# 3. Modern Methods

3.1 The polymerase chain reaction

The polymerase chain reaction (PCR) is one of the most important methods developed in molecular biology in the last 15 years. Today, PCR is one of the most fundamental methods without modern molecular biology would be impossible. The central theme is the interaction of polymerases with DNA. Unfortunately, the question of how this interaction depends on conditions and sequence of the corresponding DNA is up to today not entirely clear. Finding the right reaction conditions for an efficient PCR is therefore still a task of trial and error.
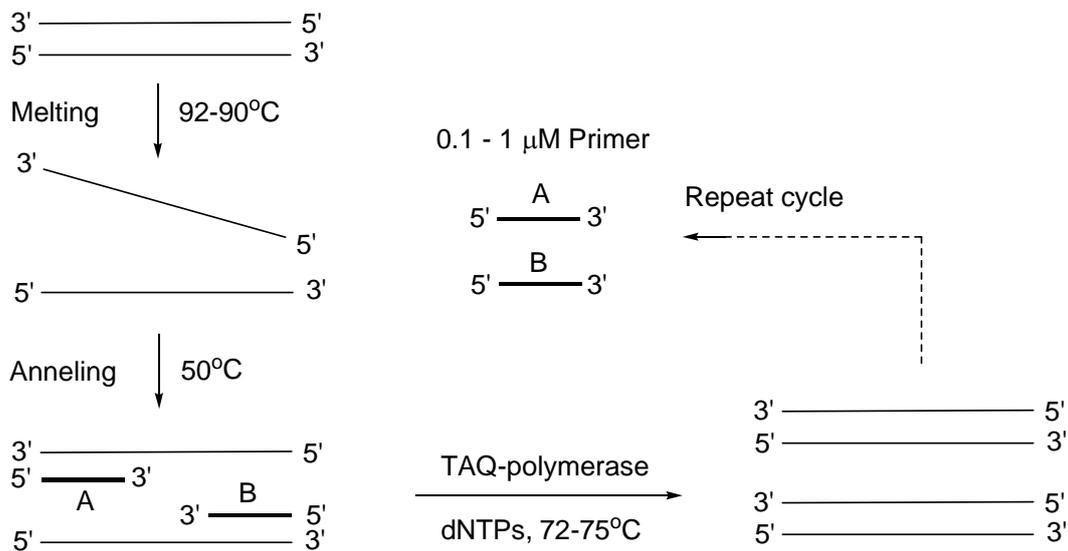
A typical PCR experiments start with heating of the sample in order to denature the DNA double strand to about 90°C. The reaction is slowly cooled in the presence of a large excess of primers, which bind to the complementary end-sequences of the target DNA. The large primer concentration ensures formation of the primer-DNA complexes. The reaction is then heated to about 70°C which is the optimal working temperature for a polymerase from a thermophilic organism (TAQ-polymerase). Now, the polymerase elongates the primer strand in the presence of all four triphosphates (dNTPs). The result is that the amount of DNA present in the sample has been doubled. The whole PCR cycle is then started from the beginning. In total one

obtains an exponential growth of the DNA sample over many PCR cycles. The replicated DNA is called the replicon.

PCR has many faces and can be used to answer a variety of important questions. One advantage of PCR is that only DNA strands are replicated which have the primer strands bound. One can therefore introduce a mixture of DNA strands into the reaction. By choosing of the sequence of the primer strands one can direct the amplification to just one DNA sequence in solution. So one can amplify DNA out of a mixture of DNA. The basis for the experiment is the selective amplification of DNA being complementary to the primer.

The amplification is possible to a final concentration of about 10 - 100 nM. Then decomposition of the polymerase will start due to the cycling at rather large temperatures of 90°C. During the experiment also the primer concentration is dropping making formation of the primer DNA complex more difficult. The concentration of the amplicon is strongly increasing making it in addition more difficult to denature the double strand.

PCR is today used for:
1. Amplification of very small amounts of DNA (1 molecule is sufficient). This is important for forensic analysis.
2. In order to amplify we have to know only a small fraction of the DNA sequence allowing us to amplify similar but not identical genes from different organisms.
3. Analysis of DNA e.g. to identify species. The PCR cyclers are aslo cheap and light so that analysis can be performed everywhere.
4. Genexpression studies can be performed using cDNA from organisms. One can use reverse transcription and PCR in the same experiment. Some polymerases have even reverse transcriptase activity.
5. If one determines the number of amplification steps (copy number) one can calculate the amount of DNA present in the sample.

3.2 PCR and the process of *in vitro random selection*

The design of biomolecules with defined structures and functions is an unreached goal. Today, we understand protein folding only incompletely. Also how catalysis is achieved is not jet fully understood. Synthetic enzyme mimics or designed proteins mimics are still inefficient catalysts in comparison to natural enzymes. Instead of rational design, more and more evolutionary approaches are used to create biomolecules with wanted functions. These methods allow preparation of large variety (library) of proteins or nucleic acids from which those with desirable function are selected and amplified. In such experiments, the molecules compete with each other and only those, which succeed will be amplified. The selection of the best molecules can be introduced into another round of selection and selective amplification so that at the end the molecule with the best properties will be obtained. In every round the selection pressure can be enhanced. Typically the process of selection and amplification is repeated up to 10 times. Between each round, the pool of molecules can also be mutated to increase again the diversity of the molecules present. So each cycle contains: 1. Mutation, 2. Selection, and 3. Amplification. Using this method it is possible to create proteins, DNA and RNA molecules with desired functions.

Since the discovery that RNA molecules fold into three dimensional objects, which can have catalytic properties the selection of RNA and DNA molecules with specific molecular recognition properties or catalytic capabilities has become a prime focus of research. These molecule are called ribozymes or DNAzymes. They can, like antibodies, bind specifically to small molecules and also larger proteins and they can bind to transition states and therefore stabilize them, which induces catalysis. DNA and RNA receptors able to bind e.g. cellular proteins with high specificity are called aptamers. If they are used inside a cell to bind a specific protein in a cellular environment they are termed intramers. The process of finding these molecules is called SELEX standing for: *systematic evolution of ligands by experimental enrichment*.

This process will be now explained in more detail. The Fig. below details the process. The first step involved preparation of the library of compounds. To this end a DNA synthesizer will be programmed to pump, starting from a certain nucleotide not a

single phosphoramidite but a mixture of all four amidites. In this way one obtaines DNA with constant end sequences but a randomized sequence somewhere in the middel. The number of molecules (X) created after N random coupling steps will be $X = 4^N$.

The first step requires the synthesis of the library. To this end the synthesizer will be programmed to couple not a single defined base but a mixture of all four bases. This randomized sequence is prepared in the middle to two defined sequences needed for primer recognition. Also, we do not want to go with the number of possible sequences beyond the number of molecules in our flask. In such a case we do not increase the diversity any more but create holes in sequence space.

━━━━━━━━━ ──────────────── ━━━━━━━━━

━━━━━━━━━   Known sequence

─────────   Randomized sequence NNNNNNNN
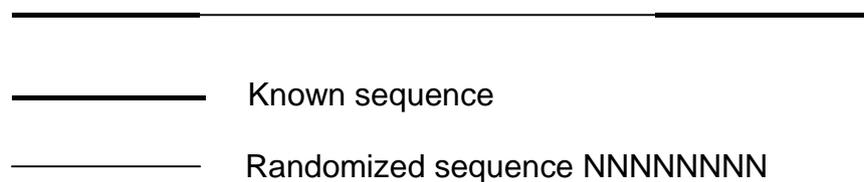
What will be obtained is a mixture of $10^{15} - 10^{16}$ individual molecules. This mixture can be put onto an affinity column, which is a column filled with a material onto which e.g. a guest molecule is bound. Mostly one uses either a polymer or hydrophobic silica gel as the matrix onto which the molecule or target that one wants to bind is bound. Those individual molecules, which have some affinity to the column and most likely to the compound bound to the surface will be retarded. These molecules are eluted e.g. by increasing the salt concentration of the eluent. The fraction is collected and the containing molecules are amplified using PCR. One obtains a pool of compounds in which binding RNA and DNA molecules are enriched. This pool is now again put onto the affinity column and the whole process of selection and amplification is repeated. The selection critera can be increased e.g. by eluting with a more shallow salt gradient and collection of only the compound that come off the column at rather high salt conditions. This will elute only the most powerful binding molecules.

