopted from (Odom 2000.)*

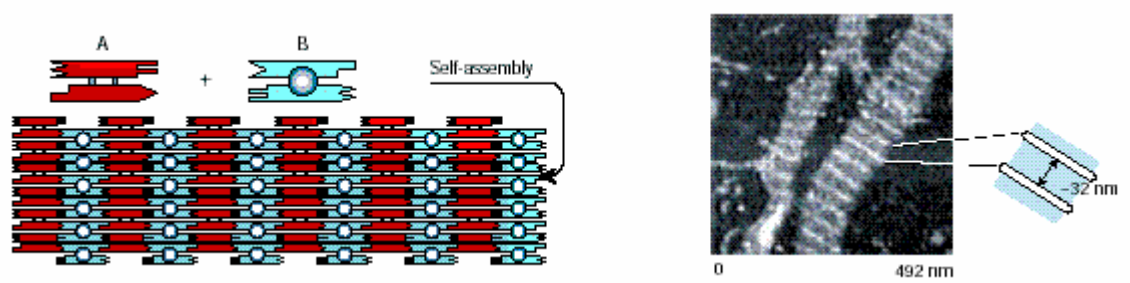


***Figure 6.4**. Left: The principle of building a 2-periodic 2D-lattice with two sets of DX molecules A and B. The dots represent "flags" of DNA attached to the molecules to make it more recognisable by AFM (right). Adopted from (Seeman 2003)*

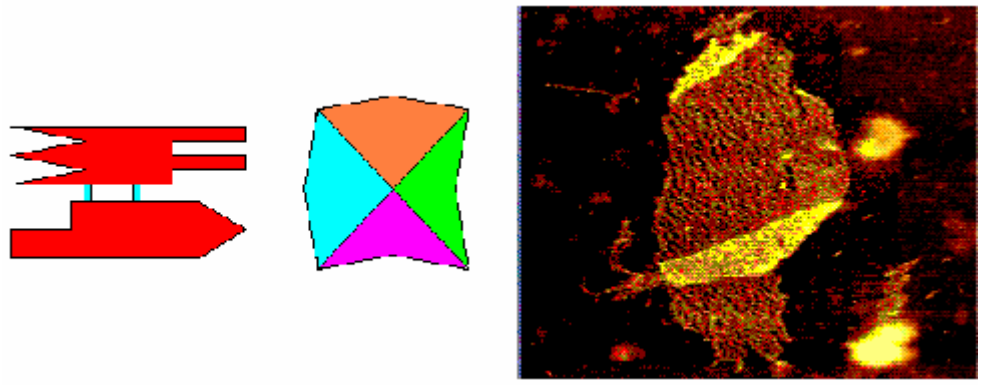


***Figure 6.5**: (Left) Two ways of modelling DNA crossover molecules as Wang tiles. In the left figure, the shapes and positions of the four ends correspond to the sticky ends of a DNA crossover molecule, in the right the colors and the protrusion/intention does. (Right) Another AFM image of a lattice. Right figure adopted from (Winfree 1998).*

6.4. Algorithmically assembled DNA structures: DNA computing

Algorithmically assembled DNA structures are the most advanced structures, and much research is to be conducted within this area during the next decades. This field represents the link between DNA nanotechnology and DNA computing. Computer scientists, chemists and biotech researchers work jointly here.
It is to be noted that as far as known, the research described here has hitherto been mostly theoretical.
The notion of "DNA computing" may be confusing to start with, as most people think of a computer as a PC containing processor, hard disk etc. This is naturally not the case here. A DNA computer consists of between a picomole and a micromole of DNA strings in a solution that perform computational tasks due to their ability to couple to each other in different manners according to how their sticky ends are designed. As an example, we will describe a "DNA-computer" that counts binary. This has been theoretically proposed (Seeman 2003).
Consider the seven DX molecules in fig 6.6. The protrusion or intention of the wang tiles equals the property "form" described earlier. Two of them are set to represent "0" and the other two "1". We use three more sets of DX molecules or Wang tiles to define the edges of the pattern. We then add a large number of each of the four "counting molecules" and the

pattern in fig 6.6 will form - *must* form. Since there are theoretically no other binding alternatives, this structure thermodynamically optimizes the system. This is not obvious as it relies on the assumption that the pairing of DX molecules with all four sticky ends complementary will almost always be preferred to pairing of molecules with only some sticky ends complimentary. That this is the case will be shown at the end of this chapter. The pattern, which shows counting from 1 to 12 would be about 50*30 nm. Furthermore, note the parallelism in DNA computing: As every part of the pattern is formed simultaneously, the "DNA computer" in this example doesn't count 1-2-3… but counts from 1 to 12 "momentarily".
Another example of algorithmic assembly would be cumulative logical operation with triple crossover molecules on a string of binary bits. Mao et al. have shown this (Mao 2000). Their idea is to create a pattern of molecules representing the cumulative XOR operation, and the principle is illustrated in fig 6.7. Readout in 2D algorithmic assembly can be managed by attaching quantum dots of different colors to different tiles.

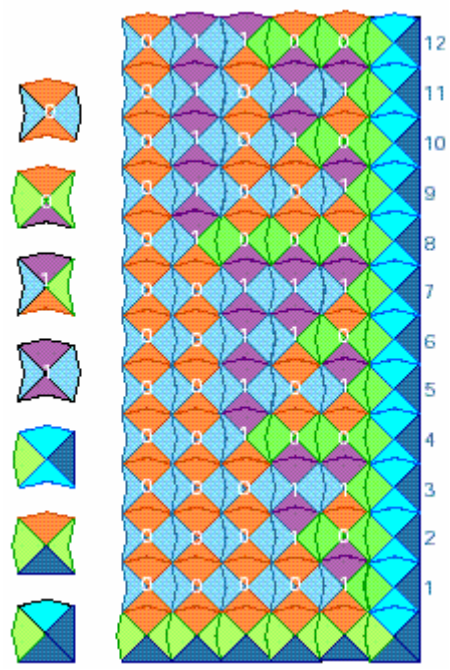


Figure 6.6. *A self-assembled pattern that shows counting from 1 to 12. Four sets of "counting tiles" and three sets of corner tiles are involved. Adopted from (Seeman 2003).*

6.5. Constructing 3D objects and lattices using DNA

The next step is going from 2D to 3D. The simplest 3D structure is obviously the cube, see fig 6.8. A problem with building a DNA-cube is that it is not rigid enough, even if DX molecules are used. Triangular-shaped objects are stiffer, and recently regular octahedrons with a diameter of 22 nm have been produced (Shih 2004). See fig 6.9. A project for the future would be to cluster the 3D objects into lattices. The power of DNA computing in a 2D lattice is already impressive, in a 3D lattice it would be astonishing, as parallel computing in this case could be performed in 3 dimensions. A way of doing this could be to make two sets of octahedrons and let some of the sticky ends fold to the other instead of folding on itself. When investigating these 3D-structures, readout would of course be a major problem, as it would be very difficult to "see" inside.

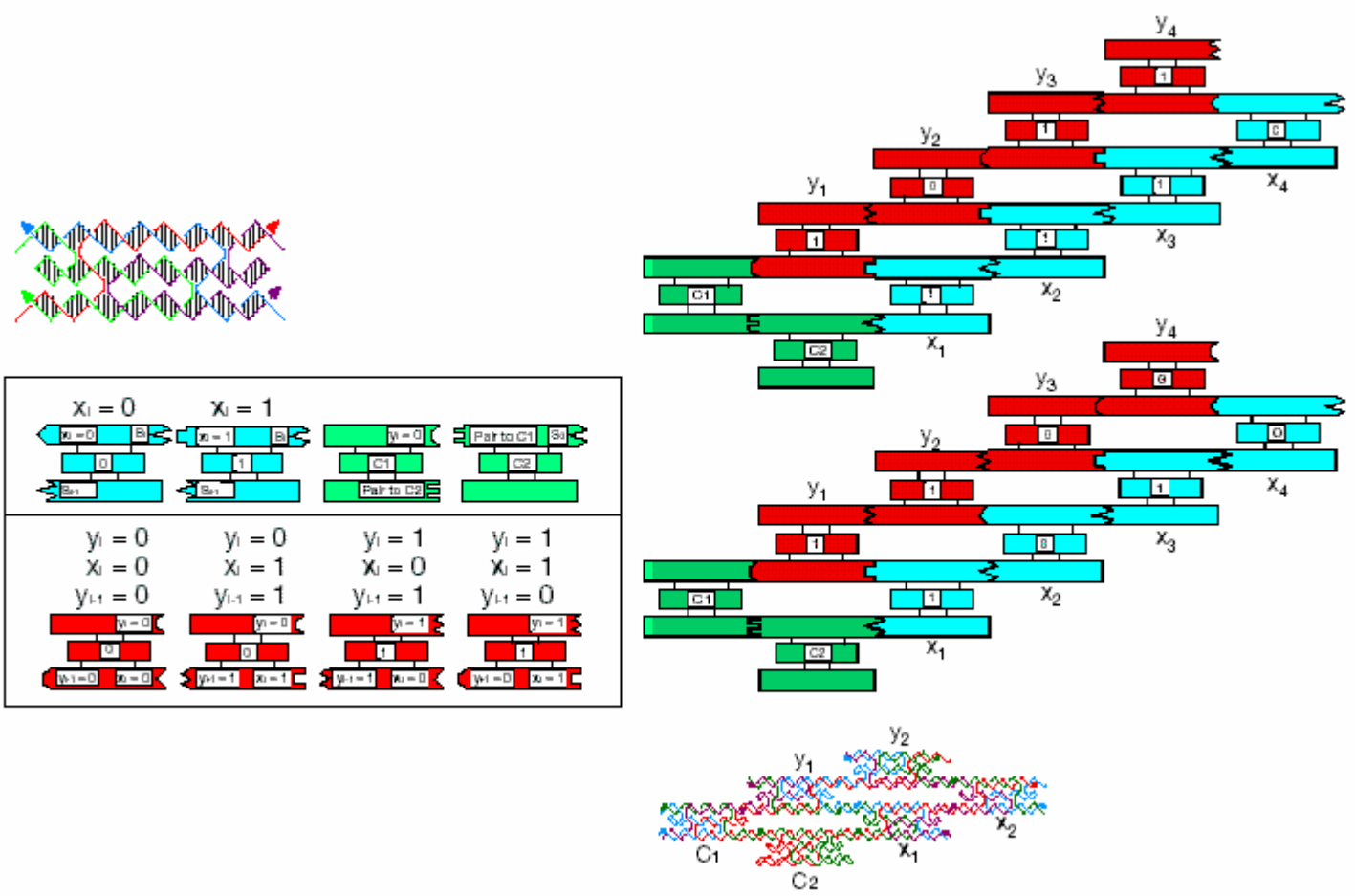


Figure 6.7. *Cumulative XOR computation. The blue triple crossover molecules are the "input tiles", the red represent the "computed" answer. Adopted from (Mao 2000).*