If the precedure is done with RNA, please remember that RNA can not be introduced into the PCR. Before each amplication step we have to copy the RNA into DNA using

a reverse transcriptase (RT). Then, the DNA will be amplified. The DNA needs to be transcribed again into RNA just before the selection. All remaining DNA can be digested using a DNAse. At the very end, the finally selected RNA will be again copied into DNA and the DNA will be sequenced.
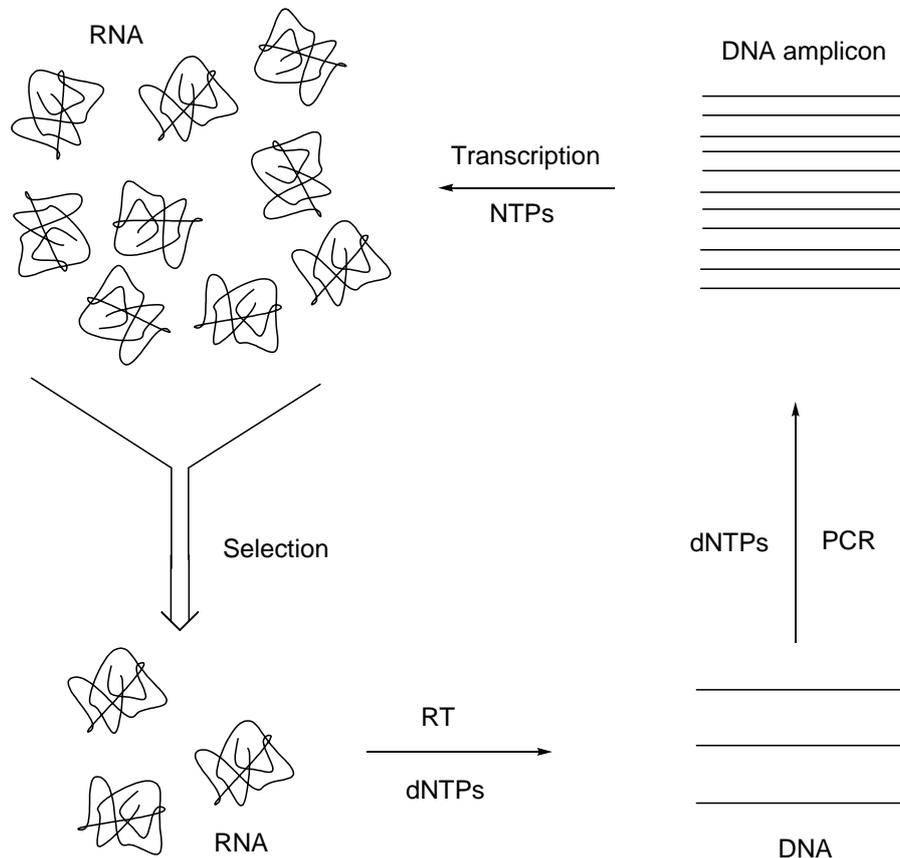
This method allows us to select DNA and RNA molecules binding to small molecules. How can we select molecules with catalytic properties. This requires a different selection protocol. One example of such a protocol is depicted below. This scheme allows selection of molecules with self-modifying properties. E.g. only molecules will be selected which are able to ligate themself to the DNA or RNA strand present in solution. Only those, which are linked are able to bind to an affinity column loaded with a DNA strand complementary to the strand present in solution. The selected RNA or DNA molecules have at the end the property that they are able to ligate themselves to a DNA or DNA strand. In consequence, they are able to catalyze formation of a phosphodiester bond.

RNA-library with self-ligating molcules

Only self-ligated molecules will be retarded on the column
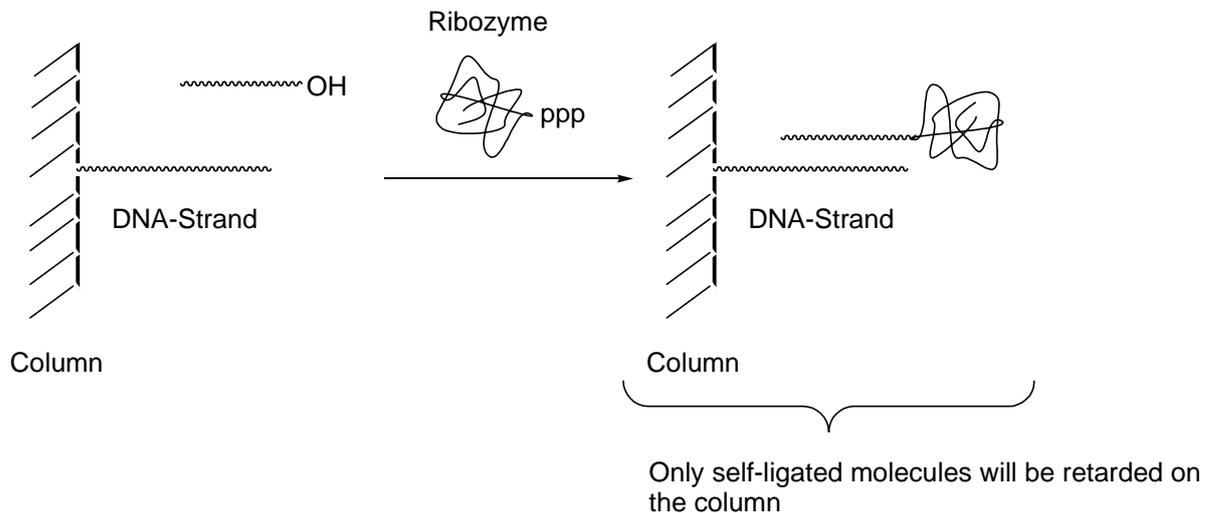
In the described example one was able to select active RNA molecules already after 3 rounds of selection and amplification. In the example the selection pressure was increased in every selection round by a reduction of the ligation time.

A special trick to find highly selective binders is to randomize each selected pool again. This is possible if the PCR reaction is modified so that the polymerase makes mistakes when it copies the DNA strand The error rate can be tunes to be between one error per 1 to every 20 base pairs.

Why is his additional mutagenic PCR so important? The complete randomization of a 220mer oligonucleotide would give $4^{220} = 10^{132}$ individual molecules. This number can not be reached because one would need tons of RNA to ensure that every molecules is present at least once in the mixture. Typically, one has enough material to generate a library of about 10^{15} molecules. Out of this pool we select the most active once. We can screen in consequence only a small fraction of the sequence space. In order to cover selection out of a larger pool we have to post-randomize the selected molecules. In this way we can fine tune and we can strongly increase the possibility to get really a strong binder. The pool is evolving in an increasing sequence room.

These types of experiments are particularly important tin respect to questions addressing the origin of life. Today it is believed that life started with RNA able to store genetic information and to catalyze biologically important reactions needed to develop life. The SELEX method us now allows us to create such molecules between genotype and phenotype and to further test this hypothesis.
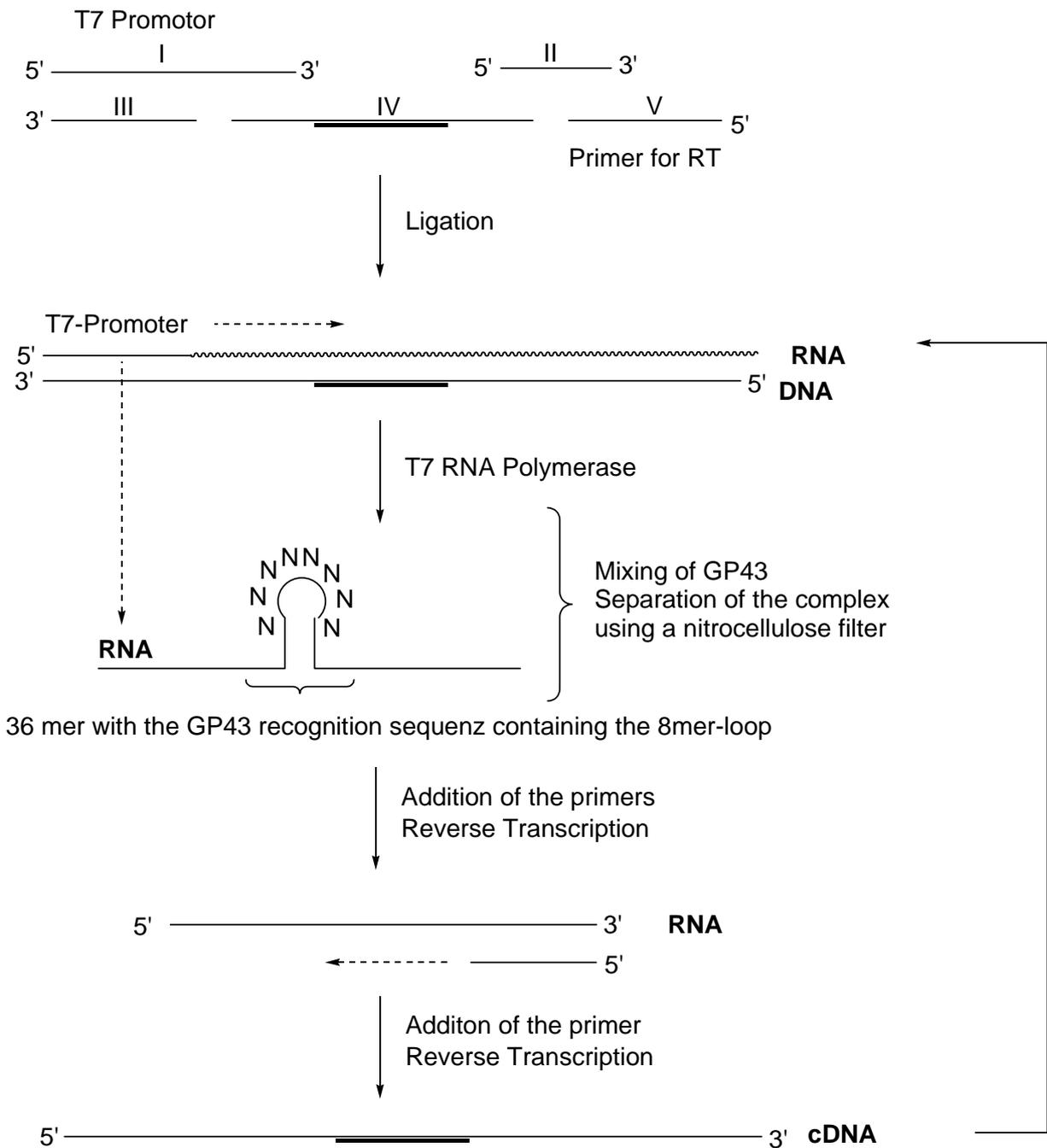
3.2.1 A more complex SELEX example

A protein was discovered with the name GP43. This protein is a T4-DNA polymerase, which recognizes a *m*-RNA which exist in a hairpin structure. The binding of the protein interupts the translation of the mRNA at the ribosome. However, the binding of GP43 to the hairpin mRNA, which is typically a 36mer RNA with a 8 nucleotide containing loop is relatively weak. It was now the plan to create another RNA motif that is better recognized by GP43. The process is by the way an autoregulation. If the polymerase binds to its own mRNA it will inhibit its own production at the ribosome.

First, a 110mer oligonucleotide was constructed out of three different fragments (III, IV und V) by ligation. The fragment IV contains a fully randomized 8mer sequence with the 36mer binding region recognized by GP43. Sequence V is the primer for reverse transcription. The obtained library is in summary rather small with 4^8 = 65'536 molecules. The oligonucleotide I needed for the ligation carries the T7-promotor. After complete ligation, with the help of oligo 1 the T7 RNA polymerase is used to prepare a piece of RNA with 92 oligonucleotides. This is now the required RNA library. This library was incubated with the protein GP43. The bound RNAs are separated by filtration through nitrocellulose, where only the complex will be bound. The rest elutes through the filter. The retarded RNA is collected and reverse transcribed into cDNA. To this end one has to add oligonucleotide V which functions as the primer for the reverse transcriptase. Reverse transcription is followed by amplification using PCR. Then the amplified cDNA is transcribed into RNA and the selection is repeated. After only 4 rounds of selection and amplification, a new RNA motif was selected with a large K_d = $5x10^{-9}$. The sequence of this RNA strand was obtained after reverse transcription and sequencing. The result:

Wildtyp: 5'-CAAG-<u>AGCCU**AAUAACUC**GGGCU</u>-AUAAACUAAGGAAU-3'

Variant: 5'-CAAG-<u>AGCCU**AGCAACCU**GGGCU</u>-AUAAACUAAGGAAU-3'

T7 Promotor

5' ———— I ———— 3'   5' —— II —— 3'

3' ———— III ————   ————— IV —————   ———— V ———— 5'

Primer for RT

Ligation

T7-Promoter   ------- →

5' ————————〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜〜—— **RNA**
3' ————————————————————————— 5' **DNA**

T7 RNA Polymerase

N N N N
N   N
N   N
N   N

RNA ————————————————

Mixing of GP43
Separation of the complex
using a nitrocellulose filter

36 mer with the GP43 recognition sequenz containing the 8mer-loop

Addition of the primers
Reverse Transcription

5' ———————————————— 3' **RNA**

←- - - - - - ———— 5'

Additon of the primer
Reverse Transcription

5' ————————————————————— 3' **cDNA**

The experiment

3.3 Aptamers and intramers

RNA and also DNA can fold like proteins into defined three dimensional structures.
The picture below shows NMR structures of some aptamers, which bind small

organic molecules and which were selected using the SELEX approach. All these aptamers bind their guests with very high selectivity. The molecules are: citrulline, B: FMN, flavinmononucleotid and C: neomycin B.

Today, we are able to select DNA and RNA molecules able to bind any small molecule. Because one obtaines not a single molecules but a variety of binding molecules, by comparing their binding motifs one get a deep insight into the prerequisites for efficient binding. If we look at the citrulline binding aptmer and we compare this aptamer with others that bind citrulline one can determine the conserved nucleobases, which are indicated in the picture with capital letters. Variable positions are labeled with small letters. Additional mutation studies can help us to decipher which residues are primarily important for the binding event. Particularly one can find out which residues change the binding selectivity from arginine to citrulline. These critical selectivity determining residues are indicated with with circles. These RNA aptamers behave consequently similar to proteins. Another smilarity is that most aptamers use an induced fit mechanism for binding. One can also learn that the aptamers can be rather small molecules. This and the ability for high selectivity has triggered interest to use aptamers for diagnostic applications.

Currently the selection of aptamers which bind a specific protein is very popular. Particularly when the aptamer inhibits the function of the protein. The aptamer can than be extended by sequences that direct the aptamer to certain points in the cell or which regulate the degradation of the aptamer. The cDNA (gene) for the aptamer can

even be inserted into the cell so that the cell produces the aptamer itself. Then the active aptamer is present in the cell and can regulate or disable certain proteins. This may be exploited as a tool for proteomics research. Such intracellular working aptamers are called intramers.

Another major interest in aptames come from application in analytics. Since aptamers bind their targets via an induced fit like mechanism, a fluorophore introduced into an aptamer might change its fluorescence upon binding of the analyte. Such aptamers might shine up when the analyte is bound. Another approach involves synthesis of aptamers with ligand induced (switchable) catalytic properties. The group of A. D. Ellington developed recently as an example an ATP binding aptamer (JACS, 2000, 122, 2469-2473). This aptamers is small enough to allow chemical synthesis and so they prepared a number of variants containing fluorophores in different positions. In fact one of the created molecules showed the desired fluorescence increase upon binding (APTR-R-Ac13). The Figure below shows the structure of this aptamer and indicates the position at which the fluorophore was placed to give a good signal. The fluorophore used was an acridine molecule (APTR-R-Ac13).

This experiment is another proof that binding of ATP changes the structure of the aptamer due to an induced fit. Acridine was the fluorophore of choice because this chromophore reacts sensitive to environmental changes. It is essential that the fluorophore does not interfere with binding but just reports the binding event
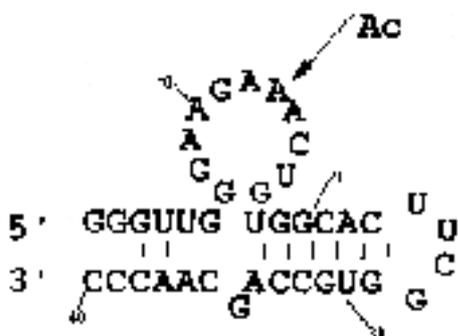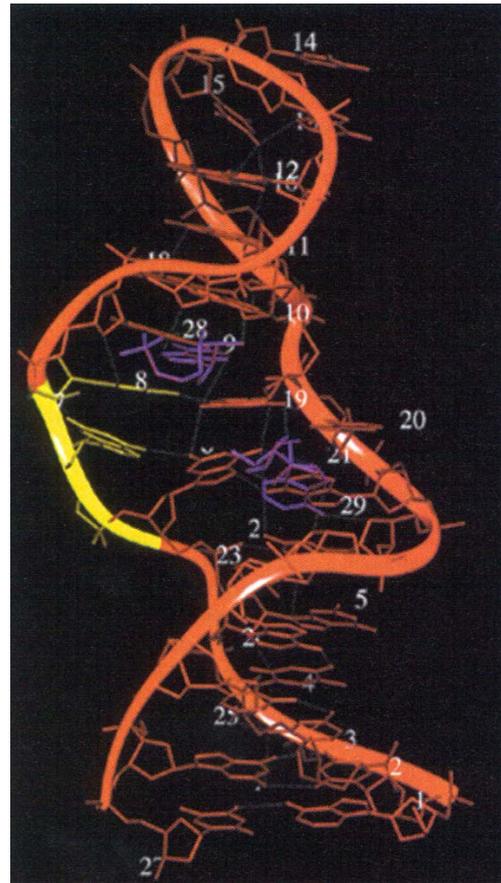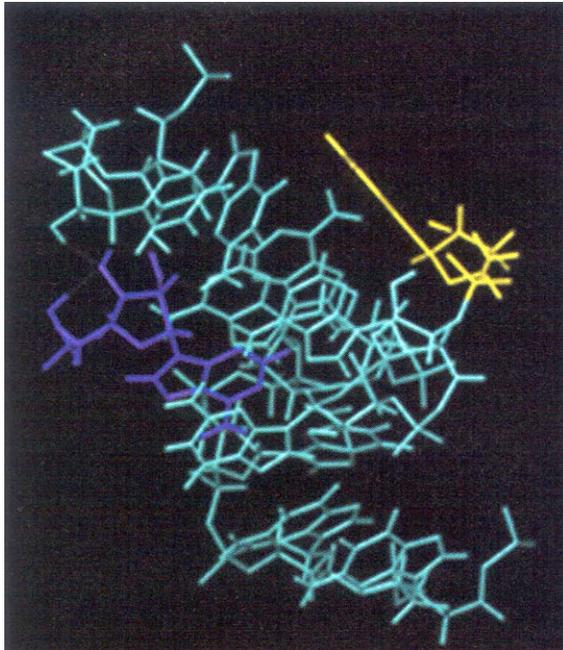
The RNA modified at position 13 with an acridine gives after ligand binding a strongly enhanced fluorescence signal of about 80%. GTP does not result in any fluorescence increase showing the selectivity of the aptamer. The fluorescence increase depends as expected on the concentration of adenosine triphosphate (ATP) in solution. (*J. Am. Chem. Soc.* **2000**, *122*, 2469).