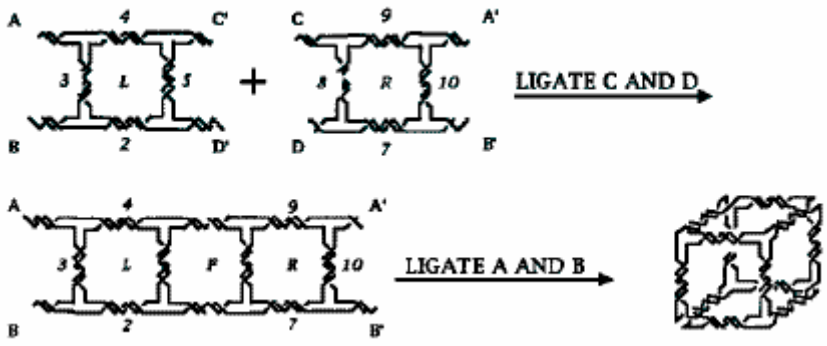


Figure 6.8*: The principle of forming a 3D-object: The cube. The object evidently contains 3-branched junctions. Adopted from (Seeman 1991).*

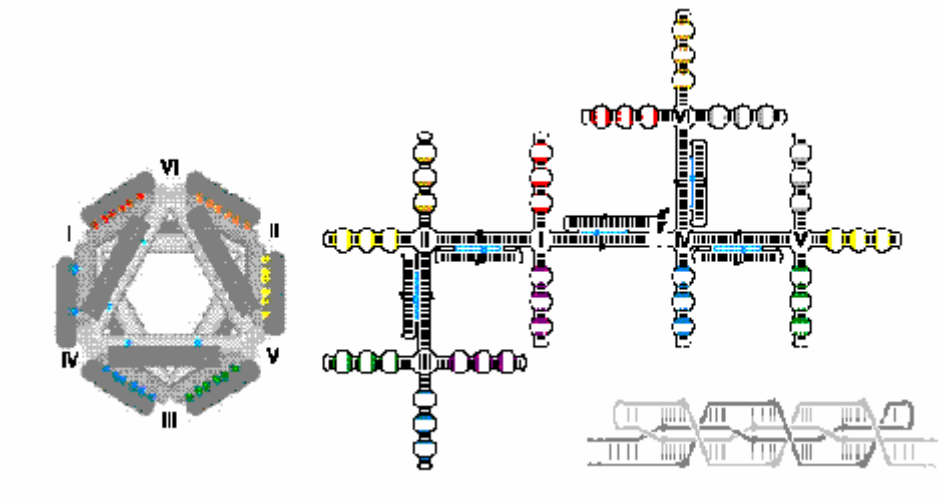


Figure 6.9*: Design of the DNA octahedron. 6 crosses are joined together by crossover junctions and pairing of sticky ends with the same color. Adopted from (Shih 2004).*

6.6. Other ways of computing with DNA

Although the research towards algorithmically assembled DNA lattices is mostly at a theoretical stage, other types of DNA assemblies have experimentally proven able to solve specific computational problems. A paper by Leonard Adleman in 1994 launched the notion of DNA-based computing (Adleman 1994). Adleman used single-stranded DNA strings to solve the directed Hamiltonian path problem. The principles of his experiment will be outlined here.

A graph G is said to have a Hamiltonian path if and only if there exists a path starting and ending at two different vertices A and B and enters all vertices exactly once. This path is the Hamiltonian path.

In his experiment, Adleman wanted to find such paths starting at 0 and ending at 6 in a directed graph according to fig 6.10 (left). This is of course not a difficult problem but it is the way of solving it with DNA strings that is interesting.

Adleman represented each vertex by a single-stranded DNA-string of 20 random bases. Call these $O(i)$, $i=1,...,6$. Call the complimentary base sequences $\bar{O}(i)$. The 2*6*6=72 edges were then represented according to fig 6.10 (right). Call these $O(i \rightarrow j)$ with obvious notation. Make special arrangements at $O(0)$ and $O(6)$. Then mix 50 pmole of each of the strands $\bar{O}(i)$ and $O(i \rightarrow j)$. .

(This makes about 10^{11} of each strand) and let them arbitrarily pair into double-stranded DNA. Ligate where necessary. This will create large chains representing different paths and lengths in th*e.g.*raph. Then follow this algorithm:

1) Keep only those paths that begin with $O(0)$ and ends with $O(6)$.
2) Keep only those DNA with exactly 140 base pairs, corresponding to 140/20=7 vertices and base pairs.
3) Keep only those DNA with each base pairing sequence $O(i) - \bar{O}(i)$ occurring at least once.

The remaining DNA oligonucleotides now represent Hamiltonian paths.

It took quite an amount of laboratory work to perform the three steps and find the winning sequence. The important thing is that this shows the principle of a way to compute with DNA.

The algorithm described is non-deterministic; *i.e.* it builds its success on the probabilistic assumption that that the winning DNA actually will form, on which there would be no risk of betting in this case, considering the huge number of DNA strands.

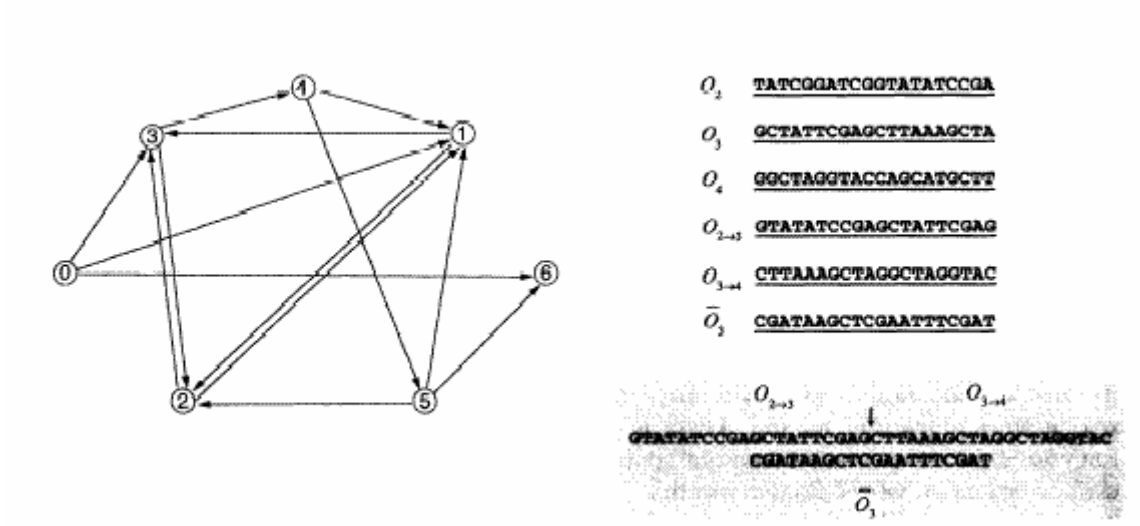


Figure 6.10: *A graph and the DNA representation of its nodes and vertices. The task is to find the directed Hamilton path, i.e. the path from 0 to 6. From (Adleman 1994).*

6.7. Quantitative assessment of DNA computing

It should be emphasized once more that the advantage of DNA computing compared to conventional lies in the parallel computing features. This is true for both ways described here: algorithmic assembly and computing by DNA strand representation as in Adleman's experiment. The key is that very many possibilities can be tested simultaneously whereas a conventional computer only may test one at a time. DNA computing is stochastic and conventional computing is deterministic.

As an example of time gain in parallel computing the computing times of the Hamiltonian path problem in conventional and DNA computing have been compared (Adleman 1994). For a conventional computer it scales as $t(n)\sim c^n$ where c is some constant $c>1$ which describes the overall speed of the processor and other circumstances. n is the number of vertices in the graph. The computing time in a DNA computer, on the other hand, would be $t(n)\sim n^d$ where $d>1$ is a time constant which is decided by the DNA computer environment. The constants c and d are normally in the range 1<c,d<10. The exact values of c and d are of no importance since it clearly shows that for sufficiently large n the computation time of the conventional computer, which is exponential, will be far larger than that of the DNA computer, which is a power law. Compare the values of n^2 and 2^n for large n, say $n=10$. In this case a perfectly working DNA computer would be $\sim 2^{10}/10^2 \sim 1000/100=10$ times faster! In this case, as $2^4=4^2$, a DNA computer would be faster for n as small as $n=5$.

For a quantitive analysis of the time advantage of DNA computing at growing n, let

$$\alpha = \frac{n_1}{n_0} \qquad (6.14)$$

where $n_1>n_0$.

Let δ_0 and δ_α be the rates of computation time (DNA computer)/computation time (comventional computer) at n_0 and after the number of vertices has increased a factor α:

$$\delta_0 = \frac{c^{n_0}}{n_0^d} \qquad (6.15)$$

$$\delta_\alpha = \frac{c^{\alpha n_0}}{\alpha^d n_0^d} \qquad (6.16)$$

Then (6.14)+(6.15)+(6.16) give:

$$\frac{\delta_\alpha}{\delta_0}(\alpha, c, d) = \frac{c^{n_0(\alpha-1)}}{\alpha^d} \qquad (6.17)$$

If, as in the example above, $c=d=2$, and n doubles from $n_0=5$ to $n_1=10$, the gain in DNA computing compared to conventional computing would be a factor $2^3=8$.

If the Hamiltonian problem were to be solved by a salesman who had to travel to every country in the world ($n\sim200$) a DNA computer would be about 10^{55} times faster, assuming c=d=2. It obviously means that by conventional computation, the salesman would not stand any chance to get the answer in his lifetime.
Adleman states that a perfectly working DNA computer is plausible to execute as much as 10^{20} operations per second at a low energy consumption rate. The flop rate of a conventional supercomputer reaches 10^{14} at its best, at an energy consumption of 10^9 operations/Joule. A DNA computer could according to Adleman achieve 10^{19} operations/Joule.
In order to solve large problems, the order of 1 micromole of DNA would be needed. It means that there would be 10^{17} single stranded DNA strings needed. If it is assumed, in correspondence to Adleman's experiment, that each strand is 20 base pairs, then there would be single stranded DNA of total length $\sim10^{18}$ base pairs. This is ~0.1 mg DNA. The number of different sets of strands needed is always $n+n(n-1)$, where the second term is the number of vertices in the graph.

6.7.1. Drawbacks

Adleman's paper is probably too enthusiastic about DNA computing. Alone the fact that it was published as far back as in 1994, without much having happened since, speaks against his optimistic views. Computer scientists are particularly interested in the parallelism features, but these may prove as prominent and easier to handle by methods being developed within other areas of nanotechnology (for example spintronics/quantum computing).

The predominant problems are two:

1) How to detect the computed DNA and separate it from false solutions?

In Adleman's case, where he solved a very simple Hamiltonian graph problem, it took him a week to find the "winning" strand (*i.e.* the one corresponding to the solution of the problem). How long time will it take when the amount of added DNA is much larger, in the order of a

micromole? Insofar DNA computing could be of use in the future, it would be if the gain in computational time were much larger than detection time. (*E.g.* solving the Hamiltonian path problem for large *n.*)
FISH cannot be used here, as this method prerequisites knowledge about the "winning" strand.

2) A solution of DNA strands in a biochemical enviromment is obviously much more inconvenient to handle than a conventional computer processor.

In addition, DNA is very expensive at large quantities and many different sequences.

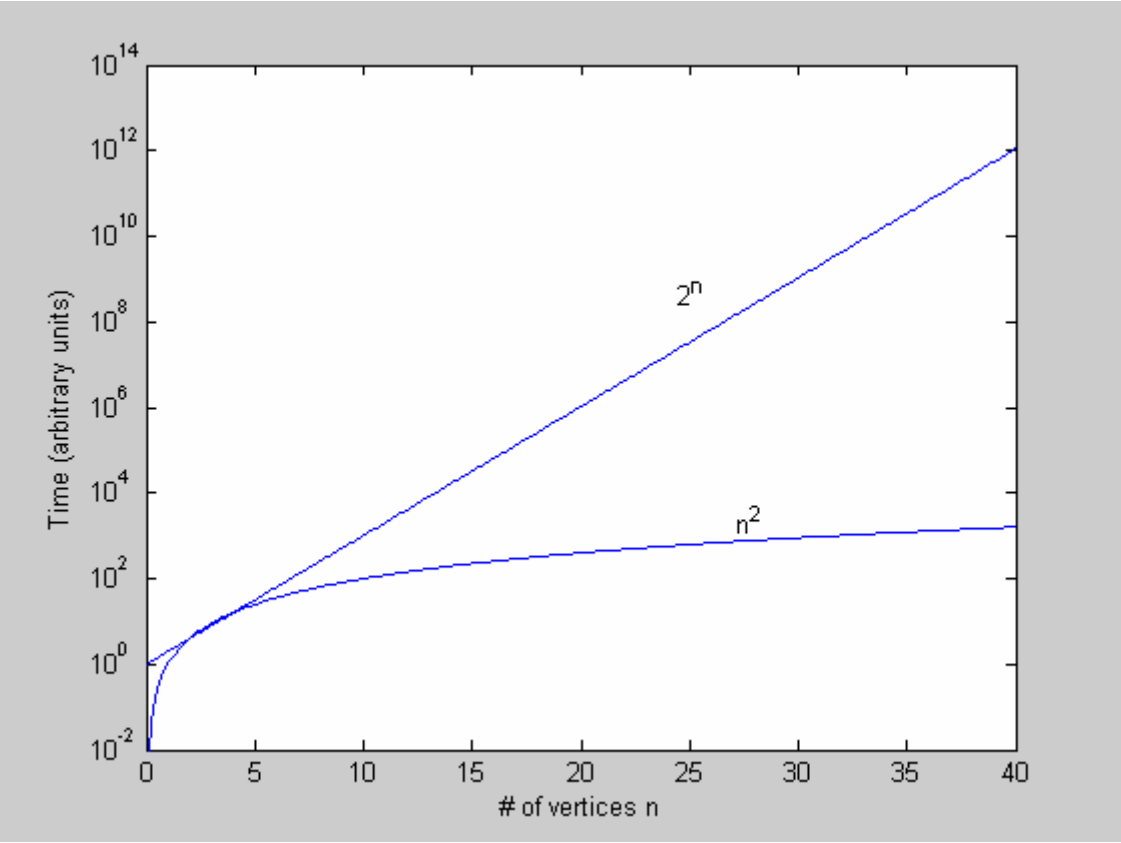


Fig. 6.11: *In the case where c=d=2, parallel computing becomes much faster as n grows beyond n=5.*

6.8. A thermodynamic model of DNA self-assembly

It may not at first sight be obvious that DNA would self-assemble only because of the sticky ends. Winfree showed that this is the case. He introduced the following thermodynamic model (Winfree 2003):

Make the following simplifying assumptions:

i) Molecule concentrations of the tiles are considered constant at all times. Molecules with different types of sticky ends (in the following referred to as different *sets* of molecules) have the same concentration. In practice this means that the free molecules always outnumber the ones attached to the pattern by several factors.

ii) Aggregates do not interact with each other. The only reactions to be considered are association and dissociation of single molecules to the aggregate.

iii) There is a time constant k_f which is the same for dissociation as well as correct (sticky ends complementary) and incorrect (sticky ends not complimentary) association.

iv) Mismatched sticky ends make no base-pair bonds at all, even if they were to have some bases in common.

Let the two parameters $G_{mc} > 0$ and $G_{se} > 0$ represent the loss in entropy when a molecule is attached to the growing lattice, and the free energy cost of breaking the bonds of one sticky end with the lattice, respectively. For reasons of energy conservation, the following must hold:

$$G_{mc} = (b - \varepsilon)G_{se} \qquad (6.1)$$

where b is the number of sticky ends with which the molecule sticks to the lattice, and ε $0< \varepsilon <1$ is an error rate constant. To simplify and clarify the discussion here, let ε=0.
Further, each set of DX molecules can be considered as occupying a thermodynamic state, and Boltzmann's state distribution formula can be used, with $G_{mc}>>kT$. Thus write the concentration rate of each set of free DX molecules as

$$[DX]=exp(-G_{mc}/kT) \qquad (6.2)$$

i.e., they are all the same because of condition i) above.

The rate of association R_f (the time it takes from when a site is allowed to when it is frozen) of a particular DX molecule to a particular site on a particular aggregate is proportional to concentration:

$$R_f=k_f exp(-G_{mc}/kT) \qquad (6.3)$$

The rate of dissociation of a molecule attached with b sticky ends is:

$$R_{r,b}=k_f exp(-bG_{se}/kT) \qquad (6.4)$$

(1)+(2)+(3) gives the ratio association/dissociation.

$$R_f/R_{r,b}=exp(G_{mc}-G_{mc})=1 \qquad (6.5)$$

resulting in no net growth. If the molecule were attached with one more sticky end, *i.e.* b+1, the same ratio would be,

$$R_f/R_{r,b+1}=e \qquad (6.6)$$

and in general,

$$R_f/R_{r,b+n}=e^n \qquad (6.7)$$

So the ratio with which association wins over dissociation grows exponentially with the number of sticky ends shared. As self-assembly is a stochastic process, this means that in the long run associations of molecules with two sticky ends complementary will always be preferred to those with only one sticky end complementary, which in turn will be preferred to those with no sticky ends that fit.

6.9. A stochastic process: The right molecule in the right place

As mentioned above, this process is entirely stochastic. Tiles that are adjacent may thermodynamically dissociate again even if their sticky ends match due to the energy in the system and *vice versa*: Tiles without every end complimentary may be adjacent for a long period. This is not critical as long as correctly assembled tiles do not surround them and the tile becomes frozen in the wrong position. In that case, error may occur. Winfree modelled the self-assembly as a stochastic Markov process, see fig 6.12. A Markov process is network of different states between which movement is possible at different probability rates. The rates may be described in a matrix $\mathbf{M}_{ij}$ where index *ij* means transition rates from state *i* to state *j*. For more on Markov processes, see *e.g.. (Rydén 1999).*

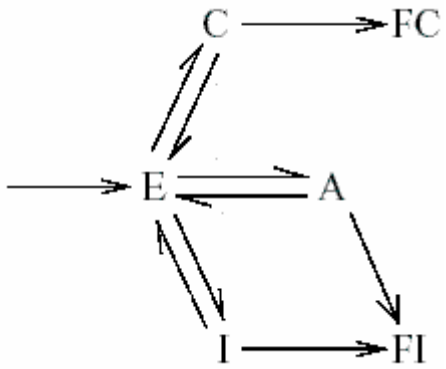


Figure 6.12. *A scheme of the DNA assembly modelled as a Markov chain.* ***E*** *represents an empty tile.* ***C*** *is a tile that becomes correctly attached with both its ends.* ***A*** *represents a tile with one correct and one incorrect connection.* ***I*** *is a tile with both ends incorrect.* ***FC*** *and* ***FI*** *are correct and incorrect frozen tiles, respectively.*
This scheme can always be used when modelling DNA self-assembly in a 2D lattice.

In fig. 6.13 the "DNA computer" described in a previous section is depicted together with the scheme above, and the probability rates of this particular assembly. Note the vacancy at the upper left corner. With notations as in fig. 6.13, this empty state (**E**) can be filled with either a **C** tile (only tile I), an **A** tile (III or IV), or an **I** tile (II, V, VI, or VII). As the concentrations of each of the 7 tiles are equal, the probability for the system in fig 6.13, left, to move from **E** to **C** is 1/7, from **E** to **A** 2/7, and from **E** to **I** 4/7. This could also be expressed as the association rates being r_f, $2r_f$, and $4r_f$, respectively. The dissociation rates are, as described previously, $r_{r,i}$. r* equals the growth rate of the pattern, $r^* \approx r_f - r_{r,2} = k_f \exp(-G_{mc}/kT) - k_f \exp(-2G_{se}/kT)$. This means that we assume growth to be overwhelmingly error-free.
In practice all this means that it's highly probable that the system moves from for example state **E** to state **I**, but the probability that it will come back to **E** is also very high. The wrong tile will be at the site for a while but will then in most cases annotate. On the other hand,

although the probability that a correct tile will attach is comparatively lower, once it has, it generally sticks to the pattern. This also highlights the assumption that the association rates are equal for all tiles, but the dissociation rates are not.

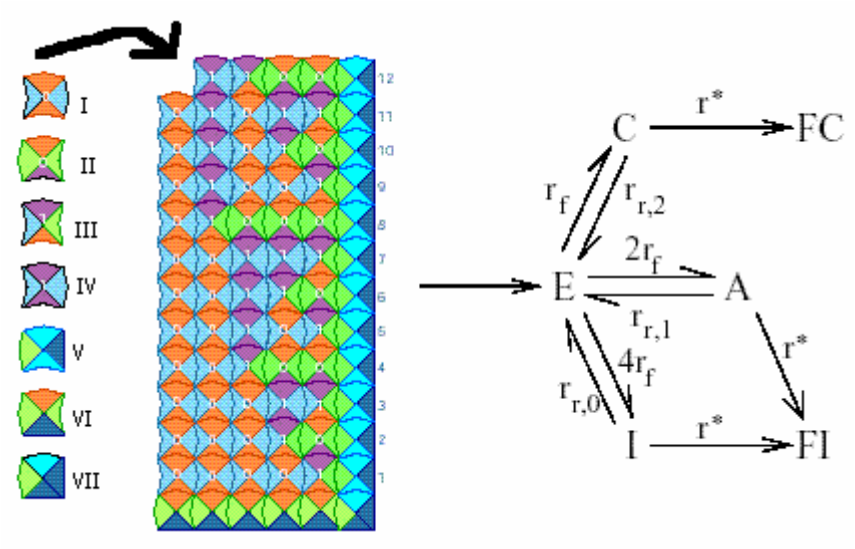


Figure 6.13. *Adopted from (Seeman 2003) and (Winfree 2003).*

In what follows, this is how Winfree analyzes the dynamics of DNA self-assembly and optimizes the parameters G_{mc} and G_{se} with respect to the maximum growth rate for a system as in fig. 6.13.

When the pattern as a whole has surrounded a site and grown past it, corrections can no longer be made. A tile, whether correct or incorrect, will "freeze" in that position. Obviously, it is desirable that correct tiles freeze and incorrect do not.