3.4 Aptamers and ribozymes

Natural ribozymes are RNA molecules, which are able to catalyze the cleavage of phosphodiester bonds. Most of the time it is an intramolecular cleavage that leads to the processing of mRNA. This reaction has no turn over and ribozymes are in this respect (lack of turn over) no enzymes. A well know example to such a ribozyme is the hammerhead-ribozyme. Here the 2'OH-group of one ribose sugar attacks the phosphodiester group by first forming a penta-coordinated intermediate which then decomposes to the cleaved product.

Hammerhead-ribozymes are in general relatively small Y-type RNA-molecules in which some of the vases are involved in non-standard H-bond base pairing. The

Figure below shows a hammerhead ribozyme. Indicated are the bases essential for catalysis. The lines represent H-bonds.
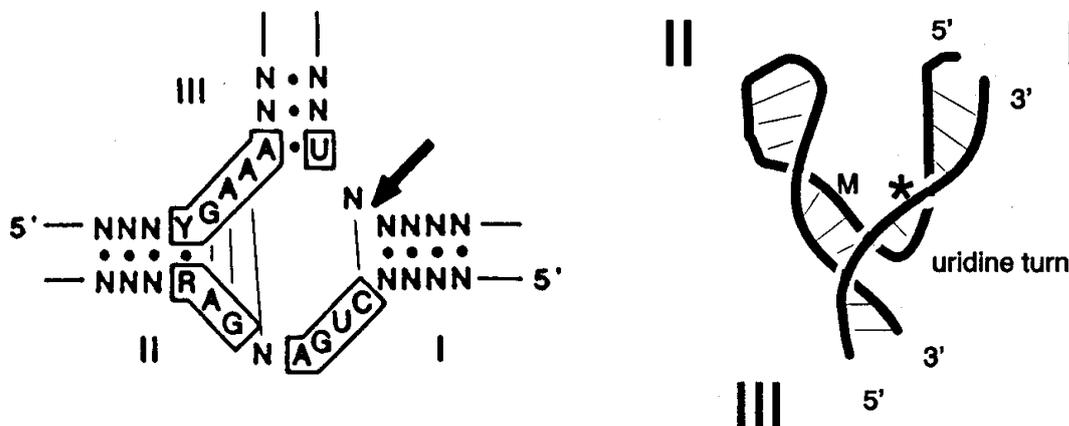
Relatively new developments are combining the recognition of a small molecule with a switching of the catalytic properties. Ribozymes are currently selected which become catalytically active when a small molecule binds and induces a conformational change. This regulation is called allosteric regulation. One major future goal is the development of intracellulary switchable intramers, which like intelligent molecules sence the cell status and become active if the concentration of a certain analyte reaches a critical limit. This mechanism of riboswitches is in fact used by nature.

A few years ago it was by R. R. Breaker (Yale university) discovered that the mRNA of proteins involved in the biosynthesis of vitamins such as thiamine, or vitamin B12, recognizes the vitamin and changes its conformation so that the information can not longer be translated at the ribozyome. Natural riboswitches are therefore allosterically autoregulated mRNAs

One more analytical development:

Ribozymes may be valuable, which are able to bind certain metal ions to sense environmental pollution at biologically important metal concentrations. Interesting sensors are for Ca^{2+}, Pb^{2+}, Zn^{2+}

A nice example was published recently *J. Am. Chem. Soc.* **2000**, *122* (42) 10467.

a.

```
                           cleavage site    substrate strand 17DS
                               ↓                    ╱
      3'- G T A G A G A A G G rA T A T C A C T C A -5'
           │ │ │ │ │ │ │ │ │ •   │ │ │ │ │ │ │ │
      5'- C A T C T C T T C T     A T A G T G A G T -3'
                       C           A A
                    C ╲          C     G
                 A  G  ╲ G  T  c G      enzyme strand 17E
                  ╲ ╲C    ╱
                 A   ╲  G
                  G  C
```

b.

The DNAzyme is composed of two DNA strands as depicted above. The first contains a fluorophore such as the TAMRA molecule and a single RNA nucleotide. Here selective cleavage can occur. The second strand contains a fluorescence quencher. Typically DABCYL is used. Are both strands aligned, than the fluorescence of TAMRA is almost completely quenched (curve II Fig. above) due to the presence of the quencher. If the solution contains Pb^{2+}, the ribozyme develops cleaving activity at the RNA unit. The TAMRA containing strand is cleaved and released in response. Since TAMRA is released from the construct, it starts to fluoresce at 580 nm (curve III Fig. above). The particular metallo ribozyme recognizes Pb^{2+} with very high selectivity as shown below. The detection limit was determined to 10 nM as shown below too.

The first Fig. shows the selectivity of the ribozyme. Only Pb^{2+} is recognized. The second Fig. Shows the fluorescence response with increasing Pb^{2+} concentration.

13

3.4.1 Dideoxysequencing

In the above mentioned examples, we have to decipher at the end the sequences of the selected molecules in order to clarify the identity of the active molecules. Sequencing of DNA and RNA is today one of the most important methods in molecular biology. Today the sequences of a number of bacteria, archea, plants and even of the human being are known. Very important was the sequencing of the total genome of many model organisms that were studied in the past such as *Escherichia coli*, *Methanobakterium thermoautothrophicum* and the fly *Drosophila melanogaster*. The complete sequencing of the human genome is a mile stone in our understanding of human genetics.

Modern sequencing methods rely today exclusively on the use of polymerases and dideoxynucleotides.
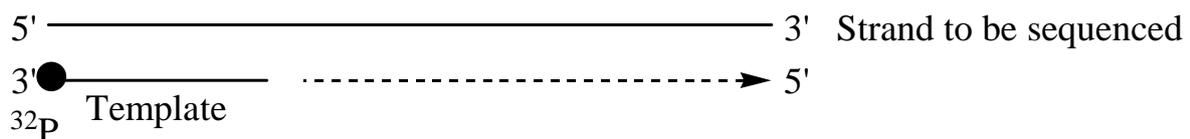
The strand that one wants to sequence is first amplified using PCR. The DNA is then distributed in four reaction vials (Eppendorf tubes). In the next step one adds into each reaction vial the template strand marked with a radioactive phosphate. One heats up and cools the mixture down to allow annealing. One adds in the next step into each vial a mixture of all four triphosphates dNTP. Finally, to each reaction vial is added a small amount of a 2', 3'-dideoxybuilding block shown below. Reaction vial 1 gets dideoxythymidine-5'-triphosphat, in reaction vial 2 one adds dideoxycytidine-5'-triphosphat, reaction vial 3 will get dideoxyadenosine-5'-triphosphat and to reaction vial 4 one will add dideoxyguanidine-5'-triphosphat. In each reaction vial, the polymerase will start to elongate the primer strand and mostly the full length replicon will be obtained. However, since each reaction vial contains a small amount of the dideoxynucleotides, this will be incorporated by the polymerase as well. Because this building block can not be elongated the polymerase reaction stops and one obtains a shorter fragment. Because the dideoxynucleotides are added only in rather small amounts, this stop of replication is a rare event. In each vial the polymerase will stop once in a while at every position, where it has a chance to incorporate the modified