A Markov chain may be represented as a matrix. The Markov graph in fig. 6.13, right, may be described as,

$$\mathbf{M}= \begin{bmatrix} -7r_f & r_{r,2} & r_{r,1} & r_{r,0} & 0 & 0 \\ r_f & -r_{r,2}-r^* & 0 & 0 & 0 & 0 \\ 2r_f & 0 & -r_{r,1}-r^* & 0 & 0 & 0 \\ 4r_f & 0 & 0 & -r_{r,0}-r^* & 0 & 0 \\ 0 & r^* & 0 & 0 & 0 & 0 \\ 0 & 0 & r^* & r^* & 0 & 0 \end{bmatrix}$$

(Adopted from (Winfree 2003))

Here, the diagonal elements are the sum of the association rate to each node, with a minus sign, and the other elements represent the torrents between the nodes. For the dynamics of this systems the following holds:

$$D\mathbf{p}/dt = Mp(t) \qquad (6.8)$$

where **p** is the state vector.
The initial value vector is $\boldsymbol{p}(0)=[1\ 0\ 0\ 0\ 0\ 0]^T$ (as a site starts as empty). It's now interesting to analyze what happens when steady state (*i.e.* when the tiles freeze) occurs. After a long time, the following holds for the system:

$$\dot{p}(\infty) = [1\ 0\ 0\ 0\ p_{FC}(\infty)\ p_{FI}(\infty)]^T = Mp(\infty) \tag{6.9}$$

which gives

$$p_{FC}(\infty) = \frac{\dfrac{1}{r^* + r_{r,2}}}{\dfrac{1}{r^* + r_{r,2}} + \dfrac{2}{r^* + r_{r,1}} + \dfrac{4}{r^* + r_{r,0}}} \tag{6.10}$$

(Adopted from (Winfree 2003))

If, in the formula above, $r^* >> r_{r,i}$, *i.e.* $r_f >> r_{r,i}$,*i.e.* $G_{mc} < G_{se}$, *i.e.* the loss in entropy when attaching is much smaller than the energy demanded for breaking the bonds, the probability for a correct step is 1/7. This extreme result is quite obvious, as it means that a tile, no matter how many sticky ends are correct, with a very large probability attaches to the pattern. And as there are only 1/7 of all tiles that are correct, this will be the result. Growing a pattern with such parameters would turn out to be quite worthless, but it can at least be regarded as a confirmation of the accuracy of this model. A system in this state is called randomly aggregated.
On the other hand, if on the other extreme $r^* \rightarrow 0$, and no growth occur at all, then $G_{mc}/G_{se} \rightarrow 2$.
For a meaningful growth to occur, it must thus hold that $1 < G_{mc}/G_{se} < 2$.
We now find the maximum growth rate r^*_{max} with respect to G_{mc} and G_{se}. From the definition of r*, standard optimization procedures give r^*_{max} on the following line;

$$G_{mc} = 2G_{se} - \ln 2 \tag{6.11}$$

So, if we could by some means control the two parameters to scale according to (11), maximum growth rate r^*_{max} would be obtained.
Winfree also calculated r^*_{max} with respect to the error rate of the growth, ε, *i.e.* the error rate when condition (6.11) is fulfilled:

$$r^*_{max}(\varepsilon) = 10^6\ \varepsilon^2\ s^{-1} \tag{6.12}$$

where ε is the error rate of the assembly.

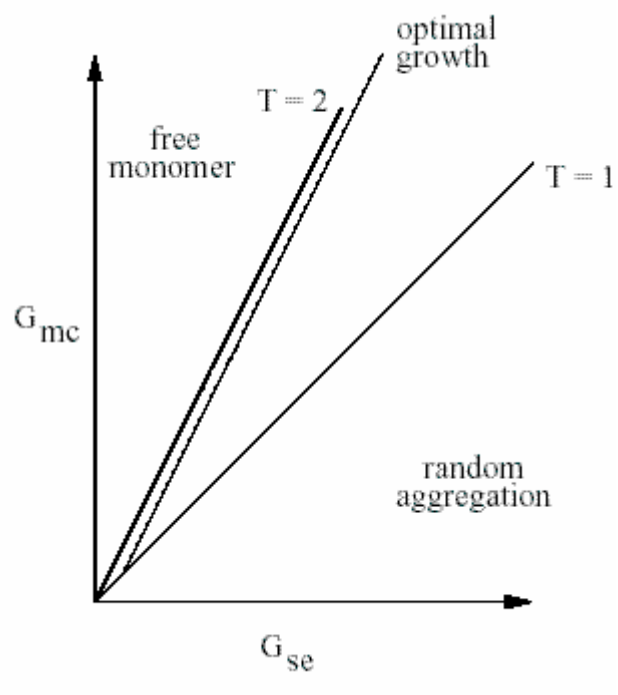


Figure 6.14. *Adopted from (Winfree 2003).* ***T*** *is the rate* G_{mc}/G_{se}.

If we could allow (statistically) 1 tile in 1000 to be incorrect, *i.e.* let $\varepsilon=0.001$, th*e.g.*rowth rate would thus be 1 tile per second. In case only one tile in 10 000 is allowed to be incorrect, we have to adopt our parameters so that the growth rate becomes 1 tile in 100 s.

It is to be noted that the thermodynamic model adopted here including the free parameters G_{mc} and G_{se} could be applied to liquids, solids, and gases of any kind. For example, a gas would then correspond to free DNA molecules as the liquid would to a DNA lattice.

6.10. DNA-mediated nanoparticle self-assembly

Methods have been developed by Mirkin *et al.* to assemble nanosized particles with the assistance of DNA strands (Mirkin 1996). The idea is simply to attach complimentary DNA strands to these particles and let DNA mediate bonding between them. The nanoparticle can be made of metal, magnetic, or semiconductor material. Mirkin's group has extensively examined a reversible process of gold particle self-assembly.
The first step consists of two different sets of 13 nm-diameter gold particles with non-complimentary DNA strands with sticky ends being mixed together. The particles will hence remain free. In the second step, another DNA strand complimentary at each end to the two sets attached to the gold is added to the solution, thus making the particles cling together. However, the formation is not well ordered enough to be regarded as a crystal.
Effects of the hybridization may be observed on the macroscopic level, as it causes the color of the solution to change from red to purple.

As mentioned, this process is reversible. Switching between the two states of hybridization on/off is achieved by varying the temperature. The dissociation temperature for a double-stranded DNA depends on its sequence, the contents of the solution etc. In this case it was 42°C. At lower temperatures the cluster will thus remain intact but will dissolve at higher temperatures. Fig 6.15 shows a spectroscopic analysis of absorbance vs. temperature at wavelength 700 nm, where there is a high sensitivity to changes in Au aggregation. At 80°C there is no change in absorbance since there are no clusters.

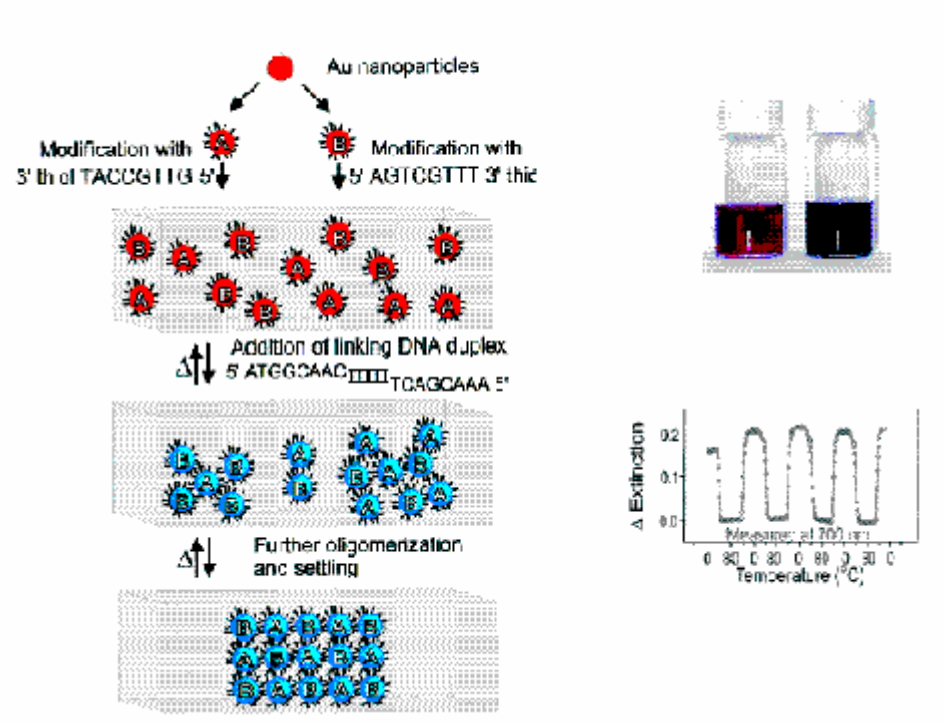


Fig 6.15. *DNA-mediated self-assembly of Au nanoparticles. From (Mirkin 1996)*

The Mirkin group developed their experiment with gold particles further in an ingenious method for DNA sequencing (Cao 2002c). In order to detect DNA of a specific sequence (the target DNA), gold nanoparticles were prepared with complimentary DNA about half the length of the target DNA. A silver-surface was prepared with DNA complimentary to the other half. When a mixture of target DNA and nanoparticles travelled over the surface the target DNA would attach them according to fig. 6.16. When exposed to laser light, the nanoparticles become SERS active and detection is possible.
This method's ability for multiplexing was also tested. Six different dye molecules were used as probes for six different DNA strands. Every dye has its own characteristical spectrum. Colors were coupled to each spectrum. Some of the six strands were mixed together in different combinations and the simultaneous detection ability of the probes was confirmed.

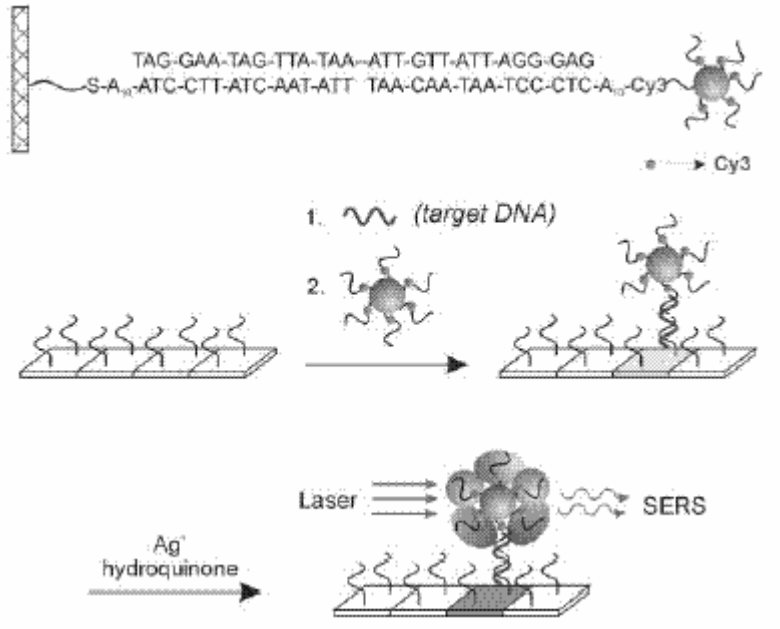


Figure 6.16. *DNA detection through SERS. From (Cao 2002c)*

6.11. Experiments to be conducted within this area of research

Lattices – what can they be used for?

As mentioned above, DNA molecules can be used for computing. It does not mean that a PC or a laptop will be replaced by a DNA solution in the near future, but it could rather be used for specific computational problems that demand a large amount of parallelism. However, in the emerging field of quantum computing, methods are being developed that might turn out to be equally adapted for parallel computing and more conveniently constructed.
Creating lattices of DNA represents a bottom-up method, in contrast to *e.g.* lithography, which is a top-down method. In the former, one tries to take advantage of molecular properties created by nature itself, for example the base-pairing in DNA, and from a nanosize building block like the DX molecule efforts are made to construct larger structures like lattices.
In a top-down method like lithography, on the other hand, efforts are made to improve existing techniques so that even smaller objects can be made. One highly interesting area for the future would be fabrication of nanoelectronic circuits (Seeman 2003). The idea is simply to attach electronic nanocomponents fabricated by top-down methods to the DNA tiles and let them self-assemble into a circuit. Nanowires have already been fabricated (Wooley 2003; Yan 2003). DNA crosses of the form described in fig 6.17 (As with the DX molecule described earlier, these are also rigid enough) were assembled into a nanorod.

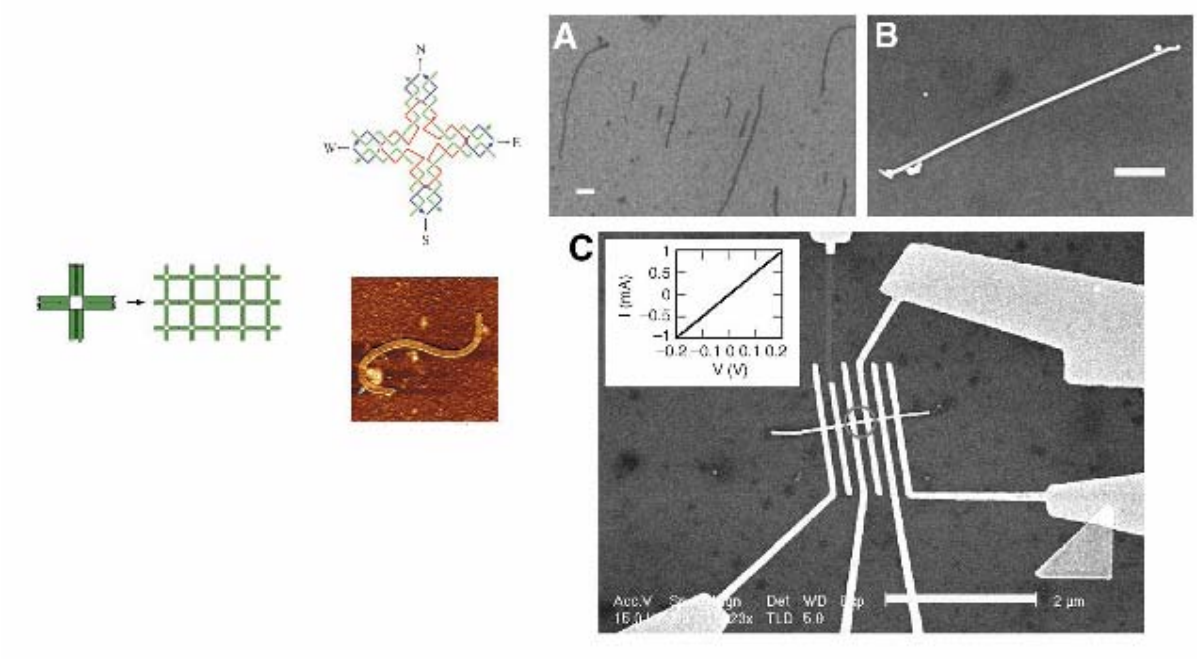


Figure 6.17 Left: *A nanowire fabricated by use of self-assembled DNA. Three sets of crosses are needed here: The mesh in the middle (at the arrow in the leftmost figure) has sticky ends at all four branches. Crosses to the left and right of the middle tile are border tiles and only have only three sticky ends.* ***Middle:*** *The DNA nanowire and the structure of the large DNA cross (not to be confused with the simple one in fig 1).* ***Right:*** *A) DNA nanowires B) Nanowire of gold coated DNA through e-beam evaporation. C) Ohmic (linear) behaviour of conductivity. Dimensions of the nanowire fabricated here are 35 (height)*43 (width)*several thousands (length) nm^3. From (Yan 2003).*

Experiments measuring the conduction of DNA show a wide range of results, from supraconducting to entirely isolating, but there seem to be better proof for the latter (Zhang 2002). In order to fabricate conductive DNA-based nanorods, these have to be coated with conductive material (in this case gold) using e-beam evaporation. Yan *et al.* have also created a 2D-lattice to which proteins were attached in an ordered array. See fig. 6.18. This allows for crystallographic examinations of the protein structure, which today only may be performed with crystallized proteins. Similar 2D-lattices with arrays of quantum dots instead of proteins are possible. Such arrays have been proposed as a base for quantum computing (Brown 2001), but could also be seen as a network of photonic wires.

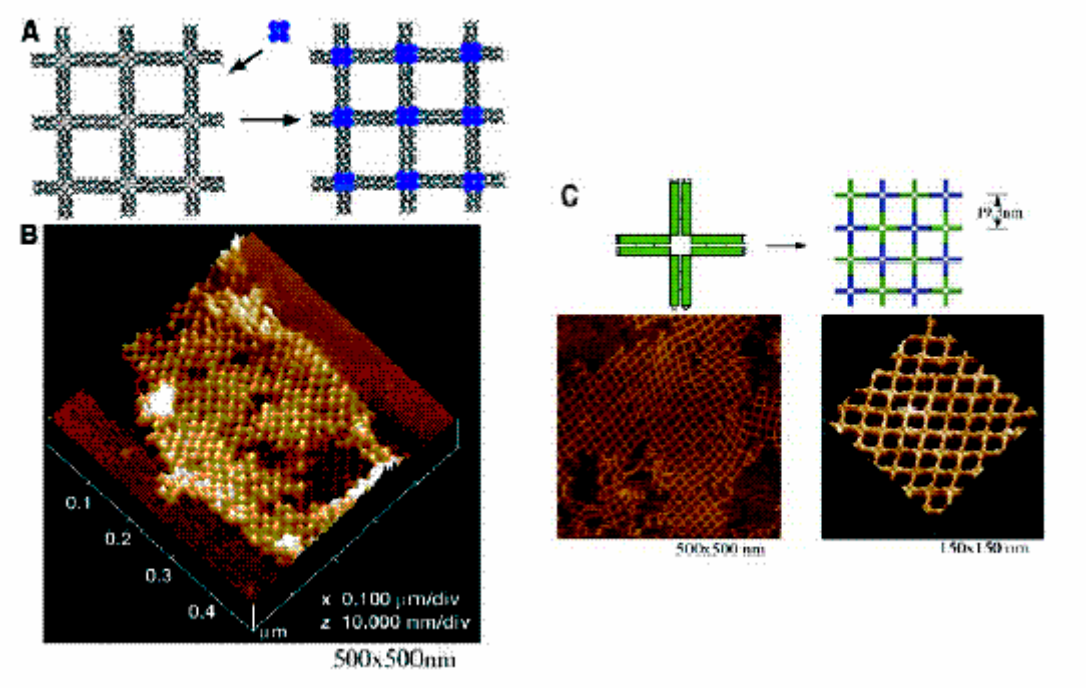


Figure 6.18 *A) and B) A protein (streptavidin) is attached to a 2D-lattice made of the same DNA cross construction as in fig 6.16. C) The DNA lattice without proteins. From (Yan 2003).*

It might be possible to position carbon nanotubes or other electronically active molecules by attaching these to some of the tiles. The procedure of attaching nanotubes to DNA has been researched (Williams 2002). See fig 6.19.

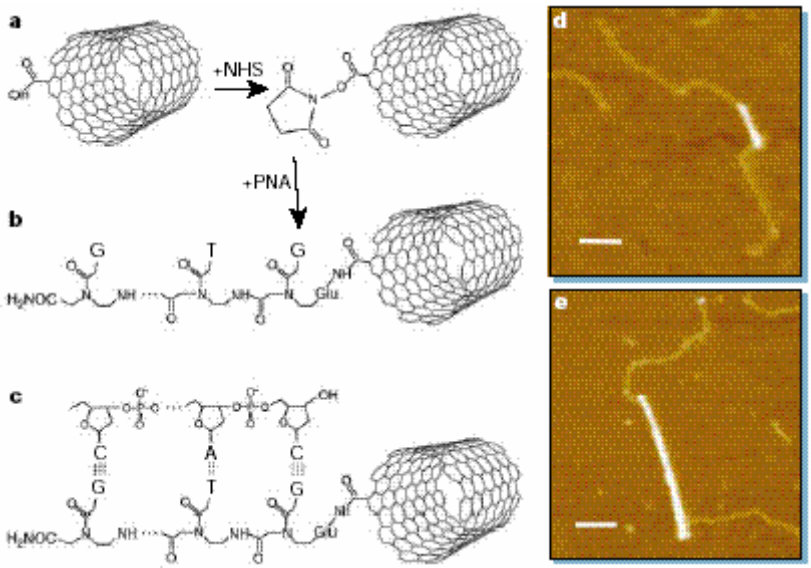


Figure 6.19: *Attaching PNA to carbon nanotubes. N-hydroxysuccinimide (NHS) acts as a linker between the nanotube and PNA. D) and E): The bright line is the nanotube, the paler are PNA. Scale bars: 100 nm. From (Williams 2002)*

Nanowires consisting of copper-coated DNA with a diameter of only 3 nm have been fabricated, fig 6.20 (Wooley 2003).

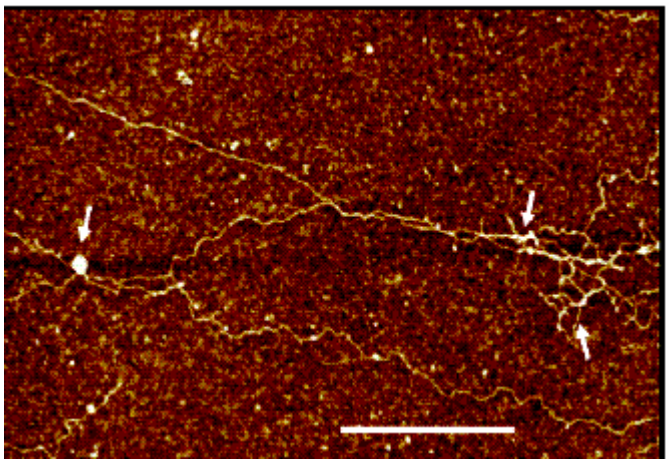


Figure 6.20 *Copper nanowires of 3 nm diameter have been fabricated on DNA. Arrows indicate slightly raised features of DNA. From (Wooley 2003).*

A project for the future would be to create even more advanced arrays, *i.e.* in 3D or by some means create 1D-0D-1D-structures on DNA, such as heterostructured nanowires. Photonic wires taking advantage of the FRET effect have so far not been combined with DNA self-assembly. It would be possible to attach FRET-active particles (flourophores or quantum dots) to tiles before letting them assemble into complex structures described previously.
It is clear that insofar DNA self-assembly is a serious research topic for the future, it is for its scaffolding properties rather than computing. Far away in the future lies of course the prospect of 3D-assembled nanocircuits!