14

base. In vial 1, the polymerase stops at all T incorporation positions. In vial 2 the stops will be at C incorporation positions and so forth. At the end of the reaction the DNA in all vials is separated by gel electrophoresis. One obtaines the so called oligonucleotide ladders.
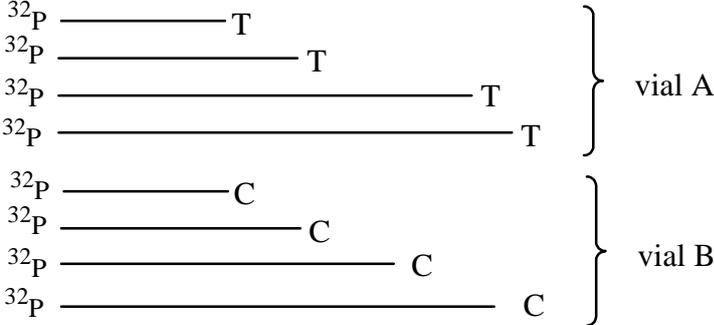

etc
Schematic representation of the shortened sequences that are obtained
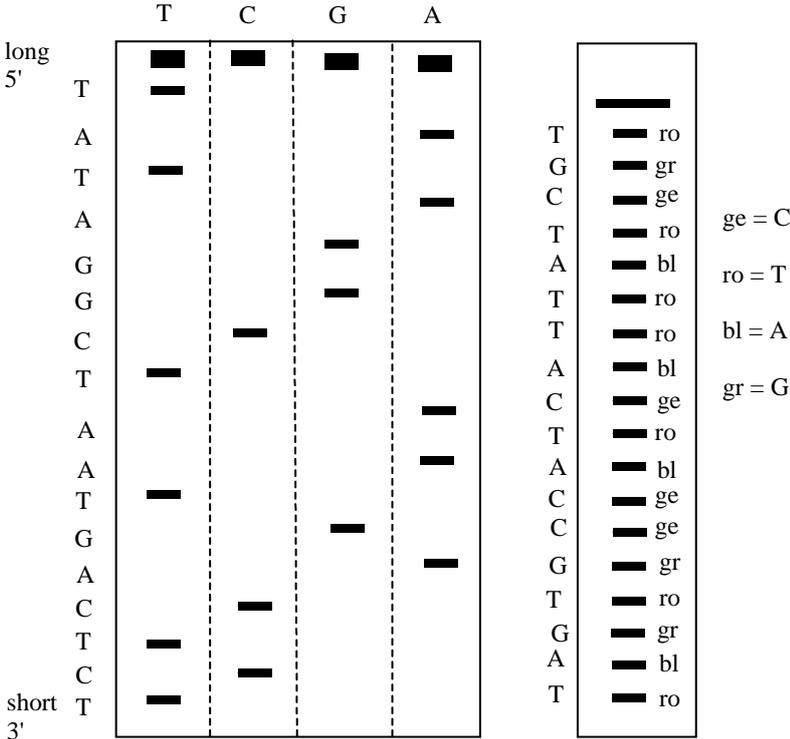
Schematic representation of sequencing gels. Left with ^{32}P labelling. Right with fluorescence labelling.

On the left side of the Figure above we have a schematic represenatation of the gel. Each lane represents the DNA found in one vial. At the top we find the long DNA strand. One can now go from the bottom to the top and can read the sequence.

Today, radioactive marking is not frequently in use any more. In addition it is time consuming to perform every sequencing in four different vials. Today the use of fluorescently labelled dideoxynucleotides circumvents both problems. Every dideoxynucleotide carries a different fluorophor. The position of the fluorophor attachment is choosen so that polymerases still accept the highly modified bases as substrates. One can perform the sequencing in one vial if fluorescence detection is used able to detect four different colors. Separation of the cutted sequences and multi color detection are today frequently performed by capillary electrophoresis.
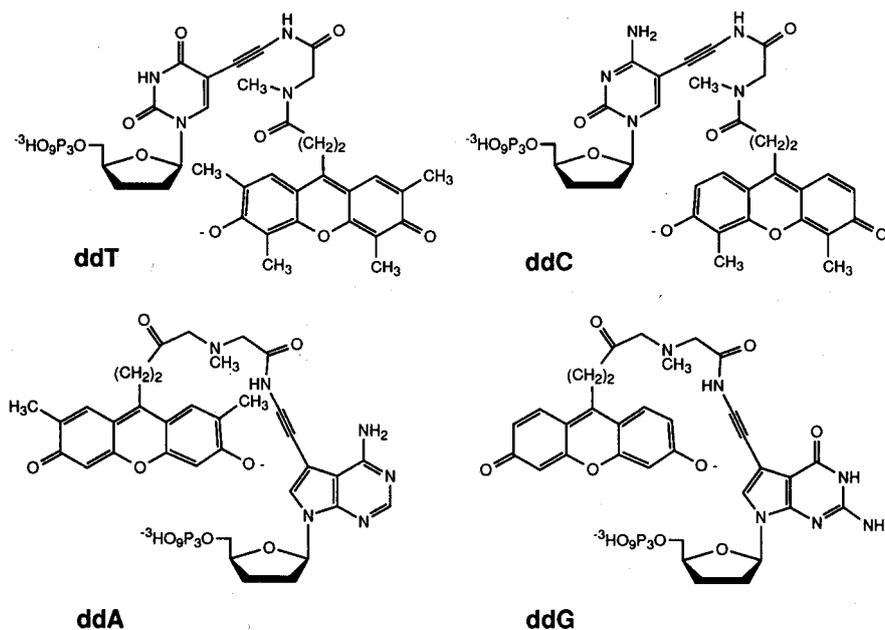
3.4.2 Fluorescence resonance energy transfer (FRET)

With the FRET method it is possible to measure distances in biomolecules. Also time resolved studied up to the single molecule level are possible allowing us to study dynamics of movements and association of proteins and other molecules to complexes. For FRET experiments we have to insert into the molecule or the complex which we wish to investigate two fluorophores. One is the so called donor

the other is the acceptor. The distances that can be measured are rather large with up to 100 Å but the accuracy is limited to about ± 5Å.

In a typical FRET experiment, we irradiated the sample at the wavelength, where the chromophore with the shortest wavelength absorption absorbs (e.g. 350 nm). This chromophore will deactivate by fluorescence. In the presence of a fluorescence acceptor however, the energy can be transferred to the acceptor which absorbs more in the red. If the acceptor is also a fluorophore, then this second acceptor will fluoresce. So one excited the first short wavelength fluorophore one but one obtains fluorescence light from the long wavelength acceptor fluorophore. This process is called donor fluorescence quenching by energy transfer. The fluorescence quenching is never 100%. Thats why one observes next to the acceptor fluorescence also a rest fluorescence of the donor. The relative fluorescence intensities between the donor and the acceptor fluorescence allows determination of the distance because the energy transfer rate depends on the distance. The efficiency of the energy transfer E_{FRET} is described by:

$$E_{FRET} = \frac{1}{\left[1 + \left(\frac{R}{R_0}\right)^6\right]}$$

R is the distance between the fluorophores. R_o is the Förster distance. It is the distance at which the fluorescence of the donor is quenched by 50%. The Förster radius is frequently determined experimentally but can also be calculated using the equation:

$$R_o^6 = 8.8 \times 10^{-23} \Phi_D \kappa^2 \eta^{-4} J(\lambda)$$

Φ_D is the fluorescence quantum yield of the donor.

κ is an orientation factor describing how the transition dipol moments of the two fluorophores are oriented. In the cases of free rotation of the chromophores it is $\kappa = 2/3$. In principle κ can take a value between 0 und 4.

η is the refraction index of the medium between the two chromophores. For water the value 1.33. Are the chromophores intercalated into DNA, this value increases to 1.75.

$J(\lambda)$ ist he spectral ovelap between the fluorescence of the donor and the absorbance of the acceptor. The energy transfer is very fast and efficient if the overlap is large. The Fig. below shows the FRET of the two chromophores fluorescein and Cyanine3.
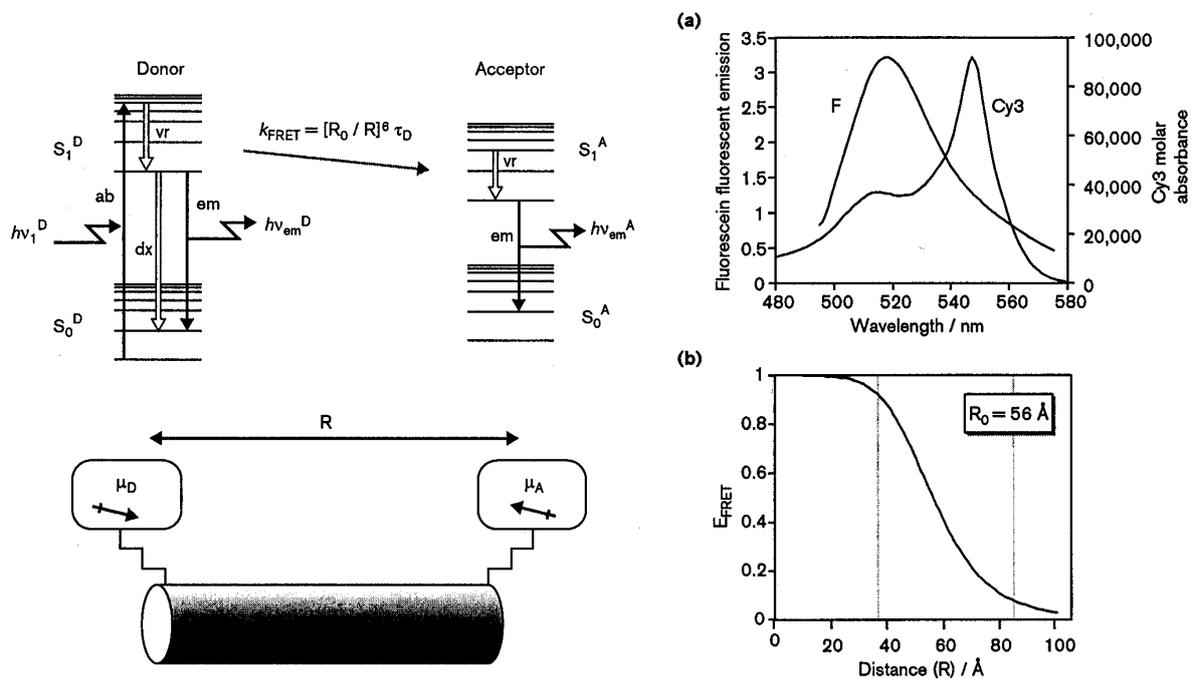
In general FRET experiments are done in a normal fluorescence spectrometer. One measures the steady state fluorescence of the donor and of the acceptor. More complex measurements, which, however, give also more information are life time measurements of the fluorescence. These life times can be directly transformed into rates and this in turn gives immediately the distance between the two chromophores if the fluorescence decrease is mono-exponentially.

For the pair fluorescin-cyanine3, the energy transfer efficiency was measured for various distances and an R_0-value of 56Å was obtained. If one needs a good measurement it is always possible to measure fluorescence changes between 0.1% and 99% Fluorescenzce. One can therefore use this FRET pair to measure distances between 35Å and 85Å. For other distances one needs to use other fluorophore pairs.

Most simple are steady state experiments in which a family of similar compounds are investigated. One is discussing in such set ups only the observed effect relative to other compounds and therefore avoids absolute measurements. These comparative measurements can be very accurate because other effects such as the influence of the environment on the fluorescence etc. are being cancel out. It is also important to

note that every set of fluorophores behaves differently. Some stick to biomolecule inhibiting free rotation. Others may intercalate e.g. into oligonucleotide with the same effect.

3.5 Recombination: An example of structural investigations with the help of FRET

Our genetic system is not a rigid polymer. It is in contrast a rather flexible unit. DNA exchange occurs frequently between different DNA strands and chromosomes. This is particularly important during meiosis, when the set of chromosomes is reduced to a haploid system. In certain phases of meiosis, genes are exchanged between the father and the mother chromosomes in order to mix up genes from parental and maternal origins. We also know today genes that hop in our genome as mobile elements. Information present on one chromosome can be transferred to the other chromosome. This is important, e.g. in case of double strand break. Then one chromosome has lost the information. It is essential to repair this loss with the help of the second chromosome. The processs of gene exchange between chromosomes is called homologous recombination.

Genetic material is exchanges between chromosome via Holliday-Junctions (Robin Holliday 1964). One of the possibilities for creating a Holliday junction is shown below. First, one of the double strands is cut. The obtained single strands then perfoms a strand invasion into the double strand containing a homologous gene sequence. This process requires in cells a number of proteins of which the key player in *E. coli* seems to be RecA (humans have functional homologs) . After formation of the heteroduplexes, the cutted DNA strands are ligated. In this way a stable cross over point is formed which can travel along the DNA (again a aprotein is needed for the translocation of the junction). This process is called branch-migration. Subsequent rearrangements at the crossing point, cutting and ligation resolves the Holliday junction afterwards. The result is an exchange of genetic material between chromosomes. Of particular interest over the last years was the question of the structure of the Holliday junctions. This could be investigated using the FRET method. To this end, DNA strands were prepared which contain fluorophores at their ends.
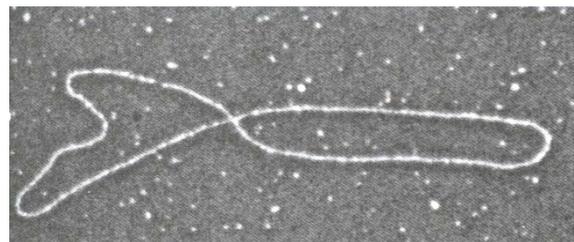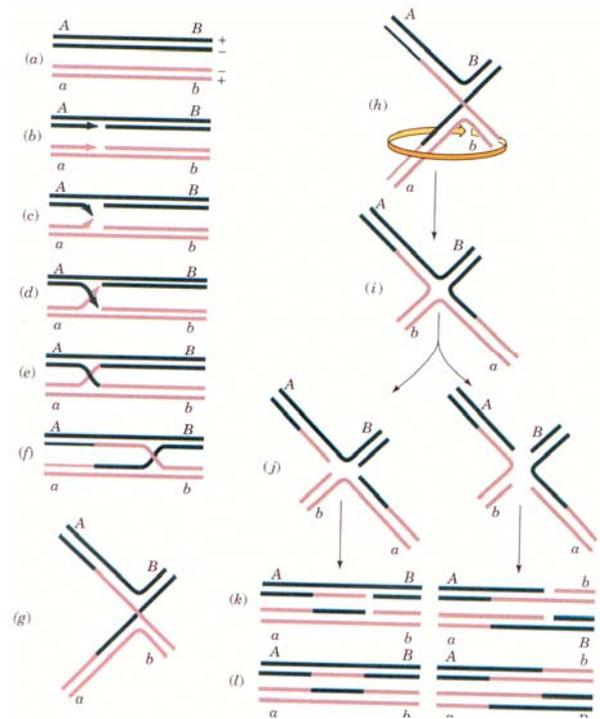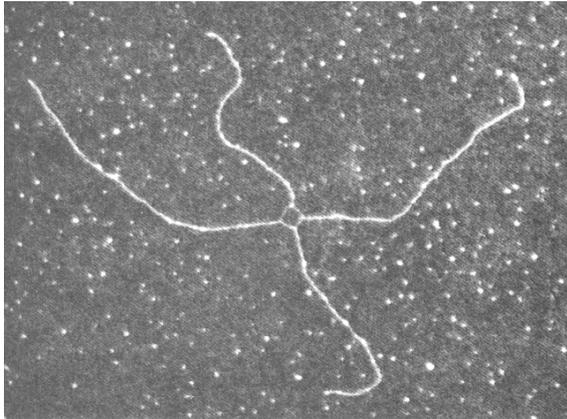
As described above, one of the major proteins involved in the process is RecA (352 amino acids). RecA is the *E.coli* enzymes but homologs have been found in yeast and humans as well. RecA binds single stranded DNA and forms a filament, which is then bind double helical DNA. This strand invasion is accompanied with the hydrolysis of ATP (RecA is therefore an ATPase).

In order to investigate this process, oligonucleotides were prepared containing a fluorescein **F** energy donor and a hexachlorofluorescein acceptor **H** (Biochemistry **1998**, *37* (33), 11692):

5'-GCACCAGATTCAGCAATTAAGCTCTAAGCC-**F**-3'
3'-CGTGGTCTAAGTCGTTAATTCGAGATTCGG-**H**-5'

The fluorescence of the F-donor is quenched in the presence of the H-acceptor. The experiments then included loading of RecA onto a 50mer single stranded oligonucleotide to build the filament. This filament was added to the double strand solution. If strand invasion takes place, than the fluorescence should increase because **F** and **H** will become separated.


(a)
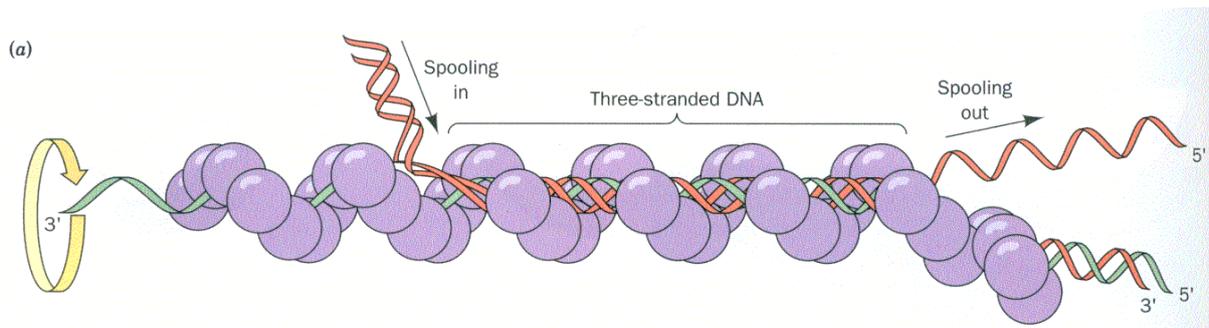
The recombination process is started after addition of cell extracts.