Playing Tetris with DNA tiles: Using nanofluidics for DNA assembly

It has proven fruitful to build a nucleus of tiles through a mediated process and then start self-mediated assembly. In this case, the tiles already have something to cling to at the beginning. This is called *directed nucleation assembly* and could possibly be performed using the positioning technique of nanofluidics. 1D-wires of DNA tiles may prove especially suitable to be fabricated by this method.
Nanochannels can be fabricated as well as prepared with soft lithography for easier flow (Tegenfeldt 2004b). Letting the tiles flow into a nanochannel not only facilitates assembly but the wire will also be stretched and ready for attachment of proteins, quantum dots, fullerenes, gold etc already at the beginning. Fabrication of 2D- and 3D-structures may also be facilitated, see fig 6.21. Narrow nanochannels which could prove suitable, have been fabricated (Cao 2002a; Tegenfeldt 2004b).

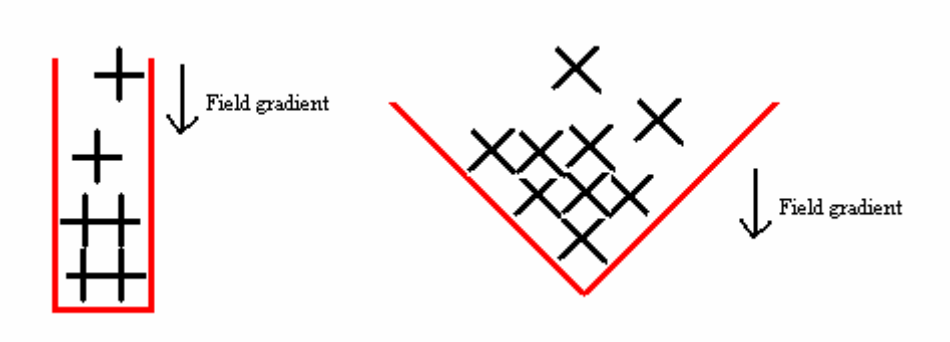


Figure 6.21 *Playing Tetris with DNA crosses. The structures shown in red should be fabricated by lithographic methods, and coated by "silicon rubber" using soft lithography. They will mediate DNA assembly in 1D (nanowires) and 2D (lattices). A sufficiently large 2D-lattice fabricated with this method may act as a start-off for more advanced, i.e. algorithmic, assembly.*

7. Genetic switches

7.1. Simple genetic switches

Genes are part of regulatory systems. There are structural genes and regulator genes. The latter code for proteins that represses or activates structural genes, but even a regulatory gene itself can be turned on and off. An entire system of structural and regulator genes is called an operon. A well-known and well-studied example of a genetic regulatory system is the control of the life cycle of the bacteria *bacteriophage λ* (Ptashne 1992). When infected by the *λ phage* virus, the viral DNA enters the so-called lysogenic mode in which it integrates into the host DNA through a protein-mediated process involving the *integrase* protein. It remains passive afterwards. This behaviour improves the chances of long-time survival of the viral DNA. The *λ phage* DNA remains there until the cell is hit by harmful UV radiation, thereby risking destruction. This makes viral DNA excite from the host DNA and enter the lytic mode, in which synthesis of the viral proteins start. The virus thus escapes its doomed host. The phenomenon is therefore referred to as a "lifeboat" response: In accordance to the fundamental biological principle, the only thing that matters is to reproduce one's own genes – and get away with it!

We will now take a closer look at the mechanisms governing this course of events, especially the switching from lysogenic to lytic mode. The phenomenon was examined in detail by Ptaschne. The two proteins *Cro* (represses all *λ* genes) and *λ repressor* (represses all *λ* genes except itself at relatively low concentrations, but all genes at high concentrations) are main actors, together with the initiation site of the gene, $\boldsymbol{o}_{\mathbf{R}}$. This site is divided into three subsites, $\boldsymbol{o}_{\mathbf{R1}}$, $\boldsymbol{o}_{\mathbf{R2}}$, and $\boldsymbol{o}_{\mathbf{R3}}$, which have different affinities, such that $A_{OR1}>A_{OR2}>A_{OR3}$. When there is no *λ repressor* present, RNA polymerase begins transcription and due to the higher affinity, it will prefer to transcribe from the $\boldsymbol{o}_{\mathbf{R1}}$ site and "rightwards" where every gene except the *λ repressor* gene itself is found. This includes the protein *excisionase*, which would switch the cycle into the lytic mode by cutting out the *λ phage* DNA from the host DNA. However, this and other proteins coded by genes found "rightwards" to the binding site $\boldsymbol{o}_{\mathbf{R}}$ will very soon be suppressed, due to production of *λ repressor* proteins whose gene is found "leftwards[4]". The *λ repressor* proteins will attach to the sites $\boldsymbol{o}_{\mathbf{R1}}$ and $\boldsymbol{o}_{\mathbf{R2}}$, which in turn silences the entire transcription of *λ phage* DNA except for the *λ repressor* protein. So finally there is such an abundance of repressors that even site $\boldsymbol{o}_{\mathbf{R3}}$ will be blocked. Evidently, this is a self-regulating process.

What has been described here is the course of events when the bacteriophage *λ* is in the lysogenic mode.

Exposure to UV light, which might be harmful to the host cell, leads to production of a protein in the host cell that stimulates the destruction of the bonded *λ repressors*. Now transcription is possible and the *excisionase* protein, among others, is synthezised. So is also the *cro* protein, described earlier, which has properties similar to *λ repressor* with certain exceptions: First, it is not destroyed by UV light, and second, for the affinities it holds that $A_{OR1}=A_{OR2}<A_{OR3}$. This means that the *cro* concentration will increase, but it will take a while before it represses the $\boldsymbol{o}_{\mathbf{R1}}$ site, making it possible for *excisionase* to separate the viral and host DNA before the *cro* becomes as abundant as it shuts the entire viral DNA off. The cycle has entered the lytic mode. The silencing of the viral protein by cro is effective only as long as the viral and host cells DNA are integrated. Viral DNA disintegrated by *excisionase* will now begin synthesising and the virus will reproduce and leave the cell.

[4] The notions of "rightwards" and "leftwards" here are of course arbitrarily chosen and serve only as orientation.

To summarize, viral DNA enters the cell, integrates into the host cell where it is more protected, and may remain silent in host cells for generations until the cell is damaged. Viruses then leave the cell in a "life-boat" response and infect new, healthy ones. This corresponds to the fundamental biological law: Every organism containing DNA, from viruses to mammals, strives to protect and reproduce its own.

7.2. Logic gates made of operons

The switching mechanisms of the genes can be even more complex. In most cases the genetic network can be described by logical components and circuits (Weiss 2003). These mechanisms exist in living organisms to respond to extra-cellular signals of which inducers or other enzymes are bearers. Proteins are produced *e.g.* when a change in environment (temperature etc.) requires them. One goal for genetic engineers is to design synthetic genetic networks with logic gates as building blocks.
Detection might in this case be made using fluorescent methods of the transcribed protein. The transcribed protein gives raise to change of color of the solution on the macro scale, and the "output" of the gate could be interpreted as the color change (or, of course, absence of color change) of the solution as a function of input.
The dynamic behaviour of the gates will not be examined here, only a few examples to show their function.

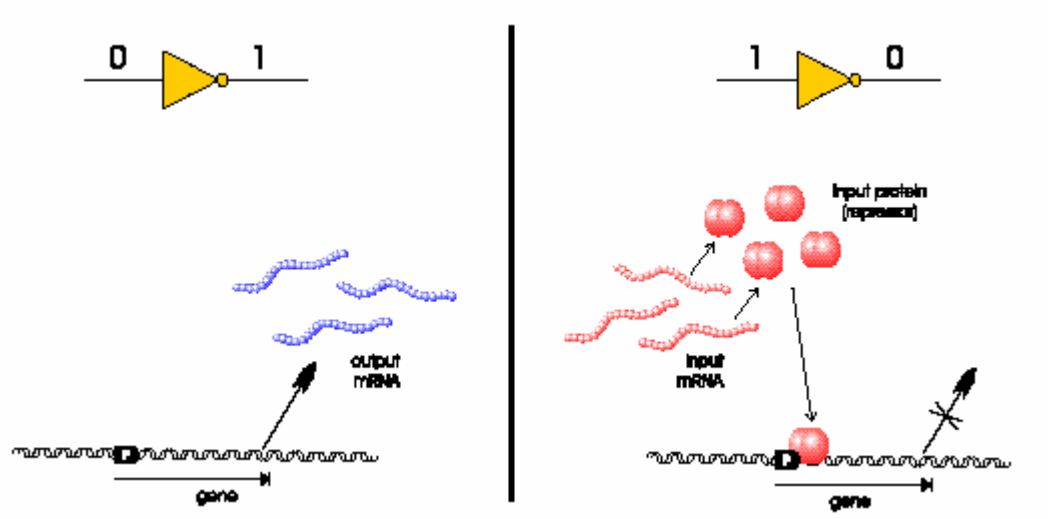


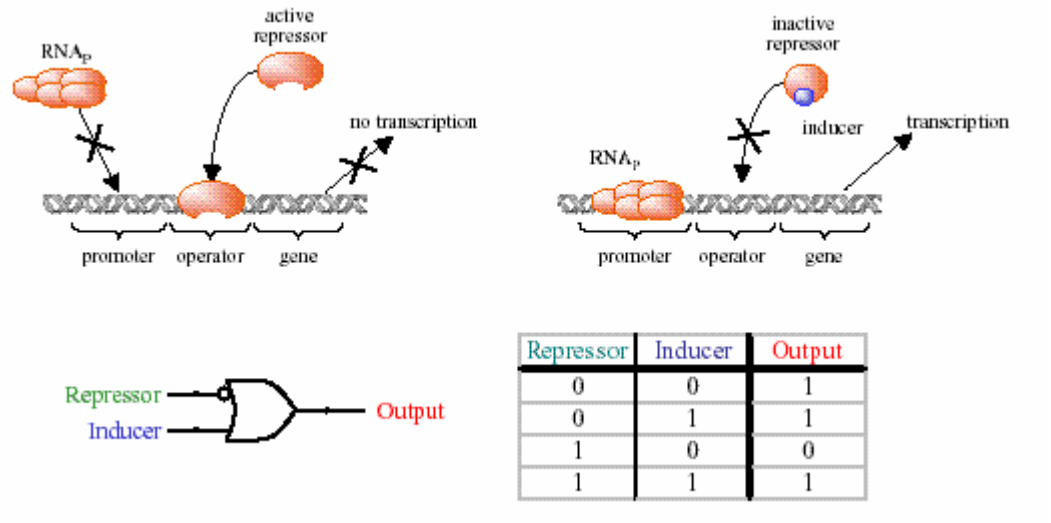
Figure 7.1 *A genetic logical inverter. From (Weiss 2003)*

7.2.1. The genetic inverter

The mechanism of the inverter is very simple. If the absence or presence of a repressor protein is regarded as an incoming 0-1-signal, and the mRNA is the outgoing, transcription will obviously occur when the incoming signal is "0", and *vice versa*. These are the characteristics of a logical inverter. See fig 7.1

7.2.2. The genetic IMPLIES gate

The IMPLIES gate, as most logical gates, requires two inputs and produces one output. The biochemical version works as follows: The input signals are represented by a repressor protein and an inducer, which comes from outside the cell. When the repressor is alone, it may as usual prevent the gene from being transcribed. However, if the repressor and the inducer are simultaneously present, the latter will bind to the former in such a way that the repressor, in turn, is unable to bind to the promoter. Hence, transcription will take place. See fig. 7.2

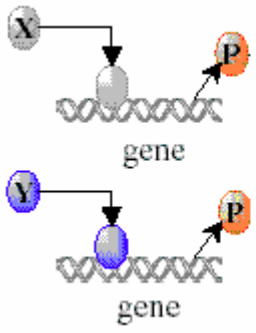


Repressor	Inducer	Output
0	0	1
0	1	1
1	0	0
1	1	1

Figure 7.2: *The genetic IMPLIES gate. From (Weiss 2003)*

7.2.3. A NAND gate

One example of a genetic NAND gate is shown in fig. 7.3. Two genes code for the same protein so the output will be "1" if either or both can transcribe, which is equivalent to both not being repressed simultaneously.