Figure A shows the emission spectrum of the donor F (dashed line, excitation at 492 nm) and the absorption spectrum of the acceptor H (solid line). Clearly visible is how well both band overlap which lets us expect that the energy transfer is very efficient. Figure B shows the emission spectrum of the double strand truncated, having just either the donor (dotted) or the acceptor (solid). The 30mer double strand having both chromophores gives the steady state spectrum depicted by the dashed line. Clearly visible is the reduced donor fluorescence and the enhanced acceptor fluorescence.

21
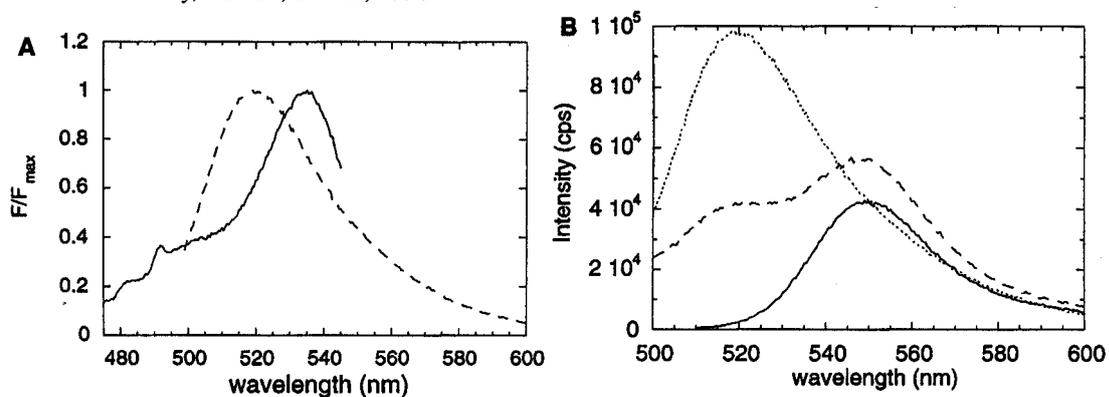
If one looks at the emission spectrum of the 30mer double strand alone and after addition of the filament one observes the dotted curve prior to mixing and the solid curve after mixing. The donor fluorescence clearly increases. This is the FRET effect. The figure below to right shows the fluorescence behaviour observed with increasing amounts of filament added. Clearly visible is that the fluorescence increases. Using this method one can observe the strand invasion process therefore in real time.
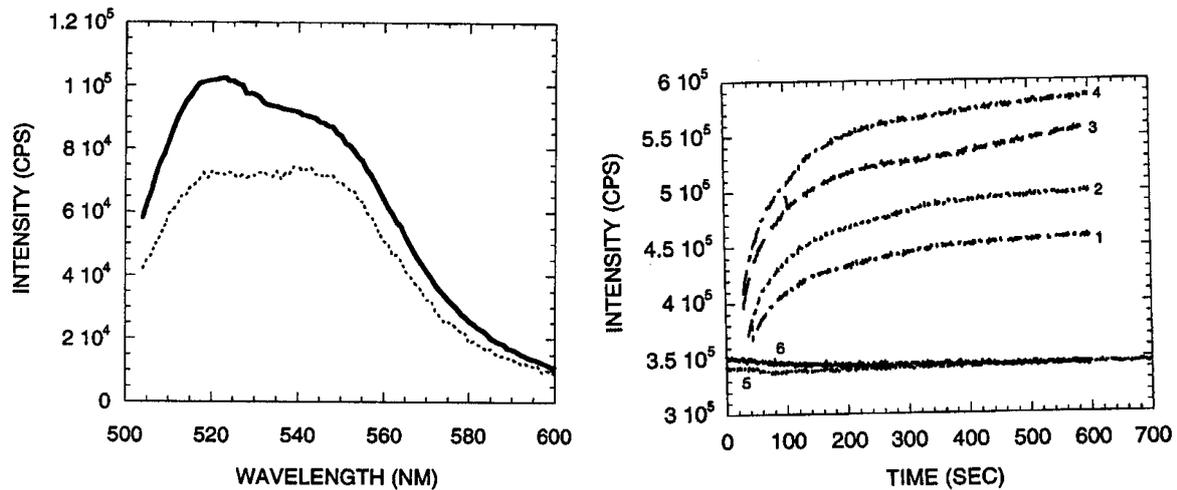

1 μM 30-mer, 50-mer Filament: 1: 0.1 μM, 2: 0.25 μM, 3: 0.3 μM and 4: 0.5 μM. The curves at the bottom visualize control experiments with poly(dT). λ_{ex}= 492 nm, λ_{em} = 520 nm.

3.6 Genomics with FRET probes

After the deciphering of the human genome it is increasingly important to develop methods that allow us to find small differences in the genome of two humans. Small genetic differences seem to be the reason, why certain cancer patients are rapidly cured by a specific chemotherpaie whereas others are not. So individual differences are responsible that medicines are not always active. These small genetic differences determine in addition the predisposition of certain individuels to certain deseases. Most genetic variants are just the replacement of one nucleotide in the genetic code by another. This single nucleotide polymorphisms (SNPs) need to be detected quickly and with high reliability. SNP detection is the beginning of a process that may lead to individualized medicine.

One of the best methods used for SNP detection is spectral genotyping, which again is based on FRET (*spectral genotyping*).

The method is based on the PCR amplification of the DNA of patients in the presence of fluorescent probes. The probes are small DNA molecules which are folded back on themselves. They have a head not involved in base pairing and a short stem structure which allows backfolding. DNA hairpins are rather small molecules, but for entropic reasons, they have rather high melting temperatures. The hairpins used to detect SNPs are called *molecular beacons*. They have at the end of the stems donor and acceptor fluorophores. In the closed conformation both are hold at a short distance, which causes quenching of the fluorescence of the donor. Very common is to use the FRET pair fluorescein or tetramethylrhodamin as the donor and Dabcyl as the acceptor. Since Dabcyl will function as an acceptor, which is not fluorescing but deactivates by electron transfer, the result of having this acceptor close to the donor will be fluorescence quenching. (An electron transfer is another possible explanation). These hairpins are added to the PCR solution. If during PCR a DNA structure is generated, which is fully complementary to the hairpin, including the head sequence, than the hairpin will open up and the fluorophore will start to fluoresc. The experimental conditions can be triggered so that even a single mismatch will inhibit this opening if the hairpin stem, although this is difficult. In any case under proper experimental conditions, single nucleotide polymorphisms can be detected.
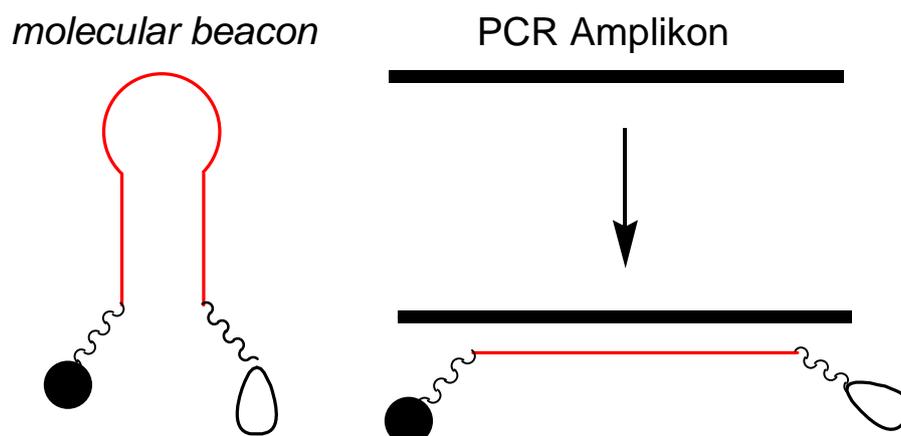
An alternative make up of the hairpins uses donor dyes which are quenched by guanine. The basis for the quenching by guanines is an electron transfer from the G-

base to the dye under formation of a G-radical cation and an acceptor radical anion. This electron transfer competes with the fluorescence. Many of the dyes used for DNA detection will exchange electrons with the guanine base. This is the basis for hole hopping studies in DNA.

molecular beacon

Again, the fluorescence quenching is only efficient, if the fluorescence donor is in close proximity to the guanines. Mostly they are stacking upon the guanine bases. As soon as the beacon opens up, the fluorophore will start to fluoresce.

Problematic is in all these analyses the background fluorescence. Even the fluorescence is almost completely quenched there will be always background and one has to measure the increase of an already present fluorescence signal. This is difficult if very small amounts of DNA are detected. One way out and good example how powerful FRET can be are the multiwavelength beacons.

24
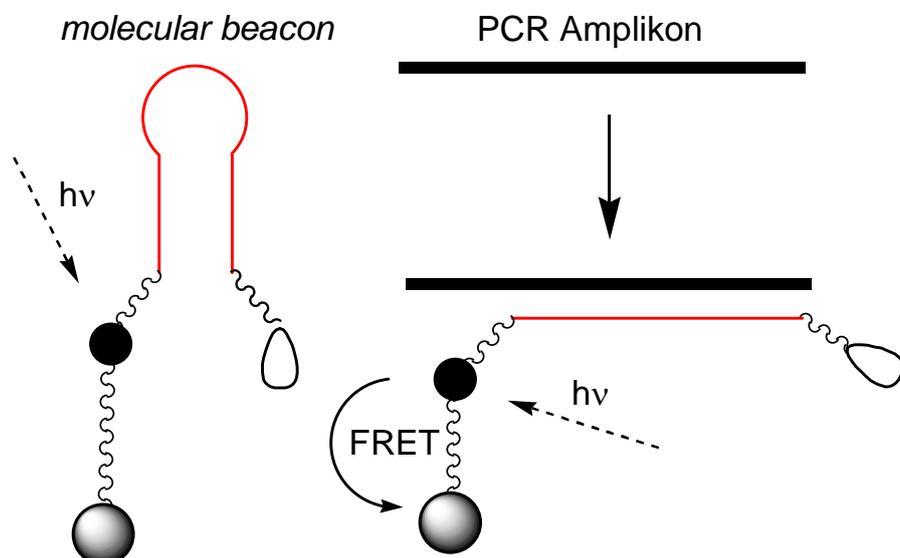
These hairpins are triple modified with fluorophores and quenchers. In the closed hairpin the fluorescence of the donor fluorophore (black ball) is quenched by the quencher. But even in the opened form, the donor is not fluorescing but is transfering the energy onto the fluorescence acceptor, which now fluoresces. Experiments have shown that such an arrangement can strongly increase the sensitivity of the beacon assay.

One expample of such a multiwavelength beacon uses Dabcyl as the quencher and fluorescein as the primary donor. The fluorescence emitter can be tetramethylrhodamine ($\lambda_{max,em}$ = 575 nm), or 6-carboxyrhodamine ($\lambda_{max,em}$ = 556 nm) or even texas red ($\lambda_{max,em}$ = 609 nm). The two fluorophores are generally separated by a few nucleotides. Optimal is a distance of about 7 Å. All the beacons are excited at the fluorescein absorption band which is 488 nm.

3.6.1 An example for genotyping

The human gene that codes for the chemokine receptor 5 contains in position 627 either a T or a C. Also position 630 can vary between a C or a T. This results if four different haplotypes --627--630--, which are –T—C--, --C—C--, --T—T— or –C—T--. Each human being has two chromosomes so that 10 different base combinations are possible in principle. The goal is to measure for a person the individual base combination. In order to do this, 4 allel discriminating *molecular beacons* are constructed. Each one can hybridise only with one of the four possible haplotypes. Each beacon has a different fluorescence emitter: Constructed were 2 normal beacons with fluorescein and tetrachlorofluorescein and two 2 FRET molecular beacons with fluorescein as the donor and either tetramethylrhodamine or texas red as the emitter fluorophore.

The DNA of various individuals is now amplified using PCR in the presence of the four beacons. Only one of the beacons will light up if the individual is homozygot. Two will shine up if the individual is one out of the six possible heterozygote combinations. So one multiplex assay allows us to decipher in detail the genetic composition of the individual

Genotyp	Fluorescein	Tetrachloro-fluorescein	Tetramethyl-rhodamin	Texas Rot	Number of human DNA samples matching this pattern
--T—C--/--T—C—	0.95	0	0.05	0	6
--C—T--/--C—T—	0	1.00	0	0	0
--C—C--/--C—C—	0	0	1.00	0	4
--T—T--/--T—T—	0	0	0	1.00	0
--T—C--/--C—T—	0.51	0.49	0	0	0
--T—C--/--C—C—	0.77	0	0.23	0	11
--T—C--/--T—T—	0.73	0	0	0.27	1
--C—T--/--C—C—	0	0.68	0.32	0	0
--C—T--/--T—T—	0	0.56	0	0.44	0
--C—C--/--T—T—	0	0	0.45	0.55	2

From Fred Russell Kramer *Nature Biotechnol*. **2000**, *18*, 1191-1196.

Further application of this technique involves the analysis of a 32 nucleotide long homozygot deletion of the β-chemokine receptor 5 gene (*CCR5*), which is responsible for some HIV resistances. The possibility to quickly analysis the *CCR5Δ32* allels enables even testing of large populations.

3.7 DNA arrays

It is for many biological investigation very desirable to analyse the protein status of a cell. Today the common approach is to separate all the proteins by 2D gel electrophoresis, to stain the proteins and to analyse the obtained spots by mass spectrometry. This give not only the molecular weight but also a fragmentation pattern can be obtained that goes towards mass spec sequencing. This approach called proteomics is still difficult and requires sophisticated mass spectrometry equipment. A way around is to analyze not the proteins but the mRNA content of the cell, although this may be only a rough approximation of the protein content. In any case, already a comparison of the mRNA status of a healthy cell with the status of a

cell featuring a disorder may be very informative. These difference allow in some cases to recognize genes which are up or down regulated.

The massively parallel analysis of the mRNA content of cells is today performed with DNA arrays (DNA chips), These are mostly glass plates which are devided in small segments. Each segment contains a specific oligonucleotide attached to the surface either covalently or by absorption. A typical chip is 2.5 x 2.5 cm large and may contain up to 50'000 segments with different oligonucleotide sequences.

The mRNA of the cells to be analysed is first reverse transcribed into cDNA. Then the DNA is amplified using PCR. The DANN is afterwards cut into smaller pieces using hydrolytic enzymes. A ligase such as the RNA ligase which attached unspecifically nucleotides to DNA and RNA is used to attach to each piece of DNA a fluorophore. The so prepared DNA is finally added onto the chip. If the fluorescence labelled oligonucleotide pieces find a complementary piece of DNA on the chip they will form the double strand. The conditions for these hybridizations are carefully controlled so that not matching oligonucleotides are not binding. Under certain conditions one can discriminate between a fully complementary oligonucleotide (binding) and one having a single mismatch (not binding). All the segment were matching oligonucleotides formed a double strand will fluoresce. This can be analyzed so that in a single experiment 50'000 DNA sequences can be checked.

The major problem is to prepare the chips and to deposite the oligonucleotides onto the chips. This is currently achieved in two different ways:

A) Synthesis of the 50'000 oligonucleotudes using a photolithographic approach directly on the chip (the known phosphoramidite DNA synthesis with light sensitive protecting groups is used)

B) Synthesis of the oligonucleotides in solution and printing of the DNA strand onto the surface using a modified ink jet printer.
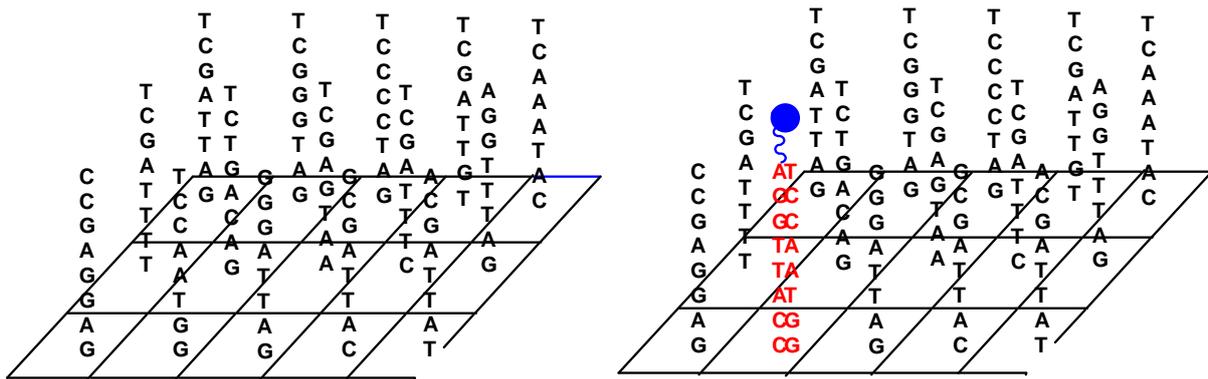
Schematic representation of the DNA array technology