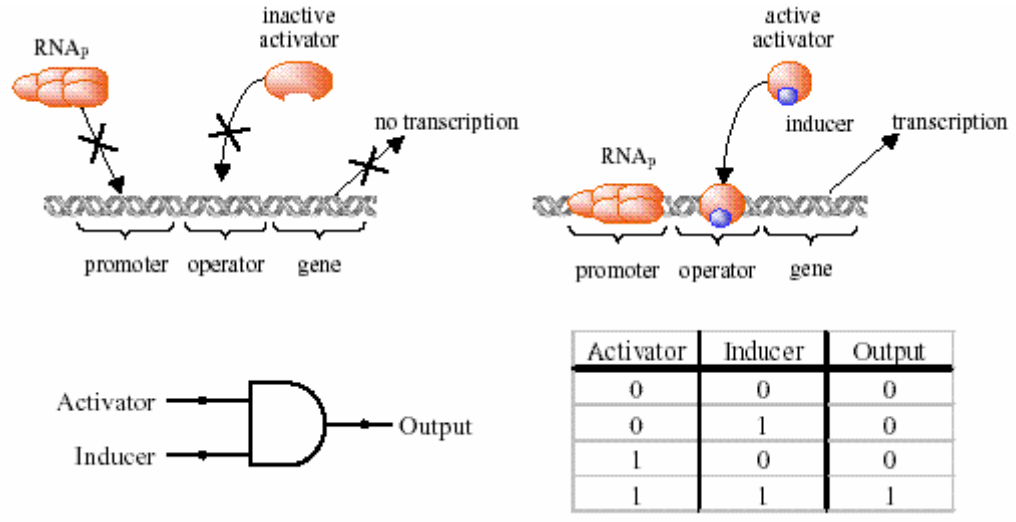


Repressor X	Repressor Y	Output
0	0	1
0	1	1
1	0	1
1	1	0

Figure 7.3: *A NAND gate consists of two genes that code for the same protein but are repressed by two different repressors. From (Weiss 2003)*

7.2.4 Two ways of making a genetic AND gate

As with the IMPLIES gate, a repressor and an inducer is involved in the AND gate. In contrast to the IMPLIES mechanism, in the AND gate the inducer activates the repressor. See fig. 7.4.



Activator	Inducer	Output
0	0	0
0	1	0
1	0	0
1	1	1

Figure 7.4: *The genetic AND gate. From (Weiss 2003)*

This AND gate is for extra cellular communication, as an inducer is involved. It's also possible to make an AND gate in another fashion. As is the case with electrical logic gates, several biochemical gates can be connected. For example, a NAND gate and an inverter also form an AND-gate, see fig. 7.5.

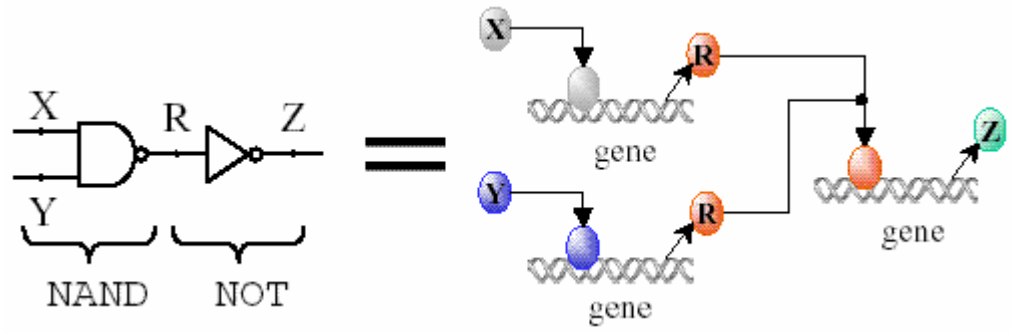


Figure 7.5: *The NAND and the inverter in series yield an AND-gate. From (Weiss 2003)*

7.3. Repressor mechanisms in looped DNA strands

As has been shown in the case of λ *phage*, there might be auxiliary as well as main repressor binding sites. Despite the fact that the auxiliary repressors may be far from each other and the main repressor in terms of linear distance, they may still be very close to each other due to the looping feature of DNA. Further, they interact in a manner that elimination of one such auxiliary binding site would reduce the repression level (which is a measure of repressor efficiency) tens or hundreds of times and produce fluctuations in protein synthesis. Vilar and Leibler examined this (Vilar 2003).
In the single site case, fig, 7.6a, transcription will switch relatively slowly between active and inactive modes, whereas in the looped case, fig 7.6b, where an auxiliary site is present, there are more thermodynamical states to choose among. It makes the biochemical process of switching faster.
Repression level in the looped case is non-linear when the number of repressors N is small but approaches constancy as it grows large, see fig 7.7. It means that the repression level is virtually independent of N when N is large. Consequently, the repression level does not change because of unwanted fluctuations in repressor concentration. The dynamics of the protein synthesis will be more stable than in the single site case. Fig 7.8. Looping in DNA is thus a biochemical way of protecting the protein synthesis from disturbances.
Looping thus has two benefits: It renders protein synthesis *faster* and *more stable.*

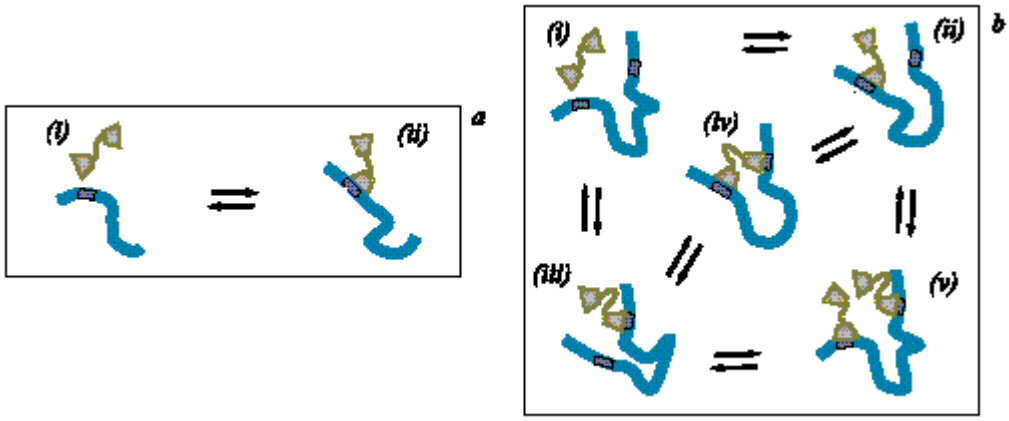


Figure 7.6: *a) Single site repressor binding. The gene may alter between repression and transcription. Switching is relatively slow. b) Looping in DNA may give rise to several thermodynamic states. State (iv) is in the middle of the chain of states and it is only here that full repression is exercised. As there are more states to choose among in this case, the probability of switching from one state to another (i.e. turning the.g.ene on and off) is higher, giving rise to faster switching. From (Vilar 2003)*

7.4. Experiments to be conducted within this area of research

It would prove interesting to be able to monitor the course of events in DNA looping and switching described here by use of the single molecular spectroscopy techniques described in other chapters. For this purpose, it is necessary to probe not only the DNA but also the regulatory proteins involved.

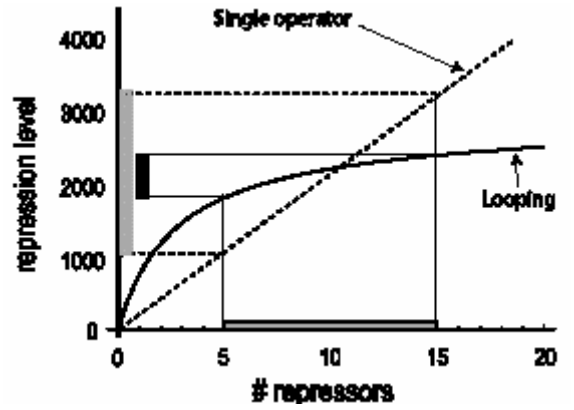


Figure 7.7: *Single operator repression level is linearly dependent on the number of repressors* N *present in the vicinity, making the protein synthesis fluctuate with* N. *With looping, on the other hand, repression level approach constancy for large N, making protein synthesis insensitive to random fluctuations in repressor concentration. From (Vilar 2003)*

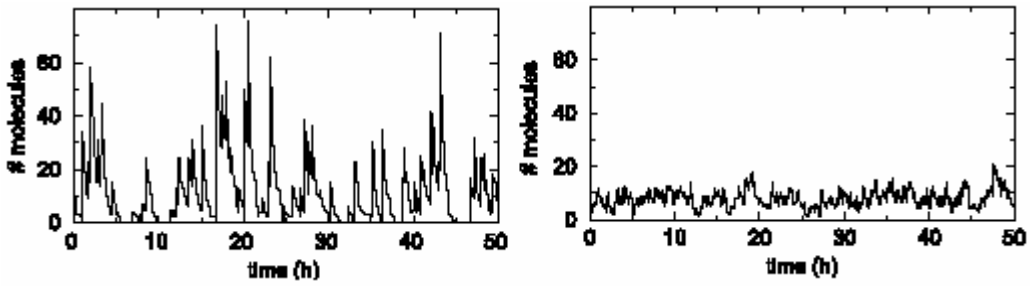


Figure 7.8: *Time series of protein synthesis for genes with one repressor site (left) and looped repression (right). As can clearly be seen, the possibility of looped repression gives a much more stable protein output. From (Vilar 2003)*

8. Conclusions and summary

It has been shown in this report that DNA nanostructures are suitable as scaffolds for proteins, FRET- and SERS active molecules, quantum dots, carbon nanotubes *etc*. A long-term goal for future research must be to synthesise scaffolding and self-assembly to a larger extent than has been done so far. Fig. 8.1. Techniques developed in nanofluidics may prove essential for this purpose.

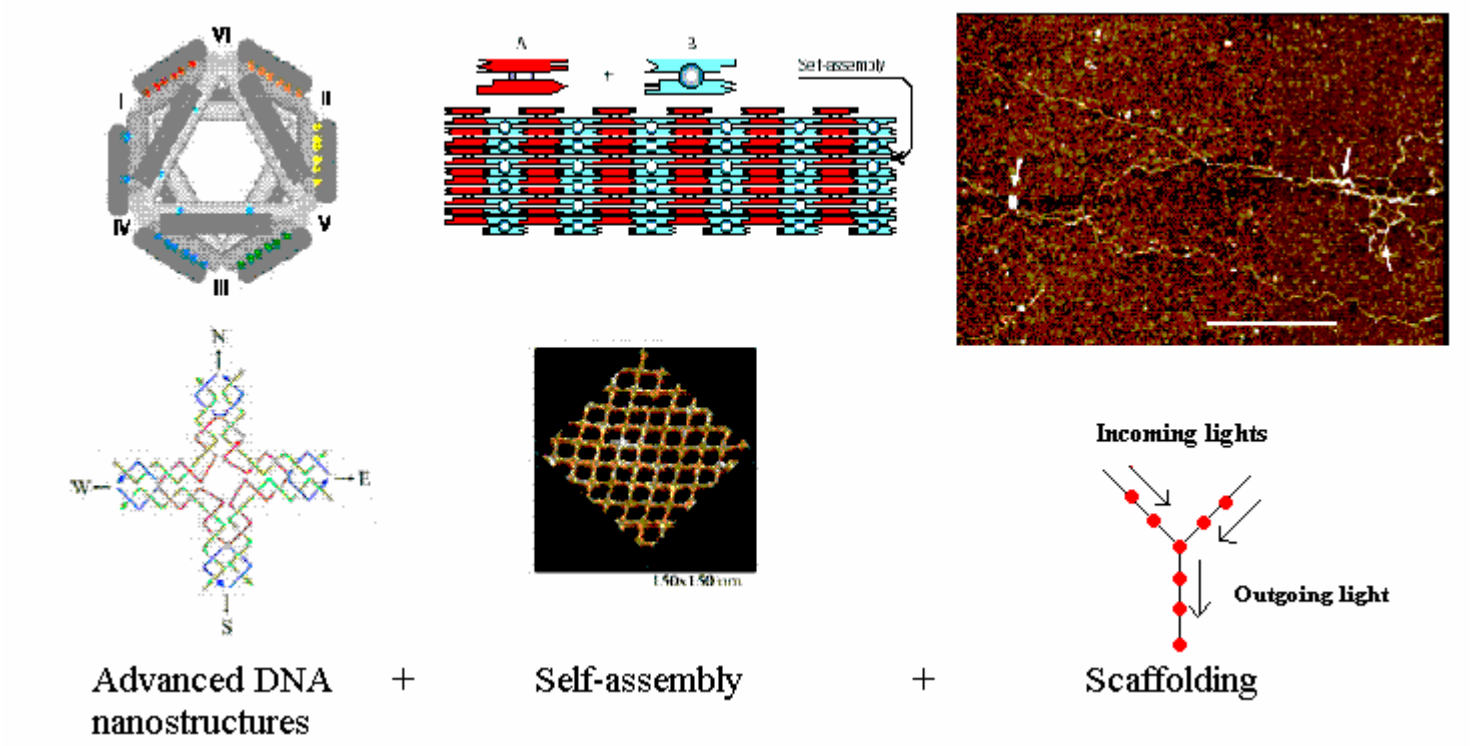


Fig 8.1. *DNA nanotechnology research should in the future focus on combining self-assembly and scaffolding properties.*

As far as the research field of DNA computing is concerned, this might not be a realistic way of accomplishing parallel computing, as it might be too callenging to detect the molecules representing the answers to the calculations. The DNA computing environment also prohibits an everyday use.
The fabrication of a biochip for fast DNA sequencing is the ultimate goal for the other field of research presented here. In order to realize it, single molecular spectroscopy methods using fluorescent, SERS, or FRET molecular probes must be integrated to work in an environment where DNA is stretched through nanochannelling techniques.
It might be possible in the future to combine DNA self assembly and DNA computing to such that the assembly itself is regulated through a DNA computing circuit.
A short overview of DNA switching mechanisms has also been presented.

Appendix: An overview of different research groups of relevance to this report

***Mirkin (Northwestern University, Chicago)*:**

SERS-based detectors, surface-plasmon based detectors.

Web page: http://www.chem.northwestern.edu/~mkngrp/

Landegren (Uppsala University)

Padlock probes, FRET-based probes, single molecule detection by biomolecular methods.

Web pages http://www.moltools.org/default.asp ,
http://www.genpat.uu.se/Forskargrupper/molme/molme.htm

Seeman (New York University, New York City)

DNA self-assembly and scaffolding.

Web page: http://seemanlab4.chem.nyu.edu/homepage.html

Winfree (Caltech, Pasadena)

DNA computing through algorithmic assembly.

Web page: http://www.dna.caltech.edu/~winfree/

Weiss (Princeton University)

Genetic networks, genetic switches.

Web page: http://swiss.csail.mit.edu/~rweiss/

Yan (Duke University, Raleigh)

DNA nanostructures, DNA computation

Web page: http://www.cs.duke.edu/~hy1/

References

Adleman, L. (1994). "Molecular Computation of Solutions to Combinatorial Problems." Science **266**: 1021.

Anderson, M. D. (2004). The University of Texas MD Anderson Cancer Center. **2004-05-03**.

Antson, D. (2001). Genotyping RNA and DNA Using Padlock Probes. Doctoral Thesis. Faculty of Medicine, Uppsala University.

Brown, K. R., et al. (2001). "Quantum Computing with Quantum Dots on Quantum Linear Supports."

Campion, A. (1998). "Surface-enhanced Raman scattering." Chemical Society Reviews **27**: 241.

Cao, H., et al. (2002a). "Fabrication of 10 nm enclosed nanofluidic channels." Applied Physics Letters **81**: 174.

Cao, H., et al. (2002b). "Gradient nanostructures for interfacing microfluidics and nanofluidics." Applied Physics Letters **81**: 3058.

Cao, Y. (2002c). "Nanoparticles with Raman Spectroscopic Fingerprints for DNA and RNA Detection." Science **297**: 1536.

Chakravarti, A. (2001). "...to a future of genetic medicine." Nature **409**: 822.

Chou, C.-F. (2002). "Electrodeless Dielectrophoresis of Single- and Double-Stranded DNA." Biophysical Journal **83**: 2170.

Deniz, A., et al. (1999). "Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations." Proc. Natl. Acad. Sci. USA **96**: 3670.

Eurogentec (2004). Peptide Nucleic Acid (PNA), http://www.eurogentec.be/module/FileLib/oligo_16.pdf.

Guha, R. (2003). Using Quantum Dots in Fluorescence Resonance Energy Transfer Studies, Pennsylvania State University.

Heilemann, M., et al. (2004). "Multistep Energy Transfer in Single Molecule Phontonic Wires." J. Am. Chem. Soc. **126**: 6514.

Huang, L R.,et al. (2002). "A DNA prism for high-speed continuous fractionation of large DNA molecules." Nat. Biotechnol. **10**: 1048

Irie, M., et al. (2002). "A digital fluorescent molecular photoswitch." Nature **420**: 759.

Isola, N. R., et al. (1998). "Surface-Enhanced Raman Gene Probe for HIV Detection." Anal. Chem. **70**: 1352.

Kaiser (2004). Raman Tutorial, http://www.kosi.com/raman/resources/tutorial/. **2004-08-23**.

Kramer, F. R. (2004). Introduction to Molecular Beacons, http://www.molecular-beacons.org. **2004**.

Landegren, U. (2004). "Molecular tools for a molecular medicine: analyzing genes, transcripts and proteins using padlock and proximity probes." Journal of Molecular Recognition **17**: 194.

Mao, R., et al. (2000). "Logical computation using algorithmic self.assembly of DNA triple-crossover molecules." Nature **407**: 493.

Mirkin, C. A., et al. (1996). "A DNA-based method for rationally assembling nanoparticles into macroscopic materials." Nature **382**: 607.

Nielsen, P. E., et al. (1991). "Sequence-Selective Recognition of DNA by Strand Displacement with a Thymine-Substituted Polyamide." Science **254**: 1497.

Odom, D., et al. (2000). "Robust charge transport in DNA double crossover assemblies." Chemistry & Biology **7**: 475.

Ptashne, M. (1992). A Genetic Switch: Phage Lambda, Blackwell Publishers.

Rydén (1999). Markovprocesser. Lund, KFS.

Seeman, N. C. (1982). "Nucleic Acid Junctions and Lattices." Journal of Theoretical Biology **99**: 237.

Seeman, N. C. (1991). "The use of branched DNA for nanoscale fabrication." Nanotechnology **2**: 149.

Seeman, N. C. (2003). "DNA in a material world." Nature **421**: 427.

Shih, W., et al. (2004). "A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron." Nature **427**: 618.

Stoneking, M., et al. (2001). "From the evolutionary past..." Nature **409**: 821.

Tegenfeldt, J. O., et al. (2004a). "The dynamics of genomic-length DNA molecules in 100-nm channels." PNAS **101**: 10979.

Tegenfeldt, J. O., et al. (2004b). "Micro- and nanofluidics for DNA analysis." Anal Bioanal Chem **378**: 1678.

Terazono, Y., et al. (2004). "Photonic Control of Photoinduced Electron Transfer via Switching of Redox Potentials in a Photochromic Moiety." J. Phys. Chem. B **108**: 1812.

Tsien, R. (1998). "The Green Fluorescent Protein." Annual Review of Biochemistry **67**: 509.

Vilar, J., et al. (2003). "DNA Looping and Physical Constraints on Transcription Regulation." J. Mol. Biol. **331**: 981.

Vo-Dinh, T., et al. (1999). "Surface-enhanced Raman Scattering (SERS) Method and Instrumentation for Genomics and Biomedical Analysis." J. Raman Spectrosc. **30**: 785.

Watson, J., Crick, F. (1953). "A structure for Deoxyribose Nucleic Acid." Nature **171**: 737.

Weiss, R. (2003). "Genetic circuit building blocks for cellular computation, communications, and signal processing." Natural Computing **2**: 47.

Williams, K., et al. (2002). "Carbon nanotubes with DNA recognition." Nature **420**: 761.

Winfree, E. (1998). "Design and self-assembly of two-dimensional DNA crystals." Nature **394**: 539.

Winfree, E. (2003). Algorithmic Self-Assembly of DNA. Doctoral Thesis. Pasadena, California Institute of Technology.

Wooley, K. R., et al. (2003). "DNA-Templated Construction of Copper Nanowires." Nano Letters **3**: 359.

Yan, H., et al. (2003). "DNA-Templated Self-Assembly of Protein Arrays and Highly Conductive Nanowires." Science **301**: 1882.

Zhang, Y., et al. (2002). "Insulating behaviour of λ-DNA on the micron scale." PRL **89:** 198102.